Genome-wide SNP identification in Fraxinus linking genetic characteristics to tolerance of Agrilus planipennis

CURRENT STATUS: POSTED

Cecelia E. Hale
Purdue University Fort Wayne

Mark A. Jordan
Purdue University Fort Wayne

Gloria Iriarte
Smithsonian Tropical Research Institute

Andrew J. Storer
Michigan Technological University

Vamsi J. Nalam
Colorado State University

Jordan Marshall
Purdue University Fort Wayne

marshalj@pfw.edu
Corresponding Author
ORCiD: 0000-0001-6024-3758

DOI: 10.21203/rs.2.14056/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
Ash, emerald ash borer, RAD-seq, tolerance, survival
Abstract

Background Ash (Fraxinus spp.) is one of the most widely distributed tree genera in North America. Populations of ash in the United States and Canada have been decimated by the introduced pest, Agrilus planipennis (Coleoptera: Buprestidae; emerald ash borer), having both negative impacts on forest ecosystems and economic interests. The majority of trees succumb to attack by A. planipennis, but some trees have been found to be tolerant to infestation despite years of exposure. Restriction site-associated DNA (RAD) sequencing was used to sequence ash individuals, both tolerant and susceptible to A. planipennis attack, in order to identify SNP patterns related to tolerance and health declines.

Results A de novo reference genome was assembled and single nucleotide polymorphisms (SNPs) were called using SAMtools. After filtering criteria were implemented, a set of 17,807 SNPs were generated. Principle component analysis (PCA) of SNPs aligned individual trees into clusters related to geography, however, five tolerant trees clustered together despite geographic diversity. A subset of 32 outlier SNPs identified within this group, as well as a subset of 17 SNPs identified based on vigor rating, are candidates for selection on host tolerance.

Conclusions Identifying genetic markers associated with host tolerance through genome-wide association has the potential to restore populations with cultivars that are able to withstand A. planipennis infestation. This study was successful in using RAD-sequencing in order to identify SNPs that are potential candidates to identify tolerance to A. planipennis. This was a first step toward uncovering the genetic basis for host tolerance to A. planipennis. Future studies are needed to identify the functionality of the loci where these SNPs occur and how they may be related to tolerance of A. planipennis attack.

Background
*Agrilus planipennis* Fairmaire (Coleoptera: Buprestidae; emerald ash borer) is a metallic green beetle native to northeastern Asia that has become a pest to North American ash (*Fraxinus* spp. L.) [1]. This pest was introduced into the Detroit/Windsor area of Michigan, USA/Ontario, Canada, and quickly dispersed via human assistance, including movement of firewood, nursery stock, and wood packing material [2,3]. In its native range, *A. planipennis* coevolved with Manchurian ash (*F. mandshurica* Rupr.) and is a secondary pest in this tree species requiring a primary stressor for successful attack [4,5]. Ash species in North America lack this natural resistance and succumb to attack, regardless of the presence of a primary stressor, often within one to four years after initial attack [4,6]. While black (*F. nigra* Marsh.), green (*F. pennsylvanica* Marsh.), and white (*F. americana* L.) ash are the most susceptible in the introduced range of *A. planipennis*, all North American ash are susceptible [2,7,8]. Blue ash (*F. quadrangulata* Michx.) has the lowest susceptibility to *A. planipennis* attack among North American ash species [9].

In North America, the life cycle of *A. planipennis* is typically completed within one year [10]. Females and males mature as they feed on canopy leaves. Males identify suitable mates via visual and contact cues, and females feed on foliage for an additional five to seven days after mating before oviposition begins [8,11,12]. Eggs are typically laid in bark cracks and crevices, with larvae subsequently tunneling into the bark to feed on the phloem and vascular cambium of the tree. Phloem consumption creates serpentine-shaped galleries, which severs photosynthate transport leading to eventual mortality. While a one-year life cycle is most common, a two-year life cycle does occur, especially in more northern latitudes, with larvae overwintering in intermediate instars within the phloem [13].

The proliferation of *A. planipennis* throughout forests in North America has caused the mortality of millions of ash trees, producing devastating ecological and economic impacts.
These impacts have created long lasting changes to North American forest ecosystems, requiring substantial restoration efforts [8,10]. Additional negative impacts include eradication of wood products produced from ash and diminished aesthetics in urban and suburban neighborhoods [8,14]. The cost of removal and replacement of ash trees in urban landscapes has been estimated at $12.5 billion from 2010–2020 [15]. Additionally, the estimated loss by timberlands in the United States is $300 billion [8].

Given the large-scale distribution of A. planipennis in both natural and urban landscapes, management options for control of the pest remain limited. Therefore, a long-term solution to preserving ash will depend on successfully identifying resilient genetic variants of ash. Resistance to wood-boring beetles is typically a function of female host selection and larval survival rate [17]. Therefore, resistance mechanisms can be placed into three general categories: antixenosis, antibiosis, and tolerance [18,19]. Antixenosis traits are aimed at decreasing preferences for feeding and/or ovipositioning, while antibiosis results from traits that negatively affect insect growth, survival, and/or fecundity. Lastly, tolerance is the ability of the host to withstand infestation while remaining relatively healthy compared to other individuals undergoing the same level of attack.

