

**Natural Selection on *S*-linked genes in *Turnera*
(Passifloraceae)**

Deanna L. Harris

A Thesis Submitted to the Faculty of Graduate Studies in Partial Fulfillment of the
Requirements for the Degree of Master of Science

Graduate Program in Biology
York University
Toronto, Ontario

August, 2015

© Deanna L. Harris, 2015

Abstract

Investigations of the evolutionary dynamics exhibited by *S*-linked loci have the potential to provide evidence concerning the particular genes that determine the expression of distyly in flowering plants. Several approaches were adopted to explore the signatures of selection on *S*-linked genes in distylous *Turnera*. While dN/dS-based results revealed pervasive purifying selection at the *S*-locus in *Turnera*, average nucleotide diversity (π) and sequence polymorphism (θ) measures were found to be elevated in two *S*-linked genes (*AP2D* and *RNABP*), suggesting the possible occurrence of balancing selection at these or closely-linked loci. Limited trans-species polymorphisms were identified in *APETALA2*, as well. Conversely, the negatively selected *S*-haplotype specific gene, *Tsstal*, also appears to be a very promising distyly gene candidate and shows significant sequence homology to known self-incompatibility proteins in *Papaver*. This study represents the first investigation of the molecular signatures of natural selection on *S*-linked genes in any heterostylous species. The implications of the results obtained for the elucidation of the genetic mechanisms that determine distyly in *Turnera* are discussed. Ultimately, it is hoped that understanding these mechanisms will, in turn, help to evaluate existing models regarding how distyly has evolved.

Dedication

For my Little Man, who liked plants too.

Acknowledgments

I owe no small measure of thanks to Dr. Joel Shore for agreeing to take me on as a graduate student and for allowing me to hang around his laboratory for the past several years. Without his supervision and encouragement, I necessarily could not have completed this work.

Thanks also to my graduate advisor, Dr. Amro Zayed, whose excellent insight and advice has certainly helped to shape the final product of my efforts. I only regret that, due to scheduling conflicts, we could not retain him for the final stages of this project. To Dr. Bridget Stutchbury, thanks ever so much for bravely stepping in at the very last moment to take his place.

To my lab-mate, Mr. Paul Chafe, I bestow much appreciation for both tolerating my presence and entertaining my many questions and queries.

To my family and friends, thanks for not always asking me why I am not “done yet” and for, instead, seeing my education as something that can only be advantageous to you in the impending zombie apocalypse and/or the future colonization of Mars.

I owe a special debt of gratitude to my wonderful mother, for allowing me to spend a large portion of my life pursuing higher education and, importantly, for not holding my apparent inability to settle on a particular profession against me. Without her absolutely unconditional love and support, I would not have had the freedom to explore those things that truly make me happiest.

And, lastly, I extend my utmost appreciation to my partner in crime of nearly a decade, Stefan Ferraro. Thanks for agreeing to endure with me the crazy schedules, constant stress, and near-permanent poverty that often accompanies graduate education. Thanks for making me work when I'd really rather not, and for telling me that I can do it when I am pretty sure that I can't. For all of your efforts – even the most subtle – I really cannot thank you enough.

TABLE OF CONTENTS

Abstract.....	ii
Dedication.....	iii
Acknowledgments.....	iv
Table of Contents.....	iv
List of Tables.....	viii
List of Figures.....	x
List of Abbreviations.....	xv
Units of Measure.....	xviii
1.0 INTRODUCTION.....	1
1.1 HETEROSTYLY.....	1
1.2 SELF-INCOMPATIBILITY (SI) SYSTEMS.....	8
1.3 GAMETOPHYTIC SI: THE S-RNase SYSTEM.....	9
1.4 GAMETOPHYTIC SI IN <i>PAPAVER</i>	11
1.5 HOMOMORPHIC SPOROPHYTIC SI IN THE BRASSICACEAE.....	13
1.6 HETEROMORPHIC SPOROPHYTIC SI.....	14
1.7 <i>PRIMULA</i> AND THE SUPERGENE MODEL.....	16
1.8 DISTYLY IN <i>FAGOPYRUM</i> AND <i>LINUM</i>	17
1.9 DISTYLY IN <i>TURNERA</i>	19
1.10 NATURAL SELECTION IN SELF/NON-SELF RECOGNITION AND SEX DETERMINATION SYSTEMS.....	25
1.11 OBJECTIVES.....	29
2.0 METHODS.....	31
2.1 PLANT MATERIAL AND DNA EXTRACTION.....	31
2.2 GENES OF INTEREST AND EXON PREDICTION: KNOWN S-LINKED GENES.....	33
2.3 CONTROL GENE SELECTION.....	38
2.4 PRIMER CONSTRUCTION.....	38
2.5 POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION.....	42
2.6 PURIFICATION OF PCR PRODUCTS FOR CLONING.....	43
2.7 CLONING, BACTERIAL CULTURE, PLATING, AND PLASMID PURIFICATION.....	43
2.8 SEQUENCING AND SEQUENCE ASSEMBLY.....	44
2.9 SEQUENCE ALIGNMENT.....	45
2.10 DESCRIPTIVE STATISTICS.....	46

2.11 PHYLOGENETIC ANALYSES AND GENE GENAOLOGIES	47
2.12 SELECTION ANALYSES	48
2.12.1 NUCLEOTIDE AND CODON SUBSTITUTION MODEL SELECTION	50
2.12.2 GLOBAL dN/dS ESTIMATES	51
2.12.3 COMPARING EVOLUTIONARY RATES.....	53
2.12.4 LOCAL dN/dS PROCEDURES	54
2.12.5 SITE-BY-SITE SELECTION ANALYSES	56
2.12.5.1 SINGLE-LIKELIHOOD ANCESTOR COUNTING (SLAC)	56
2.12.5.2 FIXED EFFECTS LIKELIHOOD (FEL)	57
2.12.5.3 MIXED EFFECTS MODEL OF EVOLUTION (MEME)	58
2.12.5.4 FAST UNCONSTRAINED BAYESIAN APPROXIMATION (FUBAR)	59
2.12.6 MACDONALD-KREITMAN TESTS (MKTs) AND TAJIMA'S D TESTS OF NEUTRALITY	60
3.0 RESULTS	62
3.1 SEQUENCING, ALIGNMENTS, AND DESCRIPTIVE STATISTICS	62
3.2 PHYLOGENETIC ANALYSES	64
3.2.1 THE TOTAL DATA SET: DNA- AND AMINO ACID-BASED PHYLOGENIES	64
3.2.2 THE TOTAL DATA SET WITH A REDUCED NUMBER OF TAXA: DNA- AND AMINO ACID-BASED PHYLOGENIES	66
3.2.3 THE <i>Tsstal</i> DATA SET WITH AN EXPANDED NUMBER OF TAXA: DNA- AND AMINO ACID-BASED PHYLOGENIES	67
3.2.4 GENE GENEALOGIES	68
3.3 SELECTION ANALYSES	70
3.3.1 GLOBAL dN/dS ESTIMATES	70
3.3.2 COMPARING dN/dS IN S-LINKED AND CONTROL GENES.....	71
3.3.3 LOCAL OR LINEAGE-SPECIFIC SELECTION ON S-LINKED GENES.....	73
3.3.4 dN/dS IN <i>Tsstal</i> AND <i>AP2D</i>	74
3.3.5 CODON-LEVEL SELECTION ON S-LINKED GENES	75
3.3.5.1 CODON-LEVEL SELECTION ON <i>APETALA2</i>	75
3.3.5.2 CODON-LEVEL SELECTION ON <i>Tsstal</i> (TOTAL DATA).....	76
3.3.5.3 CODON-LEVEL SELECTION ON <i>Tsstal</i> (WITH A REDUCED NUMBER OF TAXA)	77
3.3.5.4 CODON-LEVEL SELECTION ON <i>LEJ2</i>	77
3.3.5.5 CODON-LEVEL SELECTION ON <i>AP2D</i>	78

3.3.5.6 CODON-LEVEL SELECTION ON <i>RNABP</i>	79
3.3.5.7 CODON-LEVEL SELECTION ON <i>SCE1</i>	80
3.3.5.8 CODON-LEVEL SELECTION ON <i>FRA1</i>	80
3.3.5.9 CODON-LEVEL SELECTION ON <i>LRRK</i>	81
3.3.5.10 CODON-LEVEL SELECTION ON <i>IRX15L</i>	82
3.3.5.11 CODON-LEVEL SELECTION ON <i>FSP</i>	83
3.3.5.12 CODON-LEVEL SELECTION ON <i>NRFP</i>	83
3.3.5.13 CODON-LEVEL SELECTION ON <i>WRKY</i>	84
3.3.6 POPULATION-LEVEL ANALYSES: MACDONALD KREITMAN TESTS (MKTS) AND TAJIMA’S D TESTS OF NEUTRALITY.....	85
3.6 TABLES AND FIGURES.....	87
4.0 DISCUSSION.....	123
4.1 <i>S</i> -LOCUS GENES WITH ELEVATED NUCEOLTIDE DIVERSITY AND SEQUENCE POLYMORPHISM.....	124
4.2 GENE GENEALOGIES.....	127
4.3 SPECIES PHYLOGENIES.....	128
4.4 PERVASIVE PURIFYING SELECTION ACROSS <i>S</i> -LINKED GENES.....	130
4.5 <i>AP2D</i> : AN INTERESTING CANDIDATE GENE.....	131
4.6 <i>Tsstal</i>	134
4.7 CONCLUSIONS.....	136
LITERATURE CITED.....	139
APPENDICES.....	162
Appendix A: PCR and Sequencing Primers.....	162
Appendix B: Nucleotide Alignments for <i>S</i> -linked Genes.....	169
Appendix C: Nucleotide Alignments for Control Genes.....	212
Appendix D: Amino Acid-Based Phylogenies for Concatenated Total Data Alignments and Total <i>Tsstal</i> Data Alignment.....	233
Appendix E: DNA- and Amino Acid-Based Phylogenies for all Individual Genes and for Concatenated alignments of all <i>S</i> -linked and all Control Data.....	236
Appendix F: Supporting Likelihood Ratio Tests and Parameter Estimates.....	280
Appendix G: Adaptive Branch-Site Random Effects Likelihood (aBS-REL) Results.....	286
Appendix H: <i>Tsstal</i> in a Localized Population of <i>T. scabra</i> from the Dominican Republic (DROT)	289

List of Tables

Table 1: List of plants used in this study, along with their respective phenotypes, and ploidy levels	32
Table 2: List of <i>S</i> -linked genes of interest and their predicted functions.....	34
Table 3: List of control genes and their predicted functions.....	40
Table 4: Nucleotide and amino acid substitution models used for phylogenetic reconstruction in MEGA 6.06	49
Table 5: Nucleotide and codon substitution model selection for analyses in HYPHY.....	52
Table 6: Descriptive statistics for all alignments	88
Table 7: Proportions of polymorphic sites, synonymous and non-synonymous mutations, numbers of indel events, average indel length, and indel diversity (π_1) for all genes, obtained using a balanced data set	91
Table 8: Global dN/dS ratio estimates and likelihood ratio test of dN=dS.....	96
Table 9: Comparing of selection on <i>S</i> -linked and control genes using likelihood ratio tests.....	97
Table 10: Comparing of selection on <i>Tsst1</i> and other <i>S</i> -linked genes using likelihood ratio tests	98
Table 11: Comparing of selection on <i>AP2D</i> and other <i>S</i> -linked genes using likelihood ratio tests.....	99
Table 12: Comparison of global and local dN/dS ratio models for each nucleotide alignment.....	100
Table 13: Branch-site Unrestricted Statistical Test of Episodic Diversification (BUSTED)	101
Table 14: Integrative site-by-site selection analysis for <i>Apetala2</i>	102
Table 15: Integrative site-by-site selection analysis for <i>Tsst1</i> (total data).....	103
Table 16: Integrative site-by-site selection analysis for <i>Tsst1</i> (reduced number of taxa).....	104
Table 17: Integrative site-by-site selection analysis for <i>LEJ2</i>	105
Table 18: Integrative site-by-site selection analysis for <i>AP2D</i>	106
Table 18: Integrative site-by-site selection analysis for <i>RNABP</i>	108
Table 19: Integrative site-by-site selection analysis for <i>SCE1</i>	110
Table 20: Integrative site-by-site selection analysis for <i>FRA1</i>	111
Table 21: Integrative site-by-site selection analysis for <i>LRRK</i>	112
Table 22: Integrative site-by-site selection analysis for <i>IRX15L</i>	115
Table 23: Integrative site-by-site selection analysis for <i>FSP</i>	117
Table 24: Integrative site-by-site selection analysis for <i>NRFP</i>	118
Table 25: Integrative site-by-site selection analysis for <i>WRKY</i>	120
Table 27: McDonald-Kreitman tests and Tajima's D tests of neutrality for all <i>S</i> -linked genes of interest completed by treating diploid taxa from the <i>Turnera</i> subseries as a population sample	121

Table A1: PCR primer pairs used for the amplification of each exon(s) of interest	162
Table A2: Sequencing primers.....	166
Table F1: Parameter estimates for LRT test comparing selection on <i>S</i> -linked and control genes.....	280
Table F2: Comparison of selection on <i>S</i> -linked and control genes using LRTs and random starting values for all parameters	281
Table F3: Parameter estimates for LRTs comparing selection on <i>Tsstal</i> and all other <i>S</i> -linked genes ...	282
Table F4: Comparison of selection on <i>Tsstal</i> and other <i>S</i> -linked genes using LRTs and random starting values for all parameters	283
Table F5: Parameter estimates for LRTs comparing selection on <i>AP2D</i> and all other <i>S</i> -linked genes ...	284
Table F6: Comparison of selection on <i>AP2D</i> and other <i>S</i> -linked genes using LRTs and random starting values for all parameters	285
Table G1: Adaptive Branch-Site REL tests of Lineage-specific selection on <i>S</i> -linked genes (excepting <i>Tsstal</i>).....	286
Table G2: Adaptive Branch-Site REL tests of Lineage-specific selection on <i>Tsstal</i>	288
Table H1: The degree of <i>Tsstal</i> sequence similarity between samples from a local population of <i>T. scabra</i> (DROT) and <i>T. panamensis</i> (PAN 2S)	292

List of Figures

Figure 1: Graphical representation of distyly	3
Figure 2: Graphical representations of the recessive <i>s</i> - and dominant <i>S</i> -haplotypes, as they are known from BAC sequencing in diploid <i>T. subulata</i>	24
Figure 3: Graphical representation of trans-specific evolution.....	28
Figure 4: Gene diagrams for all <i>S</i> -linked genes based on exon predictions from NetPlantGene Server and BLAST searches	39
Figure 5: Assay of <i>Tsstal</i> in a variety of long- and short-styled individuals of the genus, <i>Turnera</i>	87
Figure 6: Nucleotide diversity (π) and Watterson's θ estimates for all alignments.....	89
Figure 7: Average nucleotide diversity (π) and Watterson's θ estimates for <i>S</i> -linked and control genes .	90
Figure 8: Molecular phylogenetic analysis by Maximum Likelihood methods using total DNA data for all genes	92
Figure 9: Molecular Phylogenetic analysis by Maximum Likelihood methods using total DNA data for all genes and a reduced number of taxa	93
Figure 10: Molecular Phylogenetic analysis by Maximum Likelihood methods using total DNA data for <i>Tsstal</i>	94
Figure 11: Gene genealogy for <i>APETALA2</i> showing evidence of trans-specific evolution in the <i>Turnera</i> subspecies clade	95
Figure B1: <i>APETALA 2</i> nucleotide alignment	169
Figure B2: <i>Tsstal</i> nucleotide alignment (total data).....	172
Figure B3: <i>Tsstal</i> nucleotide alignment with a reduced number of taxa.....	177
Figure B4: <i>LEJ2</i> nucleotide alignment	180
Figure B5: <i>AP2D</i> nucleotide alignment	181
Figure B6: <i>RNABP</i> nucleotide alignment	185
Figure B7: <i>SCE1</i> nucleotide alignment.....	187
Figure B8: <i>FRA1</i> nucleotide alignment	189
Figure B9: <i>LRRK</i> nucleotide alignment	192
Figure B10: <i>IRX15L</i> nucleotide alignment	198
Figure B11: <i>FSP</i> nucleotide alignment.....	202
Figure B12: <i>NRFP</i> nucleotide alignment.....	204
Figure B13: <i>WRKY</i> nucleotide alignment	208

Figure C1: <i>ECIP1</i> nucleotide alignment.....	212
Figure C2: <i>GAUT3</i> nucleotide alignment.....	216
Figure C3: <i>GAUT1</i> nucleotide alignment.....	220
Figure C4: <i>RNABP34</i> nucleotide alignment.....	223
Figure C5: <i>FMO1</i> nucleotide alignment.....	225
Figure C6: <i>MBD8</i> nucleotide alignment.....	227
Figure C7: <i>UNKN</i> nucleotide alignment.....	229
Figure C8: <i>POFUT</i> nucleotide alignment.....	231
Figure D1: Molecular Phylogenetic analysis by Maximum Likelihood method using total amino acid data for all genes.....	233
Figure D2: Molecular Phylogenetic analysis by Maximum Likelihood method using total amino acid data for all genes and a reduced number of taxa.....	234
Figure D3: Molecular Phylogenetic analysis by Maximum Likelihood method using total amino acid data for <i>Tsstal</i>	235
Figure E1: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>APETALA2</i>	236
Figure E2: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>APETALA2</i>	237
Figure E3: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>Tsstal</i>	238
Figure E4: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>Tsstal</i>	239
Figure E5: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>LEJ2</i>	240
Figure E6: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>LEJ2</i>	241
Figure E7: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>AP2D</i>	242
Figure E8: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>AP2D</i>	243
Figure E9: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>RNABP</i>	244

Figure E10: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>RNABP</i>	245
Figure E11: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>SCE1</i>	246
Figure E12: Molecular Phylogenetic analysis by Maximum Likelihood method using acid data for <i>SCE1</i>	247
Figure E13: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>FRA1</i>	248
Figure E14: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>FRA1</i>	249
Figure E15: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>LRRK</i>	250
Figure E16: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>LRRK</i>	251
Figure E17: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>IRX15L</i>	252
Figure E18: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>IRX15L</i>	253
Figure E19: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>FSP</i>	254
Figure E20: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>FSP</i>	255
Figure E21: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>NRFP</i>	256
Figure E22: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>NRFP</i>	257
Figure E23: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>WRKY</i>	258
Figure E24: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>WRKY</i>	259
Figure E25: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for all <i>S</i> -linked genes	260
Figure E26: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for all <i>S</i> -linked genes	261
Figure E27: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>ECIP1</i>	262

Figure E28: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>ECIP1</i>	263
Figure E29: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>GAUT3</i>	264
Figure E30: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>GAUT3</i>	265
Figure E31: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>GAUT1</i>	266
Figure E32: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>GAUT1</i>	267
Figure E33: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>RNABP34</i>	268
Figure E34: Molecular Phylogenetic analysis by Maximum Likelihood method using acid data for <i>RNABP34</i>	269
Figure E35: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>FMO1</i>	270
Figure E36: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>FMO1</i>	271
Figure E37: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>MBD8</i>	272
Figure E38: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>MBD8</i>	273
Figure E39: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>UNKN</i>	274
Figure E40: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>UNKN</i>	275
Figure E41: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for <i>POFUT</i>	276
Figure E42: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for <i>POFUT</i>	277
Figure E43: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for all control genes	278
Figure E44: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for all control genes	279

Figure H1: Alignment of *Tssta1* sequences from 17 short-styled individuals belonging to a localized population of *T. scabra* from the Dominican Republic (DROT)..... 289

List of Abbreviations

aBS-REL	Adaptive Branch-site Random Effects Likelihood
AIC	Akaike information criteria
AP2	Floral homeotic protein, <i>APETALA2</i>
AP2D	AP2 domain containing transcription factor family protein
ARC1	Arm repeat-containing protein 1
BAC	Bacterial artificial chromosome
BIC	Bayesian information criteria
BLAST	Basic Local Alignment Search Tool
BUSTED	Branch-site Unrestricted Statistical Test of Episodic Selection
CBSX1	Cystathionine beta-synthase domain containing protein
cds	Coding sequence
CI	Confidence Interval
CPREV	General Reversible Chloroplast nucleotide substitution model
csd	Complimentary sex determining locus in hymenopterans
CTAB	Cetyl-trimethylammonium bromide, (C ₁₆ H ₃₃)N(CH ₃) ₃ Br
DEAR2	DREB and EAR motif containing protein
DF	Degrees of freedom
ddH₂O	Double distilled water
dH₂O	Distilled water
dN	Nonsynonymous changes
DNA	Deoxyribonucleic acid
dNTP	Deoxyribose nucleotide triphosphate
DREB	Dehydration-responsive element binding
dS	Synonymous changes
DUF579	Domain of unknown function 579
DUS	Dihydrouridine synthase
EAR	Ethylene response factor-associated amphiphilic repression
ECIP1	EIN2 C-terminus interacting protein 1
EDTA	Ethylenediaminetetraacetic acid, C ₁₀ H ₁₆ N ₂ O ₈
EIN2	Ethylene Insensitive 2
ELF3	Early flowering 3gene (<i>Arabidopsis thaliana</i>)
ERF	Ethylene response factor
Exo70A1	A component of the exocyst complex in plants
F81	Felsenstein nucleotide substitution model
FEL	Fixed Effects Likelihood
FMN	Flavin mononucleotide
FMO1	Flavin-dependent mono-oxygenase 1
FRA1	Fragile Fibre 1, a kinesin-like motor protein
FSP	Flavonol synthase-like gene
FUBAR	Fast Unconstrained Bayesian Approximation
G	Gamma distribution
GAUT1	Galacturonosyltransferase 1
GAUT3	Galacturonosyltransferase 3
GLOBOSA	B-function MADS-box gene
GUI	Graphical user interface
GY94	Goldman-Yang codon substitution model
GPA/gpa	Putative loci of <i>Primula</i> supergene model
HCL	Hydrochloric acid

HKY85	Hasegawa-Kishino-Yano Nucleotide substitution model
HYPHY	Hypothesis Testing using Phylogenies
IDT	Integrated DNA Technologies
IGEPAL	Octylphenoxypolyethoxyethanol
IPTG	Isopropyl β -D-1-thiogalactopyranoside
IRX15L	Irregular xylem 15-like protein (previously referred to as <i>NBP1</i>)
JC	Jukes and Cantor nucleotide substitution model
JTT	Jones-Taylor-Thornton nucleotide substitution model
K2	Kimura 2-Parameter nucleotide substitution model
KCL	Potassium chloride
LB	Lysogeny broth
LEJ2	Cystathionine beta-synthase domain-containing protein (“Loss of the timing of ET and JA biosynthesis 2”)
LG	Le and Gascuel nucleotide substitution model
LH	Long homostyle
LRRK	Leucine-rich repeat protein kinase
LRR-RLK	Leucine-rich repeat receptor-like kinase
LRT	Likelihood ratio test
MAPK	Mitogen-activated protein kinase
MBD8	Methyl-CpG-binding domain 8
MCL	Maximum Composite Likelihood
MCMC	Markov Chain Monte Carlo
MEGA	Molecular Evolutionary Genetic Analysis
MEME	Mixed Effects Model of Evolution
MG94	Muse-Gaut codon substitution model
MgCl₂	Magnesium chloride
MHC	Major Histocompatibility Complex
MKT	MacDonald-Kreitman Test
MLPK	M-locus protein kinase
NaCl	Sodium chloride
NaOH	Sodium Hydroxide
NBP1	Nucleic acid binding protein
nifR3	A family of nitrogen regulation proteins
NRFP	Nitrogen regulation family protein
2OG	2-oxogluterate
PAML	Phylogenetic Analysis by Maximum Likelihood
PCR	Polymerase chain reaction
POFUT	Protein O-fucosyltransferase
PrpS	<i>Papaver rhoeas</i> pollen S
PrsS	<i>Papaver rhoeas</i> stigma S
PvGLO	Short-specific <i>GLOBOSA</i> homologue identified in <i>Primula vulgaris</i> and <i>Primula veris</i>
PvSLL1	Unknown plasma membrane protein identified in <i>P. vulgaris</i> as having apparent long- and short-specific alleles
PvSLL2	Flower timing gene identified in <i>P. vulgaris</i> as having apparent long- and short-specific alleles
PvSLP1	Short-specific RFLP identified in <i>P. vulgaris</i>
RBP-MS	RNA binding protein with multiple splicing
REL	Random Effects Likelihood
RNA	Ribonucleic-acid
RNABP	RNA binding protein
RNABP34	RNA binding protein 34

RNase	Ribonuclease
RNA-seq	RNA sequencing
RTREV	General Reverse Transcriptase nucleotide substitution model
SCE1	Small ubiquitin-related modifier (SUMO) conjugating enzyme 1
SCR	S-locus Cysteine-rich protein
SD	Standard deviation
S-ELF3	S-locus Early Flowering3 gene (<i>Fagopyrum</i>)
SH	Short-homostyle
SFB	S-haplotype specific F-Box Protein
SI	Self and intra-morph incompatibility
SLAC	Single-Likelihood Ancestor Counting
S-locus	Self-incompatibility locus
SLG	S-locus glycoprotein
SPH	S-protein homologue
SRK	S-locus receptor kinase
SRY	Sex determining region Y
SUMO	Sumo conjugating enzyme
T92	Tamura 3-parameter nucleotide substitution model
TAE	A buffer containing tris base, acetic acid, and EDTA
TAIR	The <i>Arabidopsis</i> Information Resource
TE	A buffer containing tris base and EDTA
T_m	Melting temperature
tRNADS	tRNA-dihydrouridine synthase
TSP	Trans-species polymorphism
TSSI	Unknown short (“thrum”) -style specific gene identified in <i>Linum</i>
Tsst1	<i>Turnera subulata</i> short stamen 1
UNKN	Unknown
WAG	Whelan and Goldman nucleotide substitution model
WRKY	Transcription factor named for its conserved ‘WRKY’ domain

Units of Measure

kb	Kilobase
bp	Base pair
g	Gram
mg	Milligram
ng	Nanogram
L	Litre
mL	Millilitre
μL	Microlitre
cm	Centimetre
mm	Millimetre
mM	Millimole
pmol	Picomole
x g	Relative centrifugal force, measured as a multiple of Earth's gravitational force
rpm	Revolutions per minute
V	Volts
v/v	Volume per volume
w/v	Weight per volume

1.0 INTRODUCTION

Since Darwin (1877), the question of how and why such a wide variety of flower morphologies might be maintained in nature has enjoyed a rich research tradition. Indeed, the precise arrangement of sexual organs – that is, of pistils and stamens – within the flower has been of particular interest in this regard (Vuilleumier 1967; Ganders 1979; Barrett 2002). In some species, differences in the heights of these organs distinguish individuals as belonging to one of two or three discrete mating groups – a condition known as heterostyly (Darwin 1877; Vuilleumier 1967; Ganders 1979; Barrett 1992). Here, I review the existing literature on this subject by first explaining the nature of heterostyly, with particular emphasis on distyly and the standing arguments regarding its evolution and function. I will then turn my attention to homomorphic self-incompatibility (SI) systems, where I will elucidate the two main varieties – gametophytic and sporophytic – in turn, with particular emphasis on the diversity of molecular mechanisms underlying self-pollen rejection. Heteromorphic SI, often associated with heterostyly, will then be discussed, while paying careful attention to how the system functions and is inherited in distylous *Primula*, *Fagopyrum*, and, to a lesser extent, *Linum*. Similar topics will subsequently be covered more fully with reference to *Turnera* before detailing the overall objectives of the investigation described herein.

1.1 HETEROSTYLY

Heterostyly refers to a unique flower polymorphism whereby individuals of a hermaphroditic population may be categorized into one of two (distyly) or three (tristyly) discrete mating groups depending on the particular lengths of their styles and heights of their anthers (Darwin 1877; Vuilleumier 1967; Ganders 1979; Barrett 1992). More specifically, heterostylous species are characterized by the presence of discrete morphs which exhibit a complimentary arrangement of these organs, in a condition referred to as reciprocal herkogamy (Webb and Lloyd 1986). For example, in tristylous species, individuals may be identified as being either long-styled with two sets of stamens being short or intermediate in length; mid-styled with long and short stamens; or short-styled with stamens that are intermediate and long in length (Ganders 1979). Distylous species, on the other hand, are more common and contain only

a short and a long-styled morph, each having long or short stamens, respectively (Ganders 1979; Figure 1).

A condition first coined by Hildebrand (1866) and embellished by Darwin (1862, 1877), heterostyly has since come to represent one of the fundamental model systems with which to address a range of queries in evolutionary biology (Vuilleumier 1967; Ganders 1979; Ornduff 1992; Barrett 1992; Barrett and Shore 2008). Indeed, Darwin himself considered his initial work on heterostyly to be quite important – not only as a scientific achievement – but as a personal undertaking. Indeed, in his 1872 treatise, *The Different Forms of Flowers*, he stated that distyly, in particular, “was a case to which no parallel exists” in that nowhere else in nature was a species divided into two bodies, which could not be accurately considered separate sexes, but still required complete reciprocal fertilization in order to produce viable progeny. In his 1876 autobiography he further noted that “no little discovery of mine ever gave me so much pleasure as making out the meaning of heterostyled flowers” even though, he conceded, his work to that end had been recognized by “only very few persons” (Darwin 1876). In truth, aside from those who have since endeavoured to more fully understand heterostyly, relatively few people are likely aware of his work in this area, even now.

Since Darwin’s initial investigations, several hypotheses have been advanced in order to explain the adaptive significance of heterostyly (Barrett 1990; Lloyd and Webb 1992a & b). Though he did not then call it heterostyly, Darwin was the first to suggest that the reciprocal arrangement of styles and stamens exhibited by distylous *Primula* may be an adaptation intended “namely, to favour the inter-crossing of distinct individuals” (1862). By performing several pollination experiments, he showed that the relative heights of the sexual organs of the two morphs allowed insect pollinators to leave pollen from one morph on the stigma of another. He also demonstrated that heteromorphic pollinations were more successful than those of the homomorphic variety (Figure 1; Darwin 1862 & 1877). In this way, he suggested, the occurrence of the two morphs functions similarly to the presence of individuals of separate sexes, or dioecy, in other plants, in that it discourages self-fertilization and encourages outcrossing (Darwin 1862 & 1877).

To be sure, this interpretation might be considered most popular among the majority of those who are currently concerned with the issue of heterostyly (Vuilleumier 1967; Ganders 1979; Barrett 1990; Dulberger 1992) and, moreover, it has been supported by several subsequent

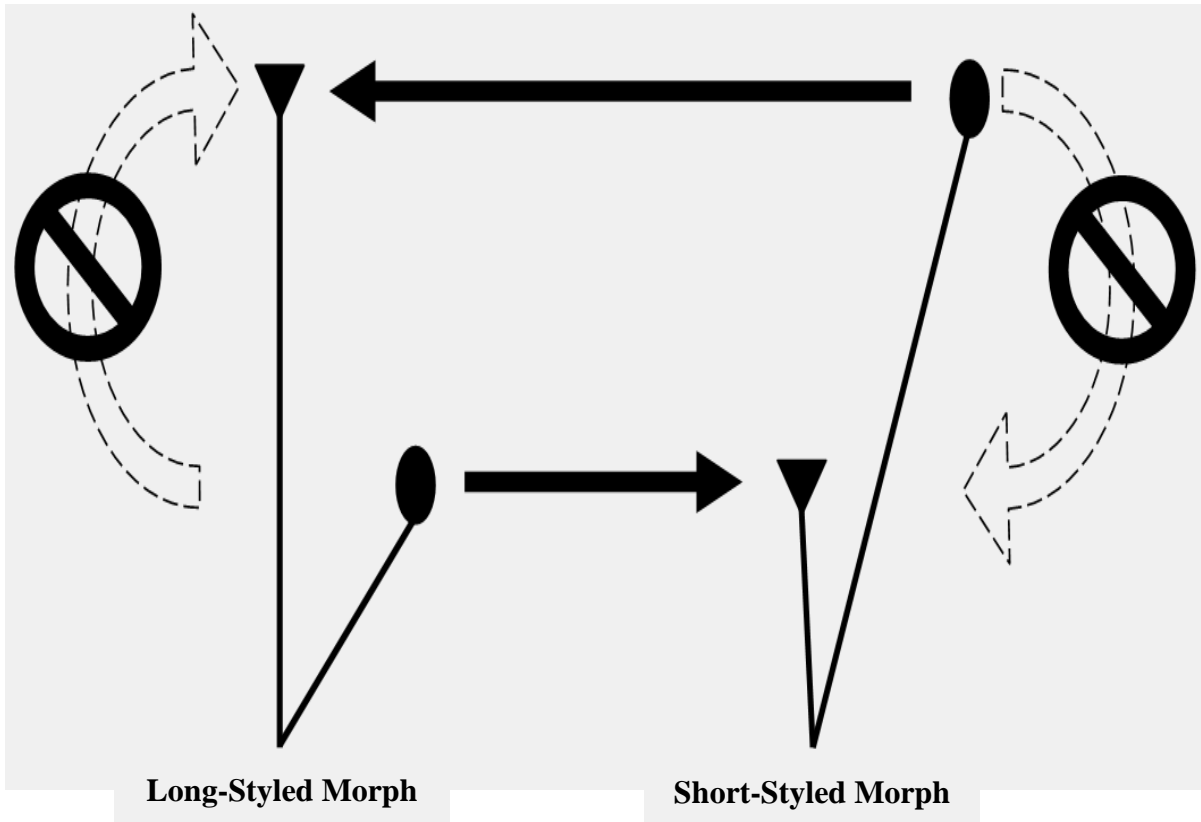


Figure 1: A graphical representation of distyly. Anthers are symbolized by ovals, while stigmas are symbolized by triangles. Pollen-stigma compatibility relationships between long- and short-styled individuals are indicated by arrows. Compatible pollinations (those that occur between individuals of different floral morphs) are shown as solid arrows, while incompatible pollinations (those that occur between individuals of the same morph) are shown as dashed arrows. As incompatible pollinations do not result in seed set, these arrows are also crossed out.

lines of evidence, including further studies on the deposition of pollen on the bodies of pollinating insects (Rosov and Serebtsova 1958; Olesen 1979; Lewis 1982; Wolfe and Barrett 1989) and pollen flow between the different flower morphs of various distylous and tristylous species (Kohn and Barrett 1992; Stone and Thomson 1994; Barrett and Glover 1985; Lau and Bosque 2003; Keller et al. 2014; Zhou et al. 2015). Indeed, these sorts of experiments are especially efficacious in distylous taxa, as their pollen is also dimorphic, making it much easier to determine which morph deposited pollen originated from (Barrett 1992; Barrett and Shore 2008).

Even so, in many cases, this system appears to be imperfect, at least as far as promoting efficient inter-morph pollen transfer goes (Barrett and Shore 2008; Cohen 2010). For instance, pollen is often identified in considerable quantities on the styles of the same morph (Barrett and Glover 1985) and, similarly, the different morphs commonly capture and transfer different amounts of compatible pollen (Ganders 1974 & 1979; Lewis 1982; Stone and Thomson 1994; Keller et al. 2014). However, it has been suggested that the occurrence of intra-flower selfing might, in part, account for these apparent inequities. Indeed, studies where flowers had been successfully emasculated in order to prevent selfing do appear to indicate that reciprocal herkogamy promotes disassortative pollen flow in heterostylous species, although pollen transference is still by no means perfectly complementary, even in these cases (Ganders 1974; Barrett and Glover 1985; Barrett 1990).

While supporting evidence has certainly been garnered, what is considered to be the first robust empirical demonstration of the hypothesis that reciprocal herkogamy promotes disassortative mating was achieved only very recently, with evidence gathered from a single population of distylous *Luculia pinceana* (Rubiaceae) (Zhou et al. 2015; Fornoni and Domínguez 2015). Interestingly, the long-styled morph of this species exhibits self- and intra-morph compatibility, while those that are short-styled cannot be selfed but can be successfully crossed with other individuals of the same mating type (Ma et. al. 2009; Zhou et al. 2012). By determining paternity in progeny that resulted from naturally-occurring matings within this population, the authors identified inter-morph pollinations as the most commonly occurring outcrossing events (62.3%). Further, using the model outlined by Lloyd and Webb (1992b), which describes the maintenance of distyly in pollen-limited and pollen-unlimited populations, it was concluded that this rate of inter-morph crossing was sufficient for the maintenance of sexual

dimorphism in the population, even in the absence of SI. That being said, more incidences of assortative mating were detected than initially anticipated based on the expected degree of pollen transfer between the morphs given their reciprocally arranged floral morphologies (Zhou et al. 2015).

The picture becomes markedly more complicated when one considers the fact that reciprocal herkogamy is often also associated with self- and intra-morph incompatibility in heterostylous systems, a condition that already ensures outcrossing (Barrett and Shore 2008). In the case of distyly, for instance, successful pollinations are often only achieved by crossing the short and long morphs, with intra-morph pollinations being incompatible and thus resulting in little or nothing in the way of seed set (Figure 1). As was also indicated by Darwin, a plant variety, in already expressing the SI trait, likely “would never have been rendered heterostyled, as this state would then have been superfluous” (1877). Indeed, explaining this apparent redundancy has been a long standing issue for those seeking to resolve the question of how a condition such as heterostyly might have evolved in the first place (Darwin 1877; Dulberger 1992; Webb and Lloyd 1986; Lloyd and Webb 1992ab; Barrett 2002; Barrett and Shore 2008).

In order to overcome this challenge, some have opted to instead consider reciprocal herkogamy and SI in so far as they might promote fitness via either male or female function, respectively (Barrett 2002; Barrett and Shore 2008). In agreement with Darwin’s original hypothesis (1877), reciprocal herkogamy would thus promote the male component of fitness by encouraging efficient cross-pollination and, in this way, discouraging pollen wastage (Baker 1964; Lloyd and Webb 1992ab; Barrett 2002; Barrett and Shore 2008). SI, on the other hand, would prevent self-fertilization, and thus inbreeding depression, thereby promoting the maternal component of fitness (Barrett 2002; Barrett and Shore 2008).

Despite the fact that there exists a general consensus in the literature that heterostyly largely functions as an outcrossing mechanism (Barrett and Shore 2008), whether or not reciprocal herkogamy preceded or followed the development of SI in the evolution of heterostyly has been and continues to be debated (Charlesworth and Charlesworth 1979a & b; Lloyd and Yates 1982; Webb and Lloyd 1986; Lloyd and Webb 1992ab; Barrett and Shore 2008). While several competing theories have been offered (Mather and de Winton 1941; Ernst 1936; Crowe 1964; Yeo 1975...etc.), those that have been presented by Charlesworth and Charlesworth

(1979b) and Lloyd and Webb (1992a & b) are more commonly held and may be considered the most well-articulated.

According to the population genetics models of Charlesworth and Charlesworth (1979b), the evolution of reciprocal herkogamy was necessitated by the earlier development of SI. In their view, the reciprocal arrangement of styles and stamens in the flower serves to reinforce a pre-existing SI system by ensuring that pollen gets to “the right place” (Charlesworth and Charlesworth 1979b; Yeo 1975). Using evidence from the literature of the day in concert with computer simulations, their justification for emphasizing SI as the first step in the evolution of heterostyly was based on the following premises: 1) Instances of heterostyly without SI were, at the time, considered to be rare and those that did exist were thought to represent a reversion to self-compatibility from SI (though this is not thought to always be the case now; Santos-Gally et al. 2015); 2) In the absence of SI, style-length polymorphisms would be difficult to maintain, as such mutations would have equal chances of either being eliminated from a population (due to a resultant reduction in pollination in the largely homostylous starting population) or reaching fixation; 3) Initial stamen-height mutations would also likely become fixed, unless selfing was maintained at uncommonly low levels; and 4) SI would not likely arise in a population expressing reciprocal herkogamy as self-fertilization would already be reduced in this case. Their model then posits that heterostyly evolved within self-compatible, homostylous ancestral populations, which first acquired recessive pollen-compatibility mutations, resulting in mutant individuals that were effectively pollen-sterile. This was then followed by subsequent and reciprocal style-compatibility mutations, which restored the fertility of the original pollen mutants. Style and stamen length dimorphism then followed, and was presumably selected for in order to reduce pollen wastage resulting from frequent incompatible self-pollinations (Charlesworth and Charlesworth 1979b). One testable prediction that emerges from their model is that distyly will only evolve if the genes determining SI and reciprocal herkogamy are linked at the *S*-locus (Charlesworth and Charlesworth 1979b; Barrett and Shore 2008).

According to the phenotypic model of Lloyd and Webb (1992a), nearly the opposite is true. Representing a revival of Darwin’s initial – though vague – conception, the authors contend that the first evolutionary steps toward heterostyly involved selection for reciprocal herkogamy in order ensure efficient cross-pollination. First, the introduction of a style length mutation into ancestral approach-herkogamous populations (monomorphic with stigmas

positioned above anthers) resulted in populations containing both approach- and reverse-herkogamous (stigmas positioned below anthers) individuals. The initial spread of the reverse herkogamous mutant was then presumably favoured due to decreased interference between male and female reproductive structures from the perspective of pollen donation (Barrett 1990). Further selection for more efficient pollen transfer between the morphs then resulted in anther-height dimorphism. The development of SI may or may not have followed depending on the relative costs of inbreeding in the population in question (Lloyd and Webb 1992a & b; Barrett and Shore 2008). Unlike the Charlesworth and Charlesworth model (1979b), the phenotypic model of Lloyd and Webb (1992 a & b) does not make any predictions about the genetic basis of distyly, and thus allows for a wide array of possible underlying genetic mechanisms, including the occurrence of genes with morph-specific functions (Barrett and Shore 2008).

As heterostyly is expected to have multiple independent origins, sometimes even within a single genus (Graham and Barrett 2004; Barrett and Shore 2008), it has been suggested that both of these proposed evolutionary pathways may have been exploited at one point or another (Fornoni and Domínguez 2015; Zhou et al. 2015). That said, while the Charlesworth and Charlesworth model cannot be outwardly rejected, most recent evidence does support the plausibility of the Lloyd and Webb model (Arroyo et al. 2002; Pérez-Barrales et al. 2006; Zhou et al. 2012; Zhou et al. 2015; Fornoni and Domínguez 2015...etc.). As was earlier described, Zhou et al. (2015) recently demonstrated that reciprocal style- and stamen-length dimorphism can be effectively maintained in a population, even in the absence of SI. This finding effectively illustrates that SI is not a necessary precondition for the evolution of heterostyly, as would be expected under the Lloyd and Webb model (Zhou et al. 2015; Fornoni and Domínguez 2015). Also, due to its emphasis on the promotion of outcrossing (as opposed to the avoidance of inbreeding) as the initial driving force behind the development of heterostyly, the Lloyd and Webb model necessarily stresses the role of pollinator behaviour in the evolution of floral morphology (Lloyd and Webb 1992 a & b; Pérez-Barrales et al. 2006; Fornoni and Domínguez 2015). In this same vein, phylogenetic character reconstructions in *Narcissus* have suggested that style-length dimorphism is ancestral to heterostyly and arose before SI in this group (Graham and Barrett 2004; Perez et al. 2004; Pérez-Barrales et al. 2006). Further, these results have also shown that the transition to heterostyly in *Narcissus* more closely mirrors historical pollinator changes than it does the evolution of SI (Pérez-Barrales et al. 2006). This evidence, of course,

leaves the question of “why SI?” unresolved, though it has been suggested that it may, perhaps, make up for the apparent inefficiencies of reciprocal herkogamy by ensuring perfectly disassortative mating (Zhou et al. 2015).

1.2 SELF-INCOMPATIBILITY (SI) SYSTEMS

SI, itself, may be divided into several varieties. Specifically, SI systems may be characterized as being either gametophytic or sporophytic in nature, and further, sporophytic SI may, in turn, be broken down into its homomorphic and heteromorphic forms (Stevens and Murray 1982; Barrett 1988). Generally, gametophytic SI refers to a scenario in which pollen is rejected by the receiving style as a result of there being a match between the haploid SI genotype of the pollen (the male gametophyte) and one of the SI alleles (hereafter referred to as *S*-alleles or *S*-haplotypes) present in the diploid tissues of the style (Newbiggin et al. 1993; Matton et al. 1994). In the case of sporophytic SI, on the other hand, pollen rejection depends not on the haploid genotype of the pollen, itself, but on the diploid genotype of its parent plant (the sporophyte producing the pollen) at the SI locus (hereafter referred to as the *S*-locus) (Matton et al. 1994). With reference to sporophytic SI, specifically, the terms homo- and heteromorphic may then be used to refer to the particular plant species in question. For instance, where a species or population is characterized by only a single flower morph, sporophytic SI is deemed to be homomorphic (Stevens and Murray 1982). In situations where a species or population is represented by a particular number of discrete flower morphs, as is the case with heterostylous systems, SI is referred to as heteromorphic (Stevens and Murray 1982). As might be imagined, the later will be the larger focus of the current review. However, gametophytic and homomorphic sporophytic SI will also be briefly considered given that considerable detail has emerged regarding the molecular and genetic mechanisms underlying their expression.

Though represented by a variety of molecular mechanisms, SI systems, both sporophytic and gametophytic, are inherited in a similar fashion. While the number and identities of the particular genes involved may differ substantially and are not thought to be homologous between systems, in all of the cases discussed, the inheritance of SI is – or is considered to be – controlled by a single polymorphic locus (the *S*-locus), now known to consist of tightly-linked genes that represent the style- and stamen-specific components of SI response (Gibbs 1986; Nasrallah and

Nasrallah 1989; Newbigin et al. 1993; Kao and Tsukamoto 2004; Barrett and Shore 2008). In those plants showing sporophytic SI, in particular, the genetic control of SI response may appear to be somewhat obscured by the complex dominance relationships that may exist between *S*-alleles (Nasrallah and Nasrallah 1989; Hatakeyama et al, 1998; Billiard et al. 2007). And moreover, the relative dominance or recessiveness of any particular *S*-allele (of which there can be >100 in a single population) may differ depending on the particular tissue in which it is expressed (Nou et al. 1993; Hatakeyama et al, 1998; Billiard et al. 2007). Species with gametophytic SI, on the other hand, may avoid this complication due to the fact that self-pollen inhibition is determined solely by the haploid genotype of the pollen itself (Newbigin et al. 1993; Matton et al. 1994; Hatakeyama et al. 1998; Billiard et al. 2007).

At the *S*-locus of heterostylous species, genes for SI are expected to be tightly linked to those controlling reciprocal herkogamy (Bateson and Gregory 1905; Lewis and Jones 1992; Barrett and Shore 2008). However, more information on the inheritance of SI in heterostylous systems, in particular, will be given in sections 1.6 and 1.7.

1.3 GAMETOPHYTIC SI: THE S-RNase SYSTEM

Gametophytic SI is the most common form of SI and is perhaps the best studied (Barrett 1988; Newbigin et al. 1993; Kao and McCubbin 1996). Unlike in the sporophytic system, all species that exhibit gametophytic SI may be represented by any number of morphologically indistinct mating groups (Barrett 1988; Newbigin et al. 1993). There are two main types of gametophytic SI that have been well described: 1) The S-RNase system that is common to the Solanaceae, Rosaceae, and Plantaginaceae (Newbigin et al. 1993; Newbigin et al. 2008); and, 2) the *S*-glycoprotein-based mechanism that is active in *Papaver* (Franklin-Tong and Franklin 2003). However, these two gametophytic SI systems are not thought to share a common evolutionary origin (Foote et al. 1994; Franklin-Tong and Franklin 2003). Indeed, there is some evidence to suggest that the S-RNase system, alone, may have multiple origins within the Rosaceae, in particular (Tao and Iezzoni 2010; Aguiar et al. 2015; Morimoto et al. 2015).

The S-RNase system is characterized by the presence of a diversity of pistil-specific glycoproteins (Anderson et al. 1986; Anderson et al. 1989; Newbigin et al. 1993; Newbigin et al. 2008). Each individual variety segregates for a particular *S*-allele, and is a product of the *S*-locus

(Anderson et al. 1989; Gebhardt et al. 1991; Newbigin et al. 1993). The glycoproteins, themselves, are catalytically active RNases. (McClure et al. 1989; Newbigin et al. 1993; Franklin-Tong and Franklin 2003). It is thought that this RNase activity is crucial to the pollen-rejection process. Namely, S-RNases are believed to act as cytotoxins, degrading the RNA associated with the pollen and pollen tubes originating from individuals of the same mating type (Franklin-Tong and Franklin 2003; Kao and Tsukamoto 2004; Goldraij et al. 2006). Indeed, transgenic studies in *Petunia inflata* have demonstrated that, when S-RNase activity was inhibited in members of this species, the ability to reject self-pollen was subsequently lost (Lee et al. 1994). Similarly, when individuals were modified to express S-RNases associated with genotypes other than their own, they then gained the ability to reject pollen from individuals having these genotypes, as well (Murfett et al. 1994; McClure 2004). Taken together, these results strongly suggest that S-RNases, alone, are sufficient to determine incompatibility reactions in the style (McClure 2010). That being said, additional genes, unlinked to the *S*-locus, have been shown to affect pollen rejection in *Nicotiana* and *Solanum* (McClure et al. 1999; O'Brien et al. 2002; McClure 2010; McClure et al. 2011).

The pollen determinants of SI in the S-RNase system were discovered somewhat more recently than their style-specific counterparts (Lai et al. 2002; Sijacic et al. 2004). Identified as F-Box proteins, they are known to be *S*-linked and have been shown to exhibit pollen-specific expression (Lai et al. 2002; Sijacic et al. 2004; McClure 2010). In other contexts, F-Box proteins are often involved in protein ubiquitination and, indeed, this is generally thought to be their role in S-RNase-based SI (McClure 2010; Chen et al. 2012). However, there is some debate regarding their precise role, as these purported pollen-*S* determinants have also been shown to exhibit far less sequence polymorphism than their associated S-RNases in some systems (Wheeler and Newbigin 2007; Newbigin et al. 2008). As some have pointed out, if the identified F-Box proteins are indeed the pollen-specific determinants of SI, their comparatively low diversity suggests that they have arisen relatively more recently than their pistil-specific partners, causing potential problems for the foundational idea that these two components of SI have coevolved (Wheeler and Newbigin 2007; Newbigin et al. 2008; McClure 2010). In *Petunia*, however, it has been shown that each type of pollen-specific F-Box protein may, instead, have the capacity to identify a particular subset of S-RNases, and that multiple F-Box proteins may work collaboratively to identify a larger suite of non-self S-RNases (Kubo et al. 2010). Indeed,

this is the view that largely prevails in the current literature (Kakui et al. 2011; Chen et al. 2012; Kubo et al. 2015; Sun et al. 2015), though how this complex system might have evolved remains an outstanding issue (Kubo et al. 2015).

Though *S*-linked F-Box proteins have been shown to interact with *S*-RNases (Qiao et al. 2004; Kubo et al. 2010), and have been implicated in their degradation (Chen et al. 2012), the precise nature of self-pollen rejection in *S*-RNase based SI systems has not been wholly elucidated. It has been suggested, however, that the specificity of the interaction is likely governed by the pollen-*S* determinants, and not the *S*-RNases (McClure 2004; Newbigin et al. 1993; Sun et al. 2015). This is evidenced by the fact that *S*-RNases do not show any special affinity for particular RNA substrates and have been demonstrated to successfully degrade RNA from a variety of sources, *in vitro* (McClure et al. 1990). Several models have been proposed to explain the precise mechanism by which pollen-*S* determinants establish the specificity of pollen rejection by *S*-RNases (Kao and Tsukamoto 2004; McClure 2004; Hua et al. 2008; Kubo et al. 2010; Chen et al. 2012). However, none have been definitively or universally demonstrated (Hua et al., 2008; Zhang et al. 2009). Even so, it is generally thought that pollen-specific F-Box proteins likely act as *S*-RNase inhibitors by targeting inter-morph *S*-RNases for ubiquitination, thus allowing for uninterrupted pollen tube growth and successful pollination (Qiao et al. 2004; Zhang et al. 2009; McClure et al. 2011; Chen et al. 2012).

1.4 GAMETOPHYTIC SI IN *PAPAVER*

An entirely different form of gametophytic SI occurs in members of the Papaveraceae family (poppy), and particularly in those belonging to the genus, *Papaver* (Zhang et al. 2009). As with the *S*-RNase system, the production of pistil-specific glycoproteins is a key aspect of SI in this group (Foote et al. 1994; Zhang et al. 2009). However, the particular glycoproteins involved are not ribonucleases and the molecular mechanisms underlying the rejection of self-pollen differ substantially (Franklin-Tong et al. 1991; Foote et al. 1994). In *Papaver rhoeas*, these proteins are referred to as *PrsS* (*P. rhoeas* stigma *S*) and are thought to act as signalling peptides involved in SI reactions (Foote 1994; Bosch and Franklin-Tong 2008; Wheeler et al. 2009; Wu et al. 2011; Eaves et al. 2014). Interestingly, *PrsS*-like genes have been shown to represent a large gene family in *Arabidopsis*, though no known function has been ascribed to them (Ride et al. 1999).

Indeed, apparent homologues have also been identified in distylous *Turnera* (Chafe et al. 2015; Shore and Chafe, Unpublished data, and see below), suggesting that this gene family may have other roles in addition to those that they play in the regulation of gametophytic SI (Ride et al. 1999).

The precise nature of the pollen-specific counterpart of *PrsS* was entirely unknown until fairly recently (Wheeler et al. 2009). Interestingly, this gene, called *PrpS* (*P. rhoeas* pollen S), appears to be entirely novel, having no known homologues (Wheeler et al. 2009; Eaves et al. 2014). The resultant protein has been associated with the plasma membrane of the pollen tube and is believed to be a transmembrane receptor protein (Bosch and Franklin-Tong 2008; Wheeler et al. 2009; Wu et al. 2011; Eaves et al. 2014). It is thought that, when non-self PrsS signalling peptides interact with these pollen-tube-bound PrpS transmembrane proteins, a complex Ca^{2+} dependent signalling cascade results, ultimately ending in self-pollen tube inhibition and programmed cell death (Thomas and Franklin-Tong 2004; Bosch and Franklin-Tong 2008; Zhang et al. 2009; Wheeler et al. 2009; Wu et al. 2011). In *Papaver*, extensive cytoskeleton modifications are triggered by SI response, and these changes have been implicated, not only in the inhibition of incompatible pollen tubes, but also in the initiation of programmed cell death, directly (Geitmann et al 2000; Snoman et al. 2002; Thomas S.G., Huang S., Li S., Staiger C.J., and Franklin Tong V.E. 2006). Indeed, the rapid acidification of the cytosol of incompatible pollen tubes appears to be a necessary step in this process (Wilkins et al 2015). Possible roles for mitogen-activated protein kinases (MAPKs) in the instigation of programmed cell death have also been elucidated (Rudd et al. 2003; Li, Samaj, and Franklin-Tong 2007).

Clearly, many facets of the SI signalling process have been identified, however, the precise mechanism by which “self” pollen is distinguished from “non-self” pollen and the exact nature of downstream signalling processes remain largely unknown (Wu et al. 2011; Eaves et al. 2014). Importantly, however, when *Papaver PrpS*'s were introduced into the genome of self-compatible *Arabidopsis thaliana*, their SI function was retained, suggesting that SI signalling targets in *Papaver* are likely ubiquitous cellular components, common to many plant taxa (de Graaf et al. 2012; Eaves et al. 2014).

1.5 HOMOMORPHIC SPOROPHYTIC SI IN THE BRASSICACEAE

As one might recall, SI in sporophytic systems is determined by the diploid genotype of the parent plant at the *S*-locus, as opposed to the haploid genotype of the pollen (Newbiggin et al. 1993; Matton et al. 1994). Homomorphic sporophytic SI, in particular, has perhaps been best described in members of the Brassicaceae family (mustard) (Takayama and Isogai 2005). Here, the main style-specific determinant of SI is an *S*-linked receptor kinase (*SRK*) (Nasrallah et al. 1987; Stein et al. 1991; Nasrallah and Nasrallah 2014a). This protein is expressed in the upper epidermal layers of the style and has been shown to determine the specificity of pollen rejection (Takasaki et al. 2000; Takayama 2001; Suzuki et al. 2003; Takayama and Isogai 2005; Nasrallah and Nasrallah 2014a).

The pollen-specific counterpart to the *SRK* consists of a small, cysteine-rich protein (referred to as the *S*-locus cysteine-rich protein, or *SCR*), which is initially expressed in the pollen tapetum and is later incorporated into the pollen coat after maturation (Takayama et al. 2001). In terms of its role in SI, *SCR*s are thought to act as ligands for style-specific *SRK*s (Schopfer et al. 1999; Takayama et al. 2000; Hiroshi et al. 2001; Takayama et al. 2001). Specifically, when an *SRK* is in the presence of an *SCR* that is associated with the same *S*-haplotype, they form a high-affinity receptor complex on the surface of the stigma (Takayama et al. 2001; Nasrallah 2011). This interaction results in a complicated signalling cascade which ultimately ends in self-pollen inhibition (Nasrallah 2011; Takayama and Isogai 2005). In the case of non-self pollinations, it is thought that the *SCR* and *SRK* simply cannot bind, and thus the signalling pathway is never initiated (Nasrallah and Nasrallah 2014a). However, several aspects of this process have yet to be fully explained. For example, how pollen-specific *SCR*s gain access to membrane bound *SRK*s and the precise mechanism by which the two molecules “recognize” each other remains mysterious. Further, the exact nature of the signal transduction pathway that results from their interaction is still largely unknown (Nasrallah 2011; Takayama and Isogai 2005; Nasrallah and Nasrallah 2014a).

That said, additional SI “modifiers” in the Brassicaceae have been identified, including what is known as the *S*-locus glycoprotein, or *SLG*, which, in some taxa, enhances *SRK* activity when in the presence of self-pollen (Takayama et al. 1987; Takasaki et al. 2000; Suzuki et al. 2003). Other components, possibly more integral to downstream signalling, have also been

discovered, including M-locus protein kinase (MLPK), arm repeat-containing protein 1 (ARC1), and Exo70A1 (Nasrallah and Nasrallah 2014a). MLPK and ARC1 are both reported to interact with SRK (Kakita et al. 2007; Indriolo et al. 2012), while Exo71, a putative component of the exocyst complex in plants, has been shown to interact with ARC1, a known ubiquitin ligase (Samuel et al. 2009). The exocyst complex is known to be involved in targeted secretion and is thus expected to be responsible for delivering water, calcium, and other factors necessary for proper pollen hydration and germination on the surface of the style (Samuel et al. 2009). Indeed, it is through this process that the three proteins listed above are anticipated to influence SI in *Brassica* and *Arabidopsis* (Nasrallah and Nasrallah 2014a). Specifically, it has been suggested that, upon encountering incompatible pollen grains, stigma-bound SRKs, in concert with MLPKs, activate ARC1. ARC1 subsequently ubiquitinates Exo70A1, resulting in its degradation. Due to the absence of Exo70A1, it has been surmised that incompatible pollen grains may then be unable to germinate due to improper hydration (Samuel et al. 2009; Nasrallah and Nasrallah 2014a). However, as the function of the exocyst complex is not well defined in plants (Samuel et al. 2009), and ARC1 has been shown to be unnecessary for the expression of strong SI in some taxa (Nasrallah and Nasrallah 2014b), whether or not these three proteins have a role to play in SI in the Brassicaceae remains controversial (Nasrallah and Nasrallah 2014b, Goring et al. 2014).

1.6 HETEROMORPHIC SPOROPHYTIC SI

In many ways, it is redundant to refer to this type of SI as both heteromorphic and sporophytic, as all known species that exhibit heteromorphic SI are of the sporophytic type (Stevens and Murray 1982). Thus, this system will hereafter be referred to simply as heteromorphic SI. Relevant to the study described herein, it is this form of SI that acts in a number of heterostylous species (Stevens and Murray 1982). However, as the genes involved remain almost entirely unknown, very few generalizations have emerged regarding how this system actually functions, biochemically (de Nettancourt 1997; Athanasiou and Shore 1997; Miljus-Dukic et al. 2004; Klein et al. 2009). At least one reason for this might be the fact that heterostyly likely evolved several times, independently (Ganders 1979; Barrett 1992; Barrett and Shore 2008). As a result, the molecular mechanisms determining self-pollen rejection might be

expected to differ between heterostylous species, thus making the discovery of a general mechanism of heteromorphic SI less likely (Dulberger 1992; Lloyd and Webb 1992ab; Athanasiou and Shore 1997). Indeed, evidence has been accumulated to suggest that, not only may the mechanisms of SI differ between heterostylous taxa (Dulberger 1992), but also between the different flower morphs represented by a single heterostylous species (Wedderburn and Richards 1990; Lloyd and Webb 1992a; Athanasiou and Shore 1997; Miljus-Dukic et al. 2004; Weller 2009; Safavian and Shore 2010). Similarly, efforts to describe the particular sites at which self and intra-morph pollen may be rejected have shown that pollen inhibition may occur in a variety of locales, including on the surface of the stigma, within the style or the ovary, or some combination of the three, depending on the species or particular morph in question (Stevens and Murray 1982; Gibbs 1986; Wedderburn and Richards 1990; Dulberger 1992; Safavian and Shore 2010).

Though their mechanisms likely differ, some generalizations may be made concerning the genetics of heteromorphic SI across heterostylous species. Indeed, considerably more success has been made in describing the particular genetical aspects of the “heterostylous syndrome” than the precise molecular mechanisms that underlie its expression (Stevens and Murray 1982; Barrett and Shore 2008). With reference to distyly, in particular, the genes determining SI are thought to be tightly linked to those responsible for the expression of reciprocal herkogamy, as part of a single diallelic genetic locus, showing typical Mendelian patterns of inheritance (Bateson and Gregory 1905; Lewis and Jones 1992; Barrett and Shore 2008). In general, the short-styled allele (*S*-allele) is most often dominant to that of the long (*s*-allele), with this relationship being reversed in only a few genera (Bateson and Gregory 1905; Baker 1966; Lewis and Jones 1992; Barrett and Shore 2008). The number, arrangement, and identities of the putative genes involved in heteromorphic SI, where they have been identified, often differ between species (Barrett and Shore 2008). With this in mind, the nature of distyly and SI will first be described in three contrasting systems, distylous *Primula*, *Fagopyrum*, and *Linum* before moving on to discuss these topics as they apply to *Turnera*.

1.7 PRIMULA AND THE SUPERGENE MODEL

Individuals of the genus, *Primula*, are perhaps the most often used subjects in investigations of distyly and the genetic and molecular mechanisms that underlie its expression (Mast and Conti 2006; Barrett and Shore 2008; Nowak et al 2015). Indeed, distyly and SI were first recognized in *Primula* by Darwin, and since then, the general pattern of their inheritance has also been deduced from experiments using members of this genus (Darwin 1862, 1877; Bateson and Gregory 1905; Ernst 1955; Dorwick 1956). Indeed, these patterns have largely been extended to other species that express the heterostylous syndrome, as well (Barrett and Shore 2008; Cohen 2010).

The recognition in *Primula* of the infrequent appearance and inheritance patterns of aberrant homostylous and heterostylous individuals having characteristics of both the long- and short-styled morphs resulted in the formulation of the supergene model of inheritance for heterostyly (Ernst 1955; Dorwick 1956; Barrett and Shore 2008; Cohen 2010). As these individuals were thought to represent *S*-locus recombinants, it was hypothesized that three tightly linked diallelic loci – one for each set of morphological and physiological characteristics represented by heterostyly – likely controlled the expression of the whole heterostylous syndrome in *Primula* (Dorwick 1956; Barrett and Shore 2008; Cohen 2010). These three loci, all located within the larger *S*-locus, are commonly referred to as *G*, *P*, and *A*, where *G* determines style length and female incompatibility, *P* determines pollen size and male incompatibility, and *A* determines anther height (Ernst 1936; Dorwick 1956; Lewis and Jones 1992; Barrett and Shore 2008; Cohen 2010). The relative order in which these loci are arranged has been debated (Dorwick 1956; Lewis and Jones 1992; Kurian and Richards 1997), though it is generally thought to be *GPA*, with the dominant *S*-haplotype being represented by three dominant alleles at these loci (*GPA*), and the recessive *s*-haplotype being represented by three recessive alleles (*gpa*) (Barrett and Shore 2008). As the short-styled morph is most commonly heterozygous, it follows that these individuals would carry the genotype, *GPA/gpa*, while the long-styled individuals would be homozygous recessive, *gpa/gpa*, at the *S*-locus in *Primula* (Dorwick 1956). Another model has suggested that the *S*-locus in *Primula* may contain up to four additional causative loci (Kurian and Richards 1997), on top of the original three suggested by Ernst and others (Ernst 1936; Dorwick 1956; Lewis and Jones 1992). This model is considered to be highly speculative

(Barrett and Shore 2008), though it has been conceded that *S*-locus gene-arrangements may vary across species (Li et al. 2011; Nowak et al. 2015). The precise size of the purported “supergene” has also yet to be defined for this genus (Manfield 2005; Li et al. 2007, 2008, 2009, 2011; Gilmartin and Li 2010; Cohen 2010; Nowak et al. 2015), though chromosome localization experiments have suggested that the *S*-locus is located near the centromere of the largest metacentric chromosome pair in *P. vulgaris* (Li et al. 2015).

After over a century of study, only four genes (known as *PvSLP1*, *PvSLL1*, *PvSLL2*, and *PvGLO*) have been shown to be potentially *S*-linked in *Primula* (Manfield et al. 2005; Li et al. 2007, 2008, 2009, & 2015; Nowak et al. 2015). However, their functions in SI, if any, remain wholly undescribed (Nowak et al. 2015). Genome sequencing efforts in distylous *P. veris* have recently been completed, representing the first genome assembled for any heterostylous species (Nowak et al. 2015). RNA-sequencing data has further identified an additional 113 gene candidates that show significant differential expression between the floral morphs. Most interestingly, of these gene candidates, the previously identified *PvGLO* is the most significantly differentially expressed and may be absent from the long-morph *S*-haplotype (Li et al. 2007; Nowak et al. 2015). Physical and genetic maps of the *P. vulgaris* *S*-locus have also been recently constructed (Li et al. 2015). Collectively, these data are expected to yield significant insight into genetic and molecular underpinnings of heterostyly in *Primula*, and perhaps other heterostylous taxa, in the very near future (Fornoni and Dominguez 2015; Li et al. 2015; Nowak et al. 2015).

1.8 DISTYLY IN *FAGOPYRUM* AND *LINUM*

The supergene model is also commonly applied to distylous *Fagopyrum* (buckwheat), though the evidence for this supposition is not considered to be as strong (Sharma and Boyes 1961; Barrett and Shore 2008; Cohen 2010). As few as two (Woo et al. 1999; Wang et al. 2005; Barrett and Shore 2008) and as many as five or more (Sharma and Boyes 1961; Matsui et al. 2003; Yasui et al. 2008) individual loci within the larger *S*-locus have been proposed for species in this genus. However, several genes outside of the *S*-locus have also been reported to affect the expression of distyly in *Fagopyrum*, though the respective identities of these reported genes have not been determined (Matsui et al. 2004).

A number of proteins that are potentially involved in distyly and SI in this genus have been identified (Miljus-Dukic et al. 2004; Yasui et al. 2012). In a comparison of protein profiles obtained from the styles of long- and short-styled morphs of *Fagopyrum esculentum* after self and cross pollination, two groups of proteins, one associated with the short- and one associated with the long-styled morph, were revealed (Miljus-Dukic et al. 2004). Though their appearance was coincident with SI response, the identities and potential functions of these proteins have not been reported (Miljus-Dukic et al. 2004; Barrett and Shore 2008).

Interestingly, when the pistils of distylous *F. esculentum* were treated with a variety of protease inhibitors after being self-pollinated, self-pollen tube inhibition was suppressed, suggesting that proteases may be involved in SI response in this species (Miljus-Dukic 2007). Similarly, when pistils were treated with phosphatase inhibitors and calcium antagonists after self-fertilization, SI was also inhibited, suggesting that calcium signalling and phosphatases might be important components of the SI response, as well (Miljus-Dukic 2003). According to the authors, these two results, in particular, may indicate that *Fagopyrum* employs molecular mechanisms very similar to those that act in other homomorphic SI systems, like *Papaver* (Miljus-Dukic 2003, 2007).

Alternatively, if the proposal made by Yasui et al. is instead adopted, the particular molecular mechanisms employed by heteromorphic SI systems may be altogether different (2012). More recently, a gene found to be closely linked to the *S*-locus and expressed only in the short-styled morph of three separate species of *Fagopyrum* was identified (Yasui et al. 2012). By examining differentially expressed genes in the styles of the two floral morphs, a novel gene was identified, showing some homology to the Early Flowering 3 (*ELF3*) gene in *A. thaliana* (Yasui et al. 2012). Taking into account this homology and its linkage to the *S*-locus, this gene was thus named *S*-Locus Early Flowering 3 (*S-ELF3*). Due to its apparent evolutionary persistence in the short-styled morph and its reported linkage to the *S*-locus, this gene was likewise assumed to have a potentially important function in determining the expression of distyly and SI in the short-styled morph of distylous *Fagopyrum* (Yasui et al. 2012). Also, as a result of its presence in only short-styled individuals, it was proposed that the dominance of the *S*-allele may be explained merely by its absence in the long-styled morph (Yasui et al. 2012).

Though limited in scope, some effort has also been made to describe the molecular events that lead to self- and intra-morph pollen rejection in distylous *Fagopyrum* (Miljus-Dukic 2003,

2007; Yasui et al. 2012). For instance, with the discovery that the *ELF3* gene produces a nuclear DNA binding protein in *Arabidopsis*, it was speculated that *S-ELF3* could possibly behave as a transcription factor in *Fagopyrum* and may thus be involved in various aspects of distyly and SI in short-styled individuals in this way (Yasui et al. 2012). However, the precise function of the protein produced by this gene in *Fagopyrum*, as well as the timing and cellular location in which it is expressed, has yet to be characterized (Yasui et al. 2012).

Interestingly, morph-specific genes have also been reportedly identified in distylous *Linum* (flax) (Ushijima et al. 2012 & 2015). While the *S*-locus in this genus is expected to be diallelic, it has not been definitively shown (Ushijima et al. 2012 & 2015). Indeed, this has proven difficult to demonstrate due to the often very subtle or even non-existent differences in stamen lengths that are observed between morphs (Barrett 1992). As a result, defining and distinguishing the floral types has been somewhat problematic (Ushijima et al. 2015). However, more recently, a clearly segregating population of *L. grandiflorum* was assembled by exploiting style to stamen length ratios in order to define morphs. This population has since been used to explore *S*-locus inheritance and identify potentially *S*-linked genes in this genus (Ushijima et al. 2015). As a result of these explorations, the short-style-specific gene, *TSSI*, was shown to be *S*-linked (Ushijima et al. 2015). *TSSI* codes for a 19 kDa basic protein with no identifiable signal peptide or transmembrane domains, as might be expected of a protein involved in SI. While there is some evidence to suggest that it may be a secreted protein, *TSSI* appears to have no significant homologues, and thus any pronouncements regarding its potential function in SI have been largely speculative (Ushijima et al. 2012).

1.9 DISTYLY IN *TURNERA*

Formerly a member of the Turneraceae family, which has more recently been incorporated into the Passifloraceae, the genus, *Turnera*, consists of species that are largely neotropical in origin (Shore et al. 2006; APG 2009). The great majority of the 143 species represented by this group are distylous and typically show strong SI (Shore et al. 2006; Arbo 2015). Indeed, molecular phylogenetic analyses also suggest that this condition is likely to have evolved independently at least once within the former Turneraceae, with no other members of the Passifloraceae having been identified as heterostylous (Shore et al. 2006).

Distyly in *Turnera* also appears to be inherited as a single, diallelic genetic locus with two alternative alleles or haplotypes (Shore and Barrett 1985b). As with *Primula* and *Fagopyrum*, the *S*-locus in *Turnera* is most often conceptualized in accordance with the ‘supergene’ model, though this has not been definitively demonstrated (Tamari et al. 2005; Barrett and Shore 2008). Even so, the majority of the genetic evidence obtained for species of this genus is, at the very least, consistent with the supergene model of *Primula* (Shore and Barrett 1985b; Tamari et al. 2001; Tamari et al. 2005; Shore et al. 2006; Labonne et al. 2010). However, there is also some evidence to suggest that pollen size in *Turnera* may be determined by more than one gene residing within or near the *S*-locus, and additional genes outside of the apparent *S*-locus may affect style and stamen length, as well (Labonne et al. 2010).

As with other distylous species, differences in protein expression in the two morphs has been exploited in *Turnera* with the intent of uncovering both the genetic and molecular basis of distyly in members of this genus (Athanasίου and Shore 1997; Khosravi et al. 2003 & 2004; Tamari and Shore 2004 & 2006; Shore et al. 2006). Two proteins, a polygalacturonase and an α -dioxxygenase, both specific to the short-styled morph of several *Turnera* species, have been identified in this way (Athanasίου et al. 2003; Khosravi et al. 2003, 2004; Tamari and Shore 2004 & 2006). In fact, two polygalacturonases were initially detected: one specific to the stylar tissue and another being pollen-specific (Athanasίου et al. 2003). However, it was later determined that the pollen-specific protein was likely unrelated to the expression of distyly in *Turnera*, generally, as it failed to be identified in the tissues of several related species of this genus (Tamari and Shore 2004). The short style-specific protein, on the other hand, was found to be expressed in all species of *Turnera* investigated, including an individual of the closely related genus, *Piriqueta* (Khosravi et al. 2003; Tamari and Shore 2004). Further, the style-specific polygalacturonase was also shown to be absent in several self-compatible homostylous species, thus suggesting that the stylar polygalacturonase perhaps has a role to play in determining reciprocal herkogamy and SI in this group (Tamari and Shore 2004). Importantly, previous results had indicated that the timing of its expression likely coincided with the onset of SI response (Athanasίου et al. 2003). While polygalacturonases generally participate in cell growth processes, and several proposals have been made to describe its potential function in the expression of characteristics associated with distyly, its precise role, if indeed it has one, remains

unknown (Athanasidou et al. 2003; Khosravi et al. 2003; Tamari and Shore 2004, 2006; Shore et al. 2006; Barrett and Shore 2008).

The short-specific α -dioxygenase was also found to be expressed in high levels in the stilar tissues of several *Turnera* species, and this tissue specific expression also appeared to be coincident with SI response (Khosravi et al. 2004). However, these proteins were shown to have homology to other α -dioxygenases involved in stress induced signalling responses in plant, thus making its potential role in distyly unclear (Khosravi et al. 2004). Moreover, the genes encoding both the α -dioxygenase and the polygalacturonase were not found to reside at the *S*-locus, itself, suggesting that they either do not have a role in determining the syndrome or, alternatively, they are upregulated in short-styled individuals by some as of yet unknown component of the distyly locus in *Turnera* (Athanasidou et al. 2003; Khosravi et al. 2004; Labonne et al. 2010).

Though little is known about the precise molecular mechanisms by which SI acts in *Turnera*, there is some evidence to suggest that the particular processes involved may be morph-specific (Tamari et al. 2001; Safavian and Shore 2010). Initial observations of pollen tube growth in several species of *Turnera* and *Piriqueta* showed that incompatible pollen tubes tended to extend somewhat further into the styles of the long-styled morph before any further growth was arrested, though the difference between the morphs was not found to be statistically significant (Tamari et al. 2001). Striking structural differences were also noted, as these incompatible pollen tubes appeared to be associated with the generation of callose plugs, which appeared only very rarely in the pollen tubes produced in the short-styled morph within 24 hours of self-pollination (Tamari et al. 2001).

These observations were later reaffirmed in a second investigation where several other ultrastructural differences between the incompatible pollen tubes of the two morphs were also identified (Safavian and Shore 2010). In addition to the distinctive presence of callous plugs at their tips, incompatible pollen tubes in the long-styled morph were found to be further characterized by the swelling and loss of mitochondrial cristae, as well as the occurrence of circular-shaped rough endoplasmic reticula (Safavian and Shore 2010). Upon self-pollination in the short-styled morph, the cell wall and plasma membrane located at the apex of pollen tubes appeared to rupture, thus rendering cellular organelles largely indistinguishable (Safavian and Shore 2010). Swollen mitochondria in the pollen tubes of the long-styled morph were hypothesized to be associated with cells that were undergoing programmed cell death, as had

been observed in other plant species (Safavian and Shore 2010). The cell membrane rupturing that was characteristic of pollen tubes in short-styled morphs was speculated to have resulted from cell necrosis, however, suggesting that the two morphs may employ different mechanisms in the rejection of self-pollen (Safavian and Shore 2010).

More recent efforts to characterize distyly in *Turnera* have focused on determining the precise identities of the genes involved by positional cloning, chromosome walking, deletion mapping, and a variety of other approaches. To do so, several molecular markers tightly linked to the *S*-locus were initially identified and their relative positions with reference to the *S*-locus were subsequently mapped (Labonne et al. 2008, 2009). As a result of these mapping efforts, alone, two putative genes and a non-LTR retroelement were discovered to be tightly associated with the *S*-locus (Labonne et al. 2009). Indeed, one gene, coding for a sulfotransferase, appeared to be somewhat differentially expressed in the short and long morphs of three *Turnera* species. Sulfotransferases have a variety of reported functions in plants. However, whether or not they play a role in SI in *Turnera* is uncertain (Labonne et al. 2009). A second gene, coding for an N-acetyltransferase, also showed some evidence of differential expression between the morphs (Labonne et al. 2009). However, the recovery of several individuals found to result from recombination events between the N-acetyltransferase gene and the *S*-locus suggested that, while *S*-linked, the N-acetyltransferase gene was not likely located within the *S*-locus, itself (Labonne et al. 2009).

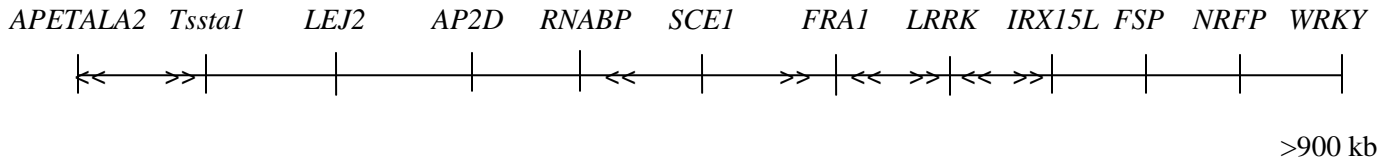
The *S*-linked non-LTR retroelement that was discovered appeared to be non-functional, on the other hand (Labonne et al. 2009). While retroelements are apparently a common feature of the *S*-locus in other species (Lai et al. 2002; Manfield 2005), their potential role, if indeed they have one, is currently unknown. However, it has been suggested that retroelements may suppress recombination at the *S*-locus (Wheeler et al. 2003), or, alternatively, they may simply accumulate in the area as a result of recombination suppression, itself (Wright et al. 2003; Labonne et al. 2009). Interestingly, previous investigations of recombination rates in *Turnera* have revealed no statistically significant evidence of recombination suppression at the *S*-locus region. However, it was conceded that, if the *S*-locus were sufficiently small in size, the tests employed may have lacked the power necessary to detect it (Labonne et al. 2007).

Despite the fact that their potential functions in distyly in *Turnera* remain elusive, the previously identified *S*-linked genes and markers have provided valuable landmarks for

subsequent chromosome walking experiments in distylous *T. subulata* (Labonne et al. 2009; Labonne and Shore 2011; Labonne 2011, PhD thesis). A bacterial artificial chromosome (BAC) library has since been constructed using genetic material obtained from a short-styled (Ss) individual. The library was subsequently screened for the presence of the already identified *S*-linked markers so as to facilitate the isolation of potentially relevant clones. Isolated BACs were sequenced and primers designed to their ends were then used to probe the library in an effort to extend the walk (Labonne and Shore 2011; Labonne 2011, PhD thesis). A series of floral mutants possessing deletions in the dominant *S*-haplotype were also exploited in order to identify BAC clones containing portions of the *S*-locus (Labonne et al. 2010; Labonne and Shore 2011; Labonne 2011, PhD thesis). From these initial efforts, a single BAC clone was found to contain the entire recessive *s*-haplotype (Labonne and Shore 2011; Labonne 2011, PhD thesis). The *s*-haplotype containing BAC has since been sequenced and additional clones containing portions of the *S*-haplotype have also been identified and sequenced in this way (Labonne 2011, PhD thesis; Shore and Chafe, Unpublished data). While the *S*-haplotype in *T. subulata* has not been fully characterized, it is known to be much larger than its recessive counterpart. While the *s*-haplotype is approximately 192kb, the portions of the *S*-haplotype that have been identified contain >900kb of sequence information, with much of the additional DNA being represented by transposable elements (Shore and Chafe, Unpublished data).

While the two haplotypes appear to be characterized by the presence of some of the same putative genes, their relative arrangements are known to differ due to one or more possible inversions (Shore and Chafe, Unpublished data; Figure 2). Further, very recently, short-specific genes have also been identified at the *S*-locus in *T. subulata* and these genes appear to be *S*-linked in a number of related species of *Turnera* and *Piriqueta*, as well (Shore and Chafe, Unpublished data). The gene we have termed *Tsstal1*, for example, was identified from a previously constructed RNA sequencing (RNA-seq) data set (Shore and Chafe, Unpublished data). This data set was constructed using RNA extracted from pooled samples of each of the following floral tissues obtained from tetraploid *T. subulata*: (1) the short styles and (2) long stamens of short-styled individuals, as well as (3) the long styles and (4) short stamens of long-styled individuals. After sequencing, transcripts were assembled, and read counts were summed for each transcript, creating 4 tissue-specific RNA-seq libraries. Transcript read-count data were then analyzed in order to detect tissue-specific differential expression in short-styled versus long-

S-Haplotype



s-Haplotype

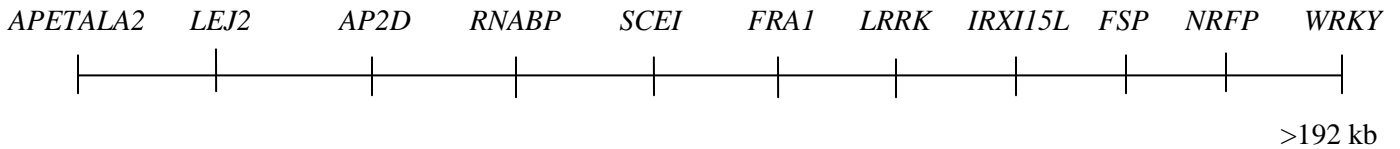


Figure 2: Graphical representations of the recessive *s*- and dominant *S*-haplotypes, based upon Bacterial Artificial Chromosome (BAC) sequencing from a single individual of diploid *T. subulata*. The approximate gene arrangements (*APETALA2*, *Tssta1*, *LEJ2*, *AP2D*, *RNABP*, *SCE1*, *FRA1*, *LRRK*, *IRX15L*, *FSP*, and *NRFP*, and *WRKY*) are known with certainty for the *s*-haplotype. To date, the entire *S*-haplotype has not been fully sequenced. Gaps in this sequence are bracketed on the left and right by “<<” and “>>” symbols. The genes that are anticipated to reside within these unsequenced sections of the dominant *S*-haplotype are indicated. The *S*-haplotype gene order shown above may be considered hypothetical, as contig orientation was assumed, in part, based on the *s*-haplotype sequence. That is, in some cases, gene order may be inverted relative to what is shown, or some genes may not be represented at all. It should also be noted that additional genes may also reside in the as of yet unsequenced regions of the *S*-haplotype. Indeed two such genes have been identified between *Tssta1* and the presumed location of *APETALA2*. The approximate size of the recessive *s*-haplotype is >192 kB. The boundaries of the recessive *s*-haplotype were determined by deletion mapping with individuals that harboured deletions in the *S*-haplotype. As a result, the recessive *s*-haplotype may itself not be fully defined. The size of the dominant haplotype is known to be much larger than its recessive counterpart based on these data, as the sequence already consists of 900 kb of genetic information. Diagrams are not drawn to scale.

styled samples. From this analysis, *Tsstal* was identified and determined to have short-stamen-specific expression (Shore and Chafe, Unpublished data).

When the sequence obtained for this transcript was submitted to BLAST searches, it was found to have considerable sequence homology to self-incompatibility family proteins, related to known stigmatic SI genes in *Papaver* (*PrS*'s, described above) (Table 2). Further investigation also showed that *Tsstal* could be amplified only in short-styled individuals of all species of *Turnera* (Figure 5) and *Piriqueta* investigated (data not shown). The gene also could not be amplified in mutants harbouring deletions in the dominant *S*-allele, indicating that *Tsstal* is *S*-linked in *Turnera* (Shore and Chafe, Unpublished data). The relative location of *Tsstal* within the apparent *S*-locus has since been confirmed by BAC sequencing. Two additional short-specific genes have been identified in this way, and also on the same BAC as *Tsstal*. Investigations of their prospective functions in distyly in *Turnera* are ongoing. To this end, in the current study, the molecular signatures of natural selection on *S*-linked genes will be exploited in an effort identify gene candidates and/or gene regions that may determine distyly in *Turnera*.

1.10 NATURAL SELECTION IN SELF/NON-SELF RECOGNITION AND SEX DETERMINATION SYSTEMS

Self/non-self recognition systems are common to many organisms, and the various SI mechanisms described in angiosperms provide good examples of such processes (Grosberg and Hart 2000; Uyenoyama 2005). Other important cases of self/non-self recognition that have been well described in the literature include the major-histocompatibility complex (MHC) of jawed vertebrates (Takahata and Nei 1990), the complimentary sex determination (*csd*) locus of honey bees (Hassleman and Beye 2004; Cho et al. 2006), and mating-type determining loci in fungi (May et al. 1999). As this list suggests, most often, self/non-self recognition is associated with the definition of mating type or innate immune system responses (Grosberg and Hart; Sanabria et al. 2008). Indeed, many interesting parallels between innate immunity and SI systems in plants, in particular, have been described, suggesting that they may share common mechanisms, at least in some species exhibiting homomorphic sporophytic SI (Sanabria et al. 2008).

Importantly, similar selective pressures have been implicated in the maintenance of most self/non-self recognition systems (Klein et al. 1998; Richman 2000; Charlesworth 2006). As

these systems often necessitate the preservation of a diversity of alleles at a locus, balancing selection is of particular importance in these contexts (Takahata and Nei 1990; Richman and Kohn 1996; Richman et al. 1996; May et al. 1999; Richman and Kohn 1999, 2000; Schierup et al. 2001; Kamu and Charlesworth 2005; Cho et al. 2006; Castric and Vekemans 2007; Roux et al. 2012). Under balancing selection, allelic diversity may be maintained in a population either by overdominance (or heterozygote advantage) or by negative frequency dependent selection (Charlesworth 2006). In homomorphic SI systems, in particular, individuals harbouring rare *S*-alleles have been shown to acquire a reproductive advantage over their more common counterparts, as they will necessarily encounter a greater proportion of compatible mating opportunities within a given population (Wright 1939; Richman et al. 1996; Schierup et al. 2001; Kamu and Charlesworth 2005; Castric and Vekemans 2007; Richman and Kohn 1999, 2000). Further, in heterostylous populations, morph ratios, stigma-height polymorphism, and reproductive success have all been shown to be maintained by negative frequency dependent selection (Eckert et al. 1996; Thompson et al. 2003; Barrett et al. 2004). Though evidence is lacking at the genetic level, *S*-linked sequence polymorphism is expected to be maintained by balancing selection in distylous species, as well (Barrett and Shore 2008), with populations possibly being characterized by the presence of *S*-alleles or haplotypes with long- and short-specific functions.

