IDENTIFYING THE ECOLOGICAL INTERACTIONS OF INVASIVE *LYMANTRIA DISPAR DISPAR* AND *AGRILUS PLANIPENNIS* USING METABARCODING TECHNIQUES

NICOLE D. BORSATO

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 2024

Abstract

The objectives of this thesis were to develop a standardized metabarcoding protocol to identify ecological interactions between insects and other taxa, as well as determine the impact of ecological correlates on the detection of these interactions. In Chapter 2, I evaluate target loci for the identification of *Lymantria dispar dispar* (spongy moth) and *Agrilus planipennis* (emerald ash borer) ecological interactions and develop a standardized protocol for the analysis of these interactions. I demonstrate that my metabarcoding protocol can be used to identify a wide diversity of interactions, including those with animals, plants, bacteria, and fungi. In Chapter 3, I assess how ecological factors (e.g., life stage, collection location, etc.) influence the presence or absence of ecological interactions in emerald ash borers using random forest models. I found that interactions were more likely to be detected in pupal specimens and in samples collected earlier in the year.

Acknowledgments

This thesis would not have been possible without the support of many individuals. First and foremost, I would like to thank my supervisor Dr. Elizabeth Clare for her continuous support, guidance, and invaluable feedback throughout my study. I also wish to thank Dr. Dirk Steinke, Dr. Eryn McFarlane, Dr. Mark Vicari, and Dr. Marina Freire-Gormaly for their insightful advice and constructive feedback on earlier drafts of my thesis. My thanks also go to Robin Floyd and the collection team at the Center for Biodiversity Genomics for their assistance in collecting and sequencing samples.

I am deeply grateful to my friends and lab mates Katherine Lunn, Nina Garrett, Gliselle Marin, Alejandro Biganzoli Rangel, Mateus Pepinelli, and Daniel Marquina for their camaraderie, support, and the countless laughs and good times we've shared both in and out of the lab. I also want to express my appreciation to my friends Cecilia Cancellara, Jaiden Muoio, Melissa Green, Ellie Burt, Kelson Young, and Blanka Maršič for helping me take breaks from my research and always being willing to lend an ear. I would also like to thank my partner Dalton Tracy for his unwavering support and belief in me, and for always reminding me of my capabilities and keeping me motivated.

A special thanks to my family, particularly my cousins Sarah and Chloe Borsato, for their support and the memorable adventures that provided much-needed relaxation. Last, but certainly not least, I am especially thankful to my parents, Linda Borsato and Russell Hurlbut, and my nonna Antonietta Borsato, for their endless love, support, and encouragement throughout my academic journey and beyond.

Tables of Contents

List of Figures

Chapter 2

Chapter 3

Appendix F

List of Tables

Chapter 2

Appendix A

Chapter 1: Introduction

Ecological Interactions: Symbiomes vs. Holobionts

Every individual interacts with other species, whether directly (e.g., one species feeding on another) or indirectly (e.g., trophic cascades; Wootton, 1994). These interactions are essential to survival and form the complex network structure of ecological communities influencing their resilience to disturbance (Angelini et al., 2016; Hensel et al., 2021) and even the diversification of life (e.g., the conversion of endosymbionts to mitochondria/chloroplast organelles; Roger et al., 2022; Stadnichuk & Kusnetsov, 2021; Thompson, 1999). These interactions include collections of organisms (i.e., bacteria, fungi, parasites, etc.) that interact with the organism of interest. When these interactions are strong, they may cause reciprocal genomic adaptations and co-evolution, but even weak interactions can have ecological consequences.

Many of these interactions result in the transfer of DNA from one species to another (e.g., an herbivore would have the DNA of the plants it consumed in its stomach), which allows DNA barcoding technology to be employed for identification of the interaction. Interactions which leave behind DNA traces have been split into two broad categories. The “symbiome” refers specifically to interactions which cause co-evolutionary changes. Those species which are simply co-occurring are referred to as “holobiont” (Tripp et al., 2017). While these are clear categories, in practice they are extremely hard to distinguish without whole genome studies to determine the degree of reciprocal co-evolution. Indeed, there will also be interactions which could be classified as both, particularly secondary

interactions where one partner is under selection and the other is not. In reality, the distinction becomes artificial in the face of the complexity of ecological interactions.

DNA Metabarcoding for the Identification of Ecological Interactions

DNA barcoding allows an organism to be identified by a short fragment of DNA (referred to as a DNA barcode; Hebert et al., 2003; Savolainen et al., 2005). DNA barcodes are generated from standardized gene loci that vary significantly between species, such as the mitochondrial gene for cytochrome c oxidase subunit 1 (COI; used for animal identifications; Hebert et al., 2003) and the rbcL gene (used for plant identifications; CBOL Plant Working Group, 2009). DNA metabarcoding works similarly, except it employs high-throughput sequencing to allow the identification of multiple taxa (i.e., DNA barcodes) simultaneously from mixed DNA templates (Compson et al., 2020).

DNA-based identification has grown in popularity for taxonomic identification in ecological assessments over traditional methods (e.g., morphological identification) as it is quick, cost efficient, and can capture more diversity (e.g., by distinguishing between cryptic species; Elbrecht et al., 2017; Grant et al., 2021; T. Wang et al., 2020). Like all taxonomic identification approaches, there are limitations. For instance, DNA-based identification relies on reference libraries that are incomplete (i.e., not all taxa have been barcoded or are represented in the reference database; Grant et al., 2021; Weigand et al., 2019) and taxon abundance cannot be determined in biodiversity assessments (Elbrecht & Leese, 2015). Despite these drawbacks, metabarcoding has comparable results to

morphological assessments (Elbrecht et al., 2017; Lejzerowicz et al., 2015) with the added advantage of identifying multiple taxa quickly.

Thesis Structure

This thesis focuses on identifying ecological interactions with holobionts and symbionts using metabarcoding techniques. In Chapter 2, I develop a general metabarcoding approach that can be applied to a variety of target insects. The objectives were to 1) evaluate different target loci for the identification of species interactions, and 2) recommend targets and methods for a general approach to symbiome and holobiont analysis. Using two invasive insects, spongy moths (*Lymantria dispar dispar*) and emerald ash borers (*Agrilus planipennis* Fairmaire, 1888), as target organisms, I evaluated six primer pairs covering five gene loci (COI, ITS, 16S, rbcL, and 18S) to identify a wide variety of interactions (e.g., fungal, microbial, parasitoid, etc.).

Using the emerald ash borer interaction data produced in Chapter 2, I assess the impact ecological factors have on these interactions in Chapter 3. My objective was to assess the influence life stage, collection location, collection date, and collection habitat have on the detection of interactions. I generated random forest models and Accumulated Local Effects (ALE) plots to determine the impact each factor has on the detection of each interaction type (animal, bacteria, fungi, and overall interactions). I also demonstrate how the presence of biological controls can be identified using metabarcoding.

Chapter 2: Methodological Approach to Symbiome Analysis

Abstract

Spongy moths (*Lymantria dispar dispar*) and emerald ash borers (*Agrilus planipennis*) are destructive invasive insects that increase tree mortality across North America, resulting in major ecological and economic impacts. Trees that are infested with spongy moths and emerald ash borers experience mortality rates which can exceed 90%, especially when exposed to other stressors. Tree mortality disrupts ecological communities and threatens the $218 billion forestry industry in North America. Given the ecological and economic impact these insects have, it is important to identify species interactions to gain a better understanding of these impacts, potential biological control methods, and of what influences their population density. These species interactions can be characterized by the collections of organisms that interact with the individuals of the organism of interest. Many of these interactions result in the transfer of DNA from one species to another, which allows metabarcoding technology to be employed for symbiome identification. My objectives were to 1) evaluate different target loci (e.g., COI, ITS, 16S, etc.) for the identification of ecological interactions in these insects, and 2) recommend targets and methods for general approaches for symbiome and holobiont analysis. I generated PCRs from spongy moths (eggs and larvae) and emerald ash borers (larvae, pupae, and adults) using multiple primers to amplify different taxonomic targets (animals, plants, microbes, and fungi) within the symbiome. All PCRs were independently barcoded and sequenced before being processed using the DADA2 pipeline in R Studio (for ITS and rbcL), QIIME2 (for 16S), mBRAVE (for COI), or Dorado/R (for 18S). Using these combined approaches, I detected the presence of taxa representing 60 orders of plants,

insects, fungi, and microbes in samples of spongy moths, and taxa representing 30 orders in the DNA extracts from individual emerald ash borers. My data suggest an extremely efficient method of assessing broad scale ecological interactions of individuals within known destructive pest species which can be generalized to other insect groups in the environment.

Introduction

Metabarcoding to Identify Symbiomes

Every individual interacts with other individual organisms in some capacity, whether it be directly (e.g., one species feeding on another) or indirectly (e.g., nutrient transfer within a food chain; Boersma et al., 2008; Fretwell, 1987). These interactions are essential to ecosystem functioning but can be challenging to identify. One suggestion is to quantify species interactions by characterizing the “symbiome” and, using molecular tools, this could be accomplished through simultaneous amplification of multiple gene loci targeting different taxa. The symbiome refers to the collection of organisms (i.e., bacteria, fungi, parasitic animals, etc.) that interact with an individual of interest (e.g., by residing in or on a host organism) and are both colocalized (co-occurring) and co-evolving. It has been subtly differentiated from a holobiont which also includes physically associated taxa but not necessarily involving co-evolution (Tripp et al., 2017). Many of these interactions, both in the holobiont and symbiome, result in the co-occurrence of DNA (e.g., an herbivore would have the DNA of the plants it consumed in its stomach, but may also carry the DNA of species in the environment), which allows DNA barcoding and metabarcoding to be employed to identify these elusive interactions. These concepts differ slightly from each

other, and the symbiome may be a specific subset of the holobiont, but may also differ from concepts which consider repeated interactions (e.g., dietary ecology; Pompanon et al., 2012) or pooling of data from individuals (Drinkwater et al., 2019) where species-level measurements are the net target and the individual is not quantified. The objective of a molecular symbiome approach is to characterize the interactions of a specific individual at a specific point in time and differentiate this from more casual co-occurrence, though in practice this is functionally difficult, and it is likely these types of interactions exist not in a discrete category but on a continuum from co-occurrence to co-evolution with gradually increasing interaction strength and selection profiles. Indeed, true co-evolution requires reciprocal genomic adaptations, a feature that is extremely hard to measure. The use of metabarcoding for the identification of ecological interactions relies on simultaneous amplification of standardized gene loci that vary significantly between species from the same DNA extract, such as the mitochondrial gene for cytochrome c oxidase subunit 1 (COI; used for animal identifications; Hebert et al., 2003) and the rbcL gene (used for genus level plant identifications; CBOL Plant Working Group, 2009). Sequencing generally employs high-throughput methods which allow for the identification of multiple taxa simultaneously from mixed samples (Compson et al., 2020). Metabarcoding can thus be used to characterize the interactions of ecologically or economically important species through the use of multiple target loci.

Targeting Insect Symbiomes

An ideal method for analysis would apply across a variety of animal targets of interest with minimal methodological variation. While a great many target genes have been

employed to diagnose species interactions (Bell et al., 2017; Clare et al., 2018; Pornon et al., 2016; Pushpakumara et al., 2023), fewer analyses have compared the recovery of these targets across taxonomic groups. In this analysis I develop a generalized analysis method for targeting ecological interactions by comparing multiple target genes for the recovery of data from insect groups of substantial economic importance in North America, invasive spongy moths (*Lymantria dispar dispar*) and emerald ash borers (*Agrilus planipennis* Fairmaire, 1888).

Spongy Moths

Spongy moths were introduced to North America in 1869 and have caused substantial damage to more than 300 species of trees by defoliation during their caterpillar stage reducing growth, root, and fruit production (Davidson et al., 1999; Elkinton & Liebhold, 1990; Kozlowski et al., 1991; Liebhold et al., 1995; Muzika & Liebhold, 1999; Nakajima, 2015). Defoliated trees are often weaker and susceptible to dying with both economic and ecological impacts (Davidson et al., 1999). Defoliation reduces timber production (Leuschner et al., 1996), and in 2021 roughly 1.8 million hectares of forests in Ontario were defoliated by spongy moths (Government of Ontario, 2014). Between timber loss and biological controls, spongy moths cost North America nearly $3.2 billion annually (Bradshaw et al., 2016).

In addition to direct economic costs, spongy moth outbreaks cause disruption of ecological communities and food webs, reducing native moth populations and altering the structure and composition of forests (Canham, 1988; Fajvan & Wood, 1996; Gurnell,

1983; Work & McCullough, 2000). Efforts to control spongy moth populations and to prevent their spread include manually removing and killing the insects, the application of pesticides, and the introduction of biological controls (i.e., natural enemies of the moths, such as the parasitic wasp *Ooencyrtus kuvanae* and the fungus *Entomophaga maimaiga*; Elkinton & Liebhold, 1990; Mcmanus & Csóka, 2007). Spongy moth populations fluctuate annually, with large-scale outbreaks occurring synchronously across large distances every 8 to 10 years, and smaller outbreaks every 4 to 5 years (Berryman, 1996; Davidson et al., 1999; Haynes et al., 2009; Peltonen et al., 2002). The cause(s) of these outbreaks have not been determined, but ecological interactions have been hypothesized to play a role (Allstadt et al., 2013; Elkinton & Liebhold, 1990). As a result, determining large-scale species interactions may facilitate better predictions of outbreaks and present mechanisms for population control.

Emerald Ash Borers

The emerald ash borer is a wood-boring beetle native to northeastern Asia. It was first detected in North America near Detroit, Michigan in 2002, although it may have been undetected since the 1990s (Cappaert et al., 2005; Siegert et al., 2014). Emerald ash borer larvae consume the phloem and cambium of ash trees (*Fraxinus* spp.) creating S-shaped galleries that disrupt nutrient transport (Cappaert et al., 2005; X.Y. Wang et al., 2010). Infestations are associated with extensive canopy dieback, especially in green ash (*Fraxinus pennsylvanica*; Anulewicz et al., 2007), causing tree mortality often exceeding 99% within 6-10 years of beetle arrival (Klooster et al., 2014; Knight et al., 2013; Natural Resources Canada, 2013). As a direct result, six North American ash species have been

added to the IUCN Red List of Threatened Species (International Union for Conservation of Nature, 2022). Since the beetle's introduction, tens of millions of trees have been lost (Cappaert et al., 2005) and they are estimated to be the costliest forest pest in North America, causing approximately $850 million in local government expenditures (i.e., tree removal, replacement, and treatment) and $380 million in lost residential property values each year in the USA alone (Aukema et al., 2011).

Ecological impacts include reduced availability of food and shelter for organisms that rely on ash trees (e.g., beavers, rabbits, deer, purple finches, and wood ducks; Schlesinger, 1990) and extinction risks in dependent species (e.g., 43 monophagous native arthropod species like the eastern ash bark beetle [*Hylesinus aculeatus*] and black-headed ash sawfly [*Tethida barda*]; Gandhi & Herms, 2010b). Like spongy moths, emerald ash borer infestations alter the structure and composition of forests, for example canopy dieback allows increased light penetration to the subcanopy, encouraging the growth of understory species and altering their distribution and abundance (Canham, 1988; Flower et al., 2013; Gandhi & Herms, 2010a).

In this study I examine potential target loci and protocols to develop a general metabarcoding approach which can be applied to a variety of insect targets. My objectives were to 1) evaluate different target loci for the identification of ecological interactions in these taxonomic groups, and 2) recommend targets and methods for general approaches for symbiome and holobiont analysis.

Methods

Sample Collection and Study Design

All samples used in this analysis were acquired from the collections of the Center for Biodiversity Genomics, University of Guelph, Canada. To simplify methodological testing, I developed the protocol using a hierarchical testing design. I started with spongy moths as pooled samples following the approach for Drinkwater et al. (2019). While this does not assess individual interactions, I used this approach to increase DNA yield for initial primer testing. I tested a subset of primers on these extracts (Table 1). I then used an individual approach for a large set of individual emerald ash borers while also reducing reaction volumes.

Spongy moths

I acquired 18 samples of spongy moth eggs and larvae. A sample consisted of 4-5 individuals from the same tree which were pooled during extraction to increase DNA yield (Drinkwater et al. 2019), thus 18 samples represented many individuals (see Table 1). These samples had been hand collected from the University of Guelph Arboretum (Guelph, Ontario, Canada) in May 2021 from a variety of host trees, including maple (n =1), larch (n =1), poplar (n =1), birch (n =2), pine (n =2), beech (n =1), and cherry (n =1). The specimens were stored in 95% ethanol and frozen.

Emerald ash borers

I acquired 277 samples of emerald ash borer larvae (3 sizes hereafter larva1, larva2, larva3), pupae, and adults (Appendix A Table S1). These samples had been hand collected from ash trees found in various locations across Brockville (n = 185), Ennotville (n = 41), Guelph (n = 41), and Puslinch (n = 10) in Ontario, Canada (Appendix A Table S2). Of the 277 specimens, there were 60 adults (6 were known to be parasitized/fungal infested), 53 pupae, 54 size 1 larvae (< 0.5in; 11 parasitized/fungal infested), 68 size 2 larvae (0.5-1in; 21 parasitized/fungal infested), 41 size 3 larvae (> 1in; 1 parasitized/fungal infested), and one specimen of an unknown life stage with an observed symbiont worm. I specifically included individuals with observed fungal or parasite interactions to increase the complexity for amplicon targeting. The specimens were stored in 95% ethanol and frozen.

DNA Extraction

For both spongy moths and emerald ash borers I extracted DNA using the Qiagen blood and tissue kit following the manufacturer's guidelines, with two modifications. First, specimens were ground using a pestle during initial lysis, and second, I decreased the final DNA elution to 100uL of Buffer AE for spongy moths and 150ul of Buffer AE for emerald ash borers to increase DNA yields. For spongy moth extractions I used 4-5 eggs or larvae from the same host tree in order to get sufficient biomass to extract DNA, thus each “sample” is a pool of 4-5 individuals (e.g. see the method used by Drinkwater et al., 2019). For emerald ash borers I extracted DNA from the posterior half of the abdomen for adult specimens, the posterior third for pupa specimens, and an approximately 6mm

fragment of the posterior end of larva specimens (or the whole specimen if classified as larvae size 1). Extraction blanks were included using sterile water.

Amplification of Target Gene Loci

Spongy Moths

A total of 100 PCRs were generated using multiple primers to amplify different taxonomic targets (animals, plants, microbes, and fungi) within each extract. Positive controls (cockroach for COI and 16S, cannabis for ITS and rbcL) and negative controls (water) were included with each amplification. All primers had been modified with CS1/CS2 tails for sequencing using the Illumina MiSeq platform and diluted to 10 uM (see Table 1).

rbcL (primers rbcl_1-8/rbcLB) and COI (primers ZBJ-ArtF1c/ZBJ-ArtR2c): Each 20 µl reaction contained 12.5 µl Qiagen multiplex PCR master mix, 4.5 µl water, 1.25 µl forward primer, 1.25 µl reverse primer, and 0.5 µl DNA template (or water for the negative control). The PCR thermocycling conditions were as follows: 5 minutes at 95℃, followed by 35 cycles of 30 seconds at 95℃, 45 seconds at 50℃, and 50 seconds at 72℃, followed by 5 minutes at 72℃. The same protocol was used for the COI primers, except only 25 cycles were run.

ITS plant and fungal target (primers ITS-S2F/ITS4): Each 25 µl reaction contained 12.5 µl Qiagen multiplex PCR master mix, 9 µl water, 1.25 µl forward primer, 1.25 µl reverse primer, and 1 µl DNA template (or water for the negative control). The PCR thermocycling conditions were as follows: 4 minutes at 94℃, followed by 34 cycles of 30

seconds at 94℃, 40 seconds at 55℃, and 1 minute at 72℃, followed by 10 minutes at 72℃ (Cheng et al., 2016).

16S microbial target (primers 799F/1115R): Each 20 µl reaction contained 12.5 µl Qiagen multiplex PCR master mix, 5 µl water, 0.75 µl forward primer, 0.75 µl reverse primer, and 1 µl DNA template (or water for the negative control). The PCR thermocycling conditions were as follows: 2 minutes at 95℃, followed by 35 cycles of 20 seconds at 95℃, 30 seconds at 48℃, and 30 seconds at 72℃, followed by 3 minutes at 72℃ (adapted from Hanshew et al., 2013).

Emerald Ash Borers

A total of 1099 PCRs were generated using multiple primers to amplify the same target lineages (animals, plants, bacteria, and fungi) within the emerald ash borer symbiome. Extraction controls and positive (cricket for COI; pepper for ITS and 16S) and negative (water) PCR controls were included with each amplification. Based on amplification success rates in spongy moths I modified the choice of primers (see Table 1), reaction volumes, and amplification protocols to reduce costs and simplify the analysis for broader application in ecological analysis. I also evaluated a variety of 18S loci to include parasite interactions.

COI insect target (primers ZBJ-ArtF1c/ZBJ-ArtR2c), ITS plant and fungal target (primers ITS3/ITS4), and 16S microbial target (primers 799F/1115R): Each 15 µl reaction contained 7.5 µl Qiagen multiplex PCR master mix, 4.5 µl water, 1 µl forward

primer, 1 µl reverse primer, and 1 µl DNA template (or water for the negative control). The PCR thermocycling conditions were as follows: 5 minutes at 95℃, followed by 35 cycles of 30 seconds at 94℃, 45 seconds at 52℃, and 1 minute at 72℃, followed by 10 minutes at 72℃. All primers had been modified with CS1/CS2 tails for sequencing using the Illumina MiSeq platform.

18S parasite target (primers SSU_F_07/SSU_R_26, formerly SSU18A and SSU26R): 18S targets present a tradeoff between shorter reads with only higher-level taxonomic assignments or longer reads with better taxonomic resolution but incompatible with most high-throughput sequencing platforms. I evaluated several 18S primers (563F/1132R, 1391F/EukBr, and PFAExF/PFAExR) but none can easily be adapted for the most common HTS platforms or lack reference collections which can identify sequences below very high order classifications (e.g., phyla). For emerald ash borers I used the protocols for high-throughput barcoding on the MinION sequencing platform (Oxford Nanopore Technologies) with primer modifications as described in Hebert et al. (2023). Each 12.5 µl reaction contained 6.25 µl 10% trehalose (Fluka Analytical), 1.25 µl 10X PlatinumTaq buffer (Thermo Scientific), 1.62 µl water, 0.625 µl of 2 µM forward primer, 0.625 µl of 100 µM reverse primer, 0.625 µl 50nM $MgCL_2$, 0.0625 µl of 10nM dNTPs (Kapa Biosystems), 0.06 µl of 5U/ul PlatinumTaq (Thermo Scientific), and 2 µl DNA template. The PCR thermocycling conditions were as follows: 2 minutes at 94℃, followed by 35 cycles of 40 seconds at 94℃, 40 seconds at 56℃, and 1 minute at 72℃, followed by 5 minutes at 72℃.

Sequencing and Analysis

Amplified DNA was sent to the Barts and the London Genome Center at Queen Mary University of London's Blizzard Institute, with the exception of 18S fragments (see below). All samples sent were processed identically including bead clean up and size selection by 0.9 X Ampure Beads, quantification, QC and normalization using Qubit nucleotide quantification, and DNA D100 Tape station. All PCRs were independently barcoded using single indexes and sequenced using a MiSeq v3 600 cycle run. Raw sequences were demultiplexed on site.

Demultiplexed ITS and rbcL PCR files were processed using the DADA2 pipeline (Callahan et al., 2016) in R Studio (R Core Team, 2023) following Garrett et al. (2023; see Appendix B). Cutadapt 3.7 (Martin, 2011) was used in paired end mode to remove primers and to separate pooled data by primer. In brief, DADA2 was used to merge paired reads, trim sequence length, and remove errors in sequence profiles using the learnErrors function. Chimeras were removed and ASV tables were generated for each primer pair and each taxonomic group individually. Resulting ASVs for ITS and rbcL sequences were compared to the nucleotide collection in GenBank using the BLAST algorithm. An identification was retained for further consideration if the identity match was ≥97% (with 100% overlap) and if the detection count was greater than the highest read count in the negative control (negative filtering). I then screened taxonomic identity against available ecological data on distributions (i.e., the organism can be found in the area the samples were collected in according to GBIF [GBIF, 2023]). Several interactions of specific

interests which met some but not all of these conservative criteria are also reported but with cautionary notes (see results).

16S reads were processed using QIIME2 (Bolyen et al., 2019; see Appendix C). Sequences were trimmed using Cutadapt 3.7 (Martin, 2011) and denoised using DADA2 (Callahan et al., 2016). The denoising parameters (trim-left-f, trim-left-r, trunc-len-f, and trunc-len-r) were as follows: 0, 0, 268, and 218 for emerald ash borers; 0, 0, 271, and 202 for spongy moths. ASV tables were then generated in R Studio (R Core Team, 2023; see Appendix D). An identification was retained if the detection count was greater than the highest read count in the negative control, if it was not associated with the positive control, and if the taxonomic identity matched known ecological data (i.e., the organism is commonly associated with soil/water/plants or is an insect pathogen according to MicrobeAtlas [v. 1.0; Matias Rodrigues et al., 2017] and is thus unlikely to be a human/lab contaminant).

COI sequences were processed in mBRAVE (Ratnasingham, 2019) using the default parameters except for the following: trim front = 30bp, trim end = 24bp, min QV = 0qv, max length (pre-trim) = 600 bp, max bases with low QV (<20) = 75%, max bases with ultra-low QV (<10) = 75%, ID distance threshold = 5%, exclude from OTU threshold = 5%, read sub-sampling - max reads per sample = 2500, read sub-sampling - max reads per contig = 200, paired end merging = merge, assembler min overlap = 10bp, and assembler max substitution = 20bp. The following system reference libraries were used: Insecta (SYS-CRLINSECTA), Non-Insect Arthropoda (SYS-CRLNONINSECTARTH), Non-

Arthropoda Invertebrates (SYS-CRLNONARTHINVERT), and Chordata (SYS-CRLCHORDATA). An identification was retained for further evaluation if the mean length fell within the expected range (i.e., expected amplicon length +/- 1bp), if “pc_sim_mean” was ≥97%, and if “id_gaps_mean” equaled zero. I used full negative filtering using the negative controls. All putative taxonomic identifications were compared against known distribution records (e.g., in GBIF [GBIF, 2023]). *Tylonotus bimaculatus* and *Hylesinus aculeatus* were detected in 13 presumed emerald ash borer samples and these were removed from further analysis as they were likely misidentified during the collection stage.

Because the 18S amplicons exceed the length which can be processed on Illumina platforms (>900 bp for the target nematodes) 18S emerald ash borer amplicons were sequenced using the MinION platform (Oxford Nanopore) with library prep carried out as described in Floyd et al. (2023) and run on a Flongle flow cell. Data were analyzed using the methods detailed in Hebert et al. (2023). To assign putative taxonomic identification to sequences, the SINTAX tool (Edgar, 2016) was employed, with a custom reference library containing the nematode sequences from 18S NemaBase (Gattoni et al., 2023) and sequences of all other taxa from the SILVA public database.

Results

Spongy Moths

I processed 1,460,234 COI, 199,979 ITS, 539,627 rbcL, and 292,782 16S raw reads. These were reduced to 281, 92,616, 377,885, and 18,091 filtered reads which were converted to 15, 167, 42, and 97 BINS (COI) or ASVs respectively (spongy moth host

reads were excluded from COI data). I identified 126 taxa across the four gene loci (ITS = 49; 16S = 41; COI = 8; rbcL = 28). Of these, 119 were unique taxa (plants identified by both rbcL and ITS were not counted twice). Most identifications were to genus, and these represented 60 different orders, with two unclassified algae and an uncultured microbe included in this count. I identified 19 orders (6 fungi and 13 plants) using the ITS locus, with eggs having 14 orders and larvae 17 orders (Appendix A Table S4). Using the 16S locus I identified 28 bacteria orders, with 20 and 22 orders detected in the egg and larvae life stages, respectively (Appendix A Table S5). I detected *Wolbachia* sp. (order Rickettsiales), an insect pathogen, in 5 samples. Using the COI target locus, I detected 6 orders of insects, other than the host itself, with 4 orders found in the egg stage and 5 orders found in the larval stage (Appendix A Table S6) representing potential parasites. I detected 15 plant orders in rbcL (8 of which were also detected using ITS), with eggs containing plant DNA from 11 orders and larval extracts containing plant DNA from 13 (Appendix A Table S7).

These detections from 60 orders accounted for 317 actual detections across the two life stages (Figure 1.1). There were 150 order level detections in the egg life stage (6 animal, 64 bacteria, 14 fungi, and 66 plant), while 167 detections were at the larval life stage (9 animal, 60 bacteria, 21 fungi, and 77 plant).

Emerald Ash Borers

I processed 1,1361,252 COI, 2,443,232 ITS, 6,173,341 16S, and 287,180 18S raw reads. These were reduced to 114,088, 942,740, 8154, and 8126 filtered reads which were

converted to 6, 36, 38, and 31 BINS (COI, 18S) or ASVs respectively (emerald ash borer host reads were excluded). I identified 65 taxa across the four gene loci (ITS = 23; 16S = 21; COI = 4; 18S = 17. These represented 61 unique identifications from 30 orders. I detected 11 orders of fungi using the ITS locus, with larva1 having a richness of 5 orders, larva2 7 orders, larva3 3 orders, pupae 6 orders, and adults 8 orders (Appendix A Table S8). Notably, I detected the fungal biocontrol agent *Beauveria* sp. (order Hypocreales) in two samples. However, it should be noted that this was also detected in one of the negative controls. I retained the identification here because it was present in very small amounts in this control compared to those detected in the samples and there is no other known source of the fungi in my processing. It was also independently verified by 18S data (described below) supporting it as a true positive. Using the 16S locus, I detected 13 orders of bacteria. I identified 5 different orders in the larva1 stage, 3 in the larva2 stage, 2 in the larva3 stage, 8 in the pupal stage, and 9 in the adult stage (Appendix A Table S9). I detected the insect pathogen *Wolbachia* sp. (order Rickettsiales) in 1 sample. I also detected 4 arthropod orders using the COI target locus including potential parasites. There were no ASVs retained from larva1, larva2 had 2 orders identified, the larva3 stage and pupal stage had 1, and the adult stage had 2. Notable identifications included Ichneumonidae parasitoid wasps (Appendix A Table S10). With the 18S locus, 9 orders were detected, with 6 of the 9 representing fungi (overlapping with those identified using ITS) and the remaining 3 representing insects, nematodes, and a parasitic alveolate protist (*Cryptosporidium* sp., order Cryptosporida). In the larva1 stage I identified 3 orders, larvae2 5 orders, larvae3 1 order, pupae 4 orders, and the adult stage 2 orders (Appendix A Table S11).

These detections from 30 orders accounted for 259 actual detections across life stages (Figure 1.2). There were 29 order level interactions detected in larva1 (3 animal, 17 bacteria, and 9 fungi), 64 in larva2 (8 animal, 24 bacteria, and 32 fungi), 10 in larva3 (2 animal, 3 bacteria, and 5 fungi), 74 in the pupae (5 animal, 31 bacteria, and 38 fungi), and 82 in the adult beetles (4 animal, 48 bacteria, 29 fungi, and 1 parasitic alveolate).

Discussion

In this chapter I compare the performance of several gene loci for the identification of individual cross species interactions. I used six primer pairs from five gene loci targeting insects, fungi, plants, microbes, and parasites to analyze the ecological interactions of commercially destructive spongy moths and emerald ash borers. My objective was to develop a standardized protocol for the analysis of ecological interactions at the individual level from single DNA extracts for the recovery of symbiont and holobiont descriptions. My data demonstrate that common DNA extracts can be successfully used to amplify a variety of targets. For simplification of approaches the ITS locus recovers a wider range of plant taxa than rbcL and provides better taxonomic resolution, while simultaneously amplifying fungal taxa, however there are taxa uniquely amplified by rbcL and not ITS. I identified common interactions, e.g., the frequent detection of *Wolbachia*, and interesting potential agents of biocontrol such as the detection of *Beauveria* in several samples of emerald ash borer. The major challenge in all loci is whether detections are true ecological interactions, a symbiome, or co-occurring DNA signals, a holobiont, which may result from

surface contamination of specimens with DNA from co-occurring taxa that do not directly interact.

Confirming Target Insect Identification

COI is an extremely efficient and well-known method for the identification of insects (Boehme et al., 2012; Jung et al., 2011; Park et al., 2011) widely used in DNA barcoding as the primary marker for taxonomic identification of animals (Hebert et al., 2003). Here I use it with a dual purpose, to confirm the identity of the host and to identify potential arthropod parasites. Among emerald ash borers COI detected 4 orders of arthropods. Unsurprisingly, the vast majority of the detections were to the host (e.g., *Agrilus planipennis*) and host DNA dominated the sequencing with only trace amounts of DNA from other organisms in terms of relative sequencing read counts. While I do not use read counts to estimate abundances, I did use strict filtering protocols from negative controls which resulted in artificially reduced detections of non-host taxa. An alternative is more relaxed filtering or not filtering but tracking potential contaminants however this raises issues around non-relevant detections (see below discussion of environmental DNA). Among emerald ash borer samples, I detected two families of mites (Oppiidae order Sarcoptiformes and Eupodidae order Trombidiformes) and one family of gall midge (Cecidomyiidae, order Diptera). Gall midges are primarily plant pests, and while some species can act as parasites of other arthropod larvae, I could not identify these further. These are likely co-occurring taxa rather than parasites. Notably, I detected a family of parasitoid wasps (Ichneumonidae, order Hymenoptera) in larva2, larva3, and pupa

samples. Ichneumonids have previously been observed to parasitize emerald ash borers (Duan et al., 2009) making these highly likely cases of true parasitism.

The Detection of Insect-Plant Interactions

I used two different ITS primers in my analysis, but only retained plant taxa after filtering using one of them. Primers ITS-S2F/ITS4 are known to cross amplify fungi and plant taxa making them a useful general marker (Chen et al., 2010; Cheng et al., 2016). This may be appropriate when plants are not a key target. For example, I used them with spongy moths where the host trees were known. In this case I did amplify a variety of plants and fungi, however, only once was the host tree identified. This was somewhat surprising and may be linked to strict filtering of my data, and the age and preservation of the specimens (e.g., extracted >2 years after collection and not frozen immediately). In contrast to ITS, rbcL does not readily cross amplify taxa outside of the plants. Among spongy moths, rbcL identified 15 orders of plants compared to 13 in ITS, but only 8 orders were common between the two. In contrast to ITS, rbcL did successfully detect the host tree in 11 of the samples.

The Detection of Insect-Fungal Interactions

ITS3/ITS4 primers are less likely to amplify plant taxa and were used to amplify fungal DNA from emerald ash borers where the tree host (ash) was known and thus devoting sequencing power to plant identification was not useful. Additionally, ITS3/ITS4 primers are known to have better specificity and amplify a greater variety of fungal taxa in comparison to other ITS primers (Yu et al., 2022). Most of the fungi detected within the

emerald ash borer samples are associated with soil (e.g., *Fusarium* sp.), decaying plant matter (e.g., *Yamadazyma* sp.), lichens (e.g., *Candelaria* sp.), or plant pathogens (e.g., *Diaporthe* sp.) and not likely the larvae, pupae, or adults themselves suggesting co-occurrence data rather than direct interactions. One of the most commonly identified fungi is *Alternaria*, a group of fungi which are plant pathogens, involved in decomposition, but also ubiquitous in the environment and found almost everywhere (and considered a general environmental contaminant of eDNA studies, *unpublished data*). However, one fungus of interest that was detected in two samples (one size 2 larva and one adult beetle) is *Beauveria* sp. (order Hypocreales). *Beauveria* is an entomopathogenic fungus that is known to successfully infect and neutralize emerald ash borers. Although *Beauveria* occurs naturally in the environment (e.g., in soil; Rehner et al., 2011), biocontrol products containing *Beauveria* (e.g., FraxiProtec) are also commercially available and have been shown to significantly reduce emerald ash borer populations (Srei et al., 2020). This detection here suggests some natural biocontrol may be ongoing and detectable by this analysis.

The Detection of Insect-Microbial Interactions

Similar to the fungi detected by ITS, most of the bacteria associated with the samples are ubiquitous throughout the environment, and can be associated with soil (e.g., *Massilia* sp. in spongy moth samples and *Terriglobus* sp. in emerald ash borer samples), plants (e.g., *Capsulimonas* sp.), lichen (e.g., *Lichenihabitans* sp.), or the rhizosphere (e.g., *Neorhizobium* sp. found in emerald ash borer samples). However, one notable bacterium found in all target taxa is *Wolbachia* sp. (order Rickettsiales). *Wolbachia* sp. are

intracellular parasites that can infect many insects, including Lepidoptera (Tagami & Miura, 2004), and are associated with reproductive abnormalities, such as the feminization of genetic males, parthenogenesis induction, and reproductive incompatibility (i.e., through cytoplasmic incompatibility between hosts infected by different strains or between uninfected/infected hosts; reviewed in Werren, 1997). *Wolbachia*'s ability to induce genetic incompatibility in populations coupled with the ability for hosts to transmit the infection vertically and horizontally means it may have potential biocontrol applications (Werren, 1997; Zabalou et al., 2004).

Challenges of Detecting Parasites from Hosts

I attempted to use both COI and 18S to detect parasitism within the symbiome. Parasite detection is extremely challenging for a number of reasons. First, differentiating true parasites from environmental DNA contamination of the body is nearly impossible. Second, without destruction of the sample, detecting internal parasites is rare and most signals will likely be from surface contamination. A challenge with COI is the swamping of secondary signals with host DNA which is both more abundant and co-amplified. An alternative target is 18S which can be used to amplify DNA from nematodes, intracellular parasites, etc. However, short 18S regions suitable for high-throughput sequencing on platforms such as Illumina have almost no reference libraries, limiting identifications to phyla, and long regions cannot be used. To address this gap, I sequenced long 18S amplicons on the MinIon (Oxford Nanopore) which provides much lower sequencing depth but can accommodate long read lengths at low cost even with a high negative detection rate.

I tested this approach on emerald ash borer specimens and interestingly the 18S locus detected the greatest diversity of interaction types (Chapter 3 Figure 2.2), including animals (nematodes and arthropods, which were expected), but also some fungi, and a parasitic alveolate. All fungal orders detected using 18S were also detected using ITS, and despite a longer amplicon length the taxonomic resolution was the same or worse likely due to poor reference libraries (i.e., taxa identified in both loci such as *Diaporthe* sp. and *Niesslia* sp. were only resolved at a broader level in 18S). This is consistent with other studies that have found the ITS locus to provide a more extensive view of fungal diversity and allow more precise identifications to be made compared to the 18S locus (J. Liu et al., 2015; Lord et al., 2002). One notable detection was Cordycipitaceae (order Hypocreales), which contains the previously noted fungal biocontrol *Beauveria* independently suggesting its presence in the same sample and providing support to its presence when I treat it with caution in ITS data. While the recovery was much lower in taxonomic richness, the lower specificity of the 18S locus might be a useful screening tool for other interactions.

The non-host arthropods detected belong to order Hymenoptera. Interactions with nematodes (order Rhabditida) detected include *Rhabditolaimus* sp. (bacteriophagous and commensal with other beetles, e.g., *Scolytus multistriatus*; Ryss & Polyanina, 2022), *Panagrellus* sp. (free-living nematodes associated with many habitats, including insect frass; Ferris, 2009; Srinivasan et al., 2013) and Neotylenchidae (members can be parasitic, e.g., *Deladenus proximus*; Zieman et al., 2015). Finally, the parasitic alveolate

detected (*Cryptosporidium* sp.) has been reported to infect humans and animals including insects (Helmy & Hafez, 2022).

Novel Challenges in Detecting Species Interactions

Environmental DNA (eDNA) refers to DNA shed by organisms (e.g., cells, excreta, etc.) that is deposited throughout the environment, including in soil (Foucher et al., 2020; Marquina et al., 2019), water (Uchida et al., 2020), and air (Clare et al., 2021, 2022). eDNA can also be spread between organisms through contact with other organisms. For example, insects leave eDNA behind on the surface of plants they inhabit/consume (Allen et al., 2023; Kudoh et al., 2020) and this can be picked up by another insect without any actual interaction. This was demonstrated by Huszarik et al. (2023), who found insect eDNA was spread to spiders who occupied the same wet pitfall trap as an exotic insect that was not consumed by the spiders. It is thus inevitable for target organisms to become coated in eDNA simply by occupying their natural environment or during collection/storage depending on the medium (e.g., pitfall traps, malaise trap, ethanol, etc.; Huszarik et al., 2023; Shokralla et al., 2010). While eDNA's ubiquitous nature allows for its use in biodiversity and community composition assessments (Clare et al., 2022; Macher et al., 2023), it can complicate symbiome metabarcoding. More specifically, eDNA contamination on the surface of insects or other target organisms poses a serious challenge to symbiome metabarcoding as surface contamination can make it difficult to determine true interactions from co-occurrence.

Some of the identifications made here were unexpected. For example, when attempting to confirm target hosts, I often detected other insects which are not likely directly associated with the target. In spongy moth samples I detected 6 insect orders, including other plant pests (e.g., *Epuraea* sp. and *Pineus* sp.) and predatory insects (e.g., *Hemerobius* sp. and *Lestodiplosis* sp.). Spongy moths may encounter other plant pests (either directly or indirectly) on the host trees and may have even been predated on by some of the insects detected like lacewings (*Hemerobius* sp., order Neuroptera). For example, green lacewings are generalist predators that have been observed to feed on Lepidopteran eggs and small larvae (Brown & Cameron, 1982; Tauber et al., 2009). However, it is impossible to confirm whether these are predatory interactions or simply other insects in the environment. Interestingly, in one spongy moth sample from a pine tree a pine-feeding aphid (*Pineus* sp.) and its predator (*Lestodiplosis* sp.) were also detected (Gagn & Havill, 2020). While it is unlikely this was a direct interaction with the spongy moth, it suggests the detection of another local interaction occurring on the host tree, a fascinating ecological effect recorded in some surface contamination by environmental DNA of nearby organisms, though unrelated to the objective of the analysis. Despite the potential accumulation of DNA signatures with time from environmental sources, the actual profile of co-occurring DNA signatures was remarkably stable with life stage (e.g., see Figure 1.3 for spongy moths).

The effect was even larger in plant identifications. I commonly identified DNA from trees and plants that were not noted during spongy moth collection in both ITS (e.g., *Salix* sp., *Poa* sp., etc.) and rbcL (e.g., *Rosa* sp., *Picea* sp., etc.). These are possibly a result of

pollen from nearby plants sticking to the fuzzy surface of egg masses or to the hairs/surface of the larvae and, in the case of larvae specifically, I may also be detecting trees that previously hosted the larvae before collection.

Most of the fungi detected are ubiquitous throughout the environment and are often found in soil (e.g., *Sporobolomyces* sp.) or plants (e.g., plant pathogens like *Alternaria* sp.). However, other fungi (i.e., *Punctelia* sp.) and algae (i.e., *Coccomyxa* sp., *Trebouxia* sp., and *Symbiochloris* sp.) that often form lichen complexes (Sanders & Masumoto, 2021) were also identified as part of the spongy moth extracts by ITS, rbcL, or both. Thus, many identifications made from whole bodies are not interactions or part of the target symbiome but reflect surface level contamination from co-occurring species through environmental DNA more characteristic of the holobiont.

There are several potential solutions to this problem of surface contamination. First, it may be prudent to wash specimens prior to DNA extraction (i.e., in a bleach solution) to remove surface DNA (Binetruy et al., 2019; Hausmann et al., 2021; Huszarik et al., 2023), particularly if they have been stored in a common location (e.g., a malaise trap collection pot). This would remove surface DNA but may limit the detection of fungi and microbes associated with that ecology. Similarly, the gut of an insect might be dissected out to determine ingested plants and gut parasitism which are more direct interactions of the symbiome than the causal contact detected by surface environmental DNA of holobionts. This would also likely decrease the false negative rate in the detection of gut parasites like nematodes by better exposing them to DNA extraction. This, however, would cause

substantial or complete destruction of the specimen. In marine systems, where diet analysis must consider actual ingested material vs. DNA in swallowed water, one suggested solution has been to analyze DNA in water samples along with DNA from stomach contents to evaluate the potential errors (W. Symondson, perss comm). Some version of this, considering body surface and gut separately or host tree tissue separately from individual pest, might similarly help establish likely error rates, though this would likely increase the false negative rate. If individual A interacts with parasite B, both may leave DNA in the environment thus removing signals also in the environment may be overly conservative. Similarly, restricting the detections only to those for which an ecology has already been established limits the discovery of new true interactions. It may instead be argued that the strict definition of symbiome vs. holobiont, interactions vs. co-occurrence, is too strict a categorization of relationships which likely fall along a continuum of effects. Strongly interacting species, those with clear coevolutionary “Red Queen” relationships fall at one end and neutral co-occurring species at the other, but between them are a range of weak interactions, secondary interactions, and tertiary interactions which form the more complex roles within ecosystems.

Conclusion

In this analysis I compared the recovery of a variety of taxonomic target loci for the description of individual ecological interactions for common insects and evaluate the performance of different target loci for the identification of ecological interactions in these taxonomic groups. My data suggest remarkable similarity in identified taxa between life stages (e.g., eggs, larvae, pupae, and adults) for spongy moths but show some difference

in the emerald ash borers. My data also highlight the extreme level of potential environmental contamination of specimens in ecological analyses generating false positive “interactions” from contact environmental DNA which require creative solutions to estimate and mitigate the challenge depending on the analysis of interest.

Table 1. Target loci for interaction analysis of *Lymantria dispar dispar* (spongy moth, LDD) and *Agrilus planipennis* (emerald ash borer, EAB). All primers were modified with CS1/CS2 tails for sequencing using the Illumina MiSeq platform with the exception of SSU_F_07/SSU_R_26 which were modified following Hebert et al. (2023) for nanopore sequencing.

Gene	Primer	Reference	Direction	Primer Sequence (5′ to 3′)	Amplicon Length (bp)	Insect Test Samples
COI	ZBJ-ArtF1c	Zeale et al., 2011	Forward	AGATATTGGAACWTTATATTTTATTTTTGG	157	LDD: 9 egg masses, 9 larvae pools EAB: 57 adults, 53 pupae, 54 size 1 larvae, 68 size 2 larvae, 41 size 3 larvae, 1 unknown specimen
	ZBJ-ArtR2c	Zeale et al., 2011	Reverse	WACTAATCAATTWCCAAATCCTCC		
rbcL	rbcl_1-8	Palmieri et al., 2009	Forward	TTGGCAGCATTYCGAGTAACTCC	~226	LDD: 9 egg masses, 9 larvae pools
	rbcLB	Palmieri et al., 2009	Reverse	AACCYTCTTCAAAAAGGTC		
ITS	ITS-S2F	Chen et al., 2010	Forward	ATGCGATACTTGGTGTGAAT	~300-400	LDD: 9 egg masses, 9 larvae pools
	ITS4	White et al., 1990	Reverse	TCCTCCGCTTATTGATATGC		
ITS	ITS3	White et al., 1990	Forward	GCATCGATGAAGAACGCAGC	~300-400	EAB: 60 adults, 53 pupae, 54 size 1 larvae, 68 size 2 larvae, 41 size 3 larvae, 1 unknown specimen
	ITS4	White et al., 1990	Reverse	TCCTCCGCTTATTGATATGC		
16S	799F	Hanshew et al., 2013	Forward	CMGGATTAGATACCCKGG	~300-400	LDD: 9 egg masses, 9 larvae pools EAB: 57 adults, 53 pupae, 54 size 1 larvae, 68 size 2 larvae, 41 size 3 larvae, 1 unknown specimen
	1115R	Reysenbach & Pace, 1995	Reverse	AGGGTTGCGCTCGTTG		
18S	SSU_F_07	Carta & Li, 2018; Floyd et al., 2002	Forward	AAAGATTAAGCCATGCATG	~1000	EAB: 57 adults, 53 pupae, 54 size 1 larvae, 68 size 2 larvae, 41 size 3 larvae, 1 unknown specimen
	SSU_R_26	Carta & Li, 2018; Floyd et al., 2002	Reverse	CATTCTTGGCAAATGCTTTCG		

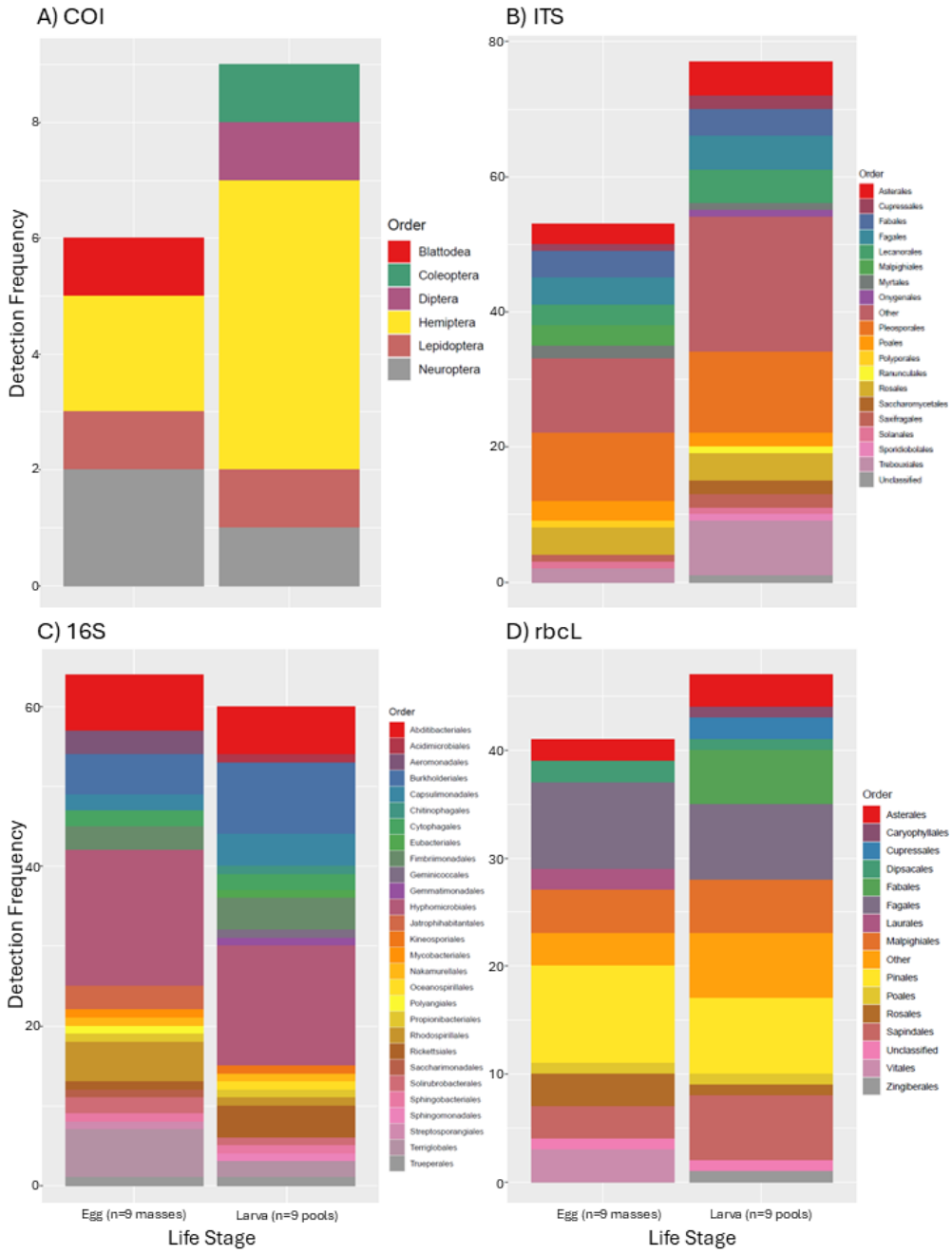

Figure 1.1. Co-amplified taxa from host spongy moths. A) in addition to the amplification of host DNA COI primers co-amplified DNA from 6 orders of insects as minor background signals. B) ITS primers recovered 19 orders of plant and fungi in the DNA extracts of moths with more taxa detected in larval extracts than egg masses. C) 16S primers targeting the microbiome detected a remarkably similar microbial community in egg vs. larval stages suggesting a highly similar core microbiome across life stage. D) rbcL primers targeting plant DNA detected a similar, but not completely overlapping plant community to ITS primers (B) with the same increased taxonomic richness in larval samples compared to egg masses.

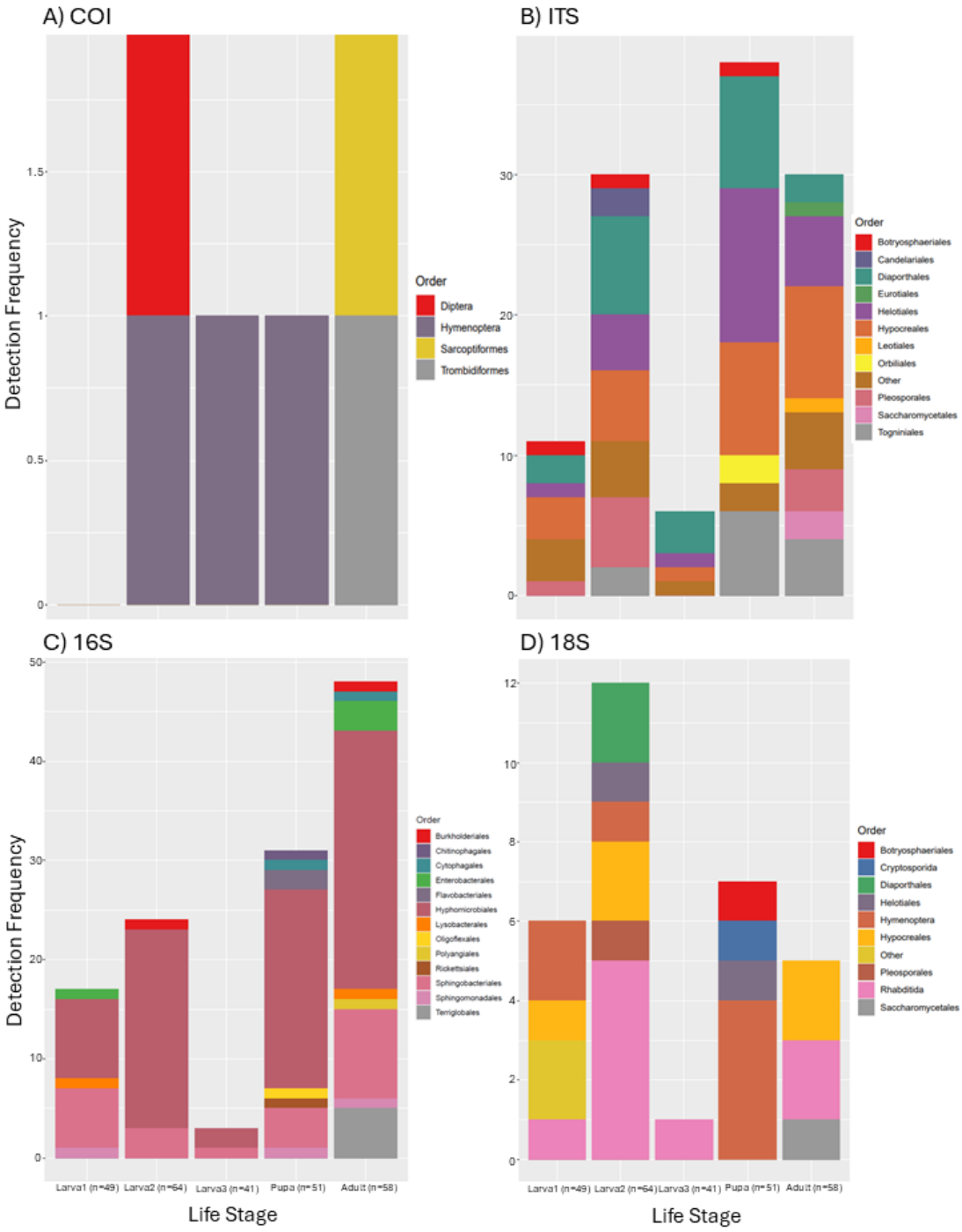

Figure 1.2. Co-amplified taxa from host emerald ash borer across life stages. A) in addition to the amplification of host DNA COI primers co-amplified 4 orders of arthropods as minor background signals. B) ITS primers targeting primarily fungal DNA recovered 11 orders in the DNA extracts of larva, pupae, and adults. C) 16S primers targeting the microbiome found an increasingly complex microbial community with life stage. D) 18S primers amplified the widest range of taxa including a series of nematodes which are likely parasites. Low taxonomic recovery from larva 3 specimens likely reflects the small sample size.

