



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/245637/>

Version: Accepted Version

Article:

Pavinato, V.A.C., Penton-Michel, E.H., Shahid, S. et al. (2026) Adaptation to plant defence in an agricultural insect pest: Integrating genome scans and gene expression in the soybean aphid reveals multi-genic pathways. *Molecular Ecology*, 35 (17). e70536. ISSN: 0962-1083

<https://doi.org/10.1111/mec.70536>

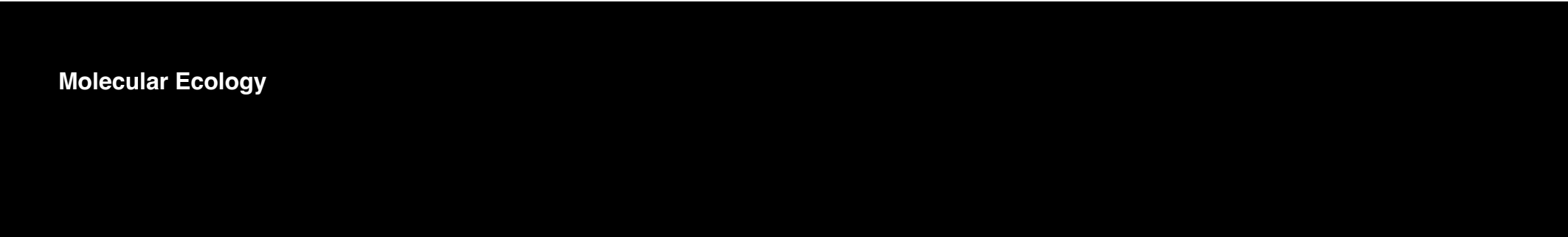
© 2026 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in *Molecular Ecology* is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Original Article

Adaptation to Plant Defense in an Agricultural Insect Pest: Integrating Genome Scans and Gene Expression in the Soybean Aphid Reveals Multi-genic Pathways

Submission ID dcf00671-c001-4209-9263-a2ee8cc467a7

Manuscript ID 6633190

Submission Version Revision 3

PDF Generation 23 Aug 2026 19:51:33 EST by Atypon ReX

Authors

Dr. Vitor A. C. Pavinato

 [ORCID](https://orcid.org/0000-0003-2483-1207)
<https://orcid.org/0000-0003-2483-1207>

Affiliations

- Dept. of Entomology, College of Food, Agricultural and Environmental Sciences, The Ohio State University, Wooster, OH, USA
- Department of Biology, Temple University, Philadelphia, PA, USA

Mrs. Erin H. Penton-Michel

Affiliations

- Dept. of Entomology, College of Food, Agricultural and Environmental Sciences, The Ohio State University, Wooster, OH, USA

Dr. Saima Shahid

 [ORCID](https://orcid.org/0000-0001-9385-0925)
<https://orcid.org/0000-0001-9385-0925>

Affiliations

- Dept. of Molecular Genetics, College of Arts & Science, The Ohio State University, Columbus, OH USA
- Donald Danforth Plant Science Center, St. Louis, MO, USA
- Plants, Photosynthesis and Soil, School of Biosciences, University of Sheffield, UK

Dr. Jacob A. Wenger

 [ORCID](https://orcid.org/0000-0001-5653-6996)
<https://orcid.org/0000-0001-5653-6996>

Affiliations

- Dept. of Entomology, College of Food, Agricultural and Environmental Sciences, The Ohio State University, Wooster, OH, USA
- Dept. of Plant Science, California State University, Fresno, CA, USA

Dr. R. Keith Slotkin

 [ORCID](https://orcid.org/0000-0001-9582-3533)
<https://orcid.org/0000-0001-9582-3533>

Affiliations

- Dept. of Molecular Genetics, College of Arts & Science, The Ohio State University, Columbus, OH USA
- Donald Danforth Plant Science Center, St. Louis, MO, USA
- Division of Biological Sciences, University of Missouri-Columbia, MO, USA
- The Center for Applied Plant Sciences, The Ohio State University, Columbus, OH USA

Dr. Andy Michel
Corresponding Author
Submitting Author

 [ORCID](https://orcid.org/0000-0002-2953-9187)
<https://orcid.org/0000-0002-2953-9187>

Affiliations

- Dept. of Entomology, College of Food, Agricultural and Environmental Sciences, The Ohio State University, Wooster, OH, USA
- The Center for Applied Plant Sciences, The Ohio State University, Columbus, OH USA

**Adaptation to Plant Defense in an Agricultural Insect Pest:
Integrating Genome Scans and Gene Expression in the Soybean Aphid Reveals Multi-genic
Pathways**

AUTHORS:

Vitor A. C. Pavinato^{1,6}, Erin H. Penton-Michel¹, Saima Shahid^{2,3,7}, Jacob A. Wenger^{1,8}, R. Keith Slotkin^{2,3,4}, Andy Michel^{1,5*}

¹Dept. of Entomology, College of Food, Agricultural and Environmental Sciences, The Ohio State University, Wooster, OH, USA

²Dept. of Molecular Genetics, College of Arts & Science, The Ohio State University, Columbus, OH USA

³Donald Danforth Plant Science Center, St. Louis, MO, USA & Division of Biological Sciences, University of Missouri-Columbia, MO, USA

⁴Division of Biological Sciences, University of Missouri-Columbia, MO, USA

⁵The Center for Applied Plant Sciences, The Ohio State University, Columbus, OH USA

⁶Current Address: Department of Biology, Temple University, Philadelphia, PA, USA

⁷Current Address: Plants, Photosynthesis and Soil, School of Biosciences, University of Sheffield, UK

⁸Current Address: Dept. of Plant Science, California State University, Fresno, CA, USA

*corresponding author: michel.70@osu.edu

Abstract

In agroecosystems, intense selection pressures cause species to adapt and spread, often leading to the evolution and persistence of pests. Understanding how pests rapidly adapt can help develop sustainable strategies for their management and improve agroecosystem health. Pest adaptation involves stable variations in DNA sequence, as well as dynamic shifts in gene expression, often mediated by non-coding regulatory elements. We examined adaptation to plant defenses in the soybean aphid, *Aphis glycines*, in which virulent aphids have overcome plant defenses and avirulent aphids have not. Previous data with laboratory colonies suggested that virulent aphids have higher overall gene expression, including transposable elements, some of which influence gene regulation. However, we lack information on how genetic variation in natural populations impacts adaptation and potentially gene regulation. We integrated population genome scans of field-collected, soybean aphid populations with gene expression profiles of virulent and avirulent laboratory colonies to uncover connections between genetic differentiation and gene regulation for virulence. Genome scan methods found 2,144 single nucleotide polymorphisms (SNPs) with significant genetic differentiation (i.e. outliers) in field-collected populations. These SNPs were near 1,004 genes, representing 5.16% of the effective number of genes. Based on previous RNA-Seq data with laboratory colonies, we found 3,160 genes and 147 long non-coding RNAs (lncRNAs) with differential expression among virulent and avirulent biotypes. By integrating both data sets, we identified 16 genes and 5 long non-coding RNAs with differential expression and that were associated with an outlier SNP (within 10kbp). We validated SNPs with additional field collected aphids and found an aphid clone with stronger virulence than our laboratory virulent colony, surviving on 2 different aphid-resistant soybean varieties. This new virulent clone had fixed allele differences at 9 SNPs compared to our avirulent and other virulent colony. Field collected soybean aphids matching the phenotype of this new virulent clone had significant genetic differentiation with 3 outlier SNPs near genes related to zinc transport and lachesin compared to field collected avirulent aphids. