DR. HILARY BARKER (Orcid ID : 0000-0002-2187-831X)
PROF. PÄR K INGVARSSON (Orcid ID : 0000-0001-9225-7521)
DR. RICHARD L. LINDROTH (Orcid ID : 0000-0003-4587-7255)

Article type : Original Article

Running title: Genes underlying aspen traits and insect communities

Linking plant genes to insect communities: identifying the genetic bases of plant traits and community composition

Hilary L. Barker*[1], Jennifer F. Riehl[2], Carolina Bernhardsson[3], Kennedy Rubert-Nason[2,4], Liza Holeski[5], Pär K. Ingvarsson[6], Richard L. Lindroth[1,2]

[1] Department of Integrative Biology, University of Wisconsin-Madison, 1630 Linden Dr., Madison, WI 53706-1598 U.S.A.

[2] Department of Entomology, University of Wisconsin-Madison, 1630 Linden Dr., Madison, WI 53706-1598 U.S.A.

[3] Department of Ecology and Environmental Science, Umeå University, Umeå, Sweden SE-901 87

[4] Current address: Division of Natural and Behavioral Sciences, University of Maine at Fort Kent, 23 University Drive, Fort Kent, ME 04743

[5] Department of Biological Sciences, Northern Arizona University, 617 S. Beaver St., Flagstaff, AZ 86011-5640 U.S.A.

[6] Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala BioCenter, Uppsala, Sweden SE-750 07

***Corresponding author:**

4622 University Ave.

Madison, WI 53707-7874 U.S.A.

hlbultman@gmail.com

Abstract

Community genetics aims to understand the effects of intraspecific genetic variation on community composition and diversity, thereby connecting community ecology with evolutionary biology. Thus far, research has shown that plant genetics can underlie variation in the composition of associated communities (*e.g.*, insects, lichen, endophytes), and those communities can therefore be considered as extended phenotypes. This work, however, has been conducted primarily at the plant *genotype* level and has not identified the key underlying genes. To address this gap, we used genome-wide association mapping (GWAS) with a population of 445 aspen (*Populus tremuloides*) genets to identify the genes governing variation in plant traits (defense chemistry, bud phenology, leaf morphology, growth) and insect community composition. We found 49 significant SNP associations in 13 *Populus* genes that are correlated with chemical defense compounds and insect community traits. Most notably, we identified an early-nodulin like protein (ENODL) that was associated with insect community diversity and the abundance of interacting foundation species (ants and aphids). These findings support the concept that particular plant traits are the mechanistic link between plant genes and the composition of associated insect communities. In putting the "genes" into "genes to ecosystems ecology", this work enhances understanding of the molecular genetic mechanisms that underlie plant-insect associations and the consequences thereof for the structure of ecological communities.

Key words: *Populus*, GWAS, plant-insect interactions, defense chemistry, community genetics

Introduction

Identification of the genetic mechanisms of species interactions and community composition is a major aim of community genetics. To date, studies have determined that different genotypes of plants have different communities of associated organisms (*e.g.*, insects, endophytes), and that community relatedness is mirrored by the genetic relatedness among plant genotypes in a common environment (Barbour et al. 2016, Keith et al. 2017, Koricheva and Hayes 2018, Kagiya et al. 2018). These associated communities are shaped by key plant traits, including morphology, phenology, and phytochemistry (Barbour et al., 2015; Barker, Holeski, & Lindroth, 2018; Robinson, Ingvarsson, Jansson, & Albrectsen, 2012; Wimp et al., 2007). To date, this research has been conducted primarily at the plant genotype level and thus the identity of the underlying plant genes has remained largely unresolved.

Previous studies that have investigated the genetics of plant resistance to insect herbivores have assessed insect-associated damage, fitness, abundance and diversity traits (Bernhardsson et al., 2013; Dewoody et al., 2013; Rönnberg-Wästljung, Ahman, Glynn, & Widenfalk, 2006; Thoen et al., 2017; Zinkgraf, Meneses, Whitham, & Allan, 2016). While many studies have explored the genetic basis of resistance in crops (*e.g.*, Tzin et al. 2015) and *Arabidopsis thaliana* (*e.g.*, Thoen et al. 2016), few have investigated the underlying plant genetics of complex insect communities (Dewoody et al. 2013, Bernhardsson et al. 2013). Dewoody *et al.* (2013) used QTL-mapping to identify genetic regions associated with foliar damage caused by various insect guilds in F_2 pedigrees of *P. trichocarpa x deltoides* hybrids. Bernhardsson and colleagues (2013) used association mapping to determine whether seven

defense-related genes in European aspen (*Populus tremula)* correlated with insect guilds and species richness. Both studies found a modest number of QTN (12-14) with small- to medium-sized effects (R^2 = 0.03-0.15 per QTL).

To significantly advance the field of community genetics, a broader experimental approach is required. First, research is needed at the genomic scale, with fine resolution (*e.g.*, genome-wide association mapping [GWAS], Ingvarsson and Street 2011), to pinpoint causative plant genes and genetic regions that structure associated communities. Second, like the work by Bernhardsson et al. (2013), plant traits (*e.g.*, secondary metabolites) that are known to shape associated communities should be included in analyses to identify genes underlying the traits and to assess whether there is overlap with genes that underlie community composition. Third, more complex association models (*e.g.*, multivariate GWAS) should be used to better capture the complexity of community traits rather than simplifying these metrics to abundance and richness values.

Here, we identified genes underlying both ecologically relevant tree traits in, and associated insect communities on, trembling aspen (*Populus tremuloides*), using a recently established genetic mapping population of 445 aspen genets. We quantified 20 tree traits, including size, growth, foliar morphology, foliar phenology, and phytochemistry, and surveyed herbivorous insect and ant communities. We chose herbivorous insects because they interact directly with the host tree, and thus should be affected by particular tree traits (*e.g.*, defense chemicals) and underlying tree genetics. We also included ant communities, since the ant species in our surveys tended aphid populations and had a pronounced effect on the canopy insect community (Barker et al. 2018, Wimp and Whitham 2012). We then used both univariate and multivariate genome-wide association analyses with a dataset of over 170,000

SNPs (single nucleotide polymorphisms) for each aspen genet to identify the underlying genes. We predicted that some of the same loci that shaped tree traits would also be associated with the composition of insect communities. We also predicted that insect species that are closely associated with the host tree (*e.g.*, leaf galling insects) would be more likely to have associations with tree genes than insects that interact less closely (*e.g.*, free feeding insects, ants).

Methods

Study system

Trembling aspen (*Populus tremuloides*) is the most widely distributed and genetically diverse tree species in North America (Mitton & Grant, 1996). It exhibits little population structure, with evidence of only two subpopulations (southwestern and northern, Callahan et al. 2013) across its range. This feature makes aspen ideal for genome wide association analyses, since population structure can mislead results with false positive associations (Ingvarsson & Street, 2011). In addition, aspen is a foundation species with substantial impacts on dependent communities (Hillstrom, 2009). Studies of aspen (Barker et al., 2018; Hillstrom, 2009) and other *Populus* species (Bangert et al., 2006; Robinson et al., 2012; Gina M Wimp, Martinsen, Floate, Bangert, & Whitham, 2005) have linked plant genotypes to unique arthropod communities, and these communities have low to high broad-sense heritabilities (0.09 to 0.78 H^2). Finally, key tree traits (tree size, bud phenology, extra-floral nectaries, and secondary chemistry) have been shown to structure insect community composition (Bangert et al., 2006; Barker et al., 2018; Robinson et al., 2012).

WisAsp Common Garden

We established a genetic mapping population of aspen in 2010 with genets collected from throughout Wisconsin (Fig. 1, latitude range: 358 km, longitude range: 186 km, corresponding to the northern subpopulation of aspen (Callahan et al., 2013)). The trees were planted in a randomized complete block design with four replicate blocks and a perimeter of non-experimental trees (to minimize edge effects) at the University of Wisconsin's Arlington Agricultural Research Station (Fig. S2). We originally planted 392 genets (N = 1568), however due to vole damage many trees died in both 2010 (N = 399) and 2011 (N = 334). We thus replanted some of our original genets and also collected and planted new genets for a total of 445 aspen genets with 3.56 replicates on average ($\pm$ 1.80 $sd, N = 1824$ in total). The garden was planted in a former grass-alfalfa field with Joy silt loam soil (USDA). Trees were planted with 2.5 m x 2.5 m spacing. The entire garden was surrounded by a 2.4 m tall electric fence to exclude deer, and the site was mowed and maintained as needed. The trees were 4-5 years old at the time of data collection (75% of the trees were at least 1.6 m and 2.6 m tall in 2014 and 2015, respectively).

Tree Trait Surveys

We measured ecologically relevant tree traits, including growth (measured 2012-2015), foliar morphology (2014-2015), phenology (2014-2015), and defense (phytochemistry 2014-2015, extra-floral nectaries 2014). To survey growth, we measured tree volume after each growing season. We recorded basal diameter (10 cm above ground level) and height (ground level to the base of the apical bud). Volume was calculated as $diameter^2 \times height$, a metric that correlates well with biomass (Stevens, Waller, & Lindroth, 2007). We calculated absolute growth as $log_{10}(treevolume_{final}/treevolume_{initial})$ and relative growth as $ln(treevolume_{final}) - ln(treevolume_{initial})$, respectively.

To measure foliar morphology, we haphazardly collected 20-30 leaves from each tree in late June/early July and scanned them on a LICOR flatbed scanner (Version 3100, Lincoln, NE). The leaves were then vacuum-dried and weighed. We calculated both average individual leaf area and specific leaf area ($leaf area\ (cm^2)/mass\ (g)$).

To assess phenology, we recorded timing of bud break and bud set using 5-point and 2-point scales, respectively. The bud break scale was adapted from Robinson et al. (2012) and varied from (1) dormant buds to (3) broken buds to (5) leaves that were flushed and completely unrolled but not yet fully expanded (Fig S1). The bud set scale measured whether the buds were still growing (0) or set and dormant (1). We examined each tree every 2-3 days for each survey. For bud break we measured the most advanced bud on the tree, following Project BudBurst (http://budburst.org/) protocols, while for bud set we measured the terminal stem bud only. In addition to measuring date of bud break and bud set, we also calculated the length of the growing season for each tree in number of days as $Julian\ bud\ set\ date - Julian\ bud\ break\ date$.

To quantify foliar defenses and nitrogen levels, we collected leaves in both late June/early July (same leaves as were used for foliar morphology assessment) and August to analyze phytochemistry and EFN density, respectively. For the June/July collection, we pulverized the dried and weighed leaves to a fine powder by ball milling, and stored them at -20 ^{o}C. We quantified foliar concentrations of nitrogen, condensed tannins, and salicinoids phenolic glycosides using near infrared reflectance spectroscopy (NIRS; see Table S1 for information on the chemical prediction models; FOSS NIRSystems, Laurel, MD, USA), as described by Rubert-Nason et al. (2013). Spectra were collected from dry, powdered leaf samples packed into 5-cm ring cup cells. After exclusion of outlier spectra with a global

Mahalanobis distance (MD) > 3, we developed calibrations relating NIR spectral bands (1100–2500 nm) to phytochemical parameters using a subset of samples (~150 – 400) chosen by the SELECT algorithm (WinISI v1.50 software Foss-Tecator, Infrasoft International LLC, State College, PA, USA) with a neighborhood MD of 1.0 (Shenk & Westerhaus, 1991; Shenk & Westerhaus, 1991). We acquired nitrogen reference values by combustion gas chromatography on a Thermo Flash EA1112 elemental analyzer (Thermo Finnigan, Milan, Italy), as described in Sollins et al. (1999). Condensed tannin reference values were measured colorimetrically (550 nm) relative to purified *P. tremuloides* condensed tannin material (Hagerman & Butler, 1980), after extraction of foliage into 70:30 v/v acetone/water and reaction with Fe(III) under acidic conditions (Porter, Hrstich, & Chan, 1986). Phenolic glycoside reference values were determined by extraction of foliage into methanol, followed by separation of extracts by ultra-high performance liquid chromatography and quantification by negative electrospray ionization single quadrupole mass spectrometry (Waters ACQUITY iClass UPLC/MS system, Milford, MA, USA), following methods adapted from Abreu et al. (2011) and Rubert-Nason et al. (2014; 2018). Samples with anomalous spectra (N~50, identified by MD>3) were also analyzed by these reference methods. We developed partial least squares regression models relating NIR spectra to phytochemical reference values (Table S1), and applied these models to predict the phytochemistry in all 1824 leaf samples from their corresponding NIR spectra.