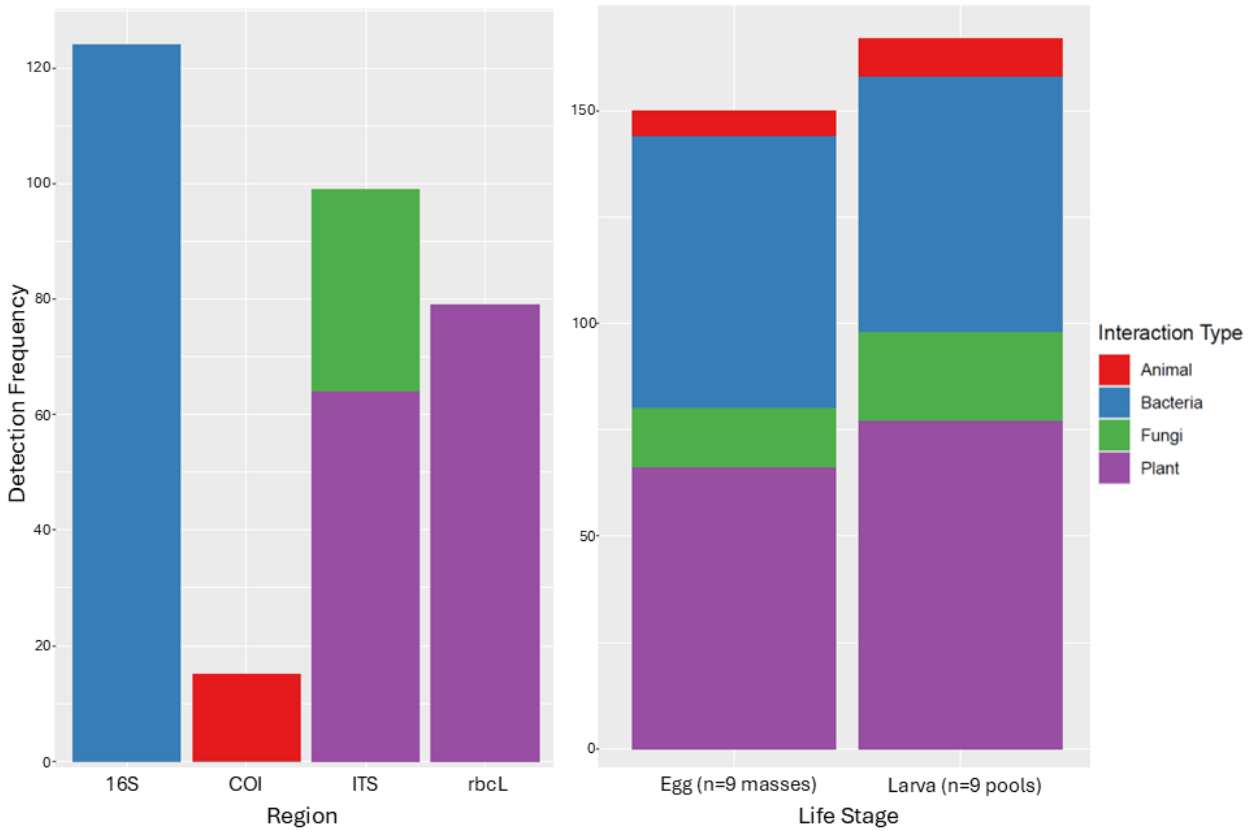

Figure 1.3. Highly similar interactions in eggs and larval spongy moths. The recovery of various targets by target amplified locus (left) shows high variability in taxonomic richness. Interestingly the profile of taxonomic recovery of co-amplified taxa is highly similar between egg masses and larval stage spongy moths (right) suggesting little change in the richness of interacting taxa over development.

Chapter 3: Emerald Ash Borer Symbiomes

Abstract

Emerald ash borers (*Agrilus planipennis*) are destructive invasive insects that increase tree mortality across North America, resulting in major ecological and economic impacts. Trees that are infested with emerald ash borers experience mortality rates (including those of mature, healthy trees) which can exceed 99%. Tree mortality disrupts ecological communities and threatens the $218 billion forestry industry in North America. Given the ecological and economic impact of these pests, it is important to identify species interactions to gain a better understanding of these impacts, potential biological control methods, and of what influences their population density. These species interactions can be characterized by DNA from the collection of organisms that interact with the individuals of the organism of interest that are co-localized and/or co-evolving and transfer DNA from one species to another, which allows metabarcoding technology to be employed to identify interactions. In Chapter 2 I establish methods for the analysis of insect ecological interactions. Here I use the data from Chapter 2 for a significant forest pest and assess the relative importance of life stages (e.g., larvae, pupae, and adults), collection location, habitat, and date on the detection of interactions within the host symbiome. In chapter 2 I generated PCRs from emerald ash borers (larvae, pupae, and adults) using multiple primers to amplify different taxonomic targets (animals, plants, microbes, fungi, and parasites) whose DNA is found in extracts from individual eggs, larva, pupae, and adults. All PCRs were independently barcoded and sequenced before being processed using the DADA2 pipeline in R Studio (for ITS and rbcL), QIIME2 (for 16S), mBRAVE (for COI), or Dorado/R (for 18S). I detected the presence of taxa representing 30 orders, including at

least five potential biocontrol agents. In this chapter, I use a random forest model to assess ecological correlates of these interactions. This model suggests the detection of symbiont and holobiont interactions can be predicted by collection date and life stage. My data suggest that interactions are more likely to be detected in pupal samples and may thus be the ideal target for future analysis.

Introduction

Emerald Ash Borers in North America

The invasive emerald ash borer (*Agrilus planipennis* Fairmaire, 1888) is a wood-boring beetle native to northeastern Asia and was likely introduced to North America from the Tianjin City/Hebei region of China (Bray et al., 2011). It was first detected in North America near Detroit, Michigan in summer 2002, although it is believed that the first individuals were introduced to the area in the 1990s and remained undetected (Cappaert et al., 2005; Siegert et al., 2014). Since its detection in 2002, the beetle has spread to five provinces in southeastern Canada (Ontario, Quebec, Manitoba, Nova Scotia, and New Brunswick) and 36 states in the United States of America, the majority of which are in northeastern regions adjacent to Canada (Natural Resources Canada, 2013; United States Department of Agriculture, 2023). The beetles' spread via stratified diffusion throughout North America has been facilitated by the flight of adults (especially mated females who have sometimes been observed to fly over 20 km a day; Taylor et al., 2010), the human-assisted movement of infested wood products (e.g., firewood, nursery stock, etc.), and the widespread presence of ash species in both natural and urban areas (BenDor et al., 2006; MacFarlane & Meyer, 2005; Muirhead et al., 2006).

Emerald ash borers have four main stages of their life cycle, beginning with an approximately 15 day egg stage, followed by a larval stage with four instars lasting roughly 300 days, pupation for roughly 3 weeks, and an adult beetle stage lasting 3 weeks during which females will lay 60-90 eggs (Government of Ontario, 2014; X.Y. Wang et al., 2010). Although their life cycle is generally completed in one year, it can sometimes be extended to two years depending on climatic conditions and host tree health, where colder climates and healthier trees are associated with delayed beetle development (Cappaert et al., 2005; Orlova-Bienkowskaja & Bieńkowski, 2016; Tluczek et al., 2011). The pests target ash trees (*Fraxinus* spp.) in North America, including black (*Fraxinus nigra*), green (*Fraxinus pennsylvanica*), white (*Fraxinus americana*), and blue ash (*Fraxinus quadrangulata*), but the beetles tend to be found on black, green, and white ash more commonly than blue ash (Anulewicz et al., 2007; Pureswaran & Poland, 2009). Additionally, green ash is preferred over white ash (Anulewicz et al., 2007).

Impacts of Emerald Ash Borers

Emerald ash borer infestations are associated with severe ecological and economic impacts. These forest pests primarily target stressed ash trees (e.g., trees that are girdled, wounded, or growing in unfavourable conditions), although healthy trees are susceptible to beetle colonization as well (Cappaert et al., 2005; McCullough, Poland, Anulewicz, et al., 2009; McCullough, Poland, & Cappaert, 2009; X.Y. Wang et al., 2010; Wei et al., 2004). The larvae consume the phloem and cambium of the ash trees they inhabit, leaving behind characteristic S-shaped galleries in their wake (Figure 2.1) that disrupt the

transportation of nutrients throughout the tree. Many of these galleries also score the outer sapwood (xylem), further disrupting the flow of water and nutrients in affected trees (Cappaert et al., 2005; X.Y. Wang et al., 2010). Infestations are also associated with extensive ash tree canopy loss, with green ash potentially more vulnerable to attack (Anulewicz et al., 2007). As a result, infestations are associated with increased ash tree mortality, with mortality rates (including those of mature, healthy trees) often exceeding 99% within the first decade of emerald ash borers arriving in an area (Klooster et al., 2014; Knight et al., 2013; Natural Resources Canada, 2013). This has had dramatic impacts and has led to six North American ash species now being included in the International Union for Conservation of Nature's (IUCN) Red List of Threatened Species, including common ash (*Fraxinus excelsior*), carolina ash (*Fraxinus caroliniana*), narrow-leafed ash (*Fraxinus angustifolia*), white ash (*F. americana*), green ash (*F. pennsylvanica*), and blue ash (*F. quadrangulata*), with the latter three designated critically endangered as of 2017 (International Union for Conservation of Nature, 2022).

Increased ash tree mortality reduces the availability of food and shelter for organisms that rely on those trees. For example, there are 43 monophagous (i.e., feed exclusively on ash trees) native arthropod species (e.g., eastern ash bark beetle [*Hylesinus aculeatus*], black-headed ash sawfly [*Tethida barda*], etc.) that are at a high-risk for becoming endangered/extinct as a result of declining ash populations (Gandhi & Herms, 2010b). In addition, a variety of mammals and birds (e.g., beavers, rabbits, deer, purple finches, and wood ducks) rely on ash trees for food (i.e., ash seeds and the bark of young ash trees) and shelter (Schlesinger, 1990). Beetle infestations also alter the structure of ecological

communities. For example, ash tree canopy dieback creates gaps within tree canopies that allow higher amounts of light to reach subcanopy levels, encouraging the growth of understory species (e.g., *Acer* spp.) over infested ash trees, thus altering the distribution and abundance of species (Canham, 1988; Flower et al., 2013; Gandhi & Herms, 2010a).

Ash trees are valued for their use in a variety of wood products, including timber, paper, and furniture. However, beetle infestations reduce the production of timber as a result of ash tree damage and mortality (Schrader et al., 2021), thus threatening the $218 billion gross domestic product forestry industry in Canada and the United States of America (Aukema et al., 2011; Natural Resources Canada, 2022; United States Department of Agriculture, 2021). Since their introduction in the 1990s, emerald ash borers have killed tens of millions of trees (Cappaert et al., 2005) and are estimated to be the costliest forest pest in North America, responsible for approximately $850 million annually in local government expenditures (i.e., tree removal, replacement, and treatment) and approximately $380 million annually in lost residential property values in the USA alone (Aukema et al., 2011). It is also believed that beetle infestations are linked to increased cardiovascular and lower-respiratory-tract illness mortality as a result of lost ecosystem services (particularly regulating services such as air quality and climate regulation which are reduced as a result of increased tree mortality; Donovan et al., 2013). Other impacted ecosystem services include nutrient cycling (i.e., due to increased ash leaf/log litter) and carbon cycling (i.e., increased ash tree mortality decreases the carbon sink capacity of some forests; Flower et al., 2013; Schrader et al., 2021; Ulyshen et al., 2011).

Management of Emerald Ash Borer Populations

There have been multiple unsuccessful attempts to eradicate emerald ash borer populations in North America since their detection in 2002. This lack of success is partly attributed to infestations being difficult to detect until the population density increases and the host begins exhibiting external symptoms of an infestation (e.g., crown dieback, bark splits, etc.; Cappaert et al., 2005). The Canadian Food Inspection Agency (CFIA) and the United States Department of Agriculture – Animal and Plant Health Inspection Service (USDA-APHIS) are the primary authorities charged with limiting the spread of emerald ash borers in North America (Hope et al., 2021). Soon after the detection of the first infestations in North America, both agencies implemented quarantine areas (where populations were established) and regulated the transportation of wood products from infested quarantine areas (both in North America and beyond) to limit the spread of the pest. For example, all off-continent wood products imported were required to be heat-treated or fumigated, and the movement of wood products (e.g., firewood) from quarantine areas was restricted (Government of Canada, 2013a; Hope et al., 2021; Removal of Emerald Ash Borer Domestic Quarantine Regulations, 2020). While these regulations are still in effect in Canada, the USDA-APHIS discontinued domestic quarantine regulations in 2021 citing their lack of effectiveness and the need to better allocate resources toward the containment and management of emerald ash borers using non-regulatory methods (discussed below; Hope et al., 2021; Removal of Emerald Ash Borer Domestic Quarantine Regulations, 2020).

Other suppression efforts make use of a combination of physical, chemical, and biological controls, however, each of these methods of control are not very effective or come with consequences. Physical controls that have been employed include creating ash-free firebreak zones and the removal of infested trees. However, adult beetles can fly far enough to cross firebreak zones or can be established outside the zone prior to creation without being noticed (Herms & McCullough, 2014; Taylor et al., 2010) which makes firebreak zones ineffective. Additionally, the removal of infested trees is dependent on the infestations having visible external effects on the host tree.

Chemical controls include the application of insecticides (e.g., imidacloprid, dinotefuran, azadirachtin, and emamectin benzoate; Herms & McCullough, 2014; McCullough, 2020) that may have unintended consequences in the environment which can limit their usefulness and areas they can be applied. For example, neonicotinoid pesticides (i.e., imidacloprid and dinotefuran) can be harmful to non-target organisms such as pollinators; imidacloprid is attributed to higher gut pathogen levels in honeybees (Pettis et al., 2012) and reduces pollen foraging efficiency, population growth, and queen production in bumblebees (Gill et al., 2012; Whitehorn et al., 2012). In other groups such as earthworms (*Eisenia fetida*) dinotefuran reduces growth and reproduction (T. Liu et al., 2018), and in aquatic invertebrates (*Pteronarcys dorsata* and *Tipula* sp.,) imidacloprid reduced leaf-litter feeding and, in some cases, caused fatalities (Kreutzweiser et al., 2007). These effects, particularly those on pollinators, resulted in the use of neonicotinoid pesticides being partially banned in the European Union in May 2013 (Wood & Goulson, 2017). Similarly, azadirachtin can harm a variety of non-target organisms (reviewed in Mužinić

& Želježić, 2018 and Kilani-Morakchi et al., 2021), including *Bombus terrestris* where mortality and the prevalence of deformities is increased and reproduction is reduced (Barbosa et al., 2015). Emamectin benzoate has also been found to negatively impact the development and fecundity of impact non-target organisms such as *Paederus fuscipes* (rove beetle; Khan et al., 2018), and has once again been found to be toxic to honeybees *Apis mellifera* (Abdu-Allah & Pittendrigh, 2018).

Biological controls include the release of natural predators and parasitoids of emerald ash borers (i.e., parasitoid wasps *Tetrastichus planipennisi*, *Spathius agrili*, *Spathius galinae*, and *Oobius agrili*) that were imported from China and Russia (Duan et al., 2018; Government of Canada, 2013b). Species native to North America, such as woodpeckers (Picidae), parasitoid wasps (e.g., *Atanycolus cappaerti*), and the fungus *Beauveria bassiana* (now commercially available as FraxiProtec), have also been linked to decreased infestation levels (Cappaert et al., 2005; Cappaert & McCullough, 2009; Lindell et al., 2008; Srei et al., 2020). However, the effectiveness of these biological controls vary, and it is unclear whether potential agents are functional in the wild. Woodpecker predation, for instance, is dependent on pest population density and the condition of the ash trees (Jennings et al., 2013; Lindell et al., 2008). Similarly, the effectiveness and horizontal transmission of *Beauveria* may depend on the density of infested trees and pest populations (Srei et al., 2020). Additionally, one of the introduced parasitoids (*S. agrili*) was found to be unable to establish self-sustaining populations in regions north of 40°N latitude and has yet to establish well in any release sites south of that latitude. However, the other three (*T. planipennisi*, *S. galinae*, and *O. agrili*) were successful in

doing so beyond their initial release sites (Duan et al., 2023; Government of Canada, 2013b) and tracking their natural dispersal is key to understanding their effectiveness.

Ecological Interactions in Biological Controls

Trophic interactions play an important role in regulating emerald ash borer populations (Cappaert et al., 2005; Cappaert & McCullough, 2009; Lindell et al., 2008; Srei et al., 2020). These trophic interactions can be characterized using the “symbiome” or ‘holobiont” approach (Chapter 2) which characterizes interactions at the level of individuals using metabarcoding techniques. The approach refers to the collection of organisms (i.e., bacteria, fungi, parasites, etc.) that interact with an individual of interest (e.g., by residing in or on a host organism) and may be colocalized (co-occurring) and co-evolving. A “symbiome” describes co-evolving taxa and has been subtly differentiated from “holobiont” which includes physically associated taxa but does not necessarily involve co-evolution (Tripp et al., 2017). Both symbiont and holobiont interactions leave behind a genetic signature that can be detected forensically, allowing this metabarcoding approach to be used.

In Chapter 2 I tested multiple target gene loci for the recovery of these interactions and generated a data set as part of methods development. The objective of this study is to use this data to classify the ecological factors that influence these ecological interactions. In particular, I assess the impact of 1) life stages (e.g., larvae, pupae, and adults), and 2) ecological correlates (e.g., collection date, location, and habitat) on the detection of ecological interactions of emerald ash borers. I also look specifically for potential agents

of biological control (i.e., viruses, fungi, parasites, etc.) or particularly vulnerable pest life stages to highlight any approaches to suppress or better target the invasive species with minimum impact on ecosystems.

Methods

The data used in this chapter are described in Chapter 2. In brief, emerald ash borer larvae (n= 54 size 1, n=68 size 2, n=41 size 3) pupae (n=53), and adults (n=60) were hand collected from ash trees found in Brockville (n = 185), Ennotville (n = 41), Guelph (n = 41), and Puslinch (n = 10) in Ontario Canada by the collection unit of the Canadian Center for Biodiversity Genomics (Appendix A Tables S1 and S2). I ground specimens using a pestle and extracted DNA using the Qiagen blood and tissue kit following the manufacturer's guidelines with a decreased final DNA elution volume of 150ul of Buffer AE to increase DNA yields. I extracted DNA from the posterior half of the abdomen for adult specimens, the posterior third for pupa specimens, a 6mm fragment of the posterior end of size 2 (larva2) and 3 larva (larva3) specimens, and the whole size 1 larva (larva1).

To confirm larva ID and identify insect parasites I amplified a partial COI barcode using the ZBJ-ArtF1c/ZBJ-ArtR2c primers (Zeale et al., 2011). Because host plant was known (ash trees) I used ITS3/ITS4 primers (White et al., 1990) which have a better recovery for fungi to amplify symbionts. I used 799F/1115R primers (Hanshew et al., 2013; Reysenbach & Pace, 1995) to amplify microbial taxa. All PCR products were sequenced on the Illumina MiSeq v3 using a 600-cycle run at the Barts and the London Genome Center at Queen Mary University of London, Blizzard Institute. To target parasites DNA

extracts were sent to the University of Guelph where a long read 18S locus was amplified using SSU_F_07/SSU_R_26 primers (Carta & Li, 2018; Floyd et al., 2002) and sequenced using the MinIon platform. All PCR primers, reaction mix, thermocycling conditions, and positive and negative controls are given in Chapter 2. Bioinformatics processing was performed as described in Chapter 2 using DADA 2 (ITS), mBRAVE (COI), QIIME2 (16S) and a purpose built platform for MinIOn Data (18S; Hebert et al., 2023).

To assess the relative influence of age, location (i.e., region and sector), habitat, sub-habitat, and collection time (day of the year) I constructed a random forest classification model (Breiman, 2001) where the presence or absence of order-level interactions was classed as a response variable, and specimen age, collection region, collection sector, collection habitat, collection sub-habitat, and collection date were treated as predictors (see Appendix A Tables S1 to S3 and Appendix E). I then considered fungi, microbes, and parasites as separate taxonomic groups and repeated the analysis using these taxa as separate response variables. The random forest models were generated using the *caret* package (Kuhn, 2008) in R Studio (R Core Team, 2023). A random forest model was selected over other models due to its robust nature (e.g., ability to handle both continuous and categorical variables and outliers), high accuracy, and ability to evaluate variable importance (i.e., which variables most influence the accuracy of the random forest model predictions).

To ensure reproducibility, the seed was initialized to 42 using the set.seed function prior to generating the models. Data were then randomly split into “training” (70%) and “testing” (30%) sets, with the former used to train the random forest model. Repeated k fold cross validation (k = 10, repeats = 5) was included in each model to estimate the performance (accuracy) of the models and to improve model accuracy by allowing more data to be used when training the model. The number of variables tried in each split (mtry) was determined using the tuneRF function in the *randomForest* package (Liaw & Wiener, 2002), where mtry = 2 was found to be optimal for the animal and overall models, while mtry = 1 and mtry = 4 were found to be optimal for the bacteria and fungi models, respectively. The testing set was used to evaluate model accuracy by generating a confusion matrix using the *caret* package (Kuhn, 2008). Variable importance (i.e., the mean decrease in the Gini coefficient) was also determined using the *caret* package (Kuhn, 2008). Accumulated local effects (ALE) plots were then generated from the training data using the *ilm* package (Molnar et al., 2018) to determine the impact each variable has on the presence or absence of interactions. A correlation matrix was also generated using the *stats* (R Core Team, 2023) and *ggcorrplot* packages (Kassambara, 2023) to determine the correlation between predictor variables.

Results

The species interaction data used here is from Chapter 2. In this chapter I used a random forest model to assess ecological correlates of these interactions. In summary, I used 65 identifications across the 4 gene loci analyzed (ITS = 23; 16S = 21; COI = 4; 18S = 17) which represented 61 unique taxa from 30 taxonomic orders. Using the COI target locus

4 arthropod orders were detected, including those of potential parasites. Due to strict filtering protocols (detailed in Chapter 2), no ASVs were retained from the larva1 samples. However, 2 orders were identified in the larva2 samples, 1 in the larva3 and pupae samples, and 2 in the adult samples. These identifications included a family of parasitoid wasps (Ichneumonidae, order Hymenoptera). I detected 13 orders of bacteria using the 16S locus, with larva1 having a richness of 5 orders, larva2 3 orders, larva3 2 orders, pupae 8 orders, and adults 9 orders. Notably, the insect pathogen *Wolbachia* sp. (order Rickettsiales) was detected in 1 sample. The ITS locus allowed the detection of 11 fungal orders, with 5 orders detected in larva1, 7 in larva2, 3 in larva3, 6 in pupae, and 8 in adults. Notably, *Beauveria* sp. (order Hypocreales), a fungal biocontrol agent, was detected in two samples. Although *Beauveria sp.* was also detected in one of the negative controls, the identification was ultimately kept due to it being present in very low amounts in this control compared to those detected in the emerald ash borer samples and there being no lab-based source of contamination. The detection of *Beauveria* sp. is also supported by the 18S data, where it was also detected at order-level in the same samples. I detected 9 orders using the 18S locus, with 6 of the 9 representing fungi (overlapping with those identified using ITS) and the remaining 3 representing nematodes, insects, and a parasitic alveolate protist (*Cryptosporidium* sp., order Cryptosporida). I identified 3 orders in the larva1 samples, 5 in larvae2, 1 in larvae3, 4 in pupae, and 2 in adults.

These identifications account for 259 order level interactions across emerald ash borer life stages (Figure 2.2). In total, I detected 29 order level interactions in the larva1 samples (3 animal, 17 bacteria, and 9 fungi), 64 in the larva2 samples (8 animal, 24 bacteria, and

32 fungi), 10 in the larva3 samples (2 animal, 3 bacteria, and 5 fungi), 74 in the pupal life stage (5 animal, 31 bacteria, and 38 fungi), and 82 in the adult life stage (4 animal, 48 bacteria, 29 fungi, and 1 parasitic alveolate). Overall, the adult life stage had the most interactions detected, and the majority of interactions detected occurred between emerald ash borers and bacteria (123 of 259 total interactions, or ~47%).

The classification accuracy of the total interactions random forest model was 68.92%, with an out-of-bag (OOB) error estimate of 31.18%. The model was found to correctly classify the testing data 69.23% of the time. The three most important features (Figure 2.3), determined by the mean decrease in the Gini coefficient (MDG), were larva3 life stage (MDG = 4.94), collection date (represented as day of the year; MDG = 3.72), and pupa life stage (MDG = 1.76). The ALE plots indicated that interactions were more likely to be present in the pupal life stage (Figure 2.4), in samples collected prior to July 14 (195^{th} day of the year), and in samples collected from the Ennotville region, Rosedale sector, woodland habitat, and forest sub-habitat (Appendix F Figures S1 to S5).

For bacterial interactions, the estimated classification accuracy of the random forest model was 68.28%. The model had an OOB error estimate of 31.72% and it was found to correctly classify the testing data 69.23% of the time. The three most important features (Figure 2.3) were collection date (MDG = 1.64), larva3 life stage (MDG = 0.17), and Tincap sector (MDG = 0.54). The ALE plots indicated that interactions with bacteria were more likely to be present in the pupal life stage (Figure 2.4), in samples collected prior to

June 22 (173rd day of the year), and in samples collected from the Guelph region, Sones sector, woodland habitat, and forest sub-habitat (Appendix F Figures S1 to S5).

The fungal interactions random forest model had an estimated classification accuracy of 68.81% and an OOB error estimate of 31.18%. The fungi random forest correctly classified the testing data 76.92% of the time, and the three most important variables (Figure 2.3) were collection date (MDG = 7.28), pupa life stage (MDG = 4.36), and larva3 life stage (MDG = 3.17). The ALE plots indicated that interactions with fungi were more likely to be present in the pupal life stage (Figure 2.4), in samples collected prior to July 6 (187th day of the year), and in samples collected from the Ennotville region, Rosedale sector, artificial habitat, and urban sub-habitat (Appendix F Figures S1 to S5).

The estimated classification accuracy of the animal interactions random forest model was 92.49%. The OOB error estimate for the model was 7.57%. It was found to correctly classify the testing data 92.41% of the time. The three most important variables (Figure 2.3) were collection date (MDG = 0.80), Rosedale sector (MDG = 0.60), and Ennotville region (MDG = 0.51). The ALE plots indicated that interactions with animals (i.e., insects, nematodes, etc.) were more likely to be present in the pupa life stage (Figure 2.4), in samples collected after July 21 (202nd day of the year), and in samples collected from the Ennotville region, Rosedale sector, artificial habitat, and arable sub-habitat (Appendix F Figures S1 to S5). The correlation between most of the predictor variables was low, although there was high correlation detected between some variables (e.g., the correlation between the Rosedale collection sector and Ennotville collection region was

high, which was unsurprising given Rosedale is a sublocation of Ennotville; Appendix F Figure S6).

Discussion

Ecological interactions play an important role in the evolution and survival of a species (Barraclough, 2015; Thompson, 1999). In the case of emerald ash borers, interactions with biological controls (e.g., woodpeckers, *Beauveria*, etc.) may result in methods of controlling their populations. These interactions can be characterized by both co-occurring (holobiont) and co-evolving (symbiont) organisms that interact with an individual of interest (Tripp et al., 2017). While species interactions are often looked at more broadly (i.e., on the scale of the entire species), symbiome and holobiont analysis allows the direct and indirect interactions of individuals to be determined and can be used to identify novel interactions.

I apply metabarcoding to assess the factors which influence the ecological interactions of commercially destructive emerald ash borers. My objectives were to characterize the interactions at different life stages and determine whether ecological correlates can predict the detection of non-target taxa within a sample. My data demonstrate that life stage targeted, collection date, and location (region and sector) can predict interactions using a random forest model, while habitat and sub-habitat play minor roles. I identified interactions with a variety of taxa (e.g., bacteria, mites, fungi, nematodes, etc.), including interesting potential agents of biocontrol (e.g., *Beauveria* sp., parasitoid wasps, etc.) in several samples of emerald ash borer.

Taxonomic Identifications

A total of 259 order level interactions were detected across emerald ash borer life stages (Chapter 2 Figure 1.2). These interactions occurred with 61 unique taxa from 30 taxonomic orders across the four markers used in this analysis (COI, ITS, 16S, and 18S). Using the COI locus, I detected 4 orders of arthropods, which included one family of gall midge (Cecidomyiidae, order Diptera), two families of mites (Oppiidae order Sarcoptiformes and Eupodidae order Trombidiformes), and notably, Ichneumonidae parasitoid wasps (order Hymenoptera, see below).

The 16S locus was used to detect bacterial interactions, however, the majority of the bacteria associated with my samples are ubiquitous throughout the environment and can be associated with soil (e.g., *Terriglobus* sp.) or the rhizosphere (e.g., *Neorhizobium* sp.). Many of these detections are thus likely co-occurring holobiont interactions rather than direct interactions (co-occurring and co-evolving) characteristic of symbionts. Notably, I detected the intracellular parasite Wolbachia sp. (order Rickettsiales) in one sample (see below). Similar to the bacteria detected by 16S, the majority of the fungi I detected using the ITS marker are associated with soil (e.g., *Fusarium* sp.), decaying plant matter (e.g., *Yamadazyma* sp.), lichens (e.g., *Candelaria* sp.), or plant pathogens (e.g., *Diaporthe* sp.) and not the emerald ash borers themselves. It is extremely likely that these represent fungi associated with the host tree or surrounding environment and would be classed as co-occurring (holobiont) rather than co-evolving or direct interactions which might

represent a symbiome. However, the fungus *Beauveria* sp. (order Hypocreales) was also detected and may represent a true symbiont interaction (see below).

The 18S locus detected the greatest diversity of interaction types, including animals (i.e., nematodes and arthropods), a parasitic alveolate, and some fungi. The fungal orders detected using the ITS locus overlapped with those identified using the 18S locus, however, the taxonomic resolution was the same or worse in the 18S batch (i.e., taxa identified in both loci such as *Diaporthe* sp. and *Niesslia* sp. were only resolved at a broader level in 18S). I also detected non-host arthropods belonging to the order Hymenoptera with the 18S marker. Detected interactions with nematodes (order Rhabditida) included Neotylenchidae (members can be parasitic, e.g., *Deladenus proximus*; Zieman et al., 2015; see below), *Panagrellus* sp. (free-living nematodes associated with many habitats, including insect frass; Ferris, 2009; Srinivasan et al., 2013), and *Rhabditolaimus* sp. (bacteriophagous and commensal with other beetles, e.g., *Scolytus multistriatus*; Ryss & Polyanina, 2022). The parasitic alveolate (*Cryptosporidium* sp.) that was also detected has been reported to infect humans and animals (including insects, Helmy & Hafez, 2022), thus this may be a true parasitic interaction (see below).

The Detection of Biological Controls and Parasites

Several potential agents of biocontrol were detected across all loci used. Ichneumonidae parasitoid wasps (order Hymenoptera) were detected in larva2, larva3, and pupae samples using the COI locus. Ichneumonids have previously been observed to parasitize emerald ash borers (Duan et al., 2009), so these are likely cases of true parasitism.

Additionally, the entomopathogenic fungus *Beauveria* sp. (order Hypocreales) that is known to successfully infect and neutralize emerald ash borers was identified in two samples (one size 2 larva and one adult beetle) using the ITS locus. Although naturally occurring in the environment (e.g., in soil; Rehner et al., 2011), biocontrol products containing *Beauveria* such as FraxiProtec are commercially available and are able to significantly reduce emerald ash borer populations (Srei et al., 2020). It is worth noting that *Beauveria* was also detected in the negative control but was ultimately included since the read count was much lower in the control and there is no external source of *Beauveria* contamination in my process. Furthermore, the detection was supported by the 18S data where I detected the family Cordycipitaceae (order Hypocreales) which contains the fungal biocontrol *Beauveria*, independently confirming its presence in the same sample.

Using the 16S locus, I detected the intracellular parasite *Wolbachia* sp. (order Rickettsiales) in one pupal sample. *Wolbachia* bacterium can infect many insects, and infections can be vertically and horizontally transmitted between hosts (Werren, 1997). Infection is associated with reproductive abnormalities, such as the feminization of genetic males, parthenogenesis induction, and reproductive incompatibility (i.e., through cytoplasmic incompatibility between hosts infected by different strains or between uninfected/infected hosts; reviewed in Werren, 1997). Therefore, the detection of *Wolbachia* is notable as it may have potential biocontrol applications (Werren, 1997; Zabalou et al., 2004).

The parasitic alveolate and nematodes detected using the 18S locus may potentially act as biocontrols as well. As mentioned, the alveolate (*Cryptosporidium* sp.) is known to infect humans and animals, including insects (Helmy & Hafez, 2022), and can therefore be considered a true parasitic interaction. Similarly, members of the Neotylenchidae family (order Rhabditida) of nematodes detected can be parasitic (e.g., *Deladenus proximus*; Zieman et al., 2015), and thus this may also represent a true parasitic interaction that may have potential biocontrol applications. The detection of all these interactions suggests that some natural biocontrol may be occurring in the collection location and detectable by metabarcoding.

Factors Influencing the Interactions Assessed Using Random Forest Models

Four random forest models were generated in total: one to predict the presence of any interaction at all, and the others to predict the presence of fungi, microbial, or animal interactions, specifically. The models were successfully able to predict the number of interactions detected using the testing dataset at least ~69% of the time, and up to ~92% of the time. Animal interactions were most likely to be classified correctly (~92% of the time), while bacteria, fungal, and total interactions were less likely to be classified correctly (~69%, ~77%, and ~69% of the time, respectively). In all models, collection date (represented as day of year) and life stage played the most important roles in correctly classifying the presence or absence of interactions, with the exception of the animal and bacteria models where location (sector and region) also played a major role.

The ALE plots generated for each interaction type were used to determine what effect, if any, each variable has on the presence or absence of interactions detected. I found that all types of interactions were more likely to be present in the pupal life stage. Collection date also seemed to play a significant role in the presence of interactions, with interactions more likely to be present in samples collected by the end of July. This differed slightly for specific interaction types (e.g., bacterial interactions were most likely to be present in samples collected by June 22, 2022), but the latest the date ranges to is July 21, 2022 (for animal interactions) before the detection rate decreases. The presence of total, fungi, and animal interactions were highest in the Ennotville collection region and Rosedale collection sector, while the presence of bacterial interactions was greatest in the Guelph region and Sones sector. Similarly, fungi and animal interactions were more likely to be present in artificial habitats, while total and bacterial interactions were more likely to be present in forest and woodland habitats. Finally, I found that bacteria and total interactions were more likely to be present in forest subhabitats, fungi interactions in urban subhabitats, and animal interactions in arable subhabitats.

Novel Challenges in Differentiating Symbiomes vs. Holobionts

One of the main challenges in determining a true interaction from co-occurrence data is the recent realization that environmental DNA (eDNA) likely contaminates all samples collected from the wild. eDNA includes all biological material (cells, excreta, etc.) shed by organisms continually and it accumulates in the environment and can be collected from air (Clare et al., 2021, 2022), water (Uchida et al., 2020), soil (Foucher et al., 2020; Marquina et al., 2019) and surfaces such as leaves (Macher et al., 2023; Valentin et al.,

2020) and the bodies of other animals through casual contact. eDNA can also accumulate and settle on biotic and abiotic surfaces. Consequently, eDNA of nearby organisms can be found on the surface of arthropods even if there was no direct interaction between the two. For example, Huszarik et al. (2023), demonstrated that eDNA of an exotic insect could be detected on spiders who occupied the same wet pitfall trap. We must thus expect that a sample or target organisms will become coated in eDNA simply from the environment or during collection and storage (e.g., pitfall traps, ethanol, etc.; Huszarik et al., 2023; Shokralla et al., 2010) and the duration and magnitude of this effect in my data is only starting to be realized. Although the ubiquitous nature of eDNA allows for its use in biodiversity and community composition assessments (Clare et al., 2022; Littlefair et al., 2023; Macher et al., 2023), it can complicate metabarcoding to determine species interactions, making it difficult to determine true interactions from co-occurrence.

The effect is suspected here. For example, most of the fungi detected are ubiquitous throughout the environment and are often found in soil (e.g., *Fusarium* sp.) or plants (e.g., *Alternaria* sp.). The fungi *Candelaria* sp. is known to form lichen complexes (Sanders & Masumoto, 2021) but was identified in the emerald ash borer samples despite likely not interacting directly with the specimens. Similarly, beetles and the family of gall midges (Cecidomyiidae, order Diptera) detected using the COI locus are unlikely to directly interact, but the DNA was co-occurring in the same samples. Thus, many identifications are not part of the symbiome, but reflect simple co-occurrence in the local environment.

While it is extremely difficult to envision an immediate solution to this problem, there are mitigation steps and data analysis steps which might reduce the challenge. First, there is some evidence that specimens can be washed (e.g., in a bleach solution) prior to DNA extraction to remove eDNA (Binetruy et al., 2019; Hausmann et al., 2021; Huszarik et al., 2023), though this would then increase the false negative rate by removing fungi and microbes associated with that surface ecology. Dissection so that only the gut undergoes DNA extraction would allow direct interactions related to feeding and gut parasitism to be determined and potentially increase the target DNA concentration relative to host DNA but would limit the scope of the interactions being assessed. However, this is very time intensive and requires destruction of the specimen. Just as important is ecological assessment of all recovered taxa to determine if they are “reasonable” given known ecology. While this will limit discovery of novel interactions, it should be encouraged so that all identifications are not automatically accepted as a true interaction.

Conclusion

In this analysis I describe the emerald ash borer symbiome at different life stages and evaluate whether the detection of ecological interactions can be predicted by collection date, location, life stage, or habitat using random forest models. My data suggests similarity in the types of interactions detected at different life stages (e.g., larvae, pupae, and adults). Additionally, my model found that interactions were more likely to be detected in the pupal specimens, indicating that the pupal life stage may be an ideal target for future analyses. I also found some support for sampling earlier in the year and regional variation in data though with smaller sample sizes this should be treated with caution.

This data also highlights the extreme level of potential environmental contamination of specimens in ecological analyses, and the difficulty in distinguishing true interactions. The risk of generating false positive “interactions” from contact environmental DNA requires creative solutions to estimate and mitigate the challenge in all measures of species interactions.

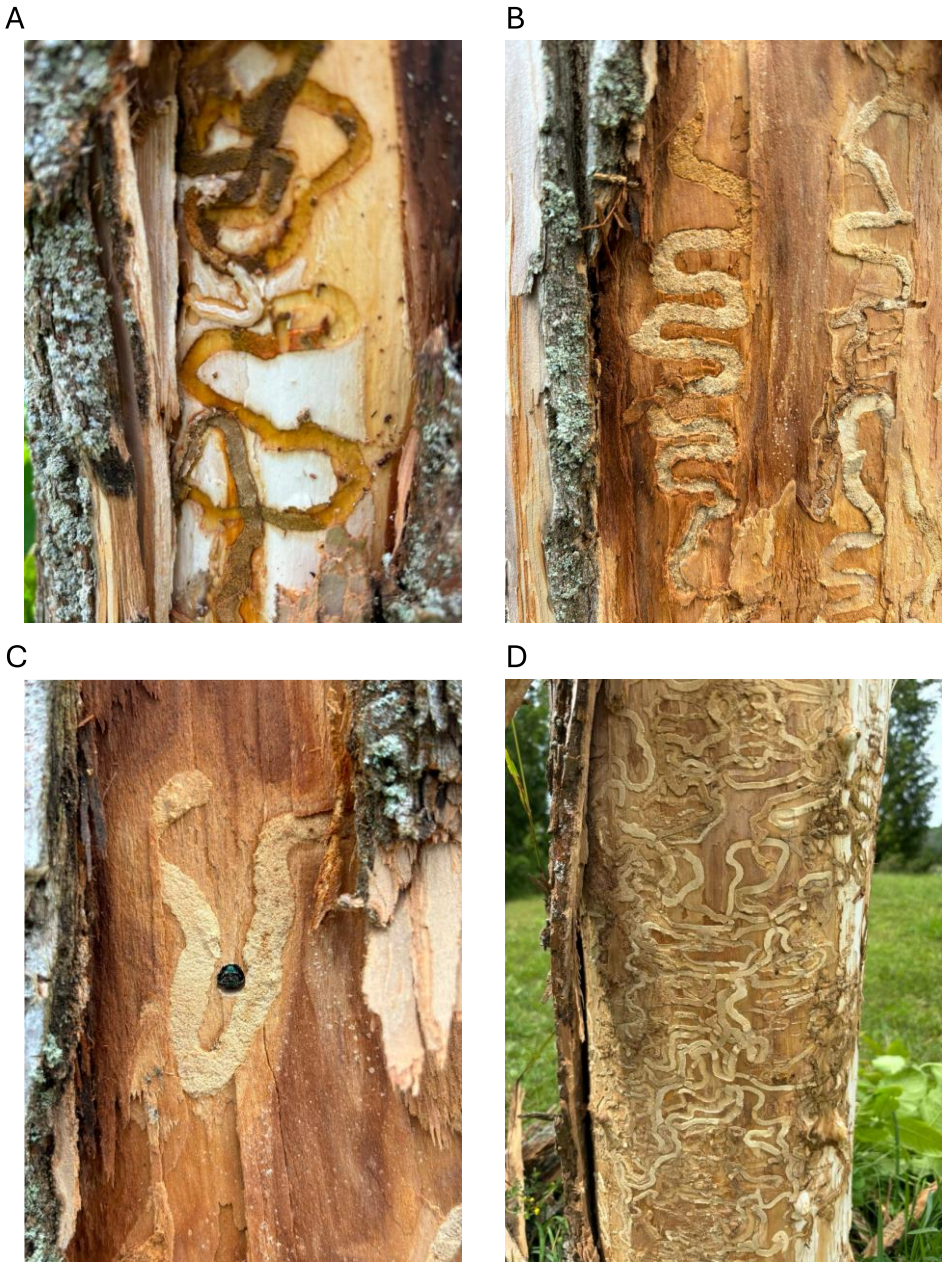

Figure 2.1. Extensive damage caused by emerald ash borers. A) a larva *in situ* cutting a trail just under the bark, B) characteristic “S” shape paths cutting off fluid and nutrient transport, C) an adult in a recessed tunnel extending through the bark, and D) extensive damage just under otherwise normal looking bark of an ash tree which had lost >90% of its canopy and was scheduled for removal. All images taken on the Bruce Peninsula, Ontario Canada July 20, 2024.

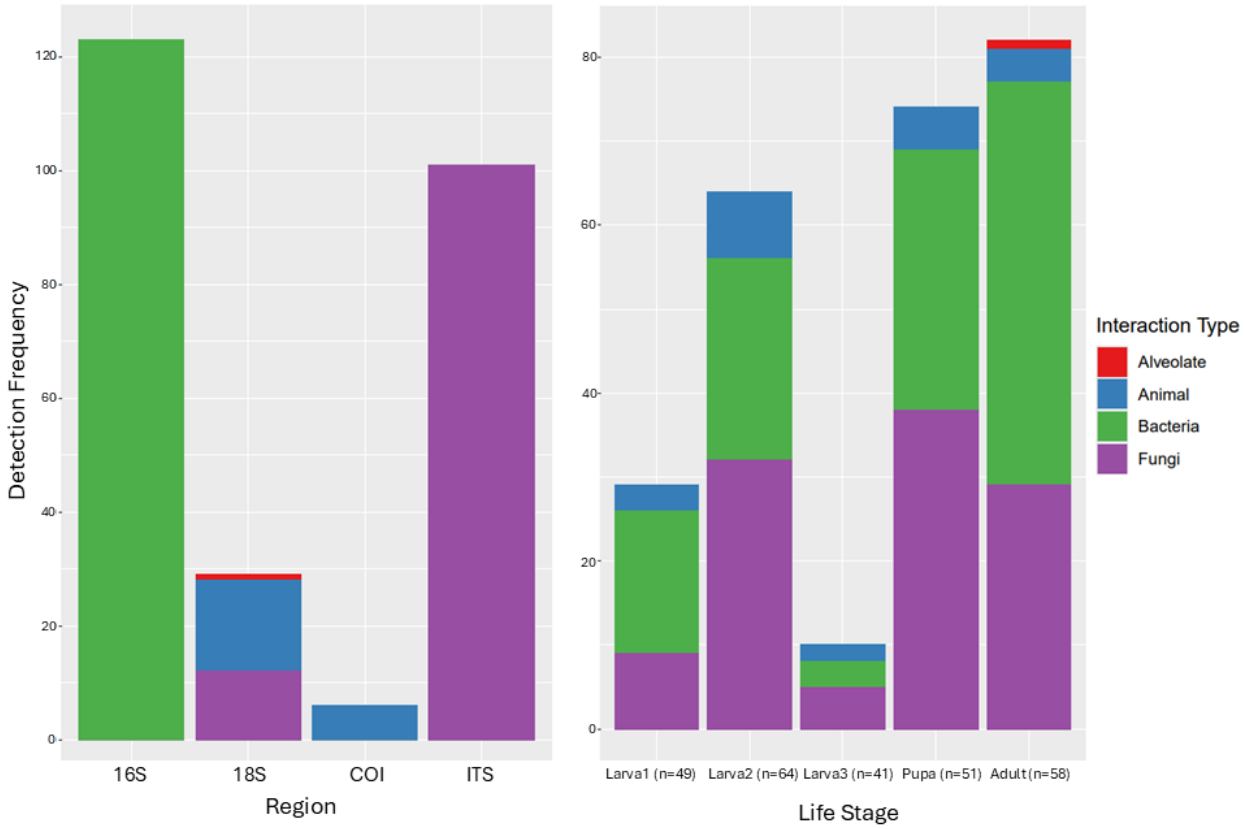

Figure 2.2. Interactions by target locus and in life stages of emerald ash borers. The recovery of various targets by target amplified locus (left) shows high variability in taxonomic richness. The profile of taxonomic recovery of co-amplified taxa is highly variable between life stages (right).

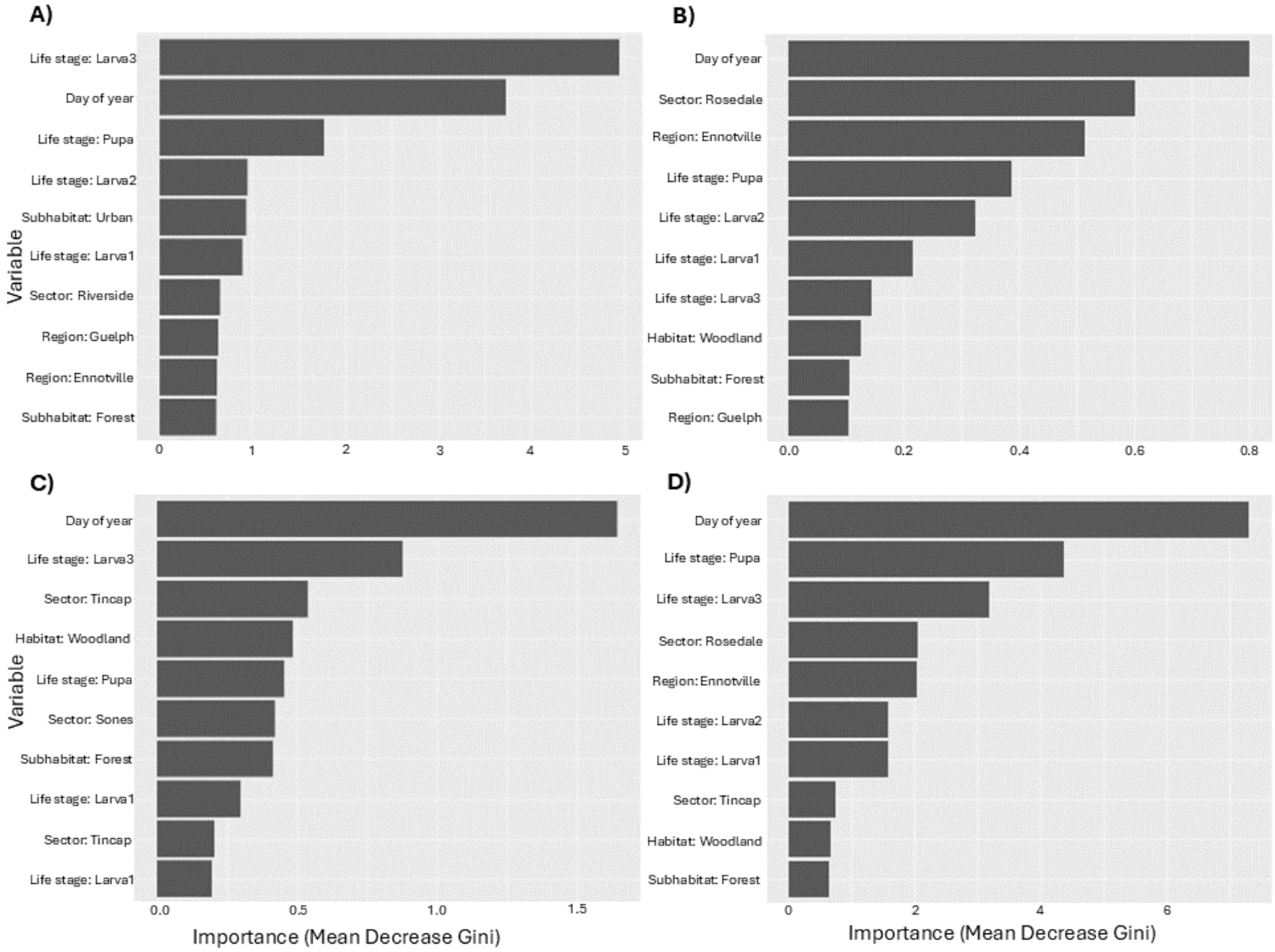

Figure 2.3. Top ten most important variables in random forest models. The ten variables with the highest importance (mean decrease Gini) are ordered top to bottom from most important to least important in classifying the presence or absence of interactions in A) the overall interaction model, B) the animal interaction model, C) the bacteria interaction model, and D) the fungi interaction model.

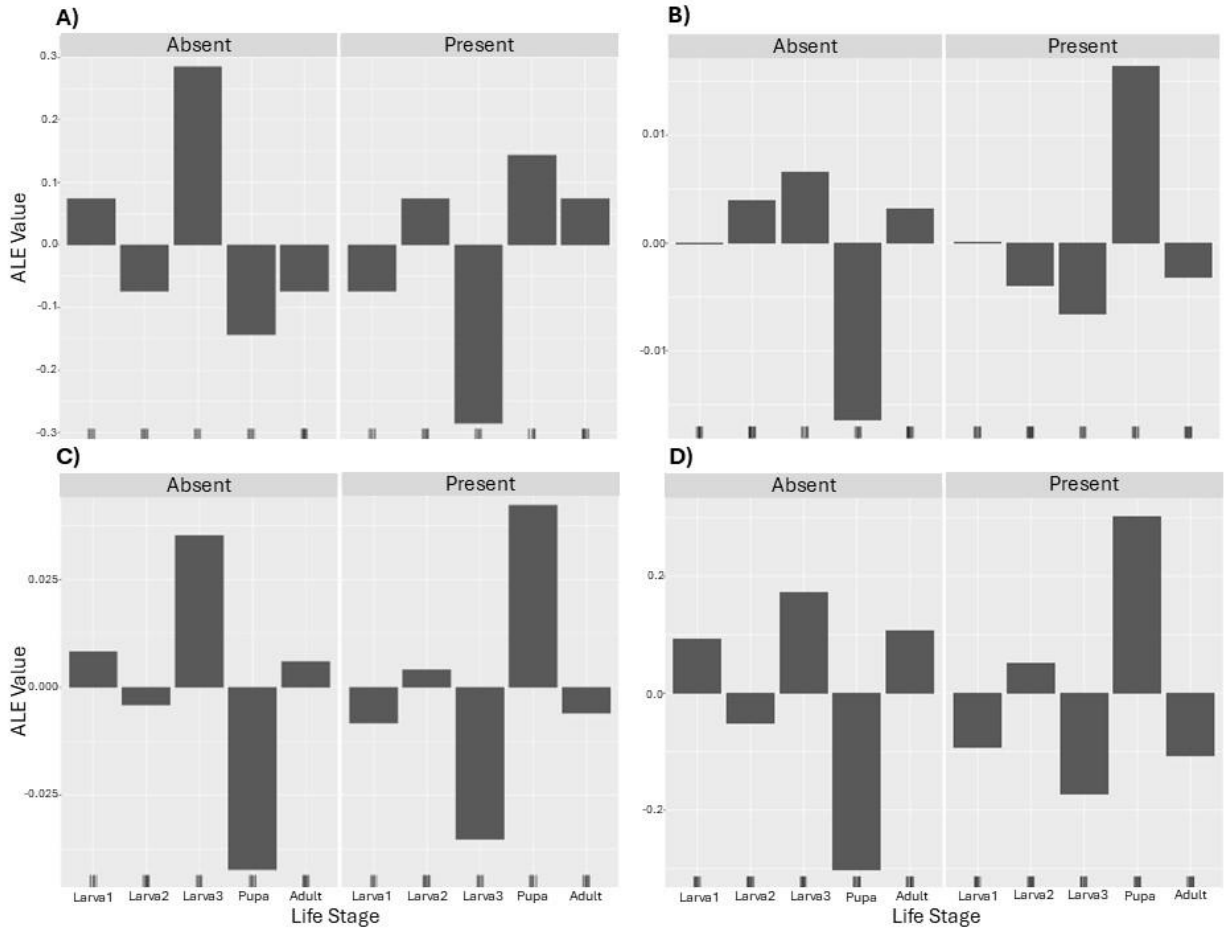

Figure 2.4. Effect of emerald ash borer life stage on the detection of interactions. Accumulated Local Effects (ALE) plots portray the effect life stage had on the presence or absence of A) overall interactions, B) animal interactions, C) bacteria interactions, and D) fungi interactions.

Chapter 4: Conclusions

The overall objective of this thesis was to develop standardized protocols for the use of metabarcoding to identify ecological interactions between insects and other taxa in the environment and to apply this to a target taxon to explore symbiome vs. holobiont interactions. In this thesis I use the invasive emerald ash borer as the focus of extensive ecological analysis and explore ecological correlates of their interactions inside a tree.

Major Conclusions from Chapter 2

My objectives for Chapter 2 were to evaluate different target loci for the identification of ecological interactions and develop a standardized protocol for the analysis of these interactions, thus allowing symbionts and holobionts to be identified. Based on the results of my analysis, I can conclude that my methods were successful in capturing a wide variety of interactions within spongy moth and emerald ash borer samples. To identify the greatest diversity of interactions each primer locus can be used in combination, however, some loci may be excluded if the analysis calls for it (e.g., rbcL was excluded from the emerald ash borer analysis as they are known to only target ash trees). While there was some overlap in the taxa detected (e.g., plants detected by both rbcL and ITS in spongy moths, or fungi detected by ITS and 18S in emerald ash borers), unique taxa were also detected in each locus and each locus contributed to the overall view of the ecological interactions of the insects. However, to simplify the approach the ITS locus (specifically the primer pair ITS-S2F/ITS4 used with spongy moths) can be used in place of the rbcL locus as it identified plants and fungi simultaneously with a greater taxonomic resolution than rbcL. This analysis also highlighted the need for caution regarding potential

environmental contamination of specimens that may generate false positive “interactions” from contact environmental DNA.

Major Conclusions from Chapter 3

The objective of Chapter 3 was to assess how ecological factors influence the presence or absence of ecological interactions. More specifically, I used random forest models and ALE plots to examine the impact that life stage, collection date, collection location, and collection habitat had on the detection of each type of interaction (split into overall, animal, bacteria, and fungi interactions). I also looked for potential agents of biological control (i.e., viruses, fungi, parasites, etc.). I found that interactions were more likely to be detected in the pupal specimens and in samples collected earlier in the year. I also found regional variation in interactions. However, my results should be treated with caution as the sample size was relatively small. My analysis also demonstrated that biological controls can be detected across life stages and potentially monitored using metabarcoding.

Future Considerations

Although it has been demonstrated that the method outlined in Chapter 2 can successfully identify a wide variety of interactions, there is still some uncertainty in differentiating symbiont (co-occurring and co-evolving) from holobiont (co-occurring) interactions. While I used negative filtering to reduce non-relevant detections (i.e., co-occurring DNA that specimens do not directly interact with), it is clear that it was not able to eliminate all instances of surface eDNA contamination (e.g., the likely irrelevant detection of a pine-

feeding aphid and its predator in spongy moth samples; see Chapter 2). An alternative approach to reduce surface contaminate detections can be to wash specimens in a bleach solution prior to extraction (Binetruy et al., 2019; Hausmann et al., 2021; Huszarik et al., 2023), however this would artificially reduce the detection of true interactions associated with the surface of insects (e.g., microbial organisms). Similarly, the strict filtering protocols used artificially reduced detections of non-host taxa that may have represented true symbiont interactions. The best approach may be to consider internal (e.g., gut) and external tissues separately, where detections from internal tissues may be considered symbionts while external tissue detections must be scrutinized further. However, it can also be argued that internal DNA signals are not necessarily true symbionts of the target organism but instead may be associated with the organism that was consumed. Or more simply, internal detections may not truly be co-evolving organisms and thus cannot be considered symbionts. It is extremely difficult to classify a symbiotic interaction when evidence of reciprocal genomic adaptations is required as proof of co-evolution, and more research must be done to address this. Alternatively, the definition of symbiome vs. holobiont, or direct interactions vs. co-occurrence, may need to be addressed to include the continuum of interactions that do not quite fit in either category but better reflect the diversity of ecological communities.

References

Abdu-Allah, G. A. M., & Pittendrigh, B. R. (2018). Lethal and sub-lethal effects of select macrocyclic lactones insecticides on forager worker honey bees under laboratory experimental conditions. *Ecotoxicology*, *27*(1), 81–88. https://doi.org/10.1007/s10646-017-1872-6

Allen, M. C., Lockwood, J. L., Kwait, R., Vastano, A. R., Peterson, D. L., Tkacenko, L. A., Kisurin, A., Stringham, O., Angle, J., & Jaffe, B. (2023). Using surface environmental DNA to assess arthropod biodiversity within a forested ecosystem. *Environmental DNA*, *5*(6), 1652–1666. https://doi.org/10.1002/edn3.487

Allstadt, A. J., Haynes, K. J., Liebhold, A. M., & Johnson, D. M. (2013). Long-term shifts in the cyclicity of outbreaks of a forest-defoliating insect. *Oecologia*, *172*(1), 141–151. https://doi.org/10.1007/s00442-012-2474-x

Angelini, C., Griffin, J. N., van de Koppel, J., Lamers, L. P. M., Smolders, A. J. P., Derksen-Hooijberg, M., van der Heide, T., & Silliman, B. R. (2016). A keystone mutualism underpins resilience of a coastal ecosystem to drought. *Nature Communications*, *7*, 12473. https://doi.org/10.1038/ncomms12473

Anulewicz, A., McCullough, D., & Cappaert, D. (2007). Emerald Ash Borer (Agrilus planipennis) Density and Canopy Dieback in Three North American Ash Species. *Arboriculture & Urban Forestry*, *33*(5), 338–349. https://doi.org/10.48044/jauf.2007.039

Aukema, J. E., Leung, B., Kovacs, K., Chivers, C., Britton, K. O., Englin, J., Frankel, S. J., Haight, R. G., Holmes, T. P., Liebhold, A. M., McCullough, D. G., & Holle, B. V. (2011). Economic Impacts of Non-Native Forest Insects in the Continental United States. *PLOS ONE*, *6*(9), e24587. https://doi.org/10.1371/journal.pone.0024587

Barraclough, T. G. (2015). How Do Species Interactions Affect Evolutionary Dynamics Across Whole Communities? *Annual Review of Ecology, Evolution, and Systematics*, *46*(1), 25–48. https://doi.org/10.1146/annurev-ecolsys-112414-054030

Bell, K. L., Fowler, J., Burgess, K. S., Dobbs, E. K., Gruenewald, D., Lawley, B., Morozumi, C., & Brosi, B. J. (2017). Applying pollen DNA metabarcoding to the study of plant–pollinator interactions. *Applications in Plant Sciences*, *5*(6), 1600124. https://doi.org/10.3732/apps.1600124

BenDor, T. K., Metcalf, S. S., Fontenot, L. E., Sangunett, B., & Hannon, B. (2006). Modeling the spread of the Emerald Ash Borer. *Ecological Modelling*, *197*(1–2), 221–236. https://doi.org/10.1016/j.ecolmodel.2006.03.003

Berryman, A. A. (1996). What causes population cycles of forest Lepidoptera? *Trends in Ecology & Evolution*, *11*(1), 28–32. https://doi.org/10.1016/0169-5347(96)81066-4

Binetruy, F., Dupraz, M., Buysse, M., & Duron, O. (2019). Surface sterilization methods impact measures of internal microbial diversity in ticks. *Parasites & Vectors*, *12*, 268. https://doi.org/10.1186/s13071-019-3517-5

Boehme, P., Amendt, J., & Zehner, R. (2012). The use of COI barcodes for molecular identification of forensically important fly species in Germany. *Parasitology Research*, *110*(6), 2325–2332. https://doi.org/10.1007/s00436-011-2767-8

Boersma, M., Aberle, N., Hantzsche, F. M., Schoo, K. L., Wiltshire, K. H., & Malzahn, A. M. (2008). Nutritional Limitation Travels up the Food Chain. *International Review of Hydrobiology*, *93*(4–5), 479–488. https://doi.org/10.1002/iroh.200811066

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852. https://doi.org/10.1038/s41587-019-0209-9

Bradshaw, C. J. A., Leroy, B., Bellard, C., Roiz, D., Albert, C., Fournier, A., Barbet-Massin, M., Salles, J.-M., Simard, F., & Courchamp, F. (2016). Massive yet grossly underestimated global costs of invasive insects. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms12986

Bray, A. M., Bauer, L. S., Poland, T. M., Haack, R. A., Cognato, A. I., & Smith, J. J. (2011). Genetic analysis of emerald ash borer (Agrilus planipennis Fairmaire) populations in Asia and North America. *Biological Invasions*, *13*(12), 2869–2887. https://doi.org/10.1007/s10530-011-9970-5

Breiman, L. (2001). Random Forests. *Mach. Learn.*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown, M. W., & Cameron, E. A. (1982). Natural enemies ofLymantria dispar [Lep.: Lymantriidae] eggs in Central Pennsylvania, U.S.A., and a review of the world literature on natural enemies ofL. dispar eggs. *Entomophaga*, *27*(3), 311–321. https://doi.org/10.1007/BF02374814

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Canham, C. D. (1988). Growth and Canopy Architecture of Shade-Tolerant Trees: Response to Canopy Gaps. *Ecology*, *69*(3), 786–795. https://doi.org/10.2307/1941027

Cappaert, D., & McCullough, D. G. (2009). Occurrence and seasonal abundance of Atanycolus cappaerti (Hymenoptera: Braconidae) a native parasitoid of emerald ash borer, Agrilus planipennis (Coleoptera: Buprestidae). 42(1&2):16–29. *Great Lakes Entomologist*, *42*(1 & 2), 16–29.

Cappaert, D., McCullough, D. G., Poland, T. M., & Siegert, N. W. (2005). Emerald Ash Borer in North America: A Research and Regulatory Challenge. *American Entomologist*, *51*(3), 152–165. https://doi.org/10.1093/ae/51.3.152

Carta, L. K., & Li, S. (2018). Improved 18S small subunit rDNA primers for problematic nematode amplification. *Journal of Nematology*, *50*(4), 533–542. https://doi.org/10.21307/jofnem-2018-051

CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 12794–12797. https://doi.org/10.1073/pnas.0905845106

Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., Zhu, Y., Ma, X., Gao, T., Pang, X., Luo, K., Li, Y., Li, X., Jia, X., Lin, Y., & Leon, C. (2010). Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLOS ONE*, *5*(1), e8613. https://doi.org/10.1371/journal.pone.0008613

Cheng, T., Xu, C., Lei, L., Li, C., Zhang, Y., & Zhou, S. (2016). Barcoding the kingdom Plantae: New PCR primers for ITS regions of plants with improved universality and specificity. *Molecular Ecology Resources*, *16*(1), 138–149. https://doi.org/10.1111/1755-0998.12438

Clare, E. L., Economou, C. K., Bennett, F. J., Dyer, C. E., Adams, K., McRobie, B., Drinkwater, R., & Littlefair, J. E. (2022). Measuring biodiversity from DNA in the air. *Current Biology*, *32*(3), 693-700.e5. https://doi.org/10.1016/j.cub.2021.11.064

Clare, E. L., Economou, C. K., Faulkes, C. G., Gilbert, J. D., Bennett, F., Drinkwater, R., & Littlefair, J. E. (2021). eDNAir: Proof of concept that animal DNA can be collected from air sampling. *PeerJ*, *9*, e11030. https://doi.org/10.7717/peerj.11030

Clare, E. L., Fazekas, A. J., Ivanova, N. V., Floyd, R. M., Hebert, P. D. N., Adams, A. M., Nagel, J., Girton, R., Newmaster, S. G., & Fenton, M. B. (2018). Approaches to integrating genetic data into ecological networks. *Molecular Ecology*, *28*(2), 503–519. https://doi.org/10.1111/mec.14941

Compson, Z. G., McClenaghan, B., Singer, G. A. C., Fahner, N. A., & Hajibabaei, M. (2020). Metabarcoding From Microbes to Mammals: Comprehensive Bioassessment on a Global Scale. *Frontiers in Ecology and Evolution*, *8*. https://www.frontiersin.org/articles/10.3389/fevo.2020.581835

Davidson, C. B., Gottschalk, K. W., & Johnson, J. E. (1999). Tree Mortality Following Defoliation by the European Gypsy Moth (Lymantria dispar L.) in the United States: A Review. *Forest Science*, *45*(1), 74–84. https://doi.org/10.1093/forestscience/45.1.74

Donovan, G. H., Butry, D. T., Michael, Y. L., Prestemon, J. P., Liebhold, A. M., Gatziolis, D., & Mao, M. Y. (2013). The Relationship Between Trees and Human Health: Evidence from the Spread of the Emerald Ash Borer. *American Journal of*

Preventive Medicine, *44*(2), 139–145. https://doi.org/10.1016/j.amepre.2012.09.066

Drinkwater, R., Schnell, I. B., Bohmann, K., Bernard, H., Veron, G., Clare, E., Gilbert, M. T. P., & Rossiter, S. J. (2019). Using metabarcoding to compare the suitability of two blood-feeding leech species for sampling mammalian diversity in North Borneo. *Molecular Ecology Resources*, *19*(1), 105–117. https://doi.org/10.1111/1755-0998.12943

Duan, J. J., Bauer, L. S., Van Driesche, R. G., & Gould, J. R. (2018). Progress and Challenges of Protecting North American Ash Trees from the Emerald Ash Borer Using Biological Control. *Forests*, *9*(3), Article 3. https://doi.org/10.3390/f9030142

Duan, J. J., Fuester, R. W., Wildonger, J., Taylor, P. B., Barth, S., & Spichiger, S. E. (2009). Parasitoids Attacking the Emerald Ash Borer (Coleoptera: Buprestidae) in Western Pennsylvania. *The Florida Entomologist*, *92*(4), 588–592.

Duan, J. J., Gould, J. R., Quinn, N. F., Petrice, T. R., Slager, B. H., Poland, T. M., Bauer, L. S., Rutledge, C. E., Elkinton, J. S., & Van Driesche, R. G. (2023). Protection of North American ash against emerald ash borer with biological control: Ecological premises and progress toward success. *BioControl*, *68*(2), 87–100. https://doi.org/10.1007/s10526-023-10182-w

Edgar, R. C. (2016). *SINTAX: A simple non-Bayesian taxonomy classifier for 16S and ITS sequences* (p. 074161). bioRxiv. https://doi.org/10.1101/074161

Elbrecht, V., & Leese, F. (2015). Can DNA-Based Ecosystem Assessments Quantify Species Abundance? Testing Primer Bias and Biomass—Sequence Relationships with an Innovative Metabarcoding Protocol. *PLOS ONE*, *10*(7), e0130324. https://doi.org/10.1371/journal.pone.0130324

Elbrecht, V., Vamos, E. E., Meissner, K., Aroviita, J., & Leese, F. (2017). Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution*, *8*(10), 1265–1275. https://doi.org/10.1111/2041-210X.12789

Elkinton, J. S., & Liebhold, A. M. (1990). Population Dynamics of Gypsy Moth in North America. *Annual Review of Entomology*, *27*.

Fajvan, M. A., & Wood, J. M. (1996). Stand structure and development after gypsy moth defoliation in the Appalachian Plateau. *Forest Ecology and Management*, *89*(1–3), 79–88. https://doi.org/10.1016/S0378-1127(96)03865-0

Ferris, H. (2009). The beer mat nematode, Panagrellus redivivus: A study of the connectedness of scientific discovery. *Journal of Nematode Morphology and Systematics*, *12*(1), Article 1.

Flower, C. E., Knight, K. S., & Gonzalez-Meler, M. A. (2013). Impacts of the emerald ash borer (Agrilus planipennisFairmaire) induced ash (Fraxinusspp.) mortality on forest carbon cycling and successional dynamics in the eastern United States. *Biological Invasions*, *15*(4), 931–944. https://doi.org/10.1007/s10530-012-0341-7

Floyd, R., Abebe, E., Papert, A., & Blaxter, M. (2002). Molecular barcodes for soil nematode identification. *Molecular Ecology*, *11*(4), 839–850. https://doi.org/10.1046/j.1365-294X.2002.01485.x

Floyd, R., Prosser, S., & Jafarpour, S. (2023). *DNA barcoding on Oxford Nanopore: Multiplexing up to 24 x 96-well plates*. https://www.protocols.io/view/dna-barcoding-on-oxford-nanopore-multiplexing-up-t-c2cnyave

Foucher, A., Evrard, O., Ficetola, G. F., Gielly, L., Poulain, J., Giguet-Covex, C., Laceby, J. P., Salvador-Blanes, S., Cerdan, O., & Poulenard, J. (2020). Persistence of environmental DNA in cultivated soils: Implication of this memory effect for reconstructing the dynamics of land use and cover changes. *Scientific Reports*, *10*(1), 10502. https://doi.org/10.1038/s41598-020-67452-1

Fretwell, S. D. (1987). Food Chain Dynamics: The Central Theory of Ecology? *Oikos*, *50*(3), 291–301. https://doi.org/10.2307/3565489

Gagn, R. J., & Havill, N. P. (2020). A New Species of Lestodiplosis (Diptera: Cecidomyiidae) Preying on Pineus (Hemiptera: Adelgidae), with a Redescription of Lestodiplosis juniperina (Felt). *Proceedings of the Entomological Society of Washington*, *122*(4), 834–849. https://doi.org/10.4289/0013-8797.122.4.834

Gandhi, K. J. K., & Herms, D. A. (2010a). Direct and indirect effects of alien insect herbivores on ecological processes and interactions in forests of eastern North America. *Biological Invasions*, *12*(2), 389–405. https://doi.org/10.1007/s10530-009-9627-9

Gandhi, K. J. K., & Herms, D. A. (2010b). North American arthropods at risk due to widespread Fraxinus mortality caused by the Alien Emerald ash borer. *Biological Invasions*, *12*(6), 1839–1846. https://doi.org/10.1007/s10530-009-9594-1

Garrett, N. R., Watkins, J., Simmons, N. B., Fenton, B., Maeda-Obregon, A., Sanchez, D. E., Froehlich, E. M., Walker, F. M., Littlefair, J. E., & Clare, E. L. (2023). Airborne eDNA documents a diverse and ecologically complex tropical bat and other mammal community. *Environmental DNA*, *5*(2), 350–362. https://doi.org/10.1002/edn3.385

Gattoni, K., Gendron, E. M. S., Sandoval-Ruiz, R., Borgemeier, A., McQueen, J. P., M. Shepherd, R., Slos, D., O. Powers, T., & L. Porazinska, D. (2023). 18S-NemaBase: Curated 18S rRNA Database of Nematode Sequences. *Journal of Nematology*, *55*(1), 20230006. https://doi.org/10.2478/jofnem-2023-0006

GBIF. (2023). *GBIF Home Page*. GBIF: The Global Biodiversity Information Facility. https://www.gbif.org/

Gill, R. J., Ramos-Rodriguez, O., & Raine, N. E. (2012). Combined pesticide exposure severely affects individual- and colony-level traits in bees. *Nature*, *491*(7422), 105–108. https://doi.org/10.1038/nature11585

Government of Canada, C. F. I. A. (2013a, May 16). *RMD-13-01: Regulated Areas for Emerald Ash Borer (EAB) (Agrilus planipennis Fairmaire)March 2014 (1st*

revision) [Reference material,policy]. https://inspection.canada.ca/plant-health/invasive-species/directives/pest-risk-management/emerald-ash-borer/eng/1368741925939/1368741926892#item5.2

Government of Canada, C. F. I. A. (2013b, June 13). *Questions and Answers: Wasps as biological control agents for Emerald Ash Borers*. https://inspection.canada.ca/plant-health/invasive-species/insects/emerald-ash-borer/wasps/eng/1371137262586/1371137530758

Government of Ontario. (2014, July 18). *Emerald ash borer*. Ontario.Ca. http://www.ontario.ca/page/emerald-ash-borer

Grant, D. M., Brodnicke, O. B., Evankow, A. M., Ferreira, A. O., Fontes, J. T., Hansen, A. K., Jensen, M. R., Kalaycı, T. E., Leeper, A., Patil, S. K., Prati, S., Reunamo, A., Roberts, A. J., Shigdel, R., Tyukosova, V., Bendiksby, M., Blaalid, R., Costa, F. O., Hollingsworth, P. M., … Ekrem, T. (2021). The Future of DNA Barcoding: Reflections from Early Career Researchers. *Diversity*, *13*(7), Article 7. https://doi.org/10.3390/d13070313

Gurnell, J. (1983). Squirrel numbers and the abundance of tree seeds. *Mammal Review*, *13*(2-4), 133–148. https://doi.org/10.1111/j.1365-2907.1983.tb00274.x

Hanshew, A. S., Mason, C. J., Raffa, K. F., & Currie, C. R. (2013). Minimization of chloroplast contamination in 16S rRNA gene pyrosequencing of insect herbivore bacterial communities. *Journal of Microbiological Methods*, *95*(2), 149–155. https://doi.org/10.1016/j.mimet.2013.08.007

Hausmann, A., Höcherl, A., Niessner, A., Zakharov, E., Dolynskyj, S., & Diller, J. (2021). Accurate decontamination of insects from bulk samples does not affect DNA sequencing success. *Spixiana*, *44*(1), 71–76.

Haynes, K. J., Liebhold, A. M., & Johnson, D. M. (2009). Spatial Analysis of Harmonic Oscillation of Gypsy Moth Outbreak Intensity. *Oecologia*, *159*(2), 249–256.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert, P. D. N., Floyd, R., Jafarpour, S., & Prosser, S. W. (2023). *Barcode 100K Specimens: In a Single Nanopore Run* (p. 2023.11.29.569282). bioRxiv. https://doi.org/10.1101/2023.11.29.569282

Helmy, Y. A., & Hafez, H. M. (2022). Cryptosporidiosis: From Prevention to Treatment, a Narrative Review. *Microorganisms*, *10*(12), 2456. https://doi.org/10.3390/microorganisms10122456

Hensel, M. J. S., Silliman, B. R., van de Koppel, J., Hensel, E., Sharp, S. J., Crotty, S. M., & Byrnes, J. E. K. (2021). A large invasive consumer reduces coastal ecosystem resilience by disabling positive species interactions. *Nature Communications*, *12*(1), 6290. https://doi.org/10.1038/s41467-021-26504-4

Herms, D. A., & McCullough, D. G. (2014). Emerald Ash Borer Invasion of North America: History, Biology, Ecology, Impacts, and Management. *Annual Review of Entomology*, *59*(1), 13–30. https://doi.org/10.1146/annurev-ento-011613-162051

Hope, E. S., McKenney, D. W., Pedlar, J. H., Lawrence, K., & MacDonald, H. (2021). Canadian efforts to slow the spread of Emerald Ash Borer (*Agrilus planipennis* Fairmaire) are economically efficient. *Ecological Economics*, *188*, 107126. https://doi.org/10.1016/j.ecolecon.2021.107126

Huszarik, M., Röder, N., Eberhardt, L., Kennedy, S., Krehenwinkel, H., Schwenk, K., & Entling, M. H. (2023). External DNA contamination and efficiency of bleach decontamination for arthropod diet analysis. *Environmental DNA*, *5*(3), 540–550. https://doi.org/10.1002/edn3.410

International Union for Conservation of Nature. (2022, Version). *The IUCN Red List of Threatened Species*. IUCN Red List of Threatened Species. https://www.iucnredlist.org

Jennings, D. E., Gould, J. R., Vandenberg, J. D., Duan, J. J., & Shrewsbury, P. M. (2013). Quantifying the Impact of Woodpecker Predation on Population Dynamics of the Emerald Ash Borer (Agrilus planipennis). *PLOS ONE*, *8*(12), e83491. https://doi.org/10.1371/journal.pone.0083491

Jung, S., Duwal, R. K., & Lee, S. (2011). COI barcoding of true bugs (Insecta, Heteroptera). *Molecular Ecology Resources*, *11*(2), 266–270. https://doi.org/10.1111/j.1755-0998.2010.02945.x

Kassambara, A. (2023). *ggcorrplot: Visualization of a Correlation Matrix using "ggplot2"* (Version R package version 0.1.4.1, p. 0.1.4.1) [Computer software]. https://CRAN.R-project.org/package=ggcorrplot

Khan, M. M., Nawaz, M., Hua, H., Cai, W., & Zhao, J. (2018). Lethal and sublethal effects of emamectin benzoate on the rove beetle, *Paederus fuscipes*, a non-target predator of rice brown planthopper, *Nilaparvata lugens*. *Ecotoxicology and Environmental Safety*, *165*, 19–24. https://doi.org/10.1016/j.ecoenv.2018.08.047

Kilani-Morakchi, S., Morakchi-Goudjil, H., & Sifi, K. (2021). Azadirachtin-Based Insecticide: Overview, Risk Assessments, and Future Directions. *Frontiers in Agronomy*, *3*. https://doi.org/10.3389/fagro.2021.676208

Klooster, W. S., Herms, D. A., Knight, K. S., Herms, C. P., McCullough, D. G., Smith, A., Gandhi, K. J. K., & Cardina, J. (2014). Ash (Fraxinus spp.) mortality, regeneration, and seed bank dynamics in mixed hardwood forests following invasion by emerald ash borer (Agrilus planipennis). *Biological Invasions*, *16*(4), 859–873. https://doi.org/10.1007/s10530-013-0543-7

Knight, K. S., Brown, J. P., & Long, R. P. (2013). Factors affecting the survival of ash (Fraxinus spp.) trees infested by emerald ash borer (Agrilus planipennis). *Biological Invasions*, *15*(2), 371–383. https://doi.org/10.1007/s10530-012-0292-z

Kozlowski, T. T., Kramer, P. J., & Pallardy, S. G. (1991). *The Physiological Ecology of Woody Plants*. Academic Press.

Kreutzweiser, D., Good, K., Chartrand, D., Scarr, T., & Thompson, D. (2007). Non-target effects on aquatic decomposer organisms of imidacloprid as a systemic insecticide to control emerald ash borer in riparian trees. *Ecotoxicology and Environmental Safety*, *68*(3), 315–325. https://doi.org/10.1016/j.ecoenv.2007.04.011

Kudoh, A., Minamoto, T., & Yamamoto, S. (2020). Detection of herbivory: eDNA detection from feeding marks on leaves. *Environmental DNA*, *2*(4), 627–634. https://doi.org/10.1002/edn3.113

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*, 1–26. https://doi.org/10.18637/jss.v028.i05

Lejzerowicz, F., Esling, P., Pillet, L., Wilding, T. A., Black, K. D., & Pawlowski, J. (2015). High-throughput sequencing and morphology perform equally well for benthic monitoring of marine ecosystems. *Scientific Reports*, *5*(1), 13932. https://doi.org/10.1038/srep13932

Leuschner, W. A., Young, J. A., Waldon, S. A., & Ravlin, F. W. (1996). Potential Benefits of Slowing the Gypsy Moth's Spread. *Southern Journal of Applied Forestry*, *20*(2), 65–73. https://doi.org/10.1093/sjaf/20.2.65

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22.

Liebhold, A. M., Gottschalk, K. W., Muzika, R.-M., Montgomery, M. E., Young, R., O'Day, K., & Kelley, B. (1995). Suitability of North American tree species to gypsy moth: A summary of field and laboratory tests. *Gen. Tech. Rep. NE-211. Radnor, PA: U. S. Department of Agriculture, Forest Service, Northeastern Forest Experiment Station. 34 p.*, *211*. https://doi.org/10.2737/NE-GTR-211

Lindell, C. A., McCullough, D. G., Cappaert, D., Apostolou, N. M., & Roth, M. B. (2008). Factors Influencing Woodpecker Predation on Emerald Ash Borer. *The American Midland Naturalist*, *159*(2), 434–444.