The molecular signatures of balancing selection are readily identifiable by two key indicators (Richman and Kohn 1996; Richman 2000; Charlesworth 2006). Long-term balancing selection, as would be expected in the case of heterostylous systems (Barrett and Shore 2008), often results in increased sequence diversity at *S*-linked loci relative to unlinked regions of the genome, with the diversity of any particular region being inversely proportional to its distance from the gene(s) that are the actual subjects of selection (Richman 2000; Charlesworth 2006). In the context of self/non-self recognition systems, elevated levels of polymorphism are also thought to be preserved, in part, as a consequence of reduced recombination at the relevant loci (Though this is not always the case. Indeed, the hymenopteran *csd* is a good example of an instance in which the opposite is true; see Beye et al. 1999). Though recombination suppression has not been definitively shown to occur at the *S*-locus in many heterostylous taxa, it is often assumed, as selection against self-fertility is presumed to be strong (Awadalla and Charlesworth 1999; Vieira et al. 2003; Charlesworth 2006; Labonne et al. 2007).

The occurrence of trans-species polymorphism (TSP) is another consequence of long term balancing selection (Klein et al 1998; Richman 2000; Charlesworth 2006). TSP refers to the appearance of like-alleles in closely related species. These alleles might also be referred to as “shared” or “ancestral” in that they will appear to have originated prior to species divergence (Figure 3; Klein et al. 1998; Richman 2000; Charlesworth 2006). As a result of shared ancestry, in gene genealogies, alleles will not cluster by species, but rather by their “functional type” (Klein et al. 1998; Richman 2000; Charlesworth 2006). The occurrence of high levels of sequence diversity and TSP among *S*-alleles has been demonstrated in several plant taxa, representing a diversity of homomorphic SI systems, including the Brassicaceae (Dwyer et al. 1991; Nou et al. 1993; Boyes et al. 1997; Schierup et al. 2001; Kamu and Charlesworth 2005; Castric and Vekemans 2007) and the *S*-RNase system of the Solanaceae, Rosaceae, and Rubiaceae (Ioerger et al. 1990; Richman et al 1996; Ishimizu et al.1998; Richman and Kohn 1999 & 2000; Sutherland et al. 2008; Nowak 2011). However, no efforts to detect balancing selection on candidate genes have been undertaken in any heterostylous species, though this information would prove valuable in identifying the genes involved in determining these systems.

However, morph-specific genes have recently been described in many heterostylous taxa, including *Turnera* (Li et al. 2008; Yasui et al. 2012; Ushijima et al. 2012 & 2015; Nowak et al. 2015; Chafe et al. 2015; Shore and Chafe, Unpublished data), and the possibility that the presence or absence of a single gene may determine floral morph-identity has been suggested (Yasui et al. 2012). If this is the case, it might be expected that genes involved in determining heterostyly will exhibit patterns of sequence polymorphism more akin to what would be observed in male-specific genes typical of X/Y sex determination systems (King et al. 2007; Uyenoyama 2005). That is, they will exhibit the evolutionary dynamics expected of Y-linked genes. As is anticipated for *S*-loci in plants, the Y chromosome experiences considerable recombination suppression and, as a result, deleterious mutations are allowed to accumulate (via Muller’s ratchet) and become fixed (via genetic hitchhiking) due to strong (largely purifying) selection acting on nearby functional sex-determination genes (Muller 1964; Rice 1987; Liu et al. 2004; Uyenoyama 2005; King et al. 2007). The primary male-determining gene in mammals and marsupials, *SRY* (Sex Determining Region Y), for example, is Y-linked and has been shown to be under the influence of widespread negative or purifying selection in order to maintain

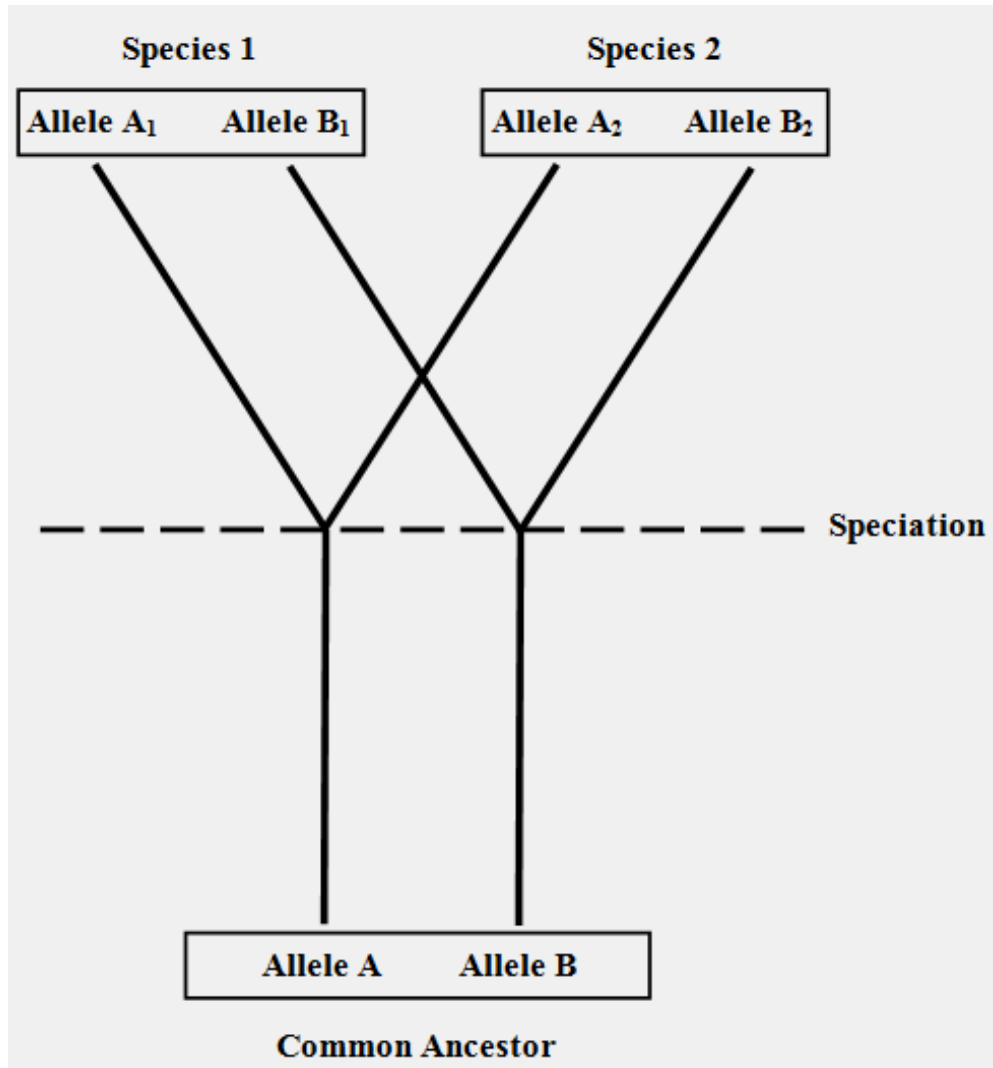


Figure 3: Graphical representation of trans-specific evolution. Due to speciation, two alleles (A and B) at a locus are transmitted to two daughter species from their common ancestor. Over time, the A and B alleles will accumulate their own unique species-specific differences. Consequently, the alleles are denoted as A_1/A_2 and B_1/B_2 depending on the daughter species in which they occur. However, as a result of common ancestry, allele A_1 will share more sequence similarities with allele A_2 than it will with allele B_1 (and vice versa), even though allele A_2 is present in a separate species (after Richman 2000 and Charlesworth 2006).

function within a variety of mammalian and marsupial orders (Graves 1998; Wang et al. 2002; King et al. 2007; Nowacka-Woszek and Switonski 2009; Murtagh et al. 2012). In any case, over time, and as a result of these processes, the Y chromosome has significantly diverged from its homologue, the X chromosome, and, other than at loci determining male-specific function, little conservation of the Y chromosome has been observed between species (Graves 1998; Liu et al. 2004; Uyenoyama 2005). Morph-specific genes in heterostylous systems might therefore be expected to show a molecular signature of strong purifying selection.

1.11 OBJECTIVES

Given that the genes for distyly are unknown, I undertake a series of exploratory analyses using various evolutionary genetic methods in an effort to identify the signatures of natural selection on known *S*-linked genes in *Turnera*. In light of theoretical expectations regarding the evolution of distyly (Charlesworth and Charlesworth 1979b; Lloyd and Webb 1992a & b) and its anticipated genetic underpinnings, my hypotheses and expectations for each set of analyses are as follows:

- 1) If any of the *S*-linked genes investigated are directly responsible for distyly and possess *S*- and *s*-specific alleles, I expect those alleles to show evidence of "trans-specific evolution" due to long-term balancing selection. That is, alleles of each lineage (*S* and *s*) are expected to have evolved prior to species divergence (Klein et al. 1998; Richman 2000; Charlesworth 2006). As such, a phylogenetic reconstruction of the evolutionary history of such a gene should reveal that all *S* forms of the gene are derived from an ancestral *S*, while the alternative *s* forms are derived from an ancestral *s*.
- 2) If alleles for causative *S*-linked genes are maintained by balancing selection (namely, negative frequency dependent selection), *S* and *s* alleles will be held in populations for longer periods of time than genes that occur elsewhere in the genome that are presumably not selectively maintained in this way. As a result, the genes determining distyly will have had greater time to accumulate neutral mutations and should show comparatively greater genetic diversity. In addition, this effect may be seen for genes very closely linked to the *S* locus, as well, with the

effect decreasing as one moves away from the causative genes as a result of recombination (Charlesworth 2006).

3) One recently discovered *S*-linked gene (*Tsstal*) is restricted to the short-styled morph and is tightly linked to the dominant *S* allele. The evolutionary dynamics of such a gene are expected to differ from those of other genes in the genome, given that it is present in only one copy (as opposed to 2) and in only half of the population of available individuals of any species/populations. If this gene directly determines distyly, it should behave as would a Y-linked sex-determining gene in mammals and marsupials, in that it will likely be maintained by strong purifying selection in order to maintain function (Graves 1998; Wang et al. 2002; King et al. 2007; Nowacka-Wozzuk and Switonski 2009; Murtagh et al. 2012).

In order to evaluate the above hypotheses and determine the forms of selection acting on known *S*-linked genes in *Turnera*, several approaches were adopted. Descriptive statistics and measures of average diversity – particularly nucleotide diversity (π), Watterson's θ , and indel diversity (π_1) – were calculated for each gene in order to assess how these measures were distributed across the *S*-locus and if any patterns could be identified. Selection on individual *S*-linked genes was examined directly using both inter-species (or “comparative”) dN/dS-based methods and population-level tests. By employing comparative dN/dS-based approaches, the molecular signatures of natural selection were investigated at the alignment-wide, lineage-specific, and codon-specific levels. Using these methods in addition to summary statistics, *S*-linked genes were compared to series of “random” genes located in other, unknown regions of the *Turnera* genome in order to determine if the nature of selection at the *S*-locus was in any way exceptional.

In part to support dN/dS-based tests, phylogenies were also constructed using the data. Phylogenies for each individual gene, also known as “gene genealogies”, were generated as well, in order to investigate the occurrence of balanced TSPs for *S*-linked genes, in particular.

Lastly, using subset of the data, representing a collection of closely related individuals, population genetic analyses were also applied, including McDonald-Kreitman tests and Tajima's *D* tests of neutrality.

2.0 METHODS

2.1 PLANT MATERIAL AND DNA EXTRACTION

Plant material was obtained from several individuals of the genus, *Turnera*. DNA samples were also acquired from individuals of the related genera, *Piriqueta* and *Erblichia*, as well as 16 additional individuals from a localized population of tetraploid *T. scabra* from the Dominican Republic (DROT). However, samples from these groups were only used to amplify *Tssta1*. A list of the plant materials used, along with their respective phenotypes and ploidy levels is presented in Table 1. It should be noted that, while the individual, TSH (*T. subulata*), has both short styles and short stamens, it behaves like a typical short-styled individual in terms of its incompatibility responses. All plants were kept under greenhouse conditions at York University, Toronto, Canada.

DNA was extracted from each of the individuals of interest using the Mini-CTAB method outlined by Labonne et al. (2009) and modified from Doyle and Doyle (1990). However, here, an intermediate sized flower bud (~ 5mm long) was removed for the purposes of DNA extraction, instead of two leaf discs (~ 6mg). Each bud was macerated with 100 μ L of CTAB isolation buffer [2% w/v CTAB (Sigma-Aldrich), 1.4 M NaCl, 0.2% v/v 2-mercaptoethanol, 20 mM EDTA, and 100 mM Tris-HCL (pH 8.0)] in an individual well of a porcelain spotting plate that had been kept cool on ice. An additional 400 μ L of CTAB buffer was then added to each sample prior to being transferred into individual 1.5 mL microcentrifuge tubes and incubated at 60°C for 20 minutes. Tubes were periodically inverted throughout the incubation period in order to ensure that their contents had been well mixed.

An equal volume of chloroform/isoamyl alcohol (24:1 v/v) was then added to each of the samples and each tube was briefly vortexed and centrifuged for 5 minutes at 15,000 x g. The supernatant was then removed and transferred into a new microcentrifuge tube to which $\frac{3}{4}$ of the volume of cold isopropanol was subsequently added in order to precipitate the DNA. After inverting the tubes a few times, they were then centrifuged for an additional 5 minutes at 15,000 x g in order to pellet out the DNA precipitate. The supernatant was removed from each of these tubes by quickly inverting them over a waste container.

Table 1: List of individual plants used in this study. The designated laboratory code name, genus, species, morph, and ploidy level of each individual is provided.

Genus	Individual	Species	Morph	Ploidy Level
<i>Erblichia</i>	EOB	<i>E. odorata</i> (Seemann)	Homostyle	Unknown
<i>Piriqueta</i>	PCARO	<i>P. cistoides ssp. caroliniana</i> (Walter)	Short-styled	Diploid
	DUART 1S	<i>P. duarteana</i> (Urban)	Short-styled	Diploid
	PMOR 137S	<i>P. morongii</i> (Rolfe)	Short-styled	Diploid
	PLIC	<i>P. plicata</i> (Urban)	Short-styled	Unknown
	PREV	<i>P. revoluta</i> (Arbo)	Short-styled	Unknown
	PNAN	<i>P. nanuzae</i> (Arbo)	Short-styled	Unknown
	PSAR	<i>P. sarae</i> (Arbo)	Short-styled	Unknown
	VIS	<i>P. viscosa</i> (Grisebach)	Short-styled	Diploid
<i>Turnera</i>	AUR	<i>T. aurelii</i> (Arbo)	Long Homostyle	Alloctoploid
	CHAM 4L	<i>T. chamaedrifolia</i> (Cambessedes)	Long-styled	Diploid
	CON 20S	<i>T. concinna</i> (Arbo)	Short-styled	Diploid
	CUN	<i>T. cuneiformis</i> (Poiret)	Long Homostyle	Alloctoploid
	DIF	<i>T. diffusa</i> (Schultes)	Long-styled	Diploid
	DEN 20S	<i>T. grandidentata</i> (Arbo)	Short-styled	Tetraploid
	DEN 54L	<i>T. grandidentata</i> (Arbo)	Long-styled	Tetraploid
	GRAN 9S	<i>T. grandiflora</i> (Urban)	Short-styled	Diploid
	TJ 29S	<i>T. joelii</i> (Arbo)	Short-styled	Diploid
	TJ 30L	<i>T. joelii</i> (Arbo)	Long-styled	Diploid
	KRAP 5S	<i>T. krapovickasii</i> (Arbo)	Short-styled	Diploid
	KRAP 12L	<i>T. krapovickasii</i> (Arbo)	Long-styled	Diploid
	OCC	<i>T. occidentalis</i> (Arbo and Shore)	Long Homostyle	Hexaploid
	TOC 139S	<i>T. oculata v. paucipilosa</i> (Story)	Homostyle	Unknown
	ORI	<i>T. orientalis</i> (Arbo)	Long Homostyle	Hexaploid
	PAN 2S	<i>T. panamensis</i> (Urban)	Short-styled	Diploid
	COLO	<i>T. subulata</i> (Smith) / <i>scabra</i> (Millspaugh)	Short-styled	Diploid
	MAN 601S	<i>T. scabra</i> (Millspaugh)	Short-styled	Diploid
	MAN 713L	<i>T. scabra</i> (Millspaugh)	Long-styled	Diploid
	DROT 41S	<i>T. scabra</i> (Millspaugh)	Short-styled	Tetraploid
	ES	<i>T. scabra</i> (Millspaugh)	Short-styled	Diploid
	MIDC 710S	<i>T. scabra</i> (Millspaugh)	Short-styled	Diploid
	F60SS	<i>T. subulata</i> (Smith)	Short-styled	Diploid
	D16L	<i>T. subulata</i> (Smith)	Long-styled	Diploid
	SL8 201S	<i>T. subulata</i> (Smith)	Short-styled	Diploid
	E 207S	<i>T. subulata</i> (Smith)	Short-styled	Tetraploid
	E 2L	<i>T. subulata</i> (Smith)	Long-styled	Tetraploid
	PA 4S	<i>T. subulata</i> (Smith)	Short-styled	Tetraploid
	TSH	<i>T. subulata</i> (Smith)	Short Homostyle	Tetraploid
	QUACO 1	<i>T. ulmifolia</i> (Linnaeus)	Homostyle	Hexaploid
	BAH	<i>T. ulmifolia v. acuta</i> (Urban)	Homostyle	Hexaploid
	VEL	<i>T. velutina</i> (Presl)	Long Homostyle	Hexaploid
	WED 2S	<i>T. weddelliana</i> (Urban and Rolfe)	Short-styled	Diploid

One hour after adding 500 μ L of wash solution (76% Ethanol, 10 mM ammonium acetate) to each of the pellet-containing tubes, the samples were then centrifuged for an additional 5 minutes at 15,000 x g. The supernatant was subsequently poured off and the pellet was allowed to air dry for at least 2 hours. Dry pellets were later resuspended in 300 μ L of TE [10 mM Tris-HCL (pH 8.0) and 1mM EDTA] so that they could be used in subsequent PCR reactions.

DNA extractions from all *Piriqueta* and *Erblichia* samples were performed by Mr. Paul Chafe.

2.2 GENES OF INTEREST AND EXON PREDICTION: KNOWN S-LINKED GENES

The recessive *s*-haplotype at the *S*-locus and portions of the dominant *S*-haplotype in *T. subulata* are known from previous sequencing projects (Labonne, 2011, PhD Thesis; Shore, Unpublished data). From this data, it was determined that some preservation of gene order between the *s*- and *S*-haplotypes could be assumed, though it has more recently been found that the dominant *S* contains additional genes that are not found on the apparent recessive *s*-haplotype (Shore and Chafe, Unpublished data). The previously described short-specific gene, *Tsstal*, is one such example (Shore and Chafe, Unpublished data).

Graphical representations of the *S*- and *s*-haplotypes, as they are currently known, are depicted in Figure 2. It should be noted, however, that the boundaries of the recessive *s*-haplotype were determined by deletion mapping that was completed with individuals that harboured deletions in the dominant *S*-haplotype. As a result, the recessive *s*-haplotype may not be properly defined. In addition, *S*-haplotype gene order, in some cases, may be inverted or otherwise arranged relative to what is shown.

All genes known to reside at or near the *S*-locus in *T. subulata* were considered in this investigation. A comprehensive list of these genes and their anticipated identities based on NCBI BLASTx (Altschul et al. 1997) and TAIR BLASTx (Lamesch et al. 2011) searches is shown in Table 2.

In lieu of sequencing whole genes, particular exons from each gene were selected and sequenced instead. In order to do this, exon positions were first predicted for each gene of

Table 2: S-linked genes. The predicted total gene and coding sequence sizes (bp), as well as the number of predicted exons is given for each entry. For each exon sequenced, the predicted exon size is also provided (bp). For each gene, the top BLASTX and TAIR BLASTX matches are shown. Based on research in *Arabidopsis*, the possible function of each gene is also detailed. Citations supporting these functions are provided. All listed genes are located in or linked to the putative *S*-locus region of *T. subulata*.

Gene Name	Gene Size (bp)	Size of Coding Sequence (bp)	Total # of Exons	Exon(s) Sequenced (#)	Exon Size (bp)	Top BLASTX Result**	Top TAIR BLASTX Result	Function in <i>Arabidopsis</i>	Citations
<i>APETALA2</i>	> 2038 *	1605*	10*	1	587	<i>Apetala 2 (AP2)-1 (Populus tomentosa; Sequence ID: AGM20693.1)</i>	<i>AP2 (Locus: AT4G36920)</i>	Encodes a member of the AP2/Ethylene Responsive Element Binding Protein (EREBP) class of transcription factors and has two AP2 DNA binding domains. It serves a diversity of functions, including roles in the specification of floral organ and meristem identity, as well as ovule and seed coat development.	Riechmann and Meyerowitz 1998, Wurschum et al. 2006, Ohto et al. 2009, Ripoll et al. 2011, Dinh et al. 2012
<i>Tsst1</i>	471	471	1	1	471	Self-incompatibility family protein (<i>Populus trichocarpa; Sequence ID: XP_002304696.1</i>)	<i>S-Protein Homologue 1 (SPH1) (Locus: AT4G16295)</i>	A member of a large gene family with significant homology to stigmatic SI genes in <i>Papaver rhoeas</i> . <i>SPH</i> 's have no known homologues in any non-plant model organisms, suggesting a plant-specific function.	Ride et al. 1999
<i>LEJ2</i>	3311	711	8	1	207	Cystathionine beta-synthase (CBS) domain-containing protein, CBSX1 (<i>Vitis vinifera; Sequence ID: XP_002283079.1</i>)	Loss of the Timing of Ethylene and Jasmonic Acid 2 (<i>LEJ2</i> or <i>CBSX1</i>) (Locus: AT4G36910)	<i>LEJ2</i> is a CBS domain containing protein and redox regulator of the ferredoxin-thioredoxin system. Specifically, it affects anther dehiscence and plant growth by regulating lignin polymerization and photosynthesis-related enzymes, respectively.	Yoo et al. 2012

Continued from previous page.

Gene Name	Gene Size (bp)	Size of Coding Sequence (bp)	Total # of Exons	Exon(s) Sequenced (#)	Exon Size (bp)	Top BLASTX Result**	Top TAIR BLASTX Result	Function in <i>Arabidopsis</i>	Citations
<i>AP2D</i>	638	634	1	1	634	AP2 domain-containing transcription factor family protein (<i>P. trichocarpa</i> ; Sequence ID: XP_002307277.1)	DREB and EAR Motif protein (<i>DEAR</i>) 2 (Locus: AT5G67190)	A member of the DREB subfamily A-5 of the Ethylene Response Factor (ERF)/AP2 transcription factor family. It contains one AP2 DNA binding domain, and has been implicated in regulating plant stress responses to dehydration and pathogen defense.	Liu et al. 1998; Tsutsui et al. 2009; Zhou et al. 2010
<i>RNABP</i>	3168	966	10	1	330	Putative RNA binding protein (<i>Ricinus communis</i> ; Sequence ID: XP_002534389.1)	RNA-binding family protein (<i>RNABP</i>) (Locus: AT2G42240)	RNA recognition motif containing protein, involved in post-transcriptional RNA processing, export, and/or stability.	Lorkovic and Barta 2002
<i>SCE1</i>	3225	489	5	3	153	Putative ubiquitin-conjugating enzyme E2 I (<i>R. communis</i> ; Sequence ID: XP_002534386.1)	Small ubiquitin-related modifier (SUMO) conjugating enzyme 1 (<i>SCE1</i>) (Locus: AT3G57870)	SCE1 directs the attachment of SUMO to target protein substrates via SUMO E3 ligases (a process referred to as sumoylation). Sumoylation is a key post-translation modification, necessary for the regulation of a variety of plant processes, including stress responses, pathogen defense, and flowering time.	Miura, Jin, and Hasegawa 2007; Miller et al. 2010
				4	114				
<i>FRA1</i>	8922	3087	25	5	188	Kinesin motor family protein (<i>P. trichocarpa</i> ; Sequence ID: XP_002310712.2)	Fragile Fibre 1 (<i>FRA1</i>) (Locus: AT5G47820)	Encodes a kinesin-like motor protein, involved in cell wall construction. Specifically, it facilitates the delivery of cell wall components, like pectin, by directing the movement of vesicles along cortical microtubules.	Zhong et al. 2002, Zhu and Dixit 2011, Zhu et al. 2015
				6	125				
				24	156				
				25	207				

Continued from previous page.

Gene Name	Gene Size (bp)	Size of Coding Sequence (bp)	Total # of Exons	Exon(s) Sequenced (#)	Exon Size (bp)	Top BLASTX Result**	Top TAIR BLASTX Result	Function in <i>Arabidopsis</i>	Citations
<i>LRRK</i>	3920	1947	2	1	1408	Leucine-rich repeat transmembrane protein kinase (<i>P. trichocarpa</i> ; Sequence ID: XP_002310125.2)	Leucine-rich repeat protein kinase family protein (<i>LRRK</i>) (Locus: AT5G67200)	Leucine-rich repeat receptor-like kinases (LRR-RLKs) have been implicated in the regulation of organogenesis, ovule development and embryogenesis, endosperm and pollen generation, and stress responses, among many other processes. Unlike animal receptor kinases, most LRR-RLKs phosphorylate serine/threonine residues.	Dievart and Clark 2003, Zhang et al. 2006
<i>IRX15L</i>	933	933	1	1	933	Predicted Irregular Xylem 15-Like (<i>IRX15-L</i>) protein (<i>P. euphratica</i> ; Sequence ID: XP_011008784.1)	<i>IRX15-L</i> (Locus: AT5G67210)	Encodes a Domain of Unknown Function 579 (DUF579)-containing protein that affects the synthesis of xylan, and is essential for its normal deposition in the primary and secondary cell walls of eudicots, grasses, and cereals. Like other DUF579 domain containing proteins, it is predicted to be a type II transmembrane protein.	Brown et al. 2011, Jensen et al. 2011, Hao and Mohnen 2014
<i>FSP</i>	3922	993	12	6	59	Putative flavonol synthase-like family protein (<i>FSP</i>) (<i>P. trichocarpa</i> ; Sequence ID: XP_002310710.1)	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein (Locus: AT3G50210)	The flavonol synthase family belongs to the larger 2-oxoglutarate-dependent dioxygenase family. They are involved in the synthesis of flavonoid pigments.	Owens et al. 2008
				7	166				

Continued from previous page.

Gene Name	Gene Size (bp)	Size of Coding Sequence (bp)	Total # of Exons	Exon(s) Sequenced (#)	Exon Size (bp)	Top BLASTX Result**	Top TAIR BLASTX Result	Function in <i>Arabidopsis</i>	Citations
NRFP	1678	1308	3	3	745	Putative tRNA-dihydrouridine synthase (<i>R. communis</i> ; Sequence ID: XP_002533266.1)	Flavin mononucleotide (FMN)-linked oxidoreductase superfamily protein (Locus: AT5G67220)	Dihydrouridine synthases (DUS) post-transcriptionally modify uridine residues located in the D-loop of tRNA molecules. The precise function of dihydrouridines in tRNAs is unknown, though it has been suggested that they may serve to stabilize the molecule, and thus afford it some degree of conformational flexibility. No plant-specific function has been identified. These proteins have previously been referred to as nitrogen regulation family proteins (nifR3).	Bishop et al. 2002, Ditt et al. 2006, Kasprzak, Czerwoniec, and Bujnicki 2012
WRKY	1578	1110	3	1	824	WRKY transcription factor 15 family protein (<i>P. trichocarpa</i> ; Sequence ID: XP_002310122.1)	WRKY DNA-binding Protein 7, (<i>WRKY7</i>) (Locus: AT4G24240)	WRKY7 is a transcription factor and negative regulator of salicylic acid-regulated defense responses to pathogen infection in <i>Arabidopsis</i> . It has also been shown to bind Ca ²⁺ -dependent calmodulin. However, the precise regulatory function of this binding is unknown.	Park et al. 2005; Kim, Fan, and Chen 2006; Agarwal, Reddy, and Chikara 2011

*Only a partial *APETALA2* sequence was obtained for *T. subulata* (D16L) after BAC Library experiments performed by Shore and Chafe (Unpublished data). What has been sequenced is 2038bp in total length (1015bp of which is apparently coding) and is predicted to consist of 7 exons. As the sequence appears to end in an intron, it is suspected that the gene is actually longer than this. This suspicion is also supported by BLAST search results. More recent RNAseq data from *T. subulata* also revealed apparent *APETALA2* transcripts, which are predicted to contain 10 exons, representing 1605bp of coding sequence (Shore and Chafe, Unpublished data).

** The top BLASTX results presented are the most informative results with the highest similarity score and lowest e-value. That is, named genes from more closely related species were prioritized.

interest using the online software, NetPlantGene Server (Hebsgaard et al. 1996). Final exon positions were determined by comparing the exon positions generated in all 3 reading frames to BLASTx results obtained for plants most closely related to *Turnera*. Diagrams of the predicted exon positions for all *S*-linked genes were produced using the online Gene Structure Display Server v. 2.0 and are depicted in Figure 4 (Hu et al. 2015).

Exons were selected for sequencing based on size, with larger exons being prioritized over smaller ones. In cases where genes appeared to be composed of a collection of small (<200 bp) exons, 2-4 exons were selected and sequenced. This latter scenario often necessitated sequencing through introns, though intron sequences were not included in any downstream analyses. Particular exons that were sequenced for each gene of interest are also identified in Figure 4. Their approximate sizes (bp) are given in Table 2.

2.3 CONTROL GENE SELECTION

Control genes were semi-randomly selected from a previously constructed RNA-seq data set (Shore and Chafe, Unpublished data). Each potential transcript was assigned a number and transcripts were then chosen using a random number generator in Excel 2007 (Microsoft). Using BLAST searches, potential gene identities were ascertained for these transcripts and the positions and sizes of their exons were determined as above. If any particular transcript was found to contain exons that were only very small in size (<200 bp), it was excluded from the pool of potential controls and another candidate was chosen. Eight suitable control genes were selected in this way. A comprehensive list of these genes and their anticipated identities as is given in Table 3. The particular exons that were chosen for sequencing and their sizes (bp) are also indicated.

2.4 PRIMER CONSTRUCTION

Primers were constructed using the free online software, Primer3 (Rozen and Skaletsky 2000). Primers were designed to locations just within the boundaries the exons chosen to be sequenced in order to obtain the largest PCR product possible. Primers were not designed outside of the identified exons in order to avoid difficulties in sequencing, which may arise as a result of

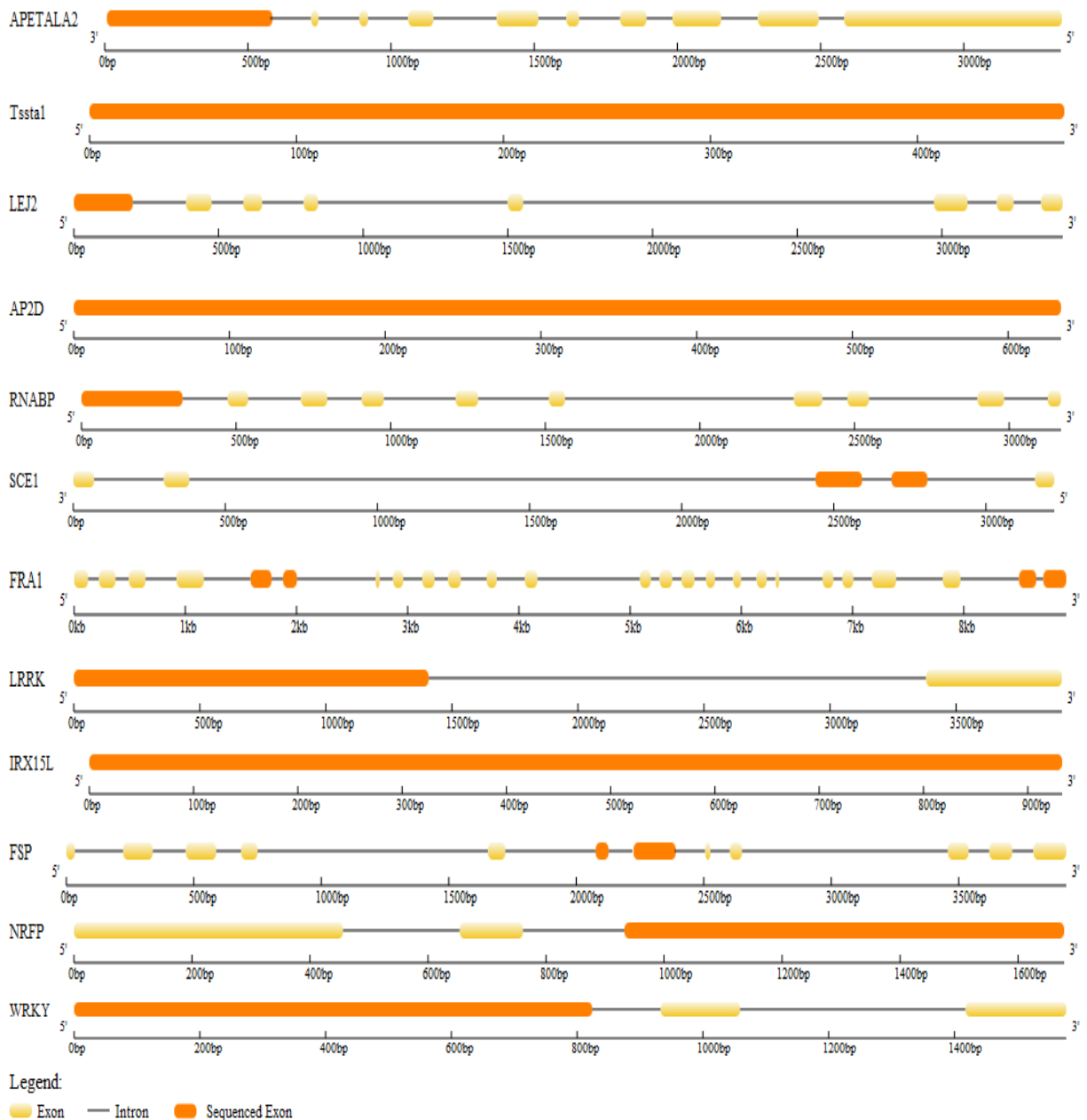


Figure 4: Schematics for all S-linked genes based on exon predictions from NetPlantGene Server and BLAST searches. Exons are depicted in yellow, while introns are shown as straight lines. Particular exons that were sequenced are displayed in orange. The size of each gene is represented by the corresponding scale shown below each diagram (in bp or kb). It should be noted that the orientation of *APETALA2* and *SCE1* is opposite, relative to the other genes. For *APATELA 2*, unknown intron sizes were determined from the most similar sequence from *Populus trichocarpa* based on BLAST searches. Diagrams were constructed using Gene Structure Display Server v. 2.0 (Hu et al. 2015).

Table 3: Control genes. The predicted total gene and coding sequence sizes (bp), as well as the number of predicted exons is given for each entry. For each exon sequenced, the predicted exon size is also provided (bp). For each gene, the top BLASTX and TAIR BLASTX matches are shown. Based on research in *Arabidopsis*, the possible function of each gene is also detailed. Citations supporting these functions are provided.

Gene Name	Total # of Exons*	Total Size (bp)*	Exon Sequenced (#)	Exon Size (bp)	Top BLASTX Result**	Top TAIR BLASTX Result	Function in <i>Arabidopsis</i>	Citation
<i>ECIP1</i>	3	2109	3	1783	MA3 domain-containing family protein (<i>Populus trichocarpa</i> ; Sequence ID: XP_002321660.1)	EIN2 C-terminus Interacting Protein 1 (<i>ECIP1</i>) (Locus: AT4G24800)	An MA3 domain containing protein. In the absence of ethylene, it degrades Ethylene Insensitive 2 (EIN2), a protein that is involved in plant responses to salt stress.	Qiao et al. 2012, Lei et al. 2011, Qiao et al. 2012
<i>GAUT3</i>	9	2043	6	946	Probable galacturonosyltransferase 3 isoform X1 (<i>Populus euphratica</i> ; Sequence ID: XP_011040798.1)	Galacturonosyltransferase 3 (<i>GAUT3</i>) (Locus: AT4G38270)	Encodes a protein with putative galacturonosyltransferase activity. In plants, galacturonosyltransferases are required for the synthesis of pectin. Unlike most GAUT proteins, which are often membrane-bound, GAUT3 contains an N-terminal signal peptide. However, its precise function is not known.	Sterling et al. 2006, Caffall et al. 2009
<i>GAUT1</i>	10	2022	4	588	Predicted polygalacturonate 4-alpha-galacturonosyltransferase-like isoform X2 (<i>P. euphratica</i> ; Sequence ID: XP_011002081.1)	Galacturonosyltransferase 1 (<i>GAUT1</i>) (Locus: AT3G61130)	In complex with GAUT7, GAUT1 synthesizes homogalacturonan, the most abundant plant pectin. GAUT1 is a type II transmembrane protein that is localized in the membrane of the golgi apparatus.	Sterling et al. 2006, Caffall et al. 2009, Atmodjo et al. 2011
<i>RNABP34</i>	6	1506	6	455	RNA recognition motif-containing family protein (<i>P. trichocarpa</i> ; Sequence ID: XP_006368382.1)	RNA-binding family protein (<i>RNABP</i>) (Locus: AT5G46840)	RNA recognition motif containing protein, involved in post-transcriptional RNA processing, export, and/or stability.	Lorkovic and Barta 2002

Continued from previous page.

Gene Name	Total # of Exons*	Total Size (bp)*	Exon Sequenced (#)	Exon Size (bp)	Top BLASTX Result**	Top TAIR BLASTX Result	Function in <i>Arabidopsis</i>	Citation
<i>FMO1</i>	5	1593	6	340	Probable flavin-containing monooxygenase 1 (<i>Populus euphratica</i> ; Sequence ID: XP_011045108.1)	Flavin-Dependent Monooxygenase 1 (<i>FMO1</i>) (Locus: AT1G19250)	<i>FMO1</i> is an essential component of acquired resistance to virulent pathogens and also promotes resistance and cell death, locally, at pathogen infection sites.	Bartsch et al. 2006, Mishina and Zeier 2006
<i>MBD8</i>	3	1575	3	395	DNA-binding family protein (<i>P. trichocarpa</i> ; Sequence ID: XP_002306571.2)	Methyl-CpG-Binding Domain 8 (<i>MBD8</i>) (Locus: AT1G22310)	<i>MBD8</i> has been shown to play a role in flowering time in some <i>A. thaliana</i> ecotypes. Though it contains a methyl-CpG-binding domain, it has not been shown to bind to methylated or unmethylated DNA in <i>Arabidopsis</i> .	Berg et al. 2003, Zemach and Grafi 2003, Strangeland et al. 2009
<i>UNKN</i>	7	1182	1	300	Uncharacterized protein LOC105107279 (<i>P. euphratica</i> ; Sequence ID: XP_010999458.1)	Unknown protein (Locus: AT2G38430)	Unknown	NA
<i>POFUT</i>	9	1650	8	332	Uncharacterized protein AT1G04910*** (<i>Populus euphratica</i> ; Sequence ID: XP_011031885.1)	O-fucosyl-transferase family protein (Locus: AT4G16650)	A member of an as of yet uncharacterised protein family in plants, related to GDP-fucose protein O-fucosyltransferases (<i>POFUTs</i>). These proteins add fucose sugars to serine and threonine residues that are located between the second and third conserved cysteins in Epidermal Growth Factor (EGF)-like repeats on Notch protein in <i>Drosophila</i> , humans, and other mammals. However, the Notch-signalling pathway does not exist in plants.	Chantha, Emerald, and Matton 2006; Chantha, Tebbji, and Matton 2007; Vodovar and Schweisguth 2008

* The total number of exons and total gene size for each entry is based on the size of full cds of the best match from *Arabidopsis*.

** The top BLASTX results presented are the most informative results with the highest similarity score and lowest e-value. That is, named genes from more closely related species were prioritized.

***AT1G04910 is also an O-fucosyltransferase family protein.

individuals being heterozygous for indels. Indeed, indels might be expected to more often occur in introns than in exons, presumably due to comparatively relaxed selective constraint in introns, in general (as shown in Shabalina and Kondrashov 1999 and Hughes and Yeager 1997, for example). All primers were subsequently synthesized by Integrated DNA Technologies (IDT, Coralville, Iowa, USA).

A compiled list of the primer pairs used for amplification, along with their sequences, melting temperatures (T_m), and anticipated PCR product sizes (bp) is provided in Appendix A, Table A1.

2.5 POLYMERASE CHAIN REACTION (PCR) AMPLIFICATION

PCR products were amplified for each gene/exon(s) of interest for the purposes of sequencing (or cloning, where required). Each PCR reaction was performed using 6 μ L of DNA (~50ng), 6 μ L of each primer (forward and reverse at 10pmol/ μ L), 12 μ L of ddH₂O, and 30 μ L of Quick-Load[®] *Taq* 2x Master Mix [10mM Tris-HCL (pH 8.6), 50 mM KCl, 1.5 mM MgCl₂, 50 units/mL *Taq* DNA Polymerase, 0.2 mM each dNTP, 5% glycerol, 0.08% IGEPAL[®] CA-630, 0.05% Tween[®] 20, 0.024% Orange G, 0.0025% Xylene Cyanol FF] (New England BioLabs, Ipswich, MA) for a 60 μ L total reaction volume. Reactions were carried out in 0.2 mL PCR tubes using a thermal cycler (Mastercycler Gradient, Eppendorf, Mississauga, ON) that had been programmed for an initial 2 minute denaturation step at 94°C. This was directly followed by an additional 30 seconds at 94°C, 30 seconds at between 50-61°C, and another 30 seconds to 1 minute at 72°C. For each particular reaction, the annealing temperature was varied depending on the T_m of the primers used. The length of the extension step was similarly adjusted according to the expected length of the PCR product (30 seconds/500 bp, approximately) (Appendix A, Table A1). This temperature sequence was cycled 35 times for the purposes of amplification. The final extension step was then performed at 72°C for an additional 5 minutes. After completion, samples were held at 4°C until removed from the thermal cycler. Samples were frozen at -20°C prior to sequencing.

PCR amplifications for all *Erblichia* and *Piriqueta* samples were performed by Mr. Paul Chafe.

2.6 PURIFICATION OF PCR PRODUCTS FOR CLONING

In some cases, sequences could not easily be obtained by sequencing PCR products, directly (usually due to the persistence of “shifted peaks” in the sequence chromatograms). As a result, some PCR products were, instead, cloned into a vector prior to sequencing. For this purpose, the relevant PCR products were purified using the EZ-10 Spin Column PCR Products Purification Kit according to the manufacturer’s instructions (BioBasic, Markham, Ontario). Prior to purification, PCR reactions were first verified by running 5 μL of the 60 μL total reaction volume out on a 0.8% agarose gel in order to ensure that the correct product had been amplified.

2.7 CLONING, BACTERIAL CULTURE, PLATING, AND PLASMID PURIFICATION

Cloning was achieved using the pGEM-T Easy Vector System (Promega, Madison, WI, USA). Each ligation reaction contained 5 μL of 2x Rapid Ligation Buffer (60mM Tris-HCl (pH 7.8), 20mM MgCl_2 , 20mM DTT, 2mM ATP, 10% polyethylene glycol), 1 μL pGEM[®]-T Easy Vector (50ng/ μL), 1 μL of T4 DNA ligase (3 Weiss units/ μL), and up to 3 μL of purified PCR product (~150ng). If less than 3 μL of PCR product was used, an appropriate amount of ddH₂O was added to the reaction in order to bring the total volume up to 10 μL . The volume of PCR product that was added to the ligation reaction was determined using a 3:1 insert to vector ratio. After mixing the contents of the ligation reaction by pipetting up and down, it was incubated at room temperature for one hour. In particularly difficult cases, reactions were instead incubated at 4°C overnight in order to ensure proper ligation.

For each PCR product, 5 μL of ligation reaction was added to 50 μL of JM109 High Efficiency Competent Cells (Promega, Madison, WI, USA) that had previously been stored at -80°C. The tubes were then gently flicked in order to combine their contents prior to being incubated for 20 minutes on ice. Cells were then heat-shocked in a 42°C water bath for 45-50 seconds before being immediately returned to the ice for an additional 2 minute incubation. Next, 200 μL of Lysogeny broth (LB) (10 g Tryptone, 10 g NaCl, and 5 g Yeast Extract/L, pH 7.5) was added to each tube of transformed cells. Tubes were then incubated for 1 hour at 37°C. For each transformation, 70-100 μL of culture was plated out on individual X-Gal plates (70 $\mu\text{g}/\text{mL}$ X-

Gal, 80 μ M IPTG, 1 μ /mL carbenicillin). These plates were subsequently incubated overnight at 37°C.

For each transformation, 12-20 individual clones were inoculated in culture tubes containing 3.5 mL of LB medium (10 g tryptone, 5 g yeast extract, and 10 g NaCl/L, pH 7.5), to which 3.5 μ L of carbenicillin (25 μ g/mL) had been added. Bacterial cultures were then grown overnight at 37°C in a shaking incubator (250 rpm).

Following overnight incubation, plasmid purification was completed using the E7-10 Spin Column Plasmid DNA Minipreps Kit in accordance with the manufacturer's protocol (BioBasic, Markham, Ontario). To ensure that the expected PCR product had been successfully cloned and inserted into the plasmid, its presence was confirmed by PCR, as described above, using a 1/100 dilution of purified plasmid and a 20 μ L total reaction volume.

2.8 SEQUENCING AND SEQUENCE ASSEMBLY

All sequencing was performed by the McGill University and G enome Qu ebec Innovation Centre (Montr eal, QC) using Applied Biosystem's 3730xl DNA Analyzer technology. For each PCR product, two 25 μ L samples of unpurified PCR reaction mixture were submitted for sequencing in both the forward and reverse directions. For each purified plasmid to be sequenced, two 10 μ L samples were similarly submitted. For each PCR sample, 15 μ L of each sequencing primer (5 pmol/ μ L) was also provided. For plasmids, however, the common primers M13F (5' - GTAAAACGACGGCCAGT - 3') and M13R (5' - GGAAACAGCTATGACCATG - 3') were supplied by the sequencing company. A list of sequencing primers is given in Appendix A, Table A2. All samples were submitted in 96-well microtiter plates or individual PCR tubes, depending on how many samples were included in the submission. Samples were kept cold on ice while in transit. Resequencing was performed as necessary.

Despite best efforts, sequences could not be obtained for the following genes (individuals): *RNABP* (KRAP 5S), *LEJ2* (DIF), *FM01* (WED 2S and DIF), and *MBD8* (CHAM 4L, WED 2S, and DIF). Also, gaps in the sequences obtained for ES and SL8 201S for *LRRK* and DROT 41S and MIDC 710S for *RNABP* could not be resolved due to persistent double peaks present in the chromatograms obtained for these samples.

Sequence chromatograms were returned by the McGill University and Génome Québec Innovation Centre (Montréal, QC) via Nanuq, a web-based data-access application. Chromatograms were subsequently downloaded and uploaded into Sequencher v. 5.0 (Gene Codes Corporation, Ann Arbor, MI, USA). For each sequence, poor quality data was first trimmed using the default criteria. Sequences were then further trimmed by sight. Where necessary, vector sequence was also excised. For each PCR product, representative sequences were assembled automatically into contigs. Once assembled, heterozygous sites were identified for each contig by calling secondary base peaks. A minimum lower peak height cut-off value of 35% relative to the height of the larger peak was employed for this task. Contigs were also checked manually in order to verify the accuracy of secondary base peak-calling, as well as to check for the presence of additional double peaks in the chromatograms.

Final sequences were exported from Sequencher in FASTA format, in preparation for subsequent analyses. Select sequences for each gene were also submitted to BLASTx searches in order to identify possible conserved domains using NCBI's Conserved Domain Database (Marchler-Bauer et al. 2015).

2.9 SEQUENCE ALIGNMENT

For each gene/exon(s) of interest, representative sequences were initially aligned in MEGA 6.06 (Tamura et al. 2013) using the MUSCLE algorithm and default settings (Edgar 2004). Alignments were further adjusted by sight, where appropriate. Intron sequence, where it existed, was removed, and alignments were further trimmed in order to maintain the proper reading frame. Stop codons were also removed, so as to ensure that only protein coding information was represented. In addition to constructing alignments for each individual gene, four additional concatenated alignments were also created: 1) one that included sequence information for all genes of interest across all taxa, 2) another identical alignment where only short-styled individuals were represented, 3) an alignment of all *S*-linked sequence information, and 4) an alignment that included all control sequences. Final alignments were exported from MEGA 6.06 in FASTA format in preparation for subsequent analyses. Final alignment figures were edited and produced using CHROMA (Goodstadt and Ponting 2001).

2.10 DESCRIPTIVE STATISTICS

Descriptive statistics for each alignment, including alignment length (bp), and the number of conserved, variable, parsimony informative, and singleton sites were generated in MEGA v. 6.06 (Tamura et al. 2013). The numbers of duplicate or identical sequences in each alignment were determined in HYPHY by selecting the “Data Operations” icon in the “Data set” window of the graphical user interface (GUI). Identical sequences were then highlighted in the alignment after selecting the “Identical Sequences” option (Pond, Frost, and Muse 2005).

Descriptive nucleotide and indel polymorphism statistics were determined for each alignment using DNAsp v. 5 (Librado and Rozas 2009). Specifically, estimates of the average pair-wise nucleotide diversity (π) (Nei 1987) and Watterson’s θ (Watterson 1975) were computed, in addition to the proportions of polymorphic sites and synonymous and nonsynonymous mutations, the number of indel events, average indel length (bp) and average indel diversity (π_1). To estimate these quantities, the individuals, DIF, WED 2S, CHAM 4L, KRAP 5S, and MAN 601S, were excluded from each alignment in order to obtain a balanced data set, thus allowing for easier comparison between genes. For *Tsstal*, this was not possible as this gene is only represented in short-styled individuals. Instead, WED2S was eliminated from this alignment and KRAP 5S was retained, in an effort to make the results more comparable to those of other genes. Accordingly, the final “balanced” *Tsstal* alignment contained genetic information from 15 individuals and 7 species (*T. subulata*, *T. scabra*, *T. grandidentata*, *T. krapovickasii*, *T. joelii*, and *T. panamensis*) while all other alignments were represented by 17 individuals of the same 7 species.

Because DNAsp does not accept alignments containing ambiguous bases for most analyses, alignments were first submitted to PHASE v. 2.1, as implemented in DNAsp v. 5, for haplotype reconstruction (Stephens et al. 2001; Stephens and Donnelly 2003; Librado and Rozas 2009). PHASE employs a coalescent-based Bayesian approach to reconstructing haplotypes. That is, prior expectations derived from population genetic and coalescent theory are used to inform haplotype reconstruction given a specific alignment of linked loci (Stephens et al. 2001). With these methods, haplotype reconstruction was completed using 100000 replicates with a burn-in of 10000 and a thinning interval of 10.

Tetraploid individuals included in each alignment (E 2L, E 207S, TSH, PA 4S, and DROT 41S) were treated as though they were diploid and phased as described above. For each of these individuals, one randomly selected haplotype was copied and entered into the alignment an additional two times. Consequently, each tetraploid individual was represented by 4 haplotypes for each gene, 3 of which were identical. The premise of this is based on the assumption that most of the observed polymorphism would be at relatively low frequency, and thus, most haplotypes might be expected to be seldom represented relative to a few that are very common. At sites where individuals appeared to be characterized by more than two bases (as was the case at one site in each of the following alignments (individuals): *IRX15L* (TSH), *NRFP3* (TSH, DEN 20S, DEN 54L), and *ECIP1* (PA 4S). In all cases, the individual(s) in question were tetraploid.), the most common two bases at that position were selected to represent that individual in order to ensure conservative estimates. Gaps resulting from sequencing difficulties in the *LRRK* (ES and SL8 201S) and *RNABP* (DROT 41S and MIDC 710S) alignments were automatically resolved by PHASE v 2.1. Base positions containing gaps at the beginnings and ends of alignments were trimmed out prior to analysis as they did not represent indel events, but, rather, were the result of differences in the lengths of sequences obtained for particular individuals at particular loci. As a result, the alignments used in the generation of these statistics were, in some cases, somewhat shorter than those that were submitted to other analyses.

For measures of indel polymorphism, the “multiallelic” option was selected in DNAsp v. 5. Standard deviations were also calculated for nucleotide diversity estimates (π). Aside from in the calculation of indel polymorphism statistics, positions in the alignment containing gaps were not considered. In addition, Mann-Whitney-U tests (Mann and Whitney 1947) were performed in STATA 12 (StataCorp. 2011) in order to determine if *S*-linked and control gene diversity and polymorphism estimates were significantly different ($p < 0.05$), on average.

2.11 PHYLOGENETIC ANALYSES AND GENE GENEALOGIES

Phylogenetic analysis was completed in MEGA v. 6.06 using Maximum Likelihood methods (Tamura et al. 2013). Phylogenies were produced using the data from each individual gene, as well as for each of four larger concatenated alignments. These nucleotide alignments were then translated and submitted to additional phylogenetic analyses.

Nucleotide and amino acid substitution models for each analysis were determined in MEGA v. 6.06 using its endogenous model selection protocol and the default settings. Substitution models employed in the production of each phylogeny are given in Table 4. Best fitting modes were determined to be those that had obtained the lowest Bayesian Information Criterion (BIC) value.

For each analysis, sites with less than 95% coverage were eliminated from the final data set. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position in the alignment. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with superior log likelihood value. Branch supports for each tree were determined using 1000 bootstrap replicates.

All trees were left unrooted, as HYPHY automatically unroots rooted phylogenies prior to performing most analyses (including all of those described here) (Pond, Frost, and Muse 2005). Final phylogenies were edited for aesthetic purposes in FigTree v. 1.4.2 (Rambaut 2007).

Alignments for individual genes were further investigated for the presence of trans-species polymorphisms (TSPs). Potential TSPs were identified by looking for nucleotide sites at which long-styled individuals were homozygous for a particular base, while short-styled individuals were heterozygous at that site. Ambiguous base codes at these heterozygous sites in short-styled individuals were then replaced with the apparently short-specific base. These alignments were subsequently submitted to phylogenetic analysis in MEGA 6.06 using the previously determined nucleotide substitution models (Tamura et al. 2013).

2.12 SELECTION ANALYSES

Species-level selection analyses were completed in HYPHY (Pond, Frost, and Muse 2005). In most cases, the nucleotide-based phylogeny that had been constructed using the total data was used in conjunction with the relevant nucleotide alignment (Figure 8). Because *Tssta1* could only be amplified in short-styled individuals, this tree was not suitable for selection analyses of this gene as more individuals were represented in the tree than were included in the alignment. As a result, a tree that was constructed using the total nucleotide data from a reduced number of

Table 4: Nucleotide and amino acid substitution models used for phylogeny reconstruction in MEGA 6.06.

Tests of model fit were completed in Mega 6.06 using maximum likelihood methods. Models that obtained the lowest Bayesian Information Criterion (BIC) values were selected and used for the relevant phylogenetic analysis.

Alignment	Model Selection for Phylogeny Reconstruction in MEGA 6.06			
	Nucleotide Substitution Model**	BIC	Amino Acid Substitution Model***	BIC
Total Data for All Genes	T92 + G	27879.505	JTT + G	14892.732
Total Data for All Genes (Reduced Taxa)*	T92 + G	24127.889	JTT + G	13501.130
APETALA2	HKY	1826.222	WAG	982.170
Tsstal (Total data)*	K2 + G	3141.331	JTT	1965.076
Tsstal	K2	1797.989	JTT	1125.493
LEJ2	JC	725.049	JTT	485.272
AP2D	JC + G	1822.265	JTT	1029.823
RNABP	JC	1457.225	JTT	852.204
SCE1	JC	1036.530	RTREV	674.148
FRA1	HKY	2164.949	RTREV	1305.091
LRRK	JC	1408.200	JTT	749.665
IRX15L	HKY + G	3308.918	JTT	1950.214
FSP	JC	747.241	CPREV	502.241
NRFP	HKY	2425.672	JTT	1451.315
WRKY	K2 + G	3317.302	JTT + G	2080.602
Total S-linked Data	K2 + G	17122.322	JTT + G	9098.434
ECIP1	HKY + G	3478.610	LG	1920.685
GAUT3	T92	2966.538	LG	1834.846
GAUT1	HKY	1901.796	LG	1145.597
RNABP34	HKY	1861.044	Dayhoff	1172.746
FMO1	HKY	1335.266	JTT	884.949
MBD8	JC	1571.346	JTT	1098.829
UNKN	JC	1204.912	JTT	769.245
POFUT	HKY	1492.680	JTT	909.208
Total Control Data	T92 + G	11064.523	JTT + G	6181.410

* The phylogenetic trees obtained using these alignments were used for the analysis of *Tsstal* in HYPHY only. All other analyses were performed using the “Total Data for All Genes” nucleotide-based tree.

**Abbreviations: T92 = Tamura 3-Parameter Model (1992); JC = Jukes and Cantor Model (1969); HKY = Hasegawa-Kishino-Yano Model (1985); K2 = Kimura 2-Parameter Model (1980); G = Gamma distributed rates among sites.

*** Abbreviations: JTT = Jones-Taylor-Thornton Model (1992); Dayhoff = Dayhoff Model (Dayhoff et al. 1978); LG = Le and Gascuel Model (2008); CPREV = General Reversible Chloroplast Model (Adachi et al. 2000); RTREV = General Reverse Transcriptase Model (Dimmic et al. 2002); WAG = Whelan and Goldman Model (2001); G = Gamma distributed rates among sites.

individuals was employed (Figure 9). And, further, because *Tsstal* was of special interest in this study, it was amplified in a greater number of individuals, including representatives of *Piriqueta* and *Erblichia*. In order to analyse these data in HYPHY, the tree constructed using the total *Tsstal* data alone was used (Figure 10). All analyses in HYPHY were completed using the “Universal Genetic Code” option.

Additional McDonald-Kreitman Tests (MKTs) and Tajima’s D tests of neutrality were performed in DNAsp v. 5 using a subset of the data, as described in section 2.12.6 (Librado and Rozas 2009).

2.12.1 NUCLEOTIDE AND CODON SUBSTITUTION MODEL SELECTION

Unlike substitution model selection procedures employed by MEGA, which consider only standard, or named, nucleotide substitution models, HYPHY evaluates all 203 possible models when estimating nucleotide biases in alignments (Pond, Frost, and Muse 2005). As a result, for the purposes of subsequent analyses in HYPHY, best fitting nucleotide substitution models were again determined for each alignment using HYPHY’s endogenous model selection procedure (the batch file, NucModelCompare.bf). Here, the best fitting model was defined as that which most accurately described the data but with the fewest parameters according to Akaike’s Information Criterion (AIC). Using this procedure, models were fitted using global parameters. Branch length estimates were reused (as opposed to being re-estimated) when assessing each model. A model rejection level of 0.0002 was selected in order to account for increases in error rate that arise as the result of multiple tests. Nucleotide substitution models selected for all alignments are given in Table 5, along with their AIC values.

In HYPHY, some analyses require the selection of a codon substitution model, two of which are available in HYPHY: MG94 (Muse and Gaut 1994) and GY94 (Goldman and Yang 1994). In cases where one is not explicitly chosen, the MG94 model is employed by default. In order to determine which model was more appropriate for each alignment, likelihood ratio tests were performed to compare them. The results of these tests are also given in Table 5, including their respective LRT statistics and p-values. In all cases the MG94 model was preferred, but not always in a statistically significant way. As a result, in all analyses where a codon substitution

model was required, MG94 was used. GY94 is the model popularly employed in the PAML package (Yang 1997), and differs from MG94 in only one major respect: MG94 allows for variation in dS while GY94 sets dS, the synonymous evolutionary rate, to be equal to 1 (Pond, Poon, and Frost 2009). While there are other minor differences between the models, they have been found to have only minor effects on the estimation of evolutionary rates (Pond, Poon, and Frost 2009).

2.12.2 GLOBAL dN/dS ESTIMATES

For each alignment, global, or alignment-wide, dN/dS (dN = nonsynonymous substitution rate; dS = synonymous substitution rate) ratios were estimated as described in Pond, Poon and Frost (2009) using the AnalyzeCodonData.bf batch file located in the Basic Analyses submenu. The MG94CUSTOM codon model was selected in all cases. This option allows the user to cross the MG94 codon model with the appropriate nucleotide substitution model (indicated in Table 5). When prompted, the global model fit option was selected.

The appropriate tree and alignment combination were then submitted to this analysis. Rather than re-estimating branch length parameters for each analysis, they were assumed to be proportional to that of the input tree.

In order to determine 95% confidence intervals for each estimate, the dN/dS value was selected in the likelihood parameter table, and the “Covariance, Sampler, and CI” option was chosen from the Likelihood drop-down list. “Asymptotic Normal (finer)” was selected as the confidence interval estimation method and the upper and lower bound global dN/dS estimates were given in the resultant table.

An additional test was performed for each alignment in order to determine if the global dN/dS estimate was statistically different from 1, which would be expected if a sequence were evolving neutrally, on average. To do this, likelihood ratio tests (LRTs) were performed in order to compare the fit of a model where dN and dS values were constrained to be equal (the null hypothesis) to another where all parameters were estimated independently from the data (the alternate hypothesis). In HYPHY, this procedure is accomplished using the LRT.bf batch file that is available in the “Miscellaneous” sub-menu as part of the larger Phylohandbook.bf batch

Table 5: Nucleotide and codon substitution model selection for analyses in HYPHY. Nucleotide substitution models were selected for each alignment using the endogenous model selection procedure available for HYPHY. Nucleotide models are presented in terms of the 6 character model designations. For models that are named in the literature, the common model name also appears in parentheses. The Akaike Information Criterion (AIC) value for the best fitting model is provided. Further, some HYPHY analyses require the selection of a codon substitution model. Two codon substitution models are used by HYPHY: the GY94 and MG94. A likelihood ratio test (LRT) was performed in order to compare the fit of the models to each alignment. In all cases, the MG94 model was preferred, though, its fit was not always found to be significantly different from that of the GY94 model. Significant p-values are shown in bold.

Alignment	Nucleotide Substitution Model Selection		Codon Model Selection		
	Nucleotide Substitution Model*	AIC	GY94 vs. MG94 LRT		
			Preferred Codon Model*	LRT Statistic	p-value
<i>APETALA2</i>	010010 (HKY)	2403.078	MG94	9.211	0.230
<i>Tssta1</i> (Total Data)	010110	3377.518	MG94	88.702	0.000
<i>Tssta1</i>	010110	2006.052	MG94	38.520	0.000
<i>LEJ2</i>	000000 (F81)	1486.252	MG94	10.466	0.140
<i>AP2D</i>	000121	3648.502	MG94	97.950	0.000
<i>RNABP</i>	000000 (F81)	2557.446	MG94	64.792	0.010
<i>SCE1</i>	000000 (F81)	768.749	MG94	5.394	0.090
<i>FRA1</i>	011010	2380.178	MG94	10.225	0.080
<i>LRRK</i>	010111	6902.758	MG94	125.011	0.000
<i>IRX15L</i>	011020	3775.998	MG94	19.028	0.060
<i>FSP</i>	000000 (F81)	1006.646	MG94	9.022	0.040
<i>NRFP</i>	010010 (HKY)	2670.647	MG94	13.956	0.000
<i>WRKY</i>	010010 (HKY)	3524.644	MG94	6.488	0.320
Total S-locus Data	010010 (HKY)	32759.508	MG94	84.322	0.000
<i>ECIP1</i>	010010 (HKY)	3444.289	MG94	51.239	0.000
<i>GAUT3</i>	010010 (HKY)	2774.875	MG94	39.713	0.000
<i>GAUT1</i>	010110	1834.421	MG94	21.773	0.000
<i>RNABP34</i>	010010 (HKY)	1664.739	MG94	31.030	0.000
<i>FMO1</i>	000000 (F81)	1288.990	MG94	22.520	0.020
<i>MBD8</i>	001102	1490.452	MG94	10.726	0.060
<i>UNKN</i>	011110	1415.236	MG94	15.500	0.040
<i>POFUT</i>	010000	1357.795	MG94	31.807	0.000
Total Control Data	010010 (HKY)	15004.412	MG94	153.193	0.000

*Nucleotide substitution model abbreviations: HKY = Hasegawa-Kishino-Yano Model (1985) ; F81 = (Felsenstein 1981).

** Codon substitution model abbreviations: MG94 = Muse and Gaut 1994; GY94 = Goldman and Yang 1994

file. As with the above, the MG94 codon substitution model, crossed with the relevant nucleotide substitution model, was used. When prompted, the “Custom” alignment option was selected. This option allows for the submission of user-supplied alignments and phylogenetic trees. After hypothesis testing was completed, LRT and p-values for each test were outputted, along with the number of estimated parameters and log likelihood values for each model. Significant p-values (<0.05) indicate that the independent model fit is superior to that of the constrained model, suggesting that dN/dS is significantly different from 1 for the alignment in question.

2.12.3 COMPARING EVOLUTIONARY RATES

To compare evolutionary rates between alignments, the `dNdSDistributionComparison.bf` batch file was employed, as described in Pond, Poon, and Frost (2009). In HYPHY, this procedure is available in the “Codon Selection Analyses” submenu. This batch file completes four LRTs in order to determine if two alignments are different from one another with respect to the distribution of particular rate parameters. Unlike the previously described LRT, which was concerned only with alignment-wide dN/dS , these tests consider dN/dS at the level of individual codons by assigning each site in the alignment to one of four rate classes (one positive, one neutral, and two negative). The value of dN/dS that is assigned to each of these rate classes is determined from the data, itself.

The first of the four tests attempts to determine if all estimated evolutionary rate parameters may be shared by the two alignments. These parameters include the estimation of dN , dS , and dN/dS for each rate class, as well as the proportion of sites within each alignment that belong to each of the four rate classes. First, a model where all parameters were estimated independently for each alignment is fitted to the data. A model where the values of all parameters are constrained so as to be shared by both alignments is then fitted and models are compared using an LRT.

The remaining three LRTs are aimed at determining if there are any differences between the alignments in terms of positive (or diversifying) selection, in particular. These tests were performed by fitting a model that constrained the value of dN/dS for the positively selected rate class, the proportion of positively selected sites, or both values simultaneously, so that they were

shared by both alignments. These three models were then compared, in turn, to the independent model described above using an LRT.

For these analyses, two alignments must be submitted so that they may be compared. Two phylogenies must also be submitted; however, the same tree may be used to represent both alignments. In all cases, branch lengths were estimated using a codon substitution model (MG94) with the relevant nucleotide substitution model correction. Initial tests were run with default starting values for all rate parameter estimates. Due to the complexity of the models being fitted, each comparison was completed a second time with random starting values for rate parameters in order to ensure that rate estimates converged on similar results.

This procedure was used to perform 3 different alignment comparisons: 1) the entire concatenated alignment for all *S*-linked genes to the total control alignment; 2) the *Tsstal* alignment to an alignment of the remaining *S*-linked genes; and 3) the *AP2D* alignment to an alignment of all other *S*-linked genes.

2.12.4 LOCAL dN/dS PROCEDURES

Strong alignment-wide signatures of negative selection can dilute signals of positive selection, especially if positive selection has affected only particular lineages or specific codon-sites in an alignment (Pond, Poon, and Frost 2009). And, moreover, it may not be safe to assume that evolutionary rates (measured here by dN/dS) have remained constant over time. In order to investigate this possibility in the current context, several tests of lineage-specific selection were completed using sequence information from *S*-linked genes. For each *S*-linked alignment, global and local (or branch-wise) dN/dS ratio models were compared in the HYPHY, as described in Pond, Poon, and Muse (2009). This procedure employs an LRT to compare models where all rate parameters are constrained so as to be equal across branches in the phylogeny (the null hypothesis) to another where rates are allowed to vary across branches (the alternate hypothesis). After hypothesis testing is completed, the program outputs LRT statistic values, degrees of freedom, and p-values for each test, as well as the number of parameters estimated under each model. Significant p-values (<0.05) indicate that the local model fit is superior to that of the global model, suggesting that dN/dS rates vary across lineages.

If positive selection has occurred in only very few lineages or on only a small selection of codon sites, evidence of diversifying selection in an alignment can be subtle, and thus very difficult to detect (Murrell et al. 2015). As a result, individual *S*-linked gene alignments were also investigated for incidences of episodic diversifying selection across individual lineages and codon sites using a Branch-site Unrestricted Statistical Test of Episodic Selection (BUSTED) (Murrell et al. 2015). BUSTED analysis can be completed in HYPHY using the BUSTED.bf batch file that is available in the “Positive Selection” submenu of the “Standard Analyses” file. All branches were included in the analysis, as there were no *a priori* reasons to believe that any particular lineage would have been disproportionately affected by positive selection. LRTs are then performed, comparing models where dN/dS is constrained so as to be less than 1 across all sites and all lineages to a model where all parameters are estimated independently. If no evidence of positive selection is obtained under the null (constrained) model (i.e.: the proportion of sites with dN/dS>1 across a subset of lineages was essentially zero), then an LRT is not performed, as it is already apparent that the null cannot be rejected.

In order to determine if any particular branches in the phylogeny have experienced episodic positive or diversifying selection on any of the *S*-linked genes, additional tests were performed using the Adaptive Branch-site Random Effects Likelihood (aBS-REL) approach (Smith et al. 2014; Pond et al. 2011). This was completed in HYPHY using the BranchSiteREL.bf batch file that is available in the “Positive Selection” submenu of the “Standard Analyses” file. For this analysis, the option that allows synonymous substitution rates to vary from site to site was selected and all possible branches were chosen to be tested. As before, the MG94 codon model was employed and locally fit so as to compute branch-wise estimates of dN/dS. This procedure first sorts branches by length, longest to shortest, and subsequently defines rate-classes for each branch. The number of rate-classes defined is limited by AIC value and additional classes are only added if it results in an AIC improvement (i.e.: if it is reduced). The proportion of sites belonging to each rate-class, or its weight, is also computed. Each branch is then tested in order to determine if there is a proportion of sites with dN/dS>1 whose removal from the data set would result in a significant reduction in the log-likelihood value obtained for that branch. That is, the locally fit MG94 model is compared to one where dN/dS is constrained to be less than or equal to 1 using an LRT. If no rate classes with dN/dS>1 are defined for a particular branch, then this test is not performed. Once these tests are complete, the Holm-Bonferroni multiple testing

correction is applied in order to correct for multiple tests. Corrected p-values for each branch, where tests are completed, are then outputted by HYPHY.

2.12.5 SITE-BY-SITE SELECTION ANALYSES

When positive selection affects only a small proportion of sites, it can be difficult to detect, especially when negative selection dominates an alignment (Pond, Poon, and Muse 2009). In order to contend with this possibility, tests of site-by-site variation in evolutionary rates were performed for all *S*-linked sequences. To do this, several methods were employed, including: Single-likelihood Ancestor Counting (SLAC) (originated by Suzuki and Gojobori 1999 and modified by Pond and Frost 2005), Fixed Effects Likelihood (FEL) (Pond and Frost 2005), and Fast Unconstrained Bayesian Approximation (FUBAR) (Murrell et al. 2013). In order to detect episodic *diversifying* selection on particular codon sites, specifically, the Mixed Effects Model of Evolution (MEME) method was employed (Murrell et al. 2012). While each method has its particular strengths, as will be discussed, it has been suggested that several methods be used in concert for an integrative approach to site-by-site selection detection – particularly when sequence data is limiting, as may be the case here (Pond, Poon, and Frost 2009).

Batch files for performing SLAC, FEL, and MEME analyses are available in the “Positive Selection” submenu of the “Standard Analyses” file, and are part of the larger QuickSelectionDetection.bf batch file. As with all other analyses, an alignment and a representative phylogenetic tree are submitted, and the best fitting nucleotide substitution model is selected. For these tests, codon models are approximated using branch lengths from the input tree and nucleotide substitution rate parameter estimates indicated by the selected nucleotide substitution model (Pond, Poon, and Frost 2009; Pond and Frost 2005). During the optimization of the approximate codon model, the global parameter, dN/dS, is estimated along with its 95% CI values in all cases.

2.12.5.1 SINGLE-LIKELIHOOD ANCESTOR COUNTING (SLAC)

SLAC analysis was completed by selecting the “Single Ancestor Counting” option in the QuickSelectionDetection.bf batch file submenu. The following SLAC-specific options were also

selected: ancestral reconstruction was applied to the whole tree, sequence ambiguities in ancestral reconstructions were averaged over all possible codon states (with particular states being weighted relative to their frequency in the alignment), and p-values were determined using an extended binomial distribution, as described below.

With this procedure, ancestral sequences are reconstructed using nucleotide and codon substitution model parameter estimates (including branch lengths, nucleotide substitution rates, and global dN/dS) following a maximum likelihood-based approach (as opposed to a parsimony-based or Bayesian approaches, for example) (Pond and Frost 2005; Pond, Poon, and Frost, 2009). These reconstructions represent the ancestral sequences for each site at each internal node of the tree. Reconstructed ancestral sequences are then fixed as known variables and are used to determine the observed proportions of non-synonymous and synonymous *changes* and expected proportions of synonymous and nonsynonymous *sites* for each codon position (which should be equivalent to the expected proportion of changes if a sequence is evolving neutrally). This process excludes the introduction of stop codons from the calculation. Observed numbers are computed by counting substitutions, directly, at a given site and averaging over all possible shortest paths when multiple substitutions are required. Expected numbers are obtained by computing the mean proportions of synonymous and nonsynonymous sites at a given codon across all branches. A test of whether the observed proportion of synonymous substitutions is significantly different from the expected proportion of synonymous sites is then performed. p-values are determined using the two-tailed extended binomial distribution (following Durrett 2005). If the test is significant and the expected proportion of changes is greater than the observed proportion of sites, then the site is said to be positively selected. If, instead, the opposite scenario is observed, then the site is said to be negatively selected (Pond and Frost 2005; Pond, Poon, and Frost, 2009).

2.12.5.2 FIXED EFFECTS LIKELIHOOD (FEL)

FEL analysis was completed in HYPHY using the “Two rate FEL” option. As above, the entire tree was submitted for analysis, as opposed to a custom subset of branches. This procedure estimates dN and dS independently at each codon site, directly, in order to accommodate site-to-site variation in rates (Pond and Frost 2005). This method contrasts with Random Effects

Likelihood (REL)-type methods, such as those implemented in PAML, which, instead, determine synonymous and nonsynonymous substitution rates from a distribution of rates that is inferred from the data (Pond and Frost 2005; Pond, Poon, and Frost, 2009).

In HYPHY, FEL processes data in a number of stages. First, global alignment parameters are fitted to the data. This includes such things as the selected nucleotide and codon substitution models, global dN/dS, and branch lengths. Once fitted, these parameters are treated as known and are shared across branches. As a result, each branch can then be considered as an independent representation of the substitution process at each site, as described by the global parameters (Pond and Frost 2005; Pond, Poon, and Frost, 2009). LRTs of dN=dS are then performed at every site with 1 degree of freedom. This is completed by comparing the fit of a constrained model (dN=dS) to another where dN and dS are estimated independently from the data. Here, the asymptotic χ^2 distribution is employed in order to determine significance. If the test is significant ($p < 0.05$) and dN>dS at a particular site, then it is assumed to be under positive or diversifying selection. If, instead, dN<dS, then the site is determined to be under negative or purifying selection (Pond and Frost 2005; Pond, Poon, and Frost, 2009).

2.12.5.3 MIXED EFFECTS MODEL OF EVOLUTION (MEME)

Unlike the previously discussed methods, MEME can not only detect pervasive site-specific positive selection, but is also able to determine if a particular codon has been affected by transient episodes of diversification across a proportion of branches in the phylogenetic tree (Murrell et al. 2012). In fact, MEME represents an extension of earlier discussed branch-site REL methods (Pond et al. 2011; Smith et al. 2014).

In HYPHY, MEME analysis is completed by selecting the “Meme” option in the QuickSelectionDetection.bf batch file. Initial data analysis proceeds similarly to FEL. Global parameters are first fitted to the data, including the selected best-fitting codon and nucleotide models, global dN/dS, and branch lengths. However, unlike FEL, dN/dS at each site is not fixed across branches. Rather, MEME is referred to as a “Mixed Effects” model because it allows the distribution of rates to vary from site to site as a fixed effect (i.e.: rates are estimated independently and directly from the data), and from branch to branch at each site as a random effect (i.e.: each branch is assigned to one of a discrete number of rate-categories). Specifically,

at each site, dS is estimated independently, while each branch is assigned to one of two categories of dN : $dN \leq dS$ or $dN > dS$ (or $dN2$). At each site, a null model of $dN2 \leq dS$ is then fitted and compared to the unconstrained model using an LRT. Here, significance values are drawn from a mixture asymptotic χ^2 distribution (χ^2_0, χ^2_1 , and χ^2_2) (after Self and Liang 1987). If $dN2$ is greater than dS and the test is significant, positive selection is inferred. As earlier indicated, unlike FEL, MEME allows rate variation across branches, and, by binning dN values across branches at each site, it also allows for the pooling of information, thus improving ones ability to detect relatively weak signatures of selection at a particular site in an alignment (Murrell et al. 2012; Pond et al. 2011).

2.12.5.4 FAST UNCONSTRAINED BAYESIAN APPROXIMATION (FUBAR)

In HYPHY, the FUBAR.bf batch file is available in the “Selection/Recombination” submenu of the “Standard Analyses” file. The following FUBAR-specific options were also applied: 400 grid points (20x20) were used (as described below), 5 Markov Chain Monte Carlo (MCMC) chains were run, each chain was 2000000 steps in length, the first 1000000 steps were discarded as burn in, 100 samples were drawn from each chain (i.e.: each chain was sampled every 10000 steps), and the concentration of the Dirichlet prior concentration parameter was set at 0.5.

FUBAR is similar to other REL-type methods, in that it bins rate-values into discrete classes that are defined from a distribution of possible dN and dS values, which is, itself, determined from the data (Murrell et al. 2013). In this way, FUBAR shares REL’s ability to pool information in order aid in the detection of transient positive selection. However, as opposed to employing only relatively very few rate categories, sites are instead assigned to one of as many as 400 discrete classes, depending on grid-size (which is indicated *a priori*). Because the distribution of rates is so finely discretized, it is considered to be essentially unconstrained, meaning that it also shares some of the advantages of FEL-type methods, which are entirely unconstrained (Murrell et al. 2013; Pond et al. 2005).

The grid, itself, represents the distribution of dN and dS values for an alignment (Murrell et al 2013). Within this grid, 70% of points along a given axis represent $dN/dS < 1$, one point

represents $dN/dS=1$, and the remaining points represent $dN/dS>1$ (to a maximum of $dN/dS=50$). Weights (i.e.: the number of sites belonging to each rate-class) are determined for each point in the grid using a hierarchical Bayes approach. A symmetric Dirichlet prior probability distribution is assumed for rate class weights. A concentration parameter is defined in order to determine how dispersed or clumped the prior distribution of weights will be (i.e.: how evenly distributed the weights are across all dN and dS value combinations). As a result, the probability of each dN - dS value combination is represented by a general discrete bivariate distribution, with each combination having a particular weight.

MCMC sampling is used to sample weights for each point on the grid in light of a particular alignment in order to approximate the posterior probability distribution (Murrell et al. 2013). In HYPHY, this is achieved using the Metropolis algorithm (Pond et al. 2005). For each sample, a posterior distribution of dN and dS is calculated for each site using Bayes' Theorem. Posterior probabilities of positive selection for each site are determined by averaging across samples. Multiple MCMC chains are run in parallel in order to ensure convergence on similar site-specific posterior probabilities of positive selection (Murrell et al. 2013). For the purposes of this study, posterior probabilities > 0.9 are considered to be significant.

2.12.6 MACDONALD-KREITMAN TESTS (MKTs) AND TAJIMA'S D TESTS OF NEUTRALITY

Population-level selection analyses were performed using genetic data for a subset of individuals. To do this, data for all diploid plants of the *Turnera* subseries were considered and, as a result, individuals of the species *T. subulata*, *T. scabra*, *T. krapovickasii*, and *T. concinna* were represented. Tetraploid data were excluded because, where sequence ambiguities existed, haplotypes could not be reliably reconstructed in PHASE v. 2.1 (Stephens et al. 2001; Stephens and Donnelly 2003). Also, as their tetraploid status already more-or-less prevents the exchange of genetic information with other non-tetraploid members of the *Turnera* subseries, together, they could not be properly considered as part of a "population" sample.

Alignments containing sequence ambiguities (heterozygous sites) were submitted to PHASE v. 2.1 for the purposes of haplotype reconstruction, as described above (Stephens et al. 2001; Stephens and Donnelly 2003; Librado and Rozas 2009). McDonald-Kreitman tests (MKTs)

were then performed on all 12 phased *S*-linked gene data sets, using sequences obtained for *T. panamensis* as the inter-specific sample. The *Tsst1* alignment was not found to contain any sequence ambiguities, and, as a result, an MKT was performed on this alignment, directly, as haplotype reconstruction was not required. An additional MKT test was performed on an alignment of *Tsst1* sequences obtained from 17 individuals, representing a very localized population of tetraploid *T. scabra* from the Dominican Republic (DROT). These analyses were also performed in DNAsp v. 5.

As proposed by McDonald and Kreitman (1991), MKTs evaluate the neutral expectation that the ratio of nonsynonymous to synonymous fixations and polymorphisms should be equal within a given alignment, with significant deviations from this expectation indicating the action of selection. Here, significance is determined using a 2-tailed Fisher's Exact Test. In addition to p-values, Neutrality Index (Rand and Kann 1996) and α values (Smith and Eyre-Walker 2002) are also outputted using the MKT protocol available in DNAsp v. 5. However, it should be noted that the numbers of synonymous and nonsynonymous fixations and polymorphisms were not calculated at codon sites containing alignment gaps in sequences for any individual(s), as per the particular constraints of DNAsp v. 5.

Tajima's D tests of neutrality were also performed in DNAsp v. 5 using this same data (excepting the inclusion of *T. panamensis*). This test compares the average number of pair-wise sequence differences (Tajima's π) to the number of segregating sites (S or Watterson's θ) in an alignment (Tajima 1989). If a particular set of sequences are evolving neutrally, then these population and pair-wise measures of diversity should be approximately equal, as they are generated by the same process (i.e.: drift). In this case, $D \approx 0$. A value of $D < 0$, which arises when the number of segregating sites is larger than the average number of pair-wise differences, can indicate a recent selective sweep, while $D > 0$ may indicate the action of balancing selection (i.e.: There are many alleles of intermediate frequency in the population, thus making $\pi > \theta$). Statistical significance is then determined from the beta distribution (Tajima 1989).

3.0 RESULTS

3.1 SEQUENCING, ALIGNMENTS, AND DESCRIPTIVE STATISTICS

In total, nearly 9.5 kb of sequence information was obtained and analysed, 6.1 kb of which originated from *S*-linked genes (Table 6). Alignments for individual genes ranged from 156 bp (*LEJ2*) to 1188 bp (*LRRK*) in length. Nucleotide alignments for all *S*-linked genes are shown in Appendix B (Figures B1-B13), while alignments for all control genes are provided in Appendix C (Figures C1-C8).

For most genes, sequences were successfully obtained from all 24 individuals of interest (or 34, in the case of the *Tsstal* total data alignment) (Table 1, Table 6). However, PCR products could not be amplified or sequenced for the following genes (individuals): *LEJ2* (DIF), *RNABP* (KRAP 5S), *FMO1* (DIF and WED 2S), and *MBD8* (DIF, WED 2S, and CHAM 4L).

Accordingly, the alignment for control gene *MBD8* included the fewest species, as representative sequences from *T. diffusa*, *T. weddelliana*, and *T. chamaedrifolia* were not obtained. Similarly, gaps in alignments for *LRRK* and *RNABP* could not be resolved due to persistent “shifted peaks” in the sequence chromatograms for ES and SL8 201S, and DROT 41S and MIDC 710S, respectively. Further, an early stop codon was identified in the *AP2D* sequence that was obtained for CHAM 4L (*T. chamaedrifolia*), resulting in a truncated sequence of only 414 bp (138 codons) for this individual.

For alignments of individual genes, sequence conservation across taxa ranged from 69% (for *Tsstal* (total data)) to 93% (for *GAUT1*). The *SCEI* alignment contained the greatest number of identical sequences (13), however, followed by the *GAUT1* and *Tsstal* (total data) alignments (8 each). In total, 84.7% sequence conservation was observed across all *S*-linked genes, 90.3% across all control genes, and 86.6% across all genes, combined (Table 6).

The overall percentage of variable sites in each gene alignment ranged from 7% (in the *GAUT3* alignment) to 30% (in the *Tsstal* (total data) alignment) (Table 6). In general, most variation was unique to a single sequence. That is, singleton sites often outnumbered parsimony informative sites. When all of the genetic data was considered (9489 bp in total), of the 1137 sites that were found to vary, 544 represented shared polymorphisms (or parsimony informative sites), while 589 were found in only one individual.

When the number of representative taxa in each alignment was reduced so as to be more or less equal across genes, estimates of nucleotide diversity (π) and sequence polymorphism (θ) were also obtained (Figure 6). Of all of the genes considered here, the highest nucleotide diversity and sequence polymorphism values were observed in the *RNABP* alignment ($\pi= 0.037$ (SD=0.004) and $\theta= 0.049$ (SD=0.014)), followed by *AP2D* ($\pi= 0.030$ (SD=0.003) and $\theta= 0.035$ (SD=0.010)), which is located adjacent to *RNABP* at the *S*-locus (Figure 6). *LRRK* also obtained similar values, though the standard deviations for these values were generally higher ($\pi= 0.029$ (SD=0.011) and $\theta= 0.039$ (SD=0.011)). *SCEI* showed the lowest nucleotide diversity and Watterson's θ values ($\pi= 0.006$ (SD=0.001) and $\theta= 0.006$ (SD=0.005)). Though the largest and smallest values were obtained by *S*-locus genes, estimates of π and θ were found to be very similar across control and *S*-linked alignments, on average. Indeed, no particular pattern in terms of the distribution of average pair-wise divergence (π) and alignment-wide sequence polymorphism (θ) across the *S*-locus was immediately detectable, though a general increasing trend in these values can be observed from *Apetala2* to *RNABP* before dropping off precipitously at *SCEI* (Figure 6). When compared, *S*-linked and control genes exhibited no significant difference in average π or θ ($p>0.05$; mean $\pi_{S\text{-linked}} = 0.020$; mean $\pi_{\text{control}} = 0.018$; mean $\theta_{S\text{-linked}} = 0.029$; mean $\theta_{\text{control}} = 0.024$) (Figure 7).

In agreement with the above estimates, *RNABP* also showed the greatest proportion of polymorphic sites (22.2%), with the proportions of synonymous and non-synonymous mutations being roughly equal (48.2% and 51.8% , respectively) (Table 7). The *RNABP* alignment also contained the greatest number of indel events (8) and exhibited the largest average indel length (9.437 bp). However, indel diversity estimates (π_1) were highest for *AP2D* ($\pi_1=0.005$), followed by *APETALA2* ($\pi_1=0.004$), *RNABP* ($\pi_1=0.003$), and *LRRK* ($\pi_1=0.003$). Only one 3 bp long indel event was identified across all control gene alignments (*RNABP34*). When the *S*-linked and control gene datasets were compared, the average number of indel events and indel length per alignment were found to be significantly different ($p<0.05$) (Table 7). However, indel diversity measures (π_1) were not found to be significantly different ($p>0.05$). In general, control genes exhibited a somewhat smaller proportion of polymorphic sites (11.0% versus 13.4% in *S*-linked genes, on average), with most mutations resulting in synonymous amino acid changes (62.4% versus 57.8% in *S*-linked genes, on average) (Table 7). However, these differences were not found to be statistically significant by Mann-Whitney-U tests ($p>0.05$). The average number of

nonsynonymous mutations per gene was also not found to differ significantly between data sets. *SCE1* exhibited the lowest proportion of polymorphic sites and the highest proportion of synonymous mutations (5.1% and 88.9%, respectively), while the control gene, *UNKN* (unknown function), exhibited the greatest proportion of non-synonymous mutations, at 61.8% (Table 7).

3.2 PHYLOGENETIC ANALYSES

Several phylogenies were computed using the data. Six trees were constructed for the purposes of analyses to be performed in HYPHY (Pond, Frost, and Muse 2005). These phylogenies were constructed using the total data for all genes and all individuals of interest, the same data with a reduced number of representative taxa, and *Tsstal* data that had been collected from an expanded number of species. Trees were constructed using both the DNA and amino acid alignments. Those that were obtained using DNA data are shown in Figures 8-10. Those computed using translated DNA data sets are supplied in Appendix D (Figures D1-D3). Phylogenies produced using nucleotide and amino acid data for each individual gene, and concatenated alignments of all *S*-linked genes and all control genes, separately, are provided in Appendix E (Figures E1-E44).

3.2.1 THE TOTAL DATA SET: DNA- AND AMINO ACID-BASED PHYLOGENIES

The total DNA data obtained for all genes and all individuals (excepting those that were specific to the *Tsstal* total data set) were used to produce the phylogeny shown in Figure 8 by maximum likelihood methods. After eliminating sites that were represented by >5% gaps and ambiguous bases, a total of 6526 informative base positions were retained and used for phylogenetic reconstruction.

In this analysis, the monophyly of the *Turnera* series – represented here by members of the *Turnera* and *Umbilicatae* subseries – was supported by strong bootstrap values (100%) (Figure 8). Though little species-level resolution was obtained within the *Turnera* subseries, this group also obtained considerable support (98%). Interestingly, the individual, DEN 20S, a short-styled

plant of the species *T. grandidentata*, appears as sister taxon to all other members of the Turnera subseries (including a long-styled plant of the same species, DEN 54L), and with high bootstrap support (98%). The Umbilicatae subseries, represented here by *T. joelii* (TJ 29S and 30L), obtained high branch support (100%) and is depicted as the sister group to the Turnera subseries (100%).

T. chamaedrifolia and *T. diffusa*, members of series Papilliferae and Microphylae, respectively, are strongly supported as a single clade, with 100% bootstrap support (Figure 8). Within the tree, they are presented as the most closely related group to the Turnera series. Representative members of series Salicifoliae (*T. panamensis* and *T. weddelliana*) are also a strongly supported monophyletic group (99%) and are sisters to all other taxa.

The total DNA data were translated and the resultant amino acid alignment was then used to generate another tree by maximum likelihood methods (Figure D1). A total of 1976 amino acid sites were retained and used to inform this phylogeny.

Analysis of the total amino acid data resulted in the production of a phylogenetic tree with very similar topology to that which was computed with the total DNA data (Figure D1). As was the case in Figure 8, the monophyly of the Turnera series was strongly supported by bootstrap values (100%), as was the position of *T. chamaedrifolia* and *T. diffusa* as the group in the tree to which it is most closely related (98%). Within the Turnera series, the Turnera and Umbilicatae subseries were also strongly supported groups, showing 90% and 100% bootstrap support, respectively. As with the above, members of the Salicifoliae series were depicted as sisters to all other groups.

As was shown in Figure 8, when the amino acid data were submitted to phylogenetic analysis, the position of the individual, DEN 20S (*T. grandidentata*), also appeared to be somewhat segregated from the remaining members of the Turnera subseries with strong bootstrap support (90%) (Figure D1). Again, no species-level resolution was obtained within the Turnera subseries when the translated data set was employed.

3.2.2 THE TOTAL DATA SET WITH A REDUCED NUMBER OF TAXA: DNA- AND AMINO ACID-BASED PHYLOGENIES

In order to perform tests of selection in HYPHY (Pond, Frost, and Muse 2005), it is necessary that the taxa included in the alignment submitted for analysis match those that are represented in the accompanying phylogenetic tree. Because the *Tsstal* gene appears only to be represented in the genomes of short-styled and short-homostyled individuals of the species of interest, it was necessary to compute an additional set of phylogenetic trees using the total genetic data that had been obtained for individuals of these types (16 individuals, total). Because plant material was only available for long-styled individuals of *T. diffusa* and *T. chamaedrifolia*, no representatives of these species could be included in phylogenetic reconstruction. In HYPHY, these trees were used for tests of selection on the *Tsstal* alignment only.

A phylogeny constructed using the total DNA data from all genes for only short-styled individuals by maximum likelihood methods is shown in Figure 9. A total of 6467 nucleotide sites were retained for this analysis. As would be expected, even after a reduction in representative taxa, the resultant phylogeny retained a similar topology to those described earlier. The *Turnera* series was again strongly supported as a monophyletic group (100%), with *T. joellii*, the only species from the Umbilicatae subseries represented here, appearing as its sister. Similarly, *T. weddelliana* and *T. panamensis*, both members of the Salicifoliae series, were again positioned as sister taxa to all other individuals included here.

In this tree, the *Turnera* subseries was also strongly supported as a clade (99%) (Figure 9). However, in most cases, individuals did not appear to cluster by species within this group. For instance, all included individuals of the *Turnera* subseries, with the exception of TSH (tetraploid *T. scabra*), E 207S (tetraploid *T. subulata*), and DEN 20S (*T. grandidentata*), formed a reasonably well supported clade (85%). Within this clade, MAN 601S and ES, both of which are *T. scabra*, also formed a moderately well supported sub-group with F60SS, a member of the species, *T. subulata* (80%). All other such relationships within the *Turnera* subseries received bootstrap supports of $\leq 75\%$.

When the DNA data set was translated and submitted to phylogenetic analysis, many of the same relationships were identified (Figure D2). A total of 1922 amino acid sites were considered in this analysis. Again, DEN 20S and E 207S were shown as sister taxa to the *Turnera* subseries clade (96%), followed by TSH (47%). The remaining members of the *Turnera* subseries,

including representatives of *T. krapovickasii*, *T. concinna*, *T. subulata*, and *T. scabra* also formed a moderately well supported group within this clade (73%).

3.2.3 THE *Tsstal* DATA SET WITH AN EXPANDED NUMBER OF TAXA: DNA- AND AMINO ACID-BASED PHYLOGENIES

Because *Tsstal* is only represented in short-styled and short-homostyled individuals of distylous *Turnera*, and also has significant sequence homology to genes known to be involved in SI response in *Papaver*, it is of particular interest in this study. As a result, *Tsstal* was sequenced in an increased number of species, including members of related genera, *Piriqueta* and *Erblichia*, as well as additional species from the *Turnera* series. This larger data set included sequence information for 34 individuals. Phylogenies constructed using these data were subsequently employed in various selection analyses of the *Tsstal* gene (Figures 10 and D3).

When the DNA data for *Tsstal* across an increased number of taxa were considered, a total of 348 nucleotide base positions were retained for phylogenetic analysis. In the resultant phylogeny, the *Turnera* series is again strongly supported as a monophyletic clade (100%) (Figure 10). However, within this group, *T. joellii* (subseries Umbilicatae) is situated within the *Turnera* subseries, in an intermediate position between *T. grandiflora* (GRAN 9S) and the remaining members of this group (84%). Within the *Turnera* subseries clade, *T. subulata*, *T. scabra*, *T. krapovickasii*, *T. concinna*, and *T. grandidentata* all cluster together at the top of the tree. However, this grouping obtained only very modest bootstrap support (64%). Within this particular collection of individuals, species-level resolution was also not achieved. Rather, members of two species, in particular, *T. subulata* and *T. scabra*, often appeared as less closely related to members of their own species than they were to individuals of the opposing species, as was also characteristic of many of the earlier described phylogenies (Figures 8-9).

With attention, again, to the *Turnera* subseries, individuals of the species *T. occidentalis*, *T. Aurelii*, *T. orientalis*, and *T. cuneiformis* formed a reasonably well-supported clade within this tree (83%) (Figure 10). Indeed, it should be noted that sequences obtained for *T. occidentalis*, *T. orientalis*, and *T. cuneiformis* were found to be identical at this locus (Appendix B, Figure B2). In any case, this collection of individuals is positioned as a close sister group to *T. oculata*, *T. ulmifolia*, and *T. velutina* (68%), with 77% bootstrap support.

As was also the case in the phylogenies described in Figures 8 and 9, in Figure 10, *T. weddelliana* and *T. panamensis* (series *Salicifoliae*) are depicted as most closely related *Turnera* series clade, with 77% and 71% bootstrap support, respectively. Members of the genus, *Piriqueta*, then form a sister group to *Turnera* in the tree. The monophyly of *Piriqueta* is supported by a 91% bootstrap value. The *Piriqueta* group is then separated into three distinct clades, containing the following species: 1) *P. cistoides ssp. caroliniana* (91%), 2). *P. sarae*, *P. revoluta*, and *P. nanuzae* (99%), and 3). *P. viscosa*, *P. plicata*, *P. morongii*, and *P. duarteana* (100%). Within the third clade, *P. plicata*, *P. morongii*, and *P. duarteana* form a subgroup with 79% bootstrap support. That being said, the branch supporting the subdivision between *Piriqueta* species groups 2 and 3 only obtained 50% bootstrap support. *Erblichia odorata* (EOD) is then positioned as the sister taxon to all other groups in the phylogeny.

When the total *Tssta1* data set was translated, and the amino acid alignment was subsequently submitted to phylogenetic analysis using maximum likelihood methods, 116 amino acid positions were included in the final data set (Figure D3). This phylogeny suggests a very similar topology to that which was observed when the DNA data was considered, though bootstrap supports for all clades were found to be uniformly reduced, as was the general resolution of the tree. However, some key differences in tree topology can be readily identified. Perhaps most interesting is the position of *T. weddelliana* within the phylogeny. Here, it appears in an intermediate position between *E. odorata* and the *Piriqueta* group, and completely separated from other species of the *Turnera* genus.

3.2.4 GENE GENEALOGIES

Individual nucleotide and amino acid alignments for each sequenced gene/exon were submitted to phylogenetic analysis, in turn, using maximum likelihood methods. The total data for all *S*-linked and control genes were also analyzed separately in this way. These analyses were completed, in part, to determine if any further species-level resolution, particularly within the *Turnera* subseries, could be achieved. Additionally, these phylogenies were computed in order to identify any possible pattern(s) in terms of the relationships between the alleles for each gene. The results of these analyses are displayed in Appendix E (Figures E1-E44).

These results were found to be largely consistent with those that were obtained using the total nucleotide and amino acid data sets, though often with lower bootstrap support, as might be expected due to insufficient data. However, there were a few notable exceptions. For instance, taxa within the nucleotide and amino acid-based trees for *AP2D* were found to be somewhat unusually arranged (Figures E7 and E8). In the nucleotide tree, *T. chamaedrifolia* and *T. diffusa* are positioned within the *Turnera* subseries (Figure E7). And, further, when the amino acid data for this gene were considered, *T. grandidentata* is repositioned as the sister group to all other taxa, with *T. chamaedrifolia*, *T. diffusa*, *T. weddelliana*, and *T. panamensis* becoming fully subsumed within the *Turnera* series (Figure E8). However, bootstrap supports for these particular species arrangements were quite meagre in most cases (<50%).

The trees obtained using the data for *RNABP* were similarly remarkable, with branches often obtaining much more significant bootstrap supports than were observed for *AP2D* (Figures E9 and E10). For instance, *T. chamaedrifolia* was again integrated within the *Turnera* series, with 98% and 88% support in the DNA and amino acid trees, respectively. The individual, DEN 54L, a long-styled member of the species, *T. grandidentata*, was also shown as being closely related to *T. diffusa*, with considerable bootstrap support, especially in the amino acid-based tree (92%). However, DEN 20S, the short-styled representative of *T. grandidentata*, is positioned as expected, within the *Turnera* subseries.

Of the phylogenetic results obtained using alignments of control genes, the phylogenies produced for *FMO1* deviated most from what had been previously observed (Figures E35 and E36). In these trees, *T. diffusa* and *T. weddelliana* are situated within the *Turnera* subseries with considerable bootstrap support (91% and 95%, in the nucleotide and amino acid trees, respectively). However, it should be noted that most aberrations from expectations are likely explained by insufficient data (i.e.: short alignments, with few differences between taxa). Indeed, when total *S*-linked and total control data were considered separately, the phylogenies converged on a similar topology to those that were observed in Figures 8-10 (Figures E25-26 and E43-44).

Individual gene alignments were also investigated for the presence of trans-species polymorphisms (TSPs). However, evidence of trans-specific evolution was only obtained for one gene: *APETALA2* (Figure 11). For this gene, short-styled individuals of the *Turnera* subseries appeared to form a single clade, while long-styled individuals fell into a polytomy from which the short-styled clade emerges. However, the TSP signal appeared to break down outside of this

subseries. *Turnera joelii*, *T. chamaedrifolia*, *T. diffusa*, *T. weddelliana*, and *T. panamensis* fell into their series/subseries-specific clades. Importantly, the TSP pattern that is evident in the gene genealogy for *APETALA2* is driven by only 3 sites (base positions 75, 376, and 406, as indicated in Appendix B, Figure B1). At site 75, both TSH (short homostyle, *T. subulata*) and DROT 41S (*T. scabra*) share the *s*-specific base. TSH also shares the *s*-specific base at position 406. At nucleotide sites 376 and 406, morph-specific base differences within the *Turnera* subseries result in amino acid substitutions. Due to base differences at site 376, long-styled individuals possess an isoleucine, while shorts have a leucine at the corresponding codon position. Due to base differences at site 406, longs possess a valine, while shorts have an isoleucine at that codon site. At site 75, both morphs possess a glutamate residue.

3.3 SELECTION ANALYSES

All selection analyses were performed in HYPHY v. 2.2.3 (Pond, Frost, and Muse 2005) and DNAsp v. 5 (Librado and Rozas 2009). For these analyses, in most cases, the phylogeny based on the total nucleotide data for all genes/exons sequenced here was used in conjunction with the relevant nucleotide alignment (Figure 8). For analyses of *Tssta1*, in particular, either the nucleotide-based phylogeny depicted in Figure 9 or Figure 10 was employed, depending on the number of taxa that were included in the relevant alignment.

3.3.1 GLOBAL dN/dS ESTIMATES

Global, or alignment-wide, estimates of dN/dS were calculated for each gene/exon sequenced, as well as their asymptotic 95% confidence intervals (CI) (Table 8). Estimates were also obtained for concatenated alignments of all *S*-linked genes and all control genes, separately. For most alignments, global dN/dS estimates were found to be less than 1, suggesting an alignment-wide deviation from neutral evolution in the direction of negative or purifying selection. However, the global dN/dS estimate for *LEJ2* (exons 1-2), was calculated to be 1.59 (\pm 1.07), suggesting that selection across this alignment is not significantly different from neutral expectations, or dN/dS=1, on average.

In order to provide further support to these results, additional likelihood ratio tests (LRTs) were performed in order to evaluate the null hypothesis of $dN=dS$ (or $dN/dS=1$) across each alignment (Table 8). Significant p-values ($p<0.05$) were obtained for all but two alignments: *LEJ2* and *UNKN* ($p = 0.15$ and 0.08 , respectively), suggesting that the model where dN/dS was estimated from the data did not fit the data significantly better than the model where dN and dS values were constrained to be equal. This result supports the conclusion that the particular exons represented in these alignments are evolving more or less neutrally, in general.

For all other alignments, LRTs of $dN/dS=1$ resulted in p-values that were $\ll 0.05$, suggesting that dN/dS is significantly different from 1 in these cases (Table 8). These results, combined with the earlier estimations of global dN/dS , indicate that, not only is the evolutionary rate significantly non-neutral for these alignments, but also that they show statistically significant signatures of purifying selection in general, on an alignment-wide scale.

3.3.2 COMPARING dN/dS IN *S*-LINKED AND CONTROL GENES

In order to determine if *S*-linked genes are uniquely influenced by selection, their evolutionary rates were compared to those of a semi-random selection of genes located elsewhere in the *Turnera* genome. When the total genetic data that had been obtained for all *S*-linked genes were considered, global dN/dS was found to be $0.25 (\pm 0.03)$, indicative of purifying selection across the alignment (Table 8). Importantly, when sequence data for *LEJ2* (which, in isolation, obtained high global dN/dS value relative to other *S*-linked genes) was removed from the total *S*-linked gene alignment, global rate estimates remained very similar, at $0.23 (\pm 0.02)$. Global dN/dS values that were obtained for the total control data were similar but somewhat lower, at $0.15 (\pm 0.03)$. In addition, when these alignments were submitted to LRTs of $dN/dS=1$ significant p-values were obtained ($p\ll 0.05$), suggesting that evolutionary rates deviate from the neutral expectation in both cases. Importantly, the global dN/dS value ranges for *S*-linked and control genes do not overlap, as is indicated by their 95% CIs. This suggests that the strength of selection, while negative in both cases, is significantly different between alignments.

In order to determine if the control and *S*-linked genetic data were significantly different from each other in terms of positive selection, in particular, four LRTs were performed. Specifically, these tests were employed to determine if any of the following rate parameters were

significantly different between alignments (Table 9): 1) All rate parameters, including the values of dN, dS, and dN/dS for all four rate classes (1 neutrally, 2 negatively, and 1 positively selected), as well as the proportion of codon sites belonging to each rate class; 2) the strength of positive selection (i.e.: the values of dN, dS, and dN/dS for the positively selected rate class); 3) the proportion of positively selected sites; and 4) the positive selective regime (the strength of positive selection and the proportion of positively selected sites). Rate parameter estimates for these tests can be found in Appendix F, Table F1.

When a model where the values of all rate parameters were constrained so as to be shared by both alignments was compared to another where all parameters were estimated independently, the *S*-linked and control data sets were found to be significantly different from each other ($p=0.002$; LRT Statistic = 27.83, DF = 10; Table 9). This suggests that the independent model fit was superior to that of the constrained model in this case, and further indicates that the data sets differ with regards to the distribution of rate parameters in some way. However, when rate parameters related to positive selection, in particular, were investigated, no significant differences were detected by any test, indicating that strength of positive selection, the proportion of positively selected sites, and the positive selective regime were largely the same across data sets ($p \gg 0.05$) (Table 9).

As indicated above, *LEJ2* obtained a global dN/dS value that was considerably higher than other *S*-linked genes (1.59 ± 1.074 ; Table 8). Interestingly, when *LEJ2* was removed from the total *S*-locus alignment and then compared to the total control alignment using the LRTs described above, the results were found to be more extreme. While the distributions of the various rate parameters were still found to be significantly different ($p \ll 0.05$), the positive selective regime and proportion of positively selected sites were also found to differ significantly in the two alignments ($p=0.04$ and 0.02 , respectively). The strength of positive selection, however, was found to be shared by the two alignments ($p > 0.05$), as was the case when *LEJ2* sequence data were included.

Due to the complexity of the models being fitted to the data, the LRTs described above were repeated using random starting values for all parameters (*LEJ2* included). Importantly, these tests converged on nearly identical results (Appendix F, Table F2).

3.3.3 LOCAL OR LINEAGE-SPECIFIC SELECTION ON S-LINKED GENES

Because strong alignment-wide signatures of purifying selection, like those that have been detected here, can overpower weaker signals of positive selection, especially when it affects only particular lineages, and because evolutionary rates may be expected to vary over time (Pond, Poon, and Frost 2009; Murrell et al. 2015), several tests of lineage-specific selection were also completed using sequence information obtained from *S*-linked genes. Using LRTs, alignment-wide or global dN/dS ratio models were first compared to local or lineage-specific dN/dS models for each alignment, independently (Table 12). In the lineage-specific model, dN/dS was estimated separately for each branch in the phylogenetic tree, while for the global model all branches were made to share the same evolutionary rate value. When these two models were compared for each alignment, significant results were not obtained in most cases ($p \gg 0.05$). However, global and local models were found to be significantly different for the *AP2D* alignment, indicating that the fit of the local model was superior to that of the global model for this alignment ($p = 0.01$; LRT statistic = 69.17, DF = 44).

Individual *S*-linked gene alignments were also investigated for the presence of more subtle signals of episodic diversifying selection across individual lineages and codon sites using Branch-site Unrestricted Statistical Tests of Episodic Selection (BUSTED) (Table 13). However, significant results were not obtained for any alignment ($p \gg 0.05$). Further, many alignments showed no evidence of positive selection (i.e.: proportion of sites with $dN/dS > 1$ was essentially 0). As a result, LRTs of $dN/dS < 1$ for all sites, across all lineages were not completed for some alignments (*SCEI*, *FRA1*, *FSP*, and *NRFP*), as it was already clear that this hypothesis could not be rejected in those cases.

Like most lineage-specific tests of selection, BUSTED is not recommended as a method of identifying particular lineages that have experienced positive selection when there is no *a priori* reason to believe that any particular lineages have been influenced by it (Pond, Poon, and Frost, 2009; Murrell et al. 2015). As a result, additional branch-site tests, which correct for the effect of multiple comparisons (aBS-REL), were performed in order to identify particular lineages that may have been influenced by episodic positive or diversifying selection in each alignment (Pond et al. 2011) (Tables G1 and G2, Appendix G). However, these tests did not identify any lineages that may have been inordinately influenced by positive selection in any alignment (corrected

$p > 0.05$ for all branches tested in all alignments). For *AP2D*, in particular, aBS-REL results obtained for node 5 (the node that separates D16L, DROT 41S, MAN 601S, ES, SL8, F60SS, and MAN 713L from MIDC 710S, COLO, PA 4S, TSH, E207S, and E 2L in the Turnera subseries clade in Figure 8) were approaching significance ($p = 0.08$). That being said, this node obtained 0% bootstrap support in the corresponding phylogeny (Figure 8).

3.3.4 dN/dS IN *Tssta1* AND *AP2D*

As was earlier discussed, *Tssta1* is naturally a very interesting gene in the context of this study because of its sequence similarity to genes known to be involved in SI in other genera, as well as its presence in the genomes of only short-styled individuals. However, in light of the above results, *AP2D* also emerges as an interesting candidate. Consequently, alignments of these two genes were investigated further. Specifically, the same series of LRTs that were performed in order to compare the strength and quality of positive selection on *S*-linked and control gene alignments were employed in order to compare the selection on alignments of *Tssta1* and *AP2D* with all other *S*-linked genes (Tables 10 and 11).

When *Tssta1* was compared to all remaining *S*-linked genes, no significant results were obtained for any test, indicating that the two alignments were not significantly different from each other with regards to the distribution of the rate parameters, dN, dS, dN/dS, and the proportion of sites belonging to any one of the four rate classes ($p \gg 0.05$) (Table 10). Parameter estimates for these tests can be found in Appendix F, Table F3. When these tests were repeated using random starting values for parameters, similar results were also obtained (Appendix F, Table F4).

When the *AP2D* alignment was then compared to an alignment of sequence data from the remaining *S*-linked genes, significant results were obtained for one of the four LRTs (Table 11). Specifically, when the fit of the fully constrained model was compared to the independent model, they were found to be significantly different ($p < 0.05$; LRT statistic = 37.33, DF = 10). However, tests of shared positive selection parameters between the alignments (i.e.: shared positive dN/dS and proportions of positively selected sites) yielded no significant results ($p > 0.05$). Parameter estimates that were obtained for these tests can be found in Appendix F,

Table F5. As with previous tests of this type, LRTs completed with random starting values for all parameters converged on very similar results (Appendix F, Table F6).

3.3.5 CODON-LEVEL SELECTION ON S-LINKED GENES

When positive selection affects only a small proportion of codons, it can be difficult to detect, especially when negative selection dominates the alignment (Pond, Poon, and Frost 2009; Murrell et al. 2012). As a result, several tests of site-by-site variation in evolutionary rates were performed on all *S*-linked sequences, using Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Mixed Effects Model of Evolution (MEME) methods. These results are summarized for each alignment in Tables 14-26. Sites that had been detected as positively or negatively selected by two or more methods are also indicated in the nucleotide alignments for each gene, as depicted in Appendix B, Figures B1-B13.

3.3.5.1 CODON-LEVEL SELECTION ON *APETALA2*

When site-by-site tests of selection were applied to the *APETALA2* exon 1 alignment, 18 codons were identified as negatively selected by at least one method, while 4 were found to be influenced by positive selection, in total (Table 14). However, only five codons were identified as being negatively selected (codons 15, 50, 79, 81, and 83), and one as being positively selected (codon 40), by more than two methods. These sites are highlighted in the alignment shown in Figure B1 (Appendix B).

Codon 40 was identified as positively selected by 3 of the 4 methods employed here. Indeed, only SLAC failed to detect positive selection at this site, though it is considered to be the most conservative test of codon-level diversifying selection (Table 14; Pond and Frost 2005). In addition, according to MEME results, 100% of the branches in the tree were shown to evolve with a dN value greater than dS at codon 40 (dN2=10.91; dS = 0.00), which further suggests pervasive positive selection at this site (Table 14). With reference to the alignment, individuals may possess one (or two, if they are heterozygous) of 3 amino acids at this codon position:

glycine, glutamic acid, or valine (Appendix B, Figure B1). *T. panamensis*, however, may be represented by an indel at this position (depending on how one resolves the gap in the sequence in this area). And, further, this particular site codes for a valine only in *T. weddelliana*.

It should be noted that codon 40 is located near the end of a particularly gap-rich portion of the *APETALA2* alignment, due to apparent insertion/deletion events (Appendix B, Figure B1; base positions 87-123 of the alignment). Indeed, the alignment could be alternatively arranged in this region, as a result. If the alignment were rearranged in this area, another site may have been identified instead, or alternatively, no positively selected sites may have been detected in the area at all. That being said, the presence of indels does suggest a certain lack of constraint in this region, in general. Interestingly, in this particular area of the alignment, it is not uncommon for individuals of the same species to be characterized by indels of differing sizes. For instance, in this particular indel, F60SS (*T. subulata*) is less 1 codon relative to D16L, which is its sibling. Similarly, MAN 713L (*T. scabra*) is represented by 3 extra codons relative to MAN 601S, which originates from the same population.

Based on BLAST searches, no conserved domains were identified in the *APETALA2* alignment, using the sequence information obtained here. Based on sequence information that was previously obtained for this gene in *T. subulata*, *APETALA2* is predicted to contain two AP2 DNA binding domains downstream of exon 1, the portion of the gene that was sequenced here (Accession: cd00018; E-values: 7.03e-03 and 8.34e-13, respectively, using the D16L sequence).

3.3.5.2 CODON-LEVEL SELECTION ON *Tssta1* (TOTAL DATA)

Site-by-site selection analyses performed using the *Tssta1* (total data) alignment resulted in the detection of 18 negatively and 1 positively selected site (codon 9) (Table 15). However, when only those sites that were detected by two or more methods were considered, these numbers were reduced to 14 negatively selected and no positively selected sites. These sites are highlighted in the alignment depicted in Figure B2 (Appendix B). Interestingly, all 14 of the negatively selected codons were found to be located within the region of the alignment identified as homologous to plant self-incompatibility domain family S1 (Accession: pfam05938; E-value: 2.88e-06, using the F60SS sequence). The sequence location of this domain is also highlighted in the alignment (base positions 133-447).

3.3.5.3 CODON-LEVEL SELECTION ON *Tssta1* (WITH A REDUCED NUMBER OF TAXA)

When tests of codon-level selection were reapplied to the *Tssta1* alignment with a reduced number of taxa, fewer potentially positively/negatively selected sites were consequently identified (Table 16). Indeed, almost no overlap in results was obtained between the two sets of results (Tables 15 and 16). Only codon sites 114 and 125 were identified as negatively selected in both cases. However, when the number of included taxa was reduced, these sites were only detected using one method (FEL).

No potentially positively selected sites were identified for *Tssta1* when fewer individuals were included in the alignment, and only 3 sites were found to be negatively selected by more than 1 method (codons 37, 51, and 99) (Table 16). These sites are highlighted in the alignment shown in Figure B3 (Appendix B). Unlike in the previous analysis of *Tssta1*, not all of the negatively selected sites detected here were located in the region of the conserved SI domain (Appendix B, Figure B3). However, one of the three was located just outside of this region.

3.3.5.4 CODON-LEVEL SELECTION ON *LEJ2*

In the *LEJ2* exon 1 and 2 alignment, 3 negatively and 5 positively selected codons were identified by at least one method (Table 17). However, only 1 site of each type was identified by two or more methods (Codon 33 and 12, respectively). These two sites are underlined in the *LEJ2* alignment depicted in Figure B4 (Appendix B).

Codon 12, identified as positively selected by all site-by-site selection detection methods employed with the exception of SLAC, is located at the beginning of the alignment, where sequence information was not obtained for 8 of the 23 individuals that were included in the analysis (also recall that no sequence was obtained for *T. diffusa* for this gene) (Table 17; Appendix B, Figure B4). However, those individuals that are represented at this particular site, are characterized by one of three different codons (CCC, TCC, or GCC), all of which code for different amino acids (proline, serine, and alanine, respectively). Indeed, in some cases, members of the same species are represented by a different codon at this position, as is the case for D16L and F60SS, which code for proline and alanine, respectively. Only representatives of *T. joellii* and *T. chamaedrifolia* code for serine at this codon. Further, MEME estimates suggest that 100%

of the branches in the tree evolve with a dN value greater than dS at this site (dN2=19.78; dS = 0.00) (Table 17).

Due to the small size of the alignment (156 bp, or 52 codons), no significant matches to any conserved domains in the region of *LEJ2* that was sequenced here were identified. According to previously obtained genetic data, in *T. subulata*, the *LEJ2* gene is predicted to contain a CBS domain, but the coding sequence for this domain is not located in exons 1 and 2, which were sequenced here (Accession: cd02205; E-value: 4.86e-23, using the D16L sequence).

3.3.5.5 CODON-LEVEL SELECTION ON *AP2D*

Site-by-site selection analysis revealed 21 negatively and 7 positively selected codons in the *AP2D* alignment (Table 18). However, only 10 negatively and 1 positively selected site (codon 20) were identified by more than one detection method. These sites are emphasized in the *AP2D* nucleotide alignment presented in Figure B5 (Appendix B).

Nearly the entire *AP2D* gene was sequenced here, for all of the individuals of interest. According to BLAST results, it contains a conserved AP2 DNA binding domain (Accession: cd00018; E-value = 4.82e-20, using the D16L sequence), located between base positions 91 and 264 in the alignment (Appendix B, Figure B5). Half of the negatively selected sites detected here are contained within this region. Specifically, this domain contains 11 conserved residues that, together, form a DNA binding site. These conserved residues have been mapped to the *Turnera AP2D* sequence alignment shown in Figure B5 (Appendix B). In this alignment, 10 of the 11 conserved sites are 100% identical across taxa. The remaining site (Codon 32) was detected as negatively selected using the methods FEL and FUBAR (Table 18).

Codon 20 was identified as potentially positively selected by all four codon-level selection tests employed here (Table 18). MEME results also estimate that 100% of the branches in the phylogeny evolve with dN>dS at this site (dN2=11.50; dS=0.00). At this codon position, individuals may possess an alanine, a threonine (or both alanine and threonine, if heterozygous), or a glycine. *T. weddelliana*, however, is represented by an indel at this position. Further, only representatives from the species, *T. joellii*, possess a glycine. Interestingly, codon 20 is located in a region of the alignment that is adjacent to and includes indels in the sequences obtained for some taxa (Appendix B, Figure B5). Sequences for some individuals contain no gaps in this area,

while others are represented by one or two fewer codons relative to those without gaps. Members of the same species are often characterized by different indels at this position, as is the case for *T. subulata* and *T. scabra*. Indeed, the *AP2D* alignment, in general, is characterized by several indels, particularly near the 3' end of the gene (between the 387 and 561bp positions in the alignment). Further, within this region, the sequence obtained for *T. chamaedrifolia* (CHAM 4L) was also found to contain an early stop codon, which truncated the coding sequence for this individual to just 414 bp. Interestingly, 3 of the 10 sites that were predicted to be negatively selected in this alignment are located just upstream of the position at which the stop codon was identified (codons 135, 136, and 138).

3.3.5.6 CODON-LEVEL SELECTION ON *RNABP*

In the *RNABP* exon 1 alignment, 22 negatively and 3 positively selected codons were found (Table 19). Of these, 17 negatively and 2 positively selected sites (codons 52 and 66) were detected by more than one method. These sites have been underlined in Figure B6 (Appendix B).

No sequence information for the individual, KRAP 5S (*T. krapovickasii*), was obtained for this gene. In addition, sequences for the individuals, DROT 41S and MIDC 710 S (both of which are tetraploid *T. scabra*) could not be resolved fully due to persistent shifted peaks in the alignment (from the 135 to 185 bp position in the alignment, approximately). It should be noted that the unresolved portions of these sequences are located within a region of the alignment that includes indels (Appendix B, Figure B6). This might indicate that DROT 41S and MIDC 710S are heterozygous for an indel somewhere within this region of the gene, which may have caused sequencing difficulties at this locus.

Importantly, both potentially positively selected sites are located in this same area of the alignment, in addition to 4 sites that are possibly negatively selected (codons 46, 60, 60, and 64) (Appendix B, Figure B6). At codon 52, which was determined to be possibly positively selected, individuals either code for an alanine or proline residue, with particular plants sometimes being heterozygous at this position. According to MEME results, 100% of the lineages represented are expected to evolve with $dN > dS$ at this codon ($dN=2.51$; $dS=0.00$). At codon 66, sequences for individuals either code for a glycine, a valine, or are instead represented by an indel. Again, the

particular codon assignment follows no apparent clade- or species-specific pattern. In addition, at this site only 58% of lineages are predicted to evolve with $dN > dS$ ($dN2=12.07$; $dS=0.00$), according to MEME results.

According to BLAST search results, exon 1 of *RNABP* contains sequence with homology to an RNA binding protein with multiple splicing (RBP-MS)-like RNA-recognition motif (Accession: cd12420; E-value $6.04e-10$, using the D16L sequence). The homologous region is indicated in Figure B6 (Appendix B), and is located between base positions 234 and 330 in the alignment. Five of the 17 potentially negatively selected sites identified here lie within this region (codons 84, 86, 90, 102, and 103). The 8 remaining negatively selected sites that were identified are located within the first 110 bases of the *RNABP* exon 1 alignment.

3.3.5.7 CODON-LEVEL SELECTION ON *SCEI*

Site-by-site selection analysis in exons 3 and 4 of *SCEI* revealed only 4 potentially negatively selected sites (Table 20). Of these, only one was identified as significant by more than one method (Codon 22). This site is highlighted in Figure B7 (Appendix B).

According to BLAST search results, exons 3 and 4 of *SCEI* have homology to a ubiquitin conjugating enzyme, E2, catalytic domain (Accession: cd00195; E-value: $1.28e-13$, using the D16L sequence), including its conserved cysteine active site as well as 19 of 21 of its thioester intermediate interaction residues. These sites are also emphasized in the alignment (Appendix B, Figure B7). Nearly all of the codon sites that correspond to these residues are 100% conserved at the DNA level across all taxa included in the alignment, with the exception of the 6th and 7th thioester intermediate interaction sites (codons 14 and 16). However, at the amino acid level, all sites are 100% conserved, with the 6th and 7th thioester intermediate interaction sites coding for Leucine and Isoleucine, respectively. The single potentially negatively selected site that was detected also resides within this region, but is not, itself, one of the conserved active sites.

3.3.5.8 CODON-LEVEL SELECTION ON *FRAI*

FRAI represents an interesting nucleotide alignment because it was constructed by concatenating sequencing products from two different parts of the gene: exons 5-6 and 24-25

(Appendix B, Figure B8). Because of this, gaps are present in the middle of the alignment that do not represent indels, but rather, are the result of differences in the amount of sequence information that was obtained at the beginning of exon 24 or the end of exon 6 for certain individuals. This should be kept in mind when interpreting the alignment and site by site selection analysis results.

No potentially positively selected sites were detected in the *FRA1* alignment, and, of the 20 potentially negatively selected sites that were detected, only 9 were identified as significant by more than one method (Table 21). These sites are underlined in the alignment in Figure B8 (Appendix B).

Exons 5-6, in particular, show significant sequence homology to a KIF4-like subfamily kinesin motor domain according to BLAST search results (Accession: cd01372; E-value: 1.16e-26, using the D16L sequence). Five of the 9 potentially negatively selected sites are located in this region (Appendix B, Figure B8). The remaining 4 negatively selected sites were found to be located in exons 24 and 25. However, no conserved domains were identified in this section of the alignment, according to BLAST searches.

3.3.5.9 CODON-LEVEL SELECTION ON *LRRK*

The *LRRK* exon 1 alignment is the largest alignment investigated here, at 1188 bp (Table 6; Appendix B, Figure B9). Unfortunately, portions of the sequences for SL8 201S (*T. subulata*) and ES (*T. scabra*) could not be fully resolved due to persistent shifted peaks in the chromatograms for these sequences (Appendix B, Figure B9). These unresolved sections are located at approximately the 620-895 bp positions in the alignment. This portion of the alignment includes indels that these individuals may, perhaps, have been heterozygous for, thus resulting in difficulties sequencing the PCR products that were obtained.

Codon-level selection analyses resulted in the identification of 57 potentially negatively selected sites, and only 3 possibly positively selected sites (codons 239, 288, and 290) (Table 22). However, when only those codons that had been detected by more than one method were considered, no positively selected sites were detected, and only 32 potentially negatively selected sites were identified. These sites are underlined in the alignment in Figure B9 (Appendix B).

Exon 1 of *LRRK* contains sequence with significant homology to a provisional Leucine-rich repeat receptor-like protein kinase multi-domain according to BLAST search results (Accession: PLN00113; E-value: 5.41e-12, using the D16L sequence) (Appendix B, Figure B9). Eight of the 32 potentially negatively selected sites are located within this region, while an additional 3 sites are located upstream of this region. The remaining 21 sites are located downstream of this, and tend to be positioned directly adjacent to other potentially negatively selected sites (often in groups of 2-3). Interestingly, many potentially negatively selected sites tend to be located near regions of the alignment that contain indels.

3.3.5.10 CODON-LEVEL SELECTION ON *IRX15L*

Site-by-site selection analysis of *IRX15L* revealed 38 potentially negatively selected sites, and 3 that were possibly positively selected (Table 23). When those sites that had been identified by more than one method were considered, 24 negatively and 2 positively selected sites were detected. These sites are emphasized in the alignment in Figure B10 (Appendix B).

Importantly, almost the entire *IRX15L* gene was sequenced here, in all taxa of interest (Appendix B Figure B10). According to BLAST search results, the predicted *IRX15L* gene in *Turnera* has considerable sequence similarity to an uncharacterized plant-specific domain in the polysaccharide biosynthesis domain superfamily (Accession: TIGR01627; E-value: 5.71e-52, using the D16L sequence) (base positions 145-774 in the alignment). All but 6 of the negatively selected sites identified here were found to be contained within this region.

Interestingly, both positively selected sites are also located within this conserved domain (Codons 143 and 200). Most individuals code for a threonine residue at codon position 143. However, within the *Turnera* subseries, the short styled individual of the species, *T. joellii*, appears to be heterozygous at this position, coding either for threonine or alanine. *Turnera chamaedrifolia*, *T. weddelliana*, and *T. diffusa* all code for glycine at this position, while *T. panamensis*, which is most closely related to *T. diffusa*, shares the same codon as the majority of members of the *Turnera* subseries. Indeed, an almost identical pattern is seen at codon 200, with the majority of individuals coding for leucine, *T. joellii* and *T. chamaedrifolia* coding for glutamine, and *T. weddelliana* and *T. diffusa* coding for a proline residue, while *T. panamensis* shares the more typical leucine-coding sequence. According to MEME results, 16% and 100% of

the lineages represented here are expected to evolve with $dN>dS$ at each of these sites, respectively (Codon 143: $dN2=39.98$; $dS=0.00$; Codon 200: $dN2=5.47$; $dS=0.00$).

3.3.5.11 CODON-LEVEL SELECTION ON *FSP*

No positively selected sites were detected in the *FSP*, exon 5-6 alignment (Table 24). Rather, only 3 potentially negatively selected sites were identified, of which only one was shown to be significant by more than one method (Codon 27). This site was detected by all four methods employed here and is emphasized in the alignment in Figure B11 (Appendix B).

According to BLAST search results, exon 5 and 6 of *FSP* has considerable sequence similarity to an undefined oxidoreductase-related domain (Accession: PLN02485; E-value: $3.29e-27$, using the D16L sequence) (Appendix B, Figure B11, base positions 7-183). Codon 27, the only statistically significant selected codon identified here, is located within this region (exon 6).

3.3.5.12 CODON-LEVEL SELECTION ON *NRFP*

When site-by-site variations in evolutionary rate in the *NRFP* exon 3 alignment were investigated, 27 negatively and 1 positively selected codon were detected (Table 25). Of these, only 8 potentially negatively selected sites and no positively selected sites were identified by more than one method. The locations of these sites are identified in the alignment in Figure B12 (Appendix B).

Based on BLAST search results, exon 3 of *NRFP* has considerable sequence homology to a dihydrouridine synthase like FMN-binding domain, including all 4 of the conserved residues that form its phosphate binding site (BLAST Accession: cd02801; E-value: $4.50e-22$, using the D16L sequence) (Appendix B, Figure B12, base positions 37-339). Five of the 8 negatively selected sites that were detected here lie within this region. None of the 4 binding site residues were recognized as being negatively selected, though they are all 100% conserved across all taxa included in the alignment at the amino acid level. While a minor amount of sequence variation is present at codon 103 (the third conserved codon in the phosphate binding site), sequences for all individuals code for the same amino acid at this position (alanine).

3.3.5.13 CODON-LEVEL SELECTION ON *WRKY*

Seventeen negatively and 6 positively selected codons were identified in the *WRKY* exon 1 alignment (Table 26). Of these, 8 potentially negatively selected and 2 potentially positively selected sites (codons 100 and 107) were detected by more than one method. These 10 sites have been highlighted in the alignment in Figure B13 (Appendix B).

According to BLAST searches using the DNA data obtained here, exon 1 of *WRKY* does not code for any known conserved domains. According to previously obtained genetic data, in *T. subulata*, the *WRKY* gene is predicted to contain a *WRKY* DNA binding domain (Accession: smart00774; E-value: 8.22e-12, using the D16L sequence), as well as an associated plant-specific zinc cluster domain (Accession: pfam10533; E-value: 5.60e-08, using the D16L sequence). However, the coding sequence for these domains is located downstream of exon 1, which was sequenced here.

Codon 100 was determined to be positively selected using all methods except SLAC (Table 26). According to MEME results, 50% of the branches in the phylogenetic tree are expected to evolve with $dN > dS$ ($dN2=18.90$; $dS=0.00$) at this codon. Individual sequences were found to code for one of three amino acids: alanine, threonine, or isoleucine (Appendix B, Figure B13). However, only individuals of the species, *T. chamaedrifolia*, *T. diffusa*, *T. weddelliana*, and *T. panamensis*, possess an isoleucine at this site.

Codon 107 was also identified as potentially positively selected by all methods, excepting SLAC (Table 26). According to MEME results, 100% of lineages are predicted to evolve with $dN > dS$ at this codon ($dN2=8.53$; $dS=0$). At this site in the alignment, codons for the amino acids alanine, valine, and proline are represented (Appendix B, Figure B13). However, the individual from the species, *T. weddelliana*, is the only one that codes for valine at this site.

It is also interesting to note that the potentially positively selected codons identified here are located on either side of an indel(s) in the alignment (and, particularly, ones that are specific to *T. panamensis*, *T. diffusa*, and *T. chamaedrifolia*) (Appendix B, Figure B13). The negatively selected sites, on the other hand, are distributed fairly evenly across the alignment.

3.3.6 POPULATION-LEVEL ANALYSES: MACDONALD KREITMAN TESTS (MKTS) AND TAJIMA'S D TESTS OF NEUTRALITY

As was earlier discussed, little to no species-level resolution was obtained within the Turnera subseries group when the genetic information that had been obtained for these individuals was submitted to various phylogenetic analyses (Figures 8-10; Appendix D, Figures D1-D3; Appendix E, Figures E1-E44). This result, in combination with the fact that these species also occupy a continuous habitat range in nature (Arbo 2005) and are capable of interbreeding with some degree of fertility (Shore and Barrett 1985a; Arbo and Fernandez 1987; Fernandez and Arbo 1990; Fernandez and Arbo 1993; Lopez et al. 2013) suggests that the application of population-level selection analyses to these data may be appropriate. To do this, data for all diploid individuals of this group were considered. Specifically, genetic information obtained from representatives of the species *T. subulata*, *T. scabra*, *T. krapovickasii*, and *T. concinna* were included. While *T. grandidentata* is also a member of the Turnera subseries, the representative individuals of this species are tetraploid, and were thus not included in the analysis. This was also the case for some individuals of *T. subulata* and *T. scabra* (Table 1). Specifically, tetraploid data were excluded because, where sequence ambiguities existed, haplotypes could not be reliably reconstructed (Stephens et al. 2001; Stephens and Donnelly 2003). Also, as their tetraploid status already more-or-less prevents the exchange of genetic information with other non-tetraploid members of the Turnera subseries, together, they could not be properly considered as part of a “population” sample.

McDonald-Kreitman tests (MKTs) were performed for all 12 *S*-linked gene data sets of interest, using sequences obtained for *T. panamensis* as the inter-specific sample (Table 27). However, no significant results were obtained for any alignment ($p > 0.05$). Rather, the ratio of non-synonymous to synonymous variation within and between species was not found to be significantly different. In the case of *FSP*, a test of significance could not be completed due to a lack of polymorphism within the population sample.

Interestingly, however, neutrality index values obtained for *APETALA2*, *Tssta1*, *SCE1*, *IRX15L*, *FSP*, and *WRKY* were all found to be < 1 , indicating a deviation from the neutral expectation in the direction of positive selection for these alignments (Table 27). According to values of α given for each gene, approximately 45.8%, 75%, 100%, 11.6%, 50%, and 36.4% of amino acid substitutions are expected to be driven by positive selection for each of these

alignments, respectively. However, as indicated by the insignificant Fisher's exact test results described above, ratios of synonymous and non-synonymous substitutions that are between species and polymorphic within the population do not deviate significantly from the neutral expectation.

In order to further investigate population-level selection on *Tsst1*, sequence data that had been obtained from a more local population of tetraploid *T. scabra* was considered (17 individuals from the Dominican Republic). As no sequence ambiguities existed in the data, haplotypes did not need to be reconstructed, and the data were directly submitted to an MKT. As with the above, *T. panamensis* was again used as an outgroup sequence. However, within the population sample, no sequence variation was identified within the coding sequence for this gene (Appendix H, Figure H). As a result, a significance test could not be completed. When the population sample sequences were compared to that of *T. panamensis*, however, most fixed differences were found to result in synonymous substitutions (Appendix H, Table H1).

Tajima's D tests of neutrality were also not found to be significant for any alignment ($p > 0.1$; Table 27), suggesting that evolutionary rates did not differ significantly from the neutral expectation for an S-linked gene.

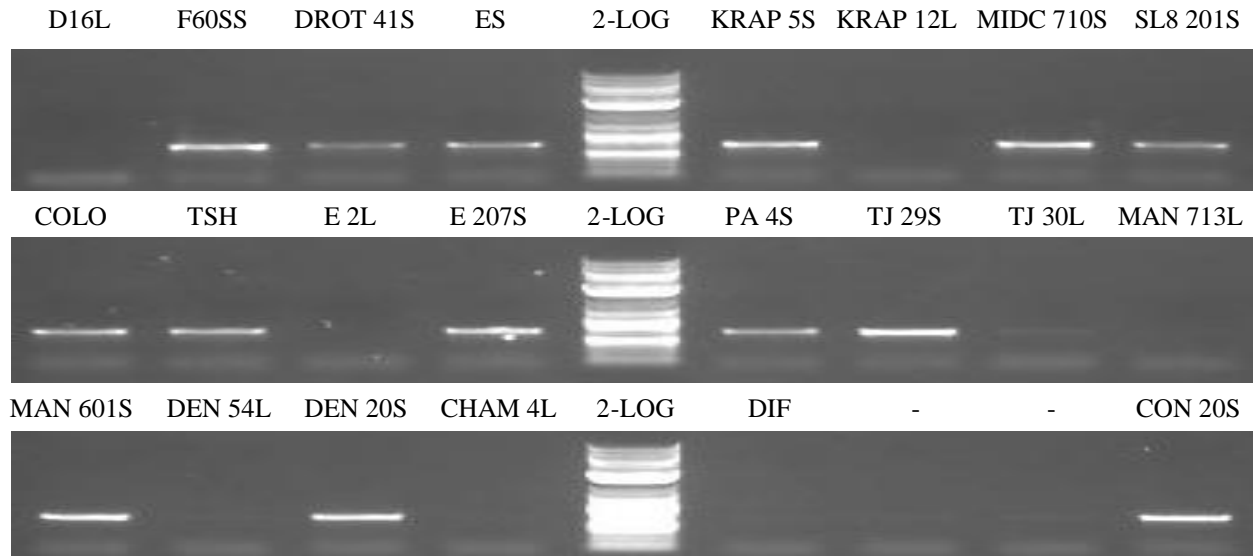


Figure 5: Assay of *Tssta1* in a variety of long- and short-styled individuals of the genus, *Turnera*. PCR products were amplified using *Tssta1* primer pair F3/R1 and run on a 0.8% agarose gel. The expected PCR product amplified only in short-styled individuals, regardless of species or population of origin. Negative controls were run in lanes identified with “-“. 2-log DNA ladder was run in the centre lane of each row in order to determine approximate PCR product sizes. PCR products shown are ~500 bp in length.

Table 6: Descriptive statistics for alignments of all genes. For each gene, the total number of taxa, alignment length (bp), and number of conserved, variable, parsimony informative, and singleton sites is given. The number of identical sequences in each alignment is also provided.

Alignment	DESCRIPTIVE STATISTICS						
	Total # of Taxa Included	Length of Alignment (bp)	# of Conserved Sites (%)	# of Variable Sites (%)	# of Parsimony Informative Sites	# of Singletons	# of Identical Sequences
Total Data	24*†	9489	8226 (86.6)	1137 (12.0)	544	589	0
<i>APETALA2</i>	24	444	380 (85.6)	61 (13.7)	20	41	1
<i>Tssta1</i> (Total Data)	34	447	311 (69.6)	136 (30.4)	70	66	8
<i>Tssta1</i>	16	447	388 (86.8)	59 (13.2)	19	40	5
<i>LEJ2</i>	23†	156	119 (76.3)	35 (22.4)	13	22	5
<i>AP2D</i>	24‡	621	516 (83.2)	87 (14.0)	45	40	0
<i>RNABP</i>	23*†	333	253 (76.0)	68 (20.4)	39	28	3
<i>SCE1</i>	24	177	164 (92.7)	13 (7.3)	6	7	13
<i>FRA1</i>	24	543	500 (92.1)	43 (7.9)	19	24	0
<i>LRRK</i>	24*	1188	922 (77.6)	181 (15.2)	96	84	0
<i>IRX15L</i>	24	777	701 (90.2)	73 (9.4)	40	33	1
<i>FSP</i>	24	183	157 (85.8)	26 (14.2)	9	17	3
<i>NRFP</i>	24	603	543 (90.0)	60 (10.0)	26	34	0
<i>WRKY</i>	24	651	545 (83.7)	103 (15.8)	45	58	0
Total S-Locus Data	24*†	6123	5188 (84.7)	809 (13.2)	377	428	0
<i>ECIP1</i>	24	798	730 (91.5)	68 (8.5)	37	31	1
<i>GAUT3</i>	24	663	607 (91.6)	56 (8.4)	26	30	5
<i>GAUT1</i>	24	465	433 (93.1)	32 (6.9)	17	15	8
<i>RNABP34</i>	24	300	244 (81.3)	56 (18.7)	24	32	4
<i>FMO1</i>	22†	261	226 (86.6)	35 (13.4)	21	14	2
<i>MBD8</i>	21†	354	321 (90.7)	33 (9.3)	25	8	2
<i>UNKN</i>	24	234	200 (85.5)	34 (14.5)	18	16	4
<i>POFUT</i>	24	291	263 (90.4)	28 (9.6)	13	15	6
Total Control Data	24†	3366	3038 (90.3)	328 (9.7)	167	161	0

*Unresolved gaps are present in alignment. The *LRRK* alignment contains a 281 bp gap for ES and a 283 bp gap for SL8 201S. For *RNABP*, a 51 bp gap in the alignment is present for DROT 41S, as well as a 37 bp gap for MIDC 710S. These gaps are also contained in the Total Data and Total S-locus data alignments.

† The alignment is missing sequences for certain taxa. No sequences were obtained for the following individuals (genes): DIF (*LEJ2*, *FMO1*, *MBD8*), KRAP 5S (*RNABP*), WED 2S (*FMO1*, *MBD8*), CHAM 4L (*MBD8*).

‡ An early stop codon was found in the *AP2D* sequence that was obtained for CHAM 4L at the 139th codon position in the alignment.

Figure 6: Nucleotide diversity (π) and Watterson's θ estimates for all genes. A) Nucleotide diversity (π) and B) Watterson's Estimator (θ). Error bars represent standard deviations (SDs). *S*-linked genes are presented in the order in which they are found at the *S*-locus, as it is currently known from the *s*-haplotype. Control genes are shown separately and are in no particular order. In all cases, excepting *Tssta1*, estimates were determined from sequence information obtained from 17 individuals representing 7 species (*T. subulata*, *T. scabra*, *T. grandidentata*, *T. krapovikkassi*, *T. concinna*, *T. joelii*, *T. panamensis*). For *Tssta1*, the same species were represented by 15 individuals.

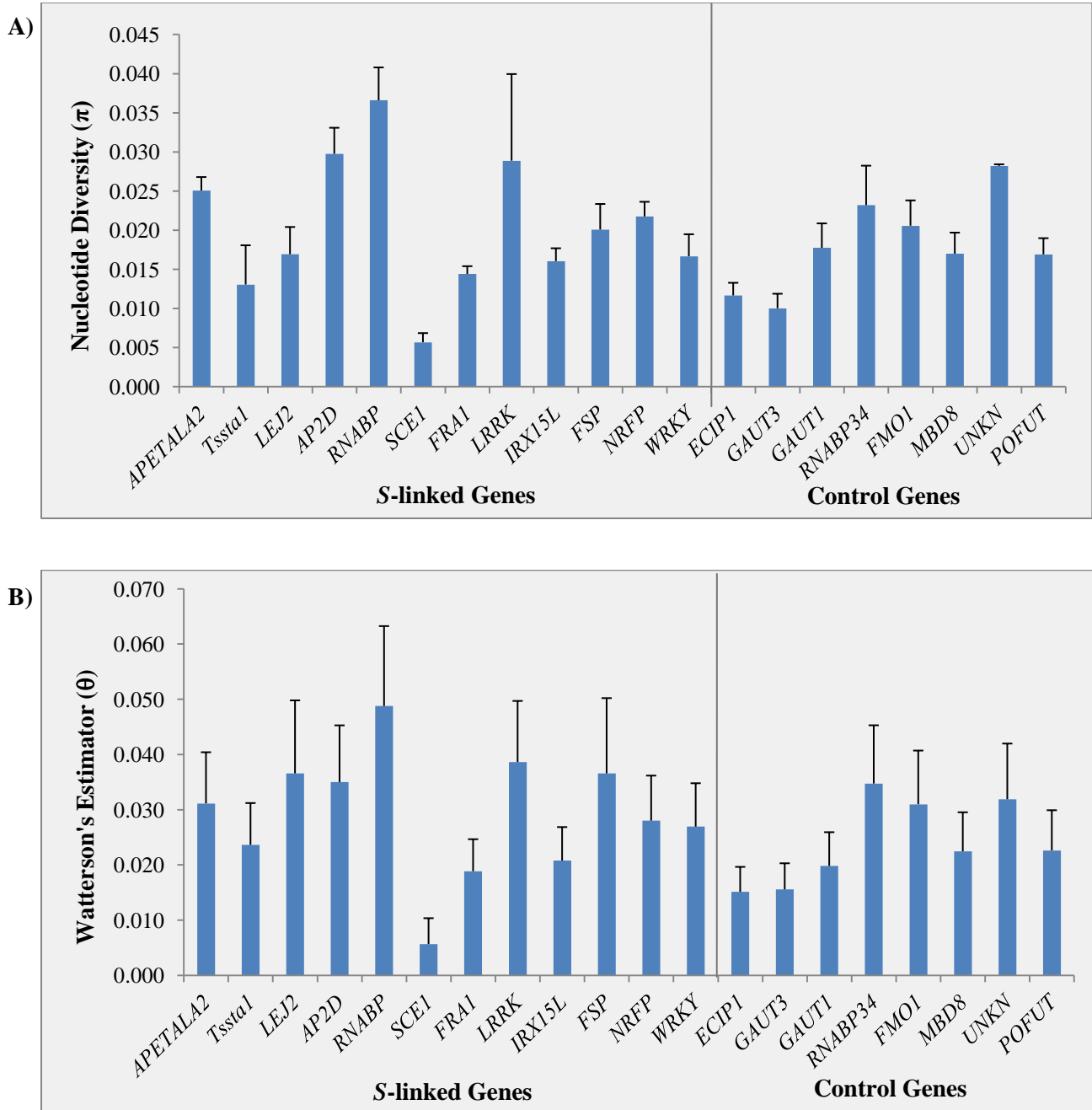


Figure 7: Average nucleotide diversity (π) and Watterson's θ estimates for S-linked and control genes. Error bars represent standard deviations (SDs). When averages were compared using a Mann-Whitney U test, no significant differences were observed between S-linked and control gene data sets in terms of average π or θ ($p > 0.05$). In all cases, excepting *Tssta1*, estimates were determined from sequence information obtained from 17 individuals representing 7 species (*T. subulata*, *T. scabra*, *T. grandidentata*, *T. krapovikkassi*, *T. concinna*, *T. joelii*, *T. panamensis*). For *Tssta1*, the same species were represented by 15 individuals.

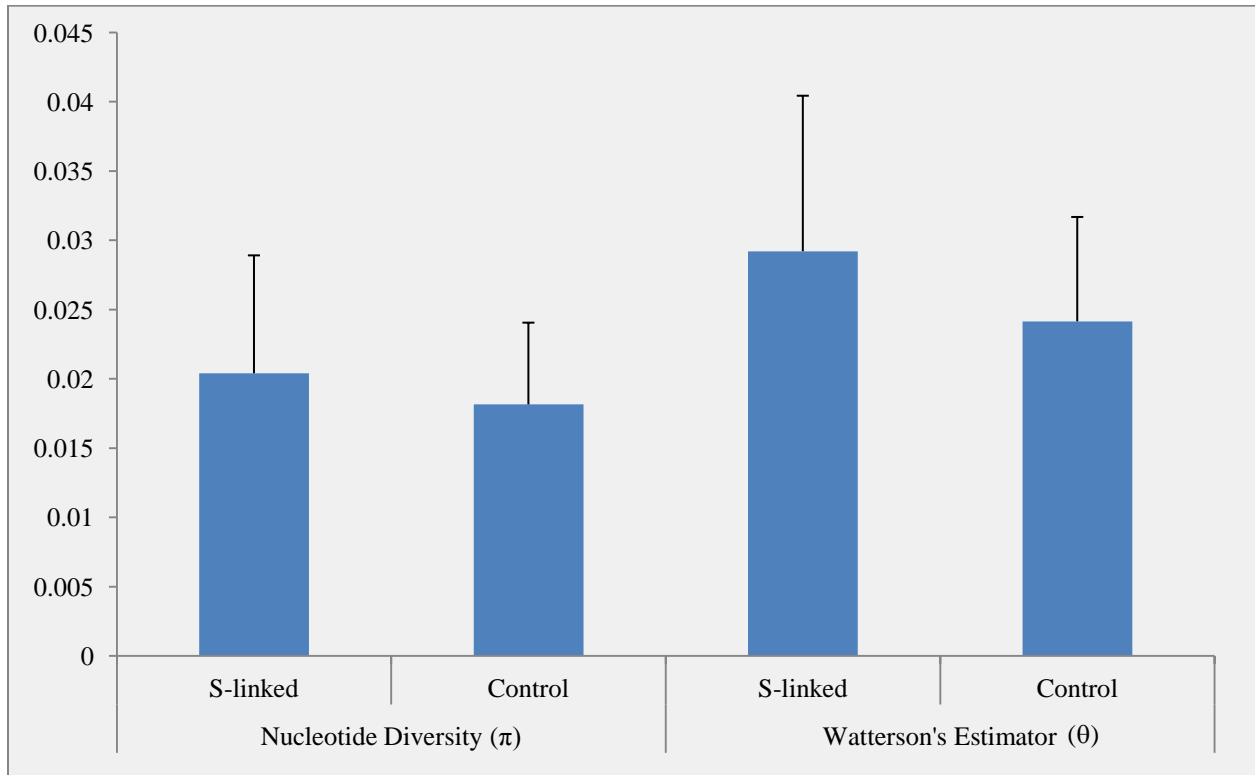


Table 7: Diversity measures for all genes. The total number of base positions (excluding gaps), proportions of polymorphic sites, synonymous and nonsynonymous mutations, as well as the number of indel events, average indel length, and indel diversity per site (π_1) were determined for each alignment. In all cases, excepting *Tsstal*, estimates were determined from sequence information obtained from 17 individuals, representing 7 species (*T. subulata*, *T. scabra*, *T. grandidentata*, *T. krapovikkassi*, *T. concinna*, *T. joelii*, *T. panamensis*). For *Tsstal*, the same species were represented by 15 individuals. Mann-Whitney-U tests were performed in order to determine if mean values were significantly different between *S*-linked and control gene data sets. Significantly different means ($p < 0.05$) are shown in bold.

Data set	Gene	# of Sites (Excluding Gaps)	Proportion Polymorphic Sites	Proportion Synonymous Mutations	Proportion Nonsynonymous Mutations	Number of Indel Events	Average Indel Length (bp)	Indel Diversity (π_1)
S-linked	<i>APETALA2</i>	435 (402)	0.142	0.466	0.534	8	8.739	0.004
	<i>Tsstal</i>	447 (438)	0.100	0.591	0.409	2	6.000	0.000
	<i>LEJ2</i>	96	0.167	0.438	0.563	0	0.000	0.000
	<i>AP2D</i>	492 (468)	0.158	0.515	0.485	7	4.500	0.005
	<i>RNABP</i>	333 (270)	0.222	0.482	0.518	8	9.437	0.003
	<i>SCE1</i>	177	0.051	0.889	0.111	0	0.000	0.000
	<i>FRA1</i>	489	0.086	0.619	0.381	0	0.000	0.000
	<i>LRRK</i>	654 (636)	0.175	0.609	0.391	6	4.684	0.003
	<i>IRX15L</i>	777 (771)	0.095	0.744	0.256	2	3.000	0.000
	<i>FSP</i>	84	0.167	0.500	0.500	0	0.000	0.000
	<i>NRFP</i>	603	0.128	0.592	0.408	0	0.000	0.000
	<i>WRKY</i>	651 (603)	0.123	0.494	0.506	8	4.839	0.001
	S-linked (Mean)	-	0.134	0.578	0.422	3.417	3.433	1.42E-03
Control	<i>ECIP1</i>	798	0.069	0.885	0.115	0	0.000	0.000
	<i>GAUT3</i>	663	0.071	0.739	0.353	0	0.000	0.000
	<i>GAUT1</i>	465	0.090	0.705	0.295	0	0.000	0.000
	<i>RNABP34</i>	300 (297)	0.158	0.510	0.490	1	3.000	0.002
	<i>FMO1</i>	255	0.141	0.605	0.395	0	0.000	0.000
	<i>MBD8</i>	342	0.102	0.486	0.514	0	0.000	0.000
	<i>UNKN</i>	234	0.145	0.382	0.618	0	0.000	0.000
	<i>POFUT</i>	291	0.103	0.676	0.324	0	0.000	0.000
		Control (Mean)	-	0.110	0.624	0.388	0.125	0.375

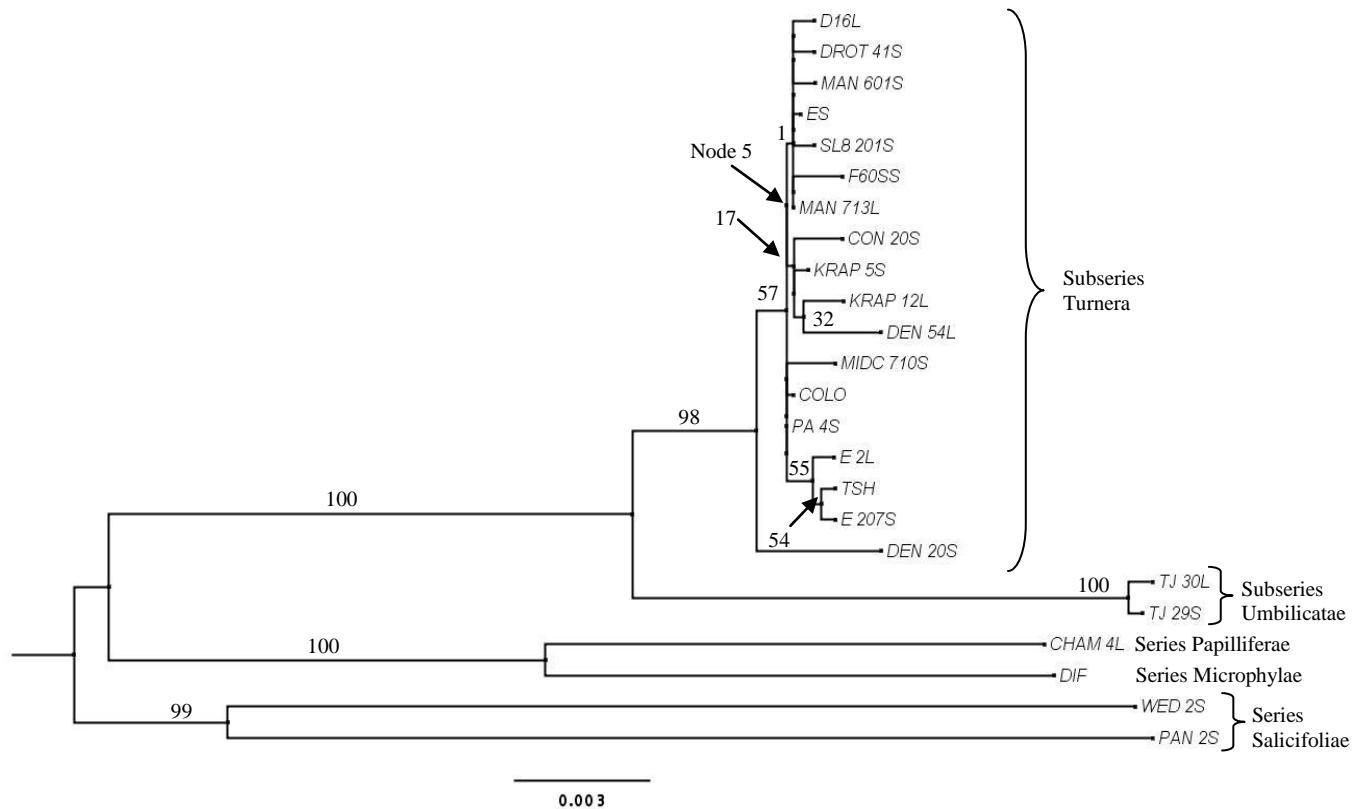


Figure 8: Molecular phylogeny constructed using sequence data for all genes. Evolutionary relationships were inferred using the maximum likelihood method based on the Tamura 3-parameter nucleotide substitution model (Tamura 1992). The tree with the highest log likelihood (-13664.792) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all branch supports are shown. Those that are not shown are $\leq 10\%$. Nodes are indicated by small “■” symbols. Node 5 is indicated for reference. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.126)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 6526 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). Series/Subseries membership is indicated to the right of the tree. Note that subseries Turnera and subseries Umbilicatae both belong to series Turnera.

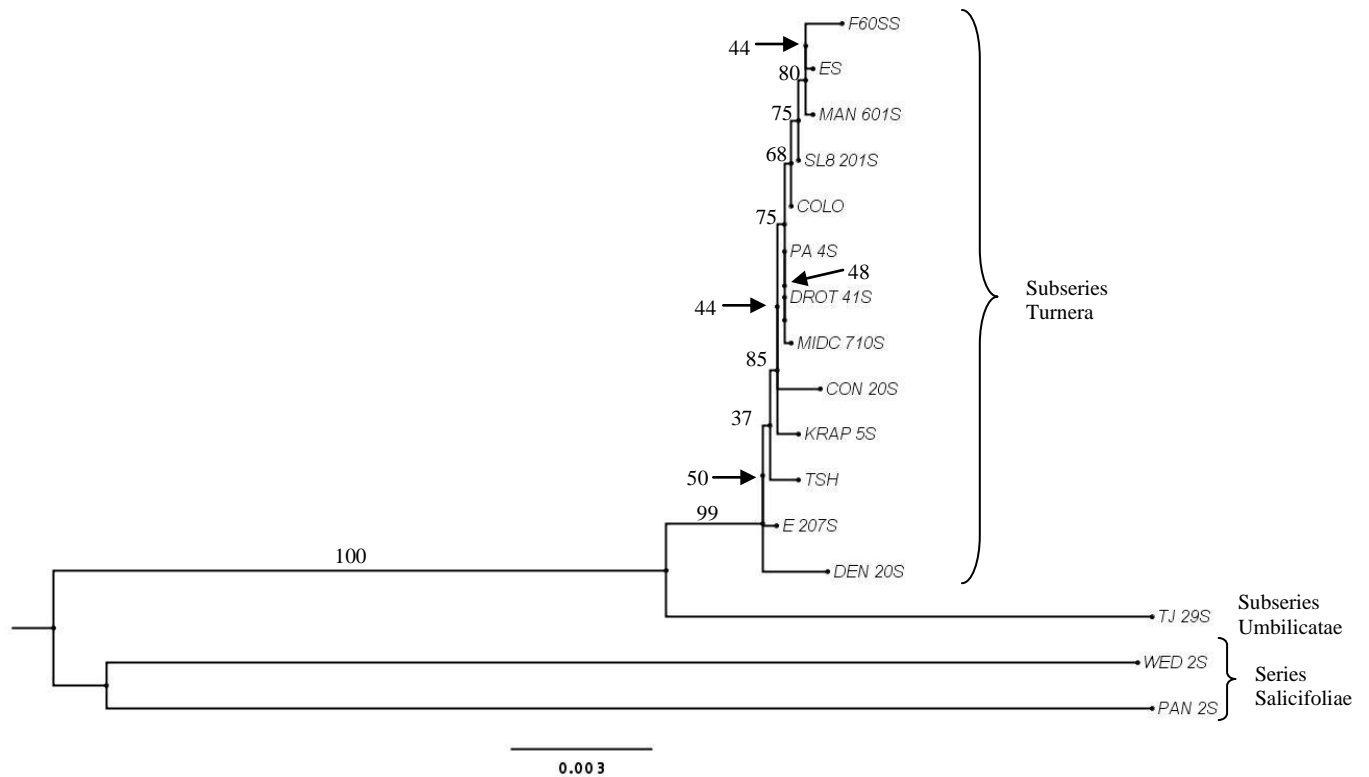


Figure 9: Molecular phylogeny constructed using sequence data for all genes, but with a reduced number of taxa. This tree was computed specifically for analyses of *Tsst1* in HYPHY. The data used to compute this phylogeny are identical to that which were used to produce the tree in Figure 8, except that data from long-styled individuals were excluded. Evolutionary relationships were inferred using the maximum likelihood method based on the Tamura 3-parameter nucleotide substitution model (Tamura 1992). The tree with the highest log likelihood (-11879.077) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all branch supports are shown. Those that are not shown are $\leq 25\%$. Nodes are indicated by small “■” symbols. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.292)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 6467 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). Series/Subseries membership is indicated to the right of the tree. Note that subseries Turnera and subseries Umbilicatae both belong to series Turnera.

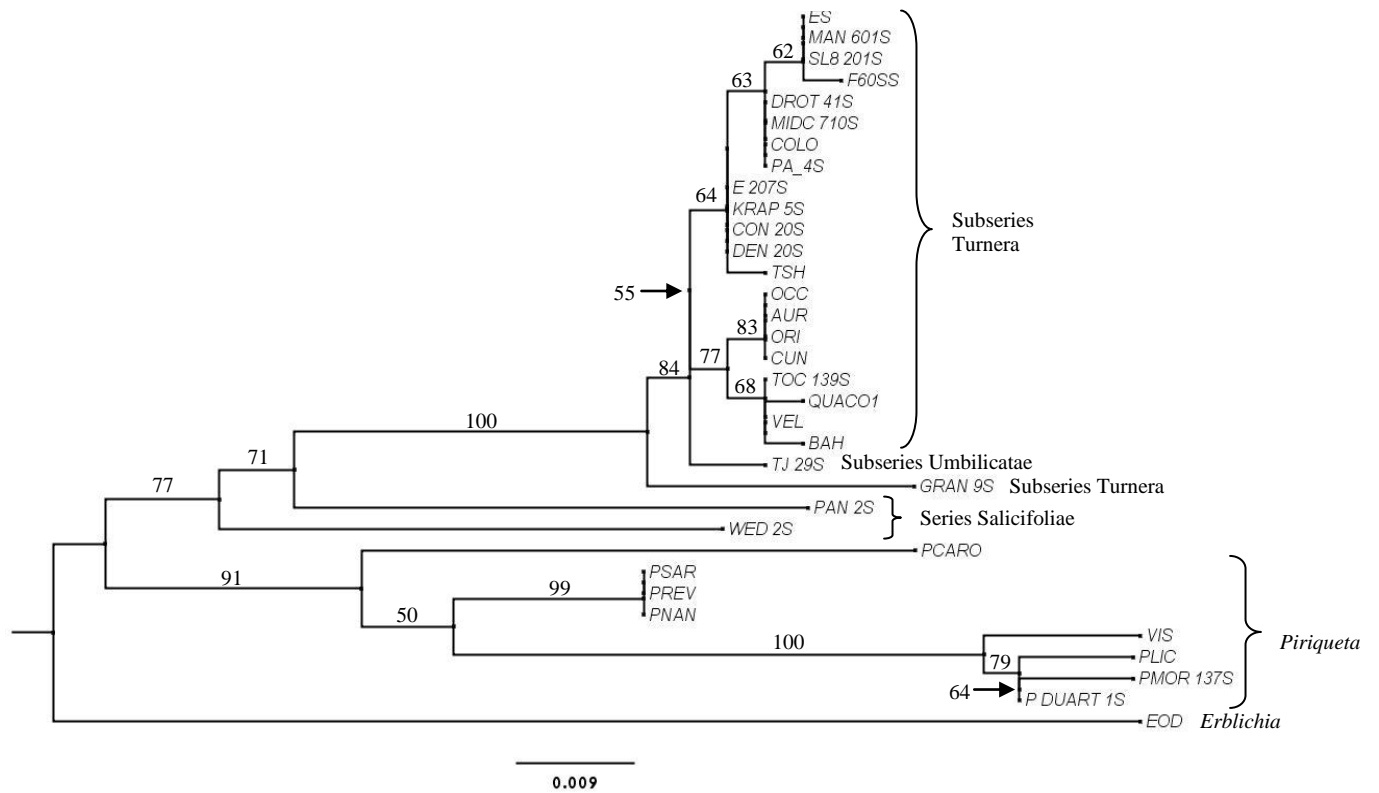


Figure 10: Molecular phylogeny constructed using all sequence data for *Tsst1*. Evolutionary relationships were inferred using the maximum likelihood method based on the Kimura 2-parameter nucleotide substitution model (Kimura 1980). The tree with the highest log likelihood (-1266.836) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all branch supports are shown. Those that are not shown are < 50%. Nodes are indicated by small “■” symbols. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.880)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 34 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 348 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al 2013). Series/Subseries/Genus membership is indicated to the right of the tree. Note that subseries Turnera and subseries Umbilicatae both belong to series Turnera.

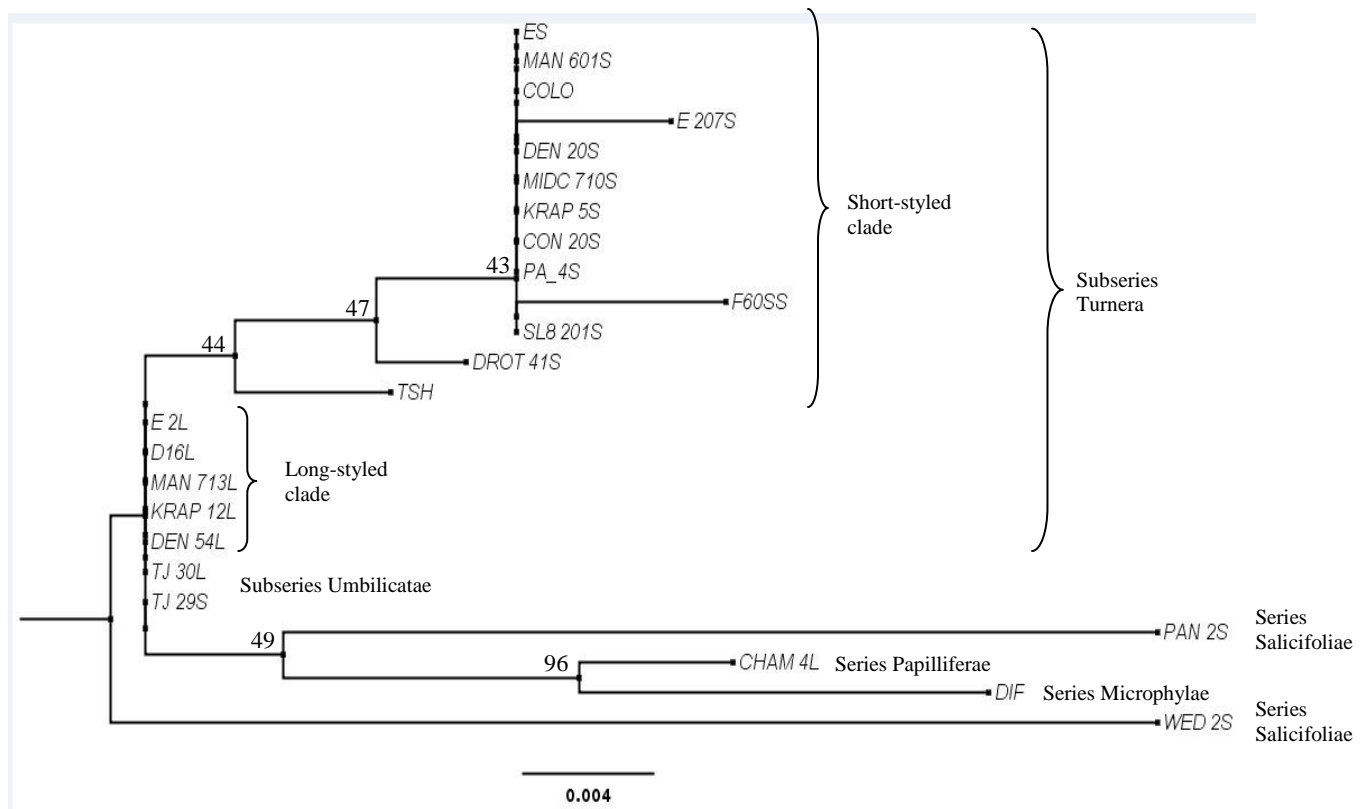


Figure 11: Gene genealogy for *APETALA2* showing evidence of trans-specific evolution in the Turnera subseries clade. Evolutionary relationships were inferred by using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano model (Hasegawa et al. 1985). The tree with the highest log likelihood (-727.2067) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all branch supports are shown. Those that are not shown are < 50%. Nodes are indicated by small “■” symbols. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. Codon positions included were 1st+2nd+3rd+Noncoding. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 371 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). Series/Subseries/Genus membership is indicated to the right of the tree. Note that subseries Turnera and subseries Umbilicatae both belong to series Turnera. Within subseries Turnera, short-styled and long-styled clades are also indicated.

Table 8: Global dN/dS ratio estimates and likelihood ratio tests of dN=dS. Global dN/dS estimates are given for each alignment, along with their associated asymptotic normal 95% CI values. For each alignment, the null hypothesis that dN=dS (or dN/dS=1) was tested against the alternative hypothesis that dN ≠ dS using a likelihood ratio test (LRT). Log likelihood values and numbers of estimated parameters for each model are stated, along with the corresponding LRT statistic for each test. Significant p-values are indicated in bold, suggesting that dN ≠ dS for the relevant alignments.

Alignment	Global dN/dS		Likelihood Ratio Test of dN=dS					
	Global dN/dS Estimate	95% CI (±)	Log Likelihood (dN=dS)	# of Parameters (dN=dS)	Log Likelihood (dN ≠ dS)	# of Parameters (dN ≠ dS)	LRT Statistic	p-value
<i>APETALA2</i>	0.350	0.143	-1132.418	55	-1121.840	56	21.156	0.000
<i>Tssta1</i> (Total Data)	0.318	0.100	-1601.089	73	-1576.319	74	49.539	0.000
<i>Tssta1</i>	0.227	0.120	-936.179	39	-936.179	40	28.253	0.000
<i>LEJ2</i>	1.592	1.074	-468.793	54	-467.746	55	2.095	0.148
<i>AP2D</i>	0.369	0.126	-1747.544	56	-1732.856	57	29.377	0.000
<i>RNABP</i>	0.249	0.085	-1169.629	54	-1137.326	55	64.607	0.000
<i>SCE1</i>	0.023	0.047	-329.407	54	-313.711	55	31.393	0.000
<i>FRA1</i>	0.088	0.055	-1119.120	55	-1085.764	56	66.712	0.000
<i>LRRK</i>	0.179	0.045	-3110.113	55	-3022.774	56	174.678	0.000
<i>IRX15L</i>	0.063	0.030	-1798.575	56	-1716.091	57	164.968	0.000
<i>FSP</i>	0.293	0.198	-448.004	54	-441.697	55	12.614	0.000
<i>NRFP</i>	0.190	0.093	-1258.169	55	-1233.177	56	49.985	0.000
<i>WRKY</i>	0.503	0.178	-1670.271	55	-1663.413	56	13.715	0.000
Total S-Locus Data	0.245	0.029	-16126.245	55	-15859.210	56	534.070	0.000
<i>ECIP1</i>	0.036	0.025	-1635.448	55	-1551.620	56	167.656	0.000
<i>GAUT3</i>	0.100	0.059	-1297.590	55	-1262.841	56	69.499	0.000
<i>GAUT1</i>	0.009	0.269	-850.374	55	-805.580	56	89.587	0.000
<i>RNABP34</i>	0.350	0.171	-765.584	55	-757.126	56	16.915	0.000
<i>FMO1</i>	0.174	0.110	-585.060	54	-571.356	55	27.409	0.000
<i>MBD8</i>	0.307	0.001	-678.931	56	-674.408	57	9.047	0.003
<i>UNKN</i>	0.577	0.315	-646.256	55	-644.749	56	3.016	0.083
<i>POFUT</i>	0.056	0.050	-619.673	55	-589.271	56	60.804	0.000
Total Control Data	0.150	0.031	-7325.466	55	-7153.880	56	343.171	0.000

Table 9: Comparing selection on S-linked and control genes. Four likelihood ratio tests were performed in order to determine if the S-linked genes of interest had experienced different selective pressures when compared to a random assortment of control genes. For each model (constrained and independent), the log likelihood value and number of estimated parameters is provided. For each test, the corresponding degrees of freedom (DF) and LRT statistic are shown. Significant p-values are indicated in bold. p-values < 0.05 suggest that the independent model fit is superior to that of the constrained model and, further, that the control and experimental gene data sets are significantly different with respect to the relevant constrained parameter(s). All four tests were completed twice, with random and default starting values for parameters, respectively. The results obtained using default starting values are shown below. Tests completed with random starting values converged on nearly identical results (see Appendix F, Table F2).

Constrained Parameter(s)	Log Likelihood (Constrained Model)	# of Parameters (Constrained Model)	Log Likelihood (Independent Model)	# of Parameters (Independent Model)	DF	LRT Statistic	p-value
All rate parameters	-22540.562	116	-22526.647	126	10	27.831	0.002
Selective regimes (+ dN/dS and proportion of + selected sites)	-22527.618	124			2	1.941	0.379
Strength of positive selection (+ dN/dS)	-22526.688	125			1	1.549	0.461
Proportion of positively (+) selected sites	-22527.421	125			1	0.081	0.775

Table 10: Comparing selection on *Tsst1* and other *S*-linked genes. Four likelihood ratio tests were performed in order to determine if *Tsst1* had experienced different selective pressures when compared to other *S*-linked genes. For each model (constrained and independent), the log likelihood value and number of estimated parameters is provided. For each test, the corresponding degrees of freedom (DF) and LRT statistic are shown. p-values < 0.05 suggest that the independent model fit is superior to that of the constrained model and, further, that the *Tsst1* and other *S*-linked gene data sets are significantly different with respect to the relevant constrained parameter(s). No significant results were obtained. All four tests were completed twice, with random and default starting values for parameters, respectively. The results obtained using default starting values are shown below. However, tests completed with random starting values converged on similar results (see Appendix F, Table F4).

Shared Parameters	Log Likelihood (Constrained Model)	# of Parameters (Constrained Model)	Log Likelihood (Independent Model)	# of Parameters (Independent Model)	DF	LRT Statistic	p-value
All rate parameters	-13212.030	84	-13208.923	94	10	6.212	0.797
Selective regimes (+ dN/dS and proportion of + selected sites)	-13208.992	92			2	0.137	0.934
Strength of positive selection (+ dN/dS)	-13208.968	93			1	0.088	0.957
Proportion of positively (+) selected sites	-13208.959	93			1	0.070	0.791

Table 11: Comparing selection on *AP2D* and other *S*-linked genes. Four likelihood ratio tests were performed in order to determine if *AP2D* had experienced different selective pressures when compared to other *S*-linked genes. For each model (constrained and independent), the log likelihood value and number of estimated parameters is provided. For each test, the corresponding degrees of freedom (DF) and LRT statistic are shown. Significant p-values are indicated in bold. p-values < 0.05 suggest that the independent model fit is superior to that of the constrained model and, further, that the *AP2D* and *S*-linked gene data sets are significantly different with respect to the relevant constrained parameter(s). All four tests were completed twice, with random and default starting values, respectively. The results obtained using default starting values are shown below. However, tests completed with random starting values converged on similar results (see Appendix F, Table F6).

Parameters constrained in constrained model	Log Likelihood (Constrained Model)	# of Parameters (Constrained Model)	Log Likelihood (Independent Model)	# of Parameters (Independent Model)	DF	LRT Statistic	p-value
All rate parameters	-15349.666	117	-15331.000	127	10	37.331	0.000
Selective regimes (+ dN/dS and proportion of + selected sites)	-15333.756	125			2	5.511	0.064
Strength of positive selection (+ dN/dS)	-15333.181	126			1	4.361	0.113
Proportion of positively (+) selected sites	-15332.584	126			1	3.168	0.075

Table 12: Comparison of global and local dN/dS ratio models for each alignment. Global and local dN/dS ratio model fits were compared for each alignment using a likelihood ratio test (LRT). Log likelihood values are shown for each model and the degrees of freedom (DF) and LRT statistics are given for each test. Significant p-values (<0.05), shown in bold, suggest that the local dN/dS ratio model fit is superior to that of the global model for the corresponding alignment.

Alignment	Log Likelihood (Global)	Log Likelihood (Local)	DF	LRT Statistic	p-value
<i>APETALA2</i>	-1117.093	-1100.590	44	33.006	0.888
<i>Tssta1</i> (Total Data)	-1566.870	-1547.556	62	38.628	0.991
<i>Tssta1</i>	-926.134	-932.486	28	8.006	1.000
<i>LEJ2</i>	-461.437	-453.479	44	15.916	1.000
<i>AP2D</i>	-1730.612	-1696.027	44	69.171	0.009
<i>RNABP</i>	-1133.299	-1123.864	44	18.868	1.000
<i>SCE1</i>	-311.390	-309.679	44	3.421	1.000
<i>FRA1</i>	-1083.328	-1074.880	44	16.897	1.000
<i>LRRK</i>	-3012.665	-2997.724	44	29.882	0.949
<i>IRX15L</i>	-1715.775	-1705.548	44	20.454	0.999
<i>FSP</i>	-435.157	-421.231	44	27.851	0.973
<i>NRFP</i>	-1232.025	-1219.722	44	24.606	0.992
<i>WRKY</i>	-1648.140	-1662.016	44	27.751	0.974

Table 13: Branch-site Unrestricted Statistical Test of Episodic Diversification (BUSTED). For BUSTED analysis, a likelihood ratio test (LRT) was used to compare the fits of two models: one where all parameters were estimated independently from the data and another where the value of dN/dS was constrained to < 1 at all sites, across all branches of the phylogenetic tree. For each model for each alignment, the log-likelihood value and the number of estimated parameters are given. If no evidence of positive selection was detected under the unconstrained model (i.e.: the proportion of sites with dN/dS>1 was approximately equal to 0), then no constrained model was fitted to the data and an LRT was not performed (NA). For each test, an LRT statistic and p-value were computed. Significant p-values suggest that there is evidence of transient positive selection on a portion of codon sites over a certain proportion of branches in the phylogenetic tree. For these tests, no branches suspected of being under the influence of transient positive selection were indicated, *a priori*.

Alignment	BUSTED						
	Log Likelihood (Independent)	# of Parameters (Independent)	Proportion of sites with dN/dS>1	Log Likelihood (dN/dS<1)	# of Parameters (dN/dS<1)	LRT Statistic	p-value
<i>APETALA2</i>	-1115.828	65	0.125	-1115.922	64	0.187	0.911
<i>Tssta1 (Total)</i>	-1559.586	83	0.011	-1560.867	82	2.561	0.278
<i>Tssta1</i>	-925.357	49	0.115	-925.738	48	0.762	0.683
<i>LEJ2</i>	-459.275	65	0.274	-461.884	64	5.217	0.071
<i>AP2D</i>	-1728.477	65	0.098	-1771.964	64	2.947	0.230
<i>RNABP</i>	-1132.556	65	0.096	-1132.550	64	-0.011	1
<i>SCE1</i>	-311.137	65	0.000	NA	NA	NA	NA
<i>FRA1</i>	-1082.817	65	0.000	NA	NA	NA	NA
<i>LRRK</i>	-3011.648	65	0.024	-3011.658	64	0.020	0.990
<i>IRX15L</i>	-1706.072	65	0.002	-1708.881	64	5.617	0.060
<i>FSP</i>	-435.057	65	0.000	NA	NA	NA	NA
<i>NRFP</i>	-1231.175	65	0.000	NA	NA	NA	NA
<i>WRKY</i>	-1662.092	65	0.101	-1662.092	64	0.001	1.000

Table 14: Integrative site-by-site selection analysis for *APETALA2*. Selection at the level of individual codons within the *APETALA2* alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic and pervasive positive or diversifying selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches in the phylogenetic tree that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*). Conflicting results are indicated in italics.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2- α	P(dN2)	p-value
Negative	9	-0.011	0.005	0.000	1.000	-0.125	0.566	-	-	-
	15*	-2.000	0.111	-7.560	0.030	-3.877	0.932	-	-	-
	29	-0.011	0.004	0.000	1.000	-0.196	0.612	-	-	-
	50*	-4.185	0.014	-5.190	0.030	-1.717	0.908	-	-	-
	65	-0.071	0.010	0.000	1.000	-0.299	0.673	-	-	-
	74	-0.006	0.009	0.000	1.000	-0.155	0.590	-	-	-
	75	-0.044	0.018	0.000	1.000	-0.166	0.602	-	-	-
	79*	-8.737	0.001	-32.300	0.000	-27.139	1.000	-	-	-
	81*	-5.436	0.008	-14.990	0.010	-10.812	0.996	-	-	-
	83*	-6.134	0.031	-24.090	0.020	-23.661	0.994	-	-	-
	84	-0.011	0.004	0.000	1.000	-0.171	0.591	-	-	-
	90	-0.044	0.018	0.000	1.000	-0.177	0.591	-	-	-
	91	-0.011	0.004	0.000	1.000	-0.137	0.576	-	-	-
	106	-4.774	0.025	-4.110	0.290	-1.695	0.760	-	-	-
	109	-0.011	0.004	0.000	1.000	-0.138	0.577	-	-	-
	128	-0.016	0.007	0.000	1.000	-0.125	0.566	-	-	-
139	-2.093	0.141	-6.300	0.040	-1.796	0.888	-	-	-	
143	-0.011	0.004	0.000	1.000	-0.138	0.575	-	-	-	
Positive	39	2.355	0.296	9.000	0.070	4.605	0.931	138.400	0.600	0.060
	40*	3.105	0.122	10.910	0.020	6.622	0.990	10.910	1.000	0.020
	95	<i>-0.241</i>	<i>0.774</i>	0.530	0.910	<i>-0.101</i>	<i>0.568</i>	324.200	0.160	0.050
	103	1.012	0.450	3.120	0.160	0.584	0.754	27.200	0.120	0.030

Table 15: Integrative site-by-site selection analysis for *Tsst1* (total data). Selection at the level of individual codons within the *Tsst1* (total data) alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic and pervasive positive or diversifying selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches in the phylogenetic tree that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*).

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2- α	P(dN2)	p-value
Negative	39	-1.970	0.169	-1.190	0.080	-0.699	0.907	-	-	-
	51	-1.984	0.168	-1.900	0.070	-0.733	0.924	-	-	-
	60*	-1.774	0.199	-3.100	0.040	-0.919	0.918	-	-	-
	68*	-2.000	0.111	-2.350	0.030	-0.811	0.950	-	-	-
	75*	-3.000	0.038	-4.230	0.010	-1.720	0.977	-	-	-
	78*	-3.259	0.042	-4.010	0.020	-0.860	0.953	-	-	-
	85*	-2.382	0.078	-2.930	0.020	-1.423	0.936	-	-	-
	89*	-2.000	0.111	-2.160	0.030	-0.789	0.948	-	-	-
	91*	-2.000	0.111	-2.320	0.030	-0.797	0.942	-	-	-
	112*	-2.705	0.131	-4.090	0.030	-1.022	0.925	-	-	-
	114*	-2.579	0.138	-3.980	0.030	-0.926	0.923	-	-	-
	119	-1.983	0.178	-2.050	0.070	-0.800	0.909	-	-	-
	122*	-2.000	0.111	-2.360	0.030	-0.821	0.945	-	-	-
	125*	-1.629	0.218	-2.790	0.050	-0.896	0.911	-	-	-
	128*	-1.743	0.162	-2.070	0.040	-0.658	0.927	-	-	-
	129*	-2.000	0.111	-1.980	0.030	-0.726	0.945	-	-	-
134*	-1.939	0.118	-2.480	0.030	-0.791	0.928	-	-	-	
139	-1.599	0.209	-1.850	0.080	-0.679	0.905	-	-	-	
Positive	9	2.415	0.296	2.470	0.150	0.561	0.764	18.530	0.150	0.040

Table 16: Integrative site-by-site selection analysis for *Tsstal* (reduced number of taxa). Selection at the level of individual codons within the *Tsstal* alignment was evaluated using three different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), and Fast Unconstrained Bayesian Approximation (FUBAR). For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. For *Tsstal*, no potentially positively selected sites were detected by any method, including Mixed Effects Model of Evolution (MEME). Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*).

Selective Regime	Codon	SLAC		FEL		FUBAR	
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability
Negative	37*	-3.550	0.035	-20.390	0.010	-14.821	0.976
	51*	-1.925	0.173	-21.310	0.010	-10.401	0.936
	63	-1.632	0.204	-8.360	0.040	-4.258	0.894
	66	-1.763	0.189	-8.270	0.040	-4.173	0.890
	99*	-1.994	0.112	-8.170	0.030	-6.313	0.935
	114	-2.451	0.144	-15.070	0.020	-4.259	0.897
	125	-1.632	0.216	-8.360	0.050	-4.282	0.884

Table 17: Integrative site-by-site selection analysis for *LEJ2*. Selection at the level of individual codons within the *LEJ2* alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic and pervasive positive or diversifying selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*). Conflicting results are indicated in italics.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2- α	P(dN2)	p-value
Negative	33*	-3.540	0.043	-13.900	0.010	-7.964	0.952	-	-	-
	41	-2.018	0.118	-8.970	0.040	-2.675	0.894	-	-	-
	46	-0.012	0.005	0.000	1.000	-0.122	0.536	-	-	-
Positive	7	2.828	0.359	13.220	0.170	6.319	0.966	<i>-66.870</i>	<i>0.650</i>	<i>1.000</i>
	8	2.329	0.346	7.900	0.130	2.718	0.927	13.840	0.710	0.130
	12*	2.158	0.198	19.790	0.030	14.678	0.993	19.780	1.000	0.030
	31	1.097	0.459	5.610	0.210	1.312	0.794	43.860	0.140	0.040
	39	1.572	0.287	8.090	0.120	3.144	0.908	8.180	0.990	0.120

Table 18: Integrative site-by-site selection analysis for *AP2D*. Selection at the level of individual codons within the *AP2D* alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic and pervasive positive or diversifying selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*). Conflicting results are indicated in italics.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2- α	P(dN2)	p-value
Negative	16*	-2.672	0.074	-5.010	0.030	-2.992	0.925	-	-	-
	22	-0.011	0.005	0.000	1.000	-0.095	0.541	-	-	-
	24*	-3.725	0.037	-8.820	0.020	-8.528	0.986	-	-	-
	31*	-3.871	0.043	-6.800	0.010	-5.499	0.950	-	-	-
	32*	-2.973	0.061	-18.850	0.000	-15.831	0.986	-	-	-
	61*	-4.000	0.019	-24.260	0.000	-22.475	1.000	-	-	-
	63	-1.000	0.333	-4.790	0.050	-1.969	0.852	-	-	-
	66	-0.044	0.018	0.000	1.000	-0.283	0.603	-	-	-
	67*	-3.250	0.040	-6.610	0.010	-5.724	0.981	-	-	-
	74	-0.044	0.018	0.000	1.000	-0.151	0.574	-	-	-
	82	-0.041	0.023	0.000	1.000	-0.277	0.600	-	-	-
	84*	-6.838	0.002	-16.010	0.000	-16.782	1.000	-	-	-
	105	-0.012	0.005	0.000	1.000	-0.277	0.600	-	-	-
	111	-0.017	0.003	0.000	1.000	-0.494	0.621	-	-	-
	121	-0.023	0.002	0.000	1.000	-0.201	0.588	-	-	-
	132	-0.023	0.002	0.000	1.000	-0.253	0.614	-	-	-
	135*	-9.141	0.001	-14.660	0.000	-16.133	1.000	-	-	-
	136*	-4.143	0.030	-6.650	0.010	-5.064	0.952	-	-	-
138*	-6.357	0.005	-9.790	0.000	-10.176	0.993	-	-	-	
140	-0.035	0.006	0.000	1.000	-0.261	0.589	-	-	-	
155	-2.086	0.276	-11.780	0.040	-6.610	0.875	-	-	-	

Continued from previous page.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2- α	P(dN2)	p-value
Positive	17	3.208	0.125	6.740	0.080	8.188	0.982	6.770	1.000	0.080
	20*	4.897	0.027	11.050	0.000	11.149	0.999	11.050	1.000	0.000
	26	1.906	0.459	3.790	0.190	2.592	0.909	3.790	1.000	0.190
	77	0.552	0.569	-0.280	0.930	0.083	0.520	2782.650	0.040	0.020
	141	2.865	0.293	4.970	0.130	4.501	0.958	4.970	1.000	0.130
	146	2.361	0.320	4.650	0.080	4.298	0.967	4.650	1.000	0.080
	186	1.993	0.394	4.370	0.130	2.911	0.892	262.800	0.080	0.010

Table 19: Integrative site-by-site selection analysis for *RNABP*. Selection at the level of individual codons within the *RNABP* alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic and pervasive positive or diversifying selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*).

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2-dS	P(dN2)	p-value
Negative	5*	-12.397	0.000	-17.060	0.000	-13.768	0.999	-	-	-
	10*	-4.483	0.020	-5.000	0.010	-3.906	0.976	-	-	-
	15*	-2.003	0.111	-2.280	0.020	-1.062	0.943	-	-	-
	18*	-6.376	0.018	-10.760	0.000	-10.254	0.985	-	-	-
	29*	-4.322	0.013	-4.680	0.000	-6.453	0.996	-	-	-
	31*	-3.004	0.064	-3.030	0.010	-2.412	0.970	-	-	-
	32*	-7.254	0.007	-13.730	0.000	-12.073	1.000	-	-	-
	37*	-5.007	0.004	-5.100	0.000	-4.350	0.996	-	-	-
	45	-0.019	0.001	0.000	1.000	-0.483	0.709	-	-	-
	46*	-1.577	0.212	-1.930	0.050	-1.117	0.916	-	-	-
	60*	-2.710	0.111	-2.640	0.040	-1.066	0.909	-	-	-
	61*	-2.710	0.111	-2.230	0.030	-0.941	0.932	-	-	-
	64*	-7.917	0.003	-38.910	0.000	-27.293	1.000	-	-	-
	65	-0.052	0.002	0.000	1.000	-0.539	0.691	-	-	-
	81	-0.009	0.007	0.000	1.000	-0.065	0.527	-	-	-
	84*	-7.135	0.001	-12.420	0.000	-9.863	1.000	-	-	-
	86*	-5.118	0.007	-5.990	0.000	-9.313	0.999	-	-	-
	87	-2.003	0.111	-2.070	0.040	-0.706	0.890	-	-	-
	90*	-3.004	0.037	-3.400	0.010	-1.487	0.970	-	-	-
	97	-0.037	0.001	0.000	1.000	-0.513	0.740	-	-	-
102*	-3.188	0.105	-4.070	0.020	-2.380	0.938	-	-	-	
103*	-3.905	0.020	-3.380	0.010	-1.262	0.942	-	-	-	

Continued from previous page.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2-dS	P(dN2)	p-value
Positive	51	2.180	0.3477	4.230	0.170	3.677	0.958	4.230	1.000	0.170
	52*	2.391	0.169	2.510	0.050	1.115	0.913	2.510	1.000	0.050
	66*	3.661	0.088	4.130	0.020	2.244	0.972	12.070	0.580	0.010

Table 20: Integrative site-by-site selection analysis for *SCE1*. Selection at the level of individual codons within the *SCE1* alignment was evaluated using three different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), and Fast Unconstrained Bayesian Approximation (FUBAR). For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. For *SCE1*, no potentially positively selected sites were detected by any method, including Mixed Effects Model of Evolution (MEME). Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*).

Selective Regime	Codon	SLAC		FEL		FUBAR	
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability
Negative	3	-2.618	0.127	-9.830	0.040	-4.883	0.842
	8	-0.011	0.004	0.000	1.000	-0.115	0.516
	18	-2.618	0.127	-7.930	0.050	-3.850	0.841
	22*	-2.000	0.113	-7.250	0.030	-7.485	0.948

Table 21: Integrative site-by-site selection analysis for *FRAI*. Selection at the level of individual codons within the *FRAI* alignment was evaluated using three different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), and Fast Unconstrained Bayesian Approximation (FUBAR). For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. For *FRAI*, no potentially positively selected sites were detected by any method, including Mixed Effects Model of Evolution (MEME). Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*).

Selective Regime	Codon	SLAC		FEL		FUBAR	
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability
Negative	20	-0.027	0.002	0.000	1.000	-0.707	0.600
	24	-0.027	0.002	0.000	1.000	-0.707	0.600
	26*	-8.328	0.003	-593.570	0.000	-36.389	1.000
	33	-0.242	0.019	0.000	1.000	-0.739	0.606
	42	-0.098	0.039	0.000	1.000	-0.194	0.554
	50*	-2.000	0.111	-6.900	0.030	-4.482	0.934
	60	-0.011	0.004	0.000	1.000	-0.199	0.551
	72*	-6.079	0.006	-16.390	0.000	-10.967	0.974
	73	-0.011	0.004	0.000	1.000	-0.202	0.558
	76*	-2.000	0.111	-7.130	0.050	-8.250	0.957
	84	-2.214	0.178	-11.580	0.030	-9.682	0.899
	86	-0.016	0.007	0.000	1.000	-0.185	0.548
	88*	-2.501	0.136	-25.790	0.020	-10.812	0.904
	106	-0.012	0.005	0.000	1.000	-0.228	0.566
	115*	-3.244	0.040	-12.330	0.010	-9.850	0.991
	124	-0.024	0.002	0.000	1.000	-0.492	0.585
	126*	-4.696	0.012	-25.600	0.000	-29.479	1.000
157*	-6.657	0.007	-18.550	0.000	-21.115	0.998	
161	-2.412	0.138	-13.850	0.020	-7.141	0.884	
164*	-3.498	0.078	-11.490	0.020	-14.223	0.975	

Table 22: Integrative site-by-site selection analysis for *LRRK*. Selection at the level of individual codons within the *LRRK* alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic and pervasive positive or diversifying selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*). Conflicting results are indicated in italics.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2- α	P(dN2)	p-value
Negative	1	-0.018	0.016	0.000	1.000	-0.300	0.680	-	-	-
	3*	-4.538	0.083	-9.990	0.010	-7.256	0.984	-	-	-
	19*	-8.944	0.033	-44.900	0.000	-24.042	0.995	-	-	-
	24*	-5.265	0.127	-12.960	0.000	-4.810	0.915	-	-	-
	42	-0.025	0.006	0.000	1.000	-0.344	0.706	-	-	-
	75	-0.022	0.005	0.000	1.000	-0.305	0.689	-	-	-
	116	-0.022	0.005	0.000	1.000	-0.309	0.690	-	-	-
	117*	-3.254	0.111	-5.890	0.030	-3.425	0.954	-	-	-
	118*	-3.209	0.119	-4.810	0.040	-2.059	0.940	-	-	-
	119*	-3.749	0.167	-10.990	0.020	-3.296	0.926	-	-	-
	123	-0.022	0.005	0.000	1.000	-0.305	0.690	-	-	-
	130	-0.238	0.006	0.000	1.000	-0.668	0.738	-	-	-
	136*	-14.128	0.000	-80.110	0.000	-18.781	0.996	-	-	-
	142*	-3.804	0.039	-5.030	0.020	-2.980	0.957	-	-	-
	143*	-3.694	0.071	-6.010	0.010	-2.802	0.973	-	-	-
	156*	-2.661	0.125	-7.910	0.030	-3.129	0.931	-	-	-
	163*	-3.698	0.031	-6.290	0.010	-4.903	0.988	-	-	-
	164	-5.217	0.005	-1.340	0.140	-0.647	0.887	-	-	-
	167	-0.011	0.004	0.000	1.000	-0.245	0.689	-	-	-
	168*	-1.651	0.163	-3.610	0.030	-1.742	0.942	-	-	-
169*	-4.305	0.015	-4.100	0.020	-2.429	0.954	-	-	-	
172	-1.000	0.333	-1.790	0.090	-0.767	0.909	-	-	-	
185*	-3.060	0.123	-16.520	0.010	-4.865	0.941	-	-	-	
189	-1.560	0.259	-2.040	0.090	-0.657	0.900	-	-	-	

Continued from previous page.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2- α	P(dN2)	p-value
Negative	191*	-2.661	0.131	-7.310	0.020	-2.810	0.942	-	-	-
	196	-0.011	0.004	0.000	1.000	-0.247	0.690	-	-	-
	198*	-5.206	0.024	-2.240	0.070	-0.778	0.910	-	-	-
	214	-4.957	0.044	-4.510	0.070	-2.122	0.872	-	-	-
	222	-0.013	0.005	0.000	1.000	-0.305	0.730	-	-	-
	230*	-2.766	0.073	-4.660	0.020	-2.601	0.961	-	-	-
	263	-0.073	0.014	0.000	1.000	-0.307	0.721	-	-	-
	270	-0.016	0.005	0.000	1.000	-0.277	0.717	-	-	-
	277*	-5.880	0.007	-33.530	0.000	-29.317	1.000	-	-	-
	281*	-2.671	0.125	-3.780	0.030	-1.087	0.933	-	-	-
	282	-0.013	0.005	0.000	1.000	-0.246	0.689	-	-	-
	284*	-2.923	0.052	-6.020	0.010	-3.062	0.965	-	-	-
	292	-2.847	0.155	-3.810	0.060	-1.152	0.906	-	-	-
	295*	-2.313	0.153	-5.380	0.040	-1.270	0.917	-	-	-
	296*	-4.016	0.012	-5.100	0.000	-3.453	0.991	-	-	-
	297*	-8.312	0.014	-27.690	0.000	-23.108	0.997	-	-	-
	304*	-2.134	0.113	-2.350	0.040	-0.968	0.937	-	-	-
	305*	-3.784	0.044	-4.410	0.050	-3.831	0.968	-	-	-
	306*	-7.721	0.001	-29.590	0.000	-29.386	1.000	-	-	-
	309	-2.647	0.114	-4.730	0.070	-3.419	0.935	-	-	-
	330*	-7.156	0.002	-20.340	0.000	-19.034	1.000	-	-	-
	331	-1.730	0.207	-2.220	0.080	-0.952	0.915	-	-	-
	346	-3.186	0.148	-19.360	0.100	-21.955	0.942	-	-	-
	348	-0.050	0.019	0.000	1.000	-0.231	0.680	-	-	-
352	-0.013	0.006	0.000	1.000	-0.194	0.650	-	-	-	
354*	-2.677	0.077	-5.090	0.010	-1.609	0.958	-	-	-	
360	-1.852	0.187	-2.810	0.060	-1.026	0.919	-	-	-	
364*	-2.265	0.113	-3.900	0.030	-1.994	0.949	-	-	-	
379	-0.014	0.005	0.000	1.000	-0.257	0.695	-	-	-	

Continued from previous page.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2- α	P(dN2)	p-value
Negative	380*	-2.148	0.111	-3.810	0.030	-1.708	0.947	-	-	-
	383*	-3.094	0.056	-7.630	0.000	-2.950	0.972	-	-	-
	386*	-7.568	0.001	-50.340	0.000	-33.735	1.000	-	-	-
	394	-3.426	0.125	-2.810	0.070	-0.935	0.907	-	-	-
Positive	239	<i>-1.267</i>	<i>0.821</i>	<i>-0.600</i>	<i>0.830</i>	<i>-1.173</i>	<i>0.292</i>	10000.000	0.140	0.000
	288	1.779	0.283	2.530	0.130	0.734	0.796	20.910	0.280	0.040
	390	1.425	0.402	1.840	0.220	0.361	0.693	91.620	0.110	0.020

Table 23: Integrative site-by-site selection analysis for *IRX15L*. Selection at the level of individual codons within the *IRX15L* alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR), and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic diversifying positive selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*). Conflicting results are indicated in italics.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2-dS	P(dN2)	p-value
Negative	3	-0.0101	0.004	0.000	1.000	-0.176	0.613	-	-	-
	4*	-3.628	0.033	-24.610	0.000	-20.614	0.999	-	-	-
	17*	-4.582	0.019	-8.090	0.000	-9.228	0.999	-	-	-
	19*	-9.251	0.000	-15.030	0.000	-17.494	0.997	-	-	-
	20*	-7.991	0.000	-8.950	0.000	-13.780	0.999	-	-	-
	25*	-5.000	0.004	-10.340	0.000	-12.513	0.999	-	-	-
	43	-0.056	0.013	0.000	1.000	-0.341	0.642	-	-	-
	46*	-3.151	0.039	-7.510	0.020	-5.095	0.959	-	-	-
	49	-1.500	0.259	-7.080	0.080	-5.707	0.902	-	-	-
	59*	-4.635	0.022	-3.710	0.050	-2.364	0.924	-	-	-
	60	-0.011	0.004	0.000	1.000	-0.250	0.632	-	-	-
	63	-0.011	0.005	0.000	1.000	-0.260	0.640	-	-	-
	65	-0.014	0.003	0.000	1.000	-0.350	0.637	-	-	-
	70*	-4.720	0.011	-6.880	0.010	-7.167	0.989	-	-	-
	78*	-2.942	0.047	-3.110	0.050	-2.093	0.927	-	-	-
	89*	-9.103	0.000	-52.380	0.000	-26.501	0.998	-	-	-
	91*	-1.971	0.114	-2.680	0.050	-1.563	0.918	-	-	-
	93*	-4.705	0.010	-6.520	0.000	-6.373	0.988	-	-	-
	100	-0.011	0.004	0.000	1.000	-0.175	0.612	-	-	-
	101	-0.011	0.004	0.000	1.000	-0.274	0.649	-	-	-
104*	-2.099	0.113	-5.590	0.030	-4.555	0.949	-	-	-	
121	-1.047	0.330	-5.970	0.040	-2.841	0.895	-	-	-	
125	-2.558	0.162	-8.760	0.050	-4.513	0.894	-	-	-	
128*	-3.109	0.046	-4.160	0.030	-2.740	0.936	-	-	-	

Continued from previous page.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	P-value	dN-dS	Posterior Probability	dN2-dS	P(dN2)	p-value
Negative	148	-0.011	0.004	0.000	1.000	-0.287	0.652	-	-	-
	149	-3.382	0.043	-4.410	0.060	-2.172	0.889	-	-	-
	165*	-2.124	0.175	-5.970	0.030	-3.075	0.903	-	-	-
	169*	-2.000	0.127	-5.230	0.020	-3.506	0.947	-	-	-
	172*	-8.851	0.001	-19.760	0.010	-17.773	0.988	-	-	-
	177*	-5.790	0.005	-85.040	0.000	-36.028	1.000	-	-	-
	183	-0.014	0.004	0.000	1.000	-0.330	0.640	-	-	-
	186*	-1.369	0.245	-5.700	0.050	-3.766	0.900	-	-	-
	199	-0.011	0.005	0.000	1.000	-0.217	0.613	-	-	-
	211	-0.044	0.018	0.000	1.000	-0.256	0.642	-	-	-
	215*	-4.729	0.015	-8.190	0.000	-9.331	0.999	-	-	-
	221	-0.044	0.018	0.000	1.000	-0.162	0.600	-	-	-
	230*	-2.000	0.111	-6.930	0.020	-4.890	0.961	-	-	-
	253*	-8.265	0.001	-10000.000	0.000	-41.623	1.000	-	-	-
Positive	48	1.591	0.240	-2.420	0.200	0.919	0.750	2633.750	0.050	0.010
	143*	1.961	0.227	5.790	0.070	4.959	0.963	39.980	0.160	0.010
	200*	2.562	0.080	5.470	0.040	5.205	0.971	5.470	1.000	0.040

Table 24: Integrative site-by-site selection analysis for *FSP*. Selection at the level of individual codons within the *FSP* alignment was evaluated using three different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), and Fast Unconstrained Bayesian Approximation (FUBAR). For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. For *FSP*, no potentially positively selected sites were detected by any method, including Mixed Effects Model of Evolution (MEME). Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*).

Selective Regime	Codon	SLAC		FEL		FUBAR	
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability
Negative	8	-0.040	0.002	0.000	1.000	-0.172	0.580
	27*	-2.365	0.094	-7.260	0.020	-4.847	0.962
	57	-2.000	0.164	-5.780	0.040	-1.821	0.875

Table 25: Integrative site-by-site selection analysis for *NRFP*. Selection at the level of individual codons within the *NRFP* alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR) and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic and pervasive positive or diversifying selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*).

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2-dS	P(dN2)	p-value
Negative	9	-0.006	0.009	0.000	1.000	-0.217	0.574	-	-	-
	10	-0.014	0.003	0.000	1.000	-0.179	0.574	-	-	-
	26	-0.014	0.004	0.000	1.000	-0.239	0.576	-	-	-
	31	-0.056	0.014	0.000	1.000	-0.444	0.593	-	-	-
	32*	-3.942	0.017	-7.590	0.030	-4.855	0.932	-	-	-
	42	-0.011	0.004	0.000	1.000	-0.217	0.574	-	-	-
	46*	-4.200	0.019	-83.610	0.010	-22.122	0.984	-	-	-
	48	-0.050	0.038	0.000	1.000	-0.136	0.549	-	-	-
	53*	-2.520	0.077	-12.110	0.010	-10.901	0.968	-	-	-
	55	-0.038	0.022	0.000	1.000	-0.210	0.583	-	-	-
	58*	-7.335	0.001	-79.180	0.000	-32.430	1.000	-	-	-
	59*	-4.276	0.016	-15.770	0.010	-9.678	0.962	-	-	-
	72	-0.098	0.039	0.000	1.000	-0.132	0.554	-	-	-
	75	-0.011	0.004	0.000	1.000	-0.197	0.580	-	-	-
	83	-0.044	0.018	0.000	1.000	-0.249	0.590	-	-	-
	84	-0.014	0.003	0.000	1.000	-0.411	0.595	-	-	-
	88	-0.011	0.004	0.000	1.000	-0.215	0.573	-	-	-
	89	-0.010	0.005	0.000	1.000	-0.181	0.574	-	-	-
	103	-0.044	0.018	0.000	1.000	-0.245	0.586	-	-	-
124	-1.540	0.251	-11.810	0.080	-9.175	0.909	-	-	-	
127	-0.125	0.030	0.000	1.000	-0.292	0.583	-	-	-	
134	-0.006	0.010	0.000	1.000	-0.036	0.510	-	-	-	

Continued from the previous page.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2-dS	P(dN2)	p-value
Negative	136	-0.014	0.005	0.000	1.000	-0.262	0.564	-	-	-
	137*	-3.044	0.038	-12.080	0.020	-11.759	0.974	-	-	-
	142*	-3.954	0.023	-47.120	0.020	-16.953	0.971	-	-	-
	153*	-2.924	0.049	-7.610	0.030	-4.947	0.941	-	-	-
	197	-0.011	0.004	0.000	1.000	-0.129	0.552	-	-	-
Positive	117	3.597	0.054	5.760	0.110	3.603	0.902	14.690	0.540	0.100

Table 26: Integrative site-by-site selection analysis for *WRKY*. Selection at the level of individual codons within the *WRKY* alignment was evaluated using four different methods: Single-likelihood Ancestor Counting (SLAC), Fixed Effects Likelihood (FEL), Fast Unconstrained Bayesian Approximation (FUBAR) and Mixed Effects Model of Evolution (MEME). Unlike the other methods, MEME only detects evidence of episodic and pervasive positive or diversifying selection. For each codon, the value of dN-dS is given, as it is common for dN/dS values to be undefined (i.e.: when dN=0). Codons are arranged in positively (dN>dS) and negatively (dN<dS) selected groups. Significant p-values (<0.05) and posterior probabilities (>0.9) are highlighted in bold. For MEME results, the proportion of branches that are estimated to evolve with dN2 (i.e.: where the value of dN is greater than dS) at each identified site is also indicated. Codons that were identified as positively/negatively selected by 2 or more methods are starred (*). Conflicting results are indicated in italics.

Selective Regime	Codon	SLAC		FEL		FUBAR		MEME		
		dN-dS	p-value	dN-dS	p-value	dN-dS	Posterior Probability	dN2-dS	P(dN2)	p-value
Negative	3*	-3.026	0.038	-10.970	0.010	-9.860	0.987	-	-	-
	11	-0.011	0.004	0.000	1.000	-0.137	0.549	-	-	-
	13	-0.014	0.004	0.000	1.000	-0.240	0.579	-	-	-
	15	-0.011	0.004	0.000	1.000	-0.192	0.565	-	-	-
	53*	-2.588	0.116	-6.810	0.050	-3.793	0.918	-	-	-
	70	-0.011	0.004	0.000	1.000	-0.172	0.561	-	-	-
	87	-0.011	0.004	0.000	1.000	-0.109	0.532	-	-	-
	89*	-2.067	0.113	-6.170	0.040	-3.907	0.917	-	-	-
	128	-0.018	0.003	0.000	1.000	-0.411	0.592	-	-	-
	129*	-2.000	0.111	-8.260	0.020	-5.679	0.944	-	-	-
	130*	-3.225	0.056	-11.870	0.020	-8.835	0.963	-	-	-
	140*	-2.035	0.112	-7.740	0.030	-5.202	0.935	-	-	-
	143*	-2.000	0.111	-8.570	0.020	-7.115	0.955	-	-	-
	158	-0.011	0.004	0.000	1.000	-0.161	0.554	-	-	-
	193	-0.018	0.003	0.000	1.000	-0.318	0.576	-	-	-
201	-0.044	0.021	0.000	1.000	-0.120	0.531	-	-	-	
203*	-2.000	0.111	-6.110	0.030	-3.173	0.920	-	-	-	
Positive	100*	1.917	0.234	6.880	0.050	5.057	0.970	18.090	0.500	0.040
	107*	2.500	0.132	8.530	0.050	6.810	0.985	8.530	1.000	0.050
	109	0.997	0.447	3.430	0.200	0.983	0.769	29.390	0.110	0.030
	127	0.420	0.793	1.180	0.510	<i>-0.040</i>	<i>0.565</i>	515.380	0.050	0.050
	141	2.298	0.161	7.780	0.060	5.348	0.965	7.780	1.000	0.060

Table 27: McDonald-Kreitman tests and Tajima's D tests of neutrality for all S-linked genes completed by treating diploid taxa from the *Turnera* subseries as a population sample. McDonald-Kreitman tests (MKTs) and Tajima's D tests of neutrality were performed for all genes using a reduced number of taxa in DnaSP v. 5 (Librado and Rozas 2009). With the exception of the MKT performed for *Tsstal1*, only genetic data from diploid individuals of the species, *T. subulata*, *T. scabra*, *T. krapovickasii*, and *T. concinna* were included. For MKTs, sequence information from *T. panamensis* was employed as the inter-specific sample. For each MKT, the number of fixed and polymorphic synonymous and non-synonymous sites is given. Neutrality index, α values, and Fisher's Exact Test p-values are also provided. Tajima's D statistics and corresponding p-values are shown for each alignment. No significant results were obtained for any test.

Alignment	McDonald-Kreitman				Tajima's D		
	Fixed	Polymorphic	Neutrality Index	α	Fisher's Exact Test p-value	Tajima's D	p-value
<i>APETALA2</i>							
Synonymous	3	18	0.458	0.542	0.328	-0.390	>0.1
Nonsynonymous	8	22					
<i>Tsstal1</i>							
Synonymous	20	5	0.250	0.750	0.374	-0.670	>0.1
Nonsynonymous	16	1					
<i>LEJ2</i>							
Synonymous	2	2	1.000	0.000	1.000	-0.639	>0.1
Nonsynonymous	4	4					
<i>AP2D</i>							
Synonymous	13	12	1.393	-0.393	0.751	-0.451	>0.1
Nonsynonymous	7	9					
<i>RNABP</i>							
Synonymous	10	13	1.026	-0.026	1.000	0.306	>0.1
Nonsynonymous	9	12					
<i>SCEI</i>							
Synonymous	3	4	0.000	1.000	1.000	-1.457	>0.1
Nonsynonymous	1	0					
<i>FRAI</i>							
Synonymous	6	13	3.692	-2.692	0.371	-0.080	>0.1
Nonsynonymous	1	8					

Continued from previous page.

Alignment	McDonald-Kreitman				Tajima's D		
	Fixed	Polymorphic	Neutrality Index	α	Fisher's Exact Test p-value	Tajima's D	p-value
<i>LRRK</i>							
Synonymous	36	33	1.091	-0.091	0.854	0.002	>0.1
Nonsynonymous	25	25					
<i>IRX15L</i>							
Synonymous	14	19	0.884	0.116	1.000	0.353	>0.1
Nonsynonymous	5	6					
<i>FSP</i>							
Synonymous	3	3	0.500	0.500	1.000	-0.401	>0.1
Nonsynonymous	2	1					
<i>NRFP</i>							
Synonymous	12	20	1.050	-0.050	1.000	0.439	>0.1
Nonsynonymous	8	14					
<i>WRKY</i>							
Synonymous	14	8	0.636	0.364	0.548	-1.069	>0.1
Nonsynonymous	22	8					

4.0 DISCUSSION

This study represents the first investigation of the molecular signatures of natural selection on *S*-linked genes in any heterostylous species. Importantly, the particular molecular “footprints” of selection that are expected to be exhibited by genes involved in determining heterostyly will largely depend on the particular genetic mechanisms that underlie the expression of reciprocal herkogamy and SI in heterostylous species. While evidence from homomorphic systems, in particular, indicates that alleles for SI genes with pollen- and pistil-specific functions are maintained by balancing selection, recent evidence from distylous *Fagopyrum*, *Linum*, and *Primula* may suggest an altogether different scenario. In species exhibiting homomorphic SI – and particularly those possessing S-RNase-based or Brassicaceae-type systems – the action of balancing selection is strongly suggested by the presence of identifiable trans-species polymorphisms (TSPs) and increased nucleotide diversity at *S*-linked loci (Ioerger et al. 1990; Dwyer et al. 1991; Nou et al. 1993; Richman et al. 1996; Boyes et al. 1997; Ishimizu et al. 1998; Richman and Kohn 1999 & 2000; Schierup et al. 2001; Kamu and Charlesworth 2005; Castric and Vekemans 2007; Sutherland et al. 2008; Nowak 2011). More generally, balancing selection is thought to be a common feature of many self/non-self recognition systems, like SI, that are possessed by a diverse array of organisms, including the mating-type determination system of fungi (May et al. 1999), the complimentary sex-determination (*csd*) locus of honey bees (Hassleman and Beye 2004; Cho et al. 2006), and the vertebrate MHC (Takahata and Nei 1990; Klein et al. 1998; Richman 2000; Charlesworth 2006).

More recently, morph-specific genes – and particularly short-specific genes – have been identified in a few distylous species, which likely possess SI mechanisms with differing evolutionary histories (Li et al. 2008; Yasui et al. 2012; Ushijima et al. 2012 & 2015; Nowak et al. 2015; Chafe et al. 2015; Shore and Chafe, Unpublished data). It has also been proposed that the mere presence or absence of these genes may be responsible for determining floral-morph identity in these taxa (Yasui et al. 2012). If this is the case, then the evolutionary dynamics exhibited by genes responsible for determining distyly may be very different from what is observed in homomorphic SI systems. Instead, under these circumstances, patterns of sequence evolution should be analogous to that which is observed in male-specific genes typical of X/Y sex determination systems (King et al. 2007; Uyenoyama 2005). For instance, the primary Y-

linked male-determining gene in mammals and marsupials, *SRY*, has been shown to experience pervasive purifying selection in order to maintain its function (Graves 1998; Wang et al. 2002; King et al. 2007; Nowacka-Woszuik and Switonski 2009; Murtagh et al. 2012). If causative morph-specific genes in distylous systems behave like Y-linked genes, we might expect them to show molecular signatures of purifying selection, as well.

The genes involved in determining heterostyly are not known with certainty for any species. However, recent efforts in the genus, *Turnera*, have identified a series of candidate “distyly genes” known to be *S*-linked in *T. subulata*, including those that are represented on both the *S*- and *s*-haplotype, as well as others that are *S*-haplotype specific (Labonne and Shore 2011; Shore and Chafe, Unpublished data). Given that the genetic basis of distyly in *Turnera* is unknown, several approaches were adopted in order to explore the nature of selection on *S*-linked genes with the aim of assessing their potential involvement in determining distyly in this genus. Here, I will summarize and interpret the main findings of these analyses and communicate any conclusions that may be drawn from them, including their implications for future study.

4.1 S-LOCUS GENES WITH ELEVATED NUCEOLTIDE DIVERSITY AND SEQUENCE POLYMORPHISM

The distribution of nucleotide diversity (π) and sequence polymorphism (Watterson’s θ) estimates across the *S*-locus indicate that levels of diversity are somewhat elevated in regions directly adjacent to the short-specific gene, *Tsstal*, relative to other regions of the genome. Indeed, diversity measures appear to increase fairly consistently as one moves out from the *Tsstal* gene, before values drop off considerably at the *SCEI* gene. Interestingly, based on recent BAC sequencing results, *SCEI* is expected to be located farther away from *RNABP* on the *S*-haplotype, perhaps in the vicinity of *WRKY* (Figure 6), as a result of one or more inversions (J.S. Shore, Personal communication). In fact, the region represented by *IRX15L* through to *WRKY* is anticipated to be inverted in the *S*-haplotype relative to the recessive *s*-haplotype, with the specific positions of *APETALA2*, *LRRK*, and *FRAI* currently being uncertain (J.S. Shore, Personal communication). If the gene order presented in Figure 6 is subsequently reconfigured to reflect the newly anticipated gene order of the *S*-haplotype in light of these suspected rearrangements, diversity measures are then shown to gradually increase from *Tsstal* until they

peak in *RNABP*, and then decrease gradually as they move outward toward *SCEI*. If *S*- and *s*-specific alleles are maintained by balancing selection, this pattern in the distribution of diversity measures might be expected for genome regions containing causative distyly genes. If this were the case, the gene or genes that are subject to selection are likely to be located in the region exhibiting the highest diversity, which, in this case, is in the vicinity of *RNABP* and *AP2D*. Indeed, two newly discovered *S*-haplotype specific genes have recently been discovered in this region (J.S. Shore, Personal communication).

Increased nucleotide diversity at SI-determining loci that are maintained by balancing selection has been demonstrated in other SI systems. Estimates of nucleotide diversity in homomorphic systems, for example, are generally very high, with populations obtaining π estimates of between 0.13 and 0.46 (Castric and Vekemans 2004). Clearly, none of the π estimates obtained here approached these values, as the greatest value of π observed represented only a few percent (0.037 for *RNABP*). While these generally low values may suggest that none of the candidate genes investigated here are likely the direct targets of balancing selection, it also may not be appropriate to compare the diversity measures obtained for *S*-linked genes in a distylous system to those garnered for SI-determining genes in homomorphic systems. Plants that possess homomorphic SI systems are generally characterized by the presence of a diversity of morphologically indistinct “mating types”, with populations containing between a dozen and a hundred or more different SI alleles, each allele representing a different pollen-stigma specificity (Nou et al. 1993; Castric and Vekemans 2007). In distylous systems, on the other hand, only two allelic lineages are thought to occur for genes involved in determining distyly: one that is *S*-specific and another that is *s*-specific (that is, if their maintenance by balancing selection is assumed). Under these circumstances, diversity estimates might be expected to be comparatively low. Also, estimates of average nucleotide diversity per site (π) are not found to be high in all self/non-self recognition systems. Indeed, at the MHC locus of jawed vertebrates, π values have been measured to be only a few percent ($\pi=0.021-0.093$), which is comparable to what was observed here (Raymond et al. 2005). Similar values were also obtained for the honey bee complimentary sex-determining locus (*csd*) (Cho et al. 2006).

While the absolute levels of nucleotide diversity obtained here appear to be lower than in other SI systems, it is, perhaps, more useful to compare *S*-linked genes to unlinked genes in the same system in order to determine the significance of sequence diversity measures. In the

context of the current study, estimates of the mean level of nucleotide diversity (π) and sequence polymorphism (θ) did not differ significantly between *S*-linked and control genes in *Turnera*, as determined by Mann-Whitney-U tests. However, not all candidate genes are expected to be involved in distyly. Indeed, it is possible that none of the genes investigated here serve any such function. If the latter is the case, we might still expect to see increased levels of sequence diversity for the genes most closely linked to those that do have a function in determining distyly, as balancing selection will cause elevated diversity at neutral sites closely linked to a gene under balancing selection (Charlesworth et al. 1997; Charlesworth 2006). There is, indeed, elevated diversity at two genes (*RNABP* and *AP2D*) that are immediately adjacent to perhaps the best candidate gene, *Tsstal* (as will be discussed in more detail below).

Interestingly, *S*-linked genes were found to exhibit a significantly greater number indels relative to controls. Indels were also significantly larger at the *S*-locus than they were elsewhere, on average. In general, the presence of indels in the coding region of a gene suggests reduced selective constraint (though they are apparently common in “essential” protein-coding genes (Chan et al. 2007; Ajawatangawong and Badauf 2013)). Indeed, site-by-site selection analyses tended to identify potentially positively selected sites near regions characterized by the presence of indels, as well. In addition, those genes in which the greatest number of indels had been identified also tended to have the largest nucleotide diversity (π) and sequence polymorphism (θ) estimates. Recently, indels have been implicated in increasing mutation rates at closely linked sites (Tian et al. 2008). A pattern that was observed across and within several disparately related species, it was proposed that the presence of indels may promote mutation in surrounding sequences, with indel-associated rates of mutation decreasing in inverse proportion to sequence distance from indels (Tian et al. 2008; Hollister et al. 2010). Importantly, this phenomenon has also been demonstrated in angiosperms, and the effects of indel-associated mutation appear to be more severe in outcrossing species, such as those studied here (Hollister et al. 2010). In light of this, it may be possible that the increased values of π and θ that were observed in *AP2D* and *RNABP*, in particular, may be related to the presence of many indels in these genes. However, this does not explain why *S*-linked genes might contain more and larger indels than are characteristic elsewhere in the genome, nor does it explain how they might arise or be allowed to persist in the coding regions of a genes that appear to be otherwise well-conserved across a number of species of *Turnera*.

Alternatively, the increased number and size of indels observed in *S*-linked genes may be indicative of the occurrence of increased neutral sequence diversity at these loci (in this instance, small deletions or insertions) that results from their linkage to another site that is the direct subject of balancing selection (Charlesworth 2006). Finally, I cannot discount the possibility that there is something unusual about the genomic region in which the candidate genes reside, such that it is subject to increased insertion/deletion events. Indeed, there is a great preponderance of transposable elements in the *S*-locus region that also involve insertion events (J.S. Shore, Personal communication).

Notably, one of the control genes, *RNABP34*, showed somewhat higher nucleotide diversity and sequence polymorphism estimates, as well. In fact, it was the only control gene that was shown to contain an indel, suggesting that RNA binding protein coding genes may simply exhibit high sequence diversity in general, regardless of their particular function. Indeed, this may be the case, as RNA binding proteins represent a remarkably diverse protein family with a variety of RNA targets (Glisovic et al. 2008). It is therefore conceivable that elevated diversity at the *S*-linked *RNABP* locus may simply be confirmation of this gene family characteristic.

In general, I caution that much of the above is built upon the base assumption that mating type in distylous *Turnera* is determined by the Mendelian segregation of two alleles at the causative distyly locus/loci (Barrett and Shore 2008). As will be discussed in further detail below, it is possible that distyly is determined by *S*-specific genes, in a system that is more akin to the genetic basis of sex determination in mammals and marsupials than it is to any SI system in plants. Indeed, this possibility is exemplified here by the occurrence of the *Tsstal* gene, which remains a strong distyly-gene candidate. Due to its presence in only one morph, it will necessarily exhibit evolutionary dynamics that will not yield increased genetic diversity at the *Tsstal* gene, itself. This possibility will be further discussed below.

4.2 GENE GENEALOGIES

If the genetic basis of distyly involves genes, each possessing two alleles (*S* and *s*), as is suggested by its apparently simple Mendelian inheritance pattern, we can make predictions about the relationship these alleles should show across species. As the causative alleles are expected to be maintained by balancing selection, they will have been preserved in populations for extended

periods of time, persisting even after speciation events (Klein et al. 1998; Richman 2000; Charlesworth 2006, Barrett and Shore 2008). As a result, we might expect short-specific alleles at these loci to form a single clade, while long-specific alleles fall into a second clade (Klein et al. 1998; Richman 2000; Charlesworth 2006). In such cases, gene genealogies will therefore not reflect species phylogenies, and will instead take on a pattern characteristic of trans-specific evolution.

Gene genealogies were constructed for all *S*-linked genes investigated here as a means to explore the incidence of trans-species polymorphism (TSP) and to provide evidence for its underlying cause (i.e.: balancing selection). Interestingly, only *APETALA2* showed evidence of TSP. For this gene, the TSP pattern occurred within the *Turnera* subseries and was driven by linkage disequilibrium at three nucleotide positions. Because this pattern breaks down in *T. joelii* and more distantly related species, it is not likely that *APETALA2* is directly involved in the determination of distyly. That is, the specific polymorphisms (which do, in fact, result in amino acid substitutions at 2 of the 3 sites) cannot be responsible for determining morph-identity because short-styled plants outside of subseries *Turnera* possess the *s*-specific nucleotides at these sites. Therefore, the TSPs observed in *APETALA2* are likely neutral in nature (Klein et al. 1998). However, their occurrence does suggest that there is reduced recombination in this gene region and that *APETALA2* is perhaps closely linked to the gene(s) that do determine distyly. Indeed, while the position of *APETALA2* is known for the *s*-haplotype, its position on the dominant *S*-haplotype remains uncertain due to one or more gene rearrangements within this haplotype, at least in *T. subulata* (P. Chafe and J.S. Shore, Personal communication). It would be of value to more fully explore the incidence to TSP in *APETALA2* using the more costly and time consuming approach of sequencing cloned PCR products so that haplotype-specific information can be obtained. It would be interesting to sequence the entire *APETALA2* gene, as well, in order to see if other such informative sequence positions occur.

4.3 SPECIES PHYLOGENIES

The phylogenies computed here were not explicitly constructed for the purposes of evaluating the phylogenetic relationships between the various species of *Turnera*. Rather, they were constructed so that they might be used to inform subsequent dN/dS based selection

analyses. Nevertheless, to that end, it is important to assess their accuracy relative to what is currently known about the evolutionary relationships between the representative species in order to ensure that selection analyses were not adversely affected by errors in phylogeny. Generally, all of the series of *Turnera* represented here appear to be placed in their expected arrangements in the phylogenies produced and reflect species relationships presented in the published literature (Truyens et al. 2005; Arbo 2015). Though relationships between members of the *Turnera* subseries were not well resolved, the subseries, itself, does form a well supported monophyletic clade in all of the trees that were employed for the purposes of subsequent selection analyses. Importantly, the species included that belong to the *Turnera* subseries are very closely related and are capable of interbreeding to varying extents (Shore and Barrett 1985a; Arbo and Fernandez 1987; Fernandez and Arbo 1990; Fernandez and Arbo 1993). As a result of potential gene exchange and recent common ancestry, their evolutionary relationships may be difficult to discern using genetic data that is generally lacking in sequence polymorphism.

Over the last decade, several attempts have been made to deduce the phylogenetic relationships among various *Turnera* species and their close relatives using molecular data (Truyens et al. 2005; Thulin et al. 2012; Lopez 2013; Arbo et al. 2015). While reasonable series-level resolution has been obtained (particularly, with the addition of morphological data), subseries resolution has proven more difficult to achieve – particularly with reference to the placement of *T. scabra* and *T. subulata* (Lopez 2013; Arbo et al. 2015). According to early taxonomic classifications, these two species were initially considered to be taxonomic varieties of a single species (*T. ulmifolia*) (Urban 1883). Recent molecular investigations have been unable to resolve the phylogenetic relationships between these species and the possibility that hybridization has occurred between them has been suggested in order to explain the lack of monophyly observed (Lopez 2013; Arbo et al. 2015).

In terms of the phylogenies obtained here, the placement of DEN 20S (tetraploid *T. grandidentata*) within the *Turnera* subseries is also somewhat interesting. That is, in many of the trees presented, it is usually positioned as sister to all other members of the *Turnera* subseries. Interestingly, *T. grandidentata* is also a segmental allotetraploid, and it is suspected that *T. concinna* and perhaps *T. subulata* (though that is less certain) may have acted as its genome donors (Fernandez and Arbo 1993; Lopez et al. 2013). This reticulate origin may explain the placement of DEN 20S as the most “basal” taxon in the *Turnera* subseries lineage.

While the low resolution obtained within the *Turnera* subseries might perhaps be remedied by the addition of increased sequence data, this finding provides at least a further rationale for treating these taxonomic species as if they are all a single species for some of the analyses that were performed here. Fortunately, in practice, dN/dS methods are also quite resilient to potential “errors” in phylogeny, so long as the tree in question is not completely inaccurate (Pond et al. 2009). Indeed, based on their general similarity to existing phylogenies in the published literature, it is thought that the trees employed here are more than sufficient for the current purposes (Truyens et al. 2005; Arbo 2015).

4.4 PERVASIVE PURIFYING SELECTION ACROSS S-LINKED GENES

Though the distribution of nucleotide diversity (π) and Watterson’s θ estimates, in addition to the increased incidence and size of indels in *S*-linked genes, may indicate the possible influence of balancing selection at a few loci, dN/dS-based tests of selection suggest the occurrence of pervasive purifying selection across the *S*-locus in *Turnera*. Indeed, this appears to be true at all levels according to alignment-wide (global), lineage-specific (local), and codon-specific analyses. While a few codon sites have been identified as potentially affected by diversifying selection, they were largely located near indels, and thus the significance of these particular sites in terms of selective processes involved in evolution of distyly in *Turnera* is doubtful. However, the observation of wide-spread purifying selection across known *S*-linked genes does not negate the possibility that balancing selection may maintain diversity at other linked sites. Indeed, purifying selection is expected to act on genes that code for functional proteins, as most mutations will likely erode that function (Pond et al. 2009; Pybus and Shapiro 2009). As a result, purifying selection is more commonly detected using dN/dS methods (Pond et al. 2009). Neutrally evolving sites, on the other hand, are observed to experience the effects of balancing selection more keenly and, similarly, the extent of silent polymorphism in selectively constrained genes might be affected by balancing selection that acts on nearby sites, as well (Kreitman and Akashi 1995; Charlesworth et al. 1997; Charlesworth 2006). This effect will not likely be detected using dN/dS methods, which prioritize nonsynonymous mutations in the detection of diversifying selection (Pond et al. 2009).

If the *S*-locus is indeed largely affected by purifying selection, it would provide support to the notion that morph-specific genes are important in the determination of distyly in this genus. In fact, when concatenated alignments of all *S*-linked and all control genes were compared using likelihood ratio tests (LRTs), they were found to differ significantly in terms of the distribution of evolutionary rate parameters. While these tests also indicate that parameters related to positive selection appear to be largely the same between the two data sets, global dN/dS values do suggest that the control genes are perhaps more constrained by purifying selection than are *S*-linked genes, on average, as control genes obtained somewhat lower global dN/dS values. This finding, perhaps, does not support the view that genes at the *S*-locus are inordinately influenced by purifying selection, and may, in fact, suggest that the opposite is true.

The estimation of dN/dS can be adversely affected by uneven taxon sampling, particularly when closely related individuals are included in a sample (Kryazhimskiy and Plotkin 2008). For several species, including *T. subulata*, *T. scabra*, *T. krapovickasii*, *T. grandidentata*, and *T. joelii*, more than one individual was included in the analysis. This was done, initially, so as to include genetic information from each of the floral morphs, as well as from individuals from different populations in the data set (though, no more than two individuals were represented from any one population). However, dN/dS methods assume that individuals are more distantly related, and therefore polymorphism within any one population is not considered (Kryazhimskiy and Plotkin 2008). Rather, each sequence difference between individuals is treated as a fixed difference between species. Under these circumstances, dN/dS may be underestimated, as synonymous polymorphisms within any one population will be treated as synonymous fixations and these will tend to drive the value of dS up (Kryazhimskiy and Plotkin 2008). As a result, the values of dN/dS obtained here may be somewhat conservative. However, when closely related individuals were subjected to more statistically powerful population-genetic analyses (Pybus and Shapiro 2009), no signatures of positive or diversifying selection were detected, thus suggesting that perhaps the dN/dS estimates obtained were not inordinately conservative after all.

4.5 *AP2D*: AN INTERESTING CANDIDATE GENE

From the results obtained here, *AP2D* emerged as an interesting gene candidate, as dN/dS-based tests of selection indicate that evolutionary rates for this gene vary in some way across

lineages (i.e.: over evolutionary time). Moreover, when selection on *AP2D* was compared to all other *S*-linked genes, *AP2D* was found to differ significantly, though apparently not in terms of the strength or prevalence of positive selection at the codon-level across the alignment. While evolutionary rates appear to have varied across lineages for this gene, no particular lineages were identified as being differentially affected by diversifying selection, in particular, suggesting that the differences detected may have resulted from changes in the strength of purifying selection or the proportion of sites affected by it. However, it is not clear why evolutionary rates for *AP2D* may be expected to vary over time, and its relevance to determining distyly in *Turnera*, if it has any, is unknown.

In addition, *AP2D* is also located in the region of the *S*-locus that appears to be characterized by somewhat elevated nucleotide diversity (π) and sequence polymorphism (θ). It also contains more and larger indels than the average *S*-linked gene and, further, the sequence for this gene that was obtained for one species, *T. chamaedrifolia* (a tetraploid), appeared to be truncated (though this is possibly because the sequence obtained was from a pseudogene that arose as a result of a duplicate locus due to allopolyploidy). All of these things, taken together, may indicate the influence of possible balancing selection on this gene or a gene immediately adjacent to it, making *AP2D* an interesting candidate. In general, however, *AP2D* exhibits alignment-wide signatures of purifying selection, like all other *S*-linked genes investigated here (with the exception of *LEJ2*, perhaps, which appeared to be evolving more or less neutrally). Further, when codon-level analyses were applied, only one potentially positively selected site was detected, indicating that the high diversity estimates obtained for this gene relative to other *S*-linked genes may not be the result of diversifying selection. Diversifying selection is notoriously difficult to detect using dN/dS methods, however, unless its molecular signal is very strong (Yang and Bielawski 2000; Wong et al. 2004; Pybus and Shapiro 2009). Moreover, because dN/dS-based tests of selection tend to be conservative, they will often fail to detect instances of balancing selection, in particular, unless a considerable excess of non-synonymous substitutions are produced as a result (Yang and Bielawski 2000). As indicated above, if balancing selection is an important force in the maintenance of distyly-determining genes, only two allelic lineages are expected for those genes that are involved. Due to this low number of expected allelic types, it may be the case that the signature of balancing selection on these genes is simply not strong enough to detect using dN/dS-based methods. As a result, the estimation of

nucleotide diversity for this gene may more readily detect the effects of balancing selection on adjacent linked genes.

Using comparative dN/dS methods, the molecular signatures of positive and balancing selection also cannot be distinguished, as both forces are expected to result in higher proportions of nonsynonymous relative to synonymous substitutions (Pond et al. 2009). Population-level tests of neutrality, like Tajima's D, on the other hand, can differentiate between these two processes and are considered to be more powerful tests of selection than dN/dS-based approaches (Pybus and Shapiro 2009). However, no significant results were obtained using these tests for any gene examined here, including *AP2D*, suggesting that balancing selection has perhaps not been a major factor in the direct evolution of these genes in *Turnera*. The occurrence of balancing selection at still other unknown *S*-linked loci remains possible, however.

The particular function of *AP2D* is also unknown in *Turnera*. Based on sequence homology to known transcription factors in *Arabidopsis*, *AP2D* is presumed to code for a member of the DREB subfamily A-5 of the ethylene response factor (ERF)/AP2 transcription factor family referred to as *DEAR2*. Proteins in this family have been implicated in plant stress responses to dehydration as well as pathogen defense (Tsutsui et al. 2009; Zhou et al 2010). While this designation does not illustrate a clear role for *AP2D* in the determination of distyly, it has been suggested that SI and innate immunity responses in plants may share some common mechanisms in other systems (Sanabria et al. 2008). Indeed, more generally, innate immunity and SI – both of which rely on self/non-self recognition mechanisms – are thought to be moulded by similar selective pressures (Nasrallah 2005; Sanabria et al. 2008). However, upon pathogen infection, *DEAR2*, in particular, was not found to be transcriptionally induced in *Arabidopsis*, suggesting that it may not be involved in responses to biotic stress in this genus (Tsutsui et al. 2009). Interestingly, Athanasiou et al. 2003 and Khosravi et al. 2004 previously discovered two other pathogen-response genes that were found to be specific to the styles of the short-styled morph of several *Turnera* species. Neither of these genes is linked to the *S*-locus, however, and their role in distyly is currently unknown.

4.6 *Tsstal*

In *Turnera*, *Tsstal* represents, perhaps, the most promising distyly gene candidate for several reasons: 1) It exhibits morph- and tissue-specific expression; 2) Its expression has been identified as being short-stamen-specific in all species that have been investigated to date; 3) It is also expressed in homostyles with short-like pollen compatibility behaviour; 4) According to BAC sequencing and deletion mutant screens, *Tsstal* is located in the *S*-locus region; and 5) *Tsstal* also has sequence homology to a known self-incompatibility protein in *Papaver*. Here, *Tsstal* was also shown to be highly conserved across populations and species of *Turnera* and its sister genus, *Piriqueta*. dN/dS-based tests of selection further indicated that *Tsstal* appeared to be evolving significantly non-neutrally, in the direction of purifying selection and signatures of purifying selection were also identified for this gene at the lineage- and codon-specific levels. When selection on *Tsstal* was compared to the remaining *S*-linked genes, no significant differences were observed, indicating that, while *Tsstal* may be negatively selected, it was perhaps not inordinately affected by purifying selection relative to other *S*-linked genes. This, however, does not necessarily indicate that *Tsstal* is not involved in distyly, only that it is conserved in the same manner as other functional genes.

In *Papaver*, PrsS's act as the style-specific determinants of SI and are small secreted proteins that interact with pollen-specific transmembrane proteins called PrpS's (Foote et al. 1994; Bosch and Franklin-Tong 2008; Wheeler et al. 2009; Wu et al. 2011; Eaves et al. 2014). In *Turnera*, however, the apparent *PrsS* homologue, *Tsstal*, has stamen-specific expression in the short-styled morph. While its homology to a known SI protein is certainly intriguing, its function in the stamen tissues of short-styled *Turnera* remains unknown. If the general function of *PrsS* is maintained in *Turnera*, and it is indeed a secreted signalling protein involved in SI response cascades, the other proteins that it may interact with remain unidentified. Alternatively, it is also possible that *Tsstal* has no function in distyly in *Turnera* at all, as a large number of *PrsS* homologues have also been identified in species of *Arabidopsis* that lack SI altogether and have relatives with SI mechanisms that are not homologous to that which is presented in *Papaver* (Ride et al. 1999). However, this scenario is unlikely as *Tsstal* appears to serve a short-stamen specific function in *Turnera*. Further, its sequence does not seem to have been degenerated in self-compatible homostylous species, which are homozygous for *Tsstal* due to the fact that they

can inbreed. As a result, *Tsstal* is free to recombine in these species, and – should *Tsstal* not serve a function – it might have been expected to have mutated to the point of being non-functional in these individuals (J.S Shore, Personal communication).

The occurrence of morph-specific genes for SI in addition to some of the peculiar characteristics exhibited by the *S*-haplotype in *Turnera* might also fit with the predictions of Uyenoyama (2004) regarding the evolution of gene regions that are tightly linked to mating type. In plants and other systems, large swaths of genes with unrelated functions tend to co-segregate with mating type (Wang et al. 2003 & 2004; Li et al 2004; Uyenoyama 2004). This observation, she proposes, may be a symptom of the limited recombination that is thought to characterize gene regions that have key functions in the determination mating type, with the function of reduced recombination being so that type-specific alleles at this locus may remain distinct over time (Uyenoyama 2004). While reduced recombination can ensure heterozygosity, it also reduces ones ability to purge deleterious mutations, and particularly the accumulation of transposable elements, that occur in closely linked regions via a process referred to as Muller's Ratchet (Muller 1964; Uyenoyama 2004). The reduced effective population size of mating-type determining alleles relative to recombining alleles at other loci is also expected to enhance the effect of these “degenerative processes” due to genetic drift (Uyenoyama 2004). In the context of *Turnera*, this might help to explain the large size of the *S*-haplotype (>900 kb) relative to the recessive *s* (192 kb), the presence of inversions and transposable elements that are apparently specific to the dominant *S*-haplotype (Shore and Chafe, Unpublished data), and also the larger number and size of indels observed in *S*-linked genes. However, this scenario does not explain the origins of *Tsstal*, which may itself have gotten “stuck” in a non-recombining region of the *S*-locus by chance, perhaps being transmitted there via a transposable element. However, the fact that *Tsstal* remains highly conserved across a number of species and populations of *Turnera* and *Piriqueta* suggests that it does serve some function in determining the stamen- or pollen-related aspects of SI in this group.

Interestingly, when a local population of tetraploid *T. scabra* from the Dominican Republic was investigated, absolutely no sequence polymorphism was identified in the coding region *Tsstal*. When levels of isozyme variation were previously surveyed in tetraploid populations of *T. scabra* from the same region, low levels of diversity were also detected in comparison with mainland populations (Shore 1991). As a result, it was suggested that lack of genetic diversity

within these populations could be due to a founder effect resulting from island colonization (Shore 1991). Indeed, this explanation may explain the lack of sequence diversity observed in the data obtained here.

4.7 CONCLUSIONS

If the expression of distyly in *Turnera* is determined by genes with *S*- and *s*-specific alleles that are maintained by balancing selection, patterns of trans-specific evolution are expected to be observed in gene genealogies constructed for causative loci (Klein et al. 1998; Richman 2000; Charlesworth 2006). Here, evidence of trans-specific evolution was obtained for one of the genes that were investigated (*APETALA2*). In general, gene genealogies tended to reflect species phylogenies, with observed deviations likely resulting from close evolutionary relationships between species in combination with insufficient genetic data necessary to obtain optimal phylogenetic resolution. In the case of *APETALA2*, in particular, it would be interesting to sequence the entire gene in order to see if more sequence positions showing evidence of TSP can be identified, as the occurrence of TSPs across this gene suggests that *APETALA2* is closely linked to gene(s) that determine distyly.

The detection of elevated nucleotide diversity and sequence polymorphism in two genes (*AP2D* and *RNABP*) closely linked to *Tssta1* may indicate that they are influenced by balancing selection, directly (Charlesworth 2006). This, in turn, may implicate them as possibly having some function in the expression of distyly in *Turnera*. The general signature of purifying selection that was identified across the *AP2D* and *RNABP* gene alignments does suggest that they are perhaps not directly involved in determining distyly, however. That being said, dN/dS-based results indicate that selective pressures on *AP2D*, in particular, are in some way different when compared to other *S*-linked genes. Its sequence similarity to *DEAR2* in *Arabidopsis* is also interesting, as genes in this family have been implicated in responses to pathogen infection (Tsutsui et al. 2009; Zhou et al 2010). Indeed, proteins related to innate immune responses in plants are thought to perhaps play a role in SI response in other systems (Nasrallah 2005; Sanabria et al. 2008). Other pathogen response genes have also been shown to have morph-specific expression in *Turnera* (Athanasiou et al. 2003 and Khosravi et al. 2004). However, their precise functions remain unknown.

Alternatively, it might be possible that the increased sequence diversity observed at these loci are symptoms of the effects of balancing selection on other closely linked genes that are responsible for the expression of distyly in *Turnera*. Indeed, balancing selection has been observed to increase neutral sequence diversity at linked sites (Kreitman and Akashi 1995; Charlesworth et al. 1997; Charlesworth 2006). Though dN/dS results identified extensive purifying selection across *S*-linked loci, these methods may fail to identify balancing selection unless its molecular signal is very strong (Yang and Bielawski 2000; Wong et al. 2004; Pybus and Shapiro 2009). In addition, as dN/dS methods emphasize increased nonsynonymous mutations as indicators of diversifying selection (Pond et al. 2009), the effect of balancing selection on synonymous mutation rates in genes that are largely maintained by purifying selection but are, perhaps, closely linked to loci that are maintained by balancing selection, will not likely be detected.

The *S*-haplotype-specific gene, *Tsstal* is possibly the most promising “distyly gene” candidate investigated here. This gene has been previously shown to have short-stamen specific expression and exhibits significant sequence homology to a known SI protein in *Papaver* (Shore and Chafe, Unpublished data). Further, *Tsstal* has also been demonstrated to be located in the *S*-locus region according to BAC sequencing and deletion mutant screens (Shore and Chafe, Unpublished data). If *Tsstal* is involved in determining distyly in *Turnera*, it is expected to exhibit evolutionary dynamics typical of Y-linked sex-determining genes in mammals and marsupials. Specifically, it would be expected to show signals of purifying selection in order to maintain function (Graves 1998; Wang et al. 2002; King et al. 2007; Nowacka-Wozzuk and Switonski 2009; Murtagh et al. 2012). Indeed, here, this was demonstrated to be the case for *Tsstal* using dN/dS methods. These results are further supported by the fact that *Tsstal* appears to be well conserved across many species and populations of *Turnera* and *Piriqueta*.

The involvement of morph-specific genes – and particularly short-specific ones like *Tsstal* – in the expression of distyly in *Turnera* might also help to explain some of the peculiar characteristics that are exhibited by the dominant *S*-haplotype, more generally. For instance, if the gene region containing causative *S*-haplotype specific genes is maintained by purifying selection and experiences reduced recombination, it might be expected to accumulate deleterious mutations in closely linked sites (Uyenoyama 2004). If this is the case in *Turnera*, it might account for the increased size of the dominant *S*-haplotype relative to the recessive *s*-haplotype

as well as the greater proportion of transposable elements, inversions, and indels that have been identified there (Uyenoyama 2004; Shore and Chafe, Unpublished data).

It is important to note, however, that the genes investigated here may not represent the “distyly genes” at all, and still other as-of-yet undiscovered causative genes might exist. To this end, two other short-specific genes have recently be uncovered in the *S*-locus region (Shore and Chafe, Unpublished data), providing support for the notion that distyly may perhaps be influenced not by one, but by an entire suite of morph-specific genes.

It is also possible that the structure of the *S*-locus differs in different species of *Turnera*. However, larger scale BAC and RNA sequencing efforts have mainly been focused on one species (*T. subulata*). It would be interesting to know if the gene-content of the *S*-locus is the same across species of *Turnera* and if gene expression profiles, particularly among differentially expressed genes, are analogous. Future selection analyses that include a wider variety of more evenly sampled taxa might serve to bolster dN/dS-based results, as well.

Several approaches were adopted to explore the signatures of selection on *S*-linked genes in distylous *Turnera*. While no clear signature of balancing selection was detected, evidence of increased nucleotide diversity and sequence polymorphism for two genes (*AP2D* and *RNABP*) that are located immediately adjacent to a strong candidate gene, *Tsstal1*, was observed. Limited TSPs were also identified in another *S*-linked gene (*APETALA2*). Purifying selection on key gene candidates, such as *Tsstal1*, was found to occur as well. While these results, perhaps, do not provide a full elucidation of the genetic underpinnings of distyly in *Turnera*, some very promising gene candidates – such as *AP2D* and *Tsstal1* – did emerge. Investigations of the evolutionary dynamics exhibited by *S*-linked genes have the potential to yield clues regarding the genetic structure that underlies the expression of distyly. It is hoped that understanding these processes and mechanisms, in turn, will help to evaluate existing models regarding how distyly is presumed to have evolved (Charlesworth and Charlesworth 1979b; Lloyd and Webb 1992 a & b).

LITERATURE CITED

- Adachi J, Waddell P., Martin W., and Hasegawa M. 2000. Plastid genome phylogeny and model of amino acid substitution for proteins encoded by chloroplast DNA. *Journal of Molecular Evolution*. 50: 348-358.
- Aguiar B., Viera J., Cunha A.E., Fonseca N.A., Iezzoni A., van Nocker S., Vieira C.P. 2015. Convergent evolution at the gametophytic self-incompatibility system in *Malus* and *Prunus*. *PLoS ONE*. 10(5):1-24.
- Ajawatangawong P. and Badauf S.L. 2013. Evolution of protein indels in plants, animals, and fungi. *BMC Evolutionary Biology*.13(140): 1-15.
- Altschul S. F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., and Lipman D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research*. 25: 3389-3402.
- Anderson M.A., McFadden G.I., Bernatzky R., Atkinson A., Orpin T., Dedman H., Tregear G., Fernley R., and Clarke A.E. 1989. Sequence variability of three alleles of the self incompatibility gene of *Nicotina glauca*. *The Plant Cell*. 1: 483-491.
- Anderson M.A., Cornish E.C., Mau S.-L., Williams E.G., Hoggart R., Atkinson A., Bonig I., Grego B., Simpson R., Roche P.J., Haley J.D., Penschow J.D., Niall H.D., Tregear G.W., Coghlan J.P., Crawford R.J., and Clark A.E. 1986. Cloning of cDNA for stylar glycoprotein associated with expression of self-incompatibility in *Nicotina glauca*. *Nature*. 32: 38-44.
- Angiosperm Phylogeny Group (APG). 2009. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society*. 161(2): 105–121.
- Arbo M.M. 2005. Estudios sistematicos en *Turnera* (Turneraceae). III Series Anomaliae Y Turnera. *Bonplandia*. 14(3-4): 115-318.
- Arbo M. M. and Fernandez Y.A. 1987. Cruzamientos intra e interespecificos en *Turnera*, serie Canaligerae. *Bonplandia*. 6(1): 23-38.
- Arbo M.M., Gonzalez A.M., and Sede S.M. 2015. Phylogenetic relationships within the Turneraceae based on morphological characters with emphasis on seed micromorphology. *Plant Systematics and Evolution*. 301(7): 1907-1926.
- Arroyo J., Barrett S.C.H, Hidalgo R., and Cole W.W. 2002. Evolutionary maintenance of stigma-height dimorphism in *Narcissus papyraceus* (Amaryllidaceae). *American Journal of Botany*. 89(8): 1242-1249.

- Athanasίου A., Khosravi D., Tamari F., and Shore J.S. 2003. Characterization and localization of short-specific polygalacturonase in distylous *Turnera subulata* (Turneraceae). *American Journal of Botany*. 90: 675-682.
- Athanasίου A. and Shore J.S. 1997. Morph-specific proteins in pollen and styles of distylous *Turnera* (Turneraceae). *Genetics*. 146: 669-679.
- Awadalla P., Charlesworth D. 1999. Recombination and selection at *Brassica* self-incompatibility loci. *Genetics*. 152: 413-425.
- Baker H.G. 1964. Variation in style length in relation to outbreeding in *Mirabilis* (Nyctaginaceae). *Evolution*. 18(3): 507-509.
- Baker H.G. 1966. The evolution, functioning and breakdown of heteromorphic incompatibility systems. I Plumbaginaceae. *Evolution*. 20: 349-368.
- Barrett S.C.H. 1988. The evolution, maintenance, and loss of self-incompatibility systems. In: Doust J.L. and Doust L.L. (ed.) *Plant Reproductive Ecology: Patterns and Strategies*. Oxford University Press, New York, pp. 98-124.
- Barrett S.C.H. 1990. The evolution and adaptive significance of heterostyly. *Trends in Ecology and Evolution*. 5(5): 144-148.
- Barrett S.C.H. 1992. Heterostylous genetic polymorphisms: Model Systems for Evolutionary Analysis. In: Barrett S.C.H. (ed.) *Evolution and Function of Heterostyly*. Springer, Berlin, Heidelberg, pp. 1-29.
- Barrett S.C.H. 2002. The evolution of plant sexual diversity. *Nature Reviews Genetics*. 3: 274-284.
- Barrett S.C.H. and Glover D.E. 1985. On the Darwinian hypothesis of the adaptive significance of tristylous. *Evolution*. 39(4): 766-774.
- Barrett S.C.H., Harder L.D., Cole W.W. 2004. Correlated evolution of floral morphology and mating-type frequencies in a sexually polymorphic plant. *Evolution*. 58(5): 964-975.
- Barrett S.C.H. and Shore J.S. 2008. New insights on heterostyly: Comparative biology, ecology and genetics. In: Franklin-Tong V. (ed) *Self-Incompatibility in Flowering plants: Evolution, Diversity and Mechanisms*. Springer-Verlag, Berlin, pp. 3-32.
- Bateson W. and Gregory R.P. 1905. On the inheritance of heterostyly in *Primula*. *Proceedings of the Royal Society of London Series B*. 76: 581-586.
- Beye M., Hunt G.J., Page R.E., Fondrk M.K., Grohmann L., Moritz R.F.A. 1999. Unusually high recombination rate detected in the sex locus region of the honey bee (*Apis mellifera*). *Genetics*. 153: 1701-1708.

- Billiard S., Castric V., and Vekemans X. 2007. A general model to explore complex dominance patterns in sporophytic self-incompatibility systems. *Genetics* 175: 1351-1369.
- Bosch M. and Franklin-Tong V.E. 2008. Self-incompatibility in *Papaver*: Signalling to trigger PCD in incompatible pollen. *Journal of Experimental Botany*. 59(3): 481-490.
- Boyes D.C., Nasrallah M.E., Vrebalov J., and Nasrallah J.B. 1997. The self-incompatibility haplotypes of *Brassica* contain highly divergent and rearranged sequences of ancient origin. *The Plant Cell*. 9: 237-247.
- Casselmann A.L., Vrebalov J., Conner J.A., Singhal A., Giovannoni J., Nasrallah M.E., and Nasrallah J.B. 2000. Determining the physical limits of the *Brassica S* locus by recombinational analysis. *The Plant Cell*. 12: 23-33.
- Castric V. and Vekemans X. 2004. Plant self-incompatibility in natural populations: A critical assessment of recent theoretical and empirical advances. *Molecular Ecology*. 13: 2873-2889.
- Castric V. and Vekemans X. 2007. Evolution under strong balancing selection: How many codons determine specificity at the female self-incompatibility gene SRK in Brassicaceae? *BMC Evolutionary Biology*. 7: 132-147.
- Chafe P.D.J., Lee T., Shore J.S. 2015. Development of a genetic transformation system for distylous *Turnera joelii* (Passifloraceae) and the characterization of a self-compatible mutant. *Plant Cell Tissue and Organ Culture*. 120: 507-517.
- Chan S.K., Hsing M., Hormozdiari F., and Cherkasov A. 2007. Relationship between insertion/deletion (indel) frequency of proteins and essentiality. *BMC Bioinformatics*. 8 (227): 1-13.
- Charlesworth B. and Charlesworth D. 1979a. The maintenance and breakdown of distyly. *American Naturalist*. 114(4): 499-513.
- Charlesworth B. and Charlesworth D. 1979b. A model for the evolution of distyly. *American Naturalist*. 114(4): 467-498.
- Charlesworth B., Nordborg M., and Charlesworth D. 1997. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical Research*. 70: 155-174.
- Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*. 2(4): 379-384.

- Chen G., Zhang B., Liu L., Li Q., Zhang Y., Xie Q., and Xue Y. 2012. Identification of the ubiquitin-binding structure in the *S*-locus F-Box protein controlling S-RNase-based self-incompatibility. *Journal of Genetics and Genomics*. 39: 93-102.
- Cho S., Huang Z.Y., Green D.R., Smith D.R., and Zhang J. 2006. Evolution of complementary sex-determination of honey bees: Balancing selection and trans-species polymorphisms. *Genome Research*.16: 1366-1375.
- Cohen J.I. 2010. "A case to which no parallel exists": The influence of Darwin's *Different Forms of Flowers*. *American Journal of Botany*. 97(5): 701-716.
- Crowe L.K. 1964. The evolution of outbreeding in plants. I. The angiosperms. *Heredity*. 19:435-457.
- Darwin C. 1862. On the two forms, or dimorphic condition, in the species of *Primula*, and on their remarkable sexual relations. *Journal of the Proceedings of the Linnean Society of London (Botany)*. 6(22): 77-96.
- Darwin C. 1876. Recollections of the development of my mind and character (1876–1881). In: Secord JA (ed). *Charles Darwin: Evolutionary Writings Including the Autobiographies* [2010]. New York: Oxford University Press. pp. 355-425.
- Darwin C. 1877. The different forms of flowers on plants of the same species. John Murray, London, U.K.
- Dayhoff M.O., Schwartz R.M., and Orcutt B.C. 1978. A model for evolutionary change in proteins. *Atlas of Protein Sequence and Structure*. 5: 345–352.
- de Graaf B.H.J. Vatovec S., Juarez-Daiz J.A., Chai L., Kooblall K., Wilkins K.A., Zou H., Forbes T., Franklin F.C.H., and Franklin-Tong V.E. 2012. The *Papaver* self-incompatibility pollen *S*-determinant, PrpS, functions in *Arabidopsis thaliana*. *Current Biology*. 22: 154-159.
- de Nettancourt D. 1997. Incompatibility in angiosperms. *Sexual Plant Reproduction*. 10: 185-199.
- Dimmic M. W., Rest J.S., Mindell D.P., and Goldstein R.A. 2002. rtREV: An amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *Journal of Molecular Evolution*. 55: 65-73.
- Dolgin E.S. and Charlesworth B. The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics*. 178: 2169-2177.
- Doyle J.J. and Doyle J.L. 1990. Isolation of plant DNA from fresh tissue. *Focus*. 12:13-15.
- Dorwick V.P.G. 1956. Heterostyly and homostyly in *Primula obconica*. *Heredity*. 10: 219-236.

- Dulberger R. 1992. Floral polymorphisms and their functional significance in the heterostylous syndrome. In: Barrett S.C.H. (ed.) *Evolution and Function of Heterostyly*. Springer, Berlin, Heidelberg, pp. 41-86.
- Durrett R. 2005. Mathematical flaws in Suzuki and Gojobori's test for selection. *Molecular Biology and Evolution*. (in press).
- Dwyer K.G., Balent M.A., Nasrallah J.B., and Nasrallah M.G. 1991. DNA sequences of self-incompatibility genes from *Brassica campestris* and *Brassica oleracea*: polymorphism predating speciation. *Plant Molecular Biology*. 16:181-186.
- Eaves D.J., Flores-Ortiz C., Haque T., Lin Z., Teng N., and Franklin-Tong V. 2014. Self-incompatibility in *Papaver*: Advances in integrating the signalling network. *Biochemical Society Transactions*. 42(2): 370-376.
- Eckert C.G., Manicacci D., Barrett S.C.H. 1996. Frequency-dependent selection on morph ratios in tristylous *Lythrum salicaria* (Lythraceae). *Heredity*. 77: 581-588.
- Edgar R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32(5): 1792-1797.
- Ernst A. 1936. Weitere untersuchungen zur phananalyse zum fertilitatsproblem und zur genetic heterostyler primeln. II *Primula hortensis*. Wettstein. Arch. Julius Klaus-Stift. Vererbungsforsch. Socialanthropol. Rassenhyg. 11:1-280.
- Ernst A. 1955. Self-fertility in monomorphic Primulas. *Genetica* 27: 391-448.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*. 17(6): 368-376.
- Fernandez A. and Arbo M.M. 1990. Gametas no reducidas y relaciones genomicas en tres especies de *Turnera* (Turneraaceae). *Darwiniana*. 30(1-4): 21-26.
- Fernandez A. and Arbo M.M. 1993. Citogenetica de hibridos entre *Turnera grandidentata* (4x) y *T. subulata* y *T. scabra* (2x) (Turneraceae). *Bonplandia*. 7(1-4): 119-127.
- Foote H.C.E., Ride J.P., Franklin-Tong V.E., Walker E.A., Lawrence M.J., and Franklin F.C.H. 1994. Cloning and expression of a distinctive class of self-incompatibility (S) gene from *Papaver rhoeas* L. *Proceedings of the National Academy of Science of the United States of America*. 91: 2265-2269.
- Fornoni J. and Dominguez C.A. 2015. Beyond the heterostylous syndrome. *New Phytologist*. 206: 1191-1192.

- Franklin-Tong V.E., Atwal K.K., Howell E.C., Lawrence M.J., and Franklin F.C.H. 1991. Self-incompatibility in *Papaver rhoeas*: there is no evidence for the involvement of stigmatic ribonuclease activity. *Plant, Cell and Environment*. 14(4): 423-429.
- Franklin-Tong V.E. and Franklin F.C.H. 2003. The different mechanisms of gametophytic self-incompatibility. *Philosophical Transactions of the Royal Society of London B*. 358: 1025-1032.
- Ganders F. R. 1974. Disassortative pollination in the distylous plant *Jepsonia heterandra*. *Canadian Journal of Botany*. 52: 2401-2406.
- Ganders F.R. 1979. The biology of heterostyly. *New Zealand Journal of Botany*. 17(4): 607-635.
- Gebhardt C., Ritter E., Barone A., Debener T., Walkemeier B., Schachtschable U., Kaufmann H., Thompson R.D., Bonierbale M.W., Ganai M.W., Tanksley S.D., and Salamini F. 1991. RFLP maps of potato and their alignment with the homoeologous tomato genome *Theoretical and Applied Genetics*. 83(1): 49-57.
- Geitmann A., Snowman B.N., Emons A.M., and Franklin-Tong V.E. 2000. Alterations in the actin cytoskeleton of pollen tubes are induced by the self-incompatibility reaction in *Papaver rhoeas*. *Plant Cell*. 12: 1239-1251.
- Gibbs P.E. 1986. Do homomorphic and heteromorphic self-incompatibility systems have the same sporophytic mechanism? *Plant Systematics and Evolution*. 154: 285-223.
- Gilmartin P.M. and Li J. 2010. Delineation of the *S*-locus in *Turnera subulata*: Homing in on heterostyly. *Heredity*. 105: 161-162.
- Glisovic T., Bachorik J.L., Yong J., and Dreyfuss G. 2008. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters*. 582(14): 1977-1986.
- Goldraij A., Kondo K., Lee C.B., Hancock C.N., Sivaguru M., Vazquez-Santana S., Sunran K., Phillips T.E., Cruz-Garcia F., and McClure B.A. 2006. Compartmentalization of S-RNase and HT-B degradation in self-incompatible *Nicotiana*. *Nature*. 439: 805-810.
- Goodstadt L. and Ponting C. 2001. CHROMA: consensus-based colouring of multiple sequence alignments for publication. *Bioinformatics*. 9: 845-846.
- Goring D.R., Indriolo E., and Samuel M.A. 2014. The ARC1 E3 ligase promotes a strong and stable self-incompatibility response in *Arabidopsis* species: Response to the Nasrallah and Nasrallah commentary. *The Plant Cell*. 26: 3842-3846.
- Graham S.W. and Barrett S.C.H. 2004. Phylogenetic reconstruction of the evolution of stylar polymorphisms in *Narcissus* (Amaryllidaceae). *American Journal of Botany*. 91: 1007-1021.