To determine extra-floral nectary (EFN) density (number per leaf), we haphazardly collected 12 leaves from each tree in August 2014 and stored them on ice in the field. We then digitally scanned the upper surface of each leaf, and counted the number of EFNs that were present on each leaf in the scanned images. EFNs in aspen are located at the leaf/petiole juncture.

Trait variation, broad-sense heritability estimates, and relationships (genetic correlations) are all reported by Barker *et al.* (2018).

Insect Community Surveys

We visually surveyed herbivorous insect and ant species on the originally planted trees (N = 989, 328 aspen genets, all planted in 2010) in mid-July to early August 2014-15. We standardized our surveys by time to account for variation in tree canopy sizes. To ensure that our methods captured the complete insect community, we first constructed rarefaction curves of number of insect species observed per time interval (30 sec intervals with a total of 20 min surveyed for each tree) for 20 trees distributed throughout the garden (early July 2014 and 2015). The rarefaction curves revealed that species richness saturated at approximately 3 min of survey time for small trees (< 1.5 m tall) and 6 min for large trees (>2.5 m tall). We therefore surveyed each tree for at least three minutes and larger trees were surveyed for additional time units (3 min increments). We then standardized the insect counts by minutes surveyed (*e.g.*, number of insects per species per minute). We stopped the survey time to collect, identify, and record insects.

In 2014, we surveyed the entire tree. In 2015, due to increased tree size, we divided the canopy into lower, middle, and upper sections and surveyed each for the same interval of time. We limited the maximum time interval to 4 min (12 min of survey time for a tree), which allowed us to survey a large subset of the canopy, but not necessarily the entire canopy. We surveyed the insect communities from 8:30AM to 4PM each day, and we conducted the surveys only on days with fine weather ($25.6℃ \pm 7.1$ average maximum temperature, $2.4\ \mathrm{m/s} \pm 1.2$ average wind speed, sunny to partly cloudy).

One author (H.L. Barker) trained a team of 6-7 surveyors for each insect survey (2014-15), and was present each day of the survey to address insect identification questions. Insects were identified using field guides and keys, and specimens were collected as needed for further identification in the lab. All common insects were identified to species and rare insects to morpho-species (family-level). We also surveyed insect-inflicted damage, including leaf mines, leaf and petiole galls, and leaf rolls and tents. Vouchers of common insect species are preserved in the UW-Madison Wisconsin Research Insect Collection, and H.L. Barker has provided a website (https://aspeninsects.wordpress.com/) that displays information (life history, key identification features, etc.) for the common insect species found at WisAsp.

Insect community variation, broad-sense heritability estimates, and insect relationships (genetic correlations) are all reported by Barker *et al*. (2018).

Genetic Analyses

Foliar DNA was extracted from one ramet in each genet (leaves were collected in June 2012-14 and freeze-dried prior to extraction). Probes (N = 65,000, 120 bp in length) were designed by Rapid Genomics (Gainesville, FL) and Nathaniel Street (Umeå Plant Science Center, Sweden) to align to each gene (exome capture) from the *P. tremula* genome assembly v1.1, which contains 9,789 gene bearing scaffolds and 36,322 gene models (Lin et al., 2018; Sjödin, Street, Sandberg, Gustafsson, & Jansson, 2009). These probes were then tested on a population of 24 *Populus* genets to identify probes that were suitable for sequencing (*e.g.*, where sequencing reads mapped to a unique genomic region). This test resulted in 45,934 probes from 5,478 scaffolds and located in 20,483 gene models, with an average of 2.3 probes/gene for sequencing. Extracted DNA from all 434 genets were then

sent to Rapid Genomics (Gainesville, FL) for paired-end sequencing (2x100bp) on an Illumina HiSeq 2000 with a minimum sequencing depth of 15x per sample. Based on technical replicates, the sequencing error rate was estimated to be 0.2%. The raw sequencing reads were mapped to the *P. tremula* v1.1 genome assembly (Lin et al., 2018) using BWA-MEM v. 0.7.12 (Li & Durbin, 2009) and sorted with Samtools v. 1.2 (Li et al., 2009). Optical duplicates were marked using Picard v. 1.127 (http://broadinstitute.github.io/picard/). Local realignments around indels and per individual genotyping (Haplotypecaller in gVCF mode) were performed with GATK v. 3.4-46 (DePristo et al., 2011; Mckenna et al., 2010; Van der Auwera et al., 2013) with a diploid ploidy setting and otherwise default settings. Individual g.vcf files were combined into batches of ~200 samples using GATK CombineGVCFs to hierarchically merge them into a single gVCF. Finally, a joint call over all samples was conducted using GATK GenotypeGVCFs with a standard emit confidence of 10 and a standard call confidence of 20.

A subset of genets (N = 11) had been sequenced previously with whole-genome sequencing (for complete details, see Wang et al. 2016). Single nucleotide polymorphisms from these 11 genets were merged with SNPs from probe sequencing (434 genets) and filtered for genotype and sample quality metrics using VCF and BCFtools (Danecek et al. 2011, see full SNP filtering pipeline in Table S2). After the SNPs were filtered, missing genotype information was imputed using LinkImpute (Money et al., 2015). This SNP filtering pipeline resulted in a dataset of 173,520 SNPs distributed across 5,332 scaffolds and 20,483 genes with eight SNPs per gene on average.

Finally, seven genets were removed after pruning full-sibs from our dataset prior to GWA analyses. After full-sib removal, genetic relatedness among the remaining individuals in the WisAsp population was extremely low, with a mean of -0.002 (± 0.007 SD, Fig. 1).

Statistical Methods

Several factors can influence the success of genome-wide association (GWA) analyses to detect significant genes underlying traits of interest (Ingvarsson and Street 2011), including measurement error and population structure. Measurement error can bias the trait data, impacting the ability to find SNP associations, whereas underlying genetic structure can lead to the detection of spurious SNP associations. To deal with potential biases, we first assessed structure within the WisAsp population using the multi-locus approach, Admixture (Alexander, Novembre, & Lange, 2009). Following recommendations of the Admixture manual, the filtered SNPs were first pruned by linkage disequilibrium before analyses (pairs of SNPs within a 50 bp window were pruned if they exhibited an r^2 value greater than 0.2 using PLINK), resulting in 139,338 SNPs. This set of SNPs was then analyzed with various population structures (K = 1 through 5) and cross-validation to identify the number of populations that best explain variation in allele frequencies among genets.

We then performed both univariate and multivariate GWA analyses with the 173,520 SNPs dataset to identify genomic regions in aspen that are associated with phenotypic traits (*i.e.*, phenology, growth, leaf morphology, phytochemistry, and extra-floral nectaries) and insect populations (*i.e.*, presence or abundance of particular species), families, guilds, and community metrics (*i.e.*, species richness, abundance, nonmetric multidimensional scaling [NMDS] axes for Bray-Curtis community dissimilarity matrices; for complete list of traits see Table S3). We assessed both the presence/absence and abundance of various insect species

and groups (families, guilds) to see if the same or difference SNPs are associated with these metrics. We first regressed each trait and metric (x_{jklm}, *e.g.*, species richness) on covariates, including experimental block (b_j) and year (y_j) in which the data were collected (both as fixed effects) and genet (g_l, as a random effect) using lme4 in R (Bates, Maechler, Bolker, & Walker, 2015; see below). The grand mean is u, and e_{jklm} is the error. To best meet the assumptions of the linear mixed models, each trait and metric (x_{jklm}) was boxcox transformed using the MASS package in R (Venables and Ripley 2002). From these models, we extracted the best linear unbiased predictors (BLUP) for each genet and rank-transformed these values for GWA analyses (Goh and Yap 2009).

$$x_{jklm} = u + b_j + y_k + g_l + e_{jklm}$$

To both identify gene associations and conduct sensitivity analysis for our GWAS models, we used two statistical packages: PLINK v.1.9 (Purcell et al., 2007) for both univariate and multivariate traits and Genome-wide Efficient Mixed Model Association (GEMMA v.0.96, Zhou and Stephens 2012) for univariate traits (results shown in supplemental data). PLINK uses a simple linear regression without corrections for relatedness among individuals, while GEMMA uses a compressed mixed linear model (Zhang et al., 2010), which controls for relatedness among individuals with a centered kinship matrix. Kinship does not appear to be an important factor for our WisAsp population, as suggested by the low relatedness of individuals after removal of full-sibs. Therefore, inclusion of a kinship matrix in the GWA analyses had little effect on the results (90% of the significant associations remained, Table S4). We also did not include any covariates for population structure, since Admixture results indicated that our sample population is panmictic (Fig. 1). For data analyses, we used SNPs with a minor allele frequency of at least 0.05 and we corrected for multiple testing using a false discovery rate (FDR) of 0.10 (which is similar to previous FDR cut-off values for Salicaceae GWAS studies, *e.g.*, Hallingbäck et al. 2016) for

each GWAS model calculated with the qvalue package in R (Bass, Storey, Dabney, & Robinson, 2015). Data from this manuscript can be found in the Data Dryad Repository (Barker et al. 2019).

To further elucidate the function of particular genetic regions and the mechanism by which they may influence insects, we compared significant associations with the annotated *P. trichocarpa* genome v3.0 (Tuskan et al., 2006), *P. tremuloides* genome v1.1 (Lin et al. 2018) and *Arabidopsis thaliana* genome TAIR10 (Swarbreck et al., 2007) using the *Populus* Genome Integrative Explorer (PopGenIE, Sjödin et al. 2009). To determine whether particular tree traits were important in structuring insect communities, we included various standardized tree traits as covariates in the insect community GWA analyses in PLINK. If the significant insect-associated SNPs disappeared with inclusion of the tree trait covariates, we inferred that these tree traits were important in shaping the SNP-related variation in the insect phenotype.

To link gene functions, products, and processes to associated traits, we conducted gene set enrichment analysis using the generic gene ontology mapper (http://go.princeton.edu/cgi-bin/GOTermMapper, Gene Ontology Consortium et al. 2004, Boyle et al. 2004), including the top 0.1% most significant SNPs (N = 174) from each GWA test. These gene sets were compared to the background set of genes that were included in our probe set (based on *Arabidopsis* homologs) and only unique gene names were used to control for differences in gene size and the number of SNPs/gene. These tests were run with a Bonferroni p-value cutoff of 0.10.