There is evidence of antixenotic traits in the interaction between A. planipennis and hosts. Adults of A. planipennis express variation in both feeding and oviposition host preferences. When given a choice, adult beetles preferentially feed on white, green, and black ash compared to Manchurian, blue, and European ash (F. excelsior L.) [20]. North American ash species receive more eggs compared to Manchurian ash, suggesting a female choice of susceptible hosts in order to increase larval performance [21,22]. Within North American ash species, inter- and intraspecific variation of volatile emissions and oviposition preferences of A. planipennis have been shown to play a role in resistance [23–25]. The bark of blue ash has a phenolic composition that may contribute to its resistance
relative to white, green, and black ash [5]. Bark smoothness, as a phenotypic characteristic, may be a limiting factor in oviposition locations and subsequently limits the number of larvae that could attack a tree at a given time [26]. Additionally, variability in ash growth rates have been related to susceptibility to *A. planipennis*, with trees tolerant of attack having more rapid and constant growth compared to susceptible trees [27]. Antibiosis interactions also exist in larval development. Mechanisms that affect larval performance mainly focus on variation in phenolic and defense protein chemistry [5,28-30]. Previous studies comparing phenolic and lignin profiles of ash species found that Manchurian ash contains unique profiles that may contribute to their resistance to *A. planipennis* [5,28,30]. Four potential defense-related proteins are expressed more than five-fold higher in Manchurian ash than in other species, and may contribute to resistance [30].

Mechanisms of tolerance are more difficult to quantify and therefore have not been as well studied. Identifying the genetic variants that allow these surviving trees in North America to tolerate infestation would greatly aide in the conservation of ash [29]. Even with severe levels of ash mortality in the introduced range, certain trees have been able to survive after years of repeated exposure [26]. This has led to the identification of trees with differing apparent tolerance levels to *A. planipennis* attack. Trees classified as tolerant survive in spite of signs of *A. planipennis* attack and damage [26]. The objectives of this study were to (1) identify ash single nucleotide polymorphisms associated with the tolerance-susceptibility gradient to *A. planipennis*, (2) identify phenotypic and genotypic relationships between trees relative to this tolerance-susceptibility gradient, and (3) test the hypothesis that tolerance and susceptibility are linked to identifiable genetic markers.

**Results**
Phenotypic Classification

The 47 ash trees selected for this study were sampled from six different geographic locations (Figure 1A). Thirty-eight green ash and nine white ash individuals were used for analysis. These trees were classified into two major groups: tolerant or susceptible to *A. planipennis* attack using both vigor and dieback. This categorization resulted in 28 trees being classified as tolerant and 19 as susceptible. Twenty-two (57.9%) green ash trees were categorized as tolerant, while six (66.7%) of the white ash trees were categorized as tolerant. Categorization (tolerant or susceptible) was independent of species (green or white) ($\chi^2 = 0.23$, df = 1, P-value = 0.630). Pooled across species, vigor rating categories were evenly represented within each category (Figure 1B). However, dieback was skewed right, with fewer trees having greater values of dieback (Figure 1C). Overall, signs of *A. planipennis* infestation (i.e. bark splits, exit holes, woodpecker damage, epicormic sprouts) were present in 37 individuals, 31 of which had more than one sign present. Of the ten trees lacking signs, eight were categorized as tolerant and two were categorized as susceptible.

RAD-Sequencing

Restriction site-associated DNA sequencing was used to sequence the genomes of 47 ash trees in order to identify SNPs that are correlated with tolerant or susceptible phenotypes. BLASTN analysis of the 1000 random sequences from each individual revealed that 60.7% of the genomic sequences were potentially ash-specific (no hits discovered). A high percentage of mitochondrial reads (18.0%) were found, potentially diverting reads from nuclear loci. Of sequences mapped, 17.7% aligned to *Populus tremula*, 0.5% *Sesamum* spp., and 0.5% *Olea europaea*. Due to the high mitochondrial hit counts, a reference of just polymorphic loci was created and BLASTN analysis was performed on those.
sequences. Out of this entire subset, only eight loci mapped to mitochondrial sequences. After SNP calling and initial filtering, a set of 23,243 SNPs were produced. Application of the more stringent filtering criteria (Table 1) generated a final set of 17,807 SNPs. Individuals were assessed for read depth. Mean read depth over all individuals was 91.5 and means ranged from 30–150 (Figure 2).

**Principal Component Analysis**

For major clusters (labeled as right, lower, upper, and middle) were identified in the PCA based on 17,807 SNPs (Figure 3A). Diverging substantially from all other groups, the cluster on the right contained trees from different geographic locations, including Houghton, Kensington, Oakwoods, and Willow. Furthermore, all the individuals in this cluster were classified as tolerant to *A. planipennis* attack. Clustering within the PCA did not appear to be influenced by tree species, with white and green ash occurring together in two of the four clusters (Figure 3C).

**Outlier SNPs**

Outlier detection based on the four PCA clusters identified 32 outlier SNPs within 28 different loci (Table 2). Outlier detection based on vigor rating identified 17 outlier SNPs with 13 different loci (Table 3). All outlier $F_{ST}$ values were skewed to the right and outliers were relatively high, suggesting directional selection (Figure 4). There were no outliers detected by analysis when trees were grouped based on tolerance and susceptibility. Out of all 41 outlier loci, only one matched to a known sequence from the NCBI nucleotide database (locus 30133_137; top blast hit: XM_011468015.1). This locus mapped to a PTI1-like tyrosine-protein kinase receptor.