- Graves J.A.M. 1998. Evolution of the mammalian Y chromosome and sex-determining genes. *The Journal of Experimental Zoology*. 281: 472-481.
- Grosberg R.K. and Hart M.W. 2000. Mate selection and the evolution of highly polymorphic self/nonself recognition genes. *Science*. 289: 2111-2114.
- Hasegawa M., Kishino H., and Yano T. 1985. Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*. 22:160-174.
- Hasselmann M. and Beye M. 2004. Signatures of selection among sex determining alleles of the honey bee. *Proceedings of the National Academy of Science of the United States of America*. 101: 4888-4893.
- Hatakeyama K., Watanabe M., Takasaki T., Ojima K., and Hinata K. 1998. Dominance relationships between S-alleles in self-incompatible *Brassica campestris* L. *Heredity*. 80: 241-247.
- Hebsgaard S.M., Korning P.G., Tolstrup N., Engelbrecht J., Rouze P., and Brunak S. 1996. Splice site prediction in *Arabidopsis thaliana* DNA by combining local and global sequence information. *Nucleic Acids Research*. 24(17): 3439-3452. Available from: <http://www.cbs.dtu.dk/services/NetPGene/>.
- Hildebrand, F. 1866. Experimente zur Dichogamie und zum Dimorphismus. *Botanische Zeitung*. 23: 13.
- Hiroshi S., Takayama S., Iwano M., Shimosatao H., Funato M., Nakagawa T., Che F-S., Suzuki G., Watanabe M., Hinata K., and Isogai A. 2001. A pollen coat protein, SP11/SCR, determines the pollen S-specificity in the self-incompatible *Brassica* species. *Plant Physiology*. 125: 2095-2103.
- Hollister J.D., Ross-Ibarra J., and Gaut B.S. 2010. Indel-associated mutation rate varies with mating system in flowering plants. *Molecular Biology and Evolution*. 27(2): 409-416.
- Hu B., Jin J., Guo A-Y., Zhang H., Luo J., and Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 31(8): 1296-1297.
- Hua Z-H., Fields A., and Kao T-h. 2008. Biochemical models for S-RNase-based self-incompatibility. *Plant Molecular Biology*. 1(4): 575-585.
- Hughes A. L. and Yeager M. 1997. Comparative evolutionary rates of introns and exons in murine rodents. *Journal of Molecular Evolution*. 45: 125-130.
- Indriolo E., Tharmapalan P., Wright S.I., and Goring D.R. 2012. The ARC1 E3 ligase gene is frequently deleted in self-compatible Brassicaceae species and has a conserved role in *Arabidopsis lyrata* self-pollen rejection. *Plant Cell*. 24: 4607-4620.

- Ioerger T.R., Clark A.G., and Kao T-H. 1990. Polymorphism at the self-incompatibility locus in Solanaceae predates speciation. *Proceedings of the National Academy of Sciences of the United States of America*. 87: 9732-9735
- Jones D.T., Taylor W.R., and Thornton J.M. 1992. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*. 8: 275-282.
- Jukes T.H. and Cantor C.R. 1969. *Evolution of Protein Molecules*. New York: Academic Press. pp. 21–132.
- Kakita M., Murase K., Iwano M., Matsumoto T., Watanabe M., Shiba H., Isogai A., and Takayama S. 2007. Two distinct forms of *M*-locus protein kinase localize to the plasma membrane and interact directly with *S*-locus receptor kinase to transducer self-incompatibility signalling in *Brassica rapa*. *Plant Cell*. 19: 3961-3973.
- Kakui H., Kato M., Ushijima K., Miyoko K., Kato S., and Hidenori S. 2011. Sequence divergence and loss-of-function phenotypes of *S*-locus F-box brothers genes are consistent with recognition by multiple pollen determinants in self-incompatibility of Japanese pear (*Pyrus pyrifolia*). *The Plant Journal*. 68: 1028-1038.
- Kamu E. and Charlesworth D. 2005. Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. *Current Biology*. 15(19): 1773-1778.
- Kao T-H. and Tsukamoto T. 2004. The molecular genetic basis of S-RNase-based self-incompatibility. *The Plant Cell*. 16: S72-S83.
- Kao T.-H. and McCubbin A.G. 1996. How flowering plants discriminate between self and non-self-pollen to prevent inbreeding. *Proceedings of the National Academy of Science of the United States of America*. 93: 12059-12065.
- Keller B., Thomson J.D., and Conti E. 2014. Heterostyly promotes disassortative pollination and reduces sexual interference in Darwin's primroses: Evidence from experimental studies. *Functional Ecology*. 28: 1413-1425.
- Khosravi D., Joulaie R., and Shore J.S. 2003. Immunocytochemical distribution of polygalacturonase and pectins in styles of distylous and homostylous Turneraceae. *Sexual Plant Reproduction*. 16: 179-190.
- Khosravi D., Yang E.C.C., Siu K.W.M., and Shore J.S. 2004. High level of α -Dioxygenase in short styles of distylous *Turnera* species. *International Journal of Plant Sciences*. 165(6): 995-1006.
- Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. 16 (2): 111-120.

- King V., Goodfellow P.N., Wilkerson A.J.P., Johnson W.E., O'Brien S.J., and Pecon-Slattery J. 2007. Evolution of the male-determining gene *SRY* within the cat family Felidae. *Genetics*. 175: 1855-1867.
- Klein D.E., Freitas L., and Cunha M.D. 2009. Self-incompatibility in a distylous species of Rubiaceae: Is there a single incompatibility response of the morphs? *Sexual Plant Reproduction*. 22: 121-131.
- Klein J., Sato A., Nagl S., and O'hUigin C. 1998. Molecular trans-species polymorphism. *Annual Review of Ecology and Systematics*. 29: 1-21.
- Kohn J.R. and Barrett S.C.H. 1992. Experimental studies on the functional significance of heterostyly. *Evolution*. 46(1): 43-55.
- Kreitman M. and Akashi H. 1995. Molecular evidence for natural selection. *Annual Review of Ecology and Systematics*. 26: 403-422.
- Kryazhimskiy S. and Plotkin J.B. 2008. The population genetics of dN/dS. *PLoS Genetics*. 4(12): 1-10.
- Kubo K., Entani T., Takara A., Wang N., Fields A.M., Hua Z., Toyoda M., Kawashima S., Ando T., Isogai A., Kao T., and Takayama S. 2010. Collaborative non-self recognition system in S-RNase-based self-incompatibility. *Science*. 330(6005): 796-799.
- Kubo K., Paape T., Hatakeyama M., Entani T., Takara A., Kajihara K., Tsukhara M., Shimizu-Inatsugi R., Shimizu K.K., and Takayama S. 2015. Gene duplication and genetic exchange drive the evolution of S-RNase-based self-incompatibility in *Petunia*. *The Plant Journal*. (1): 1-9.
- Kurian V. and Richards A.J. 1997. A new recombinant in the heteromorphy 'S' supergene in *Primula*. *Heredity*. 78: 383-390.
- Labonne J.D.J. 2011. Genetic mapping and positional cloning of the S-locus of distylous *Turnera subulata* (Turneraceae) [PhD thesis]. Toronto (ON): York University.
- Labonne J.D.J., Goultiaeva A., and Shore J.S. 2009. High-resolution mapping of the S-locus in *Turnera* leads to the discovery of three genes tightly associated with the S-alleles. *Molecular Genetics and Genomics*. 281: 673-685.
- Labonne J.D.J., Hilliker A.J., and Shore J.S. 2007. Meiotic recombination in *Turnera* (Turneraceae): Extreme sexual differences in rates, but no evidence for recombination suppression associated with the distyly (S) locus. *Heredity*. 98: 411-418.
- Labonne J.D.J. and Shore J.S. 2011. Positional cloning of the s-haplotype determining the floral and incompatibility phenotype of the long-styled morph of distylous *Turnera subulata*. *Molecular Genetics and Genomics*. 285: 101-111.

- Labonne J.D.J., Tamari F., and Shore J.S. 2010. Characterization of X-ray generated floral mutants carrying deletions at the *S*-locus of distylous *Turnera subulata*. *Heredity*. 105: 235-243.
- Labonne J.D.J., Vaisman A., and Shore J.S. 2008. Construction of a first genetic map of distylous *Turnera* and fine-scale map of the *S*-locus region. *Genome*. 51: 471-478.
- Lai Z., Ma W., Han B., Liang L., Zhang Y., Hong G., and Xue Y. 2002. An F-box gene linked to the self-incompatibility (*S*) locus of *Antirrhinum* is expressed specifically in pollen and tapetum. *Plant Molecular Biology*. 50: 29-42.
- Lamesch P., Berardini T.Z., Li D., Swarbreck D., Wilks C, Sasidharan R., Muller R., Dreher K., Alexander D.L., Garcia-Hernandez M., Karthikeyan A.S., Lee C.H., Nelson W.D., Ploetz L, Singh S., Wensel A, and Huala E. 2011. The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*. 1-9.
- Lau P. and Bosque C. 2003. Pollen flow in the distylous *Palicourea fendleri* (Rubiaceae): An experimental test of the disassortative pollen flow hypothesis. *Oecologia*. 135: 593-600.
- Le S. Q. And Gascuel O. 2008. An improved general amino acid replacement matrix. *Molecular Biology and Evolution*. 25(7): 1307-1320.
- Lee H-S., Huang S., and Kao T-H. 1994. S proteins control rejection of incompatible pollen in *Petunia inflata*. *Nature*. 367: 560-563.
- Lewis D. 1982. Incompatibility, stamen movement, and pollen economy in a heterostyled tropical forest tree, *Cratogeomys formosum* (Guttiferae). *Proceedings of the Royal Society of London. Series B*. 214: 273-283.
- Lewis D. and Jones D.A. 1992. The genetics of heterostyly. In: Barrett S.C.H. (ed.) *Evolution and Function of Heterostyly*. Springer, Berlin, Heidelberg, pp. 129-150.
- Li J., Dudas B., Webster M.A., Cook H.E., Davies B.H., and Gilmartin P.M. 2009. Hose in hose, an *S*-locus-linked mutant of *Primula vulgaris*, is caused by an unstable mutant at the *Globosa* locus. *Proceedings of the National Academy of Sciences of the United States of America*. 107(12): 5664-5668.
- Li J., Webster M., Dudas B., Cook H., Manfield I., Davies B., and Gilmartin P.M. 2008. The *S*-locus-linked *Primula* homeotic mutant *sepaloid* shows characteristics of B-function mutant but does not result from mutation in B-function gene. *The Plant Journal*. 56: 1-12.
- Li J., Webster M., Furuya M., and Gilmartin P.M. 2007. Identification and characterization of pin and thrum alleles of two genes that co-segregate with the *Primula S*-locus. *The Plant Journal*. 51: 18-31.

- Li J., Webster M.A., Smith M.C., and Gilmartin P.M. 2011. Floral heteromorphy in *Primula vulgaris*: Progress towards isolation and characterization of the *S*-locus. *Annals of Botany*. 108(4): 715-726.
- Li J., Webster M.A., Wright J., Cocker J.M., Smith M.C., Badakshi F., Heslop-Harrison P., and Gilmartin P.M. 2015. Integration of genetic and physical maps of the *Primula vulgaris S* locus and localization by chromosome in situ hybridization. *New Phytologist*. (2015):1-12.
- Li S., Samaj J., and Franklin-Tong V.E. 2007. A mitogen-activated protein kinase signals to programmed cell death induced by self-incompatibility in *Papaver* pollen. *Plant Physiology*. 145: 236-245.
- Librado P and Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 25: 1451-1452.
- Liu Q., Kasuga M., Sakuma Y., Abe H., Miura S., Yamaguchi-Shinozaki K., and Shinozaki K. 1998. Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *The Plant Cell*. 10: 1391-1406.
- Liu Z., Moore P.H., Ma H., Ackerman C.M., Ragiba M., Yu Q., Pearl H.M., Kim M.S., Charlton J.W., Stiles J.I., Zee F.T., Paterson A.H., and Ming R. 2004. A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature*. 427: 348-352.
- Lloyd D.G. and Webb C.J. 1992a. The evolution of heterostyly. In: Barrett S.C.H. (ed.) *Evolution and Function of Heterostyly*. Springer, Berlin, Heidelberg, pp. 151-178.
- Lloyd D.G. and Webb C.J. 1992b. The selection of heterostyly. In: Barrett S.C.H. (ed.) *Evolution and Function of Heterostyly*. Springer, Berlin, Heidelberg, pp. 179-208.
- Lloyd D.G. and Yates J.M.A. 1982. Intrasexual selection and the segregation of pollen and stigmas in hermaphrodite plants, exemplified by *Wahlenbergia albomarginata* (Campanulaceae). *Evolution*. 36(5): 903-913.
- Lopez A., Fernandez A., and Shore J.S. 2013. Inferences on the origins of polyploidy *Turnera* species (Passifloraceae) based on molecular data. *Botany*. 91: 167-175.
- Ma H., Wang Y., Li Z.H., Wan Y.M., Liu X.X., Liang N. 2009. A study of the breeding system of *Luculia pinceana*. *Forest Research*. 22: 373-378.
- Manfield, I. W., Pavlov V.K., Li J., Cook H.E., Hummel F., and Gilmartin P.M. 2005. Molecular characterization of DNA sequences from the *Primula vulgaris S*-locus. *Journal of Experimental Botany*. 56: 1177-1188.
- Mann, H. B. and Whitney D.R. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*. 18: 50-60

- Marchler-Bauer A., Derbyshire M.K., Gonzales N.R, Lu S., Chitsaz F., Geer L.Y., Geer R.C., He J., Gwadz M., Hurwitz D.I., Lanczycki C.J, Fu F., Marchler G.H., Song J.S., Thanki N., Wang Z., Yamashita R.A, Zhang D., Zheng C. and Bryant S.H. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Research*. 43(Database Issue): D222-D226.
- Mast A.R. and Conti E. 2006. The primrose path to heterostyly. *New Phytologist*. 171: 439-442.
- Mather K. 1950. The genetical architecture of heterostyly in *Primula sinensis*. *Evolution*. 4(4): 340-352.
- Mather K. and de Winton D. 1941. Adaptation and counter-adaptation of the breeding system in *Primula*. *Annals of Botany*. 5: 279-311.
- Matsui K., Nishio T., and Tetsuka T. 2004 Genes outside the *S* Supergene suppress *S* functions in buckwheat (*Fagopyrum esculentum*). *Annals of Botany*. 94: 805-809.
- Matsui K., Tetsuka T., Nishio T., and Hara T. 2003. Heteromorphic incompatibility in self-compatible plants produced by a cross between common and wild buckwheat. *New Phytologist*. 159: 701-708.
- Matton D.P., Nass N., Clarke A.E., and Newbiggin E. 1994. Self-incompatibility: How plants avoid illegitimate offspring. *Proceedings of the National Academy of Sciences of the United States of America*. 91: 1992-1997.
- May G., Shaw F., Badrane H., and Vekemans X. 1999. The signature of balancing selection: Fungal mating compatibility gene evolution. *Proceedings of the National Academy of Sciences of the United States of America*. 96: 9172-9177.
- McClure B.A. 2004. *S*-RNase and *SLF* determine *S*-haplotype-specific pollen recognition and rejection. *The Plant Cell*. 16: 2840-2847.
- McClure B.A. 2010. Darwin's foundation for investigating self-incompatibility and the progress toward a physiological model for *S*-RNase-based SI. *Journal of Experimental Botany*. 60(4): 1069-1081.
- McClure B.A., Cruz-Garcia F., and Romero C. 2011. Compatibility and Incompatibility in *S*-RNase –based systems. *Annals of Botany*. 108: 647-658.
- McClure B.A., Gray J.E., Anderson M.A., and Clarke A.E. 1990. Self-incompatibility in *Nicotiana glauca* involves degradation of pollen rRNA. *Nature*. 347: 757-760.
- McClure B.A., Haring V., Ebert P.R., Anderson M.A., Simpson R.J., Sakiyama F., and A.E. Clarke. 1989. Style self-incompatibility gene products of *Nicotiana glauca* are ribonucleases. *Nature*. 342:955-957.

- McClure B.A., Mou B., Canevascini S., and Bematzky R. 1999. A small asparagine-rich protein required for *S*-allele-specific pollen rejection in *Nicotiana*. *Proceedings of the National Academy of Sciences of the United States of America*. 96: 13548-13553.
- Miljus-Dukic J.D., Ninkovic S., and Nesovic M. 2003. Effects of protein phosphatase inhibitors and calcium antagonists on self-incompatible reaction in buckwheat. *Biologica Plantarum*. 46(3): 475-478.
- Miljus-Dukic J.D., Ninkovic S., Radovic S.R., Maksimovic V.R., Brkljacic J., and Neskovic M. 2004. Detection of proteins possibly involved in self-incompatibility response in distylous buckwheat. *Biologica Plantarum*. 48(2): 293-296.
- Miljus-Dukic J.D., Radovic S.R., and Maksimovic V.R. 2007. Treatment of isolated pistils with protease inhibitors overcomes the self-incompatibility response in buckwheat. *Archives of Biological Sciences, Belgrade*, 59(1): 45-49.
- Moreira M.A.M. 2002. SRY evolution in Cebidae (Platyrrhini: Primates). *Journal of Molecular Evolution*. 55: 92-103.
- Morimoto T., Akagi T., and Tao R. 2015. Evolutionary analysis of genes for S-RNase-based self-incompatibility reveals *S*-locus duplications in ancestral Rosaceae. *The Horticulture Journal*. Online Advance Publication: 1-10.
- Muller H.J. 1964. The relation of recombination to mutational advance. *Mutation Research*. 1(1): 2-9.
- Murfett J., Atherton T.L., Mou B., Gasser C.S., and McClure B.A. 1994. S-RNase expressed in transgenic *Nicotiana* causes *S*-allele specific pollen rejection. *Nature*. 367: 563-566.
- Murrell B., Moola S., Mabona A., Weighill T., Sheward D., Pond S.L.K., and Scheffler K. 2013. FUBAR: A fast, unconstrained Bayesian approximation for inferring selection. *Molecular Biology and Evolution*. 30(5): 1196-1205.
- Murrell B., Weaver S., Smith M.D., Wertheim J.O., Murrell S., Aylward A., Eren K., Pollner T., Martin D.P., Smith D.M., Scheffler K., and Pond S.L.K. 2015. Gene-wide identification of episodic selection. *Molecular Biology and Evolution*. 32(5): 1365-1371.
- Murrell B., Wertheim J.O., Moola S., Weighill T., Scheffler K., Pond S.L.K. 2012. Detecting individual sites subject to episodic diversifying selection. *PLOS Genetics*. 8(7): 1-10.
- Murtagh V.J., O'Meally D., Sakovic N., Delbridge M.L., Kuroki Y., Boore J.L., Toyoda A., Jordan K.S., Pask A.J., Renfree M.B., Fujiyama A., Graves J.A.M., and Waters P.D. 2012. Evolutionary history of novel genes on the tammar wallaby Y chromosome: Implications for sex chromosome evolution. *Genome Research*. 22: 498-507.

- Nasrallah JB. 2005. Recognition and rejection of self in plant self-incompatibility: comparisons to animal histocompatibility. *Trends in Immunology*. 26: 412–418.
- Nasrallah J.B. 2011. Self-incompatibility in the Brassicaceae. In: Schmidt R., and Bancroft I (eds) *Genetics and Genomics of the Brassicaceae*, Plant Genetics and Genomics: Crops and Models 9. Springer Science and Business Media, pp. 389-411.
- Nasrallah and Nasrallah. 1989. The molecular genetics of self-incompatibility in *Brassica*. *Annual Review of Genetics*. 23:121-139.
- Nasrallah J.B. and Nasrallah M.E. 2014a. *S*-locus receptor kinase signalling. *Biochemical Society Transactions*. 42: 313-319.
- Nasrallah J.B. and Nasrallah M.E. 2014b. Robust self-incompatibility in the absence of functional *ARC1* gene in *Arabidopsis thaliana*. *The Plant Cell*. 26: 3838-3841.
- Nasrallah J.B, Kao T-h., Chen C-H., Goldberg M.L., and Nasrallah M.E. 198. Amino-acid sequence of glycoproteins encoded by three alleles of the *S*-locus of *Brassica oleracea*. *Nature*. 326:617-619
- Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press.
- Newbigin E., Anderson M.A., Clarke A.E. 1993. Gametophytic self-incompatibility systems. *The Plant Cell* 5: 1315-1324.
- Newbigin E., Paape T., and Kohn J.R. 2008. RNase-based self-incompatibility: Puzzled by pollen *S*. *The Plant Cell*. 20: 2286-2292.
- Nou, I.S., Watanabe, M., Isuzugawa, K., Isogai, A. and Hinata, K. 1993. Isolation of *S*-alleles from a wild population of *Brassica campestris* L. at Balcesme, Turkey and their characterization by *S*-glycoproteins. *Sexual Plant Reproduction*. 6: 71–78
- Nowacka-Woszek J. and Switonski M. 2009. Differentiated evolutionary conservatism and lack of polymorphism of crucial sex determination genes (*SRY* and *SOX9*) in four species of the family Canidae. *Folia Biologica*. 57(3-4): 171-176.
- Nowak M.D., Davis A.P., Anthony F., and Yoder A.D. 2011. Expression and trans-specific polymorphism of self-incompatibility RNases in *Coffeae* (Rubiaceae). *PLoS ONE*. 6(6): 1-11.
- Nowak M.D., Russo G., Schlapbach R., Huu C.N., Lenhard M., and Conti E. 2015. The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biology*. 16:1-17.

- O'Brien M., Kapfer C., Major G., Laurin M., Bertrand C., Kondo K., Kowyama Y., Matton D.P.. 2002. Molecular analysis of the stlyar expressed *Solanum chacoense* asparagine-rich protein family related to the HT modifier of gametophytic self-incompatibility in *Nicotiana*. *The Plant Journal*. 32: 1–12.
- Olesen J.M. 1979. Floral morphology and pollen flow in the heterostylous species *Pulmonaria obscura* Dumort (Boraginaceae). *New Phytologist*. 82(3): 757-767.
- Ornduff R. 1992. Historical perspectives on heterostyly. In: Barrett S.C.H. (ed.) *Evolution and Function of Heterostyly*. Springer, Berlin, Heidelberg, pp. 31-39.
- Pérez R., Vargas P., and Arroyo J. 2004. Convergent evolution of flower polymorphism in *Narcissus* (Amaryllidaceae). *New Phytologist*. 161: 235–252.
- Pérez-Barrales R., Vargas P., and Arroyo J. 2006. New evidence for the Darwinian hypothesis of heterostyly: breeding systems and pollinators in *Narcissus* sect. Apodanthi. *New Phytologist*. 171: 553-567.
- Pond SLK and Frost S.D.W. 2005. Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*. 22: 1208-1222.
- Pond S.L.K., Frost S.D.W., and Muse S.V. 2005. Hyphy: Hypothesis testing using phylogenies. *Bioinformatics*. 21(5): 676-679.
- Pond S.L.K., Murrell B., Fourment M., Frost S.D.W., Delport W., and Scheffler K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution*. 28(11): 3033-3043.
- Pond S.L.K., Poon A.F.Y., and Frost S.D.W. 2009. Estimating selection pressures on alignments of coding sequences. In: Lemey P., Salemi M., and Vandamme A-M. (eds) *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (2nd Ed.). Cambridge University Press, New York, pp. 419-490.
- Pybus O.G. and Shapiro B. 2009. Natural selection and adaptation of molecular sequences. In: Lemey P., Salemi M., and Vandamme A-M. (eds) *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing* (2nd Ed.). Cambridge University Press, New York, pp. 419-490.
- Qiao H., Wang H, Zhao L, Zhou J, Huang J, et al. 2004. The F-box protein Ah-S2 physically interacts with S-RNases that may be inhibited by the ubiquitin/26S proteasome pathway of protein degradation during compatible pollination in *Antirrhinum*. *Plant Cell*. 16:582–95
- Rambaut, A. 2007. FigTree, a graphical viewer of phylogenetic trees. Available from: <http://tree.bio.ed.ac.uk/software/figtree>.

- Rand D.M. and Kann L.M. 1996. Excess amino acid polymorphism in mitochondrial DNA: Contrasts among genes from *Drosophila*, mice, and humans. *Molecular Biology and Evolution*. 13(6): 735-748.
- Raymond C.K., Kas A., Paddock M., Qui R., Zhou Y., Subramanian S., Chang J., Palmeieri A., Haugen E., Kaul R., and Olson M.V. 2005. Ancient haplotypes of the HLA Class II Region. *Genome Research*. 50: 1250-1257.
- Rice W.R. 1987. Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. *Genetics*. 116: 161-167.
- Richman A.D. 2000. Evolution of balanced genetic polymorphism. *Molecular Ecology*. 9: 1953-1963.
- Richman A.D. and Kohn J.R. 1996. Learning from rejection: The evolutionary biology of single-locus incompatibility. *TREE*. 11(12): 497-502.
- Richman A.D. and Kohn J.R. 1999. Self-incompatibility alleles from *Physalis*: Implications for historical inference from balanced genetic polymorphisms. *Proceedings of the National Academy of Science of the United States of America*. 96: 168-172.
- Richman A.D. and Kohn J.R. 2000. Evolutionary genetics of self-incompatibility in Solanaceae. *Plant Molecular Biology*. 42: 169-179.
- Richman A.D., Uyenoyama M.K., and Kohn J.R. 1996. Allelic diversity and gene genealogy at the self-incompatibility locus in the Solanaceae. *Science*. 273:1212-16
- Rosov S.A. and Serebtsova N.D. 1958. Honey bees and the selective fertilization of plants. XVII International Beekeeping Congress. 2: 494-501.
- Roux C., Pauwels M., Ruggiero M-V., Charlesworth D., Castric V., and Vekemans X. 2012. Recent and ancient signature of balancing selection around the *S*-locus in *Arabidopsis halleri* and *A. lyrata*. *Molecular Biology and Evolution*. 30(2): 435-447.
- Rozen S. and Skaletsky H.J. 2000. Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S and Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386. Available from: <http://frodo.wi.mit.edu/primer3/>.
- Rudd J.J., Osman K., Franklin F.C., and Franklin-Tong V.E. 2003. Activation of a putative MAP kinase in pollen is stimulated by the self-incompatibility (SI) response. *FEBS Letters*. 547: 223-227.
- Safavian D. and Shore J.S. 2010. Structure of styles and pollen tubes of distylous *Turnera joelii* and *T. scabra* (Turneraceae): Are there different mechanisms of incompatibility between the morphs? *Sexual Plant Reproduction*. 23: 225-237.

- Samuel M.A., Chong T.Y., Haasen K.E., Aldea-Brydges M.G., Stone S.L., and Goring D.R. 2009. Cellular pathways regulating responses to compatible and self-compatible pollen in *Brassica* and *Arabidopsis* stigmas intersect at Exo70A1, a putative component of the exocyst complex. *Plant Cell*. 21: 2655-2671.
- Sanabria N., Goring D., Nurnberger T., and Dubery I. 2008. Self/Nonsel self perception and recognition in plants: A comparison of self-incompatibility and innate immunity. *New Phytologist*. 178: 503-514.
- Santos-Gally R., de Castro A., Pérez-Barrales R., and Arroyo J. 2015. Styler polymorphism on the edge: Unusual flower traits in Moroccan *Narcissus broussonetii* (Amaryllidaceae). *Botanical Journal of the Linnean Society*. 117: 644-656.
- Schierup M.H., Mable B.K., Awadalla P., and Charlesworth D. 2001. Identification and characterization of a polymorphic receptor kinase gene linked to the self-incompatibility locus of *Arabidopsis lyrata*. *Genetics*. 158: 387-399.
- Schopfer C.R., Nasrallah M.E., and Nasrallah J.B. 1999. The male determinant of self-incompatibility in *Brassica*. *Science*. 286: 1697-1700.
- Seavey S.R. and Bawa K.S. 1986. Late-acting self-incompatibility in angiosperms. *The Botanical Review*. 52(2): 195-219.
- Self S.G. and Liang K-L. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *Journal of the American Statistical Association*. 82(398): 605-610.
- Sequencher® version 5.0 sequence analysis software, Gene Codes Corporation, Ann Arbor, MI USA. Available from: <http://www.genecodes.com>.
- Sharma K.D. and Boyes J.W. 1961. Modified incompatibility of buckwheat following irradiation. *Canadian Journal of Botany*. 39: 1241-1246.
- Shore J.S. 1991. Tetrasomic inheritance and isozyme variation in *Turnera ulmifolia* vars. *elgans* Urb. and *intermedia* Urb. (Turneraceae). *Heredity*. 66: 305-312.
- Shore J.S., Arbo M.M., and Fernandez A. 2006. Breeding system variation, genetics and evolution in the Turneraceae. *New Phytologist*. 171: 539-551.
- Shore J.S. and Barrett S.C.H. 1985a. Morphological differentiation and crossability among populations of the *Turnera ulmifolia* L. complex (Turneraceae). *Systematic Botany*. 10(3): 308-321.
- Shore J.S. and Barrett S.C.H. 1985b. The genetics of distyly and homostyly in *Turnera ulmifolia* L. (Turneraceae). *Heredity*. 55: 167-174.

- Sijacic P., Wang X, Skirpan A.L., Wang Y., Dowd P.E., McCubbin A.G., Huang S., and Kao T.-H. 2004. Identification of the pollen determinant of S-RNase-mediated self-incompatibility. *Nature*. 429: 302-305.
- Smith N.G.C. and Eyre-Walker A. 2002. Adaptive protein evolution in *Drosophila*. *Nature*. 415: 1022-1024.
- Snowman B.N., Kovar D.R., Shevchenko G., Franklin-Tong V.E., and Staiger C.J. 2002. Signal mediated depolymerization of actin in pollen during the self-incompatibility response. *Plant Cell*. 14: 2613-2626.
- StataCorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP.
- Stein J.C., Howlett B., Boyes D.C., Nasrallah M.E., and Nasrallah J.B. 1991. Molecular cloning of a putative receptor protein kinase gene encoded at the self-incompatibility locus of *Brassica oleracea*. *Proceedings of the National Academy of Science of the United States of America*. 88: 8816-8820.
- Stevens V.A.M. and Murray B.G. 1982. Studies on heteromorphic self-incompatibility systems: Physiological aspects of the incompatibility system of *Primula obconica*. *Theoretical and Applied Genetics*. 61: 245-256.
- Stone J.L. and Thomson J.D. 1994. The evolution of distyly: Pollen transfer in artificial flowers. *Evolution*. 48(5): 1595-1606.
- Sun P., Li S., Lu D., Williams J.S., and Kao T.-H. 2015. Pollen S-locus F-box proteins of *Petunia* involved in S-RNase-based self-incompatibility are themselves subject to ubiquitin-mediated degradation. *The Plant Journal*. 83(2): 213-223.
- Sutherland B.G., Tobutt K.R., and Robbins T.P. 2008. Trans-specific S-RNase and SFB alleles in *Prunus* self-incompatibility haplotypes. *Molecular Genetics and Genomics*. 279: 95-106
- Suzuki G., Kakizaki T., Takada Y., Shiba H., Takayama S., Isogai A., and Watanabe M. 2003. The S-haplotypes lacking *SLG* in the genome of *Brassica rapa*. *Plant Cell Reports*. 21: 911-915.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 123: 585-595.
- Takahata N. and Nei M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics*. 124: 967-978.
- Takasaki T., Hatakeyama K., Suzuki G., Watanabe M., Isogai A., and Hinata K.. 2000. The S receptor kinase determines self-incompatibility in *Brassica stigma*. *Nature*. 403:913-916.

- Takayama S. and Isogai A. 2005. Self-incompatibility in plants. *Annual Review of Plant Biology*. 56: 467-489.
- Takayama S., Isogai A., Tsukamoto C., Ueda Y., Hinata K., Okazaki K., and Suzuki A. 1987. Sequences of *S*-glycoproteins products of *Brassica campestris* self-incompatibility locus. *Nature*. 326:102-105.
- Takayama S., Shiba H., Iwano M., Shimosato H., Che F-S., Kai N., Watanabe M., Suzuki G., Hinata K., Isogai A. 2000. The pollen determinant of self-incompatibility in *Brassica campestris*. *Proceedings of the National Academy of Sciences of the United States of America*. 97(4): 1920-1925.
- Takayama S., Shimosato H., Shiba H., Funato M., Che F-S., Watanabe M., Iwano M., and Isogai A. 2001. Direct ligand-receptor complex interactions controls *Brassica* self-incompatibility. *Nature*. 431: 534-538.
- Tamari F., Athanasiou A., and Shore J.S. 2001. Pollen tube growth and inhibition in distylous and homostylous *Turnera* and *Priquetta* (Turneraceae). *Canadian Journal of Botany*. 79: 578-591.
- Tamari F., Khosravi D., Hilliker A.J., and Shore J.S. 2005. Inheritance of spontaneous mutant homostyles in *Turnera subulata* x *kravovickasii* an in autotetraploid *T. scabra* (Turneraceae). *Heredity*. 94: 207-216.
- Tamari F. and Shore J.S. 2004. Distribution of style and pollen polygalacturonases among distylous and homostylous *Turnera* and *Priquetta* spp. (Turneraceae). *Heredity*. 92: 380-385.
- Tamari F. and Shore J.S. 2006. Allelic variation for a short-specific polygalacturonase in *Turnera subulata*: Is it associated with the degree of self-incompatibility? *International Journal of Plant Sciences*. 167(1): 125-133.
- Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. *Molecular Biology and Evolution*. 9: 678-687.
- Tamura K. and Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution*. 10(3): 512-526.
- Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*. 30: 2725-2729.

- Tao R. and Iezzoni A.F. 2010. The S-RNase-based gametophytic self-incompatibility system in *Prunus* exhibits distinct genetic and molecular features. *Scientia Horticulturae*. 10(4): 423-433.
- Thomas S.G. and Franklin-Tong V.E. 2004. Self-incompatibility triggers programmed cell death in *Papaver* pollen. *Nature*. 429: 305-309.
- Thomas S.G., Huang S., Li S., Staiger C.J., and Franklin-Tong V.E. 2006. Actin depolymerization is sufficient to induce programmed cell death in self-compatible pollen. *Journal of Cell Biology*. 174: 221-229.
- Thompson J.D., Barrett S.C.H., and Baker A.M. 2003. Frequency-dependent variation in reproductive success in *Narcissus*: Implications for the maintenance of stigma-height dimorphism. *Proceedings of the Royal Society of London B*. 270: 949-953.
- Thulin M., Razafimandimbison S.G, Chafe P., Heidari N., Kool A., and Shore J.S. 2010. Phylogeny of the Turneraceae clade (Passifloraceae s.l.): Trans-Atlantic disjunctions and two new genera in Africa. *Taxon*. 61(2): 308-323.
- Tian D., Wang Q., Zhang P. Araki H., Yang S., Kreitman M., Nagylaki T., Hudson Richard, Bergelson J., and Chen J-Q. 2008. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. *Nature*. 45 (7209): 105-108.
- Truyens S., Arbo M.M., and Shore J.S. 2005. Phylogenetic relationships, chromosome and breeding system evolution in *Turnera* (Turneraceae): Inferences from ITS sequence data. *American Journal of Botany*. 92(10): 1749-1758.
- Tsutsui T., Kato W., Asada Y., Sako K., Sato T., Sonoda Y., Kidokoro S., Yamaguchi-Shinozaki K., Tamaoki M., Arakawa K., Ichikawa T., Nakazawa M., Seki M., Shinozaki K., Matsui M., Ikeda A., and Yamaguchi J. 2009. *DEARI*, a transcriptional repressor of DREB protein that mediates plant defense and freezing stress responses in *Arabidopsis*. *Journal of Plant Research*.122: 633-643.
- Tuskan G.A., Difazio S. Jansson S., Bohlmann J., Grigorjiev I., Hellsten U., Putnam N., Ralph S., Rombauts S., Salamov A., Schein J., Sterck L., Aerts A., Bhalerao R.R., Bhalerao R.P., Blaudez D., Boerjan W., Brun A., Brunner A., Busov V., Campbell M., Carlson J., Chalot M., Chapman J., Chen G.-L., Cooper D., Coutinho P.M., Couturier J., Covert S., Cronk Q., Cunningham R., Davis J., Degroeve S., Dejardin A., dePamphillis C., Detter J., Dirks B., Dubchak I., Duplessis S, et al. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*. 3113(5793): 1596-1604.
- Urban, I. 1883. Monographie der familia der Turneraceen. *Jahrb. Konigl. Bot. Gart.* 2: 1-152.

- Ushijima K, Nakano R, Bando M, Shigezane Y, Ikeda K, Namba Y, Kume S, Kitabata T, Mori H, Kubo Y. 2012. Isolation of the floral morph-related genes in heterostylous flax (*Linum grandiflorum*): the genetic polymorphism and the transcriptional and post-transcriptional regulations of the *S* locus. *Plant Journal*. 69: 317–331.
- Ushijima K., Ikeda K., Nakano R., Matsubara M., Tsuda Y., and Kubo Y. 2015. Genetic control of floral morph and petal pigmentation in *Linum grandiflorum* Desf., a heterostylous flax. *The Horticulture Journal Preview*. (2015): 1-8.
- Uyenoyama M.K. 2005. Evolution under tight linkage to mating type. *New Phytologist*. 165: 63-70.
- Vieira C.P., Charlesworth D., Vieira J. 2003. Evidence for rare recombination at the gametophytic self-incompatibility locus. *Heredity*. 91: 262–267.
- Vuilleumier B.S. 1967. The origin and evolutionary development of heterostyly in the angiosperms. *Evolution*. 21(2): 210-226.
- Wang Y., Scarth R., and Campbell C. 2005. S^h and S_c – Two complimentary dominant genes that control self-compatibility in buckwheat. *Crop Science*. 45: 1229-1234.
- Wang X., Zhang J., and Zhang Y. 2002. Erratic evolution of *SRY* in higher primates. *Molecular Biology and Evolution*. 19(4): 582-584.
- Watterson G.A. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*. 7(2): 256-276.
- Webb C.J. and Lloyd D.G. 1986. The avoidance of interference between the presentation of pollen and stigmas in angiosperms II. Herkogamy. *New Zealand Journal of Botany*. 24: 163-178.
- Wedderburn F. and Richards A.J. 1990. Variation in within-morph incompatibility inhibition sites in heteromorphic *Primula* L. *New Phytologist*. 116: 149-162.
- Weller S.G. 2009. The different forms of flowers – what have we learned since Darwin? *Botanical Journal of the Linnean Society*. 160: 249-262.
- Wheeler D. and Newbigin E. 2007. Expression of 10 *S*-class *SLF*-like genes in *Nicotiana glauca* pollen and its implications for understanding the pollen factor of the *S*-locus. *Genetics*. 177 (4): 2171-2180.
- Wheeler M.J., Armstrong S.A., Franklin-Tong V.E., and Franklin F.C.H. 2003. Genomic organization of the *Papaver rhoeas* self-incompatibility *SI* locus. *Journal of Experimental Botany*. 54: 131–139

- Wheeler M.J., de Graaf B.H.J., Hadjiosif N., Perry R.M. Poulter N.S., Osman K., Vatovec S., Harper A., Franklin F.C.H., and Franklin-Tong V.E. 2009. Identification of the pollen self-compatibility determinant in *Papaver rhoeas*. *Nature*. 459(7249): 992-995.
- Whelan S. and Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology and Evolution*. 18(5): 691-699.
- Wolfe L.M. and Barrett S.C.H. 1989. Patterns of pollen removal and deposition in tristylous *Pontederia cordata* L. (Pontederiaceae). *Biological Journal of the Linnean Society*. 36: 317-329.
- Wong W.S.W., Yang Z., Goldman N., and Nielsen R. 2004. Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sites. *Genetics*. 168: 1041-1051.
- Woo S.H., Adachi T., Jong S.K., and Campbell C.G. 1999. Inheritance of self-compatibility and flower morphology in an inter-specific buckwheat hybrid. *Canadian Journal of Plant Science*. 79: 483-490.
- Wright S. 1939. The distribution of self-sterility alleles in populations. *Genetics*. 24: 538-552.
- Wright S.I., Agrawal N., and Bureau T.E. 2003. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Research*. 13: 1897-1903
- Wu J., Wang S., Zhang S., Publicover S.J., and Franklin-Tong V.E. 2011. Self-incompatibility in *Papaver rhoeas* activates non-specific cation conductance permeable to Ca^{2+} and K^{+} . *Plant Physiology*. 155: 963-973.
- Yang Z. and Bielawski J.P. 2000. Statistical methods for detecting molecular adaptation. *TREE*. 15(12): 496-503.
- Yang Z., Nielsen R, and Hasegawa M. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Molecular Biology and Evolution*. 15(12): 1600-1611.
- Yasui Y., Mori M., Matsumoto D., Ohnishi O., Campbell C.G., and Ota T. 2008. Construction of a BAC library for buckwheat genome research – An application to positional cloning of agriculturally valuable traits. *Genes and Genetic Systems*. 83: 393-401.
- Yasui Y., Mori M., Aii J., Abe T., Matsumoto D., Sato S., Hayashi Y., Ohnishi O., and Ota T. 2012. *S*-locus early flowering 3 is exclusively present in the genomes of short-styled buckwheat plants that exhibit heteromorphic self-incompatibility. *PLoS ONE*. 7(2): e31264.
- Yeo P.F. 1975. Some aspects of heterostyly. *New Phytologist*. 75: 147-153.

- Zhang Y., Zhao Z., and Xue Y. 2009. Roles of proteolysis and plant self-incompatibility. *Annual Review of Plant Biology*. 60: 21-42.
- Zhou W., Barrett S.C.H, Wang H., and Li D-Z. 2012. Loss of floral polymorphism in heterostylous *Luculia pinceana* (Rubiaceae): A molecular phylogeographic perspective. *Molecular Ecology*. 21(18): 4631-4645.
- Zhou W., Barrett S.C.H., Wang H., and Li D-Z. 2015. Reciprocal herkogamy promotes disassortative mating in a distylous species with intramorph compatibility. *New Phytologist*. 2015: 1-10

Appendix A: PCR and Sequencing Primers

Table A1: PCR primer pairs used for the amplification of each exon(s) of interest. The name, sequence, and melting temperature (T_m °C) for each individual primer is provided. The PCR extension time (seconds) used and approximate PCR product size (bp) for each pair is also given. Some exons were amplified with multiple primer pairs. The pair that was employed depended on the DNA sample that was used in the PCR reaction.

Gene	Exon(s) Amplified	Primer Name	Primer Sequence	T_m (°C)	Extension Time (Seconds)	PCR Product Size (bp)
<i>APETALA2</i>	1	F10	TGGGACTTGAACGACTCTCC	59.11	30	580
		R10	ATATGAGACTCCCATCGGCC	58.73		
		F11	GGA CTTGAACGACTCTCCTGA	59.03	30	570
		R9	GACTCCCATCGGCCAGTT	59.01		
<i>Tssta1</i>	1	3F	CCA ACTATTTTCATTGTGAAAGCATTTA	59.89	45	600
		1R	CCTTTCCTTTTTTCTGATATACCA	58.47		
<i>LEJ2</i>	1	F11	CTCCTTAATCCTCCACCGGG	59.24	30	445
		R16	ACCACGCGTAACTCATCTTT	57.55		
		F12	GGCTCCTCCTCCCTTTACTAC	58.96	30	420
		R16	ACCACGCGTAACTCATCTTT	57.55		
		F12	GGCTCCTCCTCCCTTTACTAC	58.96	30	400
		R15	TGGTCATAAAATCACCCACCG	58.28		
<i>AP2D</i>	1	F7	ATGGAAAGCGGGGTGGAAAA	60.18	30	620
		R10	CTCCCCATCCGACTCTTCC	58.87		
		F7	ATGGAAAGCGGGGTGGAAAA	60.18	30	600
		R9	CAGGCACCTTGTTCAAGTCA	58.32		
		F8	AAAAGAGATGGTGGCGGC	58.00	30	575
		R9	CAGGCACCTTGTTCAAGTCA	58.32		
		F4	GAAAGCGGGGTGGAAAA	60.95	30	555
		R9	CAGGCACCTTGTTCAAGTCA	58.32		
<i>RNABP</i>	1	F3	GGTGT CAGCGCATGAGAATA	59.83	30	430
		R9	TGAGTTGGGCTACGGAGATT	58.43		

Continued from previous page.

Gene	Exon(s) Amplified	Primer Name	Primer Sequence	T _m (°C)	Extension Time (Seconds)	PCR Product Size (bp)
<i>SCE1</i>	3-4	F17	ACTGATTGGGAGGGTGGTT	58.12	30	780
		R20	AGACAAGAGGAGGGTATTGCT	58.16		
		F19	TGGGAGGGTGGTTACTTTCC	58.93	30	765
		R17	GAGGGTATTGCTTGGCTTGAA	58.55		
		F20	ACTTTCCGCTTACTCTGCAC	58.20	30	755
		R18	AGGAGGGTATTGCTTGGCTT	59.00		
<i>FRA1</i>	24-25	F14	CGCAGTTCAAATAAGCACCG	58.12	30	415
		R18	CCATATCGACAAGAGTTCGAGT	57.71		
		F15	GTTCAAATAAGCACCGTGTAGAT	57.20	30	405
		R18	CCATATCGACAAGAGTTCGAGT	57.71		
		F16	CAAATAAGCACCGTGTAGATGAC	57.86	30	400
		R18	CCATATCGACAAGAGTTCGAGT	57.71		
	5-6	F6	TCGGTTCGCATGCCATATTCA	59.89	30	400
		R12	AGTTTACTGTCTCGGTAGGGA	57.55		
<i>LRRK</i>	1	F5	AAATCCCTCTTCCTCGACCG	59.18	60	1055
		R8	CGTGAATGAGATTGAAGAGGCT	58.47		
		F6	ACTTCTTCTCCGGCTCCTTC	59.10	60	1025
		R7	TGAGATTGAAGAGGCTGCCA	59.01		
		F1	TGCCTCTACTAACCAACTCAATT	57.50	60	1010
		R4	CGCAGCATTACCGTCAACTT	59.20		
		F2	CCACTCCTTCTTTTCATCTCCG	58.73	60	875
		R1	TGTTGATCAAACCGCCCC	58.65		
<i>IRX15L</i>	1	F1	CGGTAACAATAACAACACAAAGC	57.20	60	925
		R4	CAGGAAGATTTGGGTGATTTTGA	57.27		
		F2	CACAAAGCTCATCCTCCTCC	57.96	60	900
		R2	GGGTGATTTTGTATGTCCGGTTA	58.65		
		F1	CGGTAACAATAACAACACAAAGC	57.20	60	875
		R1	AGCTATTCTCATCCATCTTCTCC	57.26		

Continued from previous page.

Gene	Exon(s) Amplified	Primer Name	Primer Sequence	T _m (°C)	Extension Time (Seconds)	PCR Product Size (bp)
<i>FSP</i>	6-7	F1	GCCACTCAATCCACCAAAC	58.38	30	300
		R4	TGTCATTTTCCGGAGCATCA	57.51		
		F4	TCCACCAAACCTCAAACCACT	57.91	30	295
		R4	TGTCATTTTCCGGAGCATCA	57.51		
		F3	CTCAATCCACCAAACCTCAAACC	58.38	30	290
		R1	CGGAGCATCATGACCATTTGT	58.98		
<i>NRFP</i>	3	F5	CGGATTGCCAAACGAGGG	58.81	45	720
		R8	CCCATTGGTGAAGCTCTGG	58.82		
		F5	CGGATTGCCAAACGAGGG	58.81	45	710
		R7	CTCTGGCTACATGATTGATCTCT	57.53		
		F7	ACCTTCCGCTTGTGAAATCTC	58.57	45	630
		R5	AGTGAATCCTAACACCCAGT	58.37		
<i>WRKY</i>	1	F1	ATGGCCGTAGAGCTCATGA	58.17	60	790
		R4	CACCACACTTGCCAGAACC	58.97		
		F3	GTTATAGGAACGGTAGCTTTGTG	57.29	60	760
		R4	CACCACACTTGCCAGAACC	58.97		
<i>ECIP1</i>	3	F1	TTTCACGGTGGAGGAAGTGA	58.88	60	1000
		R1	GGCATTCCAAGGTCACGAAT	58.54		
		F2	AGGAAGTGGGGTGGAAAGTAC	58.64	60	995
		R2	GCAAGCTTCACCCACAACA	58.89		
<i>GAUT3</i>	6	F1	GCAGCTGACTATTTCCGACA	57.99	45	700
		R1	GCTTTCTCCACTCCTTCAAGTC	58.93		
		F2	CCGACATGGGTATCAAAGAAAG	57.76	45	690
		R2	TTCCGCTTTCTCCACTCCTT	58.94		
<i>GAUT1</i>	4	F1	TCCTTCCAAGCATGTTTTCCA	58.04	45	575
		R1	GCCACTTGTGATATATCCCTGTG	58.93		

Continued from previous page.

Gene	Exon(s) Amplified	Primer Name	Primer Sequence	T_m (°C)	Extension Time (Seconds)	PCR Product Size (bp)
<i>GAUT1</i>	4	F1	TCCTTCCAAGCATGTTTTCCA	58.04	45	560
		R2	TCCCTGTGATATCCTTCTTCTTC	57.12		
<i>RNABP34</i>	6	F1	AAGCTGCAAAGTTGGTGGTT	58.81	30	430
		R1	TCCTGTTTGA ACTCCCAGGG	59.23		
<i>FMO1</i>	6	F2	ACA ACTAGCAGTGATTGGGTTC	58.59	30	280
		R2	ACAAATCGGCGAAGAATCCC	58.62		
<i>MBD8</i>	3	F1	TCTTAATAATGGCAAGAACGGGA	57.89	30	355
		R1	AGTCCTTTATGTGATTTCCGAGG	58.23		
<i>UNKN</i>	1	F1	GGCTTACATCCCTCCACACA	59.38	30	240
		R1	TCTTCCACAGAAACAGGCTCA	59.23		
<i>POFUT</i>	8	F2	GCAAGCAACCCTGACAAAGA	58.97	30	315
		R2	CATTTTAGCCATGTTGCCGT	57.35		

Table A2: Sequencing primers. The name and sequence is given for each primer. PCR products for each exon were sequenced in both the forward and reverse directions, resulting in the production of at least two sequences for each individual. These sequences were then assembled in order to produce a single sequence. The anticipated size of this product is provided for each primer pair. Of course, not all sequences obtained were of the suggested length. Some exons were sequenced with multiple primer pairs. The pair that was used depended on the DNA sample from which the sequenced PCR product was amplified. For plasmid sequencing, primers M13F and M13R were provided by the G enome Qu ebec and McGill University Innovation Centre.

Gene	Exon(s) Sequenced	Primer Name	Primer Sequence	Product Size (bp)
<i>APETALA2</i>	1	F10	TGGGACTTGAACGACTCTCC	575
		R9	GACTCCCATCGGCCAGTT	
		F11	GGACTTGAACGACTCTCCTGA	570
		R9	GACTCCCATCGGCCAGTT	
<i>Tssta1</i>	1	3F	CCAACTATTTTCATTGTGAAAGCATTTA	600
		R1	CCTTTCCTTTTTTCTGATATACCA	
		1F	ATTGCGTCTTCCCATTCT	450
		1R	ATAATCCGGAGGCCAAAGTG	
<i>LEJ2*</i>	1	F12	GGCTCCTCCTCCCTTTACTAC	420
		R16	ACCACGCGTAACTCATCTTT	
		F12	GGCTCCTCCTCCCTTTACTAC	400
		R15	TGGTCATAAAATCACCCACCG	
<i>AP2D</i>	1	F8	AAAAGAGATGGTGGCGGC	575
		R9	CAGGCACCTTGTTCAAGTCA	
<i>RNABP</i>	1	F5	CGAAAACCTGGCGGCATATT	315
		R9	TGAGTTGGGCTACGGAGATT	
<i>SCE1</i>	3-4	F17	ACTGATTGGGAGGGTGTT	780
		R19	CAAGAGGAGGGTATTGCTTGG	
		F20	ACTTTCCGCTTACTCTGCAC	750
		R17	GAGGGTATTGCTTGGCTTGAA	
		F20	ACTTTCCGCTTACTCTGCAC	713
R8	TCTGCAGCATCCTGAATCAA			
<i>FRA1</i>	24-25	F16	CAAATAAGCACCGTGTAGATGAC	400
		R18	CCATATCGACAAGAGTTCGAGT	
		F15	GTTCAAATAAGCACCGTGTAGAT	400
		R17	TCGACAAGAGTTCGAGTCCT	
	5-6	F6	TCGGTCGCATGCCATATCA	400
		R12	AGTTTACTGTCTCGGTAGGGA	
		F8	CGCATGCCATATTCACCATCA	390
R11	ACTGTCTCGGTAGGGAACATG			

Continued from previous page.

Gene	Exon(s) Sequenced	Primer Name	Primer Sequence	Product Size (bp)
<i>LRRK</i>	1	F5	AAATCCCTCTTCCTCGACCG	1055
		R8	CGTGAATGAGATTGAAGAGGCT	
		F6	ACTTCTTCTCCGGCTCCTTC	1025
		R7	TGAGATTGAAGAGGCTGCCA	
		F1	TGCCTCTACTAACCAACTCAATT	940
		R2	CGGAGCAGATACGTTTACAGT	
		F7	CTCACCTCCTCGATCTCTC	925
		R6	TTTGGCCTGAAAGTAAGCGG	
		F8	TCTCCCACAACAACCTCTCC	900
		R5	AAAGTAAGCGGCCATCCTCA	
		F8	TCTCCCACAACAACCTCTCC	900
		R5	AAAGTAAGCGGCCATCCTCA	
F4	CCGCAGCACAGACAAACT	850		
R1	TGTTGATCAAAACCGCCCC			
<i>IRX15L</i>	1	F2	CACAAAGCTCATCCTCCTCC	900
		R2	GGGTGATTTTGATGTCCGGTTA	
		F2	CACAAAGCTCATCCTCCTCC	860
		R1	AGCTATTCTCATCCATCTTCTCC	
<i>FSP</i>	6-7	F1	GCCACTCAATCCACCAAACCT	295
		R3	TTTTCCGGAGCATCATGACC	
		F4	TCCACCAAACCTCAAACCACT	280
		R1	CGGAGCATCATGACCATTGT	
<i>NRFP</i>	3	F5	CGGATTGCCAAACGAGGG	720
		R8	CCCATTTGGTGAAGCTCTGG	
		F7	ACCTCCGCTTGTGAAATCTC	630
		R5	AGTGGAATCCTAACACCCAGT	
<i>WRKY</i>	1	F1	ATGGCCGTAGAGCTCATGA	790
		R4	CACCACACTTGCCAGAACC	
		F2	GGCCGTAGAGCTCATGATGA	780
		R3	ACACTTGCCAGAACCAGAT	
		F3	GTTATAGGAACGGTAGCTTTGTG	760
R4	CACCACACTTGCCAGAACC			
<i>ECIP1</i>	3	F1	TTTCACGGTGGAGGAAGTGA	1000
		R1	GGCATTCCAAGGTCACGAAT	
<i>GAUT3</i>	6	F1	GCAGCTGACTATTTCCGACA	700
		R1	GCTTTCTCCACTCCTTCAAGTC	

Continued from previous page.

Gene	Exon(s) Sequenced	Primer Name	Primer Sequence	Product Size (bp)
<i>GAUT1</i>	4	F1	TCCTTCCAAGCATGTTTTCCA	575
		R1	GCCACTTGTGATATATCCCTGTG	
<i>RNABP34</i>	6	F1	AAGCTGCAAAGTTGGTGGTT	430
		R1	TCCTGTTTGA ACTCCCAGGG	
<i>FMO1</i>	6	F2	ACA ACTAGCAGTGATTGGGTTC	280
		R2	ACAAATCGGCGAAGAATCCC	
<i>MBD8</i>	3	F1	TCTTAATAATGGCAAGAACGGGA	355
		R1	AGTCCTTTATGTGATTCCGAGG	
<i>UNKN</i>	1	F1	GGCTTACATCCCTCCACACA	240
		R1	TCTTCCACAGAAACAGGCTCA	
<i>POFUT</i>	8	F2	GCAAGCAACCCTGACAAAGA	315
		R2	CATTTTAGCCATGTTGCCGT	
Plasmids		M13F	GTAAAACGACGGCCAGT	
		M13R	GGAAACAGCTATGACCATG	

Appendix B: Nucleotide Alignments for S-linked Genes

D16L	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGA	GATGGGAGTGGC	---	-----	105
F60SS	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAG	GATGGGAGTGGC	---	-----	105
SL8 201S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
E 207S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
E 2L	AAGGGGAA	GAGGGTC	GGATCCGTGTC	SAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAA	GATGGGAGTGGC	---	-----	105
DROT 41S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGA	GATGGGAGTGGC	---	-----	105
ES	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
MIDC 710S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
MAN 601S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
MAN 713L	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGA	GATGGGAGTGGC	---	-----	105
COLO	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
PA 4S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
TSH	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAA	GATGGGAGTGGC	---	-----	105
KRAP 5S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
KRAP 12L	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAA	GATGGGAGTGGC	---	-----	105
CON 20S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
DEN 20S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAR	GATGGGAGTGGC	---	-----	105
DEN 54L	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGA	GATGGGAGTGGC	---	-----	105
TJ 30L	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAA	GATGGGAGTGGC	---	-----	105
TJ 29S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAA	GATGGGAGTGGC	---	-----	105
CHAM 4L	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGAA	GATGGGAGTGGC	---	-----	105
WED 2S	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTAAATCGAGGAT	GGATCCGAGGAAGAA	GATGGGAGTGGC	CGAC	AAGAGTAATTATTAT	105
DIF	AAGGGGAA	GAGGGTC	GGATCCGTGTC	GAAAT	TCTAGCTCCTCCGCC	GTTGTCTCGTAGGAT	GGATCCGAGGAAGAA	GATGGGAGTGGC	---	-----	105
PAN 2S	AAGGGGAA	AAGGGTC	GGATCCGTGTC	GAAAT	TCCAGCTCCTCCGCC	GTGGTCATCGAGGAT	GGATCCGAGGAAGA	GATGGGAGTGGC	CGAC	AAGAGTAATTAT---	105
SELECTION											
D16L	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	GCCATGGACAGTGAT	210			
F60SS	-----GAA	GCAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	ACCATGGACGGTGAT	210			
SL8 201S	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	GCCATGGACAGTGAT	210			
E 207S	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	ACTATGGACAGTGAT	210			
E 2L	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	RCYATGGACAGTGAT	210			
DROT 41S	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	GCCATGGACAGTGAT	210			
ES	-----GAAGSA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGY	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	RCCATGGACRGTGAT	210			
MIDC 710S	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	GCCATGGACAGTGAT	210			
MAN 601S	-----GAA	GSAGGAGGAAGATTG	CTCATCAAGAAGCGY	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	RCCATGGACRGTGAT	210			
MAN 713L	---GAAGGAGGAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGC	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	GCCATGGACAGTGAT	210			
COLO	-----GAA	GSAGGAGGARGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	ACCATGGACGGTGAT	210			
PA 4S	-----GAA	GSAGGAGGARATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	ACCATGGACGGTGAT	210			
TSH	-----GAAGGA	GGAGGAGGAWKAYTS	MTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	ACTATGGACAGTGAT	210			
KRAP 5S	-----GAAGSA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGC	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	RCCATGGACRGTGAT	210			
KRAP 12L	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGC	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	GCCATGGACAGTGAT	210			
CON 20S	-----GAAGSA	GGAGGARGAAGATTG	CTCAYCAAGAAGCGY	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	RCCATGGACRGTGAT	210			
DEN 20S	-----GAAGSA	GGAGGARGAAGATTG	CTCAYCAAGAAGCGY	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	RCCATGGACRGTGAT	210			
DEN 54L	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGY	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	RCCATGGACAGTGAT	210			
TJ 30L	-----GAAGGAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGY	ACCRATGGACAGTGAT	210			
TJ 29S	-----GAAGGAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGY	ACCRATGGACAGTGAT	210			
CHAM 4L	-----GAAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	ACCATGGACAGTGAT	210			
WED 2S	ACTGGAGGAGGAGTA	GGAGGAGGGAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	GCCATGGACAGTGAT	210			
DIF	-----GGAGGA	GGAGGAGGAAGATTG	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	ACCATGGACAGTGAT	210			
PAN 2S	ACTGGAGGAGGA---	---GGAGGAAGATT	CTCATCAAGAAGCGT	AGCAGCAGAATATTC	GGCTTCTCGGTGCC	TATGAGGAGGACAGT	ACCATGAAATAGTGAT	210			
SELECTION											

D16L	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---AGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGTGGG	315
F60SS	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTAGACAATCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
SL8 201S	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCYGTGCTGRGTRTKY	GGYGGTGSTGMTGRT	RGTRGWWGGWGGK---	315
E 207S	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCTGGTTGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
E 2L	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCTGRGTRKY	GGYGGTGSTGMTGRT	RGTRGWWGGWGGK---	315
DR0T 41S	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---AGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGTGGG	315
ES	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACRACYCCCGAA	ATGGGGGCCAYGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
MIDC 710S	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---AGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGTGGG	315
MAN 601S	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTRGACRACYCCCGAA	ATGGGGGCCAYGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
MAN 713L	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
COLO	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTAGACAATCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
PA 4S	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTAGACAATCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGMTGRT	RGTRGWWGGWGGK---	315
TSH	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACSTCC	GCTGCTGCTGGTTGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
KRAP 5S	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTRGACRACYCCMGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
KRAP 12L	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
CON 20S	GAGCCGCCAGYRACA	CGGCAGTTCTTTCCC	GTRGACRACYCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
DEN 20S	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTRGAMRAYCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
DEN 54L	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGAMGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
TJ 30L	GAGCCGCCAGTGACR	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
TJ 29S	GAGCCGCCAGTGACR	CGGCAGTTCTTTCCC	GTGGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTGCTGAT	GGTAGTGGAGGT---	315
CHAM 4L	GAGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTAGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTACTGAT	GGTRGTTGGAGGA---	315
WED 2S	GAGCCGCCAGTGACA	CGACAGTTCTTTCCC	GTAGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGTGGTGGTGGTGGT	GGTAGTGGAGGA---	315
DIF	GAACCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTAGACGACCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTACTGAT	GGTGGTGGAGGA---	315
PAN 2S	GAGGCCGCCAGTGACA	CGGCAGTTCTTTCCC	GTGGACGATCCCGAA	ATGGGGGCCACGTCC	GCTGCTGCT---TGT	GGCGGTGTTACTGAC	KTAGTATATA-----	315
SELECTION								

D16L	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCACCA	420
F60SS	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTAGAA	CTTTTCATCGCTTGCT	TCCCATCAGAAAACCC	ATCGAGGTCTCACCA	420
SL8 201S	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTWGAA	MTTTCATCGCKTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCACCA	420
E 207S	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCACCA	420
E 2L	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCACCA	420
DR0T 41S	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCACCA	420
ES	GGWGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTWGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCACCA	420
MIDC 710S	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTWGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCACCA	420
MAN 601S	GGWGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTWGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCACCA	420
MAN 713L	GGTGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCACCA	420
COLO	GGWGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTWGAA	MTTTCATCGCTTGCT	TCCCATMAAAAACCC	RTCGAGGTCTCACCA	420
PA 4S	GGWGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTWGAA	MTTTCATCGCTTGCT	TCCCATMAAAAACCC	RTCGAGGTCTCACCA	420
TSH	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCACCA	420
KRAP 5S	GGWGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCWCCA	420
KRAP 12L	GGTGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCWCCA	420
CON 20S	GGWGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCACCA	420
DEN 20S	GGWGGMGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	RTCGAGGTCTCACCA	420
DEN 54L	GGWGGMGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	MTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCACCA	420
TJ 30L	GGAGGAGGAGCTGST	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCACCA	420
TJ 29S	GGAGGAGGAGCTGST	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCACCA	420
CHAM 4L	CGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCGCCA	420
WED 2S	GGAGCTGGGGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAA	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCT	GTCGAGGTCTCTCCA	420
DIF	CGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTCGCT	TCCCATCAAAAACCC	GTCGAGGTCTCGCCA	420
PAN 2S	GGAGGAGGAGCTGGT	TTTCCTAGGGCTCAC	TGGGTTGGAGTCAAG	TTTTGCCAACTGAA	ATTTTCATCGCTTGCT	TCCCATCAAAAACCC	GTCGAGGTCTCACCA	420
SELECTION								

D16L	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
F60SS	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
SL8 201S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
E 207S	<u>TCACA</u> <u>RCC</u> <u>GTGAAG</u> <u>AAGAGCCGG</u> 444
E 2L	<u>TCACA</u> <u>RCC</u> <u>GTGAAG</u> <u>AAGAGCCGG</u> 444
DROT 41S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
ES	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
MIDC 710S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
MAN 601S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
MAN 713L	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
COLO	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
PA 4S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
TSH	<u>TCACA</u> <u>GCC</u> <u>GTGAAG</u> <u>AAGAGCCGG</u> 444
KRAP 5S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
KRAP 12L	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
CON 20S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
DEN 20S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
DEN 54L	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
TJ 30L	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
TJ 29S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
CHAM 4L	<u>TCACA</u> <u>CC</u> <u>R</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
WED 2S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGA</u> 444
DIF	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
PAN 2S	<u>TCACA</u> <u>CCG</u> <u>TTGAAG</u> <u>AAGAGCCGG</u> 444
SELECTION	

Figure B1: *APETALA2* nucleotide alignment. Alignment of DNA sequences for exon 1 of *APETALA2* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. Sequences for the individuals, TJ 29S and TJ 30L were found to be identical at this locus. Sites that show evidence of trans-specific evolution in the *Turnera* subseries are underlined and starred in red. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The length of the alignment (bp) is given in the right-most column (444 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

F60SS	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
SL8 201S	ATGTCATCCCCTACC	GGAGGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
E 207S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	GTCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTA	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
DROT 41S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
ES	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
MIDC 710S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
MAN 601S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	GTCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
COLO	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
PA 4S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
TSH	ATGTCATCCCCTACC	GGACGGGAGCTTACC	GTCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
KRAP 5S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
CON 20S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
DEN 20S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
GRAN 9S	-----	-----	-----	-----	-----	-----	TCCTTGAAGTTTGAT	105
TJ 29S	ATGTCATCCCCTACC	GGACGGGTGGCTTACC	ATC-----TATCTC	TTTTCTATTATTACA	ATCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
WED 2S	-----	-----	-----	-----	-----	---TCAGCCAATAGC	TCCTTGAAGTTTGAT	105
PAN 2S	ATGTCATCCCCTTCG	GGCGGTTGCCCTTACC	-----CTCTATCTC	TTTTCTATTATTACA	ATATGCCCATTTCTT	GTCTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
CUN	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTACTGTG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
OCC	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTACTGTG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
ORI	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTACTGTG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
TOC 139S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTACTGTG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
QUACO1	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTACTGTG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
VEL	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTACTGTG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
AUR	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTACTGTG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
BAH	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTACTGTG	GTCTTCCCATTTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
PCARO	ATGCCATCCTCTACA	TGGCTGGTGCCTAAC	ATC-----TATCTT	TTTTCTATTATTACA	ATCTTCCCATTTCTT	GCATCAGCCAATAAC	TCCTTGAAGTTTGAT	105
PMOR 137S	ATGCCATCCTCTACG	TGGCTGGTGCCTAAC	ATC-----TATCTT	TTTTCTATTATTACA	ATCTTCCCATTTCTT	GCATCAGCCAATAAC	TCCTTGAAGTTTGAT	105
VIS	-----	-----	-----	-----	-----	-----	TCCTTGAAGTTTGAT	105
PNAN	-----	-----	-----	-----	-----	---TCAGCCAATAGC	TCCTTGAAGTTTGAT	105
PSAR	-----	-----	-----	-----	-----	---TCAGCCAATAGC	TCCTTGAAGTTTGAT	105
PREV	-----	-----	-----	-----	-----	-----	TCCTTGAAGTTTGAT	105
PLIC	-----	-----	-----	-----	-----	-----GCCAATAGC	TCCTTGAAGTTTGAT	105
PDUART 1S	-----	-----	-----	-----	-----	-----GCCAATAGC	TCCTTGAAGTTTGAT	105
EOD	-----	-----	-----	-----	-----	---TCAGCCAATAGC	TCCTTGAAGTTTGAT	105
SELECTION								

F60SS	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCGAGCAATGAC	GATCTAGGGCACCAT	210
SL8 201S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCGAGCAATGAC	GATCTGGGGCACCAT	210
E 207S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
DROT 41S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
ES	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCGAGCAATGAC	GATCTGGGGCACCAT	210
MIDC 710S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
MAN 601S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCGAGCAATGAC	GATCTGGGGCACCAT	210
COLO	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
PA 4S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
TSH	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
KRAP 5S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
CON 20S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
DEN 20S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
GRAN 9S	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
TJ 29S	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGTAAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
WED 2S	CCTTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGCTACTTCATTGT	TGGTCAAGTGATGAC	GACCTGGGGCACCAT	210
PAN 2S	CTCTTTGGCAACTTA	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTTTTACTTCATTGT	TGGTCAAGTGATGAC	GACCTGGGGCACCAT	210
CUN	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
OCC	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
ORI	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
TOC 139S	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
QUACO1	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
VEL	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
AUR	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
BAH	CTCTTTGGCAACATG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTGGGGCACCAT	210
PCARO	CCCTTTGGCAACTCT	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTACTTCATTGC	TGGTCAAGTGATGAT	GCCCTAGGTCAACAT	210
PMOR 137S	CCCTTTGGCAACTTG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTAATTCATTGC	TGGTCAAGTGATGAT	GACCTGGGGCACCAT	210
VIS	CCCTTTGGCAACTTG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTAATTCATTGC	TGGTCAAGTGATGAT	GACCTGGGGCACCAT	210
PNAN	CCCTTTGGCAACGTG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTAATTCATTGC	TGGTCAAGTGATGAT	GACCTGGGTCAACAT	210
PSAR	CCCTTTGGCAACGTG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTAATTCATTGC	TGGTCAAGTGATGAT	GACCTGGGTCAACAT	210
PREV	CCCTTTGGCAACGTG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTAATTCATTGC	TGGTCAAGTGATGAT	GACCTGGGTCAACAT	210
PLIC	CCCTTTGGCAACTTG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTAATTCATTGC	TGGTCAAGTGATGAT	GACCTGGGGCACCAT	210
PDUART 1S	CCCTTTGGCAACTTG	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTAATTCATTGC	TGGTCAAGTGATGAT	GACCTGGGGCACCAT	210
EOD	CCCTTTGGCAACAA	TACAGAGTTCATGTC	ATCAATGGCTTCAGC	AGCAATGACCAACCA	TTGTTACTTCATTGC	TGGTCAATGATGAC	GATCTGGGGCACCAT	210
SELECTION								



F60SS	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
SL8 201S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
E 207S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
DROT 41S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
ES	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
MIDC 710S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
MAN 601S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
COLO	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
PA 4S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
TSH	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
KRAP 5S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
CON 20S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
DEN 20S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
GRAN 9S	AGTCTCTACATTGGG	GGCGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
TJ 29S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
WED 2S	AGCCTCTACATTGGT	GGAGACTTTAGCTTC	CACTTTGGCCTCCGG	ATTATACCTCCCGCT	ACCCGTTTCTGGTGT	GACATGAA	CAGGGGC	CCAAAATATATTCCCT	315
PAN 2S	AGCCTCTACATTGGC	GGAGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACACGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCC	315
CUN	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
OCC	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
ORI	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
TOC 139S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCATTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
QUACO1	AGTCTCTACAATGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCATTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
VEL	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCATTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
AUR	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
BAH	AGTCTATACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCTCCGG	ATTATACCTCCCTCT	ACCCATTTCTGGTGT	GACATGAA	CCGCGGC	CCAAAATATATTCCCT	315
PCARO	AGCCTCTATATTGGC	GAAGACTTTCACTTC	CACTTCGGCCTCAGG	ATTATACCTCCTTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGGGGC	CCAAAATATATTCCCT	315
PMOR 137S	AGCCTCTACATTGGA	GAAGACTTTAACTTC	CAGTTCCGGCTCAGG	ATTATACCTCCGTCA	ACCCATTTCTGGTGT	GACATGAA	CCAGGGC	CGAAAATATATTCCCT	315
VIS	AGCCTCTACATTGGA	GAAGACTTTAACTTC	CACTTCGGCCTCAGG	ATTATACCTCCGTCA	ACCCATTTCTGGTGT	GACATGAA	CCGGGGC	CGAAAATATATTCCCT	315
PNAN	ACACTCTACATTGGC	GAAGACTTTAACTTC	CACTTCGGCCTCAGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGGGGC	CCAAAATATATTCCCT	315
PSAR	ACACTCTACATTGGC	GAAGACTTTAACTTC	CACTTCGGCCTCAGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGGGGC	CCAAAATATATTCCCT	315
PREV	ACACTCTACATTGGC	GAAGACTTTAACTTC	CACTTCGGCCTCAGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAA	CCGGGGC	CCAAAATATATTCCCT	315
PLIC	AGCCTCTACATTGGA	GAAGACTTTAACTTC	CACTTCGGCCTCAGG	ATTATACCTCCGTCA	ACCCATTTCTGGTGT	GACATGAA	CCGGGGC	CGAAAATATATTCCCT	315
PDUART 1S	AGCCTCTACATTGGA	GAAGACTTTAACTTC	CACTTCGGCCTCAGG	ATTATACCTCCGTCA	ACCCATTTCTGGTGT	GACATGAA	CCGGGGC	CGAAAATATATTCCCT	315
EOD	AACCTCTGCCTTGGC	GGAGACTTTCAGTTTC	CACTTTGGCCTCAGG	ATTATACCTCCCTCT	ACGCATTTCTGGTGC	GACATGAA	CCGGGGC	CCAAAATATCTTCAT	315
SELECTION									

F60SS	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
SL8 201S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
E 207S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
DROT 41S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
ES	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
MIDC 710S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
MAN 601S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
COLO	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
PA 4S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
TSH	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
KRAP 5S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
CON 20S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
DEN 20S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
GRAN 9S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
TJ 29S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
WED 2S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
PAN 2S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	AAGACGAAGCAATGT	TATTGGCGTGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
CUN	CGAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
OCC	CGAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
ORI	CGAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
TOC 139S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
QUACO1	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
VEL	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
AUR	CGAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
BAH	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
PCARO	CAAGTTAGTGTGTTT	GACGAGGATGAGGTG	TTGCATTTGTGCAGC	CAGACGCAGCAATGT	TATTGGCGAGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
PMOR 137S	CAAGTTTCTGTGTTT	GAAGAAGATGAGGTG	TTGCATTTGTGCAGC	CATACACAGCAATGT	TATTGGCGAGGGCAA	GATAACGGACTCTAT	TTCTCTAATGATAAC	420
VIS	CAAGTTTCTGTGTTT	GAAGAATATGAGGTG	TTGCATTTGTGCAGC	CATACGCAGCAATGT	TATTGGCGAGGGCAA	GATAAGGACTCTAT	TTCTCTAATGATAAC	420
PNAN	CAAGTTACTGTGTTT	GAAGAGGATGAGGTG	TTGCATTTGTGTAGC	CACACGCAGCAATGT	TATTGGCGAGGGCAA	GATAACGGACTGTAT	TTCTGCAATGATAAC	420
PSAR	CAAGTTACTGTGTTT	GAAGAGGATGAGGTG	TTGCATTTGTGTAGC	CACACGCAGCAATGT	TATTGGCGAGGGCAA	GATAACGGACTGTAT	TTCTGCAATGATAAC	420
PREV	CAAGTTACTGTGTTT	GAAGAGGATGAGGTG	TTGCATTTGTGTAGC	CACACGCAGCAATGT	TATTGGCGAGGGCAA	GATAACGGACTGTAT	TTCTGCAATGATAAC	420
PLIC	CAAGTTTCTGTGTTT	GAAGAAGATGAGGTG	TTGCATTTGTGCAGC	CATACACAGCAATGT	TATTGGCGAGGGCAA	GATGACGACTCTAT	TTCTCTAATGATAAC	420
PDUART 1S	CAAGTTTCTGTGTTT	GAAGAAGATGAGGTG	TTGCATTTGTGCAGC	CATACACAGCAATGT	TATTGGCGAGGGCAA	GATAACGGACTCTAT	TTCTCTAATGATAAC	420
EOD	CAAGTTACTGTTT	GACGAGGATGAGGTG	TTGCATTTGTGCAGC	CGCACCCAGCGATGT	TACTGGCGGGGGCAA	TACGACGGACTCTAT	TTTTCCAATGATAAC	420
SELECTION								

MIDC 710S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
MAN 601S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
COLO	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PA 4S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
TSH	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
KRAP 5S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
CON 20S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
DEN 20S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
GRAN 9S	<u>TCCTCCTATTTCAAG</u>	TTG-----	447
TJ 29S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
WED 2S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PAN 2S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
CUN	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
OCC	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
ORI	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
TOC 139S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
QUACO1	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
VEL	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
AUR	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
BAH	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PCARO	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PMOR 137S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
VIS	<u>TCCTCCTATTTCAAG</u>	TTG-----	447
PNAN	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PSAR	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PREV	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PLIC	<u>TTCTCCTATTTCAAG</u>	TTG-----	447
PDUART 1S	<u>TCCTCCTATTTCAAG</u>	TTG-----	447
EOD	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
SELECTION			



Figure B2: *Tsstal* nucleotide alignment. Alignment of DNA sequences for *Tsstal* obtained from 34 individuals from the genus, *Turnera*. Because *Tsstal* is only represented on the dominant *S*-haplotype, sequences were only obtained from short-styled and short-homostyled individuals. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) MIDC 710S, DROT 41S, COLO, and PA 4S; 2) DEN 20S and CON 20S; 3) MAN601S and ES; 4) OCC, CUN, and ORI; and 5) PSAR and PNAN. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to plant self-incompatibility domain family S1 according to BLAST searches is indicated by arrows at the base of the alignment (↑). The length of the alignment (bp) is given in the right-most column (447 bp, total, representing almost the entire gene). The alignment is sectioned into groups of 15 bases (or 5 codons).