Littlefair, J. E., Allerton, J. J., Brown, A. S., Butterfield, D. M., Robins, C., Economou, C. K., Garrett, N. R., & Clare, E. L. (2023). Air-quality networks collect environmental DNA with the potential to measure biodiversity at continental scales. *Current Biology: CB*, *33*(11), R426–R428. https://doi.org/10.1016/j.cub.2023.04.036

Liu, J., Yu, Y., Cai, Z., Bartlam, M., & Wang, Y. (2015). Comparison of ITS and 18S rDNA for estimating fungal diversity using PCR–DGGE. *World Journal of Microbiology and Biotechnology*, *31*(9), 1387–1395. https://doi.org/10.1007/s11274-015-1890-6

Liu, T., Chen, D., Li, Y., Wang, X., & Wang, F. (2018). Enantioselective Bioaccumulation and Toxicity of the Neonicotinoid Insecticide Dinotefuran in Earthworms (Eisenia

fetida). *Journal of Agricultural and Food Chemistry*, *66*(17), 4531–4540. https://doi.org/10.1021/acs.jafc.8b00285

Lord, N. s., Kaplan, C. w., Shank, P., Kitts, C. l., & Elrod, S. l. (2002). Assessment of fungal diversity using terminal restriction fragment (TRF) pattern analysis: Comparison of 18S and ITS ribosomal regions. *FEMS Microbiology Ecology*, *42*(3), 327–337. https://doi.org/10.1111/j.1574-6941.2002.tb01022.x

MacFarlane, D. W., & Meyer, S. P. (2005). Characteristics and distribution of potential ash tree hosts for emerald ash borer. *Forest Ecology and Management*, *213*(1), 15–24. https://doi.org/10.1016/j.foreco.2005.03.013

Macher, T.-H., Schütz, R., Hörren, T., Beermann, A. J., & Leese, F. (2023). It's raining species: Rainwash eDNA metabarcoding as a minimally invasive method to assess tree canopy invertebrate diversity. *Environmental DNA*, *5*(1), 3–11. https://doi.org/10.1002/edn3.372

Marquina, D., Esparza-Salas, R., Roslin, T., & Ronquist, F. (2019). Establishing arthropod community composition using metabarcoding: Surprising inconsistencies between soil samples and preservative ethanol and homogenate from Malaise trap catches. *Molecular Ecology Resources*, *19*(6), 1516–1530. https://doi.org/10.1111/1755-0998.13071

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), Article 1. https://doi.org/10.14806/ej.17.1.200

Matias Rodrigues, J. F., Schmidt, T. S. B., Tackmann, J., & von Mering, C. (2017). MAPseq: Highly efficient k-mer search with confidence estimates, for rRNA sequence analysis. *Bioinformatics*, *33*(23), 3808–3810. https://doi.org/10.1093/bioinformatics/btx517

McCullough, D. G. (2020). Challenges, tactics and integrated management of emerald ash borer in North America. *Forestry: An International Journal of Forest Research*, *93*(2), 197–211. https://doi.org/10.1093/forestry/cpz049

McCullough, D. G., Poland, T. M., Anulewicz, A. C., & Cappaert, D. (2009). Emerald Ash Borer (Coleoptera: Buprestidae) Attraction to Stressed or Baited Ash Trees. *Environmental Entomology*, *38*(6), 1668–1679. https://doi.org/10.1603/022.038.0620

McCullough, D. G., Poland, T. M., & Cappaert, D. (2009). Attraction of the emerald ash borer to ash trees stressed by girdling, herbicide treatment, or wounding. *Canadian Journal of Forest Research*, *39*(7), 1331–1345. https://doi.org/10.1139/X09-057

Mcmanus, M., & Csóka, G. (2007). *History and Impact of Gypsy Moth in North America and Comparison to Recent Outbreaks in Europe*. 18.

Molnar, C., Casalicchio, G., & Bischl, B. (2018). iml: An R package for Interpretable Machine Learning. *Journal of Open Source Software*, *3*(26), 786. https://doi.org/10.21105/joss.00786

Muirhead, J. R., Leung, B., van Overdijk, C., Kelly, D. W., Nandakumar, K., Marchant, K. R., & MacIsaac, H. J. (2006). Modelling local and long-distance dispersal of invasive emerald ash borer Agrilus planipennis (Coleoptera) in North America. *Diversity and Distributions*, *12*(1), 71–79. https://doi.org/10.1111/j.1366-9516.2006.00218.x

Muzika, R. M., & Liebhold, A. M. (1999). Changes in radial increment of host and nonhost tree species with gypsy moth defoliation. *Canadian Journal of Forest Research*, *29*(9), 1365–1373. https://doi.org/10.1139/x99-098

Mužinić, V., & Želježić, D. (2018). Non-target toxicity of novel insecticides. *Archives of Industrial Hygiene and Toxicology*, *69*(2), 86–102. https://doi.org/10.2478/aiht-2018-69-3111

Nakajima, H. (2015). Defoliation by gypsy moths negatively affects the production of acorns by two Japanese oak species. *Trees*, *29*(5), 1559–1566. https://doi.org/10.1007/s00468-015-1237-9

Natural Resources Canada. (2013, October 25). *Emerald ash borer*. Natural-Resources.Canada.Ca; Natural Resources Canada. https://natural-resources.canada.ca/our-natural-resources/forests/wildland-fires-insects-disturbances/top-forest-insects-and-diseases-canada/emerald-ash-borer/13377

Natural Resources Canada. (2022, March 25). *How does the forest sector contribute to Canada's economy?* Natural Resources Canada; Natural Resources Canada. https://www.nrcan.gc.ca/our-natural-resources/forests/state-canadas-forests-report/forest-industry-contribute/16517

Orlova-Bienkowskaja, M. J., & Bieńkowski, A. O. (2016). The life cycle of the emerald ash borer *Agrilus planipennis* in European Russia and comparisons with its life cycles in Asia and North America: Life cycle of emerald ash borer in European Russia. *Agricultural and Forest Entomology*, *18*(2), 182–188. https://doi.org/10.1111/afe.12140

Palmieri, L., Bozza, E., & Giongo, L. (2009). Soft Fruit Traceability in Food Matrices using Real-Time PCR. *Nutrients*, *1*(2), 316–328. https://doi.org/10.3390/nu1020316

Park, D.-S., Foottit, R., Maw, E., & Hebert, P. D. N. (2011). Barcoding Bugs: DNA-Based Identification of the True Bugs (Insecta: Hemiptera: Heteroptera). *PLOS ONE*, *6*(4), e18749. https://doi.org/10.1371/journal.pone.0018749

Peltonen, M., Liebhold, A. M., Bjørnstad, O. N., & Williams, D. W. (2002). Spatial Synchrony in Forest Insect Outbreaks: Roles of Regional Stochasticity and Dispersal. *Ecology*, *83*(11), 3120–3129. https://doi.org/10.1890/0012-9658(2002)083[3120:SSIFIO]2.0.CO;2

Pettis, J. S., vanEngelsdorp, D., Johnson, J., & Dively, G. (2012). Pesticide exposure in honey bees results in increased levels of the gut pathogen Nosema. *Die Naturwissenschaften*, *99*(2), 153–158. https://doi.org/10.1007/s00114-011-0881-1

Pompanon, F., Deagle, B. E., Symondson, W. O. C., Brown, D. S., Jarman, S. N., & Taberlet, P. (2012). Who is eating what: Diet assessment using next generation sequencing. *Molecular Ecology*, *21*(8), 1931–1950. https://doi.org/10.1111/j.1365-294X.2011.05403.x

Pornon, A., Escaravage, N., Burrus, M., Holota, H., Khimoun, A., Mariette, J., Pellizzari, C., Iribar, A., Etienne, R., Taberlet, P., Vidal, M., Winterton, P., Zinger, L., & Andalo, C. (2016). Using metabarcoding to reveal and quantify plant-pollinator interactions. *Scientific Reports*, *6*(1), 27282. https://doi.org/10.1038/srep27282

Pureswaran, D. S., & Poland, T. M. (2009). Host Selection and Feeding Preference of Agrilus planipennis (Coleoptera: Buprestidae) on Ash (Fraxinus spp.). *Environmental Entomology*, *38*(3), 757–765. https://doi.org/10.1603/022.038.0328

Pushpakumara, B. L. D. U., Tandon, K., Willis, A., & Verbruggen, H. (2023). Unravelling microalgal-bacterial interactions in aquatic ecosystems through 16S rRNA gene-based co-occurrence networks. *Scientific Reports*, *13*(1), 2743. https://doi.org/10.1038/s41598-023-27816-9

R Core Team. (2023). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Ratnasingham, S. (2019). mBRAVE: The Multiplex Barcode Research And Visualization Environment. *Biodiversity Information Science and Standards*, *3*, e37986. https://doi.org/10.3897/biss.3.37986

Rehner, S. A., Minnis, A. M., Sung, G.-H., Luangsa-ard, J. J., Devotto, L., & Humber, R. A. (2011). Phylogeny and systematics of the anamorphic, entomopathogenic genus Beauveria. *Mycologia*, *103*(5), 1055–1073. https://doi.org/10.3852/10-302

Removal of Emerald Ash Borer Domestic Quarantine Regulations, 85 F.R. 81085 (2020). https://www.federalregister.gov/documents/2020/12/15/2020-26734/removal-of-emerald-ash-borer-domestic-quarantine-regulations

Reysenbach, A., & Pace, N. (1995). Reliable amplification of hyperthermophilic Archaeal 16s rRNA genes by the polymerase chain reaction. In F. Robb (Ed.), *Archaea: A Laboratory Manual* (pp. 101–107). Cold Spring Harbor Laboratory Press.

Roger, F., Ghanavi, H. R., Danielsson, N., Wahlberg, N., Löndahl, J., Pettersson, L. B., Andersson, G. K. S., Boke Olén, N., & Clough, Y. (2022). Airborne environmental DNA metabarcoding for the monitoring of terrestrial insects—A proof of concept

from the field. *Environmental DNA*, *4*(4), 790–807. https://doi.org/10.1002/edn3.290

Ryss, A. Y., & Polyanina, K. S. (2022). Life cycle and population dynamics of Rhabditolaimus ulmi (Nematoda: Rhabditida: Diplogastridae) in vitro. *Russian Journal of Nematology*, *30*(1), Article 1.

Sanders, W. B., & Masumoto, H. (2021). Lichen algae: The photosynthetic partners in lichen symbioses. *The Lichenologist*, *53*(5), 347–393. https://doi.org/10.1017/S0024282921000335

Savolainen, V., Cowan, R. S., Vogler, A. P., Roderick, G. K., & Lane, R. (2005). Towards writing the encyclopaedia of life: An introduction to DNA barcoding. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1462), 1805–1811. https://doi.org/10.1098/rstb.2005.1730

Schlesinger, R. C. (1990). Fraxinus americana L. White Ash. In R. M. Burns & B. H. Honkala, *Silvics of North America. Volume 2. Hardwoods* (pp. 333–338).

Schrader, G., Baker, R., Baranchikov, Y., Dumouchel, L., Knight, K. S., McCullough, D. G., Orlova-Bienkowskaja, M. J., Pasquali, S., & Gilioli, G. (2021). How does the Emerald Ash Borer (Agrilus planipennis) affect ecosystem services and biodiversity components in invaded areas? *EPPO Bulletin*, *51*(1), 216–228. https://doi.org/10.1111/epp.12734

Shokralla, S., Singer, G. A. C., & Hajibabaei, M. (2010). Direct PCR amplification and sequencing of specimens' DNA from preservative ethanol. *BioTechniques*, *48*(3), 305–306. https://doi.org/10.2144/000113362

Siegert, N. W., McCullough, D. G., Liebhold, A. M., & Telewski, F. W. (2014). Dendrochronological reconstruction of the epicentre and early spread of emerald ash borer in North America. *Diversity and Distributions*, *20*(7), 847–858. https://doi.org/10.1111/ddi.12212

Srei, N., Guertin, C., Lavallée, R., Lajoie, M.-È., Brousseau, C., Bergevin, R., Miller, F., McMillin, K., & Trudel, R. (2020). Microbial Control of the Emerald Ash Borer (Coleoptera: Buprestidae) Using Beauveria bassiana (Hypocreales: Cordycipitaceae) by the Means of an Autodissemination Device. *Journal of Economic Entomology*, *113*(6), 2657–2665. https://doi.org/10.1093/jee/toaa226

Srinivasan, J., Dillman, A. R., Macchietto, M. G., Heikkinen, L., Lakso, M., Fracchia, K. M., Antoshechkin, I., Mortazavi, A., Wong, G., & Sternberg, P. W. (2013). The Draft Genome and Transcriptome of Panagrellus redivivus Are Shaped by the Harsh Demands of a Free-Living Lifestyle. *Genetics*, *193*(4), 1279–1295. https://doi.org/10.1534/genetics.112.148809

Stadnichuk, I. N., & Kusnetsov, V. V. (2021). Endosymbiotic Origin of Chloroplasts in Plant Cells' Evolution. *Russian Journal of Plant Physiology*, *68*(1), 1–16. https://doi.org/10.1134/S1021443721010179

Tagami, Y., & Miura, K. (2004). Distribution and prevalence of Wolbachia in Japanese populations of Lepidoptera. *Insect Molecular Biology*, *13*(4), 359–364. https://doi.org/10.1111/j.0962-1075.2004.00492.x

Tauber, C. A., Tauber, M. J., & Albuquerque, G. S. (2009). Chapter 181 - Neuroptera: (Lacewings, Antlions). In V. H. Resh & R. T. Cardé (Eds.), *Encyclopedia of Insects (Second Edition)* (pp. 695–707). Academic Press. https://doi.org/10.1016/B978-0-12-374144-8.00190-9

Taylor, R. A. J., Bauer, L. S., Poland, T. M., & Windell, K. N. (2010). Flight Performance of Agrilus planipennis (Coleoptera: Buprestidae) on a Flight Mill and in Free Flight. *Journal of Insect Behavior*, *23*(2), 128–148. https://doi.org/10.1007/s10905-010-9202-3

Thompson, J. N. (1999). The Evolution of Species Interactions. *Science*, *284*(5423), 2116–2118. https://doi.org/10.1126/science.284.5423.2116

Tluczek, A. R., Mccullough, D. G., & Poland, T. M. (2011). Influence of Host Stress on Emerald Ash Borer (Coleoptera: Buprestidae) Adult Density, Development, and Distribution in <I>Fraxinus pennsylvanica</I> Trees. *Environmental Entomology*, *40*(2), 357–366. https://doi.org/10.1603/EN10219

Tripp, E. A., Zhang, N., Schneider, H., Huang, Y., Mueller, G. M., Hu, Z., Häggblom, M., & Bhattacharya, D. (2017). Reshaping Darwin's Tree: Impact of the Symbiome. *Trends in Ecology & Evolution*, *32*(8), 552–555. https://doi.org/10.1016/j.tree.2017.05.002

Uchida, N., Kubota, K., Aita, S., & Kazama, S. (2020). Aquatic insect community structure revealed by eDNA metabarcoding derives indices for environmental assessment. *PeerJ*, *8*, e9176. https://doi.org/10.7717/peerj.9176

Ulyshen, M. D., Klooster, W. S., Barrington, W. T., & Herns, D. A. (2011). Impacts of emerald ash borer-induced tree mortality on leaf litter arthropods and exotic earthworms. *Pedobiologia-International Journal of Soil Biology 54:261-265*, *54*, 261–265. https://doi.org/10.1016/j.pedobi.2011.05.001

United States Department of Agriculture. (2021, July 29). *The State of the Forest*. United States Department of Agriculture. https://www.usda.gov/media/blog/2019/04/22/state-forest

United States Department of Agriculture. (2023, June 15). *Emerald Ash Borer*. Aphis.Usda.Gov. https://www.aphis.usda.gov/aphis/ourfocus/planthealth/plant-pest-and-disease-programs/pests-and-diseases/emerald-ash-borer

Valentin, R. E., Fonseca, D. M., Gable, S., Kyle, K. E., Hamilton, G. C., Nielsen, A. L., & Lockwood, J. L. (2020). Moving eDNA surveys onto land: Strategies for active eDNA aggregation to detect invasive forest insects. *Molecular Ecology Resources*, *20*(3), 746–755. https://doi.org/10.1111/1755-0998.13151

Wang, T., Zhang, Y., Yang, Z., Liu, Z., & Du, Y. (2020). DNA barcoding reveals cryptic diversity in the underestimated genus Triplophysa (Cypriniformes: Cobitidae,

Nemacheilinae) from the northeastern Qinghai-Tibet Plateau. *BMC Evolutionary Biology*, *20*(1), 151. https://doi.org/10.1186/s12862-020-01718-0

Wang, X. Y., Yang, Z. Q., Gould, J. R., Zhang, Y. N., Liu, G. J., & Liu, E. shan. (2010). The biology and ecology of the emerald ash borer, Agrilus planipennis, in China. *Journal of Insect Science (Online)*, *10*, 128. https://doi.org/10.1673/031.010.12801

Wei, X., Reardon, D., Wu, Y., & Sun, J. (2004). Emerald ash borer, Agrilus planipennis Fairmaire (Coleoptera: Buprestidae), in China: a review and distribution survey. *Acta Entomologica Sinica*, *47*, 679–685.

Weigand, H., Beermann, A. J., Čiampor, F., Costa, F. O., Csabai, Z., Duarte, S., Geiger, M. F., Grabowski, M., Rimet, F., Rulik, B., Strand, M., Szucsich, N., Weigand, A. M., Willassen, E., Wyler, S. A., Bouchez, A., Borja, A., Čiamporová-Zaťovičová, Z., Ferreira, S., … Ekrem, T. (2019). DNA barcode reference libraries for the monitoring of aquatic biota in Europe: Gap-analysis and recommendations for future work. *Science of The Total Environment*, *678*, 499–524. https://doi.org/10.1016/j.scitotenv.2019.04.247

Werren, J. H. (1997). Biology of *Wolbachia*. *Annual Review of Entomology*, *42*(1), 587–609. https://doi.org/10.1146/annurev.ento.42.1.587

White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and Direct Sequencing of Fungal Ribosomal RNA Genes for Phylogenetics. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, & T. J. White (Eds.), *PCR Protocols: A Guide to Methods and Applications* (pp. 315–322). Academic Press. https://doi.org/10.1016/B978-0-12-372180-8.50042-1

Whitehorn, P. R., O'Connor, S., Wackers, F. L., & Goulson, D. (2012). Neonicotinoid Pesticide Reduces Bumble Bee Colony Growth and Queen Production. *Science*, *336*(6079), 351–352. https://doi.org/10.1126/science.1215025

Wood, T. J., & Goulson, D. (2017). The environmental risks of neonicotinoid pesticides: A review of the evidence post 2013. *Environmental Science and Pollution Research International*, *24*(21), 17285–17325. https://doi.org/10.1007/s11356-017-9240-x

Wootton, J. T. (1994). The Nature and Consequences of Indirect Effects in Ecological Communities. *Annual Review of Ecology and Systematics*, *25*, 443–466.

Work, T. T., & McCullough, D. G. (2000). Lepidopteran Communities in Two Forest Ecosystems During the First Gypsy Moth Outbreaks in Northern Michigan. *Environmental Entomology*, *29*(5), 884–900. https://doi.org/10.1603/0046-225X-29.5.884

Yu, Y., Yang, Q., Petropoulos, E., & Zhu, T. (2022). ITS3/ITS4 outperforms other ITS region and 18S rRNA gene primer sets for amplicon sequencing of soil fungi. *European Journal of Soil Science*, *73*(6), e13329. https://doi.org/10.1111/ejss.13329

Zabalou, S., Riegler, M., Theodorakopoulou, M., Stauffer, C., Savakis, C., & Bourtzis, K. (2004). Wolbachia-induced cytoplasmic incompatibility as a means for insect pest population control. *Proceedings of the National Academy of Sciences*, *101*(42), 15042–15045. https://doi.org/10.1073/pnas.0403853101

Zeale, M. R. K., Butlin, R. K., Barker, G. L. A., Lees, D. C., & Jones, G. (2011). Taxon-specific PCR for DNA barcoding arthropod prey in bat faeces. *Molecular Ecology Resources*, *11*(2), 236–244. https://doi.org/10.1111/j.1755-0998.2010.02920.x

Zieman, E. A., Reeve, J. D., Braswell, W. E., & Jiménez, F. A. (2015). Pathology, distribution, morphological and genetic identity of *Deladenus proximus* (Tylenchida: Neotylenchidae) a parasitic nematode of the woodwasp, *Sirex nigricornis* in the eastern United States. *Biological Control*, *87*, 14–22. https://doi.org/10.1016/j.biocontrol.2015.04.009

Appendix A: Supplemental Tables

Emerald Ash Borer Specimen Collection Tables

Table S1. Emerald ash borer life stages. A total of 277 emerald ash borer specimens were collected, including one specimen of an unknown life stage (not shown).

Life Stage	Tissue Sampled	Specimen Count
Larva1 (<0.5in)	Whole specimen	54
Larva2 (0.5-1in)	~6mm fragment of posterior	68
Larva3 (>1in)	~6mm fragment of posterior	41
Pupa	Posterior third	53
Adult	Posterior half of the abdomen	60

Table S2. Emerald ash borer collection locations. A total of 277 emerald ash borer specimens were collected across southern Ontario, Canada.

Region	Sector	Specimen Count
Brockville	Flood	20
	Sones	107
	Tincap	58
Ennotville	Rosedale	41
Guelph	Riverside	41
Puslinch	Hebert	10

Table S3. Emerald ash borer collection habitats. Emerald ash borer specimens were collected from a variety of habitats and subhabitats in southern Ontario, Canada.

Habitat	Subhabitat	Specimen Count
Artificial	Arable	119
	Urban	51
Woodland	Forest	107

Spongy Moth Interaction Tables

Table S4. All taxa identified using the ITS target locus across all spongy moth samples. The frequency of taxa identified in spongy moth samples (9 each of egg and larva) across the egg and larva life stages using the ITS target locus.

Kingdom	Phylum	Class	Order	Family	Genus	Egg Frequency	Larva Frequency
Fungi	Ascomycota	Dothideomycetes	Pleosporales	Didymosphaeriaceae	*Paraconiothyrium*	1	0
				Leptosphaeriaceae	*Leptosphaeria*	0	1
				Lophiostomataceae	*Lophiostoma*	0	1
					(blank)	1	0
				Pleosporaceae	*Alternaria*	6	6
					Stemphylium	2	2
				(blank)	*(blank)*	0	2
			(blank)	(blank)	*(blank)*	2	2
		Eurotiomycetes	Onygenales	Ascosphaeraceae	*Ascosphaera*	0	1
		Lecanoromycetes	Lecanorales	Parmeliaceae	*Punctelia*	3	5
		Saccharomycetes	Saccharomycetales	Pichiaceae	*Pichia*	0	2
		(blank)	(blank)	(blank)	*(blank)*	3	6
	Basidiomycota	Agaricomycetes	Polyporales	Fomitopsidaceae	*Fomitopsis*	1	0
		Microbotryomycetes	Sporidiobolales	Sporidiobolaceae	*Sporobolomyces*	0	1
	(blank)	(blank)	(blank)	(blank)	*(blank)*	5	8
Plantae	Chlorophyta	Trebouxiophyceae	Trebouxiales	Trebouxiaceae	*Asterochloris*	0	1
					Symbiochloris	0	1
					Trebouxia	2	6
			Unclassified	Unclassified	*Coccomyxa*	0	1
	Tracheophyta	Liliopsida	Poales	Poaceae	*Bromus*	0	1
					Lolium	1	0
					Poa	1	1
					(blank)	1	0
		Magnoliopsida	Asterales	Asteraceae	*Centaurea*	0	1
					Coreopsis	1	1

				Erigeron	0	1
				Sonchus	1	1
				(blank)	1	1
		Fabales	Fabaceae	*Lotus*	2	1
				Trifolium	1	2
				Vicia	1	1
		Fagales	Betulaceae	*Betula*	3	5
			Fagaceae	*Quercus*	1	0
		Malpighiales	Salicaceae	*Populus*	1	0
				Salix	2	0
		Myrtales	Onagraceae	*Chamaenerion*	1	0
				Circaea	0	1
				(blank)	1	0
		Ranunculales	Ranunculaceae	*Ranunculus*	0	1
		Rosales	Cannabaceae	*Cannabis*	0	1
			Rosaceae	*Potentilla*	1	1
				Prunus	1	0
				Rosa	1	2
				Sorbaria	1	0
		Saxifragales	Paeoniaceae	*Paeonia*	1	2
		Solanales	Solanaceae	*Calibrachoa*	1	1
	Pinopsida	Cupressales	Cupressaceae	*Juniperus*	0	1
				Thuja	1	1
(blank)	(blank)	(blank)	(blank)	*(blank)*	1	4

Table S5. All taxa identified using the 16S target locus across all spongy moth samples. The frequency of taxa identified in spongy moth samples (9 each of egg and larva) across the egg and larva life stages using the 16S target locus.

Kingdom	Phylum	Class	Order	Family	Genus	Egg Frequency	Larva Frequency
Bacteria	Abditibacteriota	Abditibacteriia	Abditibacteriales	Abditibacteriaceae	*Abditibacterium*	7	6
	Acidobacteriota	Terriglobia	Terriglobales	Acidobacteriaceae	*Bryocella*	1	0
					Granulicella	4	2
					Terriglobus	1	0
	Actinomycetota	Acidimicrobiia	Acidimicrobiales	Iamiaceae	*Aquihabitans*	0	1
		Actinomycetes	Propionibacteriales	Nocardioidaceae	*Marmoricola*	1	1
			Jatrophihabitantales	Jatrophihabitantaceae	*Jatrophihabitans*	3	0
			Kineosporiales	Kineosporiaceae	*Kineosporia*	0	1
			Nakamurellales	Nakamurellaceae	*Nakamurella*	1	1
			Streptosporangiales	Treboniaceae	*Trebonia*	1	0
			Mycobacteriales	Nocardiaceae	*Williamsia*	1	0
		Thermoleophilia	Solirubrobacterales	Patulibacteraceae	*Patulibacter*	1	0
				Baekduiaceae	*Baekduia*	1	1
	Armatimonadota	Armatimonadia	Capsulimonadales	Capsulimonadaceae	*Capsulimonas*	2	4
		Fimbriimonadia	Fimbriimonadales	Fimbriimonadaceae	*Fimbriimonas*	3	4
	Bacteroidota	Chitinophagia	Chitinophagales	Chitinophagaceae	*Segetibacter*	0	1
		Cytophagia	Cytophagales	Cytophagaceae	*Spirosoma*	2	2
		Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	*Pedobacter*	1	1
	Candidatus Saccharibacteria	Candidatus Saccharimonadia	Candidatus Saccharimonadales	Candidatus Saccharimonadaceae	*Saccharimonas*	1	0
	Myxococcota	Polyangia	Polyangiales	Labilitrichaceae	*Labilithrix*	1	0
	Pseudomonadota	Alphaproteobacteria	Hyphomicrobiales	Beijerinckiaceae	*Methylovirgula*	1	0
				Lichenihabitantaceae	*Lichenihabitans*	7	7
				Methylobacteriaceae	*Enterovirga*	0	1
					Psychroglaciecola	5	3
				Nitrobacteraceae	*Tardiphaga*	2	1
				Aurantimonadaceae	*Aureimonas*	1	1
				Phyllobacteriaceae	*Mesorhizobium*	0	1

				Rhizobiaceae	*Neorhizobium*	1	1
			Rhodospirillales	Acetobacteraceae	*Roseomonas*	5	1
			Rickettsiales	Anaplasmataceae	*Wolbachia*	1	4
			Sphingomonadales	Sphingomonadaceae	*Sphingomicrobium*	0	1
			Geminicoccales	Geminicoccaceae	*Geminicoccus*	0	1
		Betaproteobacteria	Burkholderiales	Oxalobacteraceae	*Massilia*	3	5
				Burkholderiaceae	*Caballeronia*	0	1
				Burkholderiales genera incertae sedis	*Rhizobacter*	2	3
		Gammaproteobacteria	Oceanospirillales	Oceanospirillaceae	*Marinomonas*	0	1
			Aeromonadales	Aeromonadaceae	*Pseudaeromonas*	1	0
					Tolumonas	2	0
	Bacillota	Clostridia	Eubacteriales	Oscillospiraceae	*Ruminiclostridium*	0	1
	Gemmatimonadota	Gemmatimonadia	Gemmatimonadales	Gemmatimonadaceae	*Gemmatirosa*	0	1
	Deinococcota	Deinococci	Trueperales	Trueperaceae	*Truepera*	1	1

Table S6. All taxa identified using the COI target locus across all spongy moth samples. The frequency of taxa identified in spongy moth samples (9 each of egg and larva) across the egg and larva life stages using the COI target locus.