Of the outliers detected between the PCA clusters, ten had a clear pattern of the polymorphic nucleotide being predominantly present in the five right cluster individuals.
(Figure 5). These patterns were slightly offset by similar genetic trends between the middle and right clusters, however, the right cluster clearly had the highest occurrence of these polymorphisms. Interestingly, one set of outlier SNPs, all occurring at the locus 1669_22, was present in all trees except the five in the right group and two trees from the middle group (Figure 6). These trees retained the reference nucleotide in this case, not the polymorphic nucleotide. For the outlier SNPs present at this locus, each individual either had all four polymorphic nucleotides or retained all four reference nucleotides. Of the outliers detected between groups based on vigor rating, four had a clear pattern of the polymorphic nucleotide being present exclusively in trees with high vigor (Figure 7). Three of these SNPs were present at one locus (4467_128). One outlier SNP at locus 10225_13 displayed a pattern of the reference nucleotide occurring more frequently in trees with high vigor, whereas, trees with low vigor all had the polymorphic nucleotide (Figure 8).

Phenotypic Analysis

PCA based on all phenotypic data resulted in no distinct clustering between geographic locations or species (Figure 3B). Tolerant and susceptible trees did separate in this PCA, however, this is due to the categorization being defined by the same phenotypic data used to calculate the PCA. The PC1 axes of both the SNP and phenotype PCA analyses were not correlated ($r = -0.16$, $P$-value = 0.270). There was no correlation between the SNP PC1 axis and vigor ($r = -0.22$, $P$-value = 0.132). Likewise, there was no correlation between the SNP PC1 axis and dieback ($r = -0.16$, $P$-value = 0.271). Additionally, there was overlap between green and white ash (Figure 3D).

Discussion

This study identified polymorphic loci in *Fraxinus* spp. using RAD-sequencing genotyping-
by-sequencing. The filter settings were selected to ensure high quality nucleotide data with sufficient coverage across individuals [31,32]. The resulting SNPs were used to highlight insights into a potential genetic basis for host tolerance to *A. planipennis*.

**Patterns of Genetic Variation**

PCA plots provided visual representation of genetic divergence among individuals. There was no clear relationship between the PCA clusters and geographic distribution, as three of the four clusters contained trees from multiple locations.

White and green ash did not separate out on the PCA as expected. This is not the first occurrence of genetic overlap of these two species. White ash is a polyploid species (2n = 46, 92, and 138) and hybridization appears to confound genetic results [33,34]. In some cases, white ash individuals group with green ash in phylogenetic analyses, a result that was attributed to the white ash samples likely being a polyploid hybrid with green ash [34]. Additionally, there is low genetic differentiation between white, velvet (*F. velutina* Torr.), and green ash [35]. Rapid radiation or recent exchange of genetic material could have led to these relationships [35]. The co-occurrence of white and green ash in all sampling locations presents the possibility of hybridization between the two species; therefore, low genetic differentiation could have resulted between individuals identified as the two different species based on phenotypic characters if hybridization occurred.

Across the PCA, there was little separation based on tolerance and susceptibility categories applied by field assessment data. The exception to this pattern was the right cluster, which contained five tolerant individuals from various geographic locations in Michigan. Four of those individuals were in close proximity to the *de facto* *A. planipennis* introduction epicenter [36], indicating they have been exposed to *A. planipennis* for nearly 20 years and are still able to tolerate infestation. These five trees were located in Houghton County, Kensington, and Oakwoods (the one individual from Willow was grouped
with Oakwoods for some analyses). For this reason, outlier SNPs were identified between the four clusters on the PCA to determine which SNPs were likely causing the variation in this group.

**SNP Candidates for Tolerance Selection**

All outlier SNPs detected between the PCA clusters had high $F_{ST}$ values and appeared to be responsible for the divergence of the right cluster. However, subsequent PCA on SNP variation with these outliers removed (results not shown) revealed that the five individuals in the right cluster still displayed the same pattern of divergence, indicating that these 28 loci are not the only source of variation within this group.

Throughout the outliers identified, there were similar genetic trends between trees in the middle and right clusters. The similarities between these two clusters are evident when looking at just the ten polymorphic loci that showed a pattern of almost exclusive presence in the right group. For seven of the ten loci, one to two trees from the middle cluster also had the polymorphic nucleotide. Two of these trees from the middle cluster were classified as susceptible; however, all of the trees from the middle group that had genotypic similarities with the five right group trees had no signs of *A. planipennis* attack (i.e. lacking bark spits, exit holes, woodpecker activity, and sprouting), despite being located in areas where *A. planipennis* is present.

The outlier locus 30133_137 mapped to a PTI1-like tyrosine-protein kinase 2. This protein is known to be involved in growth and development, as well as defense responses [37,38]. PTI1 serine/threonine protein kinases were described to be key components of speck disease resistance in tomatoes by amplifying signaling pathways [39]. This gene may also play a role in defense against *A. planipennis* by amplifying pathways necessary to tolerate infestation.
Locus 16669_22 is another potentially important gene for host tolerance. Four outlier SNPs at this locus had distinctive patterns in individuals with either all present as the polymorphism or all in their reference form. For the five individuals in the right cluster, the reference nucleotides were retained for all four SNPs at this locus. Unfortunately, BLAST analysis did not map this locus to any known genes. Additionally, two trees that clustered in the middle PCA group also had the reference nucleotides at this locus. These two trees were classified as susceptible, but interestingly, they were the only two susceptible trees with no signs of *A. planipennis* attack. This exemplifies the coarseness of categorizing tolerance based on phenotypic characteristics of vigor and dieback. These two trees with poor vigor and high dieback may simply be displaying other disease manifestations not associated with *A. planipennis* (i.e. Houghton County trees were along a highway and subject to salt spray).

Outliers detected between trees based on vigor rating resulted in an additional 13 loci being identified as potential candidates for host tolerance. None of these mapped to any known functional genes, however, the five outliers that did show a pattern of either the polymorphic or the reference nucleotide being present exclusively in high vigor trees are of particular interest. Future analyses on characterizing the functionality of the outlier loci detected in this study could expose the importance of these genes and the role they may play in tolerance.

A clear link between genotypic and phenotypic data was not identified. Most likely, this was due to the coarseness of phenotype classification, which then failed to correlate with complex genetic diversity. Phenotypes were defined by tree assessments, which included categorical tree health observations and presence or absence of signs of *A. planipennis* attack. A future study may have more success if these signs are quantified at finer scales (i.e. number of exit holes per square meter, area of phloem regrowth, and location of bark...
splits) as opposed to whole tree values. Detailed phenotypic data would allow for more robust analyses linking genotype and phenotype, such as a mixed-linear model. Finally, the power of genome-wide associations are affected by the genetic complexity and heritability of a trait [40]. As tolerance is expected to be a complex genetic trait, this increases the chance of false positive associations. To remedy this issue in future studies, as many genotypes as possible should be used along with high quality nucleotide data.

Conclusions

*Agrilus planipennis* has devastated populations of *Fraxinus* in North America, however, the survival of some individuals despite years of exposure to *A. planipennis* is evidence of host tolerance. Understanding the mechanisms of host tolerance through genome-wide association has the potential to restore populations with cultivars that are able to persist in the presence of *A. planipennis*. Despite the caveats presented above, this study was successful in using RAD-sequencing in order to identify SNPs that are potential candidates for tolerance to *A. planipennis*. This was a first step toward uncovering the genetic basis for host tolerance to *A. planipennis*. Future studies are needed to identify the functionality of the outlier loci detected in this study.

Methods

Study Species and Sample Collection

Trees were selected from *Fraxinus* spp. individuals within Fort Wayne, Indiana USA (*n* = 3), Huron-Clinton Metroparks, Michigan USA (*n* = 39), and Houghton County, Michigan USA (*n* = 5) (Figure 1). Within most of these locations, green ash was the dominant species with white ash being less common. However, in Houghton County, white ash dominated. Leaf, bud, and bark morphological characteristics were used to identify species (C. E. Hale and J. M. Marshall). Vouchers were not collected. Selection of trees was based on their
occurrence along an apparent gradient from high tolerance to high susceptibility (i.e. low tolerance) to *A. planipennis* attack. Apical buds were collected from trees and placed in liquid nitrogen immediately after collection. Fort Wayne and Huron-Clinton Metropark collections were made in July 2014. Houghton County collections were made in August 2016. Once returned to the lab, samples were stored at –80 °C.

**Tree Assessment**

Selected trees were assessed on vigor (overall tree health: categorical 1–5 with 1 being high vigor [crown with relatively few dead twigs; normal foliage color and density] and 5 being low vigor [more than half of crown dead]), crown dieback (percent of dead branch tips: 5–100%), and signs of *A. planipennis* attack (presence/absence: bark splits, exit holes, woodpecker damage, epicormic sprouts). Assessments followed those conducted in previous studies [26,41-44], which were derived from Millers et al. [45]. After assessment, 47 individuals were selected for analysis and given an overall categorization of tolerant or susceptible to *A. planipennis* infestation. This tolerant-susceptible categorization was similar to [46]. Individuals with a vigor ≤ 3 and dieback of ≤ 30 were considered tolerant. Individuals with a vigor of ≥ 3 and dieback > 30 were considered susceptible. Chi-squared analysis was used to test the null hypothesis that tolerance categorization was independent of species.

**DNA Extraction and Quantification**

Entire bud samples (two to three buds) were homogenized using sterile ceramic mortars and pestles, which were first cooled with liquid nitrogen. DNeasy Plant Mini Kit (QIAGEN) was used to extract total genomic DNA following the manufacturer’s protocol. DNA from each sample was quantified using UV spectrophotometry (NanoDrop 1000) absorbance. All samples were subsequently diluted to a concentration of 25 ng/μl.
Library Creation and SNP Discovery

Genomic DNA was converted into nextRAD genotyping-by-sequencing libraries (SNPsaurus, LLC) as described by Russello et al. [47]. Briefly, genomic DNA was first fragmented with Nextera reagent (Illumina, Inc), which also ligates short adapter sequences to the ends of the fragments. The Nextera reaction was scaled for fragmenting 7 ng of genomic DNA, although 14 ng of genomic DNA was used for input to compensate for the amount of degraded DNA in the samples and to increase fragment sizes. Fragmented DNA was then amplified for 25 cycles at 75 °C, with one of the primers matching the adapter and extending eight nucleotides into the genomic DNA with the selective sequence TGCAGGAG. Thus, only fragments starting with a sequence that can be hybridized by the selective sequence of the primer will be efficiently amplified. The nextRAD libraries were sequenced on a HiSeq 4000 with one lane of 150 bp reads (University of Oregon).

The genotyping analysis used custom scripts (SNPsaurus, LLC) that trimmed the reads using bbduk (BBMap tools, http://sourceforge.net/projects/bbmap/). Command was as follows:

```
bash bbmap/bbduk.sh in = $file out = $outfile ktrim = r k = 17 hdist = 1 mink = 8 ref = bbmap/resources/nextera.fa.gz minlen = 100 ow = t qtrim = r trimq = 10
```

Next, a de novo reference genome was created by collecting 10 million reads in total, evenly from the samples, and excluding reads that had counts fewer than 7 or more than 700. The remaining loci were then aligned to each other to identify allelic loci and collapse allelic haplotypes to a single representative. All reads were mapped to the reference with an alignment identity threshold of 95% using bbmap (BBMap tools). In order to assess the proportion of sequence reads that originated from *Fraxinus* spp. versus other species, 1000 high-quality reads from each sample were subject to BLASTN analysis in the NCBI database.
Genotype calling was done using SAMtools and BCFtools [48]. Command was as follows:

```bash
samtools mpileup -gu -Q 12 -t DP, DPR -f ref.fasta -b samples.txt | bcftools call -cv - > genotypes.vcf
```

The VCF file was filtered to remove alleles with a population frequency of less than 0.03. Loci were removed that were heterozygous in all samples or had more than two alleles in a sample (suggesting collapsed paralogs). The absence of artifacts was checked by counting SNPs at each read nucleotide position and determining that SNP number did not increase with reduced base quality at the end of the read. All polymorphic sequences retained were subject to BLASTN analysis in the NCBI database.

VCFtools [49] was used to further filter SNPs based on the following criteria: (1) Phred-quality score, (2) minor allele frequency, (3) maximum missing genotype, and (4) minimum mean read depth (Table 1). Loci that failed to meet the quantification threshold for any of the filtering criteria were removed and excluded from subsequent analyses.

Samples were not filtered based on Hardy-Weinberg expectations because the goal of this study was to identify polymorphic loci under selection, which are expected to deviate from equilibrium. The VCF file was converted into file formats necessary for analysis using PGDSpider 2.1.1.3 [50].

**Principal Component Analysis**

The packages `vcfR` v1.7.0 and `adegenet` v2.1.1 in R v3.4.2 [51-53] were used to perform an individual-based principal component analysis (PCA) to characterize structure based on SNP variation. PCA was also used to visualize relationships between individuals based on phenotypic characteristics (vigor, dieback, and signs of *A. planipennis* attack) using `prcomp` R base function. Pearson’s correlation was used to test the hypothesis that the first principle component axis for SNP PCA and the first principle component axis for phenotypic PCA had a linear relationship. Similarly, linear correlations were tested.
between the first principle component axis for SNP PCA with vigor and dieback values.

Detection of Markers under Selection

BAYESCAN v2.1 [57] was used to identify outlier loci based on populations defined by the PCA clusters, as well as populations defined by vigor rating. BAYESCAN uses a hierarchal-Bayesian method to estimate population-specific $F_{ST}$ coefficients, described by Beaumont and Balding [58]. A more conservative neutral model available in BAYESCAN (prior odds = 1000) was used to minimize the number of false positives. Prior odds or prior probability is the likelihood of the null hypothesis being true before the test is performed. This increase in prior odds corresponds to the selection model being 1000 times less likely than the neutral model, which was a more appropriate assumption given the number of SNPs included in this analysis [59]. After 100,000 iterations, SNPs with a posterior distribution over 0.95 were considered outliers. High $F_{ST}$ values (outliers) suggest that the locus has undergone directional selection (in contrast to balancing selection).

Abbreviations

**BLASTN:** Standard nucleotide basic local alignment search tool  
**$F_{ST}$:** Fixation index  
**PCA:** Principle component analysis  
**RAD:** Restriction site associated DNA  
**SNP:** Single nucleotide polymorphism

Declarations

Acknowledgements

The authors would like to thank Nick White for field assistance and SNPsaurus, LLC for contributing to a portion of the methods. The authors would like to thank Kirk Broders for
comments on an earlier version of the manuscript.

Funding

Financial support for this work was partially provided by Purdue University Fort Wayne Institute for Research, Scholarship, and Creative Endeavors Collaborative Grant Program; funds were used for data acquisition.

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the GenBank repository as BioProject ID: PRJNA561365, https://www.ncbi.nlm.nih.gov/bioproject/561365

Authors’ contributions

Conceived and designed the experiments: AJS, VJN and JMM. Performed the experiments: CEH, VJN and JMM. Analyzed the data: CEH, MAJ, GB, VJN and JMM. Wrote the manuscript: CEH, VJN and JMM. All authors have read and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for Publication

Not applicable.

Competing Interests

The authors declare that they have no competing interests.

References

1. McCullough DG, Katovich SA. Pest Alert. Emerald Ash Borer. NA-PR-02-04, U.S. Department of Agriculture, Forest Service. Northeastern Area, Newton Square, PA; 2004.

2. Cappaert D, McCullough DG, Poland TM, Siegert NW. Emerald ash borer in Northern
1. America: a research and regulatory challenge. Am Entomol. 2005;51:152-165.

3. Buck JH, Marshall JM. Hitchhiking as a secondary dispersal pathway for adult emerald ash borer, Agrilus planipennis. Great Lakes Entomol. 2008;41:155-157.

4. Rebek EJ, Herms DA, Smitley DR. Interspecific variation in resistance to emerald ash borer (Coleoptera: Buprestidae) among North American and Asian ash (Fraxinus spp.). Environ Entomol. 2008;37:242-246.

5. Whitehill JG, Opiyo SO, Koch JL, Herms DA, Cipollini DF, Bonello P. Interspecific comparison of constitutive ash phloem phenolic chemistry reveals compounds unique to Manchurian ash, a species resistant to emerald ash borer. J Chem Ecol. 2012;38:499-511.

6. Eyles A, Jones W, Reidl K, Cipollini D, Schwartz S, Chan K, Herms DA, Bonello P. Comparative phloem chemistry of Manchurian ash (Fraxinus mandshurica) and two North American ash species (Fraxinus americana and Fraxinus pennsylvanica). J Chem Ecol. 2007;33:1430-1448.

7. Anulewicz AC, McCullough DG, Miller DL. Oviposition and development of Emerald Ash Borer (Agrilus planipennis) (Coleoptera: Buprestidae) on hosts and potential hosts in no-choice bioassays. Great Lakes Entomol. 2006;39:99-112.

8. Poland TM, McCullough DG. Emerald ash borer: invasion of the urban forest and the threat to North America’s ash resource. J For. 2006;104:118-124.

9. Tanis SR, McCullough DG. Differential persistence of blue ash and white ash following emerald ash borer invasion. Can J For Res. 2012;42:1542-1550.

10. Herms DA, McCullough DG. Emerald ash borer invasion of North America: history, biology, ecology, impacts and management. Ann Rev Entomol. 2014;59:13-30.

11. Lelito JP, Fraser I, Mastro VC, Tumlinson JH, Böröczky K, Baker TC. Visually mediated ‘paratrooper copulations’ in the mating behavior of Agrilus planipennis (Coleoptera:
Buprestidae), a highly destructive invasive pest in North American ash trees. J Insect Behav. 2007;20:537–552.

12. Lelito JP, Böröczky K, Jones TH, Fraser I, Mastro VC, Tumlinson JH, Baker TC. Behavioral evidence for a contact sex pheromone component of the emerald ash borer, Agrilus planipennis Fairmarie. J Chem Ecol. 2009;35:104-110.

13. Wei X, Wu Y, Reardon R, Sun TH, Lu M, Sun JH. Biology and damage traits of emerald ash borer (Agrilus planipennis Fairmaire) in China. Insect Sci. 2007;14:367–373.

14. Flower CE, Knight KS, Gonzalez-Meler MA. Impacts of the emerald ash borer (Agrilus planipennis Fairmaire) induced ash (Fraxinus spp.) mortality on forest carbon cycling and successional dynamics in the eastern United States. Biol Invasions. 2013;15:931–944.

15. Kovacs KF, Mercader RJ, Haight RG, Siegert NW, McCullough DG, Liebhold AM. The influence of satellite populations of emerald ash borer on projected economic costs in US communities, 2010-2020. J Environ Manag. 2011;92:2170–2181.

16. Wagner DL, Todd KJ. Ecological impacts of emerald ash borer. In Biology and Control of Emerald Ash Borer. Forest Health Technology Enterprise Team, Morgantown, WV; 2015.

17. Hanks LM. Influence of the larval host plant on reproductive strategies of cerambycid beetles. Ann Rev Entomol. 1999;44:483-505.

18. Kogan M, Ortman EF. Antixenosis—a new term proposed to define Painter’s “nonpreference” modality of resistance. Bull Entomol Soc Am. 1978;24:175-176.

19. Painter RH. Insect Resistance in Crop Plants. The Macmillan Company, New York; 1951.

20. Pureswaran DS, Poland TM. Host selection and feeding preference of Agrilus planipennis (Coleoptera: Buprestidae) on ash (Fraxinus spp.). Environ Entomol.
20. Gripenberg S, Mayhew PJ, Parnell M, Roslin T. A meta-analysis of preference-performance relationships in phytophagous insects. Ecol Lett. 2010;13:383-393.

21. Rigsby CM, Muilenburg V, Tarpey T, Herms DA, Cipollini D. Oviposition preferences of Agrilus planipennis (Coleoptera: Buprestidae) for different ash species support the mother knows best hypothesis. Ann Entomol Soc Am. 2014;107:773-781.

22. Anulewicz AC, McCullough DG, Cappaert DL, Poland TM. Host range of the emerald ash borer (Agrilus planipennis Fairmaire) (Coleoptera: Buprestidae) in North America: results of multiple-choice field experiments. Environ Entomol. 2008;37:230-241.

23. Chen Y, Whitehill JG, Bonello P, Poland TM. Differential response in foliar chemistry of three ash species to emerald ash borer adult feeding. J Chem Ecol. 2011;37:29-39.

24. Koch JL, Carey DW, Mason ME, Poland TM, Knight KS. Intraspecific variation in Fraxinus pennsylvanica responses to emerald ash borer (Agrilus planipennis). New For. 2015;46:995-1011.

25. Marshall JM, Smith EL, Mech R, Storer AJ. Estimates of Agrilus planipennis infestation rates and potential survival of ash. Am Midl Nat. 2013;169:179-193.

26. Boyes KN, Hietala-Henschell KG, Barton AP, Storer AJ, Marshall JM. Linking tree growth rate, damage repair, and susceptibility to a genus-specific pest infestation. J For Res. 2019;xx:xxx-xxx.

27. Cipollini D, Wang Q, Whitehill JG, Powell JR, Bonello P, Herms DA. Distinguishing defensive characteristics in the phloem of ash species resistant and susceptible to emerald ash borer. J Chem Ecol. 2011;37:450-459.

28. Villari C, Herms DA, Whitehill JG, Cipollini D, Bonello P. Progress and gaps in understanding mechanisms of ash tree resistance to emerald ash borer, a model for wood-boring insects that kill angiosperms. New Phytol. 2016;209:63-79.
30. Whitehill JG, Popova-Butler A, Green-Church KB, Koch JL, Herms DA, Bonello P. Interspecific proteomic comparisons reveal ash phloem genes potentially involved in constitutive resistance to the emerald ash borer. PLoS One. 2011;6:e24863.

31. Bao R, Huang L, Andrade J, Tan W, Kibbe WA, Jiang H, Feng G. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. Cancer Informatics. 2014;13:CIN-S13779.

32. Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011;12:443.

33. Schlesinger RC. Fraxinus americana L. white ash. In Silvics of North America. Volume 2: Hardwoods. Agriculture Handbook 654. US Department of Agriculture, US Forest Service, Washington, DC; 1990.

34. Wallander E. Systematics of Fraxinus (Oleaceae) and evolution of dioecy. Plant System Evol. 2008;273:25–49.

35. Hinsinger DD, Basak J, Gaudeul M, Cruaud C, Bertolino P, Frascaria-Lacoste N, Bousquet J. The phylogeny and biogeographic history of ashes (Fraxinus, Oleaceae) highlight the roles of migration and vicariance in the diversification of temperate trees. PLoS One. 2013;8:e80431.

36. Siegert NW, McCullough DG, Liebhold AM, Telewski FW. Dendrochronological reconstruction of the epicenter and early spread of emerald ash borer in North America. Div Distrib. 2014;20:847–858.

37. Anthony RG, Khan S, Costa J, Pais MS, Bögre L. The Arabidopsis protein kinase PTI1–2 is activated by convergent phosphatidic acid and oxidative stress signaling pathways downstream of PDK1 and OXI1. J Biol Chem. 2006;281:37536–37546.

38. Floriduz M, Terzi M, Filippini F. Comparative proteome bioinformatics: identification of a whole complement of putative protein tyrosine kinases in the model flowering
39. Sessa G, D'Ascenzo M, Martin GB. The major site of the Pti1 kinase phosphorylated by the Pto kinase is located in the activation domain and is required for Pto-Pti1 physical interaction. FEBS J. 2000;267:171-178.

40. Burghardt LT, Young ND, Tiffin P. A guide to genome-wide association mapping in plants. Curr Protoc Plant Biol. 2017;2:22-38.

41. Clark RE, Boyes KN, Morgan LE, Storer AJ, Marshall JM. Development and assessment of ash mortality models in relation to emerald ash borer infestation. Arbor Urban For. 2015;41:270-278.

42. Marshall JM, Storer AJ, Fraser I, Beachy JA, Mastro VC. Effectiveness of differing trap types for the detection of emerald ash borer (Coleoptera: Buprestidae). Environ Entomol. 2009;38:1226-1234.

43. Marshall JM, Storer AJ, Fraser I, Mastro VC. Efficacy of trap and lure types for detection of Agrilus planipennis (Col., Buprestidae) at low density. J Appl Entomol. 2010;134:296-302.

44. Marshall JM, Porter MJ, Storer AJ. Predicting emerald ash borer, Agrilus planipennis (Coleoptera: Buprestidae), landing behavior on unwounded ash. Great Lakes Entomol. 2012;45:29-39.

45. Millers I, Lachange D, Burkman WG, Allen DC. North American sugar maple decline project: organization and field methods. USDA Forest Service Gen. Tech. Rep. NE-154; 1991.

46. Hietala KG. Evaluation and monitoring of ash (Fraxinus spp.) tolerant to long-term emerald ash borer (Agrilus planipennis [Coleoptera: Buprestidae]) exposure. MS thesis. USA: Michigan Technological University; 2013.
Russello MA, Waterhouse MD, Etter PD, Johnson EA. From promise to practice: pairing non-invasive sampling with genomics in conservation. PeerJ. 2015;3:e1106.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–2079.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–2158.

Lischer HEL, Excoffier L. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28:298–299.

Knaus BJ, Grünwald NJ. VCFR: a package to manipulate and visualize variant call format data in R. Mol Ecol Res. 2017;17:44–53.

Jombart T. Adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics. 2008;24:1403–1405.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; 2017 https://www.R-project.org/.

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155:945–959.

Earl DA. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Con Genet Resour. 2012;4:359–361.

Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour. 2015;15:1179–1191.

Foll M, Gaggiotti O. A genome-scan method to identify selected loci appropriate for
both dominant and codominant markers: a Bayesian perspective. Genetics. 2008;180:977-993.

58. Beaumont MA, Balding DJ. Identifying adaptive genetic divergence among populations from genome scans. Mol Ecol. 2004;13:969-980.

59. Lotterhos KE, Whitlock MC. Evaluation of demographic history and neutral parameterization on the performance of FST outlier tests. Mol Ecol. 2014;23:2178-2192.

Tables

Table 1. Filtering criteria for polymorphic loci. Loci failing to meet the quantification threshold for any of the criteria were excluded from subsequent analysis.

| Filtering Criteria                  | Quantification |
|-------------------------------------|----------------|
| Heterozygous in all samples         | False          |
| More than 2 alleles in a sample     | False          |
| Phred-like quality score            | > 20           |
| Minor allele frequency              | > 0.1          |
| Maximum missing genotype            | 0.5            |
| Minimum mean read depth             | 14             |