F60SS	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
SL8 201S	ATGTCATCCCCTACC	GGAGGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
E 207S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	GTCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
DROT 41S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
ES	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
MIDC 710S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
MAN 601S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	GTCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
COLO	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
PA 4S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
TSH	ATGTCATCCCCTACC	GGACGGGAGCTTACC	GTCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
KRAP 5S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
CON 20S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
DEN 20S	ATGTCATCCCCTACC	GGACGGGAGCTTACC	ATCTATCTCTATCTC	TTTTCTATTATTGCG	GTCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
TJ 29S	ATGTCATCCCCTACC	GGACGGGTGGCTACC	ATC-----TATCTC	TTTTCTATTATTACA	ATCTTCCCATTCTT	ACTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
WED 2S	-----	-----	-----	-----	-----	---TCAGCCAATAGC	TCCTTGAAGTTTGAT	105
PAN 2S	ATGTCATCCCCTTCG	GGCGGTTGCCTACC	-----CTCTATCTC	TTTTCTATTATTACA	ATATGCCCATTTCTT	GCTTCAGCCAGTAGC	TCCTTGAAGTTTGAT	105
SELECTION								

F60SS	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCGAGCAATGAC	GATCTAGGGCACCAT	210
SL8 201S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCGAGCAATGAC	GATCTAGGGCACCAT	210
E 207S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
DROT 41S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
ES	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCGAGCAATGAC	GATCTAGGGCACCAT	210
MIDC 710S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
MAN 601S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
COLO	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
PA 4S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
TSH	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
KRAP 5S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
CON 20S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
DEN 20S	CTATTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
TJ 29S	CTCTTCGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGTAATGACCTGCCA	TTTTTACTTCATTGT	TGGTCAAGCAATGAC	GATCTAGGGCACCAT	210
WED 2S	CCTTTTGGCAACATG	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCAACCA	TTGCTACTTCATTGT	TGGTCAAGTGATGAC	GACCTAGGGCACCAT	210
PAN 2S	CTCTTCGGCAACTTA	TACAGAGTTCATGTC	ATCAAATGGCTTCAGC	AGCAATGACCAAGCCA	TTTTTACTTCATTGT	TGGTCAAGTGATAAC	GACCTAGGGCACCAT	210
SELECTION	-	↑	-	-	-	-	-	-

F60SS	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
SL8 201S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
E 207S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
DROT 41S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
ES	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
MIDC 710S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
MAN 601S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
COLO	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
PA 4S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
TSH	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
KRAP 5S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
CON 20S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
DEN 20S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
TJ 29S	AGTCTCTACATTGGC	GGTGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACCGCGGC	CCAAAATATATTCCCT	315
WED 2S	AAGCTCTACATTGGT	GGAGACTTTAGCTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACCCGTTTCTGGTGT	GACATGAACAGGGGC	CCAAAATATATTCCCT	315
PAN 2S	AGCCTCTACATTGGC	GGAGACTTTAACTTC	CACTTTGGCCCTCCGG	ATTATACCTCCCTCT	ACACGTTTCTGGTGT	GACATGAACCGGGGC	CCAAAATATATTCCCT	315
SELECTION								

F60SS	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
SL8 201S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
E 207S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
DROT 41S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
ES	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
MIDC 710S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
MAN 601S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
COLO	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
PA 4S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	CTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
TSH	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
KRAP 5S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
CON 20S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
DEN 20S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
TJ 29S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAGCAATGC	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
WED 2S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	CACACGCAACAATGT	TATTGGCGGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
PAN 2S	CAAGTTAGTGTGTTT	GAGGAGGATGAAGTG	TTGCATTTGTGCAGC	AAGACGAGCAATGT	TATTGGCGTGGGCAA	GATAACGGACTCTAT	TTCTGCAATGATAAC	420
SELECTION								

F60SS	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
SL8 201S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
E 207S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
DROT 41S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
ES	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
MIDC 710S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
MAN 601S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
COLO	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PA 4S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
TSH	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
KRAP 5S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
CON 20S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
DEN 20S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
TJ 29S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
WED 2S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
PAN 2S	<u>TCCTCCTATTTCAAG</u>	<u>TTGTATGACTGG</u>	447
SELECTION			



Figure B3: *Tsstal* nucleotide alignment with a reduced number of taxa. Alignment of DNA sequences for *Tsstal* obtained from 16 individuals from the genus, *Turnera*. Because *Tsstal* is only represented on the dominant *S*-haplotype, sequences were only obtained from short-styled and short-homostyled individuals. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) MIDC 710S, DROT 41S, COLO, and PA 4S; 2) CON 20S and DEN 20S; and 3) MAN601S and ES. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to plant self-incompatibility domain family S1 according to BLAST searches is indicated by arrows at the base of the alignment (↑). The length of the alignment (bp) is given in the right-most column (447 bp, total, representing almost the entire gene). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	ACGCCCCGCCGCTCC	CGCCGCGTCTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
F60SS	ACGCTCCGCCGCTCC	CGTGCCATCTCCTT	GTGCCCCAGCTGCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
SL8 201S	ACGCCCCGCCGCTCC	CGCCGCGTCTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
E 207S	ACGCYCCGCCGCTCC	CGTGCCRWCTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	YCCGCCATCCCTCC	105	
E 2L	ACGCYCCGCCGCTCC	CGTGCCRWCTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
DROT 41S	-----	-----	-----	---TCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
ES	ACGCCCCGCCGCTCC	CGTGCCGTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
MIDC 710S	-----	-----	-----	---TCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
MAN 601S	ACGCCCCGCCGCTCC	CGTGCCGTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
MAN 713L	ACGCCCCGCCGCTCC	CGTGCCGTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
COLO	-----	-----	-----	---TCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
PA S4	-----	-----	-----	---TCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
TSH	ACGCTCCGCCGCTCC	CGTGCCATCTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
KRAP 5S	ACGCTCCGCCGCTCC	CGTGCCATCTCCTT	GTGCCCCAGCTGCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTTC	105	
KRAP 12L	-----	-----	-----	---TCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
CON 20S	ACGCTCCGCCGCTCC	CGTGCCGTCCTT	GTCCCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
DEN 20S	-----	-----	-----	-----	-----	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
DEN 54L	-----	-----	-----	---TCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
TJ 30L	ACGGCCGCCGCTCC	CGTGCCGTCCTT	GTCTCCCAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCG	105	
TJ 29S	-----	-----	-----	-----	---CTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCG	105
CHAM 4L	RCGCWCCGCCGCTCC	CGCCACCRTCTCCTT	GTCTCCYAGCTCCC	CTGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	-----TCC	105	
WED 2S	ACGCTCCGCCGCTCC	CGCCACCATATCCGT	GGGCCCCAGCTCCC	ATGTCTCTC	---TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	TCCGCCATCCCTCC	105	
PAN 2S	CCGCTCCGCCGACC	CGTGCCATCTCCTT	GTGCCCCAGCTCCC	CTGTATCTCT	TCC	TCCTCCCCGGACCGG	AGGCTCTTTCTCTA	CAGGCCATCCCTCC	105	
SELECTION			+							

D16L	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
F60SS	<u>CCCCGGAGGC</u> TCTCA	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
SL8 201S	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
E 207S	<u>CCCCGGAGGC</u> CCTCC	RCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
E 2L	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
DROT 41S	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
ES	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
MIDC 710S	<u>CCCCGGAGGC</u> YCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
MAN 601S	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
MAN 713L	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
COLO	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
PA S4	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
TSH	<u>CCCCGGAGGC</u> CCTCC	RCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
KRAP 5S	<u>CCCCGGAGGC</u> CTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
KRAP 12L	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
CON 20S	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
DEN 20S	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
DEN 54L	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
TJ 30L	<u>CCCCGGAGGC</u> CCTCC	GCCGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
TJ 29S	<u>CCCCGGAGGC</u> CCTCC	GCCGTGGCAGCTGGC	GGGACCTTGATTACC	AATTCG	156
CHAM 4L	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGAC	GGGACCTTGATGACC	AATTCG	156
WED 2S	<u>CCCCGGAGGC</u> CTCC	GCCGTGGCAGCTGGC	GGGACCTTGATGACC	AATTCG	156
PAN 2S	<u>CCCCGGAGGC</u> CCTCC	GCTGTGGCAGCTGGC	GGGACCTTGACGACC	AATTCG	156
SELECTION					

Figure B4: *LEJ2* nucleotide alignment. Alignment of DNA sequences for exon 1 of *LEJ2* obtained from 23 individuals from the genus, *Turnera*. No sequence information was obtained for the individual, DIF (*T. diffusa*). Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) SL8 201S and D16L; 2) MAN 601S, MAN 713L, and ES; and 3) COLO, DROT 41S, and PA 4S. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The length of the alignment (bp) is given in the right-most column (156 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	GAAAGCGGGTGGAA	AAAGAGATG	GTTGGCG	GCGGCCACAAGGAAG	AGAGCAACA	---ACA	GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105		
F60SS	-----	-----	GTGGCG	GCGGCCACAAGGAAG	AGAAACAACAGCAGCA		GGAGGAGGGGGCATG	GATAACGACAGGCC	TAC	AAGGGGATAAGG	105		
SL8 201S	GRAAGCGGGTGGAA	AAAGAGATC	GTTGGCG	GCGGCCACAAGGAAG	AGAACAACAGCAGCA		GGAGGAGGGGGCWTR	GAWAACGACAGGCC	TAC	AAGGGGATAAGG	105		
E 207S	-----	-----	-----	-----	-----	-----	GCA	GGAGGAGGGGGCATG	GATAACGACAGGCC	TAC	AAGGGGATAAGG	105	
E 2L	GAAAGCGGGTGGAA	AAAGAGATG	GTTGGCG	GCGGCCACAAGGAAG	AGAAACAACAGCTGCA		GGAGGAGGGGGCATG	GATAACGACAGGCC	TAC	AAGGGGATAAGG	105		
DROT 41S	GAARGCGGGTGGAA	AAAGAGATC	GTTGGCG	GCGGCCACAAGGAAG	AGAAACAACAGCAGCA		GGAGGAGGGGGCATG	GATAACGACAGGCC	TRC	AAGGGGATAAGG	105		
ES	GAAAGCGGGTGGAA	AAAGRGAT	S	GTTGGCG	GCGGCCACAAGGAAG	AGARCAACAGCAGCA		GGAGGAGGGGGCATG	GAWAACGACAGGCC	TAC	AAGGGGATAAGG	105	
MIDC 710S	GAAAGCGGGTGGAA	AAAGAGAT	S	GTTGGCG	GCGGCCACAAGGAAG	AGAAACAACAGCAGCA		GGAGGAGGGGGCATG	GATAACGACAGGCC	TAC	AAGGGGATAAGG	105	
MAN 601S	GAAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAGCAACA	---ACA	GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
MAN 713L	---AGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAGCAACA	---ACA	GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
COLO	-----	GAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAGCAACA	---ACA	GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105
PA 4S	GAAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAGCAACA	---ACA	GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
TSH	GAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAACAACAGCTGCA		GGAGGAGGGGGCATG	GATAACGACAGGCC	TAC	AAGGGGATAAGG	105	
KRAP 5S	GAAAGCAGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGARCAACAGCAGCA		GGAGGAGGGGGCATG	GATAACGACAGGCC	TAC	AAGGGGATAAGG	105	
KRAP 12L	GAAAGCGGGTGGAA	AAAGAGAT	S	GTTGGCG	GCGGCCACAAGGAAG	AGARCAACAGCAGCA		GGAGGAGGGGGCATG	GATAACGACAGGCC	TAC	AAGGGGATAAGG	105	
CON 20S	GAAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAGCAACA	---ACA	GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
DEN 20S	GAAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGGGCAACA	---ACA	GTAGGAGGGGGAATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
DEN 54L	GAAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGGGCAACA	---ACA	GTAGGAGGGGGAATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
TJ 30L	GAAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAACGACAGCAGGA		GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
TJ 29S	GAAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAACGACAGCAGGA		GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
CHAM 4L	GAAAGCGGGTGGAA	AAAAAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAGCAACA	---ACA	GGAGGAGGGGGCATG	GAAAACGACAGGCC	TAC	AAGGGGATAAGG	105	
WED 2S	GAAAGCGGGTGGAA	AAGGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAACAACA	-----	GGAGGAGGG---ATG	GAAAACGACAGGCC	TAT	AAGGGGATAAGG	105	
DIF	-----	GTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAACAACAGCAGCA	GGAGGAGGG---TTG	GAAAACGACAGGCC	TAT	AAGGGGATAAGG	105	
PAN 2S	GAAAGCGGGTGGAA	AAAGAGAT	G	GTTGGCG	GCGGCCACAAGGAAG	AGAACAACA	---GCA	GGAGGAGGGGGGATG	GAAAACGACAGGCC	TAT	AAGGGGATAAGG	105	
SELECTION							-	+					

D16L	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
F60SS	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
SL8 201S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
E 207S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
E 2L	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
DROT 41S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
ES	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
MIDC 710S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
MAN 601S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
MAN 713L	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
COLO	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
PA 4S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
TSH	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
KRAP 5S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
KRAP 12L	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
CON 20S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
DEN 20S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
DEN 54L	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
TJ 30L	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
TJ 29S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
CHAM 4L	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
WED 2S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
DIF	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCGGCAAGGGCCTAT	210
PAN 2S	ATGAGGAAGTGGGC	AAGTGGTGGCT	SAG	ATTAGGSAACCCAAC	AAAAGGTCGAGGATT	TGGCTCGGTCTTAT	TCCACTCCGGTGGCG	GCRGCGGGCCTAT	210
SELECTION							-	-	

D16L	CAACACCAC---AAC	CATAACCACCACCAC	CGCCGCAACCACCAG	GAGGAGCAGTATAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
F60SS	CAACACCAC---AAC	CTTAACCACCACCAC	CGC---CACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
SL8 201S	CAACAMCAC---AAC	CWTAACCACCACCAC	CGC---CACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
E 207S	CACCACCAC---MAY	CATMGCCGCCACCAC	CRC---CATCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
E 2L	CACCACCAC---AAC	CATAGCCGCCACCAC	CAC---CATCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
DROT 41S	CAACACCAC---AAC	CTTAACCACCACCAC	CGC---CACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
ES	CAACACCAC---AAC	CWTAACCACCACCAC	CGC---CACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
MIDC 710S	CAACACCAC---AAC	CTTAACCACCACCAC	CGC---CACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
MAN 601S	-----	-----	-----	-----	-----	-----	-----	-----	525
MAN 713L	CACCACCAC---AAC	CATAACCACCACCAC	CGCCGCAACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
COLO	CAACACCAC---AAC	CATAACCACCACCAC	CGCCGCCACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
PA 4S	CAACACCAC---AAC	CATAACCACCACCAC	CGCCGCCACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
TSH	CAACACCAC---AAC	CATAGCCGCCGCCAC	CAC---CATCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
KRAP 5S	CAACATCAC---AAC	CATAACCACCAC---	-----CACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
KRAP 12L	CAACAWCAC---AAC	CWTAACCACCACCAC	---CGCCGCCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
CON 20S	CAACACCAC---AAC	CTTAACCACCAC---	-----CACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
DEN 20S	CAACACCAC---AAC	CATAACCACCACCAC	-----CACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
DEN 54L	CAACACCAC---AAC	CATAACCACCAC---	-----CACCAG	GAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
TJ 30L	CRACACCACGACAGC	CATAACCATCAC---	-----CACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCAAGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
TJ 29S	CGACACCACGACAGC	CATAACCATCAC---	-----CACCACCAG	GAGGAGCAGTATAAT	CAAAATAGTCAAGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
CHAM 4L	-----	-----	-----	-----	-----	-----	-----	-----	525
WED 2S	CACCACCAC---AGC	CATCACCACCACCAC	-----	CAGGAGCAGTATAAT	CAAAATAGTCACGGG	GGT---GGT	AGTAGT	AGTAGTGATAATAGT	525
DIF	CAC-----AGC	CATAACATCACCAC	CAC-----	CAGGAGCAGTATAAT	CAAAATATTCATGGG	GGTGGTAGT	AGTAGT	AGTAGTGATAATAGT	525
PAN 2S	CAC-----AGT	CGTAACCACCACCAC	CATCATCACCACAA	CAAGAGCAGTATAT	CAAAATAGTCACGGG	GGT-----	AGTAGT	AGTAGTGATAATAGT	525
SELECTION									

D16L	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	---TCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAG</u> ---	621
F60SS	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
SL8 201S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
E 207S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GATACGACAGAG---	-----	-----	-----	-----	621
E 2L	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
DROT 41S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
ES	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	---TCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	YGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
MIDC 710S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
MAN 601S	-----	-----	-----	-----	-----	-----	-----	621
MAN 713L	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GAT-----	-----	-----	-----	-----	621
COLO	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	-----	621
PA 4S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	---TCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
TSH	<u>ATTCAAGAAAATARC</u> <u>AAC</u> -----	AACTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
KRAP 5S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	TCCTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	TGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
KRAP 12L	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	TCCTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	YGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
CON 20S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	---TCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
DEN 20S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	---TCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAT	<u>AAGGTG</u>	621
DEN 54L	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	---TCCAAAGCTATT	GATACGGCAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAT	<u>AAGGTG</u>	621
TJ 30L	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	TCCTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
TJ 29S	<u>ATTCAAGAAAATAAC</u> <u>AAC</u> -----	TCCTCCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
CHAM 4L	-----	-----	-----	-----	-----	-----	-----	621
WED 2S	<u>ACTCATGAAAATAGC</u> <u>AAC</u> -----	-----AAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CCGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
DIF	<u>ACTCAAGAAAATAAC</u> <u>AAC</u> AGCAGCAACAAC	ACCACCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAG</u> ---	621
PAN 2S	<u>ACTCAAGAAAATAAC</u> <u>AAC</u> -----	ACCACCAAAGCTATT	GATACGACAGAGTTG	AAGTCGTT	CGTGGAG	CGGGTTGACTTGAAC	<u>AAGGTG</u>	621
SELECTION								

Figure B5: AP2D nucleotide alignment. Alignment of DNA sequences for AP2D obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. No identical sequences were identified in this alignment. An early stop codon, identified in the sequenced obtained for CHAM 4L (*T. chamaedrifolia*) is shown in red. This stop codon was removed from all analyses, however. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to a plant-specific AP2 DNA binding domain according to BLAST searches is indicated by arrows at the base of the alignment (↑). Conserved residues that form the DNA binding site are indicated by boxes. The length of the alignment (bp) is given in the right-most column (621 bp, total, representing almost the entire gene). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGGCCATACCT	TCACTGGGCGCCCTC	105
F60SS	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
SL8 201S	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
E 207S	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
E 2L	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
DROT 41S	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
ES	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
MIDC 710S	GCGGCATATTACCAG	CTGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
MAN 601S	GCGGCATATTACCAG	COGCTGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
MAN 713L	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
COLO	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
PA 4S	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
TSH	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
KRAP 12L	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
CON 20S	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
DEN 54L	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
DEN 20S	GCGGCATATTACCAG	COGCTGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
TJ 30L	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
TJ 29S	GYGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
CHAM 4L	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
WED 2S	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
DIF	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
PAN 2S	GCGGCATATTACCAG	COGCCGGTTCCTCCT	CCGCCAACC---GTC	GCACACTACCCGGCC	TACTACCAAGCCACA	CGGCCGACCATACCT	TCACTGGGCGCCCTC	105
SELECTION	-	-	-	-	-	-	-	-

D16L	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCGTAC	GGG-----	210
F60SS	CCGCTCCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAGACCT	TCGCCG-----	---GGAGCCCCCGCA	GGTTTCGCATCCTAC	GTG-----	210
SL8 201S	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAGACCT	TYGSCG-----	---GGAGCCCCCGCA	GGTTTCGCATCCTAC	GTG-----	210
E 207S	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCGTAC	GGG-----	210
E 2L	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTTCGCATCCTAC	GGG-----	210
DROT 41S	CCGCTTCCGCCA---	CACCACCACCCTTAN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNGCATCCTAC	GTG-----	210
ES	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAGACCT	TYGSCG-----	---GSAGCCCCCGCA	GGTTTCGCATCCTAC	GTG-----	210
MIDC 710S	CCGCTCCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAGATNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNGCATCCTAC	GTGGCTCCTTAC---	210
MAN 601S	CCGCTCCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGCTTCGCATCGTAC	GGG-----	210
MAN 713L	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGSCG-----	---GGAGCCCCCGCA	GGTTTCGCATCCTAC	GTG-----	210
COLO	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCCTAC	GTG-----	210
PA 4S	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCCTAC	GTG-----	210
TSH	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCGTAC	GGG-----	210
KRAP 12L	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCGTAC	GGG-----	210
CON 20S	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAGACCT	TCGCCG-----	---GCAGCCCCCGCA	GGTTTCGATCCTAC	GTGGCTCCTTAC---	210
DEN 54L	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	ACGCCGGCAGCAGGA	GCGGCAGCCCCCGCG	GGTTTCGCATCGTAC	GTGGCGCCTTACGTA	210
DEN 20S	CCGCTCCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGCTTCGCATCGTAC	GGG-----	210
TJ 30L	CCGCTTCCGCCA---	CACCACCACCCATA	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCGTAY	GGG-----	210
TJ 29S	CCGCTTCCGCCA---	CACCACCACCCATA	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCGTAC	GGG-----	210
CHAM 4L	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	TTGGCG-----	---GCAGCCCCCGCA	GGTTTCGCATCCTAC	GTG-----	210
WED 2S	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAGCCG	GCAGCG-----	-----	-----	-----TACGTC	210
DIF	CCGCTTCCGCCA---	CACCACCACCCTTAC	ATCCCCCAGCAACCT	CCGCCGCCA-----	---GCCGCCCCCGCG	GGTTTCGCATCGTAC	GTGGCTCCTTACTTA	210
PAN 2S	CCGCTTCCGCCAC	CACCACCACCCTTAC	ATCCCCCAGCAGCCG	GCAGCG-----	-----	-----	-----TACGTC	210
SELECTION	-	-	-	+	-	-	+	-

D16L	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
F60SS	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTTCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
SL8 201S	-----GCTCAG	GATCAGGTGCGACA	YTGTTTCGTTGCTGGG	CTTCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
E 207S	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCKGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
E 2L	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCKGGG	CTWCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
DROT 41S	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTTCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
ES	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTTCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
MIDC 710S	-----GGGGCTCAG	CATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
MAN 601S	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
MAN 713L	-----GYTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTTCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
COLO	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTTCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
PA 4S	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTTCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
TSH	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCKGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
KRAP 12L	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
CON 20S	-----GGGGCTCAG	GATCAGGTGCGACT	CTGTTTCGTTGCTGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
DEN 54L	CCACAGGGGGCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGATGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
DEN 20S	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
TJ 30L	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
TJ 29S	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
CHAM 4L	-----GCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCTGGG	CTTCCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
WED 2S	CCGCAGGGGGCTGAG	GATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
DIF	CCACACGGGGGCTCAG	GATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGATGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
PAN 2S	CCACTGGGGGGCTGAG	GATCAGGTGCGACA	CTGTTTCGTTGCGGGG	CTACCGGAGGACGTG	AAGGCCCGGGAAATC	TACAACTCTTCCGC	GAGTTCCCGGGCTAC	315
SELECTION								

D16L	GAATCCTCCAATCTC	CGT	333
F60SS	GAATCCTCCAATCTC	CGT	333
SL8 201S	GAATCCTCCAATCTC	CGT	333
E 207S	GAATCCTCCAATCTC	CGT	333
E 2L	GAATCCTCCAATCTC	CGT	333
DROT 41S	GAATCCTCCAATCTC	CGT	333
ES	GAATCCTCCAATCTC	CGT	333
MIDC 710S	GAATCCTCCAATCTC	CGT	333
MAN 601S	GAATCCTCCAATCTC	CGT	333
MAN 713L	GAATCCTCCAATCTC	CGT	333
COLO	GAATCCTCCAATCTC	CGT	333
PA 4S	GAATCCTCCAATCTC	CGT	333
TSH	GAATCCTCCAATCTC	CGT	333
KRAP 12L	GAATCCTCCAATCTC	CGT	333
CON 20S	GAATCCTCCAATCTC	CGT	333
DEN 54L	GAATCCTCCAATCTC	CGT	333
DEN 20S	GAATCCTCCAATCTC	CGT	333
TJ 30L	GAATCCTCCAATCTC	CGT	333
TJ 29S	GAATCCTCCAATCTC	CGT	333
CHAM 4L	GAATCCTCCAATCTC	CGT	333
WED 2S	GAATCCTCCAATCTC	CGT	333
DIF	GAATCCTCCAATCTC	CGT	333
PAN 2S	GAATCCTCCAATCTC	CGT	333
SELECTION			

Figure B6: RNABP nucleotide alignment. Alignment of DNA sequences for exon 1 of *RNABP* obtained from 23 individuals from the genus, *Turnera*. No sequence information was obtained for the individual, KRAP 5S (*T. krapovickasii*). Gaps in the alignment for sequences from DROT 41S (*T. scabra*) and MIDC 710S (*T. scabra*) are also present, and are represented by a profusion of undefined bases (N's) at the 135-186 bp and 149-186 bp positions, respectively. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) PA 4S, COLO, and CHAM 4L; and 2) DEN 20S and MAN 601S. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a "+" or "-", suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to an RNA binding protein with multiple splicing (RBP-MS)-like RNA-recognition motif according to BLAST searches is indicated by arrows at the base of the alignment (↑). The length of the alignment (bp) is given in the right-most column (333 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

		*	*	*	*	*	*		
D16L	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
F60SS	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
SL8 201S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
E 207S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>ACKGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
E 2L	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>ACKGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
DROT 41S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
ES	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
MIDC 710S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
MAN 601S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
MAN 713L	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
COLO	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
PA 4S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
TSH	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>ACKGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
KRAP 5S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
KRAP 12L	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>ACGGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
CON 20S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
DEN 20S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
DEN 54L	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>ACKGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
TJ 30L	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
TJ 29S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
CHAM 4L	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
WED 2S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
DIF	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
PAN 2S	ATCCAGGACTTGCTG	<u>GACCAGCCAAATCCT</u>	<u>GCTGATCCAGCACA</u>	<u>AACTGAGGGTTATCAT</u>	<u>CTTTTATTTCAG</u>			177	
SELECTION							↑		

Figure B7: *SCE1* nucleotide alignment. Alignment of DNA sequences for exons 3-4 of *SCE1* obtained from 24 individuals from the genus, *Turnera*. All non-coding, intronic sequence information has been excised from the alignment. Boundaries between exons are indicated by “|” symbols in the alignment. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) F60SS, D16L, SL8 201S, DROT 41S, ES, MAN 601S, MAN 713L, PA4S, KRAP 5S, CON 20 S, and DEN 20S; 2) E 2L, E 207S, and TSH; and 3) TJ 29S and TJ 30L. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to a Ubiquitin Conjugating enzyme, E2, catalytic domain according to BLAST searches is indicated by arrows at the base of the alignment (↑). The conserved cysteine active site is indicated by a box. Conserved thioester intermediate interaction residues are starred at the top of the alignment. The length of the alignment (bp) is given in the right-most column (177 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	GGTAACGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
F60SS	GGTAACGTTATCAGT	GCGCTTGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
SL8 201S	GGTAACGTTATCAGT	GCGCTWGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA -----	-----	-----TCCTCT	315
E 207S	GGTAAYGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA -----	-----	-----	315
E 2L	GGTAAYGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
DROT 41S	GGTAACGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
ES	GGTAACGTTATCAGT	GCGCTWGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA -----	-----	-----TCCTCT	315
MIDC 710S	GGTAACGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
MAN 601S	GGTAACGTTATCAGT	GCGCTWGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA -----	-----	-----TCCTCT	315
MAN 713L	GGTAACGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
COLO	GGTAACGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
PA 4S	RGTAACGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
TSH	GGTAAYGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
KRAP 5S	GGTAAYGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGWAAAAGAA	GGTGTGCATGTTCCC	TACCGA -----	-----	-----TCCTCT	315
KRAP 12L	GGTAACGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCTAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
CON 20S	GGTAAYGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA -----	-----	-----	315
DEN 20S	GGTAAYGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA -----	-----	-----TCCTCT	315
DEN 54L	GGTAACGTTATCAGT	GCGCTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA -----	-----	-----	315
TJ 30L	GGTAACGTTATCAGT	GCGCTAGGAGATGAC	AAAAACGCGAAAAGAA	GGTGTTCATGTTCCC	TACCGA GTAGATGAC	ACAAGTGGTCCCTCTA	TCCCCAAGATCTCTT	315
TJ 29S	GGTAACGTTATCAGT	GCGCTAGGAGATGAC	AAAAACGCGAAAAGAA	GGTGTTCATGTTCCC	TACCGA -----GAC	ACAAGTGGTCCCTCTA	TCCCCAAGATCTCTT	315
CHAM 4L	GGTAATGTTATCAGT	GCACTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAAGTGGTCCCTCTA	TCCCCAATATCTCTT	315
WED 2S	GGTAATGTTATCAGT	GCTCTAGGAGATGAG	AAAAACGAAAAGAA	GGTGTGCATGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
DIF	GGTAATGTTATCAGT	GCACTAGGAGATGAG	AAAAACGCGAAAAGAA	GG-----	----- GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
PAN 2S	GGTAACGTTATCAGT	GCACTAGGAGATGAG	AAAAACGCGAAAAGAA	GGTGTGCACGTTCCC	TACCGA GTAGATGAC	ACAATGGTCCCTCTA	TCCCCAAGATCTCTT	315
SELECTION	-	-	-	-	↑	-	-	-

D16L	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
F60SS	CCAGCTCCAAAGCAA	CTAAAATTCTCACCT	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
SL8 201S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCY	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
E 207S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
E 2L	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
DROT 41S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCTGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
ES	CCAGCTCCAAAGCAA	CTAAAATTCTCACCY	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCWGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
MIDC 710S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCTGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
MAN 601S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCY	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCWGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
MAN 713L	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCTGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
COLO	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCTGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
PA 4S	CCRGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCTGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
TSH	CCAGCTCCAAAGCAA	STAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
KRAP 5S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
KRAP 12L	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
CON 20S	-----CCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGARGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
DEN 20S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCARCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
DEN 54L	-----CCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCARCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
TJ 30L	CCAGCTCCAAAGCAA	CTAAAATTCTCACCA	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
TJ 29S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCA	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
CHAM 4L	CCAGCTCCAAAACAA	CTAAAATTCCACCT	GGGATTGCTAATGGA	TCTGCTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACGCGAAAG ATG	GTGCCAGTTGGACAG	420
WED 2S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCWGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
DIF	CCAGCTCCAAAGCAA	STAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCAGCATTTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
PAN 2S	CCAGCTCCAAAGCAA	CTAAAATTCTCACCC	GGGATTGCTAATGGA	TCTGTAAAGAAGCA	GCWGCATTTTGAAC	CAGACACGAAAG ATG	GTGCCAGTTGGACAG	420
SELECTION	-	-	-	-	-	-	-	-

D16L	CTGTCAATGAGGAAA	GTGGCAACCGTTGGG	CAATCTGGA AA ATTG	TGGAGTGGAAAGAGG	AGTCATCATCAATGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
F60SS	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
SL8 201S	CTGTCAATGAGGAAA	GTGGCAGCYGTTGGG	CAATCTGGA AA ARTTG	TGGAGRTGGAAAGAGG	AGTCATCATCARTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
E 207S	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ARTTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
E 2L	CTGTCAATGAGGAAA	GTGGCAGCWGTTGGG	CAATCTGGA AA ATTG	TGGAGRTGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
DROT 41S	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
ES	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ARTTG	TGGAGATGGAAAGAGG	AGTCATCATCARTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
MIDC 710S	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
MAN 601S	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ARTTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
MAN 713L	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
COLO	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
PA 4S	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
TSH	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ARTTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
KRAP 5S	CTGTCAATGAGGAAA	GTGGCAGCYGTTGGG	CAATCTGGA AA ARTTG	TGGAGRTGGAAAGAGG	AGTCATCATCARTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
KRAP 12L	CTGTCAATGAGGAAA	GTGGCAACCGTTGGG	CAATCTGGA AA ATTG	TGGAGTGGAAAGAGG	AGTCATCATCAATGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
CON 20S	CTGTCAATGAGGAAA	GTGGCAGCYGTTGGG	CAATCTGGA AA ARTTG	TGGAGRTGGAAAGAGG	AGTCATCATCARTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
DEN 20S	CTGTCAATGAGGAAA	GTGGCARTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
DEN 54L	CTGTCAATGAGGAAA	GTGGCAACYGTTGGG	CAATCTGGA AA ATTG	TGGAGRTGGAAAGAGG	AGTCATCATCARTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
TJ 30L	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGTGGAAAGAGG	AGCCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
TJ 29S	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGTGGAAAGAGG	AGCCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
CHAM 4L	CTGTCAATGAGGAAA	GCTGCAGCTGTTGGA	CAATCCGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
WED 2S	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
DIF	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCWTCRGTTCAAA	TGGAAATKGCAGAAA	525
PAN 2S	CTGTCAATGAGGAAA	GTGGCAGCTGTTGGG	CAATCTGGA AA ATTG	TGGAGATGGAAAGAGG	AGTCATCATCAGTGG	TTGCTTCAGTTCAA	TGGAAATGGCAGAAA	525
SELECTION								

D16L	CCCTGGAAGCTCTCA	GAA	543
F60SS	CCCTGGAAGCTCTCA	GAA	543
SL8 201S	CCCTGGAAGCTCTCA	GAA	543
E 207S	CCCTGGAAGCTCTCA	GAA	543
E 2L	CCCTGGAAGCTCTCA	GAA	543
DROT 41S	CCCTGGAAGCTCTCA	GAA	543
ES	CCCTGGAAGCTCTCA	GAA	543
MIDC 710S	CCCTGGAAGCTCTCA	GAA	543
MAN 601S	CCCTGGAAGCTCTCA	GAA	543
MAN 713L	CCCTGGAAGCTCTCA	GAA	543
COLO	CCCTGGAAGCTCTCA	GAA	543
PA 4S	CCCTGGAAGCTCTCA	GAA	543
TSH	CCCTGGAAGCTCTCA	GAA	543
KRAP 5S	CCCTGGAAGCTCTCA	GAA	543
KRAP 12L	CCCTGGAAGCTCTCA	GAA	543
CON 20S	CCCTGGAAGCTCTCA	GAA	543
DEN 20S	CCCTGGAAGCTCTCA	GAA	543
DEN 54L	CCCTGGAAGCTCTCA	GAA	543
TJ 30L	CCCTGGAAGCTCTCA	GAA	543
TJ 29S	CCCTGGAAGCTCTCA	GAA	543
CHAM 4L	CCCTGGAAGCTATCA	GAA	543
WED 2S	CCCTGGAAGCTCTCA	GAA	543
DIF	CCCKGGAAGCTCTCA	GAA	543
PAN 2S	CCCTGGAAGCTCTCA	GAA	543
SELECTION			

Figure B8: *FRA1* nucleotide alignment. Alignment of DNA sequences for exons 5-6 and 24-25 of *FRA1* obtained from 24 individuals from the genus, *Turnera*. All non-coding, intronic sequence information has been excised from the alignment. Exons are presented in order (5- 25) and the boundaries between exons are indicated by “|” symbols in the alignment. Base positions showing 100% identity across taxa are shown in blue. No identical sequences were identified in this alignment. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-”, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to a KIF4-like subfamily kinesin motor domain according to BLAST searches is indicated by arrows at the base of the alignment (↑). The length of the alignment (bp) is given in the right-most column (543 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
F60SS	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
SL8 201S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
E 207S	CGGCCSYTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCSGTSACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
E 2L	CGGCCSYTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTSACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
DROT 41S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	MCTGTTCCGGTSACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
ES	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCSGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
MIDC 710S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	MCTGTTCCGGTSACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
MAN 713L	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
MAN601S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCSGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
COLO	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
PA 4S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTSACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
TSH	CGGCCSYTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCSGTSACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
CON 20S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCSGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
KRAP 5S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCSGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
KRAP 12L	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
DEN 20S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCSGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
DEN 54L	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
TJ 30L	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
TJ 29S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
CHAM 4L	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
WED 2S	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
DIF	CGGCCCTAAACCAG	ACCTTCTTGGTCGCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTGTTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGGCGGGCGTCCG	525
PAN 2S	-----CAG	ACATTTTGGTCTCC	TTCGACGTCTCCGGG	AACAACCTCAGCGGG	CCTATTCCGGTGACG	CCACAGCTCTCCGGG	TTCGGTGC GGCGTCCG	525
SELECTION	-	-	-	-	↑	-	-	

D16L	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCCGCTCGGTCCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGCGAATGCG	630
F60SS	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCCGCTCGGTCCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG---GCGAATGCG	630
SL8 201S	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCCGCTCGGTCCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGKCGNNNNNNNNNN	630
E 207S	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCAGCTCGGTYCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGCGAATGCG	630
E 2L	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCAGCTCGGTYCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGYGAATGCG	630
DROT 41S	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCCGCTCGGTCCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGCGAATGCG	630
ES	TTCAGCGGAACGCS	GATCTCTGCGGCGAG	ATTATKAACAGRCC	TGCMGCTCGGTYCT	CCGTTCTTYGAGAAT	---AACACCTCCTCG	TCGKCNNNNNNNNNNN	630
MIDC 710S	TTCAGCGKAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCMGCTCGGTCCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGCGAATGCG	630
MAN 713L	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCCGCTCGGTCCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGCGAATGCG	630
MAN601S	TTCAGCGGAACGCS	GATCTCTGCGGCGAG	ATTATKAACAGRCC	TGCMGCTCGGTYCT	CCGTTCTTYGAGAAT	---AACACCTCCTCG	TCGKCG-----	630
COLO	TTCAGCGGAACSCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCMGCTCGGTCCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGCGAATGCG	630
PA 4S	TTCAGCGGAACSCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCMGCTCGGTCCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGCGAATGCG	630
TSH	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCAGCTCGGTYCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCGTCGGCGAATGCG	630
CON 20S	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCAGCTCGGATTC	CCGTTCTTYGAGAAT	---AACACCTCCTCG	TCG---GCGAATGCG	630
KRAP 5S	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCAGCTCGGTTCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG---GCGAATGCG	630
KRAP 12L	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCAGCTCGRTTCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG---GCGAATGCG	630
DEN 20S	TTCAGSGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGRCC	TGCMGCTCGCRYTCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG---GCGAATGCG	630
DEN 54L	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCAGCTCGGATTC	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG---GCGAATGCG	630
TJ 30L	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGGCC	TGCAGCTCGGTTCC	CCGTTCTTTCGAGAAT	AACAACACCTCCTCG	TCG---GCGAATGCG	630
TJ 29S	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGGCC	TGCAGCTCGGTTCC	CCGTTCTTTCGAGAAT	AACAACACCTCCTCG	TCG---GCGAATGCG	630
CHAM 4L	TTCAGCGGAACGCG	GATCTCTGCGGCGAG	GTTATTAACAGACCC	TGCAGCTCGGTTCC	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG-----	630
WED 2S	TTCAGAGGAACGTG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGCAGCTCGGTTCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG---GCGAATGCG	630
DIF	TTCAGCGGAACGCG	GATCTCTGCGGCGAA	GTTATTAACAGACCC	TGCAGCTCGGTTCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG-----	630
PAN 2S	TACAGCGGAACGCG	GATCTCTGCGGCGAG	ATTATTAACAGACCC	TGTAGCTCTGTTCT	CCGTTCTTTCGAGAAT	---AACACCTCCTCG	TCG---GCGAATGCG	630
SELECTION	-	-	-	-	-	-	-	

D16L	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGAGTAAC	GTCGTCTTGGTGTTT	735
F60SS	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGACCAAC	GTCGTCTTGGTGTTT	735
SL8 201S	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	735
E 207S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGRGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGACCAAC	GTCGTCTTGGTGTTT	735
E 2L	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGRGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGASYAAC	GTCGTCTTGGTGTTT	735
DROT 41S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGASYAAC	GTCGTCTTGGTGTTT	735
ES	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	735
MIDC 710S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGACCAAC	GTCGTCTTGGTGTTT	735
MAN 713L	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGAGTAAC	GTCGTCTTGGTGTTT	735
MAN601S	-----	-----	-----	-----	-----	-----	-----	735
COLO	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGAGTAAC	GTCGTCTTGGTGTTT	735
PA 4S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGAGTAAC	GTCGTCTTGGTGTTT	735
TSH	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGMGGGGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGACCAAC	GTCGTCTTGGTGTTT	735
CON 20S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGAGTAAC	GTCGTCTTGGTGTTT	735
KRAP 5S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGAGTAAC	GTCGTCTTGGTGTTT	735
KRAP 12L	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGAGTAAC	GTCGTCTTGGTGTTT	735
DEN 20S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGACTAAC	GTCGTCTTGGTGTTT	735
DEN 54L	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCC	CCTTCGAGTGCGAAG	CACAAGCGGASTAAC	GTCGTCTTGGTGTTT	735
TJ 30L	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCCCCG	CCTTCGAGTGCGAAG	CACAAGCGGACTAAC	GTCGTCTTGGTGTTT	735
TJ 29S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCCCCG	CCTTCGAGTGCGAAG	CACAAGCGGACTAAC	GTCGTCTTGGTGTTT	735
CHAM 4L	---GCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCTCTGGCTCCG	CCTTCGAGTGCGAAG	CACAAGCGGACTAAC	GTCGTCTTGGTGTTT	735
WED 2S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCT	CCTTCGAGTGCGAAG	CACAAGCGGACTAAC	GTCGTCTTGGTGTTT	735
DIF	---GCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCTCTGGCTCCG	CCTTCGAGTGCGAAG	CACAAGCGGACTAAC	GTCGTCTTGGTGTTT	735
PAN 2S	ACGGCGCCGTTTGGG	CAGAGCGCGCAGGCG	GAAGGCGGAGTGGTG	GTGGCGCTGGCTCCT	CCTTCGAGTGCGAAG	CACAAGCGGACTAAC	GTCGTCTTGGTGTTT	735
SELECTION								

D16L	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCCCGTCGGCG	840
F60SS	GCCGTTGTCCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCTCCGTCGGCG	840
SL8 201S	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	840
E 207S	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCMCCGTCGGCG	840
E 2L	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCMCCGTCGGCG	840
DROT 41S	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCCCGTCGGCG	840
ES	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	840
MIDC 710S	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCCCGTCGGCG	840
MAN 713L	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCCCGTCGGCG	840
MAN601S	-----	-----	-----	-----	-----	-----	-----	840
COLO	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCCCGTCGGCG	840
PA 4S	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCCCGTCGGCG	840
TSH	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCACCGTCGGCG	840
CON 20S	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCTCCGTCGGCG	840
KRAP 5S	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGTCYCCGTCGGCG	840
KRAP 12L	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCCCGTCGGCG	840
DEN 20S	GCCGTCGTCTCTCG	GTCTTGRCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGWCWCGTCGGCG	840
DEN 54L	GCCGTCGTCTCTCG	GTCTTGRCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCCCGTCGGCG	840
TJ 30L	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTTRWCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCACCGTCGGCG	840
TJ 29S	GCCGTCGTCTCTCG	GTCTTGCCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTTRWCAAC	AAAAAGAGGAACAGA	---ACTGTAAACGTA	TCTGCACCGTCGGCG	840
CHAM 4L	GCCGTCGTCTCTCG	GTATTGATCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAG	AGACAGAGGAACAGA	ACTACTGTAAACGTA	TCTGCCCATCGGGC	840
WED 2S	GCCGTCGTCTCTCG	GTCTTGATCGCCCGCG	ATTTTGTGCGTCGGG	GCGGTTTTGATCAAG	AGACAGAGGAACAGA	---ACTGTAAACGTA	ACTGCCCGTCGGCG	840
DIF	GCCGTCGTCTCTCG	GTATTGATCGCCCGCG	ATTTTGTGCGTCATG	GCGGTTTTGATCAAG	AGACAGAGGAACAGA	ACTACTGTAAACGTA	TCTGCCCGTCGGCG	840
PAN 2S	GCCGTCGTCTCTCG	GTCTTGACCGCCACG	ATTTTGTGCGTCGTG	GCGGTTTTGATCAAG	AGGACAGGAACMGA	---ACTGTAAACGTA	ACTGCCCGTCGGCG	840
SELECTION								

D16L	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTTGATGGTAAAT---	-----GCTGCTGCG	GCGGCGGCGGCGAAG	ACGAAAGATATCGAG	945
F60SS	AGTCCGGTCAGGTTG	GCCGGAGCAGAACC	GCCGAGGCTGTGAAA	GTTGACGGTAAAT---	----GCTGCGGGGGCG	GCGGCGGCGGCGAAG	ACGAAAGATATCGAG	945
SL8 201S	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	-----GCKGCKGCG	GCGGCGGCGGCGAAG	ACGAAAGATATCGAG	945
E 207S	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTTGATGGTAAAT---	----GCTGCTGCTGCT	GCTGCGGCGGCGAAG	ACGAAAGATATCGAG	945
E 2L	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTTGATGGTAAAT---	----GCTGCTGCTGCT	GCTGCGGCGGCGAAG	ACRAAAGATATCGAG	945
DROT 41S	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTGGATGGTAAAT---	-----GCTGCTGCT	GCTGCTGCGGCGAAG	ACGAAAGATATCGAG	945
ES	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	NNNNNNNNNNNNNNNN	----GCTGCKGCKGCG	GCKGCGGCGGCGAAG	ACGAAAGATATCGAG	945
MIDC 710S	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTGGATGGTAAAT---	----GCTGCTGCTGCT	GCTGCGGCGGCGAAG	ACGAAAGATATCGAG	945
MAN 713L	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTTGATGGTAAAT---	GCTGCTGCTGCTGCG	GCTGCGGCGGCGAAG	ACGAAAGATATCGAG	945
MAN601S	-----	-----	-----	-----	-----	-----	-----	945
COLO	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTGGATGGTAAATGCT	GCTGCTGCTGCTGCG	GCTGCGGCGGCGAAG	ACGAAAGATATCGAG	945
PA 4S	AGTCKKGTCAAGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTGGATGGTAAAT---	-----GCTGCTGCT	GCTGCKGCGGCGAAG	ACGAAAGATATCGAG	945
TSH	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	GCCGAGGCTGTGAAA	GTTGATGGTAAATGCT	GCTGCTGCTGCTGCT	GCTGCGGCGGCGAAG	ACGAAAGATATCGAG	945
CON 20S	AGTCCGGTCAGGTTG	GCCGGAGCAGAACC	GCCGAGGCTGTGAAA	GTTGACGGTAAAT---	----GCTGCTGCTGCT	GCGGCGGCGGCGAAG	GCGAAAGATATCGAG	945
KRAP 5S	AGTCCGGTCAGGTYG	GCCGGAGMAGAACC	RCCGAGGCTGTGAAA	GTTGAYGGTAAAT---	----GCTGCTGCTGCT	GCTGCGGCGGCGAAG	ACGAAAGATATCGAG	945
KRAP 12L	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	ACCGAGGCTGTGAAA	GTTGATGGTAAAT---	GCTGCTGCTGCTGCK	GCGGCGGCGGCGAAG	ACGAAAGATATCGAG	945
DEN 20S	AGTCCGGTCAGGTCG	GCCGGAGCAGAACC	GCCGAGGCTGTGAAA	GTTGRCGGTAAAT---	-----GCTGCT	GCTGCGGCGGCGAAG	ACGAAAGATATGGAG	945
DEN 54L	AGTCCGGTCAGGTCG	GCCGGAGAAGAACC	ACCGAGGCTGTGAAA	GTTGATGGTAAAT---	-----GCTGCTGCT	GCTGCGGCGGCGAAG	ACGAAAGATATCGAG	945
TJ 30L	AGTCCGGTCAGGTCA	GCCGGAGCAGAACC	GCCGARGCTGTGAAA	GTTGACGGTAAAT---	GCTGCTGCTGCTGCT	GCGGCGGCGGCTAAG	GCGAAAGATATCGAG	945
TJ 29S	AGTCCGGTCAGGTCA	GCCGGAGCAGAACC	GCCGARGCTGTGAAA	GTTGACGGTAAAT---	----GCTGCTGCTGCT	GCGGCGGCGGCTAAG	GCGAAAGATATCGAG	945
CHAM 4L	AGTCCGGTCCGGTCC	GCCGGAGCAGAAATC	GGTGAAGCTGTGAAG	GTTGAGAGAAAT---	-----GCTGCT	GCGTGGGCGGGAAG	ACGAAGATATGGAG	945
WED 2S	AGCCCGGTAAGGTCA	GCCGGAGCAGAACC	GCCGAAGCTGTGAAG	GTCGACGGATAT---	-----GCTGCT	GTTGCGGCGGCGAAG	ACGAAAGATATCGAG	945
DIF	AGTCCGGTCCGGTCC	GCCGGA-----	GTTGAAGCTGTGAAG	GTTGAGAGAAAT---	-----GCTGCTGCT	GCT-----	-----	945
PAN 2S	AGCCCGGTCGGTCA	GCCGGAGCAGAACC	GCTGAAGCTGTGAAG	GTTGAGGGAAT---	----GCTACTGCGACG	GCGAGCGGCGCAAG	ACGAAAGATATCCAG	945
SELECTION	-	-	-	-	-	-	-	-

D16L	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
F60SS	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGA	CTAGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTGTACACATTGGAG	1050	
SL8 201S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGR	CTRGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTSTACACATTGGAG	1050	
E 207S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGS	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTSTACACATTGGAG	1050	
E 2L	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
DROT 41S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
ES	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGR	CTRGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTSTACACATTGGAG	1050	
MIDC 710S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
MAN 713L	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
MAN601S	-----	-----	-----	-----	-----	-----	-----	1050	
COLO	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
PA 4S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
TSH	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGCG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
CON 20S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGA	CTAGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTGTACACATTGGAG	1050	
KRAP 5S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGR	CTRGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTSTACACATTGGAG	1050	
KRAP 12L	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTCTACACATTGGAG	1050	
DEN 20S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCAGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTGTACACATTAGAG	1050	
DEN 54L	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGG	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GTNTACACATTAGAG	1050	
TJ 30L	GTGGGAGAAGCAGTG	GCGGTGAGGGCAGGG	AGCAAGAGCGGGGGA	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GCGTACACWTTTRGAG	1050	
TJ 29S	GTGGGAGAAGCAGTG	GCGGTGAGGGCAGGG	AGCAAGAGCGGGGGA	CTGGTGTTCGCG---	-----GGGAAGCG	CAG-----	GCGTACACWTTTRGAG	1050	
CHAM 4L	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGA	CTGGTGTTCGCG---	-----GGGAAGCG	CAGACGTACCCGTAC	CCGTACACATTAGAG	1050	
WED 2S	GTGGGAGAAGCAGTG	GCGGTGAGGGCGGGG	AGCAAGAGCGGGGGT	CTGGTGTTCGCGGG	GTTGGCGGGGAGCG	CAG-----	CCTTAC	CCTTACACATTAGAG	1050
DIF	-----	-----	-----	-----	-----	-----	-----	1050	
PAN 2S	GTGGGAGAAGGAGGG	GCGGTGAGGGCGTGG	GGCAAGAGCGGGGGT	CTGGTGTTCGCG---	-----GGGAAGCG	AAG-----	GTGTACACATTAGAG	1050	
SELECTION	-	-	-	-	-	-	-	-	

D16L	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
F60SS	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
SL8 201S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTRACG	GTGAAGAGGCTGGAC	1155
E 207S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
E 2L	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
DROT 41S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	-----	-----	-----	1155
ES	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGAYGATAGTGACG	GTGAAGAGGCTGGAC	1155
MIDC 710S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
MAN 713L	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
MAN601S	-----	-----	-----	-----	-----	-----	-----	1155
COLO	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CRGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
PA 4S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
TSH	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
CON 20S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGRTAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
KRAP 5S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGRTAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
KRAP 12L	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACRATG	GGGACGACGTACAAG	GCGGTGATGGTAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
DEN 20S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
DEN 54L	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACCATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
TJ 30L	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACSATG	GGGACGACGTACAAG	GCGGTGATGGATMAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
TJ 29S	CAGTTAATGAGAGCG	TCGGCCGAGTTGCTG	GGGAGGGGAACSATG	GGGACGACGTACAAG	GCGGTGATGGATMAT	CAGATGATAGTGACG	GTGAAGAGGCTGGAC	1155
CHAM 4L	CAGTTAATGAGAGCG	TCGGCCGAGTTGTTA	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACT	GTGAAGAGACTTGAC	1155
WED 2S	CAGTTAATGAGAGCG	TCGGCCGAGTTGTTA	GGGAGGGGAACATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACG	GTGAAGAGACTTGAC	1155
DIF	-----	-----	-----	-----	-----	-----	-----	1155
PAN 2S	CAGTTAATGAGAGCG	TCGGCCGAGTTGTTA	GGGAGGGGAACGATG	GGGACGACGTACAAG	GCGGTGATGGATAAT	CAGATGATAGTGACA	GTGAAGAGGCTGGAC	1155
SELECTION	-----	-----	-----	-----	-----	-----	-----	1155

D16L	<u>GCGAGCAAGACTGGA</u>	-----	---	1188
F60SS	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
SL8 201S	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
E 207S	<u>GCSAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
E 2L	<u>GCSAGCAAGACTGSA</u>	GTTACGAGCAGTGAA	GCG	1188
DROT 41S	-----	-----	---	1188
ES	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
MIDC 710S	<u>GCCAGCAAGACTGGA</u>	GTTACGAGCAGTGAA	GCG	1188
MAN 713L	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
MAN601S	-----	-----	---	1188
COLO	<u>GCGAGCAAGACTGGA</u>	RTTACGAGCAGTGAA	GCG	1188
PA 4S	<u>GCSAGCAAGACTGSA</u>	GTTACGAGCAGTGAA	GCG	1188
TSH	<u>GCGAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
CON 20S	<u>GCGAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
KRAP 5S	<u>GCSAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
KRAP 12L	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
DEN 20S	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	---	1188
DEN 54L	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCG	1188
TJ 30L	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGCGAA	GCG	1188
TJ 29S	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGCGAA	GCG	1188
CHAM 4L	<u>GCCAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	GCT	1188
WED 2S	<u>GCGAGC</u>	-----	---	1188
DIF	-----	-----	---	1188
PAN 2S	<u>GCGAGCAAGACTGCA</u>	GTTACGAGCAGTGAA	---	1188
SELECTION	-----	-----	---	1188

Figure B9: *LRRK* nucleotide alignment. Alignment of DNA sequences for exon 1 of *LRRK* obtained from 24 individuals from the genus, *Turnera*. Gaps in the alignment for sequences from SL8 201S (*T. subulata*) and ES (*T. scabra*) are also present, and are represented by a profusion of undefined bases (N's) at the 622-896 bp and 621-896 bp positions, respectively. Base positions showing 100% identity across taxa are shown in blue. No identical sequences were identified in this alignment. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-”, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to a provisional Leucine-rich repeat receptor-like protein kinase multi-domain according to BLAST searches is indicated by arrows at the base of the alignment (↑). The length of the alignment (bp) is given in the right-most column (1188 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	TGGCTCCTGCATTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTAATCTACACT	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
F60SS	TGGCTCCTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTAATCTATACT	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
SL8 201S	TGGCTCCTTGCMTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTAATCTAYACT	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
E 207S	TGGCTCCTTGCMTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTWATCTAYACY	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
E 2L	TGGCTCCTTGCMTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTWATCTAYACY	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
DROT 41S	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTTATCTACACC	AGAGAATCCTTGCCCT	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
ES	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTWATCTATACY	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
MIDC 710S	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTTATCTACACC	AGAGAATCCTTGCCCT	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
MAN 601S	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTWATCTATACY	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
MAN 713L	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTTATCTATACC	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
COLO	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTTATCTACACC	AGAGAATCCTTGCCY	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
PA 4S	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTTATCTAYACC	AGAGAATCCTTGCCY	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
TSH	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTWATCTAYACY	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
KRAP 5S	TGGCTCCTTGCATTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTTATCTATACC	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
KRAP 12L	TGGCTCCTTGCATTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTTATCTATACC	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
CON 20S	TGGCTCCTTGCMTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTAATCTAYACT	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
DEN 20S	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTAATCTATACT	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
DEN 54L	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTAATCTATACT	AGAGAATCCTTGCCC	ACCAAAACA---GCC	ACCAGCGCCACCATG	105
TJ 30L	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACGCTAATCTATACT	AGAGAATCCTTGCCC	ACCAAAACTGCGAGCC	ACCAGCGCCACCATG	105
TJ 29S	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACGCTAATCTAYACY	AGAGAATCCTTGCCC	ACCAAAACTGCGAGCC	ACCAGCGCCACCATG	105
CHAM 4L	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTA	ACACTAATCTACACT	AGAGAATCCTTGCCCT	ACCAAAACA---ACY	ACCAGCGCCACCATG	105
WED 2S	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACASTAATCTACACT	AGAGAATCCTTGCCCT	ACCAAAAGA---ACC	ACCAGCGCCACCATG	105
DIF	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTA	ACACTAATCTACWCT	AGAGAATCCTTGCCCT	ACCAAAACA---ACC	ACCAGCGCCACCATG	105
PAN 2S	TGGCTCCTTGCCTTT	GTATCCTTCTTCACC	ATTGCCTTCCCTCCTC	ACACTAATCTACACT	AGAGAATCCTTGCCC	ACCAAAACA---ACC	ACCAGCGCCACCATG	105
SELECTION	-	-	-	-	-	-	-	-

D16L	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCACCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
F60SS	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCGCCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
SL8 201S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCRCCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCY	210
E 207S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCRCCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCY	210
E 2L	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCACCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCK	210
DROT 41S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCACCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
ES	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCRCCAWTKCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCY	210
MIDC 710S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCACCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
MAN 601S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCRCCAWTKCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCY	210
MAN 713L	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCACCAATTGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
COLO	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCACCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
PA 4S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCACCAWTGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
TSH	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCRCCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCB	210
KRAP 5S	ACCACCGCCAGCACC	ACCGGGAAY---AAT	GCACCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
KRAP 12L	ACCACCGCCAGCACC	ACCGGGAAY---AAT	GCACCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
CON 20S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCRCCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCY	210
DEN 20S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCRCCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCY	210
DEN 54L	ACCACCGCCAGCACC	ACCGGGAACATTAAK	GCACCATTGGCCACC	RCAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
TJ 30L	ACCACCGCCAGCACC	GCSGGGAAC---AAT	GCRCCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTACGCC	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
TJ 29S	ACCACCGCCAGCACC	GCCGGGAAC---AAT	GCACCMTTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTAYGCM	TCTAGATCAAACGAC	ACCTTTCACATGCCCT	210
CHAM 4L	ACCGCYGCCAGCACC	ACCGGGAAC---AAT	GCCCCATTGGCCAACC	ACAGTCATCAGCACA	CTCCTCCACTATGCC	TCTAGATCAAACGAY	ASTTTCACATGCCCT	210
WED 2S	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCACCATTGGCCAACC	ACAGTCATCAACACA	CTCCTCCACTACGCC	TCTAGATCAAACGAC	ACTTTCACATGCCCT	210
DIF	ACCACCGCCAGCACC	ACCGGGAAC---AAT	GCCCCATTGGCCAACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCTAGATCAAACGAC	ACTTTCACATGCCG	210
PAN 2S	ACCACCGCCAGCACC	ACCGGGAAC---AGT	GCRCCATTGGCCACC	ACAGTCATCAACACA	CTCCTCCACTATGCC	TCCAGATCAAACGAC	ACTTTCACATGCCCT	210
SELECTION	-	-	-	-	-	-	-	-

D16L	ACAGCTGGT	GTTTCTT	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
F60SS	ACAGCTGGT	GTTTCTT	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
SL8 201S	ACAGCTGGT	GTTTCTT	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
E 207S	ACAGCTGGT	GTTTCTT	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
E 2L	ACAGCTGGT	GTTTCTW	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
DROT 41S	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
ES	ACAGCTGGT	GTTTCTW	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
MIDC 710S	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
MAN 601S	ACAGCTGGT	GTTTCTW	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
MAN 713L	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
COLO	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
PA 4S	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
TSH	ACAGCTGGT	GTTTCTT	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
KRAP 5S	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
KRAP 12L	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
CON 20S	ACAGCTGGT	GTTTCTW	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CAYGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
DEN 20S	ACRGTGGT	GTTTCTW	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	YAYGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
DEN 54L	ACRGTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
TJ 30L	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
TJ 29S	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
CHAM 4L	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTA	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
WED 2S	ACAGCTGGC	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTG	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
DIF	ACAGCTGGT	GTTTCTA	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTG	CACGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
PAN 2S	ACAGCTGGT	GTTTCTT	GCTAGAAGCAAGAAG	GCTAGCAATGCCAAG	ACTCACATATTTGTG	CATGATTACTACAGA	GACGTGGAGAGGGTT	TATGGGGATGAGTTC	735
SELECTION									

D16L	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
F60SS	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
SL8 201S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
E 207S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
E 2L	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
DROT 41S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
ES	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
MIDC 710S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
MAN 601S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
MAN 713L	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
COLO	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
PA 4S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
TSH	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
KRAP 5S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
KRAP 12L	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
CON 20S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
DEN 20S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
DEN 54L	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
TJ 30L	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
TJ 29S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
CHAM 4L	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
WED 2S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
DIF	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
PAN 2S	TTGTGCAGGGAGAAC	TTGGTGGACACAAAT	GACATGCTTGCA	777
SELECTION				

Figure B10: *IRX15L* nucleotide alignment. Alignment of DNA sequences for *IRX15L* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. Sequences for MAN 601S and ES were found to be identical at this locus. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to an uncharacterized plant-specific domain, and member of the polysaccharide biosynthesis domain superfamily, according to BLAST searches is indicated by arrows at the base of the alignment (↑). The length of the alignment (bp) is given in the right-most column (777 bp, total, representing most of the gene). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	CCAAACTTCAAACCA	CTCGTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
F60SS	CCAAACTTCAAACCA	CTCATGGAGAAATAT	GTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAATA	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
SL8 201S	CCAAACTTCAAACCA	CTCGTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAATW	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
E 207S	CCAAACTTCAAACCA	CTCGTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
E 2L	CCAAACTTCAAACCA	CTCRTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
DROT 41S	CCAAACTTCAAACCA	CTCGTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTRCATTAAAGT	GGTTCACCCGATGAG	105
ES	-----	-----	-----	- ---CTTTCAAGGAAA	ATTATGCGCGGAATW	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
MIDC 710S	CCAAACTTCAAACCA	CTCRTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTGGCATTAAAGK	GGTTCACCCGATGAG	105
MAN 601S	-----	-----	-----	- ---CTTTCAAGGAAA	ATTATGCGCGGAATW	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
MAN 713L	CCAAACTTCAAACCA	CTCGTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
COLO	CCAAACTTCAAACCA	CTCGTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
PA 4S	CCAAACTTCAAACCA	CTCGTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
TSH	-----	-----	-----	- ---CTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
KRAP 5S	CCAAACTTCAAACCA	CTCATGGARAAATAT	STCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTRCATTAAAGT	GGTTCACCCGATGAG	105
KRAP 12L	CCAAACTTCAAACCA	CTCATGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTGGCATTAAAGT	GGTTCACCCGATGAG	105
CON 20S	-----	-----	-----	- ---CTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
DEN 20S	-----	-----	-----	- ---CTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
DEN 54L	CCAAACTTCAAACCA	CTCRTGGAGAAATAT	CTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	GCTCTAGCATTAAAGT	GGTTCACCCGATGAG	105
TJ 30L	CCAAACTTCAAACCA	CTCATGGAGAAATAT	GTCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	TCTCTAGCATTACGT	GGTTCACCCGATGAG	105
TJ 29S	CCAAACTTCAAACCA	CTCATGGAGAAATAT	STCAACCTGTGCACA	G ACCTTTCAAGGAAA	ATTATGCGCGGAAT	KCTCTAGCATTAMGT	GGTTCACCCGATGAG	105
CHAM 4L	CCAAACTTCAAACCA	CTCATGGAGAAATAT	GTCAACCTTTGCACA	G ACCTTTCAAGGAAA	ATCATGCACGGAAT	GCTCTGGCATTAAAGT	GGATCACCCGATGAG	105
WED 2S	CCAAACTTCAAACCA	CTCATGGAGAAATAT	GTCAACCTTTGCACA	G ACCTTTCAAGGAAA	ATTATGCGTGGGAAT	GCTCTAGCATTAAAGT	GGATCACCCGATGAG	105
DIF	CCAAACTTCAAACCA	CTCATGGAGAAATAT	GTCAACCTTTGCACA	G ACCTTTCAAGGAAA	ATCATGCTCGGAAT	GCTCTAGCATTAAAGT	GGATCACCCGATGAG	105
PAN 2S	---AACTTCAAACCA	CTCATGGAGATATAT	GTCAACCTTTGCACA	G ACCTTTCAAGGAAA	ATTATACGCGGAAT	GCTCTAGCATTAAAGT	GGATCACCTGATGAG	105
SELECTION	↑							

D16L	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
F60SS	<u>TTTGAAGGGGATATA</u>	ACAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
SL8 201S	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
E 207S	<u>TTTGAAGGGGATATA</u>	KCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
E 2L	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
DR0T 41S	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
ES	<u>TTTGAAGGGGATATA</u>	RCAGGAGACGCATTT	TGG-----	-----	-----	---	183
MIDC 710S	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
MAN 601S	<u>TTTGAAGGGGATATA</u>	RCAGGAGACGCATTT	TGG-----	-----	-----	---	183
MAN 713L	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
COLO	<u>TTTGAAGGGGATATA</u>	GCTGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
PA 4S	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
TSH	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCA	-----	-----	-----	---	183
KRAP 5S	<u>TTTGAAGGGGATATA</u>	TCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
KRAP 12L	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
CON 20S	<u>TTTGAAGGGGATATA</u>	KCAGGAGACGCATTT	TGG-----	-----	-----	---	183
DEN 20S	<u>TTTGAAGGGGATATA</u>	KCAGGAGACGCATTT	TGG-----	-----	-----	---	183
DEN 54L	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
TJ 30L	<u>TTTGAAGGGGATATA</u>	GCAGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
TJ 29S	<u>TTTGAAGGGGATATA</u>	GCMGGAGACGCATTT	TGGGTATTACGCATT	ATTGGTTACCCTGGT	CTTTCAGCACAAAT	GGT	183
CHAM 4L	<u>TTTGAAGGTGATATA</u>	GCAGGAGACGCATTT	TGGGTATTGCGCATT	ATTGGTTACCCTGGT	ATTTTCGAGCACAAAT	GGT	183
WED 2S	<u>TTTGAAGGTGATATA</u>	GCAGGAGACGCATTT	TGGGTATTGCGCATT	ATTGGTTACCCTGGC	CTTTCGAGCACAAAT	GGT	183
DIF	<u>TTTCAAGGTGATATA</u>	GCAGGAGACGCATTT	TGGGTATTGCGCATT	ATTGGTTACCCTGGT	CTTTCGAGCACAAAT	GGT	183
PAN 2S	<u>TTTGAAGGTGATAGA</u>	GCAGGAGACGCATTT	TGGGTATTGCGCATT	ATTGGCTACCCTGGT	CTTTCAGCACAAAT	GGT	183
SELECTION							



Figure B11: *FSP* nucleotide alignment. Alignment of DNA sequences for exons 6 and 7 of *FSP* obtained from 24 individuals from the genus, *Turnera*. All non-coding, intronic sequence information has been excised from the alignment. Boundaries between exons are indicated by “|” symbols in the alignment. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) MAN 601 S and ES; and 2) MAN 713L, D16L, and PA 4S. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to an undefined oxidoreductase-related domain, according to BLAST searches is indicated by arrows at the base of the alignment (↑). The length of the alignment (bp) is given in the right-most column (183 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
F60SS	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCCGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
SL8 201S	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGCGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
E 207S	AAATCTCTGGTTGAA	AAASTTGCTCTTAAAC	CTTCATGTGCCYGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCM AAGATGCTG	105
E 2L	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAMTTAGT	TATGCCAAGATGCTG	105
DROT 41S	AAATCTCTGGTTGAA	AAACTTGCTCKTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
ES	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCYGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
MIDC 710S	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
MAN 601S	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCYGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
MAN 713L	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
COLO	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
PA 4S	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
TSH	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCYGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCV AAGATGCTG	105
KRAP 5S	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCYGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCM AAGATGCTG	105
KRAP 12L	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCCAAGATGCTG	105
CON 20S	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCYGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TATGCM AAGATGCTG	105
DEN 20S	AAATCTCTGGTTGRA	AAACTTGCTCTTAAAC	CTTCATGTGCCYGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTYG	GAAGATACAATTAGT	TATGCV AAGATGCTG	105
DEN 54L	AAATCTCTGGTTGRA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTYG	GAAGATACAATTAGT	TATGCV AAGATGCTG	105
TJ 30L	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAGT	TAYGCA AAGATGCTG	105
TJ 29S	AAATCTCTGGTTGAA	AAAMTTGCTCTTAAAY	CTTCATGYGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GARGATACAATTAGT	TAYGCA AAGATGCTG	105
CHAM 4L	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACGATTAGT	TATGCCAAGATGCTG	105
WED 2S	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTG	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAAAT	TATGCCAAGATGCTG	105
DIF	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACGATTAGT	TATGCCAAGATGCTG	105
PAN 2S	AAATCTCTGGTTGAA	AAACTTGCTCTTAAAC	CTTCATGTGCCGTGTA	TCTTGCAAAATCCGT	TGTTTTCCAAAAGTTG	GAAGATACAATTAAAT	TATGCCAAGATGCTG	105
SELECTION			↑					

D16L	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGCTATC	AAGGCTGTAAAAGAT	210
F60SS	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CACGGTCGAACGAGG	GACGAGAAGGATGGG	AAGAAATTCGTGCC	AATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
SL8 201S	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGAACRAGR	GACGAGAAGGATGGG	AAGAAATTCYCGKCC	RATTGGAAARGCTATC	AAGGCTGTAAAAGAT	210
E 207S	GAGGAWGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGAACRAGR	GACGAGAAGGATGGS	AAGAAATTCYCGKCC	RATTGGAAARGCTATC	AAGGCTGTAAAAGAT	210
E 2L	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAARGCTATC	AAGGCTGTAAAAGAT	210
DROT 41S	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
ES	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CACGGTCGAACRAGR	GACGAGAAGGATGGG	AAGAAATTCYCGKCC	RATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
MIDC 710S	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
MAN 601S	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGAACRAGR	GACGAGAAGGATGGG	AAGAAATTCYCGKCC	RATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
MAN 713L	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
COLO	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
PA 4S	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
TSH	GAGGAWGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGAACRAGR	GACGAGAAGGATGGG	AAGAAATTCYCGKCC	RATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
KRAP 5S	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGAACRAGR	GACGAGAAGGATGSG	AAGAAATTCYCGKCC	RATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
KRAP 12L	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
CON 20S	GAGGAWGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGAACRAGR	GACGAGAAGGATGGG	AAGAAATTCYCGKCC	RATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
DEN 20S	GAGGAWGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGRACRAGR	GACGAGAAGGATGGG	AAGAAAWTYCGKCC	RATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
DEN 54L	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGRACRAGR	GACGAGAAGGATGGG	AAGAAATTCYCGKCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
TJ 30L	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
TJ 29S	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CAYGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGGGCC	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
CHAM 4L	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGTGCT	GATTGGAAAGGCTATC	AAGGCCGTAAAAGAT	210
WED 2S	GAGGAAGCTGGCTGC	TYGCTTTTGGCTGTG	CATGGTCGACAGA	GACGAGAAGGATGGG	AAGAAATTCGTGCT	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
DIF	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GACGAGAAGGATGGG	AAGAAATTCGTGCT	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
PAN 2S	GAGGAAGCTGGCTGC	TCGCTTTTGGCTGTG	CATGGTCGAACAAGA	GTCGAGAAGGATGGG	AAGAAATTCGTGCT	GATTGGAAAGGCTATC	AAGGCTGTAAAAGAT	210
SELECTION			↑					

D16L	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
F60SS	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGGAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTCTCGCTGAGACT	315
SL8 201S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
E 207S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
E 2L	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
DROT 41S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
ES	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
MIDC 710S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGSAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
MAN 601S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
MAN 713L	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
COLO	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
PA 4S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
TSH	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
KRAP 5S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
KRAP 12L	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
CON 20S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
DEN 20S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGRAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTYTCGCTGAGACT	315
DEN 54L	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
TJ 30L	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
TJ 29S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
CHAM 4L	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
WED 2S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAGC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
DIF	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
PAN 2S	GCCCTCAGAATCCCA	GTGCTTGCCAAAGGG	AACATACGTCATATG	GATGATGTTTCGGAAAC	TGCTTGGAAAGAGACT	GGCACTGATGGGGTG	CTTTCGCTGAGACT	315
SELECTION								

D16L	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGAGGCAAGTGTA	GATGGGCAACTAGAC	CAGGCAGATTTATTG	420
F60SS	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGTGCTGTTGGTGGGA	GAAGAGGCAAGTGCA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
SL8 201S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGKGCYGTGGTGGGA	GAAGAGGCAAGTYA	GATGGRCAACTAGAC	CAGGCWATTTATTG	420
E 207S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCYGTGGTGGGA	GAAGAGGCAAGTYA	GATGGRCAACTAGAC	CAGGCWATTTATTG	420
E 2L	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGAGGCAAGTGTA	GATGGRCAACTAGAC	CAGGCAGATTTATTG	420
DROT 41S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGARAGCAGTYTA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
ES	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGTGACGTTGGTGGGA	GAAGAGGCAAGTYA	GATGGCAACTAGAC	CAGGCWATTTATTG	420
MIDC 710S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGAGGCAAGTGTA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
MAN 601S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGKGCYGTGGTGGGA	GAAGAGGCAAGTYA	GATGGCAACTAGAC	CAGGCWATTTATTG	420
MAN 713L	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGAGGCAAGTGTA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
COLO	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGARAGCAGTYTA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
PA 4S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGARAGCAGTYTA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
TSH	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCYGTGGTGGGA	GAAGAGGCAAGTYA	GATGGCAACTAGAC	CAGGCWATTTATTG	420
KRAP 5S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCYGTGGTGGGA	GAAGAGGCAAGTYA	GATGGCAACTAGAC	CAGGCWATTTATTG	420
KRAP 12L	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGAGGCAAGTYA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
CON 20S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCYRTGGTGGGA	GAAGAGGCAAGTYA	GATGGCAACTAGAC	CAGGCWATTTATTG	420
DEN 20S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCYGTGGTGGGA	GAAGAGGCAAGTYA	GATGGCAACTAGAC	CAGGCWATTTATTG	420
DEN 54L	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGAGGCAAGTYA	GATGGCAACTAGAC	CAGGCWATTTATTG	420
TJ 30L	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	YTCCGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGAGGCAAGTYTA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
TJ 29S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	YTCCGACCGCTGAG	TGGCCGTTGGTGGGA	GAAGAGGCAAGTYTA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
CHAM 4L	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGGGGA	GAAGAGGCAAGTATA	GATGGCAACTAGAC	CARGCAGATTTATTG	420
WED 2S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGCGAA	GAAGAGGCAAGTATA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
DIF	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGGGGGA	GAAGAGGCAAGTRTA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
PAN 2S	CTCCTCGAGAATCCA	GCTCTCTTTGCTGGA	TTCAGACCGCTGAG	TGGCCGTTGAGTGGGA	GAAGAGGCAAGTATA	GATGGCAACTAGAC	CAGGCAGATTTATTG	420
SELECTION								

D16L	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
F60SS	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
SL8 201S	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
E 207S	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
E 2L	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
DROT 41S	GTGGAATATTTGAAG	YTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
ES	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCRTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
MIDC 710S	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
MAN 601S	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCRTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
MAN 713L	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
COLO	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
PA 4S	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
TSH	GTGGARTATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
KRAP 5S	GTGGARTATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
KRAP 12L	GTGGATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
CON 20S	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
DEN 20S	GTGGARTATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
DEN 54L	GTGGAGTATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
TJ 30L	GTGGAGTATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
TJ 29S	GTGGAGTATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAC	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
CHAM 4L	GTGGAATATTTGAAG	CTTTGTGAAAAGTWC	CCTGTGCCGTGGAGA	ATGATCCGTGCTCAT	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
WED 2S	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCATGGAGA	ATGATCCGTGCTCAT	GTTTCATAAGCTGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
DIF	GTGGAATATTTGAAG	CTTTGTGAAAAGTAC	CCTGTGCCRTGGAGA	ATGATCCGTGCTCAT	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
PAN 2S	GTGGAATATCTGAAG	CTTTGTGAGAAGTAC	CCTGTGCCGTGGAGA	ATGATCCGTGCTCAT	GTTTCATAAGATGTTG	GGAGACTGGTTCAGG	ATCCATCCCGAAGTA	525
SELECTION								

D16L	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
F60SS	<u>AGGCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
SL8 201S	<u>AGRCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
E 207S	<u>AGRCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>YATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
E 2L	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
DROT 41S	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
ES	<u>AGRCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
MIDC 710S	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
MAN 601S	<u>AGRCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
MAN 713L	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
COLO	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
PA 4S	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
TSH	<u>AGRCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>YATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
KRAP 5S	<u>AGRCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
KRAP 12L	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
CON 20S	<u>AGRCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
DEN 20S	<u>AGRCAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
DEN 54L	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
TJ 30L	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
TJ 29S	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAAACTA</u>	<u>ACTTTTGAATTTCTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
CHAM 4L	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAGRCTT</u>	<u>ACTTTTGAATTTTTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
WED 2S	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAGACTA</u>	<u>ACTTTTGAATTTTTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
DIF	<u>AGACAGGATCTCAAT</u>	<u>GCACAATCCAGACTT</u>	<u>ACTTTTGAATTTTTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTC-----</u>	<u>---</u>	<u>---</u>	603
PAN 2S	<u>AGACAGTATCTCAAT</u>	<u>GCACAATCCAGACTA</u>	<u>ACTTTTGAATTTTTA</u>	<u>TATGATTTGGTGGAT</u>	<u>CGTCTCAGAAAAC</u>	<u>TG</u>	<u>GGT</u>	603
SELECTION								

Figure B12: *NRFP* nucleotide alignment. Alignment of DNA sequences for exon 3 of *NRFP* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. No identical sequences were identified in this alignment. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The region of the alignment that was found to be homologous to a dihydrouridine synthase-like FMN-binding domain, according to BLAST searches is indicated by arrows at the base of the alignment (↑). Conserved residues that form a phosphate binding site are indicated by boxes. The length of the alignment (bp) is given in the right-most column (603 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
F60SS	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
SL8 201S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
E 207S	GCTGCGTCKGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
E 2L	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTY	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
DR0T 41S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
ES	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATYAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
MIDC 710S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGRTTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
MAN 601S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATYAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
MAN 713L	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
COLO	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CRCCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
PA 4S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CRCCAGCATCAAGCC	AATAACAACAGCAGC	AACAACAACAAC	---	105
TSH	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGC---	ARYACAACAAC	---	105
KRAP 5S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACMACAGCAGC	AACAACAACAAC	---	105
KRAP 12L	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACMACAGCAGC	AACAACAACAAC	---	105
CON 20S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCAACAAAAC	CACCAGCATCAAGCC	AATAACAACAGC---	AACAACAACAAC	---	105
DEN 54L	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATYAGATTGCTT	TCCCAGCAACAAAAC	CACCARCATCAAGMC	AATAACAACAGC---	---ARCAACAAC	---	105
DEN 20S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATYAGATTGCTT	TCCCAGCAACAAAAC	CACCARCATCAAGMC	AATAACAACAGC---	ARCACAACAAC	---	105
TJ 30L	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATYAGATTGCTT	TCCCAGCACCAAAAC	CACCAGCATCAAGAC	ACTAACARCAGCAWC	AACAACAACAACAAC	---	105
TJ 29S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTGCTT	TCCCAGCATCAA---	-----GAC	AMTAACAACAGCAK	AACAACAACAACAAC	---	105
CHAM 4L	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTMATCAGATTACTT	TCCCAGCACCAA---	-----GAG	---AACAGCARCAAC	AACAACAACAAC	---	105
WED 2S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTACTT	TCCCATCACCAA---	-----GAG	-----AACAGCAGC	AACAACAACAAC	---	105
DIF	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTACTT	TCTCAGCACCAA---	-----GAG	AACAGCAACAGCAAC	AACAACAACAAC	---	105
PAN 2S	GCTGCGTCTGGGCTG	GAAAGTGTAAACAAG	CTCATCAGATTACTT	TCCCAGCACCAA---	-----GAG	-----AAC	AACAACAACAAC	---	105
SELECTION									

D16L	---AGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
F60SS	---AGTAATAAAGT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
SL8 201S	---AGTAATAACART	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
E 207S	---AGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
E 2L	---AGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCRAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
DR0T 41S	---AGTAATAAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
ES	---AGTAATAACART	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
MIDC 710S	---AGTAATAACART	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
MAN 601S	---AGTAATAACART	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
MAN 713L	---AGTAATAAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
COLO	---AGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
PA 4S	---AGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
TSH	---AGTGATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCAAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
KRAP 5S	---AGTAATAAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
KRAP 12L	---AGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
CON 20S	---AGTRATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCRAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
DEN 54L	---AGTAATAACAAC	AACACTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
DEN 20S	---AGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCGAGCTCA	AGAACAGCCTCCATG	GAGATGGACATGGAC	TGCAAGGCTGTTGCA	210
TJ 30L	AACAGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCWAGCTCA	AGAACAGCCTCCATG	GAGATGGAYATGGAC	TGCAAGGCTGTTGCR	210
TJ 29S	AACRGTAATAACAAT	AACATTAACCACCAA	TCATCATCACCTTCT	TCTTCTTCAAGCTCA	AGAACAGCCTCCATG	GAGATGGATATGGAC	TGCAAGGCTGTTGCA	210
CHAM 4L	---AGTAATAACAAT	AATACTAMCCACCCA	TCATCTTCACCTTCT	TCTTCTTCAAGCTCA	AGAACAGCCTCCATG	GAGATGGAAATGGAC	TGCAAGGCTGTTGCA	210
WED 2S	---AGTAATAACAAT	ACTAATAACCACCAA	TCA---TCACCTTCT	TCTTCTTCAAGTTCA	AGAACAGCCTCCATG	GAGATGGAAATGGAC	TGCAAGGCTGTTGCA	210
DIF	---AGTAATAACAAT	AATACTAACCACCAA	TCATCTTCACCTTCT	TCTTCTTCAAGCTCA	AGAACAGCCTCCATG	GAGATGGAAATGGAC	TGCAAGGCTGTTGCA	210
PAN 2S	---AGTAATAACAAT	AATAATAACCACCAA	TCA---TCACCTTCT	TCTTCTTCAAGCTCA	AGAACAGCCTCCATG	GAGATGGAAATAGAC	TGCAAGGCTGTTGCA	210
SELECTION								


```

D16L      CAGGTTTCCTTCAGCT GGTAAG 651
F60SS     CAGGTTTCCTTCAGCC GGTAAG 651
SL8 201S CAGGTTTCCTTCAGCY GGTAAG 651
E 207S    CAGGTTTCCTTTAGCT GGTAAG 651
E 2L      CARGTTTCCTTCAGCT GGTAAG 651
DROT 41S  CAGGTTTCCTTCAGCT GGTAAG 651
ES        CAGGTTTCCTTCAGCY GGTAAG 651
MIDC 710S CAGGTTTCCTTCAGCT GGTAAG 651
MAN 601S  CAGGTTTCCTTCAGCY GGTAAG 651
MAN 713L  CAGGTTTCCTTCAGCT GGTAAG 651
COLO      CAGGTTTCCTTCAGCT GGTAAG 651
PA 4S     CAGGTTTCCTTCAGCT GGTAAG 651
TSH       CAGGTTTCCTTCAGCT GGTAAG 651
KRAP 5S   CAGGTTTCCTTCAGCT GGTAAG 651
KRAP 12L  CAGGTTTCCTTCAGCT GGTAAG 651
CON 20S   CAGGTTTCCTTCAGCT GGTAAG 651
DEN 54L   CAGGTTTCCTTCAGCY GGTAAG 651
DEN 20S   CAGGTTTCCTTCAGCY GGTAAG 651
TJ 30L    CARGTTTCCTTCRGCT GGTAAG 651
TJ 29S    CAGGTTTCCTTCGGCT GGTAAG 651
CHAM 4L   CAGGTTTCCTTCAGCT GGTAAG 651
WED 2S    CAGGTTTCCTTCAGCT GGTAAG 651
DIF       CAGGTTTCCTTCAGCT GGTAAG 651
PAN 2S    CAGATTTCCTTCAGCT GGTAAG 651
SELECTION

```

Figure B13: *WRKY* nucleotide alignment. Alignment of DNA sequences for exon 1 of *WRKY* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. No identical sequences were identified in this alignment. Particular codon sites that were identified as positively/negatively selected by 2 or more site-by-site selection detection methods are underlined in the bottom-most sequence in the alignment. Below each underlined codon, the type of selection that was identified is indicated by a “+” or “-“, suggesting the action of positive/diversifying and negative selection, respectively. The length of the alignment (bp) is given in the right-most column (651 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
F60SS	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
SL8 201S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
E 207S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
E 2L	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
DROT 41S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
ES	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
MIDC 710S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
MAN 601S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
MAN 713L	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
COLO	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
PA 4S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
TSH	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
KRAP 5S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
KRAP 12L	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
CON 20S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
DEN 20S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
DEN 54L	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
TJ 30L	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCCAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
TJ 29S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCCAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
CHAM 4L	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GACGATGTTTTGGCT	CCCTTGAATCTAGAA	GAGATTGCCAGCCGA	TTGCCACCAAATTGC	735
WED 2S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GACGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735
DIF	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTATCCGA	TTGCCACCAAATTGC	735
PAN 2S	GATGCTTCAAATGAG	CTTGCTCTTTTCTT	GCTAGGGCTGTGATT	GATGATGTTTTGGCT	CCCTTGAATCTAGAG	GAGATTGCTAGCCGG	TTGCCACCAAATTGC	735

D16L	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
F60SS	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
SL8 201S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
E 207S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
E 2L	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
DROT 41S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
ES	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
MIDC 710S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
MAN 601S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
MAN 713L	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
COLO	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
PA 4S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
TSH	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
KRAP 5S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
KRAP 12L	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
CON 20S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
DEN 20S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
DEN 54L	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
TJ 30L	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCACTTATTGCTGCM	CGTCATGCGGGTGAA	AGG	798
TJ 29S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCACTTATTGCTGCM	CGTCATGCGGGTGAA	AGG	798
CHAM 4L	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
WED 2S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCACTTATTGCTGCC	CGTCATGCTGGTGAA	AGG	798
DIF	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCC	CGTCATGCGGGTGAA	AGG	798
PAN 2S	AGTGGGAGTGAGACT	GTGCACATGGCTAGA	TCGCTTATTGCTGCA	CGTCATGCGGGTGAA	AGG	798

Figure C1: *ECIP1* nucleotide alignment.

Alignment of DNA sequences for exon 3 of *ECIP1* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. Sequences obtained for MAN 713L and ES were found to be identical at this locus. The length of the alignment (bp) is given in the right-most column (798 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	TGTGGTTGGGCATT	GGAATGAACATGTT	GAC	663
F60SS	TGTGGTTGGGCATT	GGAATGAACATGTT	GAC	663
SL8 201S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
E 207S	TGTGGTTGGGCGTTC	GGAATGAACATGTT	GAC	663
E 2L	TGTGGTTGGGCGTTY	GGAATGAACATGTT	GAC	663
DROT 41S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
ES	TGTGGTTGGGCATT	GGAATGAACATGTT	GAC	663
MIDC 710S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
MAN 601S	TGTGGTTGGGCATT	GGAATGAACATGTT	GAC	663
MAN 713L	TGTGGTTGGGCATT	GGAATGAACATGTT	GAC	663
COLO	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
PA 4S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
TSH	TGTGGTTGGGCGTTY	GGAATGAACATGTT	GAC	663
KRAP 5S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
KRAP 12L	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
CON 20S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
DEN 20S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
DEN 54L	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
TJ 30L	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
TJ 29S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
CHAM 4L	TGTGGCTGGGCGTT	GGGATGAACATGTT	GAC	663
WED 2S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
DIF	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663
PAN 2S	TGTGGTTGGGCGTT	GGAATGAACATGTT	GAC	663

Figure C2: GAUT3 nucleotide alignment. Alignment of DNA sequences for exon 6 of *GAUT3* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) F60SS and D16L; 2) MAN 601S, MAN 713L, and ES; and 3) PA 4S, DROT 41S, and CON 20S. The length of the alignment (bp) is given in the right-most column (663 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
F60SS	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
SL8 201S	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
E 207S	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
E 2L	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
DROT 41S	AGAAACTTTGATCCS	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
ES	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
MIDC 710S	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
MAN 601S	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
MAN 713L	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
COLO	AGAAACTTTGATCCS	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
PA 4S	AGAAACTTTGATCCS	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
TSH	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
KRAP 5S	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
KRAP 12L	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
CON 20S	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
DEN 20S	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
DEN 54L	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
TJ 30L	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
TJ 29S	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
CHAM 4L	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
WED 2S	AGAAACTTTGATCCA	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
DIF	AGAAACTTTGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465
PAN 2S	AGAAACTTCGATCCG	AATGCTTGTGGATGG	GCATATGGGATGAAC	465

Figure C3: GAUT1 nucleotide alignment. Alignment of DNA sequences for exon 4 of *GAUT1* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) F60SS and D16L; 2) MAN 601S, MAN 713L, MIDC 710S, KRAP 5S, KRAP 12L, and CON 20S; 3) DEN 54L and DEN 20S; and 4) E 2L and E207S. The length of the alignment (bp) is given in the right-most column (465 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
F60SS	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
SL8 201S	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
E 207S	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
E 2L	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
DR0T 41S	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
ES	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
MIDC 710S	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
MAN 601S	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
MAN 713L	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
COLO	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
PA 4S	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
TSH	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
KRAP 5S	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
KRAP 12L	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
CON 20S	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
DEN 20S	AAACGTTATGGAGTA	GCYAAACCAAAATGGG	ATAAAGCGYAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCARTG	300
DEN 54L	AAACGTTATGGAGTA	GCCAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGACAAAAGCAGTG	300
TJ 30L	AACGTTATGGTGT	GCTAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGGCAAAAGCAATG	300
TJ 29S	AACGTTATGGAGTA	GCTAAACCAAAATGGG	ATAAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCCAAGAGAAAAGAA	AAGGCAAAAGCAATG	300
CHAM 4L	AAACGTTATGGAGTA	GCTAAACCAATGGG	GTAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCTAAGAGAAAAGAA	AAGGCAAAAGCAATG	300
WED 2S	AAACGTTATGGAGTG	GCCAAACCAAAATGGG	GTAAGCGTGAGGG	AAAAGACCTGCTGTT	GCTAAGAGAAAAGAA	AAGGCAAAAGCAATG	300
DIF	AAACGTTATGGAGTA	GCTAAACCAAAATGGG	GTAAGCGTGAGGGA	AAAAGACCTGCCGTT	GCTAAGAGAAAAGAA	AAGGCAAAAGCAATG	300
PAN 2S	AAACGTTATGGCGTA	GCTAAACCAAAATGGG	GTAAGCGTGAGGGA	AAAAGACCTGCTGTT	GCTAAGAGAAAAGAA	AAGGCAAAAGCAATG	300

Figure C4: RNABP34 nucleotide alignment. Alignment of DNA sequences for exon 6 of *RNABP34* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) F60SS and D16L; 2) MAN 601S, MAN 713L, and ES; and 3) KRAP 12L and KRAP 5S. The length of the alignment (bp) is given in the right-most column (300 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	CAACTGTGCAAAGAC	ATGGGTTGGAACCCR	AAAAGAAAGAAGGGA	TTC	TTC	261
F60SS	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
SL8 201S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
E 207S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCG	AAAAGAAAGAAGGGA	TTC	TTC	261
E 2L	CAACTGTGCAAAGAC	ATGGGTTGGAACCCR	AAAAGAAAGAAGGGA	TTC	TTC	261
DROT 41S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
ES	CAACTGTGCAAAGAC	ATGGGTTGGAACCCG	AAAAGAAAGAAGGGA	TTC	TTC	261
MIDC 710S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
MAN 601S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCG	AAAAGAAAGAAGGGA	TTC	TTC	261
MAN 713L	CAACTGTGCAAAGAC	ATGGGTTGGAACCCG	AAAAGAAAGAAGGGA	TTC	TTC	261
COLO	CAACTGTGCAAAGAC	ATGGGTTGGAACCCR	AAAAGAAAGAAGGGA	TTC	TTG	261
PA 4S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCR	AAAAGAAAGAAGGGA	TTC	TTC	261
TSH	CAACTGTGCAAAGAC	ATGGGTTGGAACCCR	AAAAGAAAGAAGGGA	TTC	TTC	261
KRAP 5S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
KRAP 12L	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
CON 20S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
DEN 20S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
DEN 54L	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	---	261
TJ 30L	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
TJ 29S	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
CHAM 4L	CAACTGTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261
PAN 2S	CAACTTTGCAAAGAC	ATGGGTTGGAACCCA	AAAAGAAAGAAGGGA	TTC	TTC	261

Figure C5: *FMO1* nucleotide alignment. Alignment of DNA sequences for exon 6 of *FMO1* obtained from 22 individuals from the genus, *Turnera*. No sequence was obtained for the individuals DIF (*T. diffusa*) and WED 2S (*T. weddelliana*). Base positions showing 100% identity across taxa are shown in blue. Sequences obtained for MAN 601S, MAN 713L, and ES were found to be identical at this locus. The length of the alignment (bp) is given in the right-most column (261 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
F60SS	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
SL8 201S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
E 207S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATMATAAAAAAT	GGRAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
E 2L	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATMATAAAAAAT	GGRAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
DROT 41S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGGAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
ES	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
MIDC 710S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
MAN 601S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
MAN 713L	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
COLO	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGRAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
PA 4S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGRAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
TSH	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATMATAAAAAAT	GGRAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
KRAP 5S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
KRAP 12L	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
CON 20S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
DEN 20S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGRAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
DEN 54L	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGRAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315
TJ 30L	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	CCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGGAAGTATGAATGT	CAGTTTTGYAGAAAA	CTTTTTGAGAAAGG	315
TJ 29S	AGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	CCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGGAAGTATGAATGT	CAGTTTTGYAGAAAA	CTTTTTGAGAAAGG	315
PAN 2S	CGAGCTCCAAAAAGG	CTTAGATGTGAAGCG	TCCCTCAGGGAAGAA	GTGATAATAAAAAAT	GGAAAGTATGAATGT	CAGTTTTGCAGAAAA	CTTTTTGAGAAAGG	315

D16L	CGTCGTTTTACGGT	CACCTCGGAAATCAC	GTAAAGGAC	354
F60SS	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
SL8 201S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
E 207S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
E 2L	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
DROT 41S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	-----	354
ES	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
MIDC 710S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
MAN 601S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
MAN 713L	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
COLO	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
PA 4S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	-----	354
TSH	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
KRAP 5S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
KRAP 12L	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
CON 20S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
DEN 20S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
DEN 54L	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
TJ 30L	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATAAAGGAC	354
TJ 29S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	-----	354
PAN 2S	CGTCGTTTTACGGT	CACCTCGGAAATCAC	ATA-----	354

Figure C6: *MBD8* nucleotide alignment. Alignment of DNA sequences for exon 3 of *MBD8* obtained from 21 individuals from the genus, *Turnera*. No sequence was obtained for the individuals DIF (*T. diffusa*), WED 2S (*T. weddelliana*), or CHAM 4L (*T. chamaedrifolia*). Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) SL8 201S and F60SS; and 2) MAN 713L and MAN 601S. The length of the alignment (bp) is given in the right-most column (354 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
F60SS	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
SL8 201S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
E 207S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTYGTTCCTCWG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
E 2L	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTYGTTCCTCWG	TTTAAAAGRAATCTC	AATTTTCAGATCACC	105
DR0T 41S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTYGTTCCTCTG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
ES	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTYGTTCCTCTG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
MIDC 710S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTYGTTCCTCWG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
MAN 601S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTAGCCAGAG	TTGCTYGTTCCTCTG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
MAN 713L	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTAGCCAGAG	TTGCTYGTTCCTCTG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
COLO	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTAGCCAGAG	TTGCTYGTTCCTCWG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
PA 4S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTYGTTCCTCTG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
TSH	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCTG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
KRAP 5S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCTG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
KRAP 12L	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCTG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
CON 20S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GATAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
DEN 20S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GAYAAGGGAAAARCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAG	TTTAAAAGRAATCTC	AATTTTCAGATCACC	105
DEN 54L	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GAYAAGGGAAAARCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAG	TTTAAAAGRAATCTC	AATTTTCAGATCACC	105
TJ 30L	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GGCAAGGGAAAAGCCA	TCCCCTACCCAGAA	TTGCTCGTTCCTCAG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
TJ 29S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GGCAAGGGAAAAGCCA	TCCCCTACCCAGAA	TTGCTCGTTCCTCAG	TTTAAAAGGAATCTC	AATTTTCAGATCACC	105
CHAM 4L	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GACAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAG	TTTAAAAGGAACCTG	AMTTTTCAGATCACC	105
WED 2S	TACATCCCTCCACAC	AAGCGCCATTCAAAG	GACAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAA	TTTAAAAGGAATCTG	AATTTTCAGATCACC	105
DIF	TACATCCCTCCACAC	AAGCGCCATTCAAAMG	GACAAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTTCCTCAG	TTTAAAAGGAACCTG	AATTTTCAGATCACC	105
PAN 2S	TACATCCCTCCACAC	AAGCGCCATTCAAAT	GACGAGGGAAAAGCCA	TCCCCTACCCAGAG	TTGCTCGTCCGTCTG	TTTAAAAGGAATCTG	AMTTTTCAGATCACC	105

D16L	AAACCTGGTGTAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGGTGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
F60SS	AAACCTGGTGTAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGGTGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
SL8 201S	AAACCTGGTGTAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGGTGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
E 207S	AAACCTGGTGTAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGRTGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
E 2L	AAACCTGGTRTAGAT	AGAAGYGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGATGG	TTTGCTGTTGGYTTG	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
DR0T 41S	AAACCTGGTGTAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGRTGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
ES	AAACCTGGTATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGATGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
MIDC 710S	AAACCTGGTGTAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGRTGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
MAN 601S	AAACCTGGTATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGATGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
MAN 713L	AAACCTGGTATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGATGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
COLO	AAACCTGGTGTAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGRTGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
PA 4S	AAACCTGGTRTAGAT	AGAAGTGGAAAGATT	ATCTATGMAGACAAT	GCCATTAGTCGRTGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
TSH	AAACCTGGTGTAGAT	AGAAGYGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGATGG	TTTGCTGTTGGYTTG	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
KRAP 5S	AAACCTGGTATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGATGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
KRAP 12L	AAACCTGGTATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGATGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
CON 20S	AAACCTGGTATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTAGTCGATGG	TTTGGTGTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
DEN 20S	AAACCTGRTATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTWSTCRATGG	TTTGSTGTTGGT	GATGWCAATGRCCAG	ATTCCACCTCATGTT	210
DEN 54L	AAACCTGRTATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACAAT	GCCATTWSTCRATGG	TTTGSTGTTGGT	GATGWCAATGRCCAG	ATTCCACCTCATGTT	210
TJ 30L	AAACCTGGTATAGAT	AGAAGTGGTAAGATT	ATCTATGCAGACAAT	GCCATTACTCGATGG	TTTGCAGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
TJ 29S	AAACCTGGTATAGAT	AGAAGTGGTAAGATT	ATCTATGCAGACAAT	GCCATTACTCGATGG	TTTGCAGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
CHAM 4L	AAACCTGATAYAGAT	AGAAGTGGAAAGATT	WTCTATGCAGACCAT	GCCATTACTCGATGG	TTTGCTGTTGGT	GATRACAAATGACCAG	ATTCCACCTCATGTT	210
WED 2S	AAACCTGATGTAGAT	AGAAGTGGAAAGATT	TTCTATGCAGACCAT	GCCATTACTCAATGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
DIF	AAACCTGCAACAGAT	AGAAGTGGAAAGATT	WYCTAYGCAGACCAT	GCCATTACTCGATGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210
PAN 2S	AAACCTGATATAGAT	AGAAGTGGAAAGATT	ATCTATGCAGACCAT	GCCATTACTCGATGG	TTTGCTGTTGGT	GATGACAAATGACCAG	ATTCCACCTCATGTT	210

D16L	AATCTTGAGCCTGTT	TCTGTGGAA	234
F60SS	AATCTTGAGCCTGTT	TCTGTGGAA	234
SL8 201S	AATCTTGAGCCTGTT	TCTGTGGAA	234
E 207S	AATCTTGAGCCTGTT	TCTGTGGAA	234
E 2L	AATCTTGAGCCTGTT	TCTGTGGAA	234
DROT 41S	AATCTTGAGCCTGTT	TCTGTGGAA	234
ES	AATCTTGAGCCTGTT	TCTGTGGAA	234
MIDC 710S	AATCTTGAGCCTGTT	TCTGTGGAA	234
MAN 601S	AATCTTGAGCCTGTT	TCTGTGGAA	234
MAN 713L	AATCTTGAGCCTGTT	TCTGTGGAA	234
COLO	AATCTTGAGCCTGTT	TCTGTGGAA	234
PA 4S	AATCTTGAGCCTGTT	TCTGTGGAA	234
TSH	AATCTTGAGCCTGTT	TCTGTGGAA	234
KRAP 5S	AGTCTTGAGCCTGTT	TCTGTGGAA	234
KRAP 12L	AGTCTTGAGCCTGTT	TCTGTGGAA	234
CON 20S	AATCTTGAGCCTGTT	TCTGTGGAA	234
DEN 20S	AATCTTGAGCCTGTT	TCTGTGGAA	234
DEN 54L	AATCTTGAGCCTGTT	TCTGTGGAA	234
TJ 30L	AGTCTTGAGCCTGTT	TCTGTGGAA	234
TJ 29S	AGTCTTGAGCCTGTT	TCTGTGGAA	234
CHAM 4L	AMTCTTGAGCCTGTT	TCTGTGGAA	234
WED 2S	AATCTTGAGCCTGTT	TCTGTGGAA	234
DIF	AMTCTTGAGCCTGTT	TCTGTGGAA	234
PAN 2S	AATCTTGAGCCTGTT	TCTGTGGAA	234

Figure C7: *UNKN* nucleotide alignment. Alignment of DNA sequences for exon 1 of *UNKN* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) F60SS, D16L, and SL8 201S; 2) MAN 601S and MAN 713L; and 3) TJ 29S and TJ 30L. The length of the alignment (bp) is given in the right-most column (234 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

D16L	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
F60SS	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
SL8 201S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
E 207S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
E 2L	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
DROT 41S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
ES	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
MIDC 710S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
MAN 601S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
MAN 713L	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
COLO	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
PA 4S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
TSH	TTCGGATTTTCTTCT	CGAATGGCTGCACTM	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
KRAP 5S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
KRAP 12L	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
CON 20S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
DEN 20S	TTCGGATTTTSTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
DEN 54L	TTCGGATTTTSTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
TJ 30L	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
TJ 29S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
CHAM 4L	TTCAGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
WED 2S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAGAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
DIF	TTCGGGTTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291
PAN 2S	TTCGGATTTTCTTCT	CGAATGGCTGCACTA	GACTTTATTGTTTGT	GATGAAAGTGATGTT	TTTGTCACTAACAAAC	AACGGC	291

Figure C8: *POFUT* nucleotide alignment. Alignment of DNA sequences for exon 9 of *POFUT* obtained from 24 individuals from the genus, *Turnera*. Base positions showing 100% identity across taxa are shown in blue. Groups of individuals with identical sequences at this locus are as follows: 1) F60SS, D16L, SL8 201S, PA 4S, and KRAP 12L; 2) MAN 713L and DROT 41S; and 3) COLO and MAN 601S. The length of the alignment (bp) is given in the right-most column (291 bp, total). The alignment is sectioned into groups of 15 bases (or 5 codons).