Results

Of 79 GWA analyses conducted for the various tree and insect traits, we identified significant associations for 15 of these traits (Tables 1 and S3). 49 SNPs (five synonymous and 27 nonsynonymous, 17 noncoding) from 13 different genes were identified, which were distributed across eight chromosomes (based on alignment to *P. trichocarpa* v3.0 genome). Individual SNPs explained 5.7-8.4% of the phenotypic variation for the associated trait (R^2 values in Table 1). Of the unique associated SNPs, nine were 3' UTR or intron variants that likely influence gene expression, four were missense mutations that alter the amino acid sequence of the affected gene and two were located in upstream or downstream regions (within 2kb) and could also affect gene expression. Many of the genes harboring significant SNPs are involved in modulating gene expression, protein modification or the movement of resources in and out of cells. In addition, several of the identified genes are known to be regulated by plant hormones (jasmonic acid, abscisic acid, brassinosteroids, ethylene).

SNPs associated with tree trait variation

Of the 20 univariate GWA tests for tree traits, three resulted in significant associations. These associations were for levels of tremulacin, phenolic glycosides (combined levels of tremulacin and salicortin) and total defense phytochemistry (combined levels of tremulacin, salicortin, and condensed tannins, Table 1). Two SNPs in *Potra003979g23949* that encode for an ASC1-like (Cyp1 Absence of growth Suppressor) protein were identified in GWA analyses for both tremulacin and phenolic glycosides. The associated genes accounted for 5.9%, 5.9%, and 5.7% of the variation in tremulacin, phenolic glycosides, and total defense phytochemistry, respectively. No SNPs were identified for tree growth/size metrics, bud phenology, leaf morphology, foliar nitrogen, and the density of extra-floral nectaries (EFN). Multivariate GWA analysis of uncorrelated ($r < 0.70$) tree traits (*i.e.*, relative

and absolute growth, specific leaf area, individual leaf area, EFN density, growing season length, and levels of condensed tannins, phenolic glycosides, and nitrogen) also resulted in no significant SNPs.

Of the gene set enrichment tests we performed for the 20 tree traits, eight exhibited significantly enriched gene ontology terms, including variation in tree size/growth metrics (e.g., spring volume, basal area increment), specific leaf area, and EFN density (Table 2). Many of the tree size/growth metrics were enriched for genes involved in response to misfolded proteins, specific leaf area was enriched for genes involved in defense (immune) response and EFN density was enriched for genes involved in glyceraldehyde-3-phosphate metabolism.

SNPs associated with insect variation

We identified significant associations in our GWA analyses for 10 out of the 47 univariate insect traits. These associations were for variation in the incidence of *Ecteodemia populella* (petiole-galling moth), *Phyllonorycter tremuloidiella* (blotch mining moth), *Clostera albosigma* (leaf-rolling moth), Cecidomyiidae (leaf-rolling flies), and *Lasius neoniger* (most common ant species), and the abundance of *Clostera albosigma*, *Choristoneura rosaceana* (leaf-rolling moth), Tortricidae (leaf-rolling moths), and *Lasius neoniger* (Table 1). The significantly associated genes accounted for 7.4-22.6% of the total variation for the insect traits (summing across all genes identified for each insect trait). When aphid incidence (BLUP) was added as a covariate within the GWAS model, all significant ant incidence and abundance SNP associations remained, except for *Potra002557g19270* (Table 1). No SNPs were identified for free-feeding insects, leaf-galling flies, several leaf-mining

moth species, and insect community metrics (*i.e.*, abundance, richness, Shannon index, Table S3).

To further explore the importance of particular tree traits to insect phenotypes, we conducted GWA analyses with those traits incorporated as covariates. Loss of significant SNP associations would indicate that the tree trait was important in (directly or indirectly) shaping the gene-insect relationship. Including tree trait covariates in the GWA models made many of these significant insect associations disappear (Table 3). In particular, individual leaf area, defense phytochemistry, and tree size/growth traits eliminated all of the significant associations for particular leaf-modifying insects. In addition, significant associations for leaf-rolling insect species disappeared when bud phenology was included in the analysis. No tree trait covariate could completely eliminate the 12 significant SNPs for *Lasius neoniger* abundance or incidence, but inclusion of either individual (tremulacin and salicortin) or combined levels of phenolic glycosides made three to four of the SNPs (found in *Potra002557g19270* and *Potra003286g21239*) insignificant.

We identified significant associations for two out of nine multivariate insect traits. These associations were for nonmetric multidimensional scaling (NMDS) axes for variation in both the abundance and presence/absence of common insect species (Fig. 2). Both multivariate traits were associated with the same gene, *Potra001060g09097*. The significant SNPs had variable effects on the different components of the trait, with coefficients ranging from -0.72 to 0.82 (Fig. 2 C). Specifically, aspen with the C allele at these SNP sites had a more diverse insect community (all common insects were more often present on these trees; Fig. 2 B) that was dominated by a few very abundant species (*Harmandia* sp., aphids, and aphid-tending ants; Fig. 2 A).

Of the 58 gene set enrichment tests we performed for insect phenotypes, 14 exhibited significantly enriched gene ontology terms (Table 2). Several of the enriched gene sets were associated with leaf-galling and -rolling insects and multivariate insect traits. The enriched gene ontology terms included biosynthesis of an anthocyanin-containing compound, response to mechanical stimulus, cell wall biogenesis (notably for a leaf-galling insect species), γ-aminobutyric acid (GABA) transport, and hormone biosynthesis/regulation (Table 2).

Discussion

Community genetics research has highlighted the importance of plant intraspecific variation in structuring associated communities (Barbour et al., 2016; Kagiya et al., 2018; Keith et al., 2017; Koricheva & Hayes, 2018; Schweitzer et al., 2008). That work, however, has focused primarily at the plant genotype level, and thus the identity of the causative genes remains largely unresolved. Previous studies have identified a limited set of *Populus* genes associated with insect damage (*e.g.*, mines, galls, herbivory, etc.; Dewoody et al. 2013, Zinkgraf et al. 2016) and insect community metrics (*e.g.*, abundance, richness; Bernhardsson et al. 2013). Our research advances the discipline by identifying both genes and gene functions that underlie ecologically relevant tree traits and insect communities, and exploring relationships between insect-associated genes and tree traits. Most notably, our research identified a gene that is involved in controlling a complex community trait (*e.g.*, NMDS axes), which to our knowledge is the first of its kind. Our findings also reveal several new genes associated with variation in defense compounds, including salicinoid phenolic glycosides. In addition, we identified ten new associations for variation in the abundance and incidence of leaf-modifying insects and ants. Third, our results indicate that the observed effects of many of the insect-associated genes are explained by variation in tree traits, including phytochemistry, individual leaf area, timing of bud break, and tree size. Finally, we

identified gene functions and processes that are associated with tree size, growth, and leaf morphology, and leaf-galling, leaf-rolling, and free-feeding insects, ants, and insect community composition.

Genes underlying tree traits

Our phytochemical traits had high heritability ($H^2 > 0.6$), which likely explains why we were able to identify some of the underlying genes (Beavis, 1994). Both tremulacin and total phenolic glycosides (combined tremulacin and salicortin levels) were associated with an ASC1-like protein, which is involved in protein translation and potential regulation by abscisic acid (Guo et al., 2011). In addition, total defense phytochemistry (combined condensed tannin and phenolic glycoside levels) was associated with a ribosomal protein (*Potra007960g26067*) that has a physical relationship with a syntaxin protein (*SYP121*). This syntaxin protein is regulated by various hormones, including jasmonic acid, abscisic acid, and salicylic acid, and the protein is involved in programed cell death and defense response to pathogens. Thus, both these newly identified defense-related genes are involved in altering gene expression and are regulated by plant hormones. These findings shed important light on the phytochemical pathway responsible for production of salicinoids, which remains largely unresolved (Boeckler, Gershenzon, & Unsicker, 2011; Bresadola et al., 2018; Woolbright et al., 2018; J. Zhang et al., 2018).

In addition to identifying specific genes showing significant associations with trait variation, we also observed an enrichment of gene functions in the top 0.1% of all SNPs. Tree growth and size traits were enriched for cellular response to misfolded proteins, which is essentially a response to environmental stress (*e.g.*, heat, cold, UV; Nakajima and Suzuki 2013). Specific leaf area was enriched for defense ("immune") response, and the density of

extra-floral nectaries was enriched for glyceraldehyde-3-phosphate metabolism, which is involved in glycolysis.

Genes underlying insect communities

We predicted that insects (*e.g.*, gallers) with the closest associations with the host tree would have more genetic correlations than insects (*e.g.*, free feeders and ants) that are not closely associated. Our findings support this prediction in that leaf-galling, -mining, and -rolling insects were associated with aspen genes, while free feeding insects had no significant genetic associations. However, aphid-tending ants had several SNP associations in three different genes. Those results indicate that insects in higher trophic levels may be influenced by plant genetics, and are similar to the findings of Wimp et al. (2005). Of course, our results derive from a single experimental garden; insect communities and gene-insect associations may vary in natural aspen habitats.

Upon insect herbivory, plants experience damage-induced ion imbalances, which lead to differing cell membrane potentials, calcium signaling, and oxidative stress (Maffei, Mithöfer, & Boland, 2007). These events alter kinase and phytohormone activity, which then influences gene expression (*e.g.*, altering the ratio of JAZ to DELLA proteins which can activate or suppress growth promoting genes, Maffei et al. 2007). Our insect community GWAS and gene ontology analyses revealed several candidate genes with functions that are consistent with this series of plant-insect events.

First, variation in Tenthredinidae sawflies (primarily leaf-folding *Phyllocolpa* sp.) was associated with genes enriched for sequestering iron ions (that have also been shown to respond to reactive oxygen species [Ravet et al. 2009]) possibly due to damage-induced ion

imbalances. Second, variation in leaf blotch miner (*Phyllonorycter tremuloidiella*) incidence was correlated with a vesicle transport protein (*Potra000892g07232*), which is physically located near a calmodulin-binding NAC protein (*NTL9*) and may influence calcium signaling. Third, variation in ant (*Lasius neoniger*) incidence and abundance was related to a tocopherol gene (*Potra002557g19270*), which responds to oxidative stress (Porfirova, Bergmuller, Tropf, Lemke, & Dormann, 2002). Fourth, variation in insects (incidence and composition) was related to genes involved in plant hormone regulation. For instance, petiole-galler (*Ecteodemia populella*) incidence was related to an abscisic acid receptor (*Potra001062g09110*), and insect community composition was enriched for gene ontology terms involved in hormone regulation. Fifth, both petiole-gallers and ants were associated with genes involved in modifying expression of other genes, including *Potra001062g09111* (transcriptional silencing via DNA methylation), *Potra003266g21171* (mRNA splicing factor that responds to biotic stress, Shang et al. 2017), and *AT2G40435* (*A. thaliana* homolog, transcription factor SCREAM-like protein that is involved in response to environmental stress, Liu et al. 2007).

We identified both enriched gene functions and SNPs associated with leaf-galling insects. Nabity et al. (2013) compared gene expression patterns in leaf gall tissue (*Daktulosphaira vitifoliae* galls on grape leaves) to those of regular leaf tissue. They revealed that leaf galls had upregulated the phenylpropanoid pathway, increased anthocyanin production and cell wall biogenesis, changed the expression of glycolysis/cellular respiration, and down-regulated the Calvin cycle. These findings are consistent with our gene enrichment analysis; genes associated with leaf-galling insects were enriched for flavonoid and anthocyanin biosynthesis, cell wall biogenesis, cellular respiration, and the tricarboxylic acid cycle. Also, both the *Harmandia* leaf galls and *Ecteodemia populella* petiole galls on our

trees were strongly colored with red/purple-pigment, suggesting the presence of anthocyanins. In addition, we identified an apoptosis-inducing factor, *Potra002833g20082*, that was associated with Cecidomyiidae (leaf-galling *Harmandia* flies and leaf-rolling midges), which may confer resistance to gall formation.

We also identified both enriched gene functions and SNPs associated with leaf-rolling insects. Variation in leaf-rolling insect (Tortricidae) abundance was related to genes involved in mechanical stimulus response and defense. For example, variation in rustylined leaftiers (*Clostera albosigma*, Notodontidae) was correlated with a *g*lycerol kinase (*A. thaliana* homolog, *AT4G38225*), which is involved in glycolysis. Glycolysis produces precursors for the shikimic acid pathway, which synthesizes secondary plant compounds such as condensed tannins and phenolic glycosides. In addition, variation in obliquebanded leafroller (*Choristoneura rosaceana*, Tortricidae) abundance was associated with *NOXY2* (*NONRESPONDING TO OXYLIPINS 2, AT5G11630*), a salicylic acid responsive gene that has been shown to play a role in defense against bacterial pathogens (Vellosillo et al., 2013).

Both insect community composition and variation in ant (Formicidae) abundance and incidence were associated with Gamma-aminobutyric acid (GABA) signaling (via enriched gene ontology). GABA is a non-proteinaceous amino acid that occurs in animals, plants, and bacteria (Ramesh, Tyerman, Gilliham, & Xu, 2017). In animals, GABA functions as one of the major inhibitory neurotransmitters in the central nervous system (Ramesh et al., 2017). In plants, GABA signaling influences plant growth, development, stress response, and long-distance transport (Ramesh et al., 2017). In addition, GABA signaling plays a role in insect resistance (Bown, MacGregor, & Shelp, 2006; Scholz, Reichelt, McKonnen, Ludewig, &

Mithofer, 2015), since insect-consumed GABA acts an inhibitory neurotransmitter, causing physiological stress in the insect that decreases its growth and survival (Ramesh et al., 2017).

Insect community composition (NMDS of common insect species) was also associated with an early nodulin-like (ENODL) transmembrane protein (*Potra001060g09097)*, that is thought to transport carbohydrates (Denancé, Szurek, & Noël, 2014). Wang et al. (2015) identified three ENODL proteins that putatively increased Bt rice resistance to brown planthopper infestation, thereby suggesting that ENODL proteins may influence plant-insect interactions. Here, we show that allelic variation in an ENODL gene influences insect community species diversity and the abundance of interacting foundation species: aphids and tending ants (Barker et al., 2018; Keith, Bailey, & Whitham, 2010; Lamit et al., 2015; Wimp & Whitham, 2012), consistent with the notion of an Ecologically Important Gene (EIG; Skovmand et al., 2018). The mechanism by which ENODL gene variation may influence insects in our system remains unknown, but variation in carbohydrate transport could directly influence aphids and ants via their interactions with carbohydrate-rich honeydew, and/or indirectly influence insects via numerous tree traits, including both growth (size) and defense. To our knowledge, this is the first identification of allelic variation in a plant gene that is associated with a complex insect community trait (*i.e.*, insect community composition).

Plant trait variation shapes insect communities

While we found no overlap in SNP associations across tree traits and insect phenotypes (potentially due to the limited number of tree trait-associated SNPs), our covariate analyses revealed that several tree traits explain significant insect SNP associations. These results suggest that the tree traits are important in structuring the associated insect

communities, thereby providing a mechanistic link by which plant genes shape insect community composition (*e.g.*, Bangert et al. 2006). Previous work in *Populus* and *Salix* has also shown that canopy insect communities are shaped by these particular plant traits, including plant size (Barbour et al., 2015; Barker et al., 2018; Evans et al., 2016; Robinson et al., 2012), individual leaf area (Robinson et al. 2012), timing of bud break (Barker et al., 2018; Evans et al., 2016; Falk, 2017), and defense phytochemistry (Barbour et al., 2015; Barker et al., 2018; Brito, 2017; Wimp et al., 2007). For instance, larger plants and plants with larger leaves and longer petioles have denser and more diverse insect communities (Barker et al., 2018; Robinson et al., 2012). Timing of bud phenology *differentially* affects the incidence of leaf-modifying insect species (Barker et al., 2018). Also, variation in defense phytochemistry corresponds with variation in insect communities (Bangert et al., 2006; Wimp et al., 2007). We recognize that our work was conducted in a young aspen plantation with immature trees. Insect communities are likely richer, and the plant traits structuring those communities may be different, in mature, closed-canopy aspen forests.

Previously identified *Populus* genes that correlate with insect metrics have been linked to plant traits, including defense and leaf morphology. Nine of the 13 insect-associated QTL found by Dewoody et al. (2013) contained shikimate-phenylpropanoid pathway genes (both phenolic glycosides and condensed tannins are products of that pathway), and two of the QTL were in genomic hot-spots for leaf morphology. In addition, Bernhardsson et al. (2013) found overlap between insect-associated SNPs and inducible defense genes (polyphenol oxidases and trypsin inhibitors).

Inclusion of tree trait covariates in our GWA models did not eliminate all of the ant-related genes: the SNPs in the mRNA splicing factor (*Potra003266g21171)* that responds to biotic stress remain. This implies that (1) tree traits that we did not survey or (2) more likely, biotic/environmental factors (*e.g.,* aphid populations; most of our ant species were tending aphid colonies), structure ant incidence and abundance.

While we have focused on bottom-up mechanisms underlying insect communities on aspen, we recognize that top-down factors (*e.g.*, predation, parasitism, etc.) also influence insect herbivores (van Veen, Morris, & Godfray, 2006; Katano, Doi, Eriksson, & Hillebrand, 2015; Vidal & Murphy, 2018). A meta-analysis by Vidal and Murphy (2017) revealed that top-down forces are often stronger than bottom-up effects in influencing insect herbivore fitness. Yet, these differences varied across insect groups (*e.g.*, specialist vs. generalist insects and across feeding guilds). Thus, our limited association of plant genes to insect metrics likely derived in part from the effects of other ecological interactions on structuring these insect communities.

Gene coverage

Our genetic dataset included 56% of the *P. tremuloides* genes, a coverage rate that reduced our ability to detect significant associations for some traits. To determine the extent to which our probe design included or excluded genes that are known to influence particular traits, we used Knetminer (Knowledge Network Miner, Hassani-Pak 2017) with *P. trichocarpa* homologs to compare our gene list to lists of genes that are associated with particular trait search terms (*e.g.,* "phenylpropanoid pathway", "biomass", "SLA", "insect", etc.). Our gene list covered 54-71% of the genes associated with tree traits and insect resistance (Table S5). Although our probe set did not include every gene, our findings

nonetheless reveal new gene associations that underlie both *Populus* traits and canopy insect communities. In comparison, most previous *Populus* GWAS studies captured smaller sets of genes (1,233 to 18,153 genes) with fewer SNPs (1,233 to 77,000 SNPs; McKown et al. 2014, Du et al. 2015, Hallingbäck et al. 2016, Fahrenkrog et al. 2017,). Several recent studies of *Populus* have employed greater coverage, but only examined individual tree traits (*e.g.*, budbreak (McKown, Klápště, Guy, El-Kassaby, & Mansfield, 2018) and lignin biosynthesis (Zhang et al., 2018)).

Conclusions

Over the last 15 years, community genetics perspectives linking plant intraspecific genetic variation to associated community metrics have garnered considerable attention in the literature of evolutionary ecology. Most of the relevant empirical studies, however, have been conducted at the level of plant genotypes, leaving the underlying genes unresolved. Here, we identified ten new *Populus* genes that structured associated insect communities, complementing the previously identified list of QTL from Bernhardsson et al. (12 SNPs, 2013) and DeWoody et al. (14 QTL, 2014). Our findings also reveal that ecologically-relevant plant traits structure gene-insect associations, highlighting the importance of these traits as the mechanistic bridge between plant genes and insect communities (Barbour et al., 2015; Barker et al., 2018; Harrison et al., 2018; Robinson et al., 2012).

Genetic variation in expression of key plant traits is influenced by both plant ontogeny and environmental context (Lindroth & St. Clair, 2013). Future work should address how plant genetic contributions to insect community organization may shift across plant ontogenetic trajectories (Gosney et al., 2014; Holeski, Hillstrom, Whitham, & Lindroth, 2012) and environmental gradients (Burkle, Souza, Genung, & Crutsinger, 2013).

Acknowledgements

This research was funded by the National Science Foundation (grants DGE-1256259 and DEB-0841609), and by the USDA National Institute of Food and Agriculture (McIntire-Stennis project WIS01651 and AFRI grant 20167013-25088). We thank A. Ahmad, M. Berkvam, A. Coakley, A. Flansburg, E. Gryzmala, A. Helm, R. Hughes, S. Lawrence, J. Lema, K. Keefover-Ring, D. Kuehn, D. Mahoney, M. Reitz, K. Sandel, E. Schwartz, A. Staples, A. Wenman, and M. Wrobbel for their assistance with field and lab work, S. Schoville, B. Payseur, and C. Gratton for comments on experimental approach and data analysis, and C. Cole and three anonymous reviewers for comments on the manuscript.

References

Abreu, I. N., Ahnlund, M., Moritz, T., & Albrectsen, B. R. (2011). UHPLC-ESI/TOFMS determination of salicylate-like phenolic glycosides in *Populus tremula* leaves. *Journal of Chemical Ecology*, *37*(8), 857–870. doi:10.1007/s10886-011-9991-7

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–64. doi:10.1101/gr.094052.109

Bangert, R. K., Turek, R. J., Rehill, B., Wimp, G. M., Schweitzer, J. A., Allan, G. J., … Whitham, T. G. (2006). A genetic similarity rule determines arthropod community structure. *Molecular Ecology*, *15*(5), 1379–91. doi:10.1111/j.1365-294X.2005.02749.x

Barbour, M. A., Fortuna, M. A., Bascompte, J., Nicholson, J. R., Julkunen-Tiitto, R., Jules, E. S., & Crutsinger, G. M. (2016). Genetic specificity of a plant-insect food web: Implications for linking genetic variation to network complexity. *Proceedings of the National Academy of Sciences*, *113*(8), 2128–33. doi:10.1073/pnas.1513633113

Barbour, M. A., Rodriguez-Cabal, M. A., Wu, E. T., Julkunen-Tiitto, R., Ritland, C. E.,

Miscampbell, A. E., … Crutsinger, G. M. (2015). Multiple plant traits shape the genetic basis of herbivore community assembly. *Functional Ecology*, *29*, 995–1006. doi:10.1111/1365-2435.12409

Barker, H. L., Holeski, L. M., & Lindroth, R. L. (2018). Genotypic variation in plant traits shapes herbivorous insect and ant communities on a foundation tree species. *PLOS ONE*, *13*(7), e0200954. doi:10.1371/journal.pone.0200954

Barker, H. L., Riehl, J. F., Bernhardsson, C. , Rubert-Nason, K. F., Holeski, L. M., Ingvarsson, P. K., & Lindroth, R. L. (2019). Data from: Linking plant genes to insect communities: identifying the genetic bases of plant traits and community composition. *Dryad Digit Repos*. doi: 10.5061/dryad.fr045hv

Bass, A., Storey, J., Dabney, A., & Robinson, D. (2015). qvalue: Q-value estimation for false discovery rate control. Retrieved from http://github.com/jdstorey/qvalue

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Beavis. (1994). The power and deceit of QTL experiments: Lessons from comparative QTL studies. *Proceedings of the Forty-Ninth Annual Corn and Sorghum Industry Research Conference*, 250–266.