Table 2. Summary of outlier loci detected between four clusters within PCA. Includes single nucleotide polymorphism (SNP) location within loci, reference/polymorphic nucleotide pairs, and $F_{ST}$ and q Values. False discovery rate of < 0.05 was used.
| Locus     | SNP Location | SNP   | $F_{ST}$ | q Value |
|-----------|--------------|-------|---------|---------|
| 2690_69   | 86           | G/A   | 0.472   | 0.045   |
| 3535_59   | 67           | A/G   | 0.478   | 0.024   |
| 6512_10   | 66           | A/G   | 0.472   | 0.042   |
| 8409_8    | 142          | A/G   | 0.533   | 0.002   |
| 9230_115  | 99           | C/G   | 0.479   | 0.029   |
| 12476_19  | 13           | C/A   | 0.466   | 0.019   |
| 13284_148 | 49           | G/A   | 0.504   | 0.031   |
| 13980_8   | 103          | T/C   | 0.505   | 0.006   |
| 16669_22  | 10           | T/C   | 0.532   | 0.004   |
| 16669_22  | 13           | A/G   | 0.532   | 0.003   |
| 16669_22  | 49           | T/C   | 0.531   | 0.004   |
| 16669_22  | 104          | T/C   | 0.527   | 0.004   |
| 17461_7   | 34           | C/T   | 0.441   | 0.047   |
| 18944_8   | 106          | T/C   | 0.437   | 0.05    |
| 22135_11  | 3            | T/A   | 0.476   | 0.008   |
| 24473_11  | 46           | T/A   | 0.505   | 0.002   |
| 30133_137 | 109          | G/A   | 0.486   | 0.014   |
| 31830_23  | 118          | C/T   | 0.479   | 0.026   |
| 34843_20  | 109          | G/A   | 0.465   | 0.007   |
| 37961_17  | 133          | C/T   | 0.531   | 0.001   |
| 38762_25  | 57           | G/A   | 0.474   | 0.035   |
| 39733_10  | 132          | G/A   | 0.478   | 0.041   |
| 50806_23  | 97           | C/G   | 0.563   | 0.0002  |
| 51756_10  | 107          | A/C   | 0.469   | 0.011   |
| 56570_57  | 87           | T/A   | 0.474   | 0.039   |
| 56570_57  | 132          | C/A   | 0.471   | 0.037   |
| 56747_9   | 86           | C/T   | 0.441   | 0.033   |
| 57187_42  | 84           | G/A   | 0.536   | 0.003   |
| 59640_10  | 101          | G/A   | 0.506   | 0.005   |
| 61718_43  | 96           | T/C   | 0.485   | 0.022   |
| 67007_71  | 5            | C/A   | 0.461   | 0.017   |
| 88261_24  | 115          | T/C   | 0.51    | 0.004   |

Table 3. Summary of outlier loci detected between populations based on vigor rating. Includes single nucleotide polymorphism (SNP) location within loci, reference/polymorphic
nucleotide pairs, and $F_{ST}$ and q Values. False discovery rate of < 0.05 was used.

| Locus     | SNP Location | SNP | $F_{ST}$ | q Value |
|-----------|--------------|-----|----------|---------|
| 4467_128  | 20           | A/G | 0.19     | 0.004   |
| 4467_128  | 28           | T/C | 0.199    | 0.002   |
| 4467_128  | 55           | A/G | 0.192    | 0.005   |
| 10225_13  | 51           | C/T | 0.163    | 0.047   |
| 19593_14  | 42           | T/A | 0.174    | 0.029   |
| 25780_8   | 16           | T/C | 0.186    | 0.019   |
| 39536_56  | 126          | C/T | 0.237    | 0.001   |
| 41669_24  | 113          | T/C | 0.157    | 0.033   |
| 46716_44  | 9            | C/T | 0.204    | 0.012   |
| 46716_44  | 11           | G/A | 0.206    | 0.015   |
| 49707_11  | 44           | A/G | 0.209    | 0.003   |
| 50319_14  | 27           | C/T | 0.16     | 0.037   |
| 50319_14  | 102          | C/T | 0.166    | 0.026   |
| 56942_24  | 101          | C/T | 0.179    | 0.009   |
| 58136_88  | 134          | A/G | 0.194    | 0.017   |
| 71869_12  | 51           | G/A | 0.162    | 0.042   |
| 90283_36  | 123          | C/T | 0.157    | 0.022   |

Figures
Figure 1

Location of sampled trees in Fort Wayne, Indiana USA (A); Huron-Clinton Metroparks, Michigan USA (Kensington, Lower Huron, Oakwoods, and Willow); and Houghton County, Michigan USA. Number of trees per category of vigor (B) and dieback (C) determined by field assessments.

Figure 2

Boxplot of read depth for all sampled ash individuals.
Figure 3

Principle component analysis (PCA) of filtered single nucleotide polymorphisms (SNPs) (A,C) and phenotypic data (B,D) from ash individuals in Fort Wayne, Indiana USA; Huron-Clinton Metroparks, Michigan USA (Kensington, Lower Huron, Oakwoods, and Willow); and Houghton County, Michigan USA. Closed symbols represent tolerant trees and open symbols represent susceptible trees. Ellipses are arbitrary, represent visual clustering, and are presented for labelling of cluster locations in PCA.
Frequency of FST values of all SNPs computed between populations based on the PCA clusters (A) and populations based on vigor rating (B).
Number of trees in each PCA cluster that had the polymorphic nucleotide (white bars) or the reference nucleotide (black bars) for the ten outlier loci identified as having a clear presence in the right group. Within each locus, there was only one outlier SNP identified. Sample sizes for each cluster were as follows: right (n = 5), lower (n = 7), upper (n = 30), middle (n = 5). Samples with missing data for that loci were not included.
Figure 6

Number of trees in each PCA cluster that had the polymorphic nucleotide (white bars) or the reference nucleotide (black bars) for the locus 16669_22. Four SNPs were identified as outliers within this one locus (Table 3). Sample sizes for each cluster were as follows: right (n = 5), lower (n = 7), upper (n = 30), middle (n = 5). Samples with missing data were not included.
Number of trees in each vigor rating that had the polymorphic nucleotide (white bars) or the reference nucleotide (black bars) for the four outlier loci identified as having a clear pattern of presence in the high vigor groups. Sample sizes for each group were as follows: 1 (n = 11), 2 (n = 9), 3 (n = 10), 4 (n = 8), 5 (n = 9).

Samples with missing data were not included.
Number of trees in each vigor rating that had the polymorphic nucleotide (white bars) or the reference nucleotide (black bars) for the locus 10225_13. Sample sizes for each cluster were as follows: 1 (n = 11), 2 (n = 9), 3 (n = 10), 4 (n = 8), 5 (n = 9). Samples with missing data were not included.