Kingdom	Phylum	Class	Order	Family	Genus	Egg Frequency	Larva Frequency
Animalia	Arthropoda	Insecta	Blattodea	Blattellidae	*Blattella*	1	0
			Coleoptera	Nitidulidae	*Epuraea*	0	1
			Diptera	Cecidomyiidae	*Lestodiplosis*	0	1
			Hemiptera	Adelgidae	*Pineus*	0	1
				Miridae	*Lygus*	2	4
			Lepidoptera	Erebidae	*(blank)*	1	0
				Geometridae	*(blank)*	0	1
			Neuroptera	Hemerobiidae	*Hemerobius*	2	1

Table S7. All taxa identified using the rbcL target locus across all spongy moth samples. The frequency of taxa identified in spongy moth samples (9 each of egg and larva) across the egg and larva life stages using the rbcL target locus.

Kingdom	Phylum	Class	Order	Family	Genus	Egg Frequency	Larva Frequency
Plantae	Chlorophyta	Trebouxiophyceae	Unclassified	Unclassified	*Coccomyxa*	1	1
	Streptophyta	(blank)	(blank)	(blank)	*(blank)*	1	4
	Tracheophyta	Liliopsida	Poales	Poaceae	*(blank)*	1	1
			Zingiberales	Musaceae	*(blank)*	0	1
		Magnoliopsida	Asterales	Asteraceae	*(blank)*	2	3
			Caryophyllales	Caryophyllaceae	*(blank)*	0	1
			Dipsacales	Adoxaceae	*Sambucus*	2	1
			Fabales	Fabaceae	*Medicago*	0	1
					Securigera	0	1
					Trifolium	0	1
					(blank)	0	2
			Fagales	Betulaceae	*(blank)*	5	5
				Fagaceae	*Fagus*	2	2
					(blank)	1	0
			Laurales	Lauraceae	*(blank)*	2	0
			Malpighiales	Salicaceae	*Populus*	4	3
					Salix	0	2
			Rosales	Cannabaceae	*Cannabis*	0	1
				Rosaceae	*Rosa*	2	0
				Ulmaceae	*Ulmus*	1	0
			Sapindales	Sapindaceae	*Acer*	2	6
					(blank)	1	0
			Vitales	Vitaceae	*(blank)*	3	0
			(blank)	(blank)	*(blank)*	2	2
		Pinopsida	Cupressales	Cupressaceae	*(blank)*	0	2
			Pinales	Pinaceae	*Picea*	5	5
					Pinus	3	2

Emerald Ash Borer Interaction Tables

Table S8. All taxa identified using the ITS target locus across all emerald ash borer samples. The frequency of taxa identified in emerald ash borer samples across all life stages (3 sizes of larva, pupa, and adult) using the ITS target locus.

Kingdom	Phylum	Class	Order	Family	Genus	Larva1 Freq.	Larva2 Freq.	Larva3 Freq.	Pupa Freq.	Adult Freq.
Fungi	Ascomycota	Candelariomycetes	Candelariales	Candelariaceae	*Candelaria*	0	2	0	0	0
		Dothideomycetes	Botryosphaeriales	Botryosphaeriaceae	*(blank)*	1	1	0	1	0
			Pleosporales	Cucurbitariaceae	*Parafenestella*	0	1	0	0	0
					Pyrenochaeta	1	4	0	0	1
				Pleosporaceae	*Alternaria*	0	0	0	0	2
		Eurotiomycetes	Eurotiales	(blank)	*(blank)*	0	0	0	0	1
		Lecanoromycetes	(blank)	(blank)	*(blank)*	0	1	0	0	0
		Leotiomycetes	Helotiales	Dermateaceae	*Neofabraea*	1	4	1	9	5
				Helotiales incertae sedis	*Cadophora*	0	0	0	1	0
				(blank)	*(blank)*	0	0	0	1	0
			Leotiales	Tympanidaceae	*Tympanis*	0	0	0	0	1
		Orbiliomycetes	Orbiliales	Orbiliaceae	*(blank)*	0	0	0	2	0
		Saccharomycetes	Saccharomycetales	Debaryomycetaceae	*Yamadazyma*	0	0	0	0	2
		Sordariomycetes	Diaporthales	Diaporthaceae	*Diaporthe*	0	2	0	1	0
				Valsaceae	*Valsa*	2	4	3	7	2
				(blank)	*(blank)*	0	1	0	0	0
			Hypocreales	Cordycipitaceae	*Beauveria*	0	1	0	0	1
				Hypocreales incertae sedis	*Eucasphaeria*	0	2	0	6	4
				Nectriaceae	*Fusarium*	3	2	1	1	3
				Niessliaceae	*Niesslia*	0	0	0	1	0
			Togniniales	Togniniaceae	*Phaeoacremonium*	0	2	0	6	4
			(blank)	(blank)	*(blank)*	0	0	1	0	0
		(blank)	(blank)	(blank)	*(blank)*	3	3	0	2	4

Table S9. All taxa identified using the 16S target locus across all emerald ash borer samples. The frequency of taxa identified in emerald ash borer samples across all life stages (3 sizes of larva, pupa, and adult) using the 16S target locus.

Kingdom	Phylum	Class	Order	Family	Genus	Larva1 Freq.	Larva2 Freq,	Larva3 Freq,	Pupa Freq,	Adult Freq,
Bacteria	Acidobacteriota	Terriglobia	Terriglobales	Acidobacteriaceae	*Terriglobus*	0	0	0	0	3
					Edaphobacter	0	0	0	0	1
					Granulicella	0	0	0	0	1
	Bacteroidota	Chitinophagia	Chitinophagales	Chitinophagaceae	*Ginsengibacter*	0	0	0	1	0
		Flavobacteriia	Flavobacteriales	Crocinitomicaceae	*Taishania*	0	0	0	2	0
		Cytophagia	Cytophagales	Fulvivirgaceae	*Chryseolinea*	0	0	0	1	1
		Sphingobacteriia	Sphingobacteriales	Sphingomonadaceae	*Sphingobium*	1	0	0	0	0
				Sphingobacteriaceae	*Olivibacter*	4	3	1	4	6
					Pedobacter	1	0	0	0	3
	Bdellovibrionota	Oligoflexia	Oligoflexales	Oligoflexaceae	*Oligoflexus*	0	0	0	1	0
	Myxococcota	Polyangia	Polyangiales	Labilitrichaceae	*Labilithrix*	0	0	0	0	1
	Pseudomonadota	Alphaproteobacteria	Hyphomicrobiales	Beijerinckiaceae	*Methylovirgula*	0	0	0	0	2
				Nitrobacteraceae	*Tardiphaga*	1	1	0	1	4
				Rhizobiaceae	*Neorhizobium*	7	17	1	18	19
				Aurantimonadaceae	*Aureimonas*	0	0	0	0	0
				Methylopilaceae	*Hansschlegelia*	0	2	1	1	1
			Sphingomonadales	Sphingomonadaceae	*Allosphingosinicella*	0	0	0	1	0
					Novosphingobium	1	0	0	0	1
			Rickettsiales	Anaplasmataceae	*Wolbachia*	0	0	0	1	0
		Betaproteobacteria	Burkholderiales	Oxalobacteraceae	*Duganella*	0	1	0	0	1
		Gammaproteobacteria	Enterobacterales	Pectobacteriaceae	*Acerihabitans*	1	0	0	0	3
			Lysobacterales	Rhodanobacteraceae	*Dokdonella*	1	0	0	0	1

Table S10. All taxa identified using the COI target locus across all emerald ash borer samples. The frequency of taxa identified in emerald ash borer samples across all life stages (3 sizes of larva, pupa, and adult) using the COI target locus.

Kingdom	Phylum	Class	Order	Family	Larva1 Freq.	Larva2 Freq.	Larva3 Freq.	Pupa Freq.	Adult Freq.
Animalia	Arthropoda	Arachnida	Sarcoptiformes	Oppiidae	0	0	0	0	1
			Trombidiformes	Eupodidae	0	0	0	0	1
		Insecta	Diptera	Cecidomyiidae	0	1	0	0	0
			Hymenoptera	Ichneumonidae	0	1	1	1	0

Table S11. All taxa identified using the 18S target locus across all emerald ash borer samples. The frequency of taxa identified in emerald ash borer samples across all life stages (3 sizes of larva, pupa, and adult) using the 18S target locus.

Kingdom	Phylum	Class	Order	Family	Genus	Larva1 Freq.	Larva2 Freq.	Larva3 Freq.	Pupa Freq.	Adult Freq.
Animalia	Arthropoda	Insecta	Hymenoptera	(blank)	*(blank)*	2	1	0	4	0
	Nematoda	Chromadorea	Rhabditida	Diplogasteridae	*Rhabditolaimus*	1	1	0	0	0
				Neotylenchidae	*(blank)*	0	3	1	0	1
				Panagrolaimidae	*Panagrellus*	0	1	0	0	0
				(blank)	*(blank)*	0	0	0	0	1
Fungi	Ascomycota	Dothideomycetes	Botryosphaeriales	Botryosphaeriaceae	*(blank)*	0	0	0	1	0
			Pleosporales	(blank)	*(blank)*	0	1	0	0	0
		Leotiomycetes	Helotiales	(blank)	*(blank)*	0	1	0	1	0
			(blank)	(blank)	*(blank)*	1	0	0	0	0
		Saccharomycetes	Saccharomycetales	(blank)	*(blank)*	0	0	0	0	1
		Sordariomycetes	Diaporthales	Valsaceae	*Diaporthe*	0	2	0	0	0
			Hypocreales	Cordycipitaceae	*(blank)*	0	0	0	0	1
				Nectriaceae	*(blank)*	1	0	0	0	0
				Niessliaceae	*Niesslia*	0	1	0	0	1
				(blank)	*(blank)*	0	1	0	0	0
		(blank)	(blank)	(blank)	*(blank)*	1	0	0	0	0
Unknown	Apicomplexa	Conoidasida	Cryptosporida	Cryptosporidiidae	*Cryptosporidium*	0	0	0	1	0

Appendix B: R Code for ITS and rbcL PCR File Processing

```
# Load the packages
library(dada2); packageVersion("dada2")
library(ShortRead); packageVersion("ShortRead")
library(Biostrings); packageVersion("Biostrings")
library(beepr)

### Establish the directory to work
path <- "~/Desktop/folder" # Only use one path because with two the pipeline does not work

### We receive both trimmed and raw versions of the sequences. Give the raw files a go with this pipeline

## Check that the files can be found in the input directory
list.files(path)

### Input the forward and reverse reads
## In the example the files were named with an ending "_L001_R1_001.fastq" for forward reads and names with an ending "_L001_R2_001.fastq" for reverse reads
fnFs <- sort(list.files(path, pattern = "_L001_R1_001.fastq", full.names = TRUE))
fnRs <- sort(list.files(path, pattern = "_L001_R2_001.fastq", full.names = TRUE))

### Input your Forward and Reverse primers

##rbcL primers used as an example; replace as needed
FWD <- "TTGGCAGCATTYCGAGTAACTCC"
REV <- "AACCYTCTTCAAAAAGGTC"

### Primer orientation checking
## Create all orientations of the input sequence
allOrients <- function(primer) {
  require(Biostrings)
  dna <- DNAString(primer)  # The Biostrings works w/ DNAString objects rather than character vectors
  orients <- c(Forward = dna, Complement = complement(dna), Reverse = reverse(dna),
               RevComp = reverseComplement(dna))
  return(sapply(orients, toString))  # Convert back to character vector
}
```

```
FWD.orients <- allOrients(FWD)
REV.orients <- allOrients(REV)

## Pre-filter any reads with Ns
# In the end if you don't find what you are looking for, you should think about changing this step, might remove too much of your data
fnFs.filtN <- file.path(path, "filtN", basename(fnFs))
fnRs.filtN <- file.path(path, "filtN", basename(fnRs))

# maxN = 0 is to establish that sequences with more than 0 Ns will be discarded. Multithread makes us of multiple CPU cores. Use the beepR package to get an alert when it is done. It takes a while!
filterAndTrim(fnFs, fnFs.filtN, fnRs, fnRs.filtN, maxN = 0, multithread = FALSE,matchIDs = TRUE); beep(3) # Downgrade to R 4.1.2

### Check for primer hits in the data
primerHits <- function(primer, fn) {
  # Counts number of reads in which the primer is found
  nhits <- vcountPattern(primer, sread(readFastq(fn)), fixed = FALSE)
  return(sum(nhits > 0))
}

rbind(FWD.ForwardReads = sapply(FWD.orients, primerHits, fn = fnFs.filtN[[1]]),
      FWD.ReverseReads = sapply(FWD.orients, primerHits, fn = fnRs.filtN[[1]]),
      REV.ForwardReads = sapply(REV.orients, primerHits, fn = fnFs.filtN[[1]]),
      REV.ReverseReads = sapply(REV.orients, primerHits, fn = fnRs.filtN[[1]]))

### Cutadapt section
## Install cutadapt using Miniconda and Bioconda
# Use the path.expand method
cutadapt <- path.expand("~/opt/miniconda3/bin/cutadapt") # Remember to change the system path to the cutadapt installed file in Miniconda

# Make sure it is working
system2(cutadapt, args = "--version") # The system2 command allows to run a shell/Terminal (including Python) from within R

## Trim primers
path.cut <- file.path(path, "cutadapt")
```

```
if(!dir.exists(path.cut)) dir.create(path.cut)

fnFs.cut <- file.path(path.cut, basename(fnFs))
fnRs.cut <- file.path(path.cut, basename(fnRs))

FWD.RC <- dada2:::rc(FWD)  # The ":::" symbol allows to recall an specific function or variable from within a package
REV.RC <- dada2:::rc(REV)

# Trim FWD and the reverse-complement of REV off of R1 (forward reads)
R1.flags <- paste("-g", FWD, "-a", REV.RC)

# Trim REV and the reverse-complement of FWD off of R2 (reverse reads)
R2.flags <- paste("-G", REV, "-A", FWD.RC)

## Run cutadapt
# You can use the "--discard-untrimmed" argument to discard any read where the primers have not been trimmed off. Useful for dealing with datasets containing multiple primersets
# You can use the "--minimum-length X" argument to discard reads shorter than Xbp. This will remove unexpected short reads and help speed up further analysis.
for(i in seq_along(fnFs)) {
  system2(cutadapt, args = c(R1.flags, R2.flags, "-n", 2, # -n 2 required to remove FWD and REV
                             "-m", 10, # low minimum length - cutadapt sometimes produces empty seqs
                             "-j", 0, # autodetect cores
                             "-o", fnFs.cut[i], "-p", fnRs.cut[i], # output files
                             fnFs.filtN[i], fnRs.filtN[i])) # input files
}; #beep(10)

## Sanity check to assess if the primers have been removed
rbind(FWD.ForwardReads = sapply(FWD.orients, primerHits, fn = fnFs.cut[[1]]),
      FWD.ReverseReads = sapply(FWD.orients, primerHits, fn = fnRs.cut[[1]]),
      REV.ForwardReads = sapply(REV.orients, primerHits, fn = fnFs.cut[[1]]),
      REV.ReverseReads = sapply(REV.orients, primerHits, fn = fnRs.cut[[1]]))

## Plot read quality profiles - these are the files that have been cutadapt-ed
plotQualityProfile(fnFs.cut[1:1])
```

```
plotQualityProfile(fnRs.cut[1:1])

## Specify the paths and file names of the cutadapt processed files (forward and reverse primer cleaned files). Pattern of the files following the example from the first steps
cutFs <- sort(list.files(path.cut, pattern = "_L001_R1_001.fastq", full.names = TRUE))
cutRs <- sort(list.files(path.cut, pattern = "_L001_R2_001.fastq", full.names = TRUE))

## Extract sample names
get.sample.name <- function(fname) strsplit(basename(fname), "_")[[1]][2]
sample.names <- unname(sapply(cutFs, get.sample.name))
head(sample.names)
list(sample.names)

## Place filtered files in filtered/ subdirectory
filtFs <- file.path(path, "filtered", paste0(sample.names, "_F_filt.fastq.gz"))
filtRs <- file.path(path, "filtered", paste0(sample.names, "_R_filt.fastq.gz"))

### Run filterAndTrim again for a further round of read cleaning (to remove quality dropouts at the end of the reads)
out <- filterAndTrim(cutFs, filtFs, cutRs, filtRs, maxN = 50,
                     truncQ = 2, truncLen=c(180, 180),
                     maxEE = c(2, 2), rm.phix = TRUE,
                     compress = TRUE, multithread = TRUE); #beep(10)

## Plot the Q scores of trimmed data generated above - better than first Q plot?
plotQualityProfile(paste(path, "filtered", list.files(paste(path, "filtered", sep="/"))[1:2],sep="/"))

## Check number of reads surviving filtering
system(paste("zgrep -c \'@M00\' ",path,"/filtered/*F*",sep=""))

## Check tables to see if Falses are present. Falses are the amount of broken samples
table(file.exists(filtFs))
table(file.exists(filtRs))

## Check tables again to see if Falses are present. If only True, continue!
#table(file.exists(filtFsno0s))
#table(file.exists(filtRsno0s))
```

```
### Learn the error rates
errF <- learnErrors(filtFs, multithread = TRUE);# beep(10)
errR <- learnErrors(filtRs, multithread = TRUE); #beep(10)

## Check convergence rates. Ideally the last number from left to right gets down or very close to 0
dada2:::checkConvergence(errF)
dada2:::checkConvergence(errR)

## Plot the estimated error rates
# Do not worry about the warning messages: 1: Transformation introduced infinite values in continuous y-axis 2: Transformation introduced infinite values in continuous y-axis
# Black dots (observed values) should follow the black line (expected values). Red line represents the expected error rate base on the quality score.
plotErrors(errF, nominalQ = TRUE)
plotErrors(errR, nominalQ = TRUE)

### Dereplication step
exists <- file.exists(filtFs)

# Check that all the samples are still present after filtering
derepFs <- derepFastq(filtFs[exists], verbose=TRUE); beep(10)
derepRs <- derepFastq(filtRs[exists], verbose=TRUE); beep(10)

# Name the derep-class objects by the sample names
names(derepFs) <- sample.names[exists]
names(derepRs) <- sample.names[exists]

### ASV inference with the dada2 algorithm.
## It is also possible to run samples together in pseudo-pools to increase the ability to identify ASVs of low abundances and for them to be retained in the analysis. Check https://benjjneb.github.io/dada2/pseudo.html#Pseudo-pooling for more info.
dadaFs <- dada(derepFs, err = errF, multithread = TRUE); beep(10)
dadaRs <- dada(derepRs, err = errR, multithread = TRUE); beep(10)

### Merging paired end reads
# Default minimum overlap between the two samples is 12 bp
mergers <- mergePairs(dadaFs, derepFs, dadaRs, derepRs, verbose=TRUE)
```

```
### Making the ASV matrix
seqtab <- makeSequenceTable(mergers)
dim(seqtab)

## Inspect distribution of sequence lengths
table(nchar(getSequences(seqtab)))

### Chimaera detection and removal - where sequences from different animals get put together
seqtab.nochim <- removeBimeraDenovo(seqtab, method="consensus", multithread=TRUE, verbose=TRUE)

dim(seqtab.nochim)

# Check the data proportions
sum(seqtab.nochim)/sum(seqtab)

# Re-check the range of ASVs lengths
table(nchar(getSequences(seqtab.nochim)))

### Sequence tracking sanity check
getN <- function(x) sum(getUniques(x))

track <- cbind(out, sapply(dadaFs, getN),
               sapply(dadaRs, getN),
               sapply(mergers, getN),
               rowSums(seqtab.nochim))

colnames(track) <- c("input", "filtered", "denoisedF", "denoisedR", "merged", "nonchim")
# denoised is a fancy way of saying dereplicating

rownames(track) <- sample.names

track
write.table(track, "~/Desktop/folder/track.tsv", sep="\t", quote=F, col.names=NA)

### Exporting data to BLAST+
# The column names of seqtab.nochim are actually the ASV sequences,
# Extract these and assign them to seqs vars
seqs <- colnames(seqtab.nochim)
```

```
## Make a new variable for ASV names
# With length equal to the number of ASVs
headers <- vector(dim(seqtab.nochim)[2], mode="character")

## Fill the vector with names formatted for a fasta header (>ASV_1, >ASV_2, etc.)
for (i in 1:dim(seqtab.nochim)[2]) {
  headers[i] <- paste(">ASV", i, sep="_")
}

## Join together to make a FASTA file
# First create a new folder called "dada2_output" in the parent path specified at the
beginning
fasta <- c(rbind(headers, seqs))
write(fasta, "~/Desktop/folder/ASVs.fasta")

## Make a table of sequence counts for each sample and ASV.
# First transpose the `seqtab.nochim` and assign this to the variable `mifish_tab`
tab <- t(seqtab.nochim)

# Name each row with the ASV name, omitting the '>' used in the fasta file
row.names(tab) <- sub(">", "", headers)
write.table(tab, "~/Desktop/folder/table_ASVs.tsv", sep="\t", quote=F, col.names=NA)

### STOP HERE AND GO TO BLAST.COM ###
```

Appendix C: QIIME2 Code for 16S PCR File Processing

```
### Activate conda to use QIIME2

cd ~
conda activate qiime2-2023.7

###Import data

qiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' --input-path data/testData.csv --output-path LDD16SAnalysis/paired-end-demux.qza --input-format PairedEndFastqManifestPhred33

###Summarizing data

qiime demux summarize --i-data LDD16SAnalysis/paired-end-demux.qza --o-visualization LDD16SAnalysis/data-demux-summary.qzv

### Optional: View data. If the code does not work, go to https://view.qiime2.org/

qiime tools view data-demux-summary.qzv

### Change directory

cd LDD16SAnalysis

### Trim data with cutadapt; adjust primer sequences in the code as needed

qiime cutadapt trim-paired \
--i-demultiplexed-sequences paired-end-demux.qza \
--p-front-f ACACTGACGACATGGTTCTACACMGGATTAGATACCCKGG \
--p-front-r TACGGTAGCAGAGACTTGGTCTAGGGTTGCGCTCGTTG \
--p-match-adapter-wildcards \
--p-discard-untrimmed \
--o-trimmed-sequences cut_pairedend_demux.qza

### Summarize data

qiime demux summarize --i-data cut_pairedend_demux.qza --o-visualization cut-data-demux-summary.qzv
```

```
### View data. If the code does not work, go to https://view.qiime2.org/

qiime tools view cut-data-demux-summary.qzv

### Denoise with DADA2. Adjust trim/trunc values as needed based on the data viewed in the previous step. Adjust output file names in last two lines as needed

qiime dada2 denoise-paired \
  --p-n-threads 8 \
  --i-demultiplexed-seqs cut_pairedend_demux.qza \
  --p-trim-left-f 102 \
  --p-trim-left-r 57 \
  --p-trunc-len-f 271 \
  --p-trunc-len-r 202 \
  --o-table cut-16s-table.qza \
  --o-representative-sequences cut-16s-rep-seqs.qza \
  --o-denoising-stats cut-16s-denoising-stats.qza

### Summarize feature table

qiime feature-table summarize \
  --i-table cut-16s-table.qza \
  --o-visualization cut-16s-table.qza \
  --m-sample-metadata-file Metadata_16s.txt

qiime feature-table tabulate-seqs \
  --i-data cut-16s-rep-seqs.qza \
  --o-visualization cut-16s-rep-seqs.qzv

qiime metadata tabulate \
  --m-input-file cut-16s-denoising-stats.qza \
  --o-visualization cut-16s-denoising-stats.qzv

### Phylogeny

qiime phylogeny align-to-tree-mafft-fasttree \
  --i-sequences cut-16s-rep-seqs.qza \
  --o-alignment aligned-rep-seqs.qza \
  --o-masked-alignment masked-aligned-rep-seqs.qza \
```

```
  --o-tree unrooted-tree.qza \
  --o-rooted-tree rooted-tree.qza

### Taxonomy; download classifier at https://docs.qiime2.org/2023.9/data-resources/

qiime feature-classifier classify-sklearn \
  --i-classifier gg_2022_10_backbone_full_length.nb.qza \
  --i-reads cut-16s-rep-seqs.qza \
  --o-classification taxonomy.qza

qiime metadata tabulate \
  --m-input-file taxonomy.qza \
  --o-visualization taxonomy.qzv

### Create and export biome file

qiime tools export --input-path cut-16s-table.qza --output-path exported

qiime tools export --input-path taxonomy.qza --output-path exported

qiime tools export --input-path cut-16s-rep-seqs.qza --output-path exported

cp exported/taxonomy.tsv biom-taxonomy.tsv #change the header to #OTUID, taxonomy, confidence

biom add-metadata -i exported/feature-table.biom -o table-with-taxonomy.biom -m Metadata_16s.txt --observation-metadata-fp biom-taxonomy.tsv
```

Appendix D: R Code for 16S PCR File Processing

```
# load libraries
library(BiocManager)
library("phyloseq")
library("microbiome")
library("knitr")
library("biomformat")
library("ape")
library("dplyr")
library("tidyverse")
library("bipartite")

# load data
data <- read_biom(biom_file = "table-with-taxonomy.biom")

# OTU table
otu_table <- as.data.frame(as.matrix(biom_data(data)))

# Classifications for each ASV with confidence scores
taxonomy <- observation_metadata(data)

# Remove anything which doesn't have classification passed phylum
taxonomy_filter <- subset(taxonomy, grepl("^p_", taxonomy$taxonomy2))
phylo <- ape::read.tree(file = file.choose()) #The tree is the result of the function in
QIIME2 qiime phylogeny. You can extract the file.
metadata <- sample_metadata(data)
write.csv(metadata, file = "16s_EAB_meta.csv")
metadata <- read.csv(file = "16s_EAB_meta.csv", header = T, sep = ",", row.names = 1)

# Create phyloseq object by converting each file into component phyloseq class objects
SAM <-sample_data(metadata)
SAM$samples <- rownames(SAM)
TAX <-tax_table(as.matrix(taxonomy))
PHY <-phy_tree(phylo)
OTU <- otu_table(otu_table, taxa_are_rows=TRUE)
pseq <-merge_phyloseq(OTU, TAX, SAM, PHY)

# Set the column names for the taxonomy table
colnames(tax_table(pseq)) <- c("confidence", "Kingdom", "Phylum", "Class",
```

```
                    "Order", "Family", "Genus", "Species")

# Check and understand some basic information about the data
microbiome::summarize_phyloseq(pseq)

# If required, the names can be cleaned as follows.
tax_table(pseq) <- gsub("[a-x]__", "",tax_table(pseq))
tax_table(pseq)[1:3]

# Export relative abundance with ASVs as columns, samples as rows
# Extract abundance matrix from the phyloseq object
glom <- tax_glom(pseq, taxrank = "Species")

# transpose if necessary
if(taxa_are_rows(glom)){OTU1 <- t(OTU1)}

# ASV vs sample
write.csv(OTU1, "OTU-IDs-rel.csv")

# Need taxonomy of each ASV
# Export as dataframe with sampleID as rows, Genus name as columns
OTUdf <- as.data.frame((OTU1))
Tax.Names <- as.data.frame(glom@tax_table)
names(OTUdf) <- Tax.Names$Genus[match(names(OTUdf), row.names(Tax.Names))]
write.csv(OTUdf, file = "EAB_ASVs-with-meta.csv")

# sampleID as rows, species as columns
names(OTUdf) <- Tax.Names$Species[match(names(OTUdf),
row.names(Tax.Names))]
write.csv(OTUdf, file = "EAB_ASVs-with-species.csv")

# Load in ASVs-with-species.csv
sixtS <- read.csv(file.choose(), sep = ",")
row.names(sixtS) <- sixtS[,1]
sixtS <- sixtS[,2:113]

# Transpose so that species are a column
sixtSt <- t(sixtS)
sixtSt <- as.data.frame(sixtSt)
```

```
#### Make a binary dataframe
Binary_sixtS <- sixtSt
# Turn into binary data
Binary_sixtS[Binary_sixtS > 0] <- 1
# Make NAs 0s
Binary_sixtS[is.na(Binary_sixtS)] <- 0

# Import ASvs-with-meta.csv
Cols <- read.csv(file.choose(), sep = ",", header = FALSE)
Cols <- t(Cols)
Cols <- as.data.frame(Cols)
Cols <- Cols[2:113,1]

# Bind binary dataframe together
sixtS_sub <- cbind(Cols, Binary_sixtS)

# Group_by to get a summary table by Genus
sixtS_sub1_genera_count <- group_by(sixtS_sub, Cols) %>%
  summarise(
    count = n(),
 )

# Export file
write.csv(sixtS_sub1_genera_count, "EAB_16s_genera.csv")
```