Bernhardsson, C., Robinson, K. M., Abreu, I. N., Jansson, S., Albrectsen, B. R., & Ingvarsson, P. K. (2013). Geographic structure in metabolome and herbivore community co-occurs with genetic structure in plant defence genes. *Ecology Letters*, *16*(6), 791–8. doi:10.1111/ele.12114

Boeckler, G. A., Gershenzon, J., & Unsicker, S. B. (2011). Phenolic glycosides of the Salicaceae and their role as anti-herbivore defenses. *Phytochemistry*, *72*(13), 1497–509. doi:10.1016/j.phytochem.2011.01.038

Bown, A., MacGregor, K., & Shelp, B. (2006). Gamma-aminobutyrate: Defense against

invertebrate pests? *Trends in Plant Science*, *11*(9), 424–427.

Boyle, E. I., Weng, S., Gollub, J., Jin, H., Botstein, D., Cherry, J. M., & Sherlock, G. (2004). GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics*, *20*(18), 3710–3715. doi:10.1093/bioinformatics/bth456

Bresadola, L., Caseys, C., Castiglione, S., Buerkle, C. A., Wegmann, D., & Lexer, C. (2018). Admixture mapping in interspecific *Populus* hybrids identifies classes of genomic architectures for phytochemical, morphological and growth traits. *BioRxiv*, 1–28. doi:10.1101/497685.

Brito, A. R. (2017). *Spatial and temporal variations in weather and host-plant quality affect arthropod herbviore populations in aspen canopies*. University of Wisconsin - Madison.

Burkle, L. A., Souza, L., Genung, M. A., & Crutsinger, G. M. (2013). Plant genotype, nutrients, and G × E interactions structure floral visitor communities. *Ecosphere*, *4*(9), 1–20. doi:10.1890/ES13-00039.1

Callahan, C. M., Rowe, C. A., Ryel, R. J., Shaw, J. D., Madritch, M. D., & Mock, K. E. (2013). Continental-scale assessment of genetic diversity and population structure in quaking aspen (*Populus tremuloides*). *Journal of Biogeography*, *40*(9), 1780–1791. doi:10.1111/jbi.12115

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi:10.1093/bioinformatics/btr330

Denancé, N., Szurek, B., & Noël, L. D. (2014). Emerging functions of nodulin-like proteins in non-nodulating plant species. *Plant and Cell Physiology*, *55*(3), 469–474. doi:10.1093/pcp/pct198

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V, Maguire, J. R., Hartl, C., … Daly,

M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491. Retrieved from https://doi.org/10.1038/ng.806

Dewoody, J., Viger, M., Lakatos, F., Tuba, K., Taylor, G., & Smulders, M. J. M. (2013). Insight into the genetic components of community genetics: QTL mapping of insect association in a fast-growing forest tree. *PLOS ONE*, *8*(11), e79925. doi:10.1371/journal.pone.0079925

Du, Q., Gong, C., Wang, Q., Zhou, D., Yang, H., Pan, W., … Zhang, D. (2016). Genetic architecture of growth traits in Populus revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytologist*, *209*(3), 1067–82. doi:10.1111/nph.13695

Evans, L. M., Kaluthota, S., Pearce, D. W., Allan, G. J., Floate, K., Rood, S. B., & Whitham, T. G. (2016). Bud phenology and growth are subject to divergent selection across a latitudinal gradient in *Populus angustifolia* and impact adaptation across the distributional range and associated arthropods. *Ecology and Evolution*, *6*(13), 4565–4581. doi:10.1002/ece3.2222

Fahrenkrog, A. M., Neves, L. G., Resende, M. F. R., Vazquez, A. I., de los Campos, G., Dervinis, C., … Kirst, M. (2017). Genome-wide association study reveals putative regulators of bioenergy traits in *Populus deltoides*. *New Phytologist*, *213*(2), 799–811. doi:10.1111/nph.14154

Falk, M. A. (2017). Effects of genetic variation in temporal foliar chemical patterns on phenological relationships between trembling aspen *(Populus tremuloides* Michx.*)* and gypsy moth *(Lymantria dispar* L.*)*. University of Wisconsin - Madison.

Gene Ontology Consortium, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., … Gene Ontology Consortium. (2004). The Gene Ontology (GO) database and

informatics resource. *Nucleic Acids Research*, *32*(90001), 258D–261. doi:10.1093/nar/gkh036

Goh, L. & Yap, V. B. (2009). Effects of normalization on quantitative traits in association test. *BMC Bioinformatics, 10*, 415. doi: 10.1186/1471-2105-10-415

Gosney, B. J., O Reilly-Wapstra, J. M., Forster, L. G., Barbour, R. C., Iason, G. R., Potts, B. M., … Potts, B. M. (2014). Genetic and ontogenetic variation in an endangered tree structures dependent arthropod and fungal communities. *PLOS ONE*, *9*(12), e114132. doi:10.1371/journal.pone.0114132

Guo, J., Wang, S., Valerius, O., Hall, H., Zeng, Q., Li, J.-F., … Chen, J.-G. (2011). Involvement of *Arabidopsis* RACK1 in protein translation and its regulation by abscisic acid. *Plant Physiology*, *155*(1), 370–383. doi:10.1104/pp.110.160663

Hagerman, A. E., & Butler, L. G. (1980). Condensed tannin purification and characterization of tannin-associated proteins. *Journal of Agricultural and Food Chemistry*, *28*(5), 947–952. doi:10.1021/jf60231a011

Hallingbäck, H. H. R., Fogelqvist, J., Powers, S. S. J., Turrion-Gomez, J., Rossiter, R., Amey, J., … Rönnberg-Wästljung, A.-C. (2016). Association mapping in *Salix viminalis* L. (Salicaceae) - identification of candidate genes associated with growth and phenology. *GCB Bioenergy*, *8*(3), 670–685. doi:10.1111/gcbb.12280

Harrison, J. G., Philbin, C. S., Gompert, Z., Forister, G. W., Hernandez-Espinoza, L., Sullivan, B. W., … Forister, M. L. (2018). Deconstruction of a plant-arthropod community reveals influential plant traits with nonlinear effects on arthropod assemblages. *Functional Ecology*, 1–12. doi:10.1111/1365-2435.13060

Hassani-Pak, K. (2017). *KnetMiner - An integrated data platform for gene mining and biological knowledge discovery*. Bielefeld University. Retrieved from https://pub.uni-bielefeld.de/publication/2915227

Hillstrom, M. L. (2009). *Effects of elevated carbon dioxide and ozone on forest insect abundance, diversity, and community composition*. University of Wisconsin, Madison.

Holeski, L. M., Hillstrom, M. L., Whitham, T. G., & Lindroth, R. L. (2012). Relative importance of genetic, ontogenetic, induction, and seasonal variation in producing a multivariate defense phenotype in a foundation tree species. *Oecologia*, *170*(3), 695–707. doi:10.1007/s00442-012-2344-6

Ingvarsson, P. K., & Street, N. R. (2011). Association genetics of complex traits in plants. *New Phytologist*, *189*, 909–922. doi:10.1111/j.1469-8137.2010.03593.x

Kagiya, S., Yasugi, M., Kudoh, H., Nagano, A. J., & Utsumi, S. (2018). Does genomic variation in a foundation species predict arthropod community structure in a riparian forest? *Molecular Ecology*, 1–12. doi:10.1111/mec.14515

Katano, I., Doi, H., Eriksson, B. K., & Hillebrand, H. (2015). A cross-system meta-analysis reveals coupled predation effects on prey biomass and diversity. *Oikos*, *124*(11), 1427–1435. doi:10.1111/oik.02430

Keith, A. R., Bailey, J. K., Lau, M. K., & Whitham, T. G. (2017). Genetics-based interactions of foundation species affect community diversity, stability and network structure. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1854), 20162703. doi:10.1098/rspb.2016.2703

Keith, A. R., Bailey, J. K., & Whitham, T. G. (2010). A genetic basis to community repeatability and stability. *Ecology*, *91*(11), 3398–3406. doi:10.1890/09-1236.1

Koricheva, J., & Hayes, D. (2018). The relative importance of plant intraspecific diversity in structuring arthropod communities: A meta-analysis. *Functional Ecology*, 1–14. doi:10.1111/1365-2435.13062

Lamit, L. J., Busby, P. E., Lau, M. K., Compson, Z. G., Wojtowicz, T., Keith, A. R., … Whitham, T. G. (2015). Tree genotype mediates covariance among communities from

microbes to lichens and arthropods. *Ecology*, *103*, 840–850. doi:10.1111/1365-2745.12416

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–60. doi:10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Sam, T. (2009). The sequence alignment / map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. doi:10.1093/bioinformatics/btp352

Lin, Y., Wang, J., Delhomme, N., Schiffthaler, B., Sundström, G., & Zuccolo, A. (2018). Functional and evolutionary genomic inferences in *Populus* through genome and population sequencing of American and European aspen. *Proceedings of the National Academy of Sciences*, *115*(46), E10970–E10978. doi:10.1073/pnas.1801437115

Lindroth, R. L., & St. Clair, S. B. (2013). Adaptations of quaking aspen (*Populus tremuloides* Michx.) for defense against herbivores. *Forest Ecology and Management*, *299*, 14–21. doi:10.1016/j.foreco.2012.11.018

Liu, J.-X., Srivastava, R., Che, P., & Howell, S. H. (2007). Salt stress responses in *Arabidopsis* utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. *The Plant Journal*, *51*(5), 897–909. doi:10.1111/j.1365-313X.2007.03195.x

Maffei, M. E., Mithöfer, A., & Boland, W. (2007). Before gene expression: early events in plant–insect interaction. *Trends in Plant Science*, *12*(7), 310–316. doi:10.1016/J.TPLANTS.2007.06.001

Mckenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … Depristo, M. A. (2010). The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome*, *20*, 1297–1303. doi:10.1101/gr.107524.110.20

McKown, A. D., Klápště, J., Guy, R. D., El-Kassaby, Y. A., & Mansfield, S. D. (2018).

Ecological genomics of variation in bud-break phenology and mechanisms of response to climate warming in *Populus trichocarpa*. *New Phytologist*, *220*(1), 300–316. doi:10.1111/nph.15273

McKown, A. D., Klápště, J., Guy, R. D., Geraldes, A., Porth, I., Hannemann, J., … Douglas, C. J. (2014). Genome-wide association implicates numerous genes underlying ecological trait variation in natural populations of *Populus trichocarpa*. *New Phytologist*, *203*(2), 535–553. doi:10.1111/nph.12815

Mitton, J. B., & Grant, M. C. (1996). Genetic variation and the natural history of quaking aspen. *BioScience*, *46*(1), 25–31.

Money, D., Gardner, K., Migicovsky, Z., Schwaninger, H., Zhong, G.-Y., & Myles, S. (2015). LinkImpute: Fast and accurate genotype imputation for nonmodel organisms. *G3: GENES, GENOMES, GENETICS*, *5*(11), 2383–2390. doi:10.1534/g3.115.021667

Nabity, P. D., Haus, M. J., Berenbaum, M. R., & DeLucia, E. H. (2013). Leaf-galling phylloxera on grapes reprograms host metabolism and morphology. *Proceedings of the National Academy of Sciences*, *110*(41), 16663–8. doi:10.1073/pnas.1220219110

Nakajima, Y., & Suzuki, S. (2013). Environmental stresses induce misfolded protein aggregation in plant cells in a microtubule-dependent manner. *International Journal of Molecular Sciences*, *14*(4), 7771–83. doi:10.3390/ijms14047771

Porfirova, S., Bergmuller, E., Tropf, S., Lemke, R., & Dormann, P. (2002). Isolation of an *Arabidopsis* mutant lacking vitamin E and identification of a cyclase essential for all tocopherol biosynthesis. *Proceedings of the National Academy of Sciences*, *99*(19), 12495–500. doi:10.1073/pnas.182330899

Porter, L. J., Hrstich, L. N., & Chan, B. G. (1986). The conversion of procyanidins and prodelphinidins to cyanidin and delphinidin. *Phytochemistry*, *25*, 223–230.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham,

P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. doi:10.1086/519795

Ramesh, S. A., Tyerman, S. D., Gilliham, M., & Xu, B. (2017). γ-Aminobutyric acid (GABA) signalling in plants. *Cellular and Molecular Life Sciences*, *74*(9), 1577–1603. doi:10.1007/s00018-016-2415-7

Ravet, K., Touraine, B., Boucherez, J., Briat, J.-F., Gaymard, F., & Cellier, F. (2009). Ferritins control interaction between iron homeostasis and oxidative stress in *Arabidopsis*. *The Plant Journal*, *57*(3), 400–412. doi:10.1111/j.1365-313X.2008.03698.x

Robinson, K. M., Ingvarsson, P. K., Jansson, S., & Albrectsen, B. R. (2012). Genetic variation in functional traits influences arthropod community composition in aspen (*Populus tremula* L.). *PLOS ONE*, *7*(5), e37679. doi:10.1371/journal.pone.0037679

Rönnberg-Wästljung, A. C., Ahman, I., Glynn, C., & Widenfalk, O. (2006). Quantitative trait loci for resistance to herbivores in willow: field experiments with varying soils and climates. *Entomologia Experimentalis et Applicata*, *118*(2), 163–174. doi:10.1111/j.1570-7458.2006.00371.x

Rubert-Nason, K. F., Hedman, C. J., Holeski, L. M., & Lindroth, R. L. (2014). Determination of salicinoids by micro-high-performance liquid chromatography and photodiode array detection. *Phytochemical Analysis*, *25*(3), 185–191. doi:10.1002/pca.2485

Rubert-Nason, K. F., Holeski, L. M., Couture, J. J., Gusse, A., Undersander, D. J., & Lindroth, R. L. (2013). Rapid phytochemical analysis of birch (*Betula*) and poplar (*Populus*) foliage by near-infrared reflectance spectroscopy. *Analytical and Bioanalytical Chemistry*, *405*(4), 1333–44. doi:10.1007/s00216-012-6513-6

Rubert-Nason, K., Keefover-Ring, K., & Lindroth, R. L. (2018). Purification and analysis of

Salicinoids. *Current Analytical Chemistry*, *14*(4). doi:10.2174/1573411014666171221131933

Scholz, S., Reichelt, M., Mekonnen, D. W., Ludewig, F., & Mithofer, A. (2015). Insect herbivory-elicited GABA accumulation in plants is a wound-induced, direct, systemic, and jasmonate-independent defense response. *Frontiers in Plant Science*, *6*, 1128. doi:10.3389/fpls. 2015.01128

Schweitzer, J. a., Madritch, M. D., Bailey, J. K., LeRoy, C. J., Fischer, D. G., Rehill, B. J., … Whitham, T. G. (2008). From genes to ecosystems: The genetic basis of condensed tannins and their role in nutrient regulation in a *Populus* model system. *Ecosystems*, *11*(6), 1005–1020. doi:10.1007/s10021-008-9173-9

Shang, X., Cao, Y., & Ma, L. (2017). Alternative splicing in plant genes: A means of regulating the environmental fitness of plants. *International Journal of Molecular Sciences*, *18*(2), e432. doi:10.3390/ijms18020432

Shenk, J. S., & Westerhaus, M. O. (1991). Population definition, sample selection, and calibration procedures for near infrared reflectance spectroscopy. *Crop Science*, *31*(2), 469. doi:10.2135/cropsci1991.0011183X003100020049x

Shenk, J. S., & Westerhaus, M. O. (1991). Populations structuring of near infrared spectra and modified partial least squares regression. *Crop Science*, *31*(6), 1548. doi:10.2135/cropsci1991.0011183X003100060034x

Sjödin, A., Street, N. R., Sandberg, G., Gustafsson, P., & Jansson, S. (2009). The *Populus* Genome Integrative Explorer (PopGenIE): A new resource for exploring the *Populus* genome. *New Phytologist*, *182*(4), 1013–1025. doi:10.1111/j.1469-8137.2009.02807.x

Skovmand, L. H., Xu, C. C. Y., Servedio, M. R., Nosil, P., Barrett, R. D. H., & Hendry, A. P. (2018). Keystone genes. *Trends in Ecology & Evolution*, *33*(9), 689–700. doi:10.1016/j.tree.2018.07.002

Sollins, P., Glassman, C., Paul, E., Swanston, C., Lajtha, K., Heil, J., & Elliott, E. (1999). Soil carbon and nitrogen: Pools and fractions. In G. Robertson, D. Coleman, C. Bledsoe, & P. Sollins (Eds.), *Standard soil methods for long-term ecological research* (pp. 89–105). New York, NY: Oxford University Press.

Staff, S. S., NRCS, & USDA. (2012). Web Soil Survey. Retrieved January 1, 2012, from https://websoilsurvey.sc.egov.usda.gov/

Stevens, M. T., Waller, D. M., & Lindroth, R. L. (2007). Resistance and tolerance in *Populus tremuloides*: genetic variation, costs, and environmental dependency. *Evolutionary Ecology*, *21*(6), 829–847. doi:10.1007/s10682-006-9154-4

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., … Huala, E. (2007). The *Arabidopsis* Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Research*, *36*(Database), D1009–D1014. doi:10.1093/nar/gkm965

Thoen, M. P. M., Davila Olivas, N. H., Kloth, K. J., Coolen, S., Huang, P., Aarts, M. G. M., … Dicke, M. (2017). Genetic architecture of plant stress resistance: Multi-trait genome-wide association mapping. *New Phytologist*, *213*(3), 1346–1362. doi:10.1111/NPH.14220

Tuskan, G. A., Difazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., … Rokhsar, D. (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, *313*(5793), 1596–604. doi:10.1126/science.1128691

Tzin, V., Lindsay, P. L., Christensen, S. A., Meihls, L. N., Blue, L. B., & Jander, G. (2015). Genetic mapping shows intraspecific variation and transgressive segregation for caterpillar-induced aphid resistance in maize. *Molecular Ecology*, *24*(22), 5739–50. doi:10.1111/mec.13418

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-

Moonshine, A., … DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(1110), 11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43

van Veen, F. J., Morris, R. J., & Godfray, H. C. J. (2006). Apparent competition, quantitative food webs, and the structure of phytophagous insect communities. *Annual Review of Entomology*, *51*(1), 187–208. doi:10.1146/annurev.ento.51.110104.151120

Vellosillo, T., Aguilera, V., Marcos, R., Bartsch, M., Vicente, J., Cascon, T., … Castresana, C. (2013). Defense activated by 9-lipoxygenase-derived oxylipins requires specific mitochondrial proteins. *Plant Physiology*, *161*(2), 617–627. doi:10.1104/pp.112.207514

Venables, W. N. & Ripley, B. D. (2002). Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Vidal, M. C., & Murphy, S. M. (2018). Bottom-up vs. top-down effects on terrestrial insect herbivores: A meta-analysis. *Ecology Letters*, *21*(1), 138–150. doi:10.1111/ele.12874

Wang, F., Ning, D., Chen, Y., Dang, C., Han, N.-S., Liu, Y., & Ye, G.-Y. (2015). Comparing gene expression profiles between Bt and non-Bt rice in response to brown planthopper infestation. *Frontiers in Plant Science*, *6*, 1181. doi:10.3389/fpls.2015.01181

Wang, J., Street, N. R., Scofield, D. G., & Ingvarsson, P. K. (2016). Natural selection and recombination rate variation shape nucleotide polymorphism across the genomes of three related *Populus* species. *Genetics*, *202*(3), 1185–200. doi:10.1534/genetics.115.183152

Wimp, G. M., Martinsen, G. D., Floate, K. D., Bangert, R. K., & Whitham, T. G. (2005). Plant genetic determinants of arthropod community structure and diversity. *Evolution*, *59*(1), 61–9. doi:10.1111/j.0014-3820.2005.tb00894.x

Wimp, G. M., & Whitham, T. G. (2012). Biodiversity consequences of predation and host plant hybridization on an aphid-ant mutualism. *Ecology*, *82*(2), 440–452.

Wimp, G. M., Wooley, S., Bangert, R. K., Young, W. P., Martinsen, G. D., Keim, P., … Whitham, T. G. (2007). Plant genetics predicts intra-annual variation in phytochemistry and arthropod community structure. *Molecular Ecology*, *16*(23), 5057–69. doi:10.1111/j.1365-294X.2007.03544.x

Woolbright, S. A., Rehill, B. J., Lindroth, R. L., DiFazio, S. P., Martinsen, G. D., Zinkgraf, M. S., … Whitham, T. G. (2018). Large effect quantitative trait loci for salicinoid in *Populus*: implications for gene discovery. *Ecology and Evolution*, *8*, 3726–3737. doi:10.1002/ece3.3932

Zhang, J., Yang, Y., Zheng, K., Xie, M., Feng, K., Jawdy, S. S., … Chen, J. (2018). Genome-wide association studies and expression-based quantitative trait loci analyses reveal roles of HCT2 in caffeoylquinic acid biosynthesis and its regulation by defense- responsive transcription factors in *Populus*. *New Phytologist*, *220*, 502–516. doi:10.1111/nph.15297

Zhang, Z., Liu, J., Ding, X., Bijma, P., de Koning, D.-J., & Zhang, Q. (2010). Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLOS ONE*, *5*(9), e12648. doi:10.1371/journal.pone.0012648

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824. doi:10.1038/ng.2310

Zinkgraf, M. S., Meneses, N., Whitham, T. G., & Allan, G. J. (2016). Genetic variation in NIN1 and C/VIF1 genes is significantly associated with *Populus angustifolia* resistance to a galling herbivore, *Pemphigus betae*. *Journal of Insect Physiology*, *84*, 50–59. doi:10.1016/J.JINSPHYS.2015.10.007

Data Accessibility

WisAsp SNP plink format files, phenotype data, and insect community data: Dryad doi:10.5061/dryad.fr045hv. Raw sequence data for all samples included in this study are available through the European Nucleotide Archive under accession number PRJEB30919.

Author Contributions

HLB: Conceptualization, data curation, formal analysis, investigation, project administration, visualization, writing – original draft

JFR: Project administration, data curation, investigation, writing – review and editing

CB: Data curation, investigation, formal analysis, writing – review and editing

KR-N: Project administration, investigation, writing – review and editing

LH: Conceptualization, funding acquisition, writing – review and editing

PKI: Conceptualization, writing – review and editing

RLL: Conceptualization, funding acquisition, methodology, resources, supervision, writing – review and editing

Table 1. Summary of significant single nucleotide polymorphism (SNP) associations (false discovery rate [FDR] q-value < 0.10) for tree traits and insect phenotypes for the WisAsp *Populus tremuloides* genetic mapping population (N = 445 genets for tree traits, N = 328 genets for insect traits). Genome-wide association models were analyzed in PLINK without a kinship matrix. Bolded alleles are the minor allele.