Appendix D: Amino Acid-Based Phylogenies for Concatenated Total Data Alignments and Total *Tsst1* Data Alignment

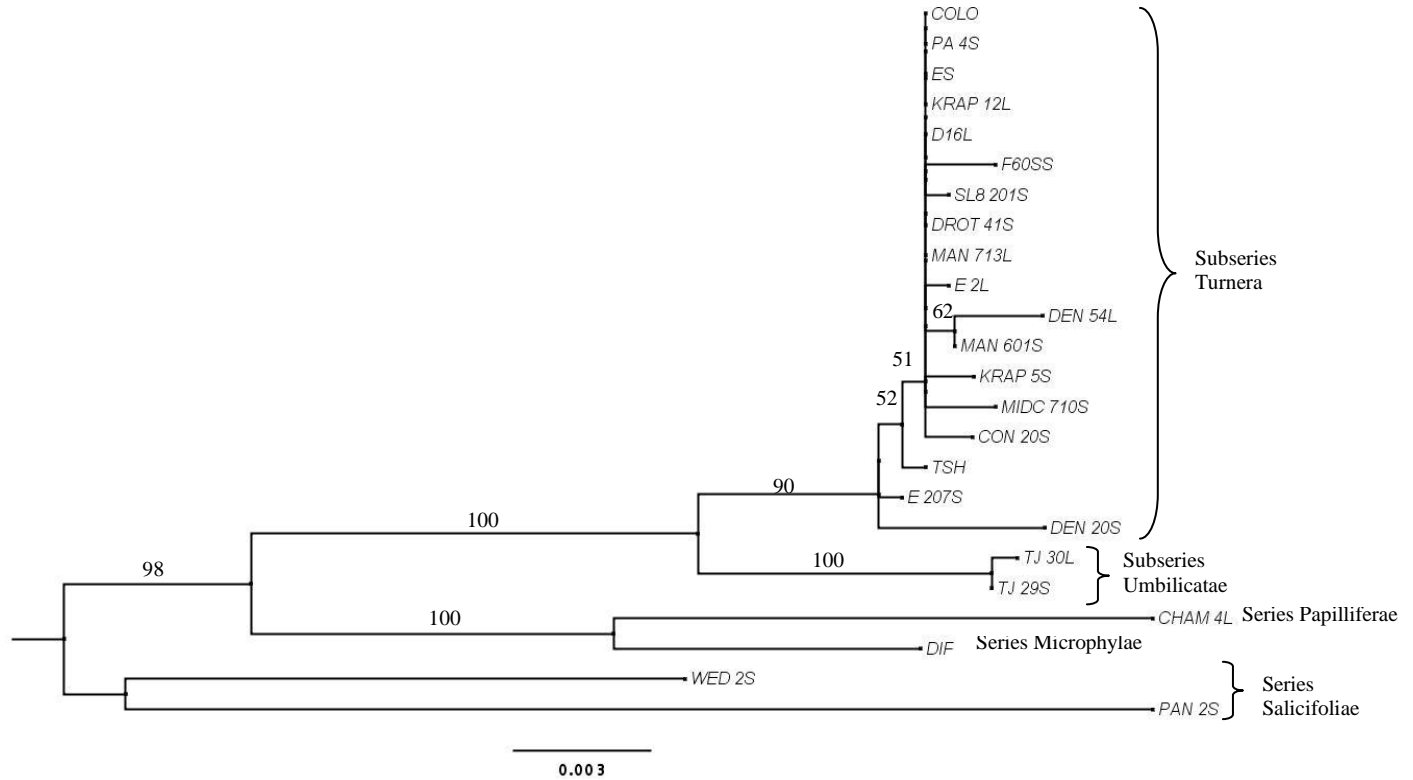


Figure D1: Molecular Phylogenetic analysis by Maximum Likelihood method using total amino acid data for all genes. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-7198.898) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.132)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 1976 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). Series/Subseries membership is indicated to the right of the tree. Note that subseries Turnera and subseries Umbilicatae both belong to series Turnera.

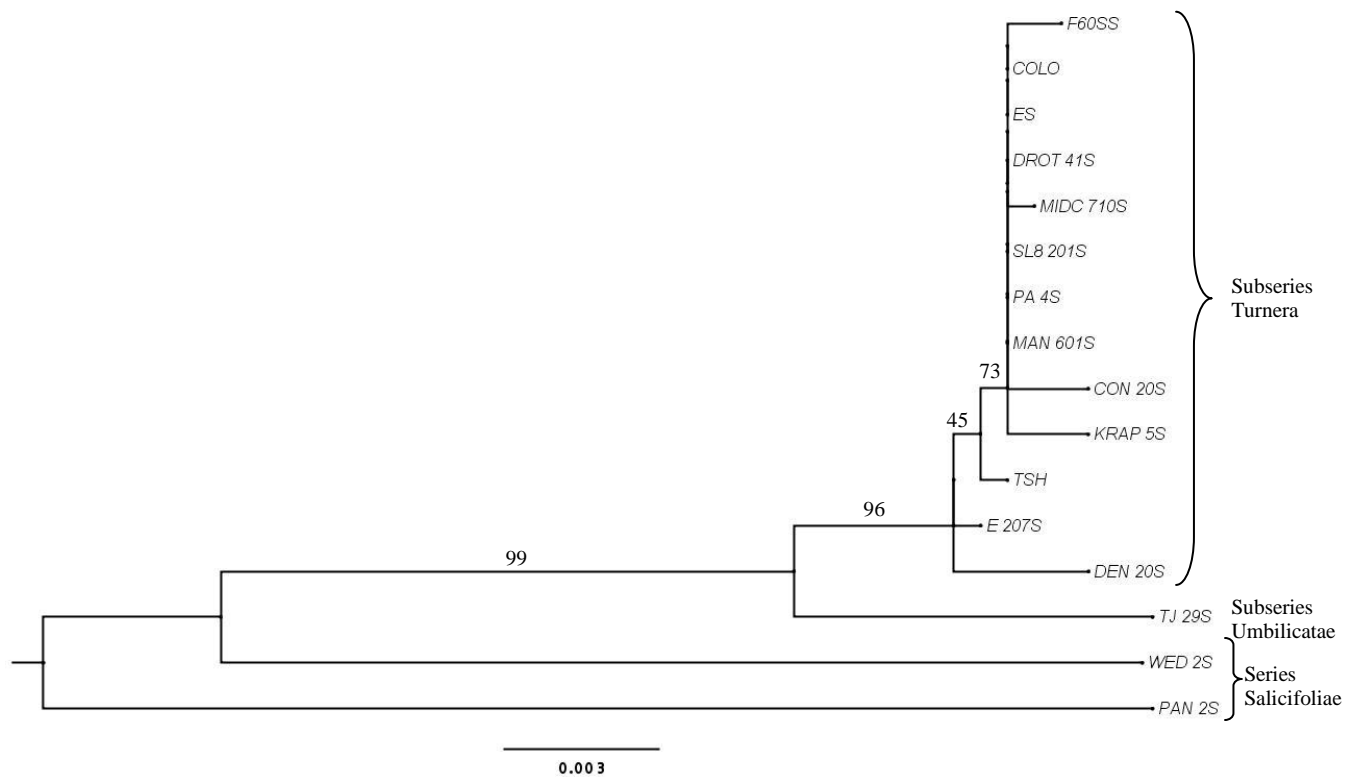


Figure D2: Molecular Phylogenetic analysis by Maximum Likelihood method using total amino acid data for all genes and a reduced number of taxa. This tree was computed specifically for analyses of *Tsstal* in HYPHY. The data used to compute this phylogeny is identical to that which was used to produce the tree in Figure 7, except that data from long-styled individuals was excluded. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-6595.560) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.116)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 amino acid sequences. The coding data was translated assuming a standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 1922 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). Series/Subseries membership is indicated to the right of the tree. Note that subseries Turnera and subseries Umbilicatae both belong to series Turnera.

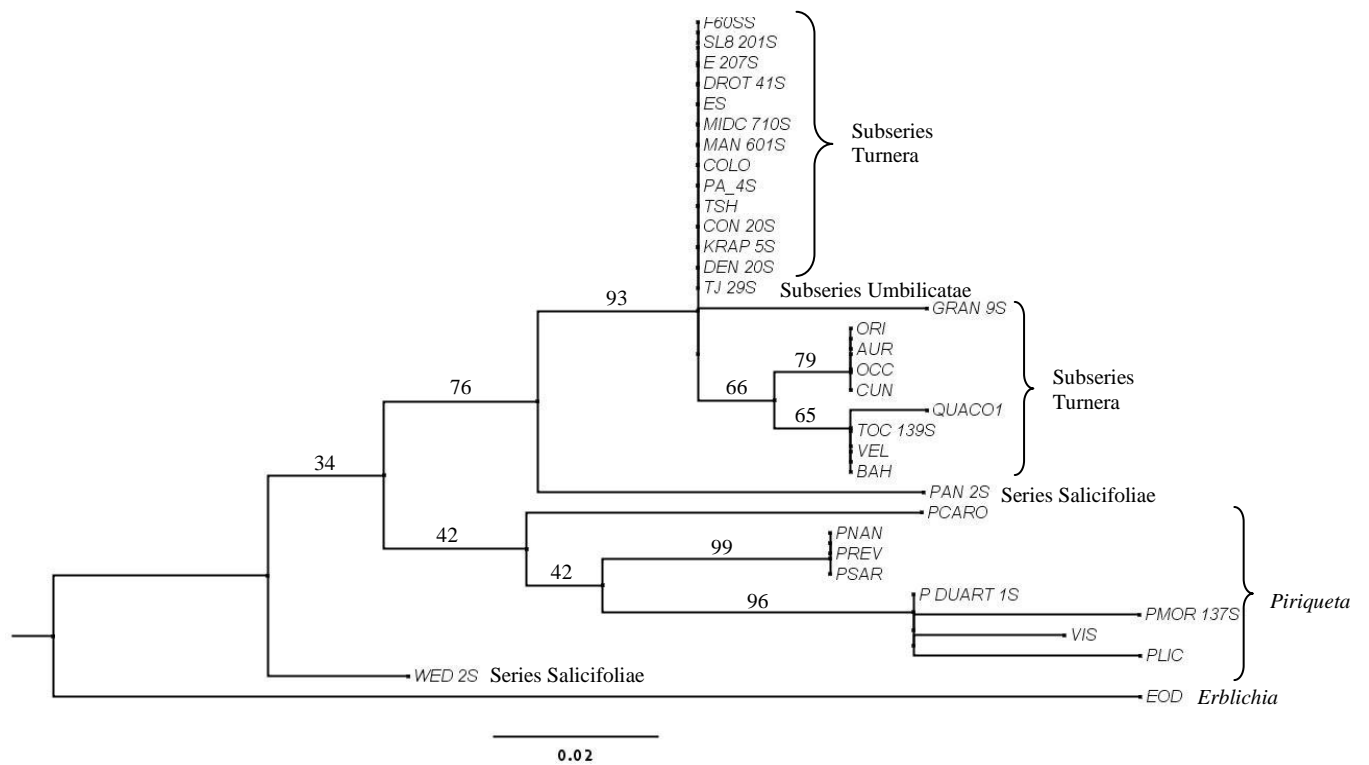


Figure D3: Molecular Phylogenetic analysis by Maximum Likelihood method using total amino acid data for *Tssta1*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-717.946) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 34 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 116 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013). Series/Subseries/Genus membership is indicated to the right of the tree. Note that subseries Turnera and subseries Umbilicatae both belong to series Turnera

Appendix E: DNA- and Amino Acid-Based Phylogenies for all Individual Genes and for Concatenated alignments of all S-linked and all Control Data

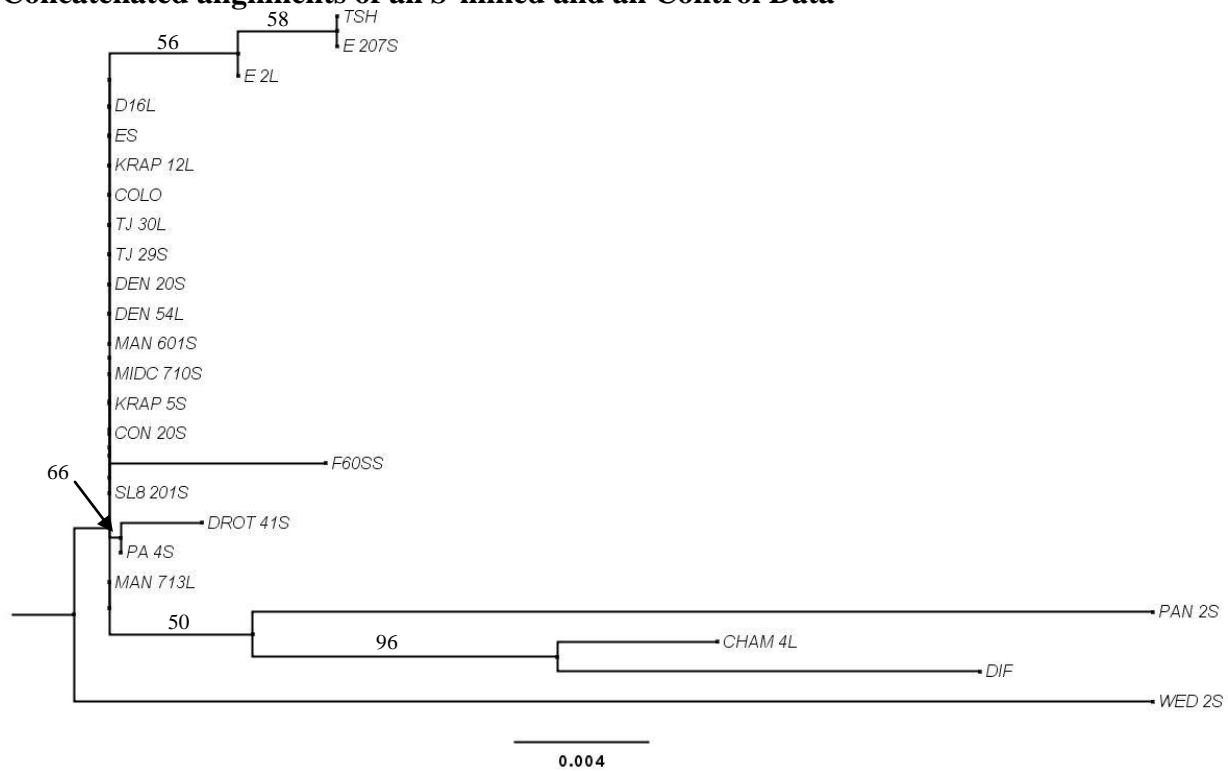


Figure E1: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *APETALA2*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-690.933) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 368 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al 2013).

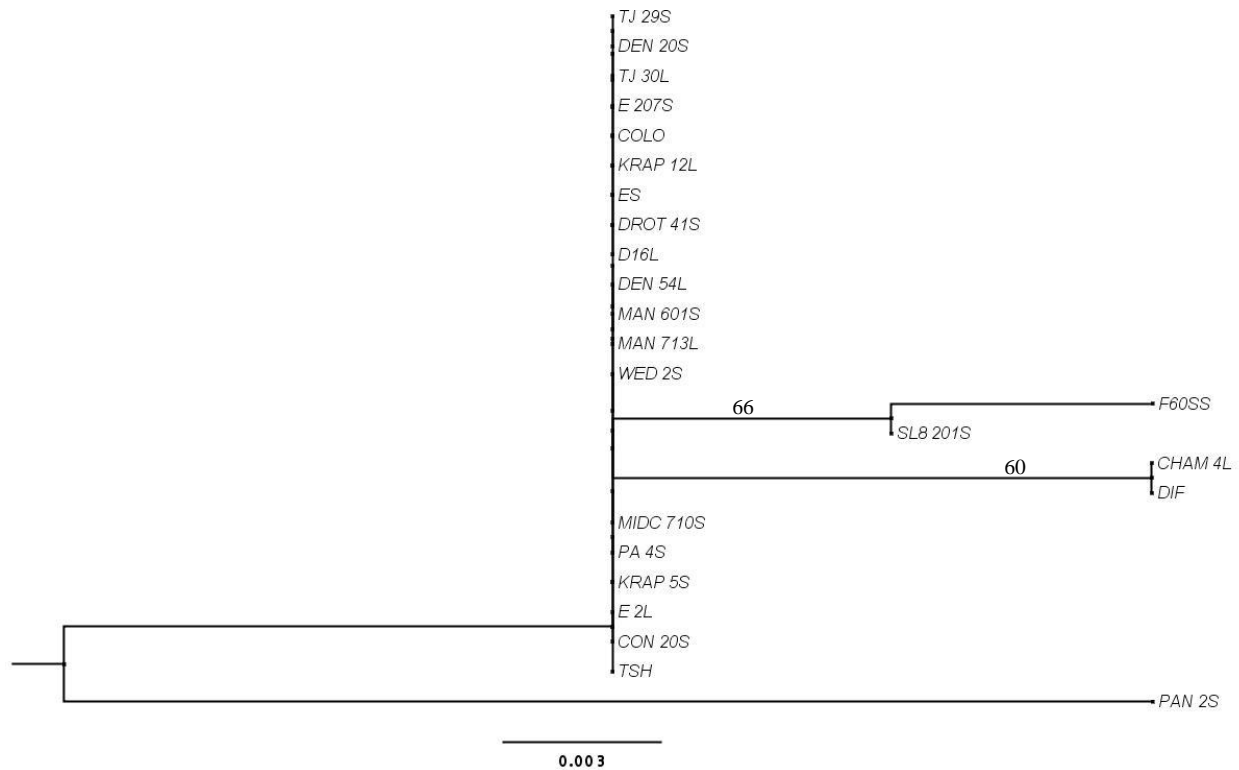


Figure E2: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *APETALA2*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the WAG amino acid substitution model (Whelan and Goldman 2001). The tree with the highest log likelihood (-316.340) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 99 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

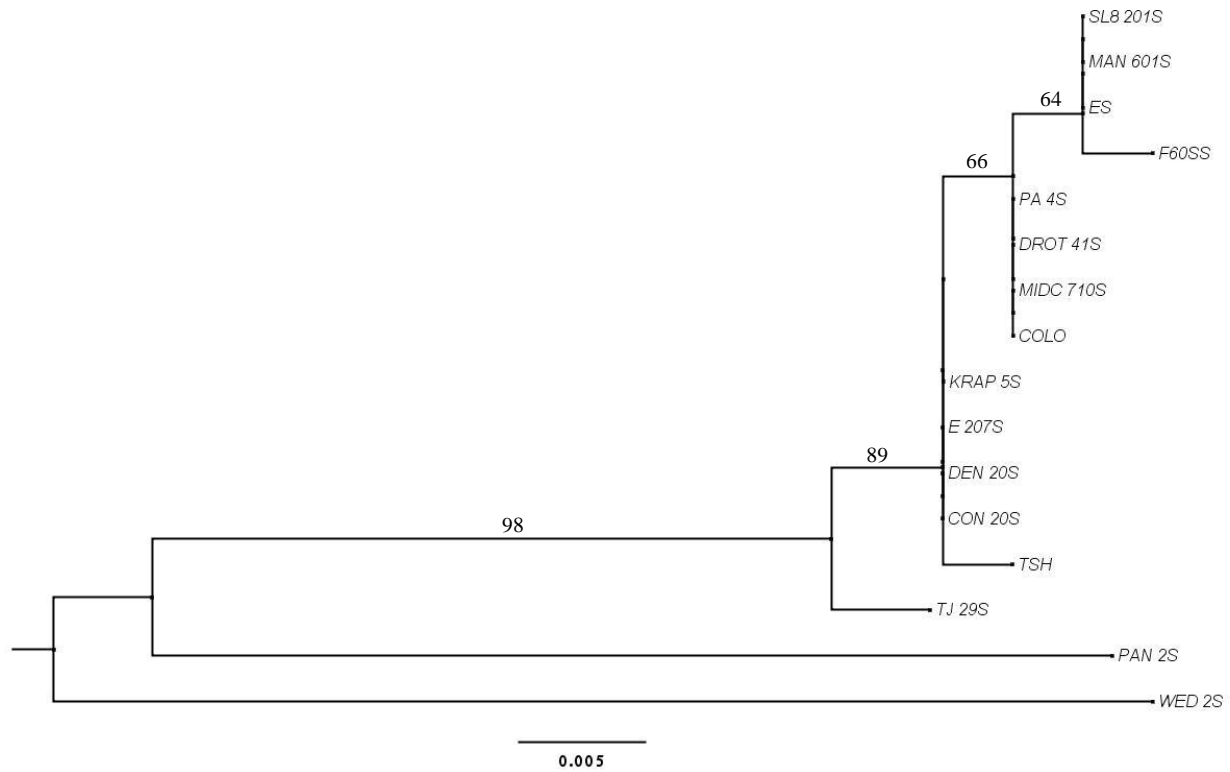


Figure E3: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *Tssta1*.

Evolutionary relationships were inferred using the Maximum Likelihood method based on the Kimura 2-parameter nucleotide substitution model (1980). The tree with the highest log likelihood (-768.753) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 369 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

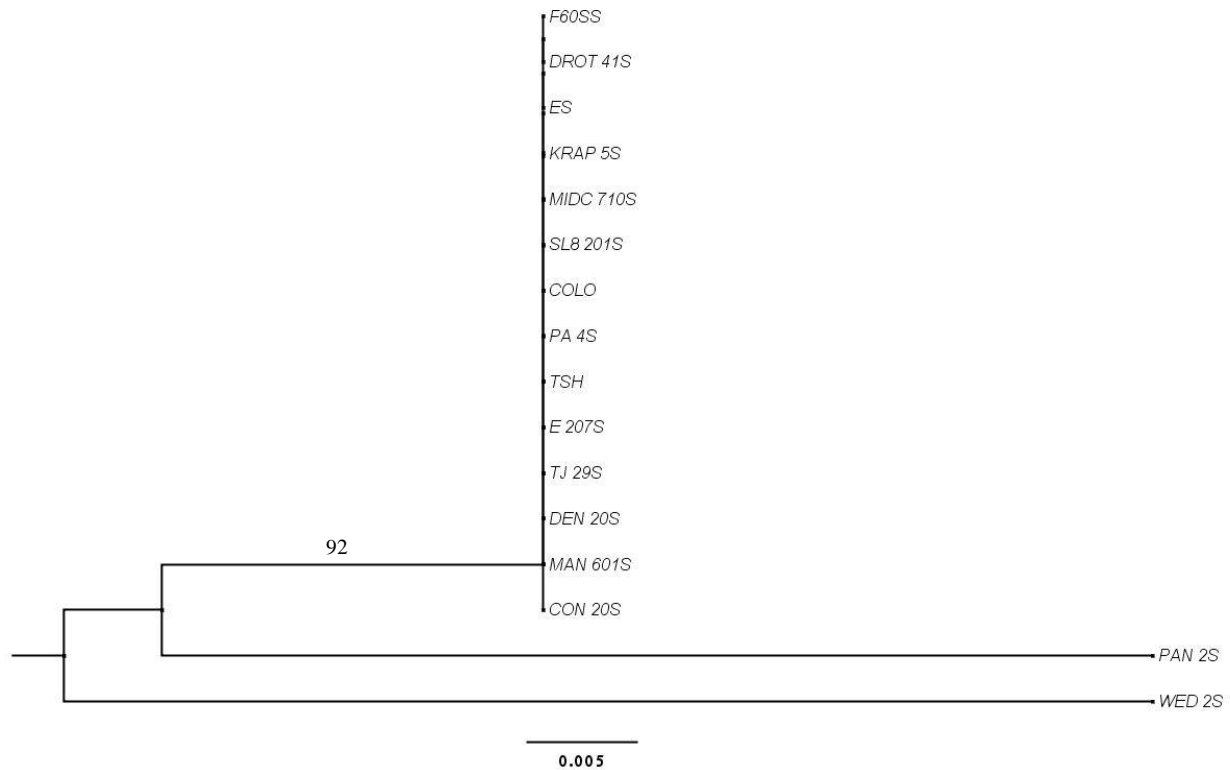


Figure E4: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *Tssta1*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-452.768) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 16 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 123 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

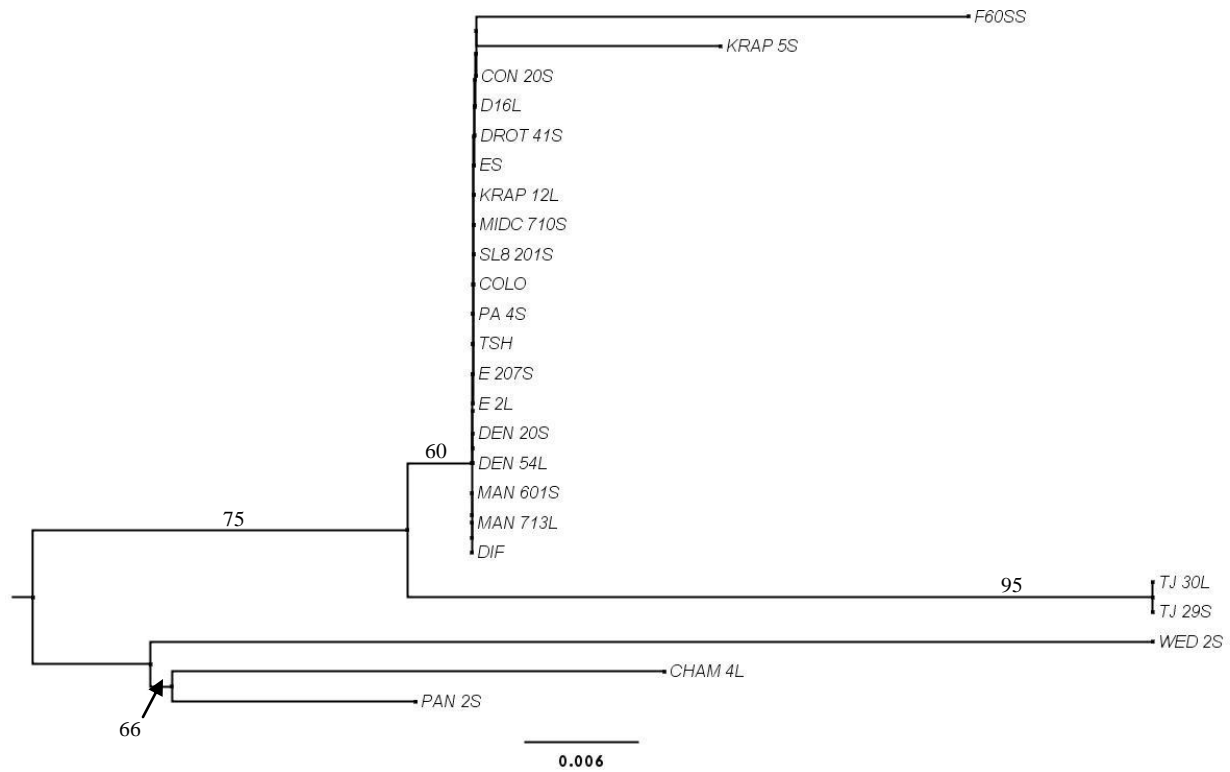


Figure E5: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *LEJ2*.

Evolutionary relationships were inferred using the Maximum Likelihood method based on the Jukes-Cantor nucleotide substitution model (1969). The tree with the highest log likelihood (-193.649) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 79 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

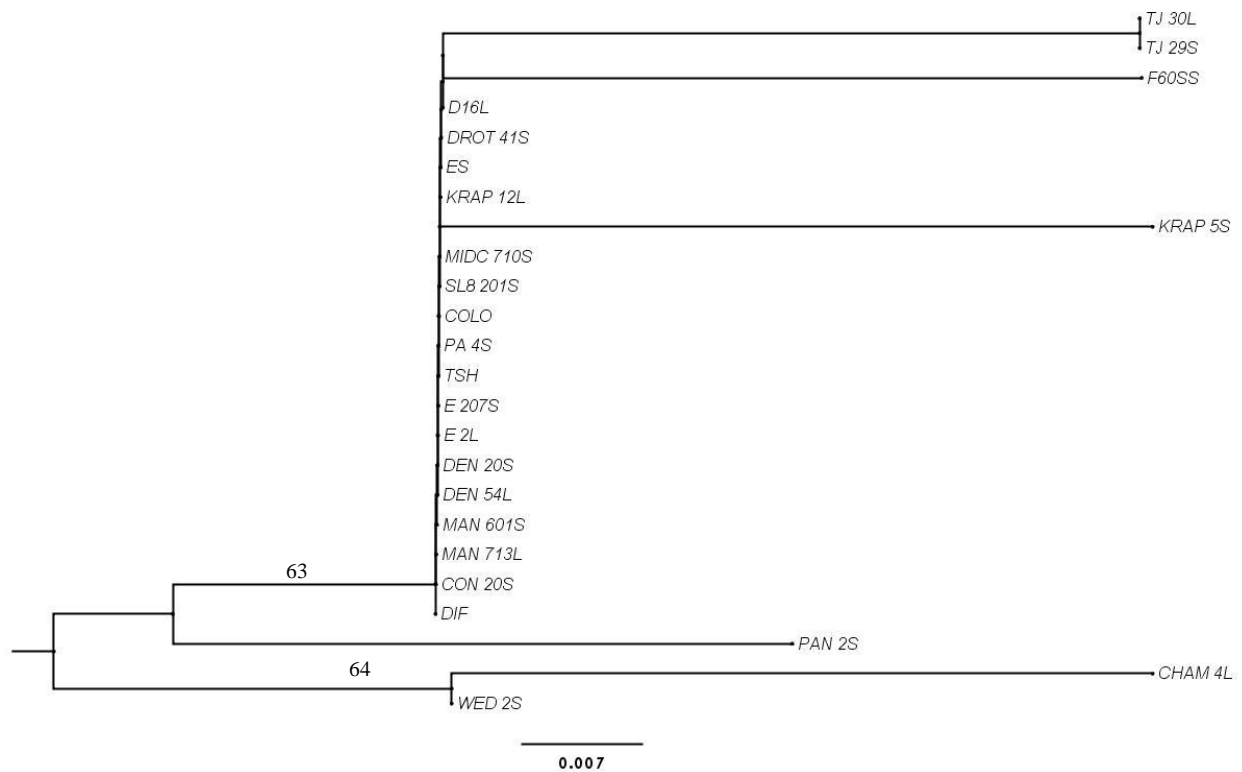


Figure E6: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *LEJ2*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-101.548) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 23 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

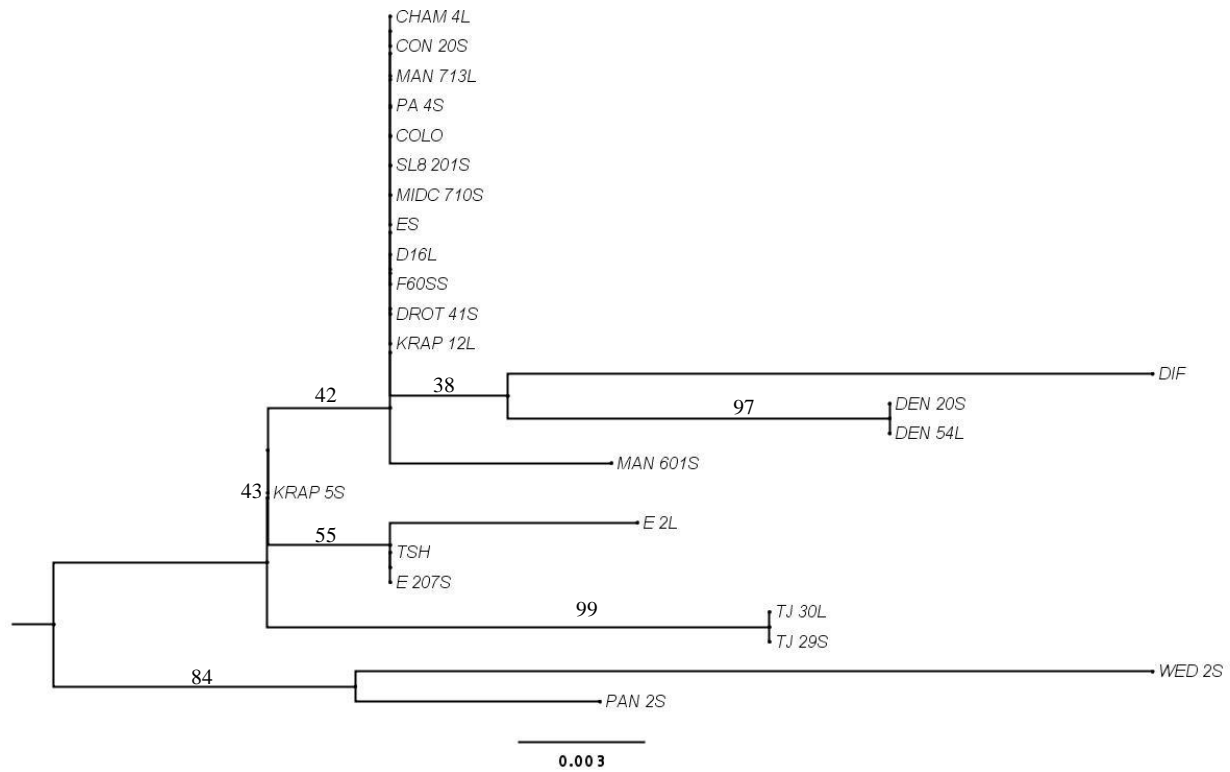


Figure E7: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *AP2D*.

Evolutionary relationships were inferred using the Maximum Likelihood method based on the Jukes-Cantor nucleotide substitution model (1969). The tree with the highest log likelihood (-682.672) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.121)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 349 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013)

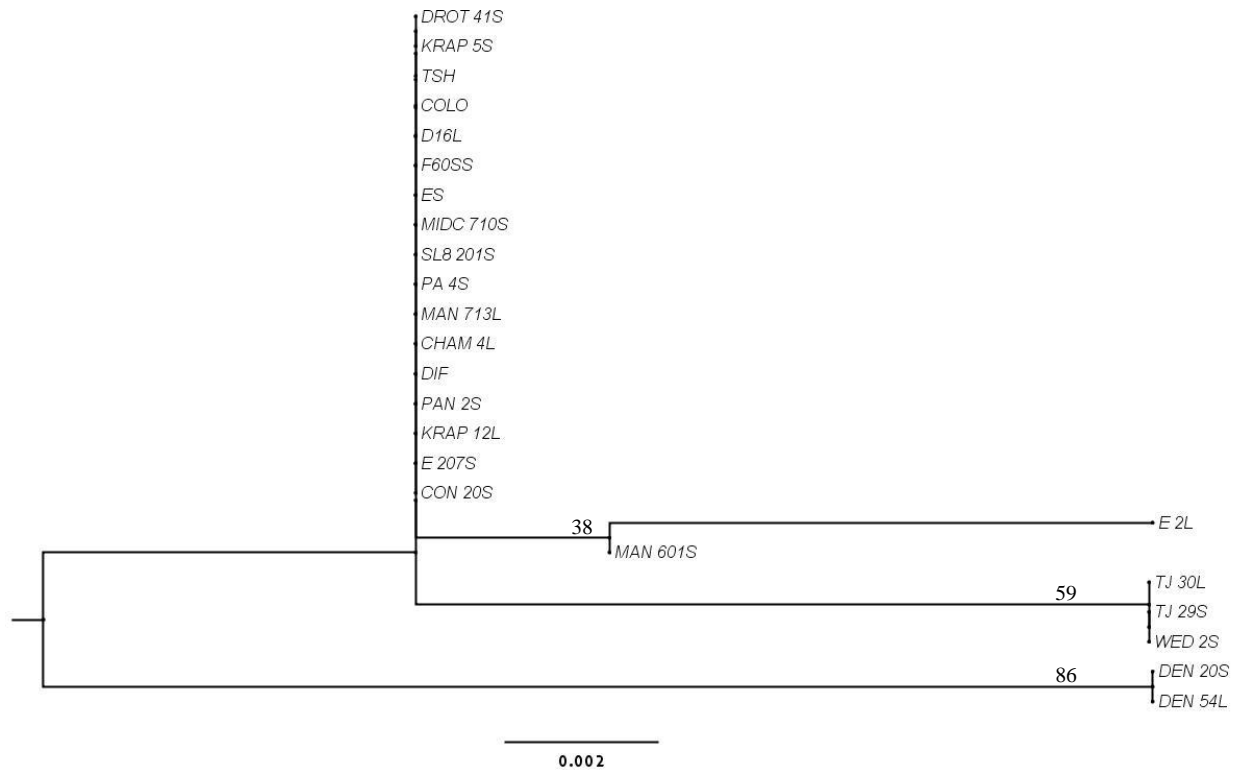


FIGURE E8: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *AP2D*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-328.268) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 104 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

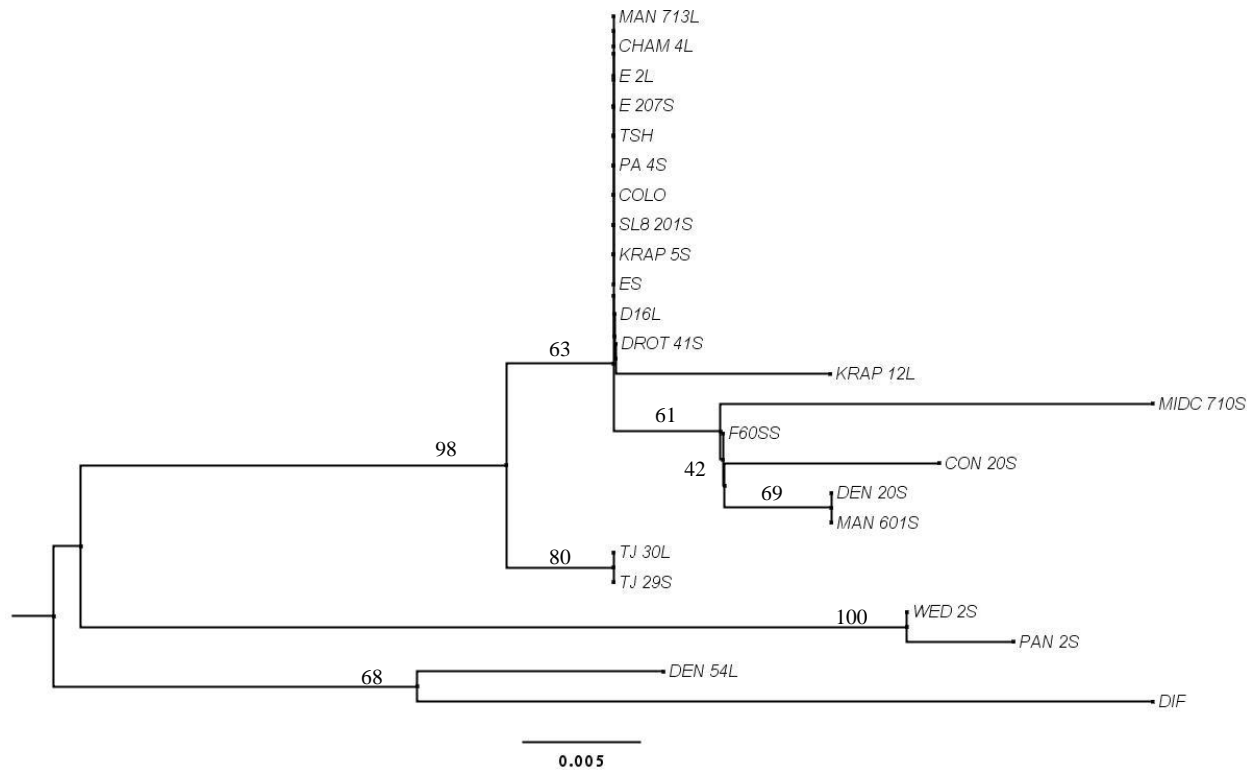


FIGURE E9: Molecular Phylogenetic analysis by Maximum Likelihood method using amino DNA for *RNABP*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Jukes-Cantor nucleotide substitution model (1969). The tree with the highest log likelihood (-536.459) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 223 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

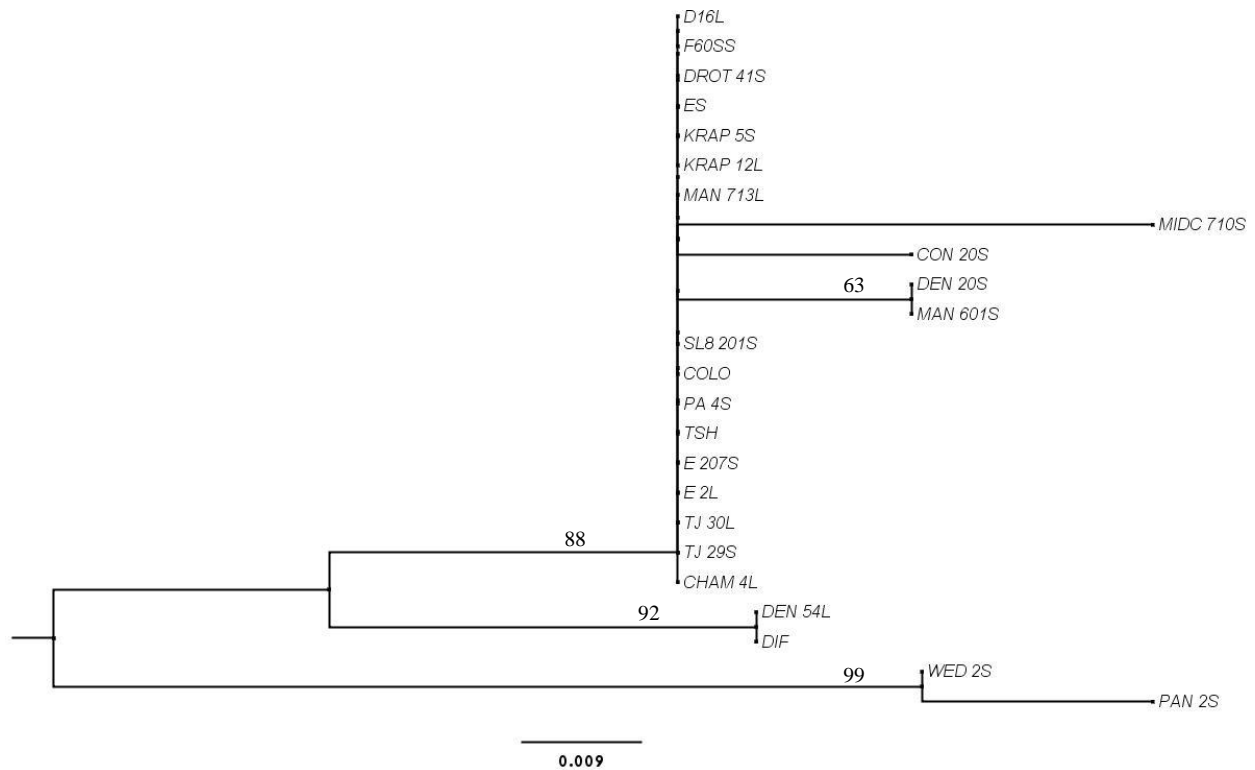


FIGURE E10: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *RNABP*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-263.017) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 61 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

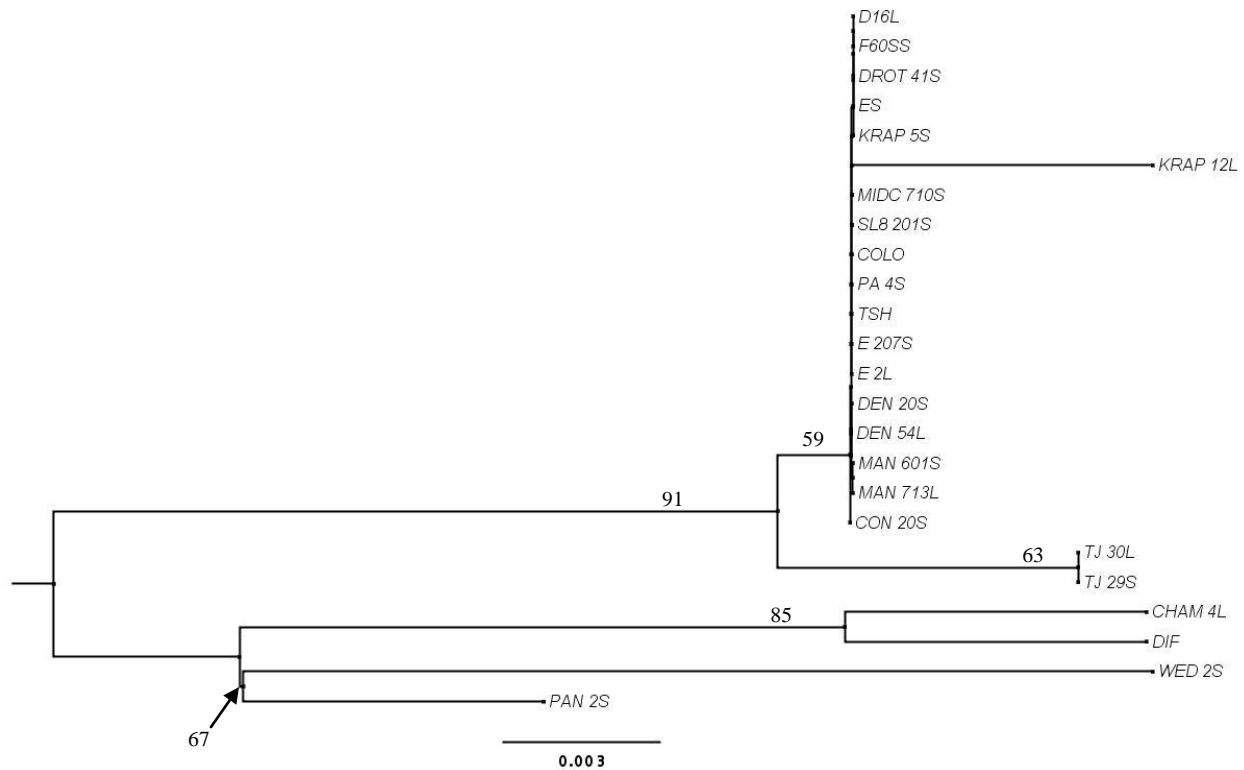


FIGURE E11: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *SCE1*.

Evolutionary relationships were inferred using the Maximum Likelihood method based on the Jukes-Cantor nucleotide substitution model (1969). The tree with the highest log likelihood (-330.698) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 176 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

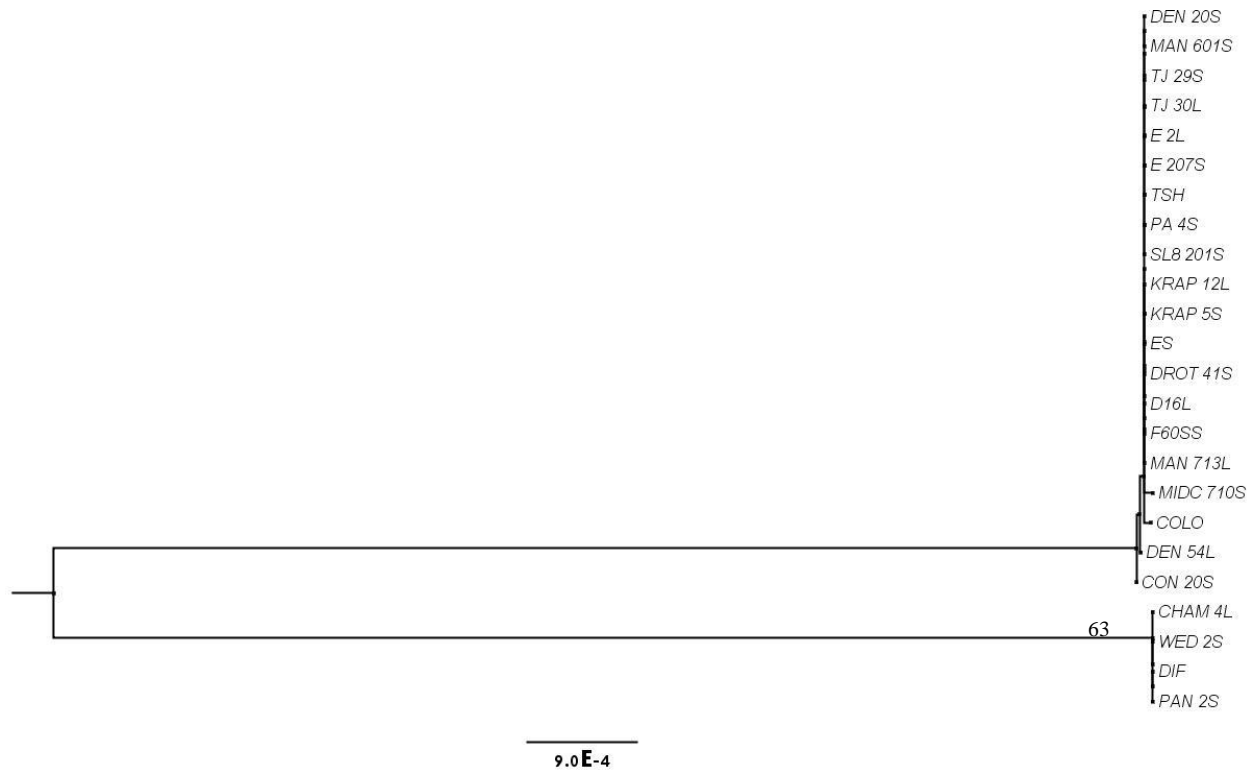


FIGURE E12: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *SCEI*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the General Reverse Transcriptase (RTREV) amino acid substitution model (Dimmic, Rest, and Mindell 2002). The tree with the highest log likelihood (-174.265) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 58 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

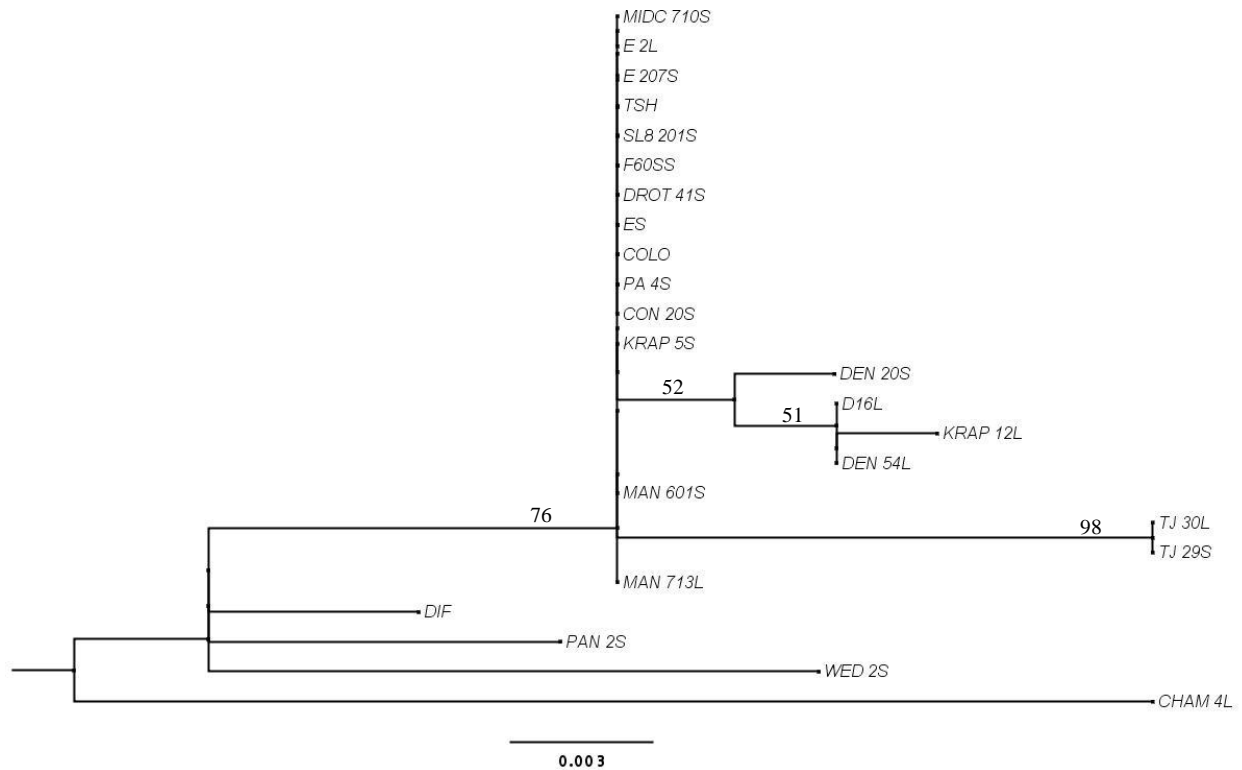


FIGURE E13: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *FRAI*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-853.086) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 478 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

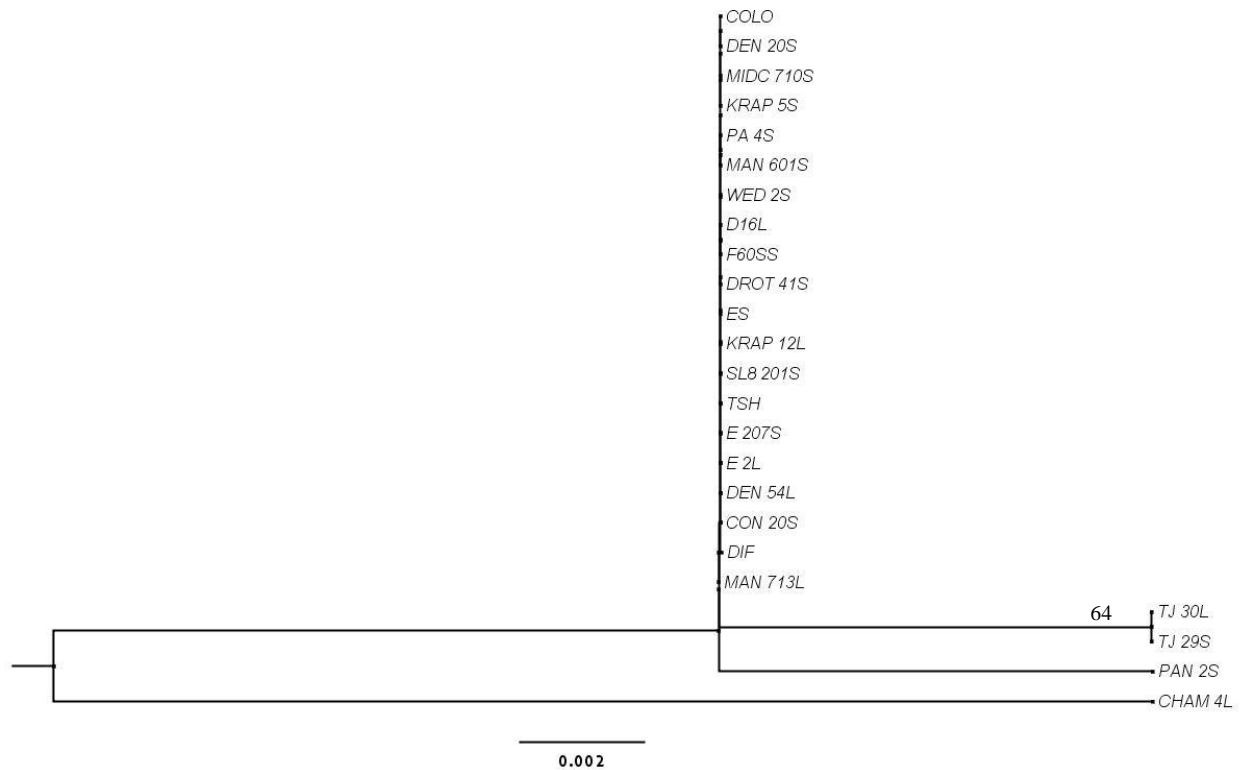


FIGURE E14: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *FRAI*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the General Reverse Transcriptase (RTREV) amino acid substitution model (Dimmic, Rest, and Mindell 2002). The tree with the highest log likelihood (-468.914) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 147 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

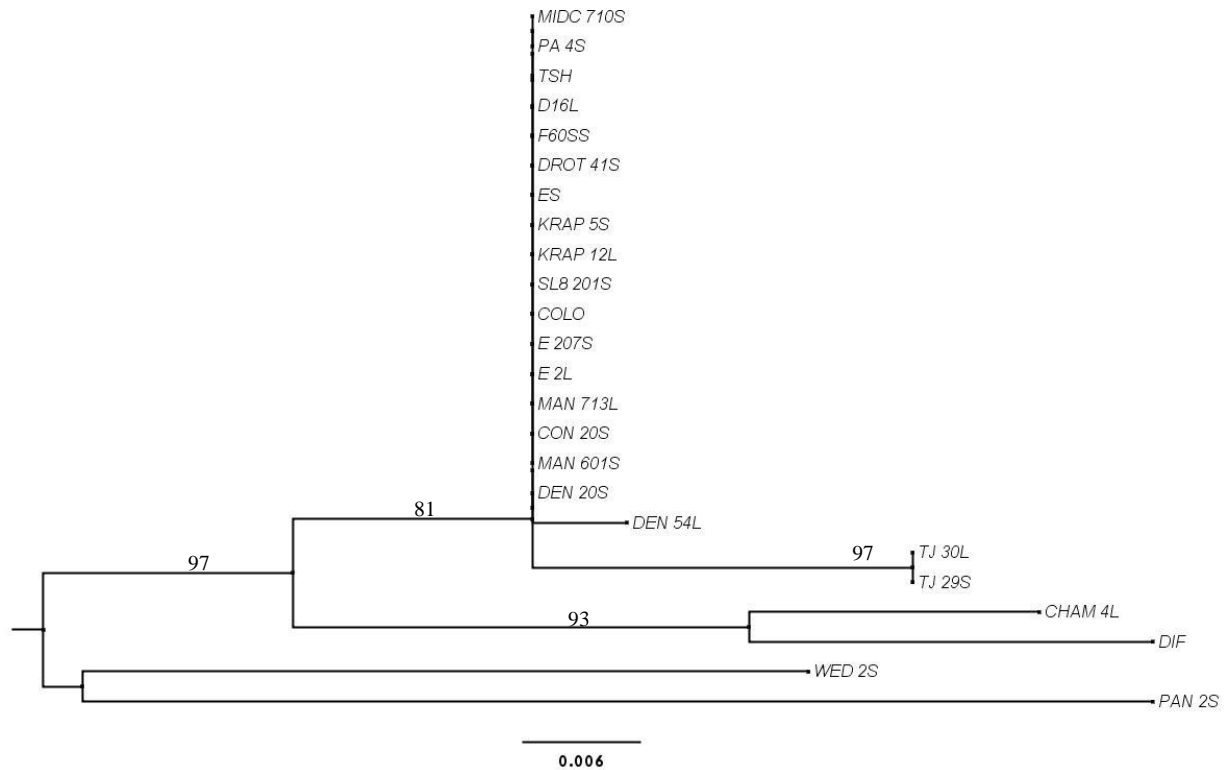


FIGURE E15: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *LRRK*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Jukes-Cantor nucleotide substitution model (1969). The tree with the highest log likelihood (-514.056) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 211 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

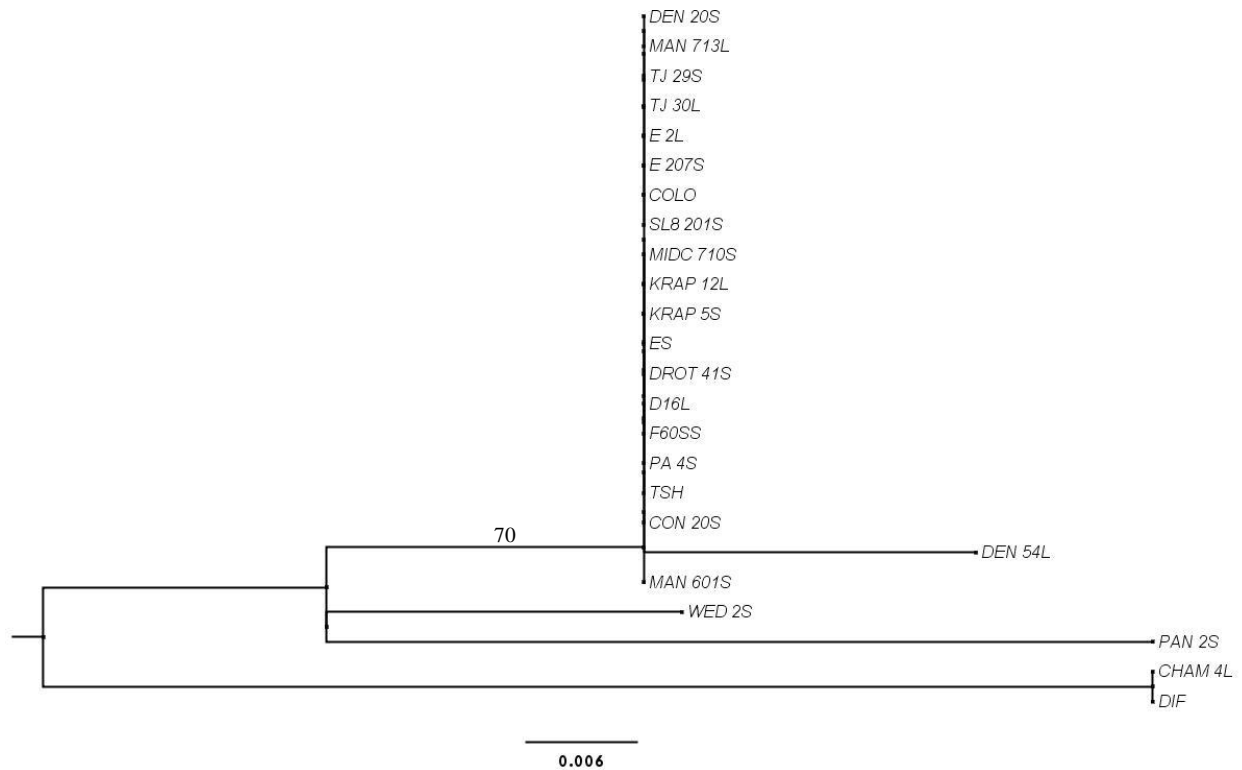


FIGURE E16: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *LRRK*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-213.401) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 55 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

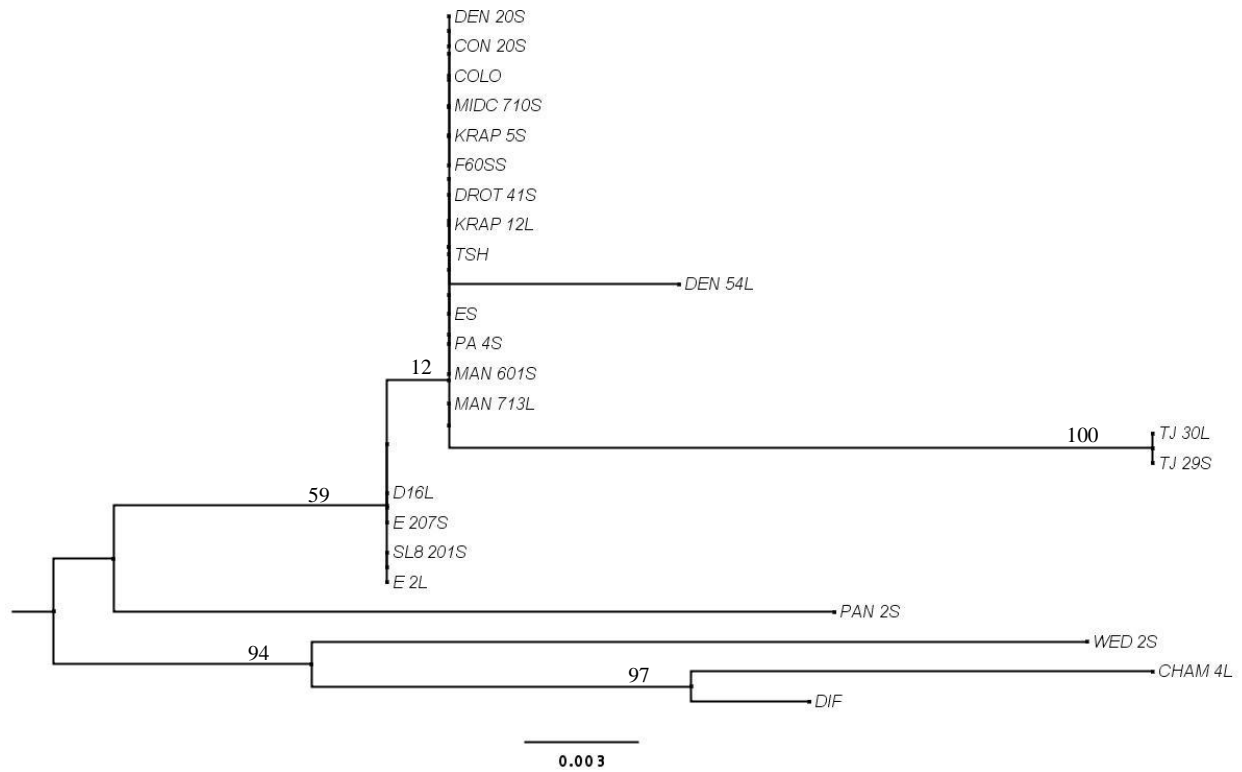


FIGURE E17: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *IRX15L*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-1409.341) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.077)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 746 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

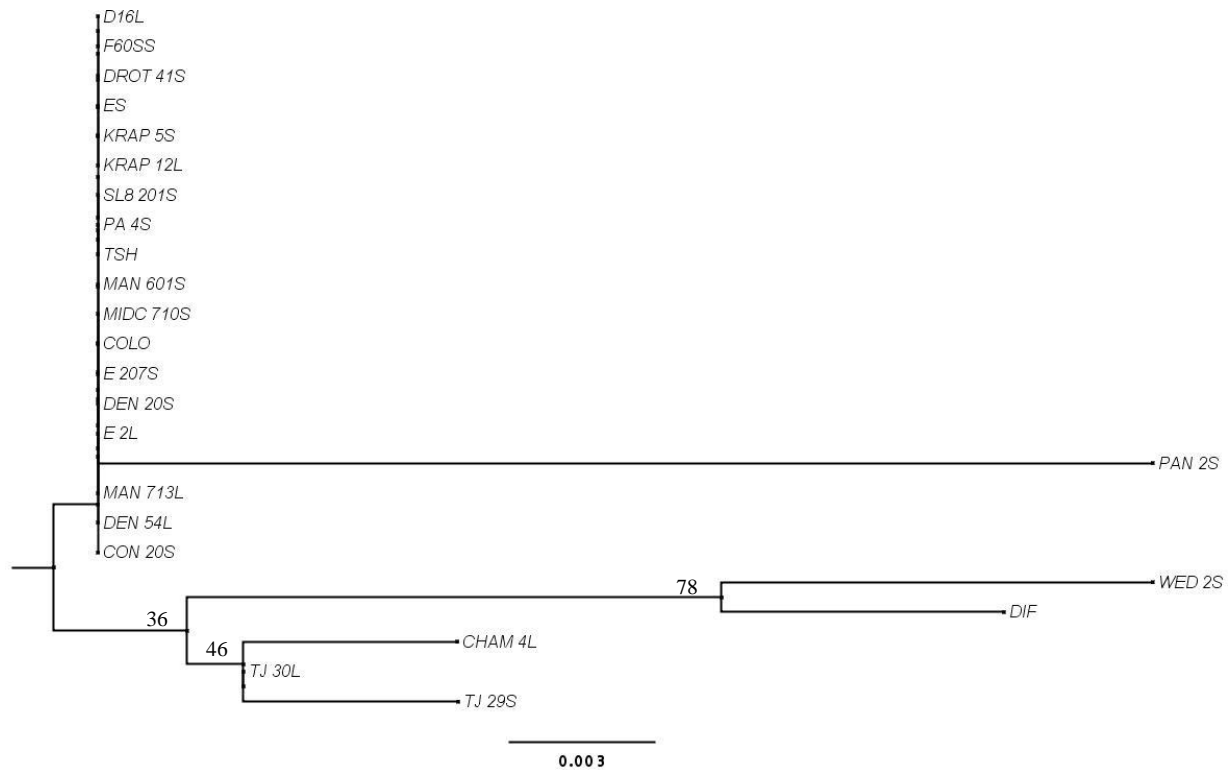


FIGURE E18: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *IRX15L*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-784.253) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 232 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

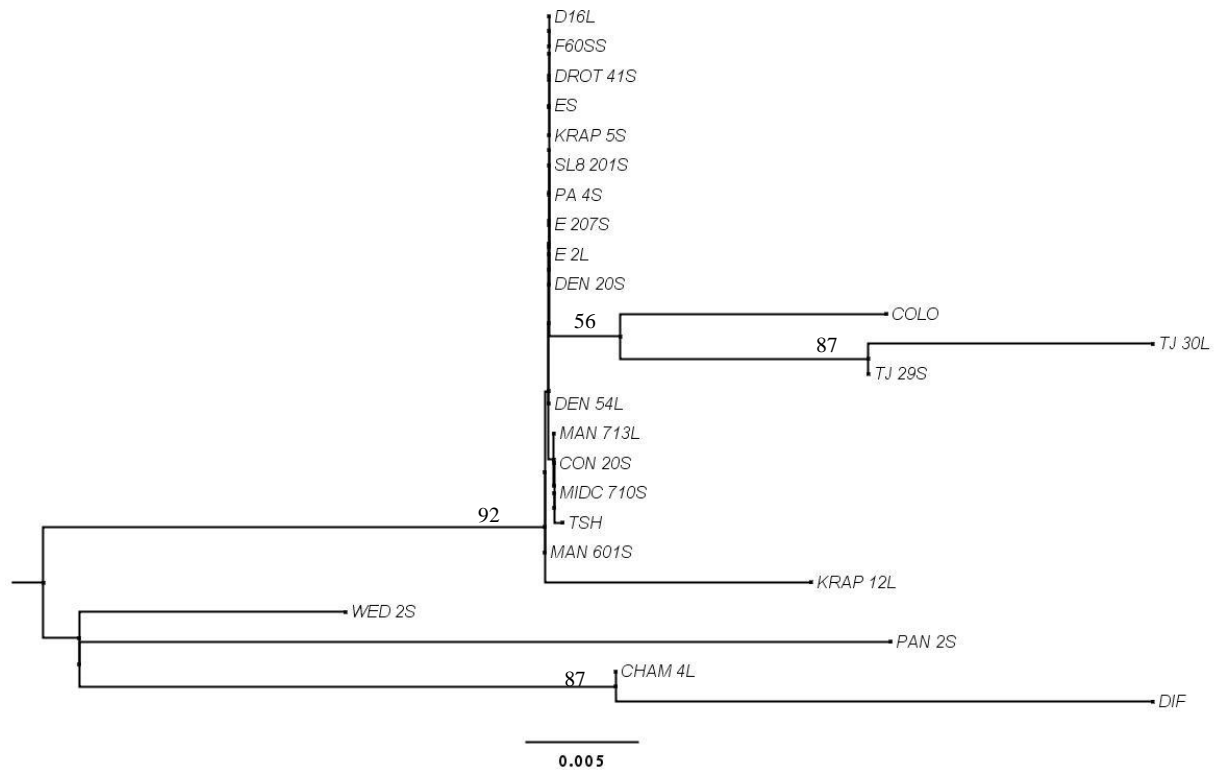


FIGURE E19: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *FSP*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Jukes-Cantor nucleotide substitution model (1969). The tree with the highest log likelihood (-202.304) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 86 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

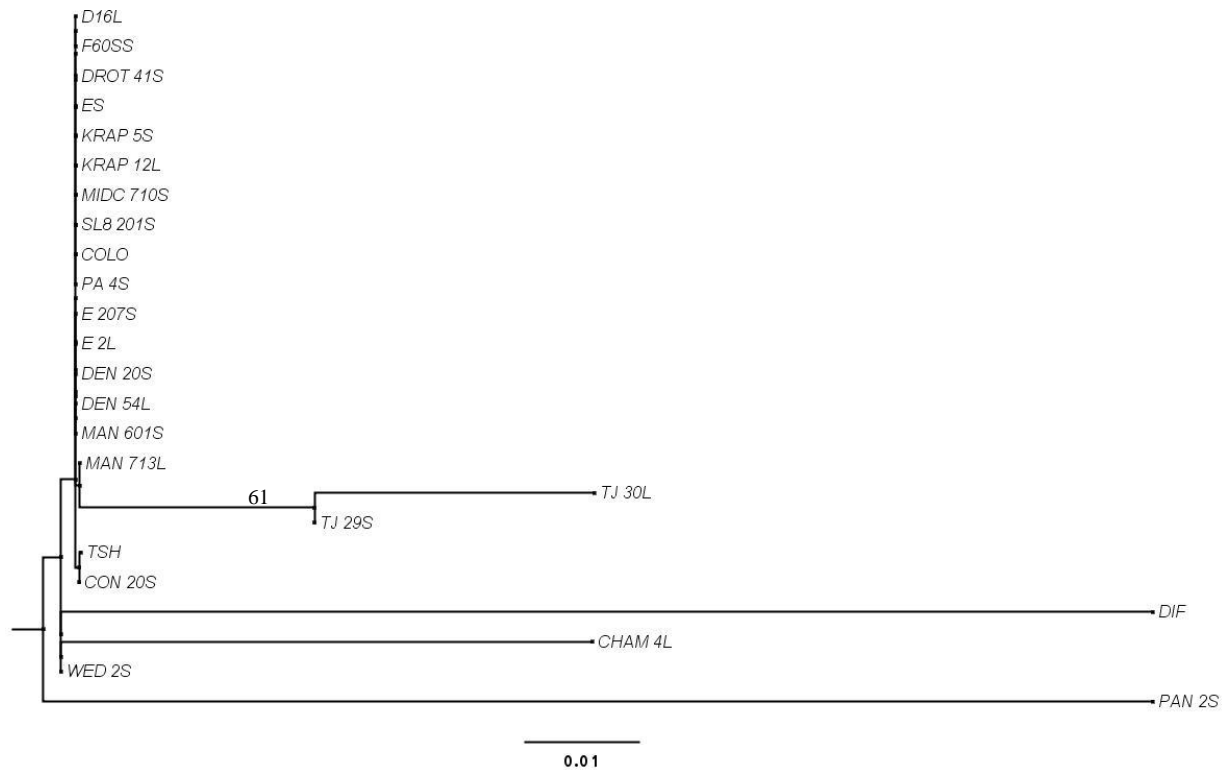


FIGURE E20: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *FSP*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the General Reversible Chloroplast (CPREV) amino acid substitution model (Adachi et al. 2000). The tree with the highest log likelihood (-107.347) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 25 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

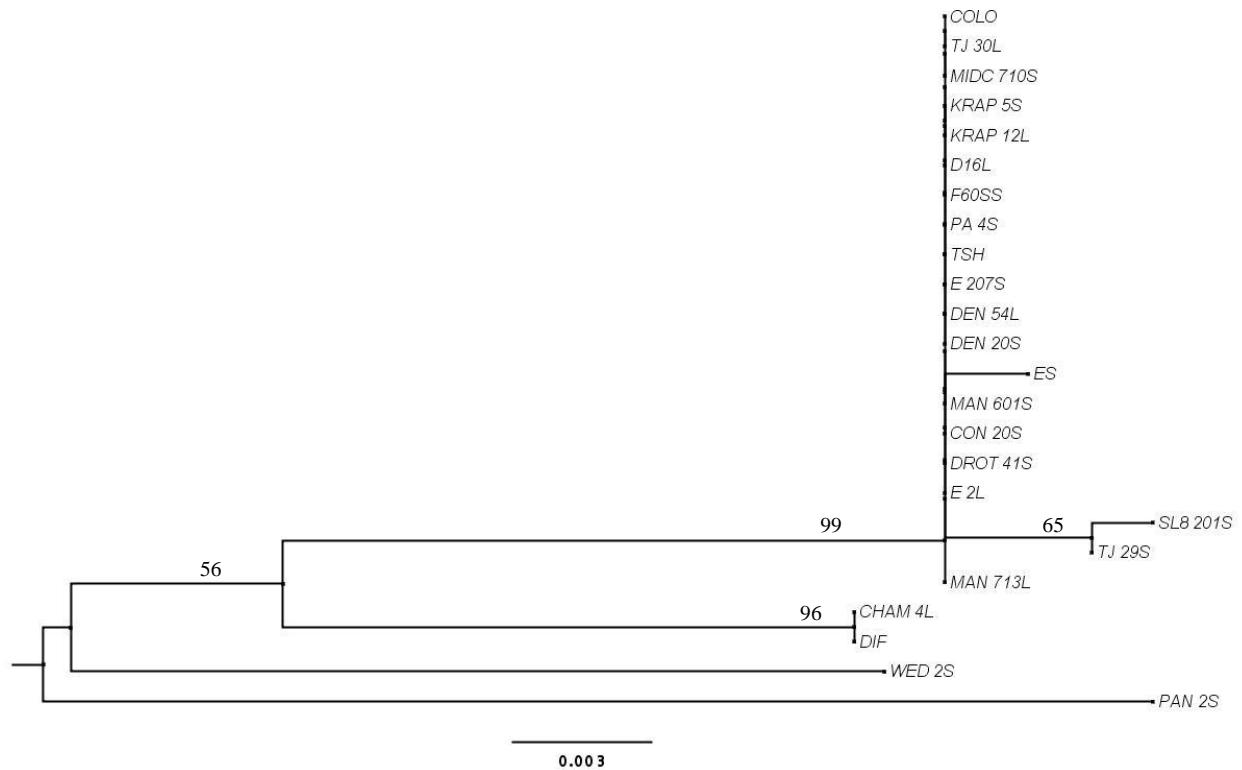


FIGURE E21: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *NRFP*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-980.419) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 563 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

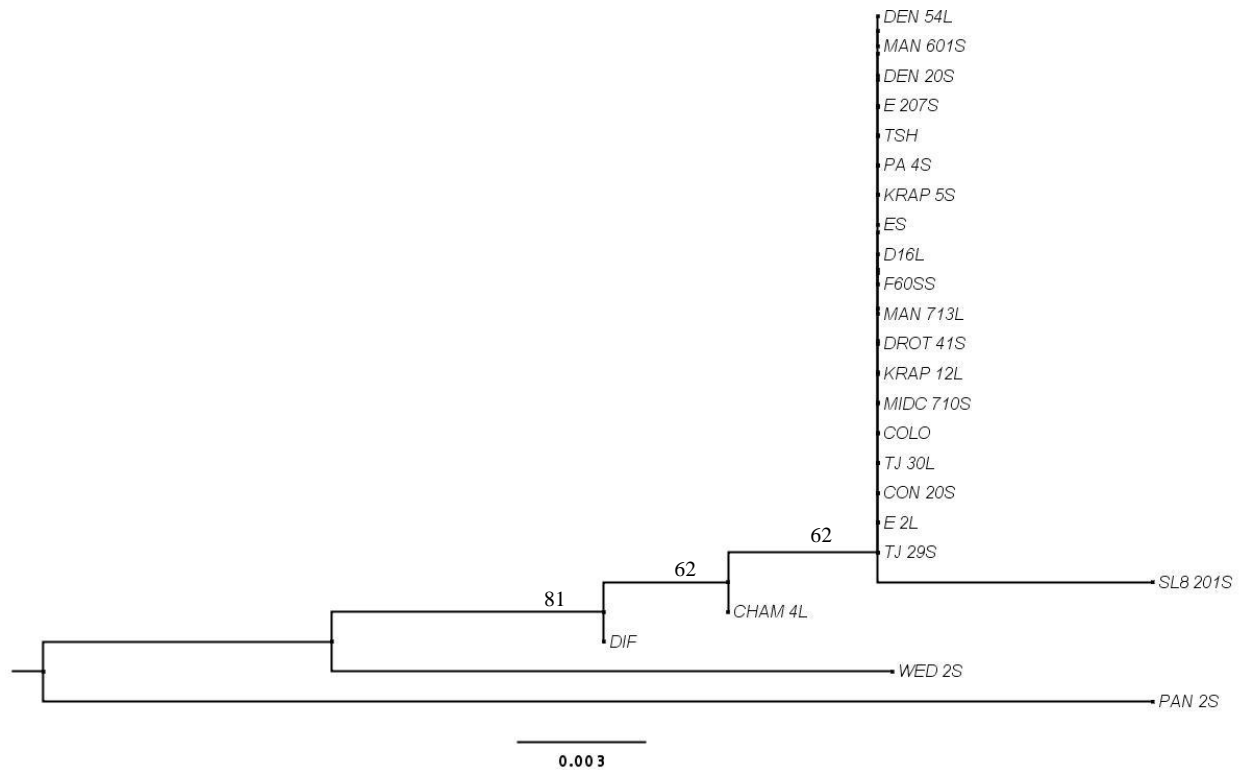


FIGURE E22: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *NRFP*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-539.965) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 161 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