Appendix E: R Code for Random Forest Models

```
##Load required packages

library(randomForest)
library(caret)
library(ggplot2)
library(dplyr)
library(iml)

##Data

#load data
Fungi_Interactions <- read.csv("EAB_fungi_pa.csv")

#rename response levels
Fungi_Interactions$Fungi_Interactions[Fungi_Interactions$Fungi_Interactions==1] <-
"Present"
Fungi_Interactions$Fungi_Interactions[Fungi_Interactions$Fungi_Interactions==0] <-
"Absent"

#convert variable to factors
Fungi_Interactions <- within(Fungi_Interactions, {
  Fungi_Interactions <- as.factor(Fungi_Interactions)
  Lifestage <- as.factor(Lifestage)
  Habitat <- as.factor(Habitat)
  Habitat_lvl2 <- as.factor(Habitat_lvl2)
  Region <- as.factor(Region)
  Sector <- as.factor(Sector)
})

#view data
summary(Fungi_Interactions)

##Make the random forest model

set.seed(42) #used for reproducibility

#define repeated k fold cross validation (arbitrary values)
repeat_cv <- trainControl(method = "repeatedcv", number = 10, repeats = 5)
```

```
#create training (70%) and testing (30%) sets
index <- createDataPartition(Fungi_Interactions$Fungi_Interactions, p = 0.7, list =
FALSE)
train = Fungi_Interactions[index,]
test = Fungi_Interactions[-index,]

#tune mtry value of random forest if needed, otherwise omit this line
tuning = expand.grid(mtry = 4)

rf <- train(
  Fungi_Interactions~.,
  data = train,
  method = "rf",
  trControl = repeat_cv,
  tuneGrid = tuning, #remove this line if tuning is not needed
  metric = "Accuracy")

print(rf)

rf$finalModel

##Tune model
tuneRF(
  x = train[,1:6],
  y = train[,7],
  stepFactor = 0.5,
  plot = TRUE,
  ntreeTry = 500,
  trace = TRUE,
  improve = 0.05)
#optimal mtry value is 4; model was adjusted above using tuneGrid

##Evaluate accuracy of model using test data
test_prediction <- predict(rf, test[,-7])
confusionMatrix(test_prediction, test$Fungi_Interactions)


##Evaluate and plot variable importance
var_imp <- varImp(rf, scale=FALSE)$importance
```

```
var_imp <- data.frame(variables=row.names(var_imp), importance=var_imp$Overall)
var_imp %>%
      arrange(importance) %>%
      top_n(10) %>% #only want top 10
      ggplot(aes(x=reorder(variables, importance), y=importance)) +
      geom_bar(stat='identity') +
      coord_flip() +
      xlab('Variables') +
      ylab("Importance (Mean Decrease in Gini Coefficient)")+
      labs(title='Top 10 Most Important Variables')

importance(rf$finalModel)  #provides importance values

##Create ALE plots
pred <- Predictor$new(rf, data = train[,-7], y = train[,7], type = "prob")

plot(FeatureEffects$new(pred, feature = "Lifestage")) +
  scale_x_discrete(limits = c("Larva1", "Larva2", "Larva3", "Pupa", "Adult"))

plot(FeatureEffects$new(pred, feature = "DOY"))

plot(FeatureEffects$new(pred, feature = "Region"))+
  scale_x_discrete(limits = c("Brockville", "Ennotville", "Guelph", "Puslinch"))

plot(FeatureEffects$new(pred, feature = "Sector"))+
  scale_x_discrete(limits = c("Flood", "Hebert", "Riverside", "Rosedale", "Sones",
"Tincap"))

plot(FeatureEffects$new(pred, feature = "Habitat"))+
  scale_x_discrete(limits = c("Artificial", "Woodland"))

plot(FeatureEffects$new(pred, feature = "Habitat_lvl2"))+
  scale_x_discrete(limits = c("Arable", "Forest", "Urban"))

#correlation plot
library(ggcorrplot)
model.matrix(~0+., data=Fungi_Interactions[,1:6]) %>%
  cor(use="pairwise.complete.obs") %>%
  ggcorrplot(show.diag=TRUE, lab=TRUE, lab_size=2.5)
```

Appendix F: Supplemental Figures

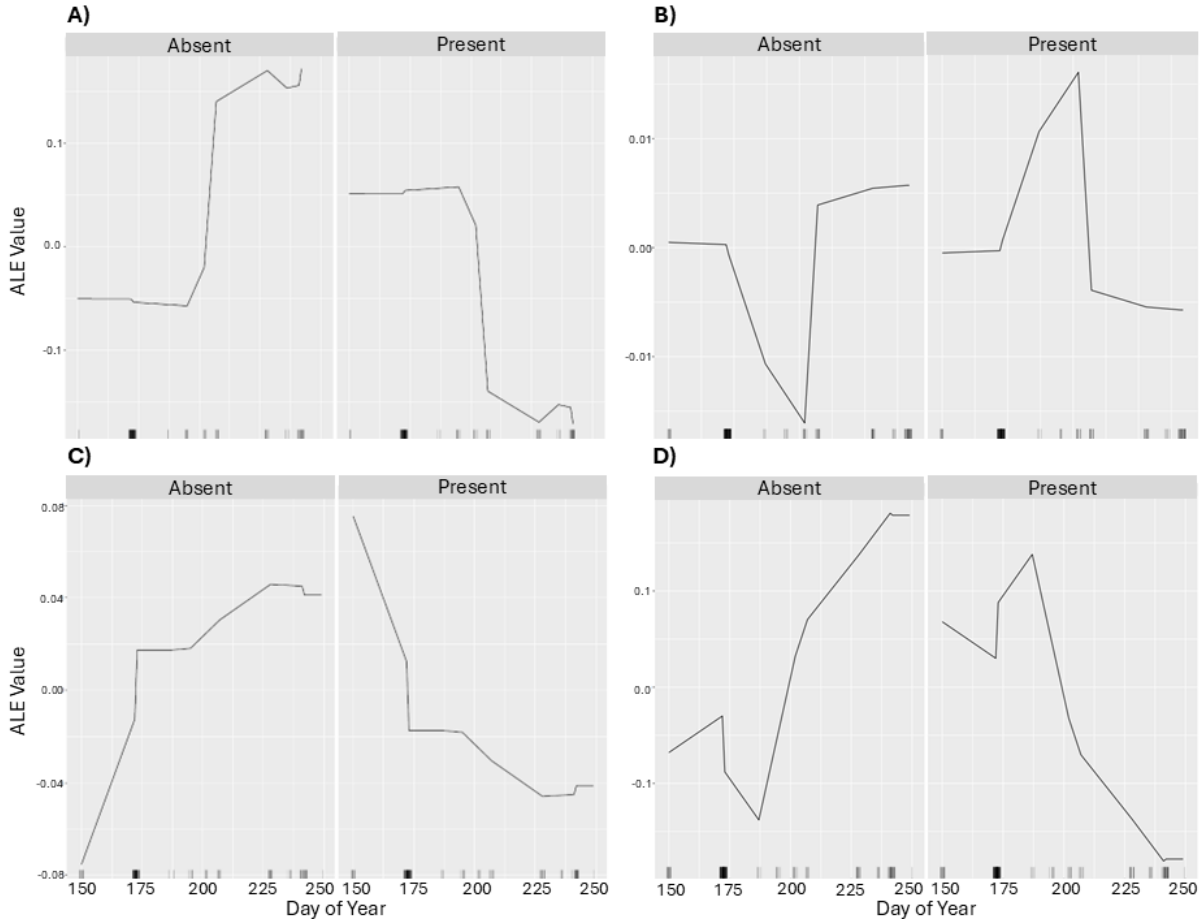

Figure S1. Effect of emerald ash borer collection date on the detection of interactions. Accumulated Local Effects (ALE) plots portray the effect collection date had on the presence or absence of A) overall interactions, B) animal interactions, C) bacteria interactions, and D) fungi interactions.

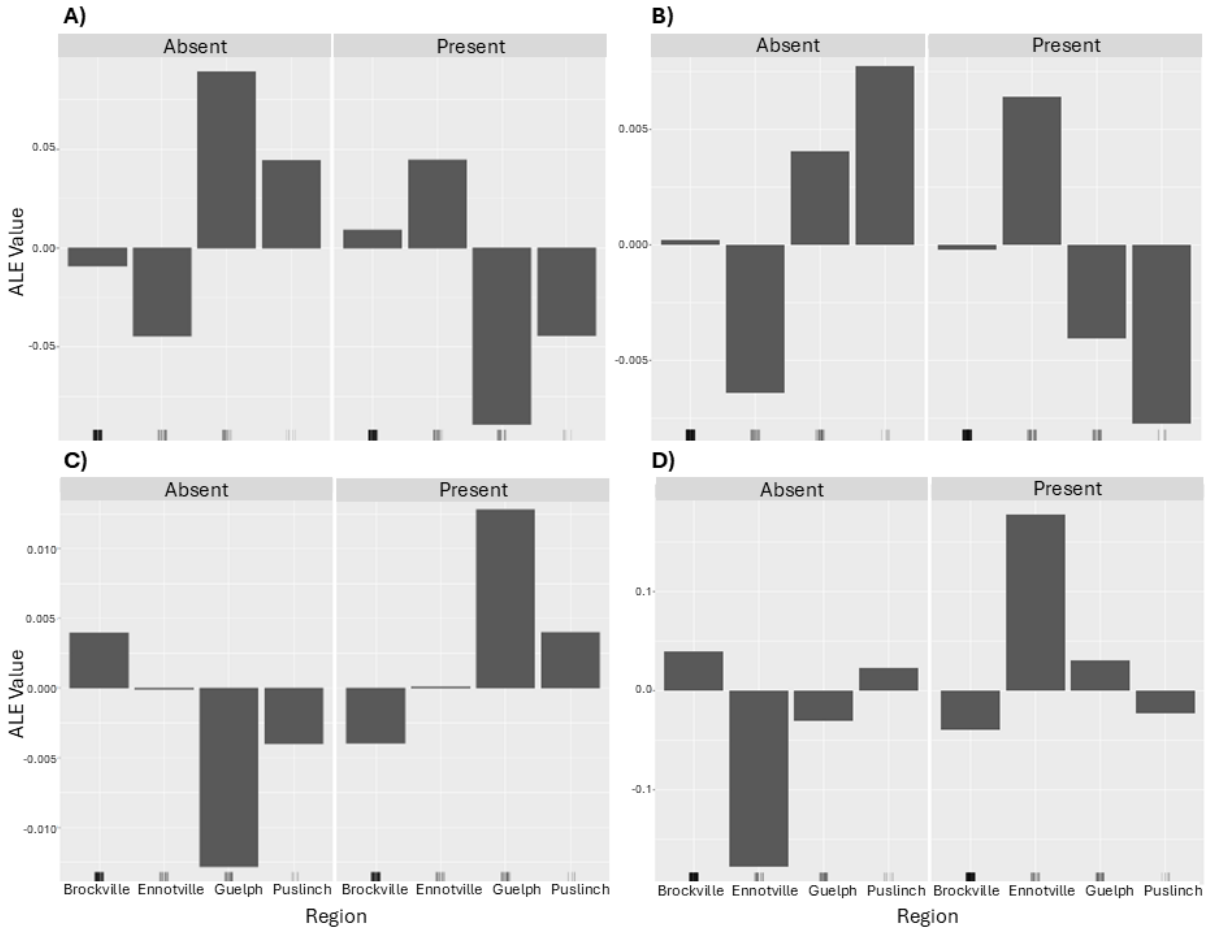

Figure S2. Effect of emerald ash borer collection region on the detection of interactions. Accumulated Local Effects (ALE) plots portray the effect collection region had on the presence or absence of A) overall interactions, B) animal interactions, C) bacteria interactions, and D) fungi interactions.

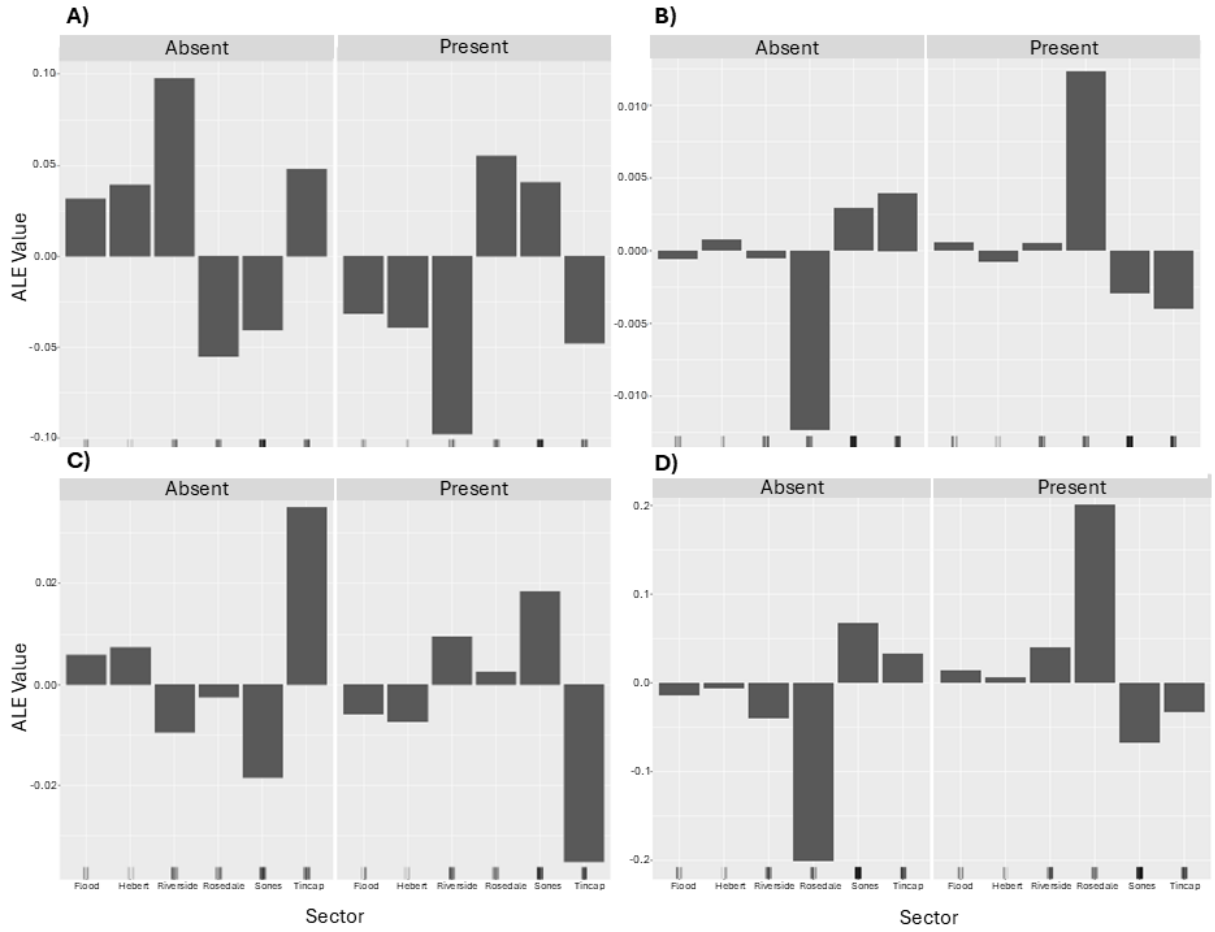

Figure S3. Effect of emerald ash borer collection sector on the detection of interactions. Accumulated Local Effects (ALE) plots portray the effect collection sector had on the presence or absence of A) overall interactions, B) animal interactions, C) bacteria interactions, and D) fungi interactions.

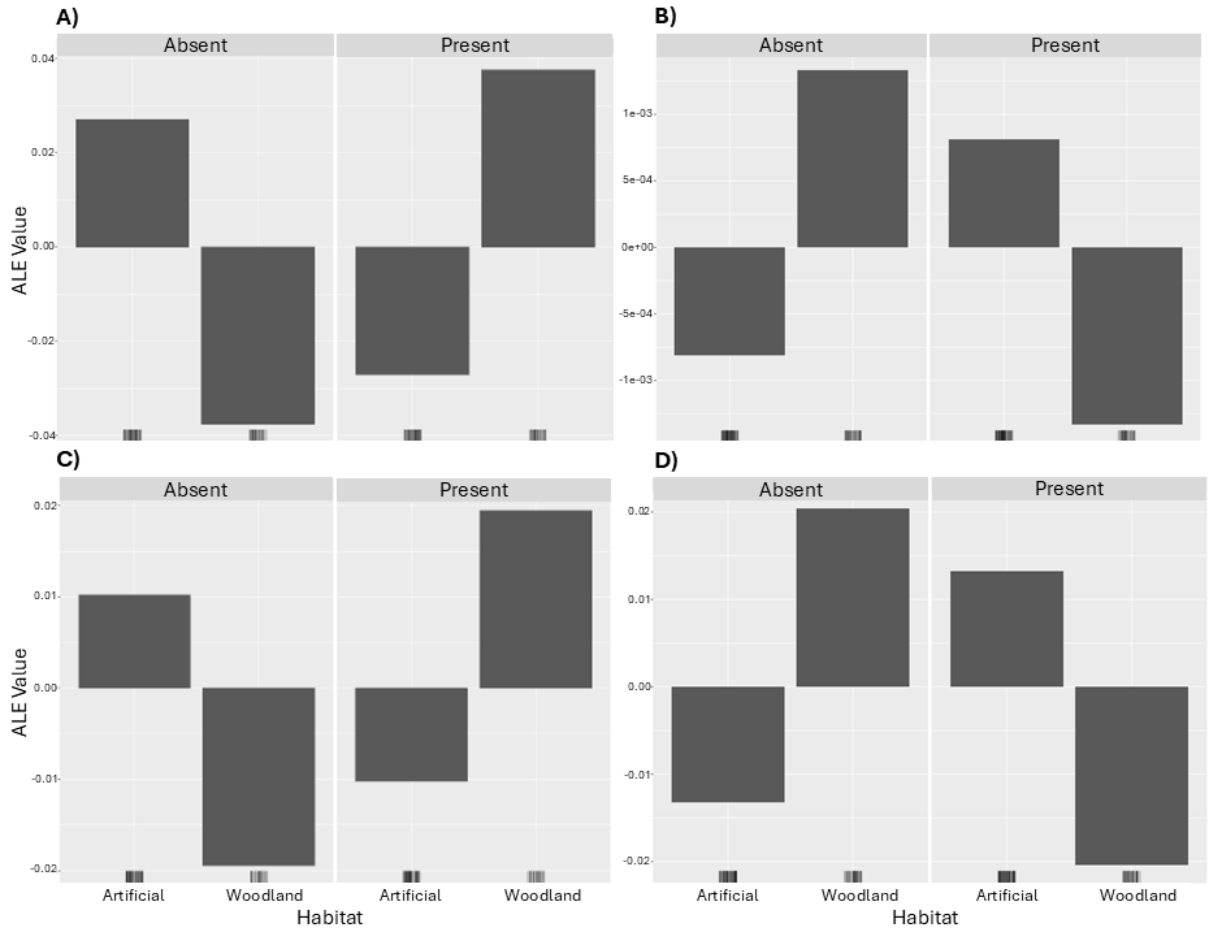

Figure S4. Effect of emerald ash borer collection habitat on the detection of interactions. Accumulated Local Effects (ALE) plots portray the effect collection habitat had on the presence or absence of A) overall interactions, B) animal interactions, C) bacteria interactions, and D) fungi interactions.

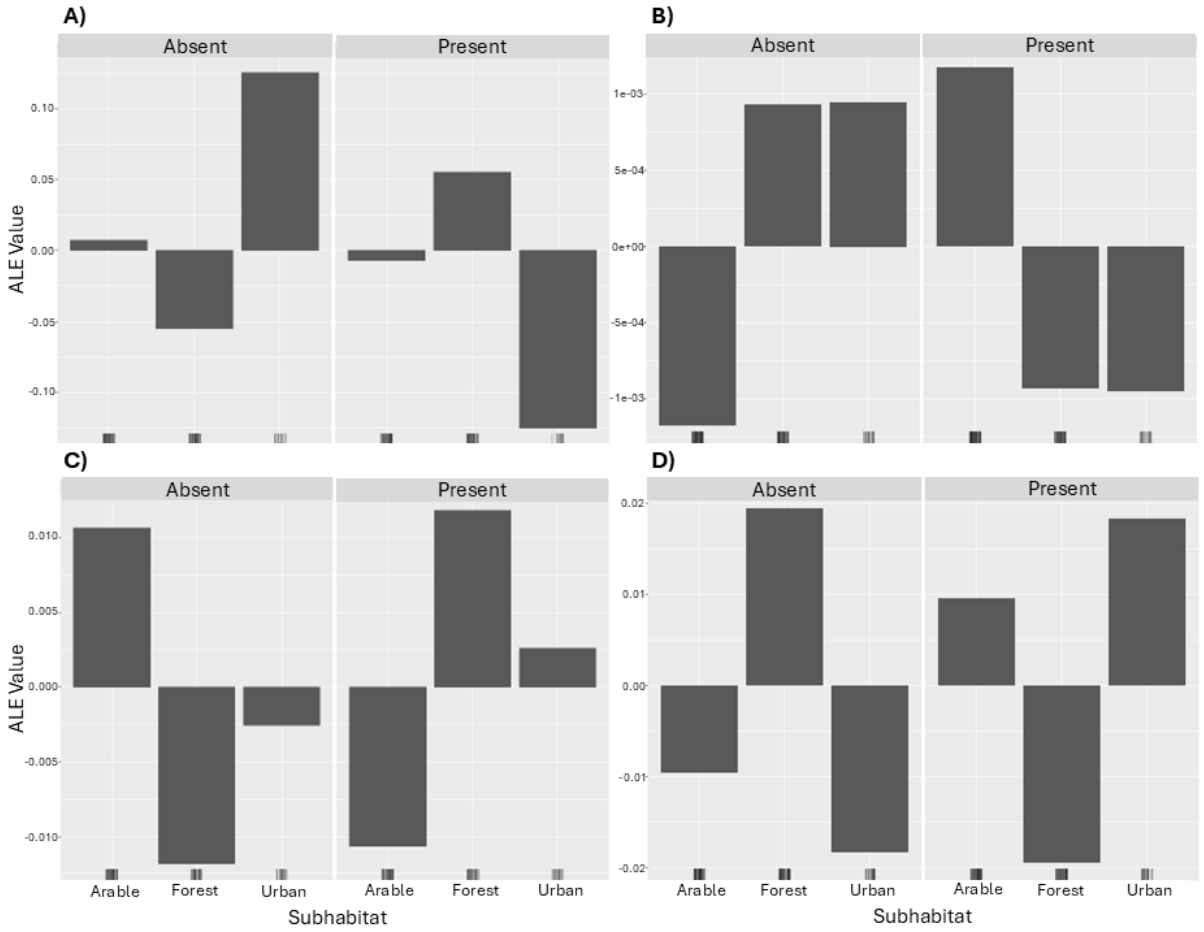

Figure S5. Effect of emerald ash borer collection subhabitat on the detection of interactions. Accumulated Local Effects (ALE) plots portray the effect collection subhabitat had on the presence or absence of A) overall interactions, B) animal interactions, C) bacteria interactions, and D) fungi interactions.

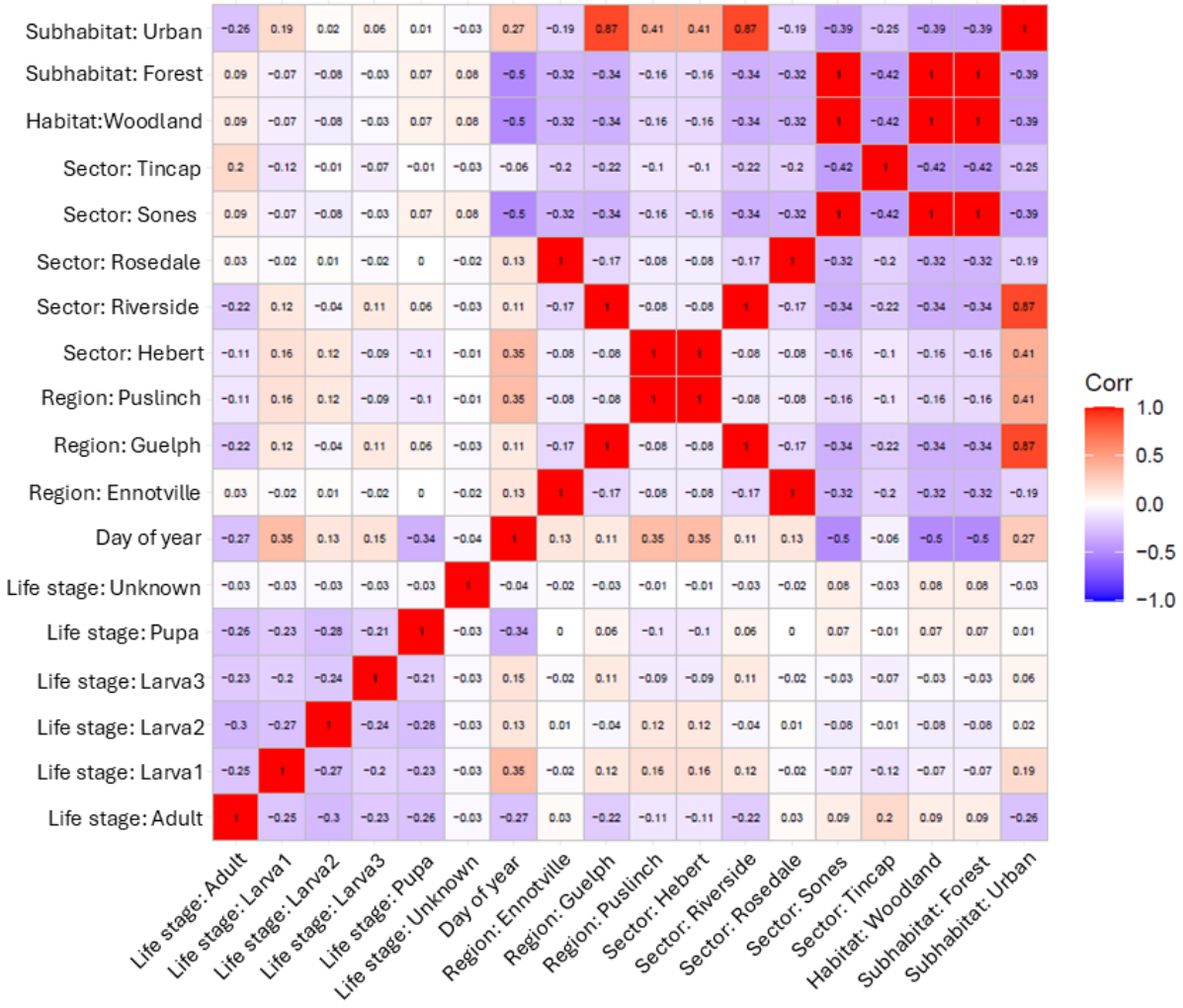

Figure S6. Correlation matrix of the predictor variables used in the random forest models. The correlation matrix portrays the correlation between life stage, collection date, collection location (region and sector), collection habitat, and collection subhabitat.