										Homologous gene			
Trait	***P. tremula* (v1.1)**	**SNP position**	**Alleles**		**MAF‡**	**P-value**	**FDR q-value**	**Beta (SE)§**	**R^2**	***P. tremuloides* (v1.1)**	***P. trichocarpa* (v3.0)**	***A. thaliana***	**Gene annotation**
Tremulacin	Potra003979g23949	37860	**A**	C	0.054	2.95E-07	0.025	0.792 (0.152)	0.059	Potrs002611g30592	Potri.004G141000	AT3G19260	ASC1-like protein
		38233	**G**	T	0.054	2.95E-07	0.025	0.792 (0.152)	0.059				
Phenolic glycosides*	Potra003979g23949	37860	**A**	C	0.054	2.50E-07	0.021	0.798 (0.152)	0.059	Potrs002611g30592	Potri.004G141000	AT3G19260	ASC1-like protein
		38233	**G**	T	0.054	2.50E-07	0.021	0.798 (0.152)	0.059				
Total defense chemistry**	Potra007960g26067	4119	**A**	G	0.462	3.93E-07	0.066	-0.341 (0.066)	0.057	Potrs039467g25117	Potri.001G218700	AT3G52580	40S ribosomal protein
Ecteodemia populella (P/A)†	Potra001062g09110	18542	**C**	T	0.059	5.64E-07	0.062	-0.887 (0.173)	0.081	Potrs042256g26529	Potri.018G046100	AT2G25770	Abscisic acid receptor
	Potra001062g09111	28996	**T**	A	0.062	1.46E-06	0.062	-0.835 (0.170)	0.075	Potrs007988g11232	Potri.018G046200	AT1G15215	Protein SAWADEE HOMEODOMAIN HOMOLOG
		29108	**A**	C	0.062	1.46E-06	0.062	-0.835 (0.170)	0.075				
		29109	**T**	A	0.062	1.46E-06	0.062	-0.835 (0.170)	0.075				
		29512	**T**	C	0.054	1.85E-06	0.063	-0.881 (0.081)	0.074				

Phyllonorycter tremuloidiella (P/A)†	Potra000892g07232	177159	**C**	T	0.249	1.70E-06	0.098	-0.459 (0.094)	0.074	Potrs009143g14209	Potri.007G124400	AT5G01430	Vesicle transport protein
		177288	**C**	T	0.249	1.70E-06	0.098	-0.459 (0.094)	0.074				
		177289	**A**	G	0.249	1.70E-06	0.098	-0.459 (0.094)	0.074				
Clostera albosigma	Potra001567g12957	32991	**T**	C	0.241	4.53E-07	0.078	-0.471 (0.091)	0.082	Potrs004442g06451	Potri.004G207400	AT4G38225	Uncharacterized protein LOC103933534 isoform
Clostera albosigma (P/A)†		32991	**T**	C	0.241	5.63E-07	0.097	0.468 (0.091)	0.081				
Choristoneura rosaceana	Potra000544g03858	259777	**C**	T	0.436	9.95E-07	0.098	-0.460 (0.092)	0.078	Potrs019807g22433	Potri.006G237800	AT5G11630	Uncharacterized protein LOC105122277
		262984	**A**	T	0.435	1.13E-06	0.098	-0.460 (0.093)	0.077				
Cecidomyiidae	Potra002833g20082	5929	**T**	C	0.043	3.30E-07	0.056	-1.076 (0.206)	0.084	Potrs005298g07564	Potri.001G217800	AT3G44190	Apoptosis-inducing factor
Cecidomyiidae (P/A)†		5929	**T**	C	0.043	3.83E-07	0.066	1.069 (0.206)	0.083				
Tortricidae	Potra000544g03858	259777	**C**	T	0.436	1.02E-06	0.100	-0.460 (0.092)	0.077	Potrs019807g22433	Potri.006G237800	AT5G11630	Uncharacterized protein LOC105122277
	Potra000544g03859	262984	**A**	T	0.435	1.15E-06	0.100	-0.460 (0.093)	0.077	Potrs042389g26631	Potri.002G045500	AT5G43680	Sec-independent protein translocase protein TatB
Lasius neoniger	Potra003266g21171	16319	**A**	T	0.305	1.93E-06	0.042	0.387 (0.080)	0.074	Potrs017980g20216	Potri.001G287600	AT3G24730	Thioredoxin-like protein
		16337	**A**	G	0.305	1.93E-06	0.042	0.387 (0.080)	0.074				
		16338	**G**	A	0.305	2.75E-06	0.053	0.378 (0.079)	0.071				

		16381	**A**	G	0.303	1.67E-06	0.042	0.391 (0.080)	0.074				
		16473	**T**	C	0.303	1.67E-06	0.042	0.391 (0.080)	0.074				
		16480	**G**	A	0.303	1.67E-06	0.042	0.391 (0.080)	0.074				
		16548	**G**	A	0.300	1.35E-06	0.042	0.400 (0.081)	0.076				
		16549	**G**	A	0.300	1.35E-06	0.042	0.400 (0.081)	0.076				
	Potra002557 g19270[β]	32237	**G**	A	0.119	4.85E-06	0.070	0.537 (0.115)	0.068	Potrs004787g0 6948	Potri.018G040200	AT4G32770	Tocopherol cyclase, chloroplastic
		32326	**T**	C	0.119	4.85E-06	0.070	0.537 (0.115)	0.068				
		32327	**C**	A	0.119	4.85E-06	0.070	0.537 (0.115)	0.068				
	Potra003286 g21239	28246	**T**	C	0.071	9.33E-07	0.042	-0.771 (0.154)	0.078	Potrs009219g1 4285	Potri.019G126800	AT2G40435	Uncharacterized protein LOC105123201
Lasius neoniger (P/A)[†]	Potra003266 g21171	16319	**A**	T	0.305	1.27E-06	0.027	-0.393 (0.079)	0.076	Potrs017980g2 0216	Potri.001G287600	AT3G24730	Thioredoxin-like protein
		16337	**A**	G	0.305	1.27E-06	0.027	-0.393 (0.079)	0.076				
		16338	**G**	A	0.305	1.82E-06	0.035	-0.393 (0.079)	0.074				
		16381	**A**	G	0.303	1.15E-06	0.027	-0.397 (0.080)	0.077				
		16473	**T**	C	0.303	1.15E-06	0.027	-0.397 (0.080)	0.077				

	16480	**G**	A	0.303	1.15E-06	0.027	-0.397 (0.080)	0.077				
	16548	**G**	A	0.300	9.17E-07	0.027	-0.406 (0.081)	0.078				
	16549	**G**	A	0.300	9.17E-07	0.027	-0.406 (0.081)	0.078				
Potra002557g19270[β]	32237	**G**	A	0.119	4.53E-06	0.066	-0.538 (0.115)	0.068	Potrs004787g06948	Potri.018G040200	AT4G32770	Tocopherol cyclase, chloroplastic
	32326	**T**	C	0.119	4.53E-06	0.066	-0.538 (0.115)	0.068				
	32327	**C**	A	0.119	4.53E-06	0.066	-0.538 (0.115)	0.068				
Potra003286g21239	28246	**T**	C	0.071	7.79E-07	0.027	0.776 (0.154)	0.079	Potrs009219g14285	Potri.019G126800	AT2G40435	Uncharacterized protein LOC105123201

* Phenolic glycosides = combined levels of salicortin and tremulacin

** Total defense chemistry = combined levels of salicortin, tremulacin, and condensed tannins

[†] P/A = presence/absence

[‡] MAF = minor allele frequency

[§] SE = standard error

[β] = Significant SNP associations (Potra002557g19270) are removed when aphids (presence/absence BLUP) are added as a covariate in the GWAS model

Table 2. Summary of WisAsp gene set enrichment analyses for tree traits and insect phenotypes. Analyses were conducted on gene lists (top 0.1% most significant SNPs, N = 174 SNPs) from PLINK genome-wide association results. Gene sets were compared to the background set of genes that were included in our probe set (based on *Arabidopsis* homologs).

Trait category	Trait	Enriched gene ontology term	Gene set frequency	Background frequency	Bonferroni p-value	FDR[§] q-value	*Arabidopsis thaliana* homologous genes
Tree size/ growth	Spring volume	Cellular response to misfolded protein	1.9	0.0	0.07	0.12	AT4G03510, AT5G52060
	Absolute growth		1.9	0.0	0.07	0.48	
	Spring basal area		1.8	0.0	0.06	0.36	
	Average basal area		1.8	0.0	0.09	0.18	
	Average volume		1.8	0.0	0.07	0.22	
	Basal area increment	Sucrose metabolic process	3.5	0.2	0.01	0.02	AT1G22710, AT5G40390, AT3G03250, AT3G22200
Leaf morphology	Specific leaf area	Immune response process	8.3	1.8	0.04	0.06	AT1G42990, AT5G65600, AT5G64930, AT4G24290,AT5G64570, AT3G04740, AT2G39200, AT3G50930, AT5G18610, AT4G37930
		Innate immune response	8.3	1.5	0.01	0.02	AT5G65600, AT5G64930, AT4G24290, AT5G64570,AT3G04740, AT2G39200, AT3G50930, AT5G18610, AT4G37930
		Immune response	8.3	1.5	0.04	0.02	AT5G65600, AT5G64930, AT4G24290, AT5G64570,AT3G04740, AT2G39200, AT3G50930, AT5G18610, AT4G37930

	EFN* density	Glyceraldehyde-3-phosphate metabolic process	3.2	0.2	0.09	0.40	AT1G71100, AT3G55440, AT4G13830, AT2G26930
Leaf galling insects	Galling insects (P/A)†	Cellular respiration	4.9	0.5	0.05	0.32	AT1G08480, AT2G20420, AT5G51060, AT5G03860, AT5G09600
		Aerobic respiration	3.9	0.3	0.10	0.60	AT1G08480, AT2G20420, AT5G51060, AT5G03860
	Ecteodemia populella (P/A)†	Anthocyanin-containing compound biosynthetic process	3.4	0.1	0.01	0.02	AT4G01900, AT3G28430, AT5G13930, AT4G14090
		Tricarboxylic acid cycle	4.2	0.2	0.07	0.10	AT1G08480, AT2G20420, AT5G03860, AT5G62575
		Citrate metabolic process	3.4	0.2	0.07	0.10	AT1G08480, AT2G20420, AT5G03860, AT5G62575
		Anthocyanin-containing compound metabolic process	3.4	0.2	0.07	0.10	AT4G01900, AT3G28430, AT5G13930, AT4G14090
		Tricarboxylic acid metabolic process	3.4	0.3	0.09	0.10	AT1G08480, AT2G20420, AT5G03860, AT5G62575
		Flavonoid biosynthetic process	4.2	0.4	0.04	0.08	AT4G01900, AT5G48930, AT3G28430, AT5G13930, AT4G14090
	Harmandia (P/A)†	Cell wall biogenesis	6.2	1.1	0.08	0.20	AT4G03210, AT3G26370, AT1G19360, AT2G33460, AT3G08900, AT5G65270, AT3G02230
Leaf rolling insects	Tortricidae	Response to mechanical stimulus	2.4	0.1	0.06	0.14	AT1G43700, AT5G61210, AT4G35920
	Tenthredinidae (P/A)†	Intracellular sequestering of iron ion	1.9	0.0	0.08	0.26	AT2G01770, AT3G11050

		Sequestering of iron ion	1.9	0.0	0.08	0.26	AT2G01770, AT3G11050
Free feeding insects	Free feeders (P/A)†	Peroxisome organization	3.4	0.2	0.03	0.08	AT1G63900, AT3G19720, AT1G48635, AT3G18160
Ants	Formicidae	Gamma-aminobutyric acid transport	2.0	0.0	0.03	0.04	AT2G01170, AT1G08230
	Formicidae (P/A)†		2.0	0.0	0.02	0.18	
Multivariate insect traits	NMDS‡ of insect species	Cellular response to acid chemical	10.2	2.3	0.01	0.00	AT4G04720, AT5G11260, AT5G66730, AT5G64930, AT1G27320, AT2G40830, AT1G78380, AT2G18470, AT3G50530, AT3G56850, AT2G18790, AT4G34220
	NMDS‡ of insect species (P/A)†	Polyol catabolic process	2.3	0.1	0.04	0.40	AT2G43900, AT2G21170, AT4G18010
	NMDS‡ of insect families	Glycosyl compound metabolism process	5.8	0.8	0.05	0.10	AT4G24340, AT2G22840, AT5G28050, AT1G27320, AT1G03110, AT4G21760
	NMDS‡ of of insect guilds**	Gamma-aminobutyric acid transport	1.9	0.0	0.03	0.34	AT2G01170, AT1G08230
	NMDS‡ of insect guilds** (P/A)†	Gamma-aminobutyric acid transport	1.8	0.0	0.03	0.22	AT2G01170, AT1G08230
		Hormone metabolic process	7.1	1.3	0.04	0.22	AT5G06300, AT1G04610, AT5G57740, AT3G30180, AT2G28305, AT1G17420, AT1G44350, AT4G02680
		Hormone biosynthetic process	6.2	1.0	0.05	0.26	AT1G17420, AT5G06300, AT1G04610, AT4G02680, AT5G57740, AT3G30180, AT2G28305

	Regulation of hormone levels	8.0	1.8	0.08	0.32	AT5G06300, AT1G04610, AT5G57740, AT3G30180, AT2G28305, AT1G17420, AT1G44350, AT4G02680, AT5G56750
Insect guilds** (P/A)†	Xyloglucan biosynthetic process	2.2	0.1	0.02	0.02	AT5G04885, AT1G14100, AT2G03220

* EFN = extra floral nectary

** Insect guilds = leaf gallers, miners, rollers, free feeders, aphids, ants

† P/A = presence/absence

‡ NMDS = nonmetric multidimensional scaling

§ FDR = false discovery rate

Table 3. Summary of tree trait covariates that eliminate or reduce the number of significant SNPs associated with particular insect traits for the WisAsp aspen (*Populus tremuloides*) genetic mapping population (N = 328 genets for insect traits). Covariates that reduce SNP associations reveal tree traits that are important in shaping the particular insect phenotype, and are indicated by an "X" below. Genome-wide association models were analyzed in PLINK without a kinship matrix and standardized tree traits (each tree trait was analyzed in separate models).

	Tree size/growth							**Leaf morphology**			**Bud phenology**			**Phytochemistry**						
Insect trait with significant SNP associations	Spring volume	Spring basal area	Average basal area	Average volume	Absolute growth	Relative growth	Basal area increment	Specific leaf area	Individual leaf area	EFN* density	Timing of bud break	Timing of bud set	Growing season length	Condensed tannins	Salicortin	Tremulacin	Phenolic glycosides**	Total defense chemistry***	Carbon:nitrogen	Nitrogen
Ecteodemia populella (P/A)†						X														
BlotchMine (P/A)†									X	X				X	X		X	X	X	
Phyllonorycter tremuloidiella (P/A)†	X	X	X			X	X		X		X				X	X	X			
Clostera albosigma (P/A)†					X	X	X		X		X				X	X	X			
Choristoneura rosaceana		X	X	X							X	X		X	X	X	X			
Cecidomyiidae					X				X	X					X		X			
Cecidomyiidae (P/A)†									X	X	X			X	X	X	X			
Tortricidae		X	X	X			X		X		X			X	X	X	X			
Lasius neoniger									X					X	X	X	X			
Lasius neoniger (P/A)†									X						X	X	X			

* EFN = extra floral nectary

** Phenolic glycosides = combined levels of salicortin and tremulacin

*** Total defense chemistry = combined levels of salicortin, tremulacin, and condensed tannins

† P/A = presence/absence

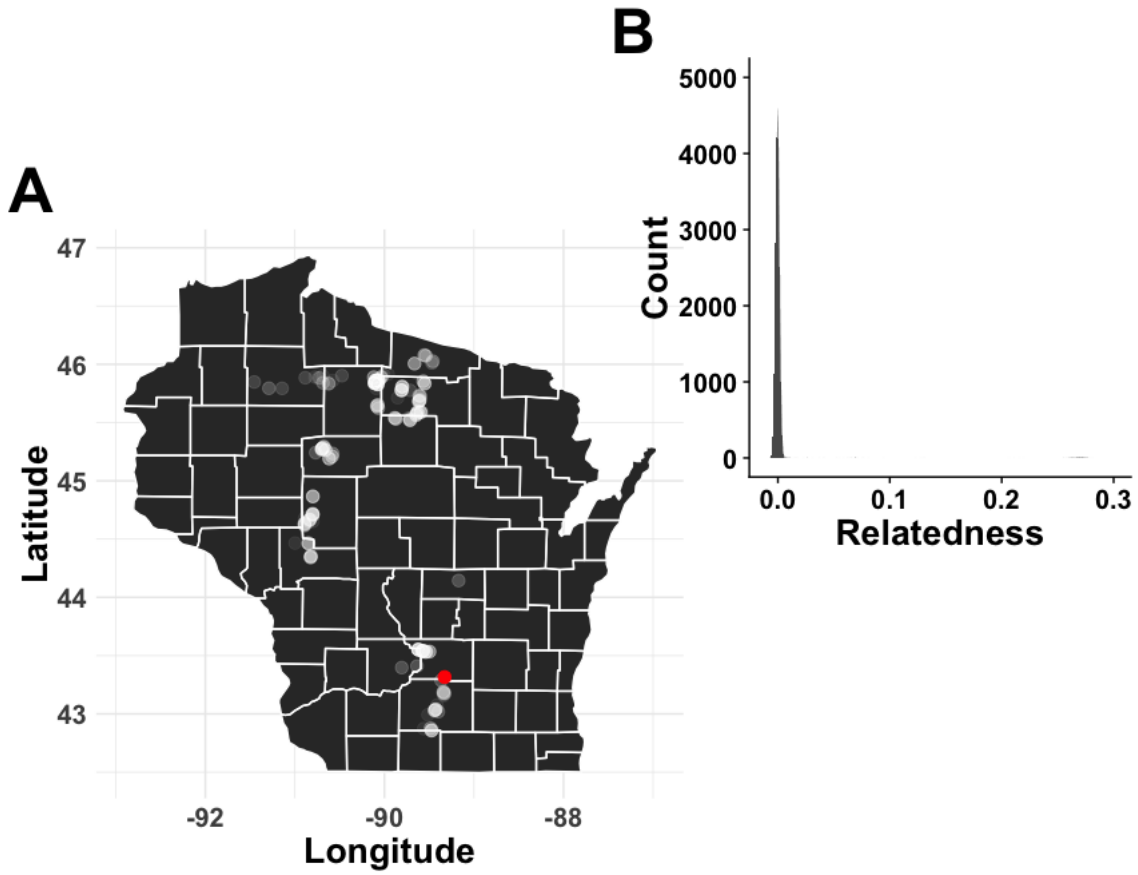
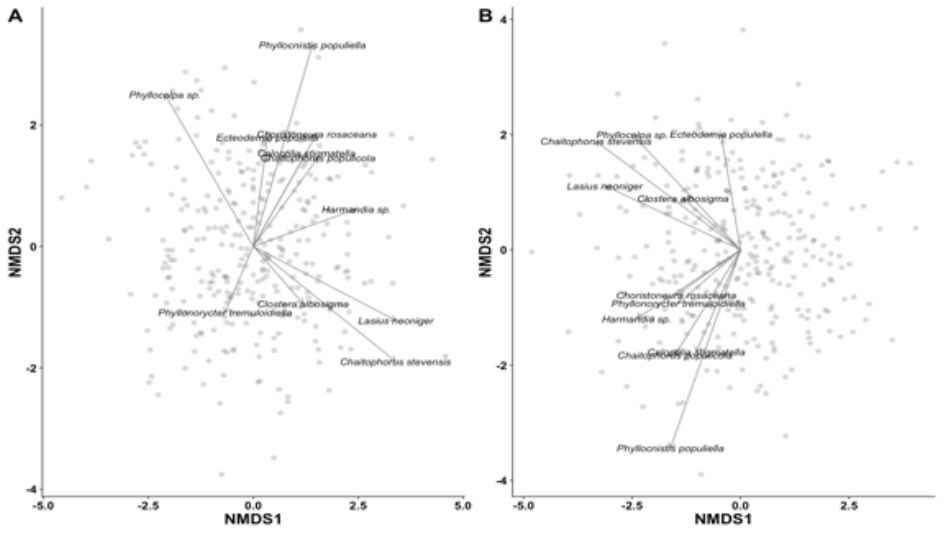


Figure 1. (A) Map of Wisconsin that displays the origin of the WisAsp aspen genets (white points, shown with 30% transparency) and the WisAsp common garden (red point). (B) Histogram of relatedness values (calculated using GEMMA, derived from a kinship matrix) for all pairwise comparisons of the 445 genets. Low values (0) indicate that the pair of genets are completely unrelated, whereas higher values (0.25) indicate that the pair of genets are more related and have a potential sibling relationship.



C

Trait	*P. tremula* (v1.1)	SNP position	Alleles	MAF	P-value	Q-value	*P. tremuloides* (v1.1)	*P. trichocarpa* (v3.0)	*A. thaliana*	Annotation	NMDS1	NDMS2	NMDS3	NMDS4	NMDS5
NMDS of insect species	Potra001060g09097	68122	C T	0.136	3.41E-07	0.029	Potrs000733g01118	Potri.003G050500	AT1G79800	nodulin-like protein	0.801	-0.174	0.060	-0.561	0.120
		68195	C T	0.136	3.15E-07	0.029					0.821	-0.180	0.095	-0.530	0.093
NMDS of insect species (P/A)	Potra001060g09097	68122	C T	0.136	5.24E-08	0.007	Potrs000733g01118	Potri.003G050500	AT1G79800	nodulin-like protein	-0.696	0.195	-0.518	-0.162	0.425
		68195	C T	0.136	7.85E-08	0.007					-0.718	0.194	-0.497	-0.168	0.410

P/A = presence/absence
NMDS = nonmetric multidimensional scaling

Figure 2. Nonmetric multidimensional scaling (NMDS) ordination for both the (A) abundance and (B) incidence of common insect species surveyed on aspen in the WisAsp common garden (2014-5). Each point represents an insect community on an aspen genet. Each vector shows the insect species-level changes in the community (i.e., along a given vector, the particular insect species is either more [A] abundant or more [B] present within the insect community). (C) Summary of significant single nucleotide polymorphism (SNP) associations (false discovery rate [FDR] q-value < 0.10) for multivariate insect phenotypes for the WisAsp Populus tremuloides genetic mapping population (N = 328 genets). Genome-wide association models were analyzed in PLINK without a kinship matrix. Bolded alleles are the minor allele.