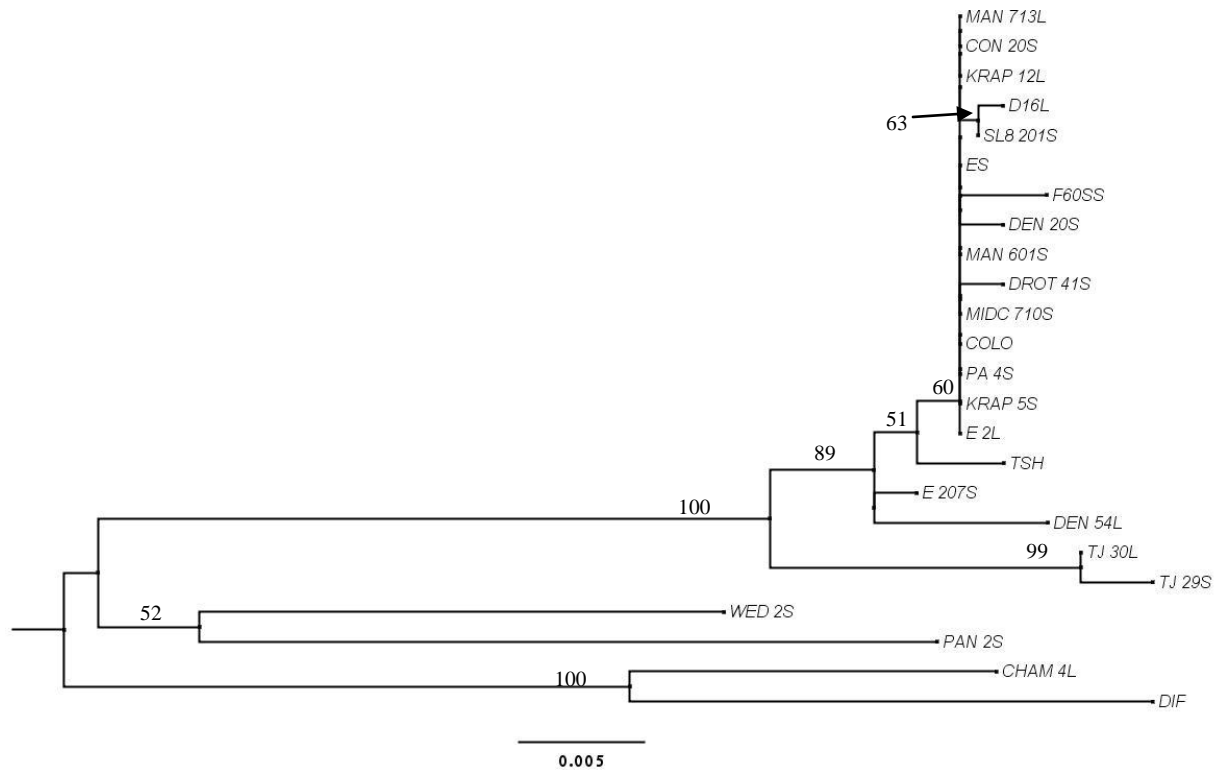


FIGURE E23: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *WRKY*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Kimura 2-parameter nucleotide substitution model (Kimura 1980). The tree with the highest log likelihood (-1433.993) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.278)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 591 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

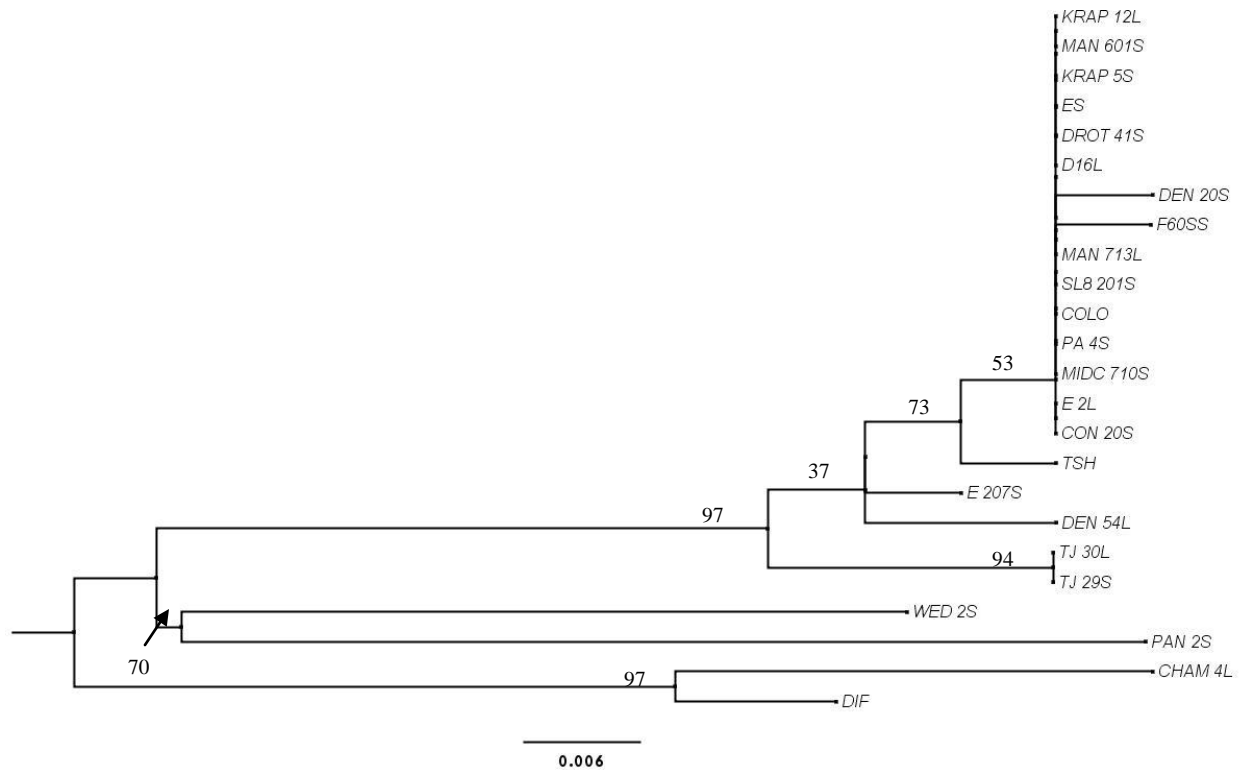


FIGURE E24: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *WRKY*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-845.125) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.270)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 184 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

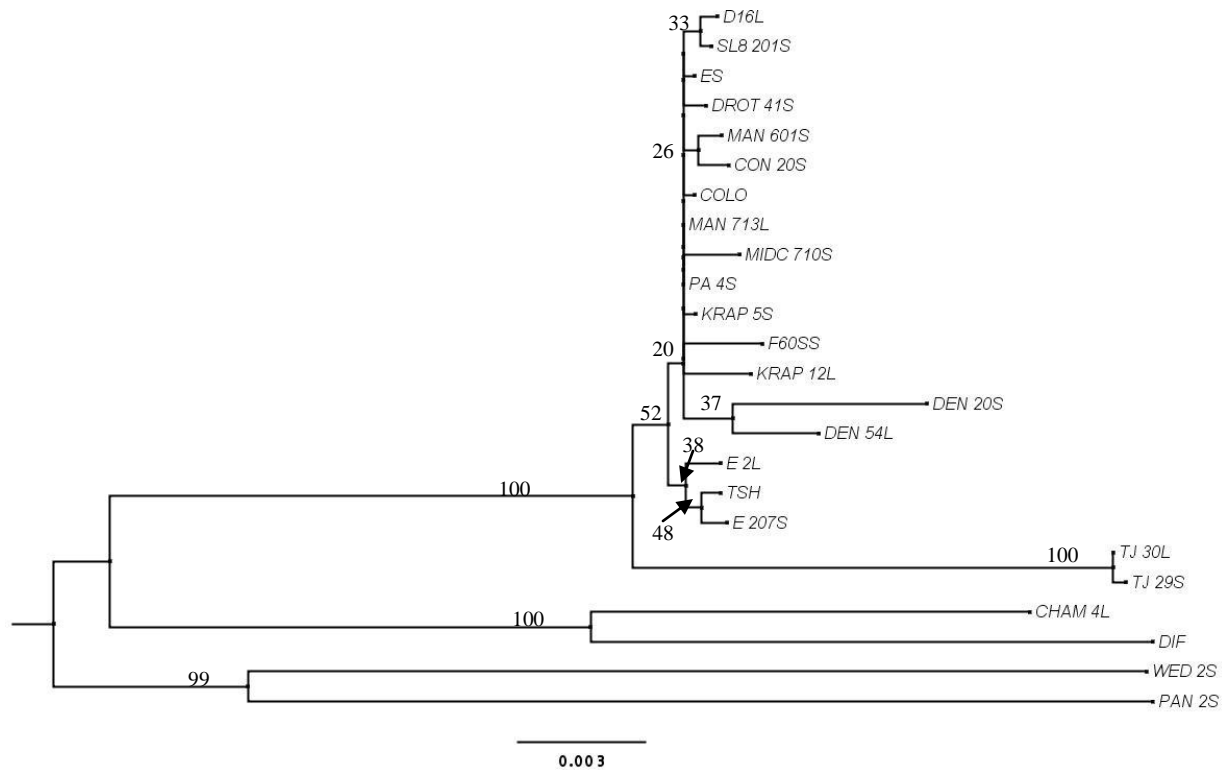


FIGURE E25: Molecular Phylogenetic analysis by Maximum Likelihood method using the total DNA data for all S-linked genes. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Kimura 2-parameter nucleotide substitution model (1980). The tree with the highest log likelihood (-8297.213) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.126)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 3872 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

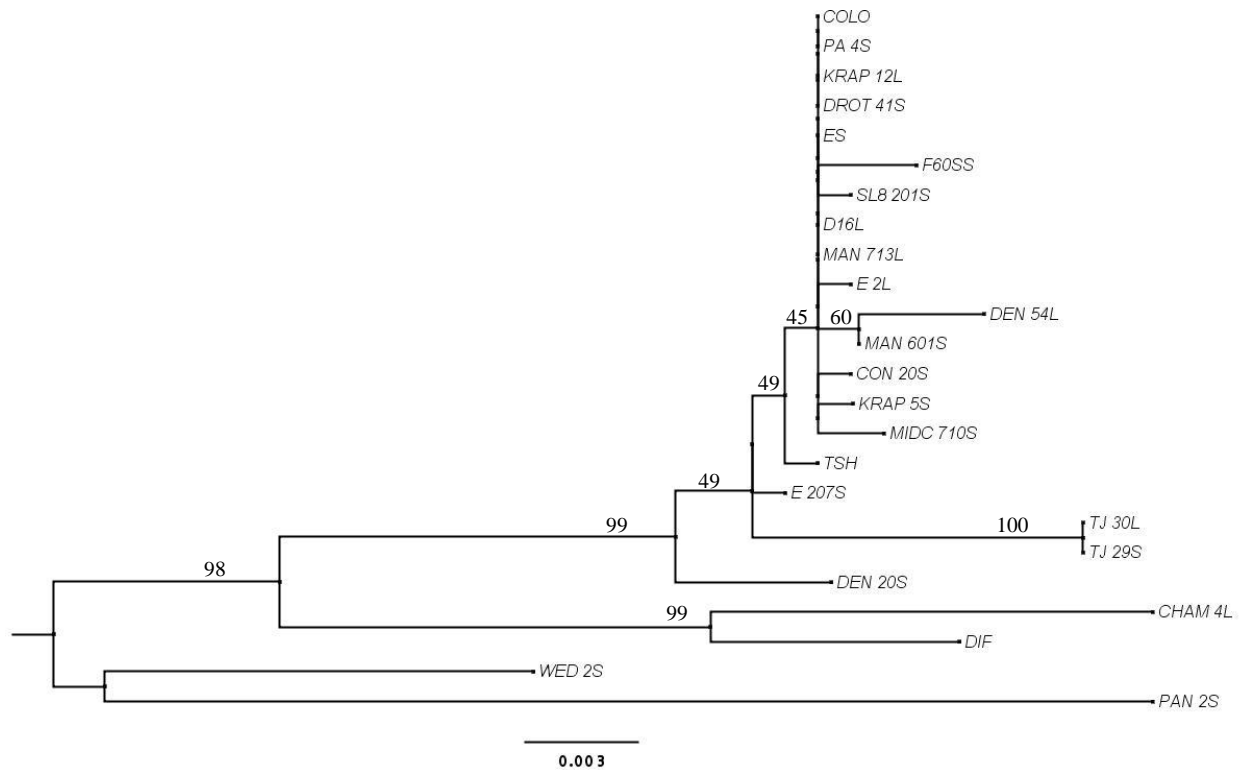


FIGURE E26: Molecular Phylogenetic analysis by Maximum Likelihood method using the total amino acid data for all S-linked genes. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-4314.262) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.132)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 1150 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

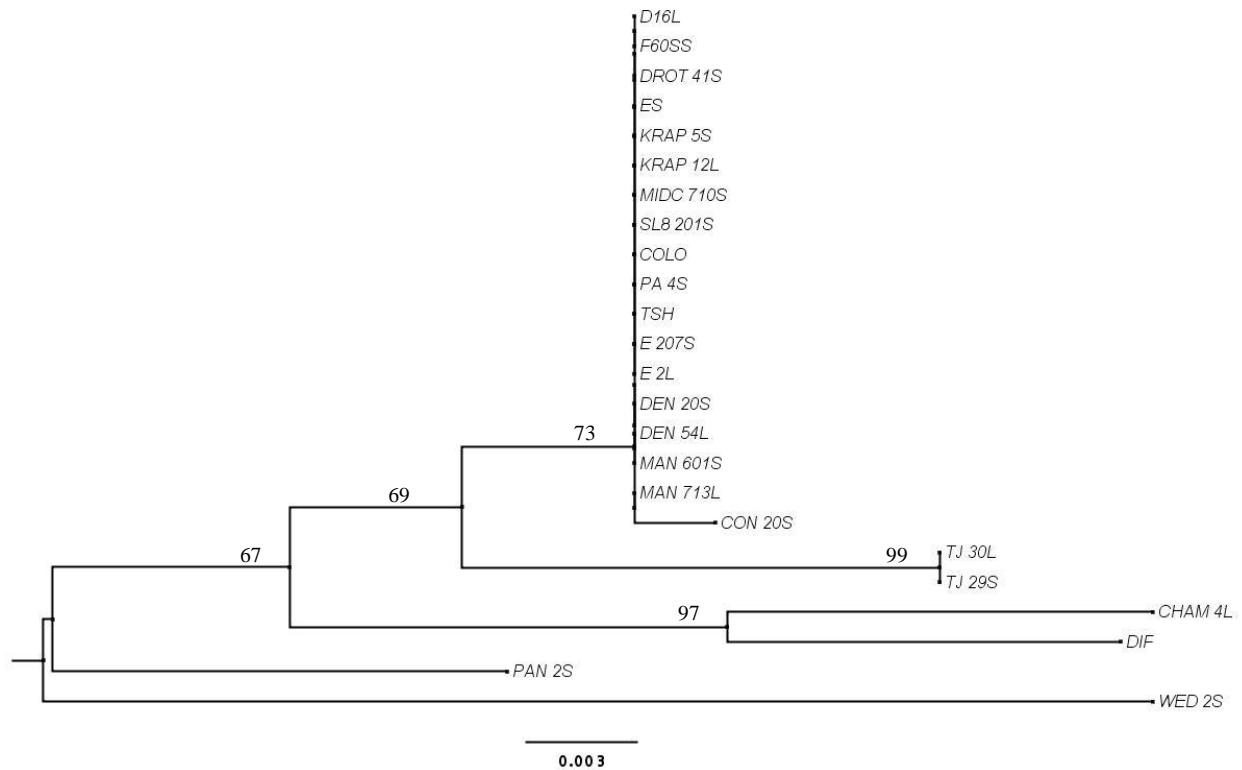


FIGURE E27: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *ECIPI*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-1493.679) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.050)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 780 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

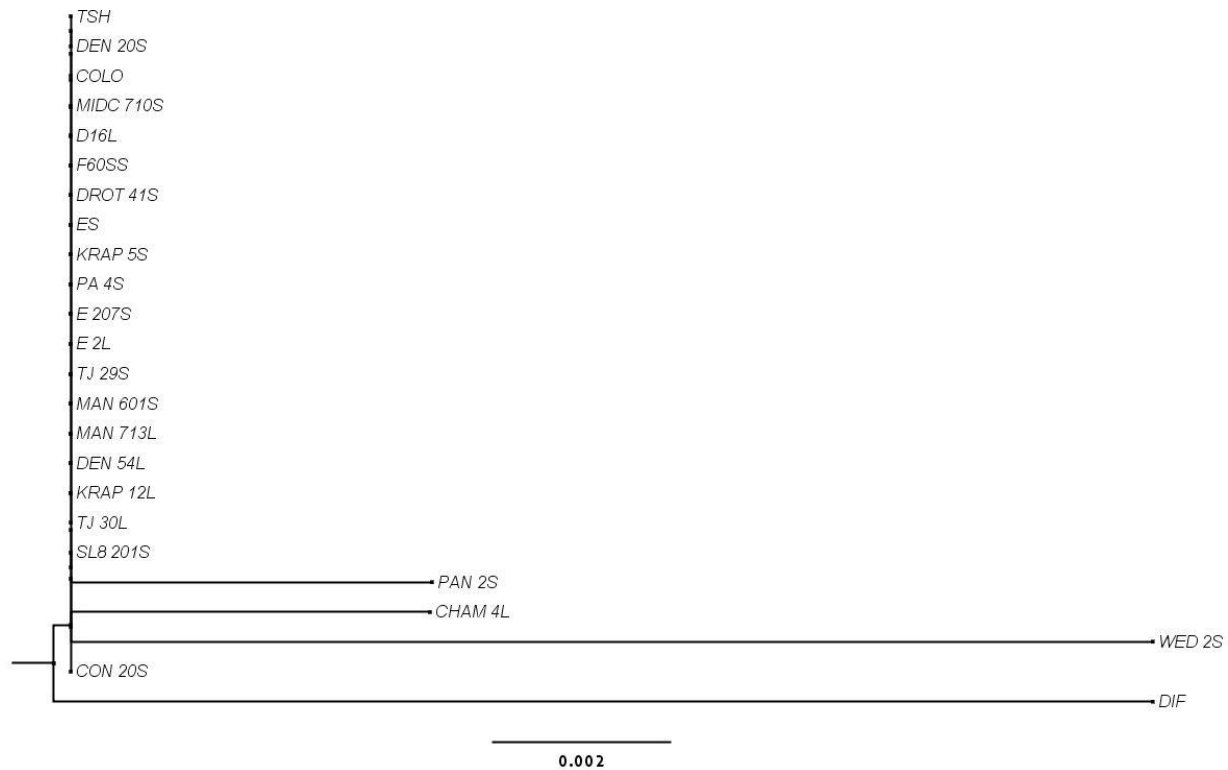


FIGURE E28: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *ECIP1*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the LG amino acid substitution model (Le and Gascuel 2008). The tree with the highest log likelihood (-764.747) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 249 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

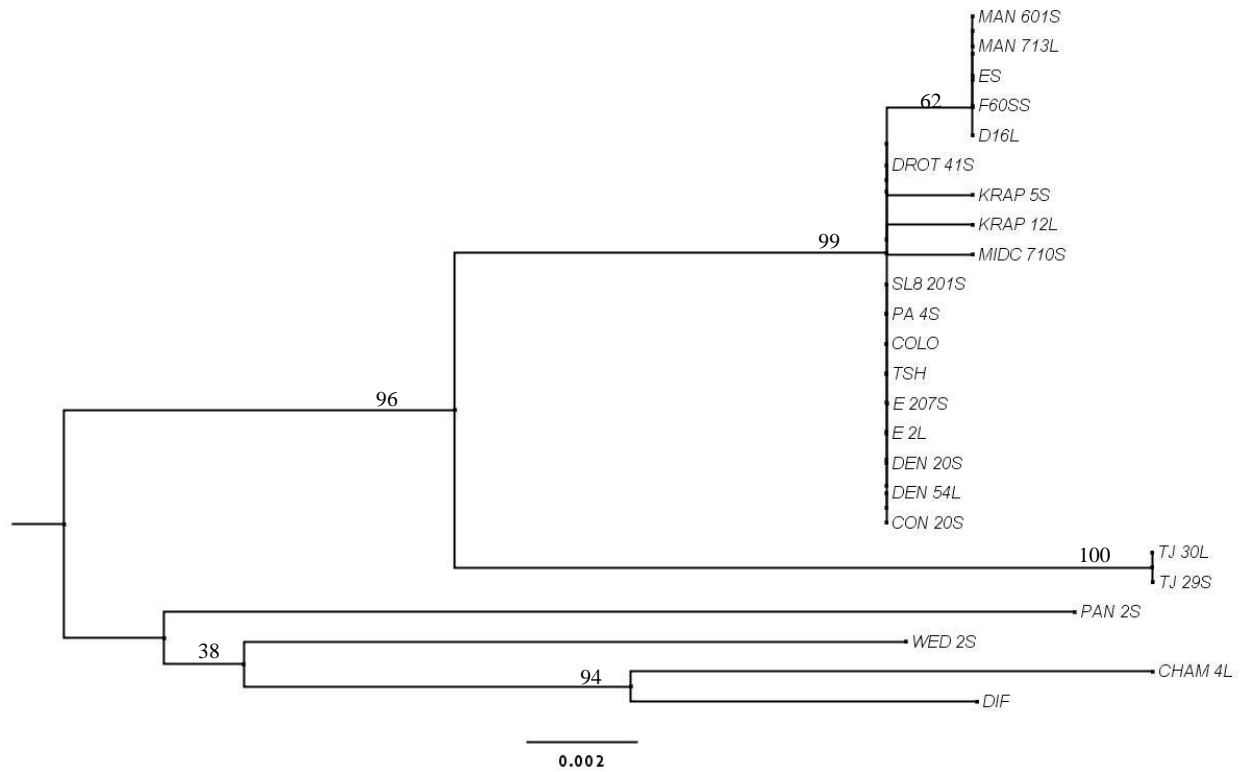


FIGURE E29: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *GAUT3*.

Evolutionary relationships were inferred using the Maximum Likelihood method based on the Tamura 3-parameter nucleotide substitution model (Tamura 1992). The tree with the highest log likelihood (-1256.347) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 651 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

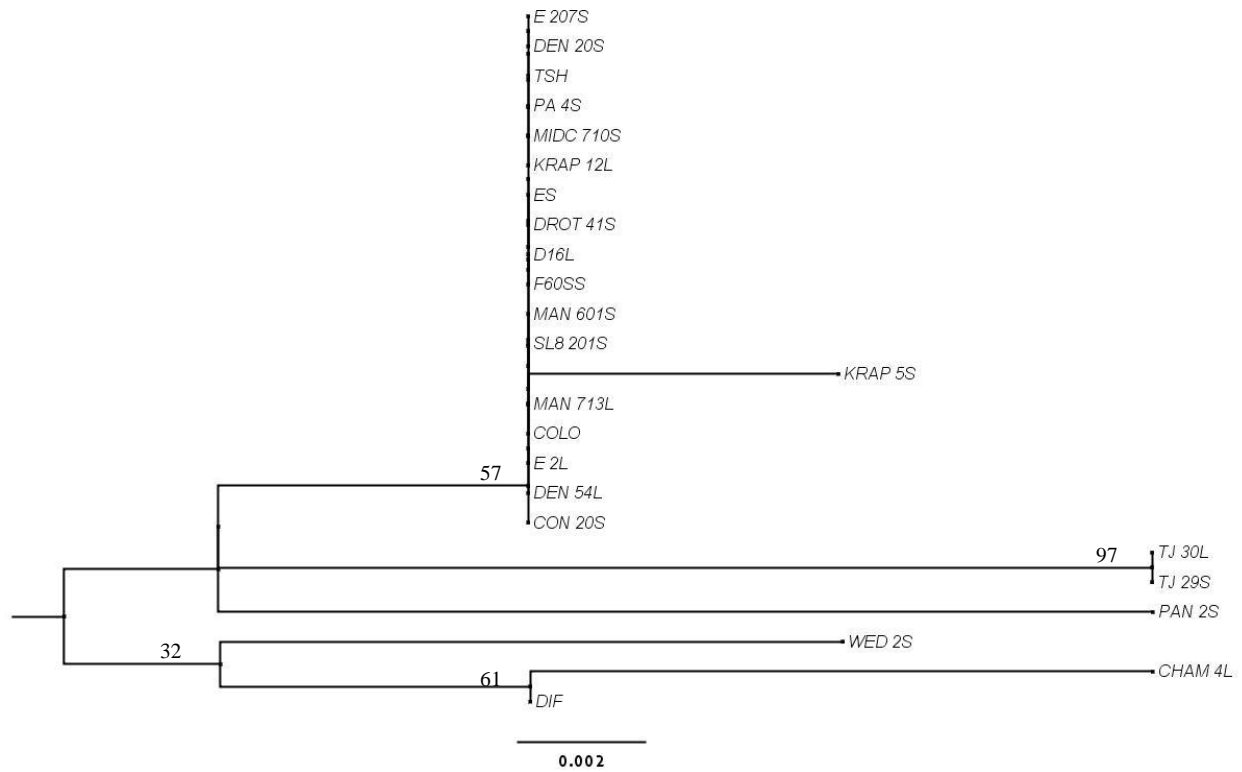


FIGURE E30: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *GAUT3*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the LG amino acid substitution model (Le and Gascuel 2008). The tree with the highest log likelihood (-725.623) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 210 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

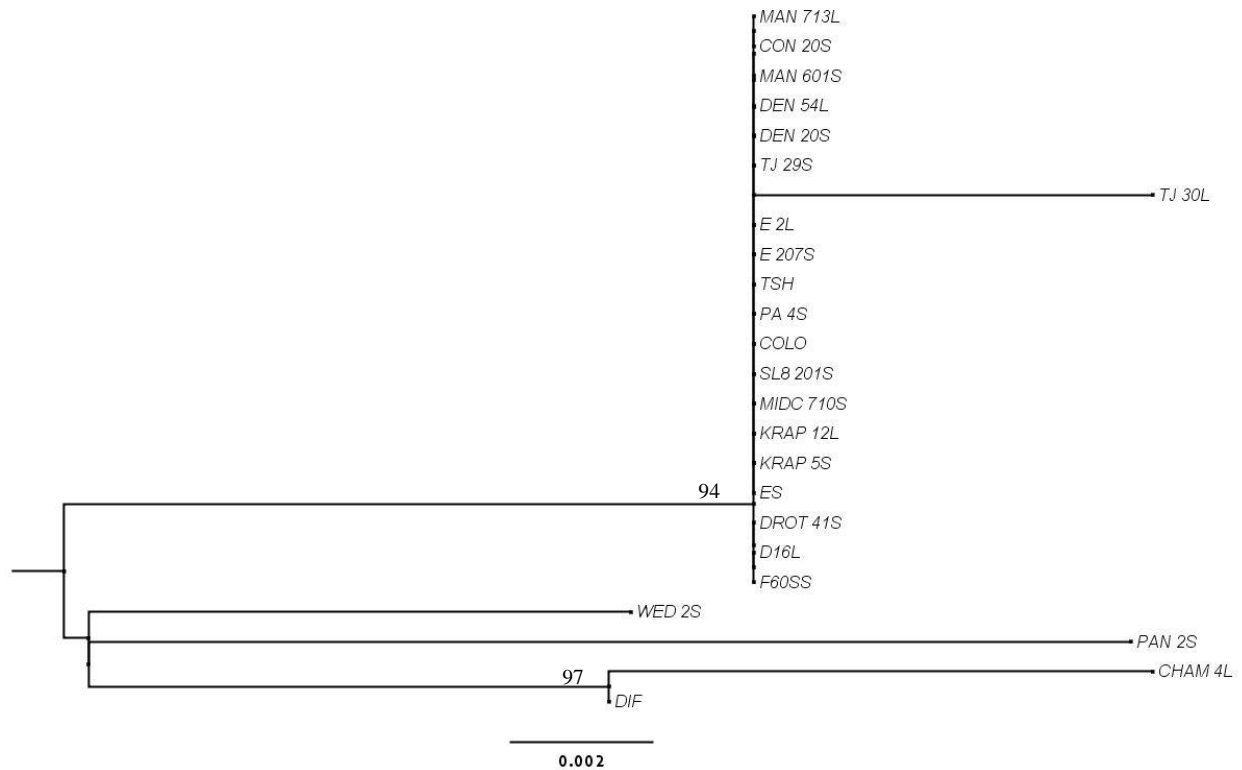


FIGURE E31: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *GAUTI*.

Evolutionary relationships were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-723.836) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 439 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

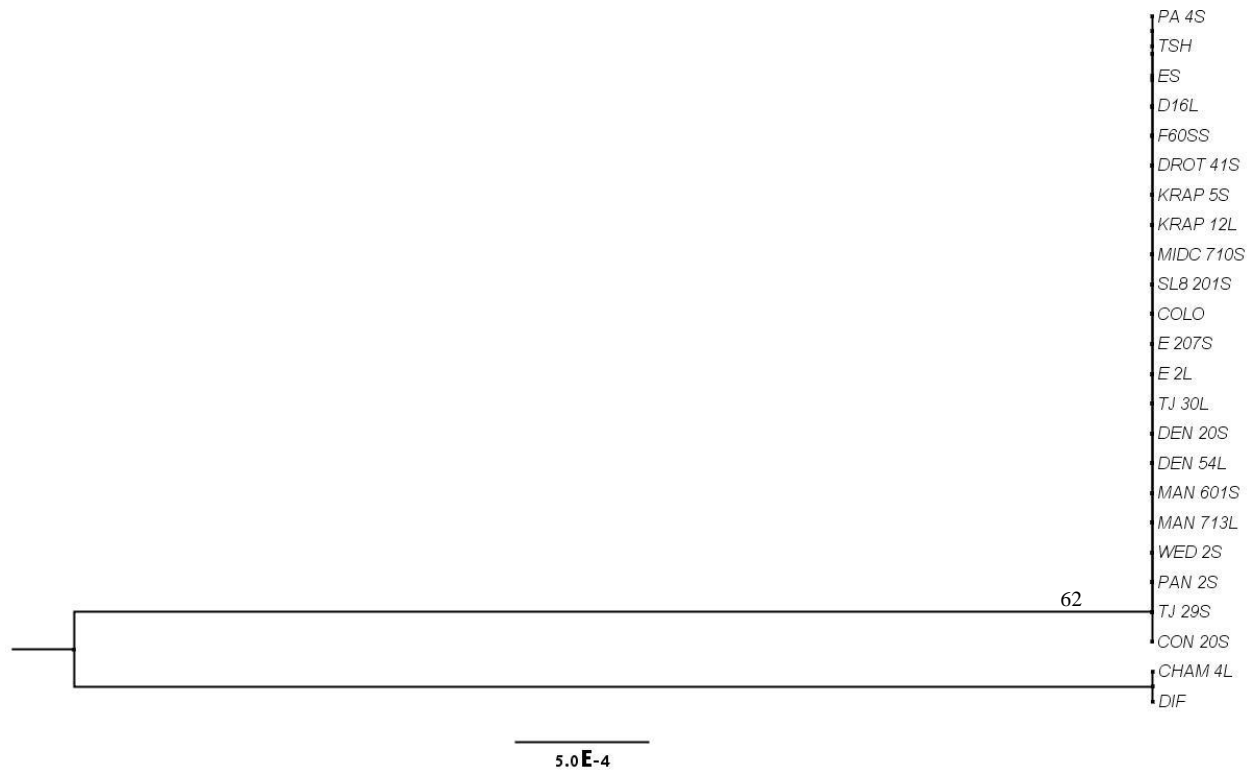


FIGURE E32: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *GAUTI*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the LG amino acid substitution model (Le and Gascuel 2008). The tree with the highest log likelihood (-392.048) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 129 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

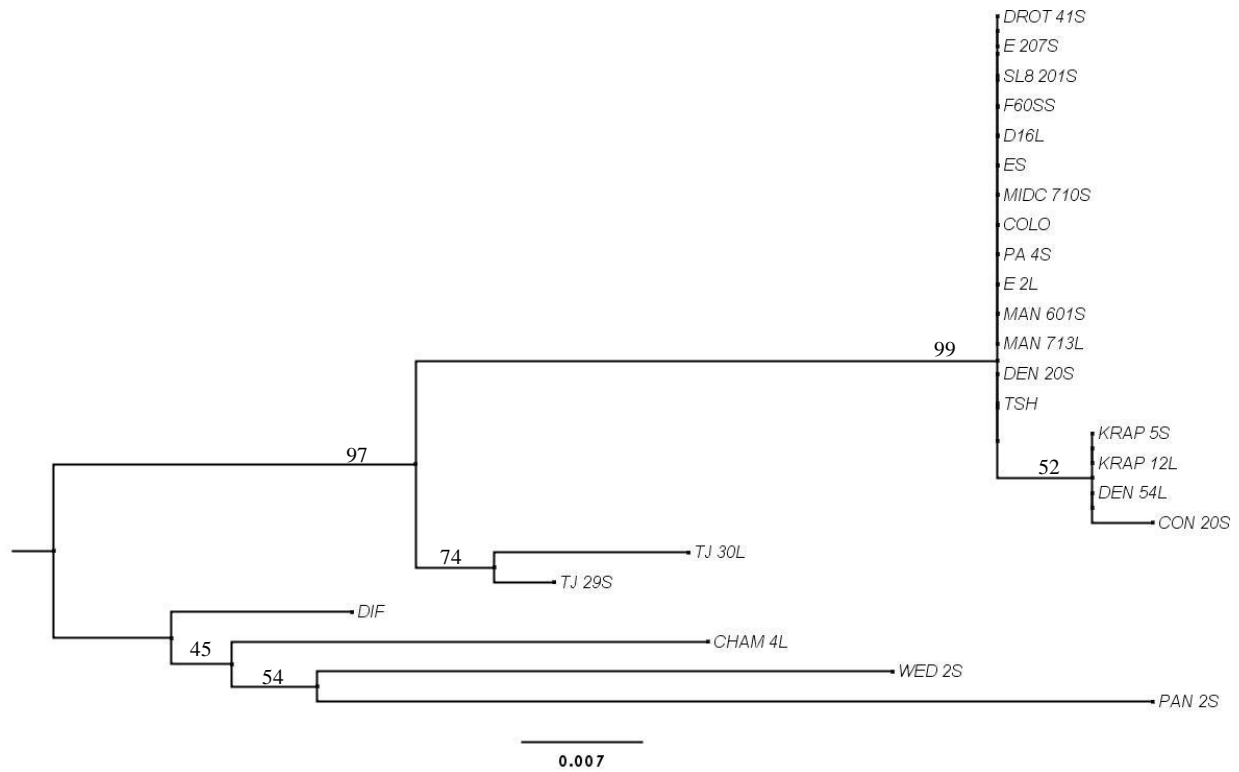


FIGURE E33: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *RNABP34*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-710.727) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 291 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

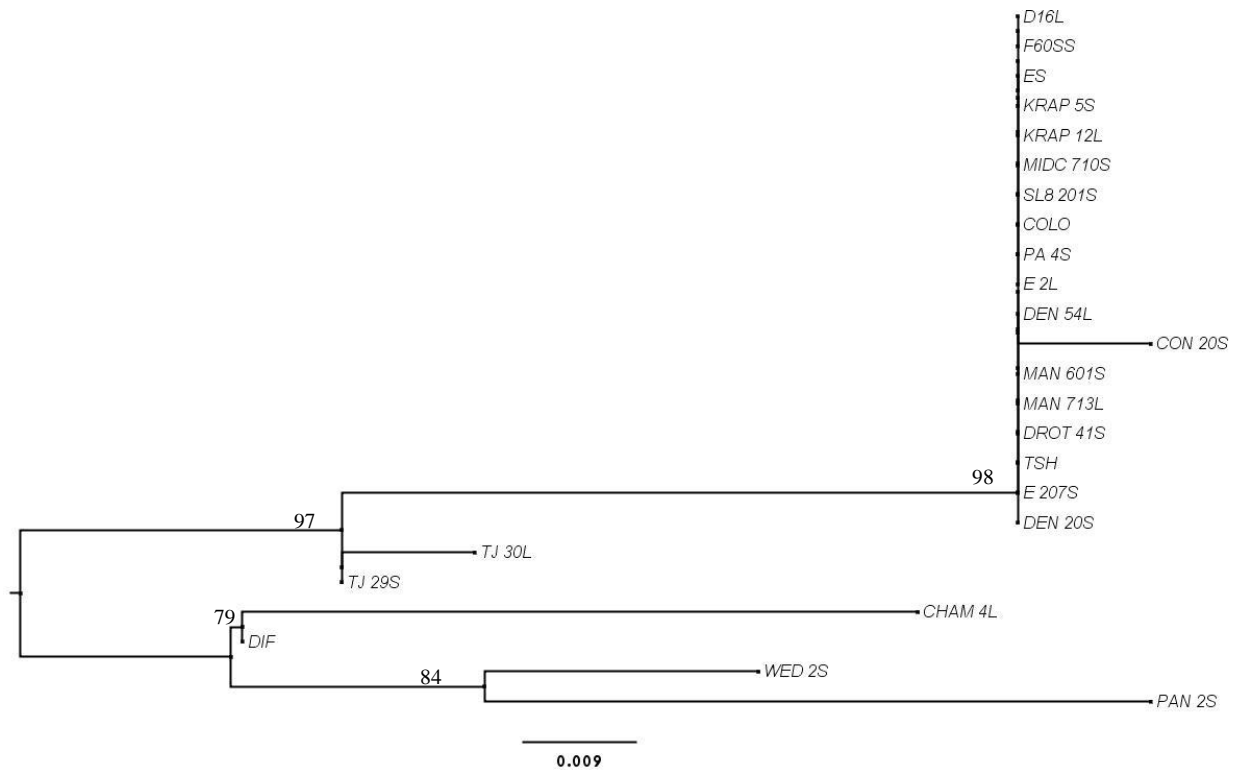


FIGURE E34: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *RNABP34*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Dayhoff matrix based amino acid substitution model (Schwarz and Dayhoff 1979). The tree with the highest log likelihood (-412.773) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 94 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

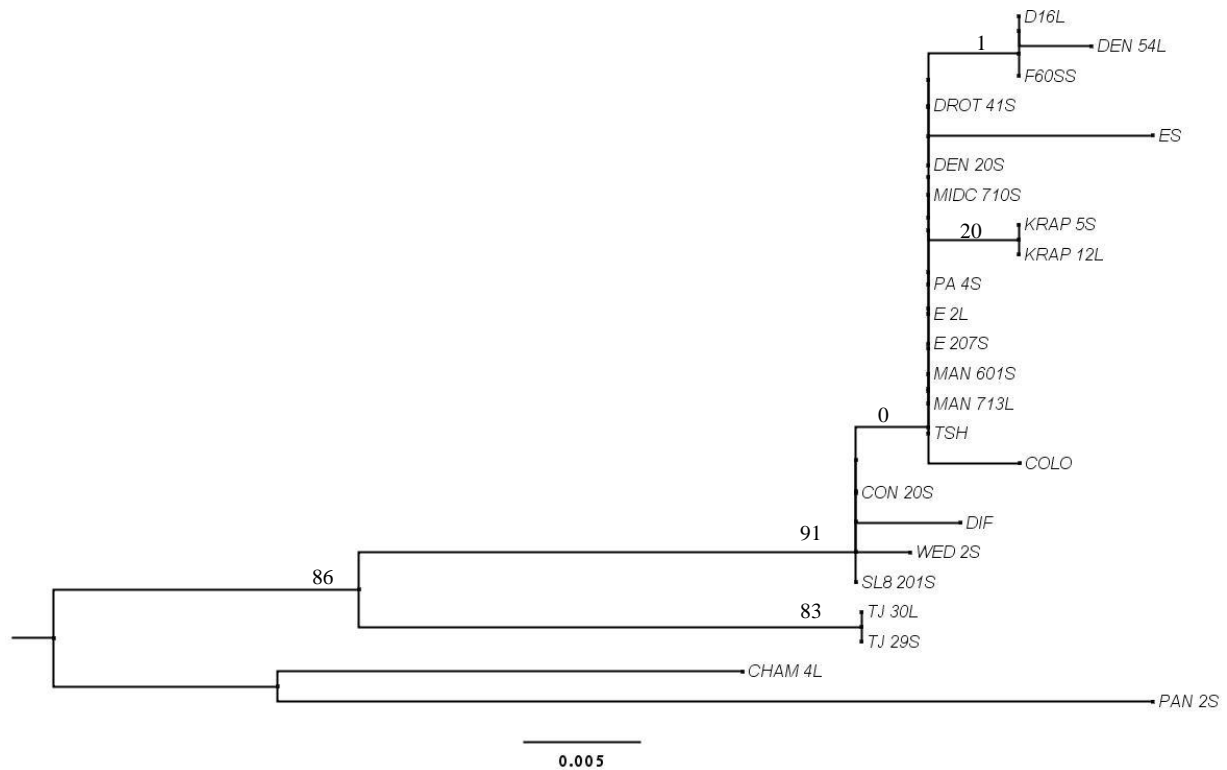


FIGURE E35: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *FMO1*.

Evolutionary relationships were inferred by using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-540.530) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. There were a total of 261 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

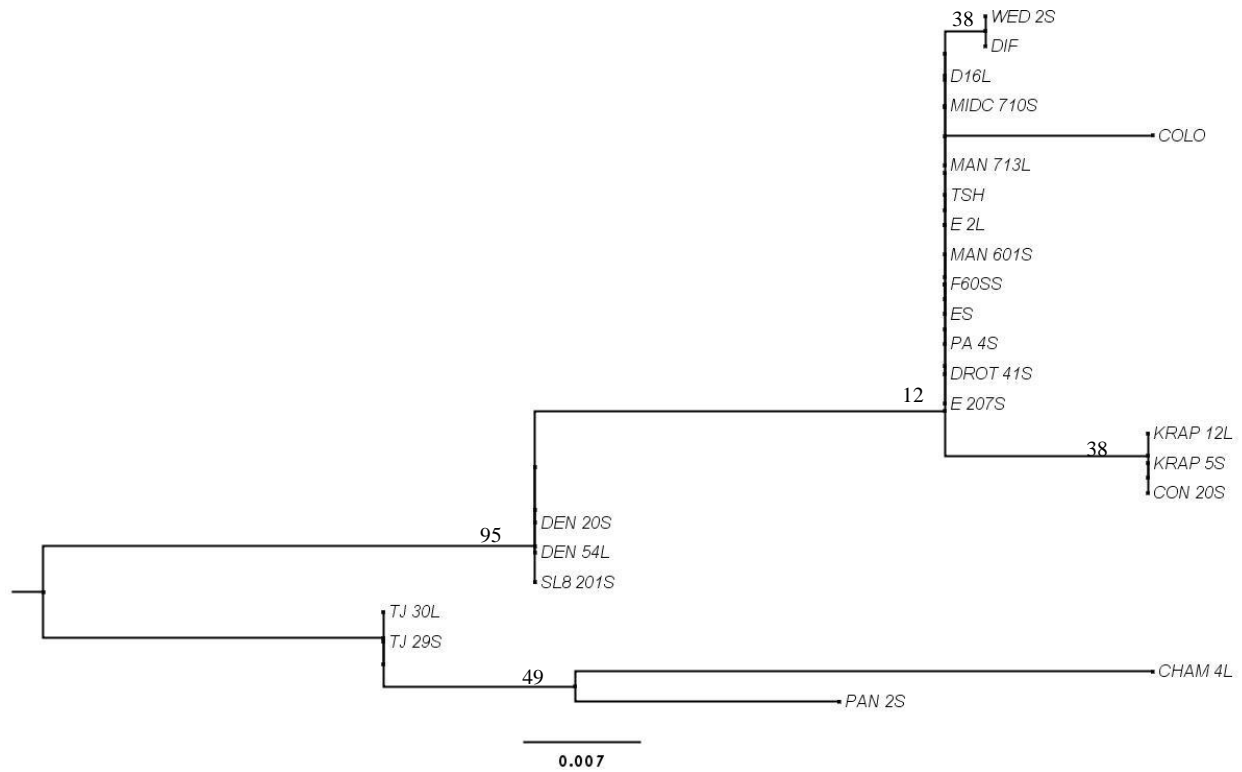


FIGURE E36: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *FMO1*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-348.872) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. There were a total of 87 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

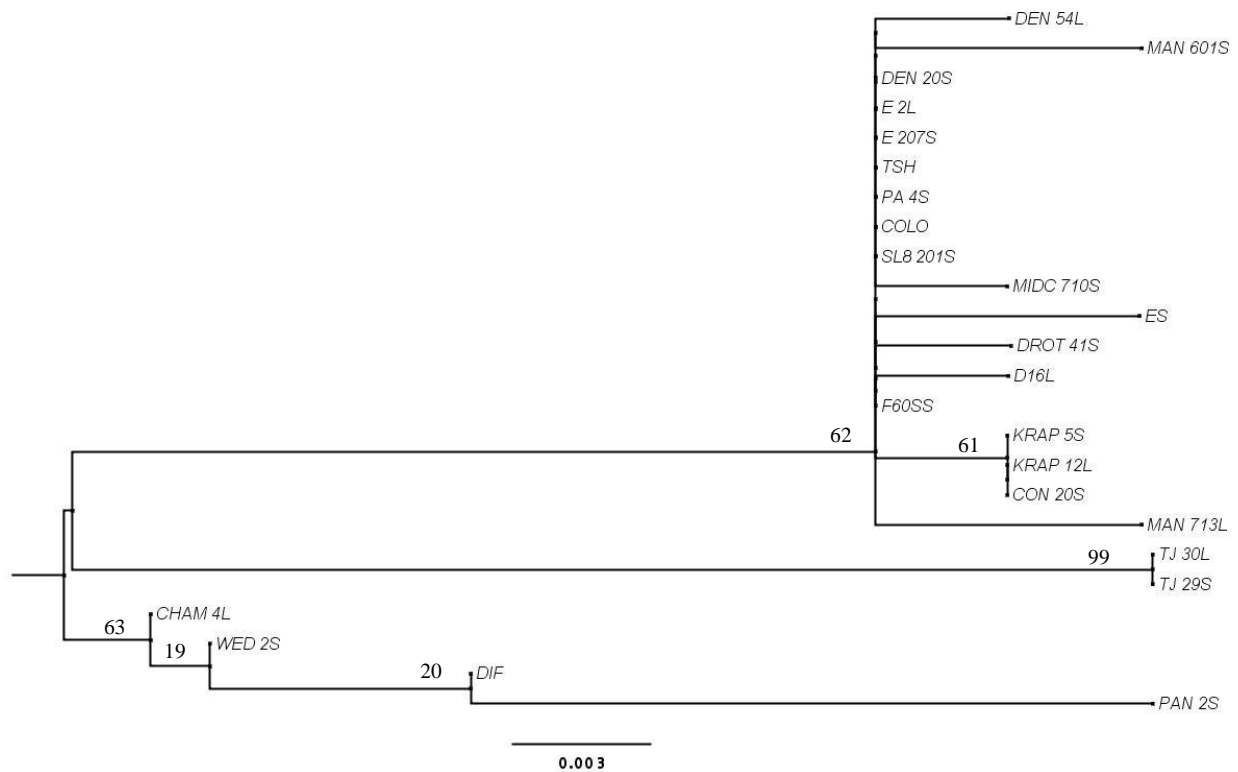


FIGURE E37: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *MBD8*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Jukes-Cantor nucleotide substitution model (1969). The tree with the highest log likelihood (-704.858) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. There were a total of 354 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

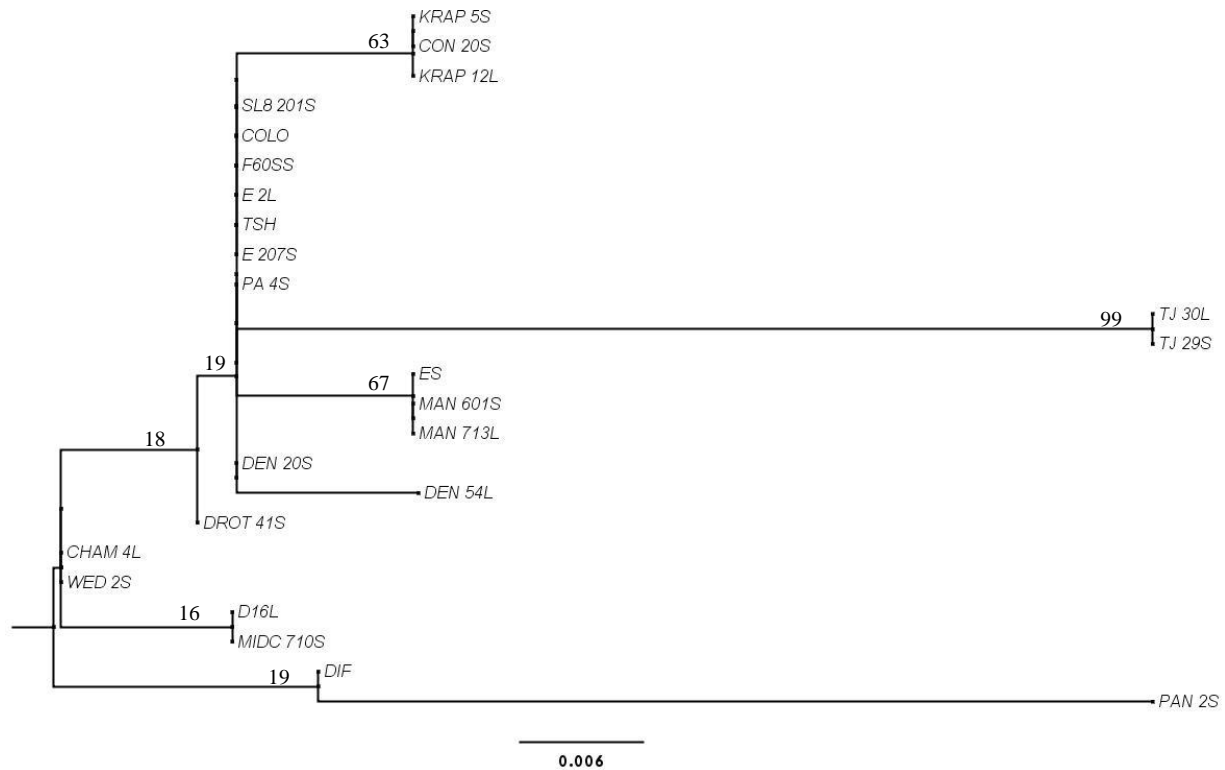


FIGURE E38: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *MBD8*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-447.942) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. There were a total of 118 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

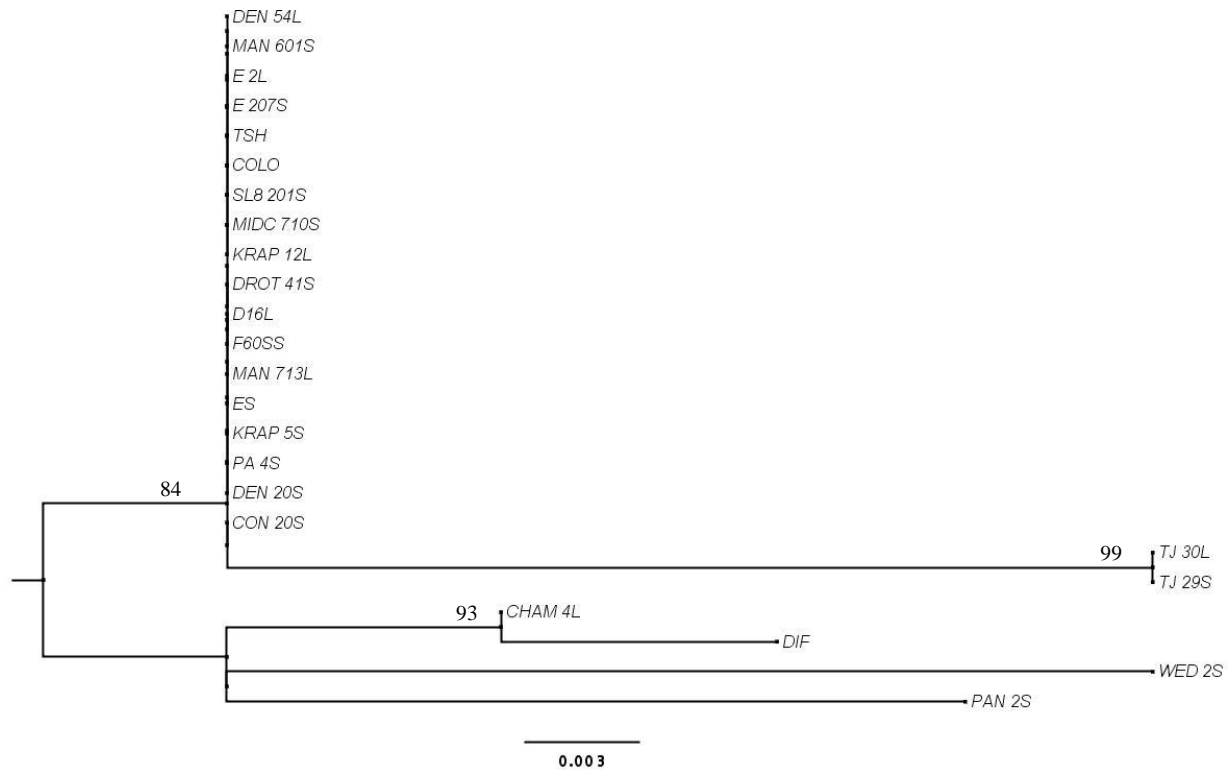


FIGURE E39: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for UNKN. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Jukes-Cantor nucleotide substitution model (1969). The tree with the highest log likelihood (-410.424) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 212 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

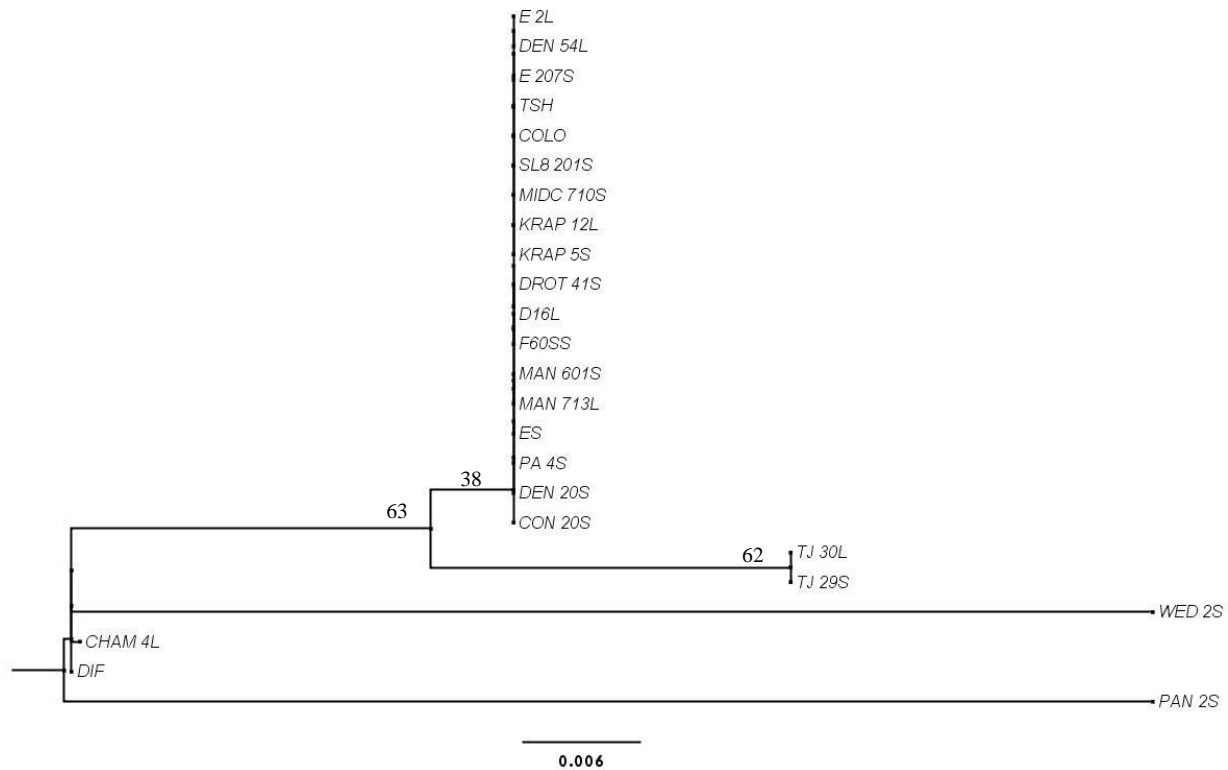


FIGURE E40: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for UNKN. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-222.025) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 58 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

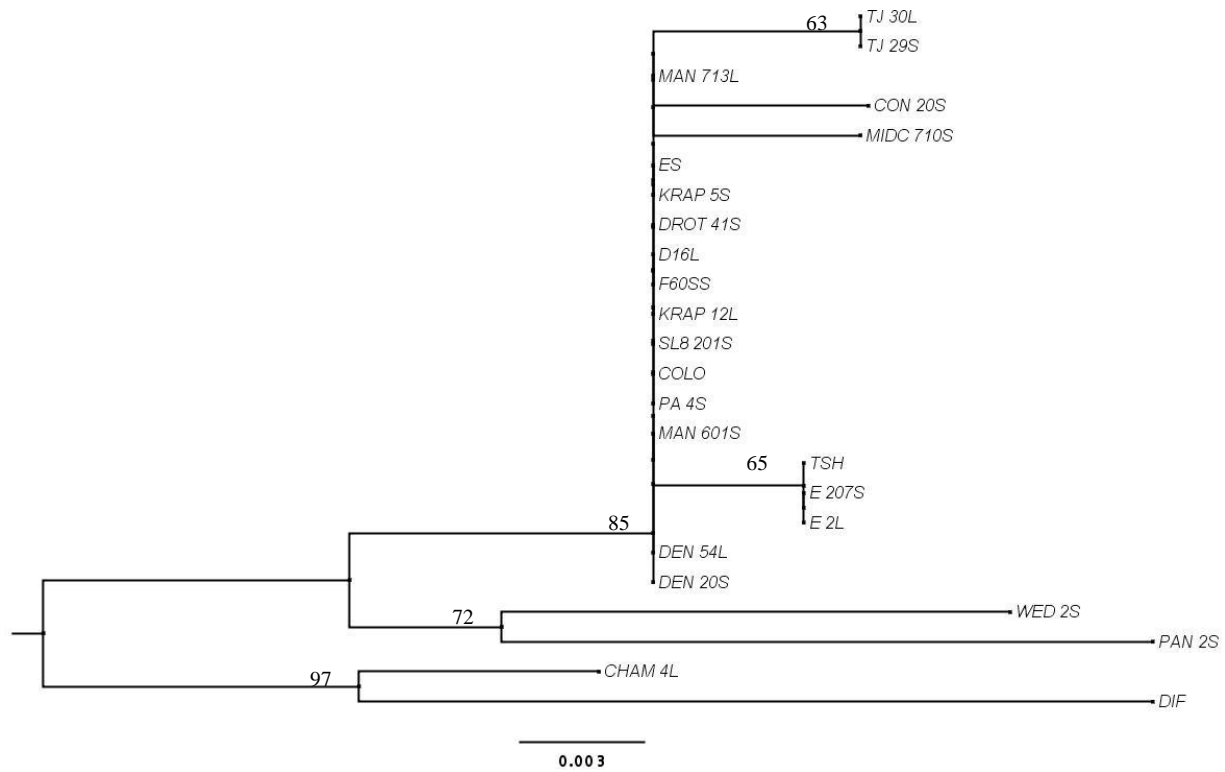


FIGURE E41: Molecular Phylogenetic analysis by Maximum Likelihood method using DNA data for *POFUT*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Hasegawa-Kishino-Yano nucleotide substitution model (1985). The tree with the highest log likelihood (-530.617) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 281 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

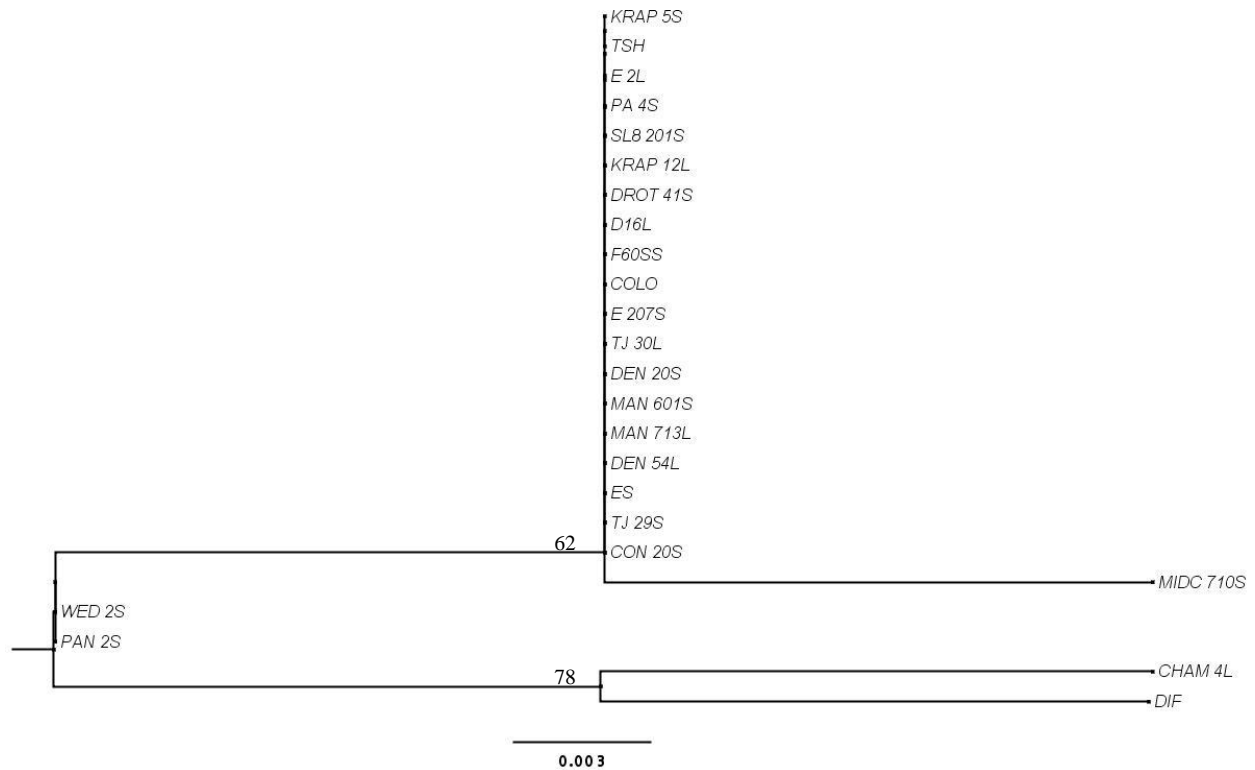


FIGURE E42: Molecular Phylogenetic analysis by Maximum Likelihood method using amino acid data for *POFUT*. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-282.992) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 86 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

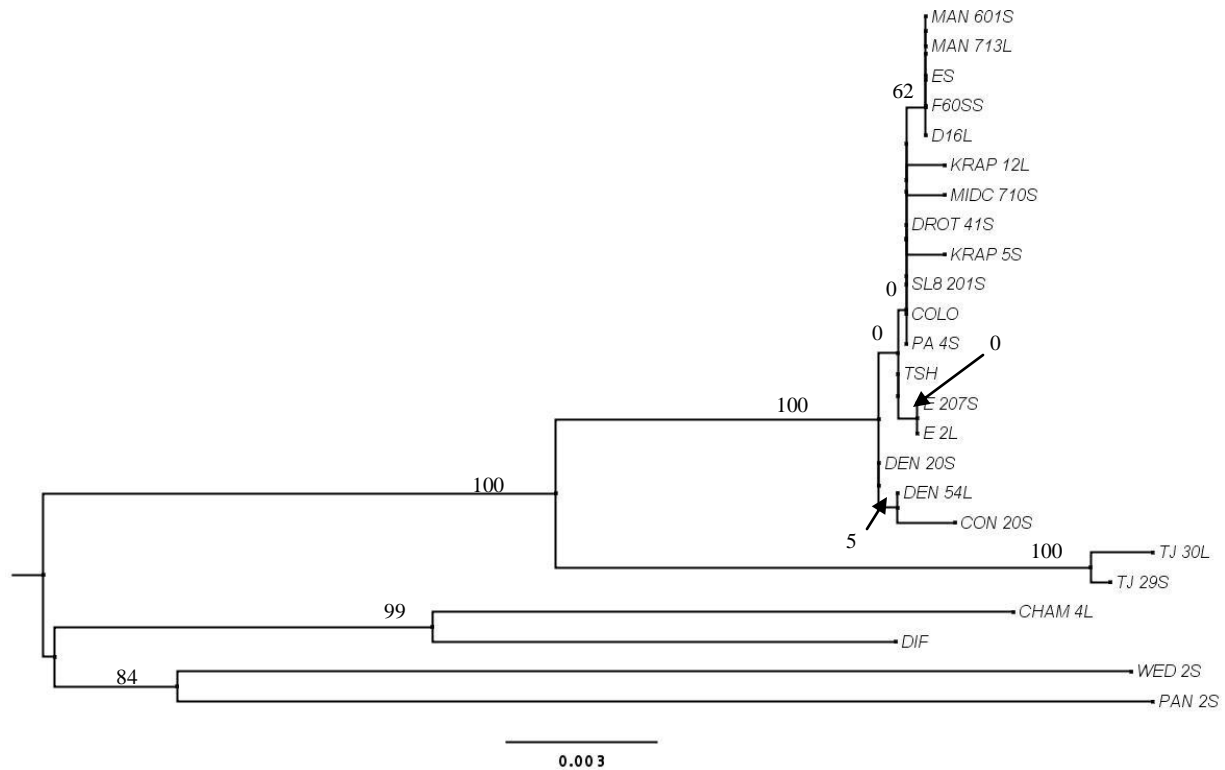


FIGURE E43: Molecular Phylogenetic analysis by Maximum Likelihood method using the total DNA data for all control genes. Evolutionary relationships were inferred using the Maximum Likelihood method based on the Tamura 3-parameter nucleotide substitution model (1992). The tree with the highest log likelihood (-5281.455) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.149)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 2654 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013)

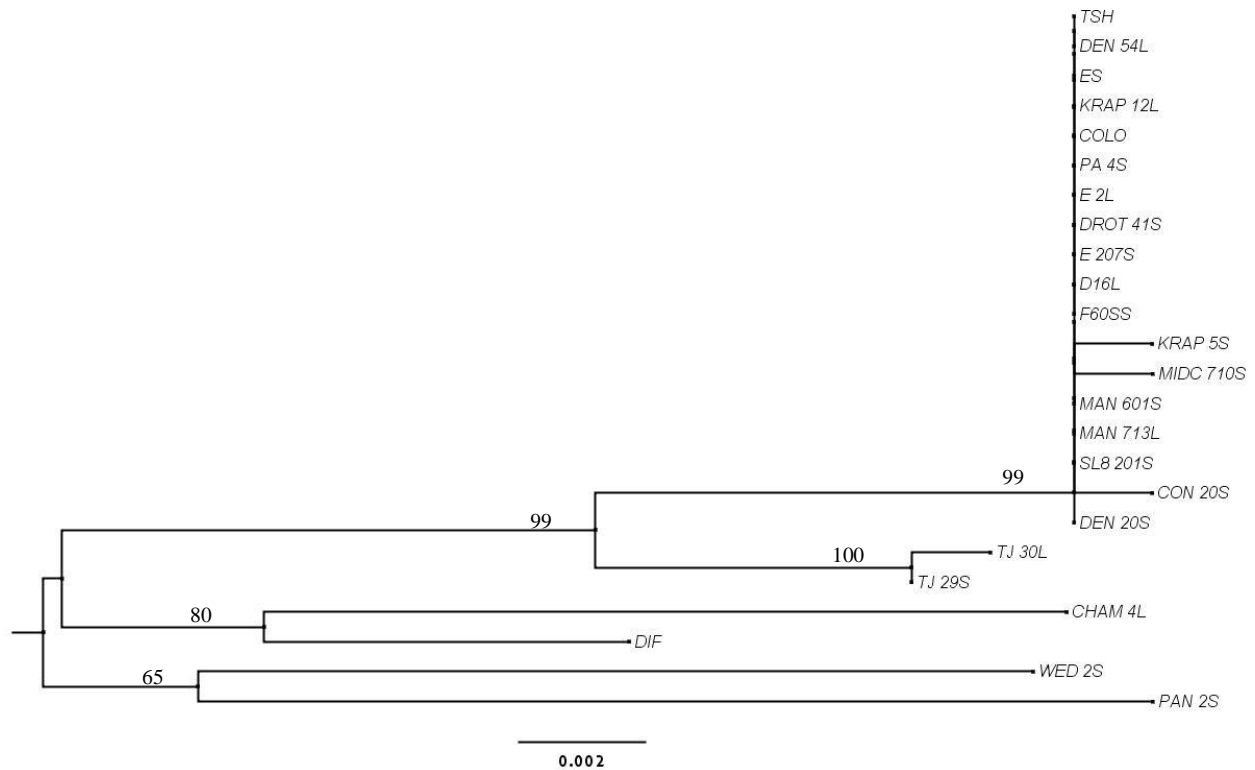


FIGURE E44: Molecular Phylogenetic analysis by Maximum Likelihood method using the total amino acid data for all control genes. Evolutionary relationships were inferred using the Maximum Likelihood method based on the JTT matrix-based amino acid substitution model (Jones, Taylor, and Thornton 1992). The tree with the highest log likelihood (-2863.213) is shown. The tree is unrooted. Bootstrap support values (%) are shown adjacent to the relevant nodes (1000 replicates). Not all supports are shown. Those that are not shown are < 50%. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pair-wise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.175)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 24 amino acid sequences. The coding data was translated assuming a Standard genetic code table. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 826 positions in the final data set. Evolutionary analyses were conducted in MEGA6 (Tamura et al. 2013).

Appendix F: Supporting Likelihood Ratio Tests and Parameter Estimates

Table F1: Parameter estimates for LRTs comparing selection on S-linked and control genes. dN/dS, dS, dN, and P (proportion of sites belonging to a given rate class) estimates for all rate classes (+, -, and neutral evolution) for all models are given. These estimates correspond to LRTs that were performed with default starting values for all parameters.

Almnt	Independent Model					Shared Positive Selection Strengths Model				Shared Positive Selection Proportions Model				Shared Positive Selection Regime Model				Shared Selection Parameters Model			
	Rate Class	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P
<i>Tssta1</i>	+	2.371	2.2	5.4	0.0	3.201	1.855	5.938	0.011	2.352	2.256	5.307	0.013	3.092	1.819	5.624	0.012	2.898	1.808	5.24	0.01
	Neut	1.000	0.9	0.9	0.1	1.000	0.999	0.999	0.138	1.00	0.921	0.921	0.149	1.000	0.958	0.958	0.142	1.000	0.887	0.88	0.13
	-	0.200	8.9	0.1	0.0	0.021	9.059	0.187	0.050	0.021	8.961	0.785	0.051	0.021	9.033	0.190	0.051	0.078	0.509	0.40	0.79
	-	0.090	0.4	0.0	0.7	0.099	0.481	0.048	0.801	0.088	0.479	0.042	0.787	0.094	0.481	0.045	0.795	0.021	7.596	0.16	0.06
<i>S-linked</i>	+	7.550	0.5	4.1	0.0	3.201	1.065	3.409	0.020	8.076	0.550	4.444	0.013	3.092	1.332	4.119	0.012	2.898	1.808	5.24	0.01
	Neut	1.000	0.5	0.5	0.1	1.000	0.503	0.503	0.107	1.00	0.641	0.641	0.110	1.000	0.638	0.638	0.114	1.000	0.887	0.88	0.13
	-	0.460	0.5	0.0	0.7	0.052	0.599	0.031	0.773	0.047	0.590	0.027	0.777	0.044	0.589	0.026	0.777	0.078	0.509	0.40	0.79
	-	0.027	4.6	0.1	0.1	0.026	4.615	0.121	0.100	0.025	4.658	0.118	0.100	0.024	4.689	0.111	0.097	0.021	7.596	0.16	0.06

Table F2: Comparison of selection on S-linked and control genes. Four likelihood ratio tests were performed in order to determine if the genes of interest were experiencing different selective pressures when compared to a random assortment of control genes. For each model (constrained and independent), the log likelihood value and number of estimated parameters is provided. For each test, the corresponding degrees of freedom (DF) and LRT statistic are shown. Significant p-values are indicated in bold. p-values < 0.05 suggest that the independent model fit is superior to that of the constrained model and, further, that the control and S-linked gene data sets are significantly different with respect to the relevant constrained parameter(s). All four tests were completed twice, with random and default starting values for parameters, respectively. The results obtained using random starting values are shown below.

Parameters constrained in constrained model	Log Likelihood (Independent Model)	# of Parameters (Independent Model)	Log Likelihood (Constrained Model)	# of Parameters (Constrained Model)	DF	LRT Statistic	p-value
Shared distributions for all parameters	-22526.646	126	-22540.563	116	10	27.830	0.002
Shared selective regimes (dN/dS and proportions)			-22527.618	124	2	1.941	0.379
Shared strength of positive selection (dN/dS)			-22527.421	125	1	1.549	0.461
Shared proportion of positively selected sites			-22526.688	125	1	0.081	0.774

Table F3: Parameter estimates for LRTs comparing selection on *Tsst1* and all other S-linked genes. dN/dS, dS, dN, and P (proportion of sites belonging to a given rate class) estimates for all rate classes (+, -, and neutral evolution) for all models are given. These estimates correspond to LRTs that were performed with default starting values for all parameters.

Almnt	Independent Model				Shared Positive Selection Strengths Model				Shared Positive Selection Proportions Model				Shared Positive Selection Regime Model				Shared Selection Parameters Model				
	Rate Class	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P
<i>Tsst1</i>	+	118.6	0.0	6.6	0.0	14.931	0.351	5.240	0.010	121.636	0.067	8.089	0.005	14.876	0.465	6.924	0.004	3.084	0.645	1.99	0.06
	Neut.	1.000	0.0	0.0	0.0	1.000	0.047	0.047	0.000	1.000	0.049	0.049	0.000	1.000	0.051	0.051	0.000	1.000	0.0117	0.01	0.43
	-	0.000	0.6	0.0	0.6	0.000	0.623	0.000	0.690	0.000	0.641	0.000	0.701	0.000	0.645	0.000	0.703	0.021	29.319	0.61	0.01
	-	0.386	1.8	0.7	0.3	0.391	1.891	0.739	0.300	0.431	1.869	0.805	0.294	0.438	1.863	0.815	0.292	0.112	1.077	0.12	0.48
S-linked	+	13.16	0.7	9.9	0.0	14.931	0.677	10.103	0.004	13.155	0.751	9.874	0.005	14.876	0.677	10.065	0.004	3.084	0.645	1.99	0.06
	Neut.	1.000	0.8	0.8	0.1	1.000	0.828	0.828	0.183	1.000	0.825	0.825	0.183	1.000	0.828	0.828	0.183	1.000	0.0117	0.01	0.43
	-	0.020	24	0.4	0.0	0.020	24.506	0.488	0.200	0.020	24.528	0.488	0.020	0.020	24.51	0.488	0.02	0.021	29.319	0.61	0.01
	-	0.023	0.4	0.0	0.7	0.024	0.455	0.011	0.793	0.023	0.456	0.011	0.792	0.024	0.456	0.011	0.793	0.112	1.077	0.12	0.48

Table F4: Comparison of selection on *Tssta1* and other S-linked genes. Four likelihood ratio tests were performed in order to determine if *Tssta1* had experienced different selective pressures when compared to other S-linked genes. For each model (constrained and independent), the log likelihood value and number of estimated parameters is provided. For each test, the corresponding degrees of freedom (DF) and LRT statistic are shown. p-values < 0.05 suggest that the independent model fit is superior to that of the constrained model and, further, that the *Tssta1* and S-linked gene data sets are significantly different with respect to the relevant constrained parameter(s). However, no significant results were obtained. All four tests were completed twice, with random and default starting values for parameters, respectively. The results obtained using random starting values are shown below.

Parameters constrained in constrained model	Log Likelihood (Independent Model)	# of Parameters (Independent Model)	Log Likelihood (Constrained Model)	# of Parameters (Constrained Model)	DF	LRT Statistic	p-value
Shared distributions for all parameters	-15402.412	94	-15401.649	84	10	-1.525	1.000
Shared selective regimes (dN/dS and proportions)			-15402.609	92	2	0.396	0.821
Shared strength of positive selection (dN/dS)			-15402.551	93	1	0.279	0.870
Shared proportion of positively selected sites			-15402.724	93	1	0.624	0.430

Table F5: Parameter estimates for LRTs comparing selection on *AP2D* and all other S-linked genes. dN/dS, dS, dN, and P (proportion of sites belonging to a given rate class) estimates for all rate classes (+, -, and neutral evolution) for all models are given. These estimates correspond to LRTs that were performed with default starting values for all parameters.

Almnt	Independent Model					Shared Positive Selection Strengths Model				Shared Positive Selection Proportions Model				Shared Positive Selection Regime Model				Shared Selection Parameters Model			
	Rate Class	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P	dN/dS	dS	dN	P
<i>AP2D</i>	+	12.7	0.4	6.34	0.0	2.480	1.904	4.722	0.030	9.372	0.532	4.983	0.057	2.396	1.614	3.868	0.059	2.539	1.691	4.29	0.01
	Neut.	1.00	1.1	1.16	0.1	1.000	1.057	1.057	0.185	1.000	1.031	1.031	0.181	1.000	0.847	0.847	0.0197	1.000	0.862	0.86	0.14
	-	0.10	0.3	0.03	0.7	0.094	0.318	0.030	0.725	0.082	0.339	0.028	0.702	0.054	0.317	0.017	0.683	0.085	0.487	0.04	0.78
	-	0.00	9.0	0.00	0.0	0.000	8.595	0.000	0.060	0.000	9.148	0.000	0.060	0.000	8.523	0.000	0.061	0.020	9.214	0.18	0.05
<i>S-linked</i>	+	2.05	0.7	1.57	0.0	2.480	0.651	1.614	0.062	2.154	0.770	1.658	0.057	2.396	0.683	1.636	0.059	2.539	1.691	4.29	0.01
	Neut.	1.00	0.0	0.05	0.5	1.000	0.054	0.054	0.531	1.000	0.059	0.059	0.539	1.000	0.056	0.056	0.535	1.000	0.862	0.86	0.14
	-	0.06	1.2	0.08	0.3	0.071	1.196	0.085	0.392	0.070	1.210	0.085	0.389	0.071	1.196	0.085	0.391	0.085	0.487	0.04	0.78
	-	0.00	31	0.27	0.0	0.010	30.646	0.297	0.015	0.009	30.773	0.277	0.015	0.009	30.62	0.289	0.015	0.020	9.214	0.18	0.05

Table F6: Comparison of selection on *AP2D* and other *S*-linked genes. Four likelihood ratio tests were performed in order to determine if *AP2D* had experienced different selective pressures when compared to other *S*-linked genes. For each model (constrained and independent), the log likelihood value and number of estimated parameters is provided. For each test, the corresponding degrees of freedom (DF) and LRT statistic are shown. Significant p-values are indicated in bold. p-values < 0.05 suggest that the independent model fit is superior to that of the constrained model and, further, that the *AP2D* and *S*-linked gene data sets are significantly different with respect to the relevant constrained parameter(s). All four tests were completed twice, with random and default starting values for parameters, respectively. The results obtained using random starting values are shown below.

Parameters constrained in constrained model	Log Likelihood (Independent Model)	# of Parameters (Independent Model)	Log Likelihood (Constrained Model)	# of Parameters (Constrained Model)	DF	LRT Statistic	p-value
Shared distributions for all parameters	-15345.260	127	-15364.038	117	10	37.555	0.000
Shared selective regimes (dN/dS and proportions)			-15347.426	125	2	4.331	0.115
Shared strength of positive selection (dN/dS)			-15346.918	126	1	3.315	0.191
Shared proportion of positively selected sites			-15346.596	126	1	2.672	0.102

Appendix G: Adaptive Branch-Site Random Effects Likelihood (aBS-REL) Results

Table G1: Adaptive Branch-Site REL tests of Lineage-specific selection on *S*-linked genes (excepting *Tsst1l*). For each gene alignment, particular lineages that may have experienced episodic diversifying selection (EDS) were identified using Adaptive Branch-Site Random Effects Likelihood (aBS-REL) methods.

Branch and node names are consistent with the phylogenetic tree depicted in Figure 8. Holm-Bonferroni-corrected p-values are given for each branch for each gene. No significant results were obtained.

Branch	Gene Alignment										
	<i>APETALA2</i>	<i>LEJ2</i>	<i>AP2D</i>	<i>RNABP</i>	<i>SCE1</i>	<i>FRA1</i>	<i>LRRK</i>	<i>IRX115L</i>	<i>FSP</i>	<i>NRFP</i>	<i>WRKY</i>
D16L	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DROT 41S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MAN 601S	1.000	1.000	0.137	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
ES	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
SL8 201S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
F60SS	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MAN 713L	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.476	1.000	1.000	1.000
Node 16	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
MIDC 710S	1.000	1.000	1.000	1.000	1.000	1.000	0.115	1.000	1.000	1.000	1.000
COLO	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PA 4S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
E2 L	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TSH	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
E 207S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 27	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Continued from previous page.

Branch	Gene Alignment										
	<i>APETALA2</i>	<i>LEJ2</i>	<i>AP2D</i>	<i>RNABP</i>	<i>SCE1</i>	<i>FRAI</i>	<i>LRRK</i>	<i>IRX115L</i>	<i>FSP</i>	<i>NRFP</i>	<i>WRKY</i>
Node 23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 5	1.000	1.000	0.076	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CON 20S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
KRAP 5S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
KRAP 12L	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DEN 54L	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 34	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 32	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 30	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DEN 20S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TJ 30L	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
TJ 29S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 38	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CHAM 4L	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
DIF	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 41	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
WED 2S	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
PAN 2S	1.000	0.991	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Node 44	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table G2: Adaptive Branch-Site REL tests of Lineage-specific selection on *Tssta1*. For each *Tssta1* alignment (reduced and total taxa), particular lineages that may have experienced episodic diversifying selection (EDS) were identified using Adaptive Branch-Site Random Effects Likelihood (aBS-REL) methods. Branch and node names are consistent with the phylogenetic trees depicted in Figures 9 and 10, respectively. Holm-Bonferroni-corrected p-values are given for each branch. No significant results were obtained.

Alignment	Branch	p-value	Alignment	Branch	p-value	Alignment	Branch	p-value
<i>Tssta1</i> (Reduced)	F60SS	1.000	<i>Tssta1</i> (Total)	Node 12	1.000	<i>Tssta1</i> (Total)	TOC139	1.000
	ES	1.000		ES	1.000		VEL	1.000
	Node 11	1.000		Node 11	1.000		BAH	1.000
	MAN601S	1.000		F60SS	1.000		Node 43	1.000
	Node 10	1.000		Node 10	1.000		Node 41	1.000
	SL8	1.000		DROT41S	1.000		Node 39	1.000
	Node 9	1.000		MIDC710S	1.000		Node 31	1.000
	COLO	1.000		COLO	1.000		Node 5	1.000
	Node 8	1.000		PA4S	1.000		GRAN9S	1.000
	PA4S	1.000		Node 21	1.000		Node 4	1.000
	DROT41S	1.000		Node 19	1.000		PAN2S	1.000
	MIDC710S	1.000		Node 17	1.000		Node 3	1.000
	Node 19	1.000		Node 9	1.000		WED2S	1.000
	Node 17	1.000		KRAP5S	1.000		Node 2	1.000
	Node 7	1.000		E207S	1.000		PCARO	1.000
	CON20S	1.000		CON20S	1.000		PSAR	1.000
	Node 6	1.000		Node 26	1.000		PREV	1.000
	KRAP5S	1.000		Node 24	1.000		Node 53	1.000
	Node 5	1.000		Node 8	1.000		PNAN	1.000
	TSH	1.000		TSH	1.000		Node 52	1.000
	Node 4	1.000		Node 7	1.000		VIS	1.000
	E207S	1.000		TJ29S	1.000		PLIC	1.000
	Node 3	1.000		Node 6	1.000		PMOR137S	1.000
	DEN20S	1.000		ORI	1.000		PDUART1S	1.000
	Node 2	1.000		AUR	1.000		Node 61	1.000
	TJ29S	1.000		Node 34	1.000		Node 59	1.000
	WED2S	1.000		OCC	1.000		Node 57	1.000
PAN5S	1.000	Node 33	1.000	Node 51	1.000			
Node28	1.000	CUN	1.000	Node 49	1.000			
Tssta1 (Total)	SL8201S	1.000	Node 32	1.000	EOD	1.000		
	MAN601S	1.000	QUACO1S	1.000				

Appendix H: *Tssta1* in a Localized Population of *T. scabra* from the Dominican Republic (DROT)

DROT 2S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 5S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 6S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 7S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 9S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 10S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 14S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 19S -----TCAGCCAATAGC TCCTTGAAGTTTGAT 105
DROT 20S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 22S -----AGC TCCTTGAAGTTTGAT 105
DROT 25S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 28S -----TCAGCCAATAGC TCCTTGAAGTTTGAT 105
DROT 31S -----TCAGCCAATAGC TCCTTGAAGTTTGAT 105
DROT 33S -----AGC TCCTTGAAGTTTGAT 105
DROT 39S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105
DROT 40S -----TAGC TCCTTGAAGTTTGAT 105
DROT 41S ATGTCATCCCCTACC GGACGGGAGCTTACC ATCTATCTCTATCTC TTTTCTATTATTGCG GTCTTCCCATTTCCTT ACTTCAGCCAGTAGC TCCTTGAAGTTTGAT 105

DROT 2S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 5S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 6S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 7S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 9S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 10S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 14S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 19S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 20S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 22S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 25S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 28S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 31S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 33S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 39S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 40S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210
DROT 41S CTATTTGGCAACATG TACAGAGTTCATGTC ATCAATGGCTTCAGC AGCAATGACCTGCCA TTTTACTTCATTGT TGGTCAAGCAATGAC GATCTGGGGCACCAT 210


```

DROT 2S      TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 5S      TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 6S      TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 7S      TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 9S      TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 10S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 14S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 19S     TCCTCCTATTTCAAG TTG----- 447
DROT 20S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 22S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 25S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 28S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 31S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 33S     TCCTCCTATTTCAAG TTG----- 447
DROT 39S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 40S     TCCTCCTATTTCAAG TTGTATGACTGG 447
DROT 41S     TCCTCCTATTTCAAG TTGTATGACTGG 447

```

Figure H1: Alignment of *Tssta1* sequences from 17 short-styled individuals belonging to a localized population of *T. scabra* from the Dominican Republic (DROT). Base positions showing 100% identity across taxa are shown in blue. Where sequence information is present, the sequences were found to be completely identical across individuals in the coding sequence for *Tssta1*. For some individuals, upstream sequence information was also obtained, from which some sequence differences between individuals could be identified (data not shown). The alignment is sectioned into groups of 15 bases (or 5 codons). The length of the alignment (bp) is given in the right-most column (447 bp, total).

Table H1: The degree of *Tssta1* sequence similarity between samples from a local population of *T. scabra* (DROT) and *T. panamensis* (PAN 2S). The local population sample (DROT) consisted of 17 short-styled individuals of the species *T. scabra* from the Dominican Republic. The numbers of synonymous and non-synonymous sites identified as being fixed between the two species and polymorphic within the population are shown below. McDonald-Kreitman tests and Tajima's D tests of neutrality could not be performed on the data due to the lack of polymorphism within the DROT population data set.

	Fixed	Polymorphic
Synonymous	17	0
Nonsynonymous	8	0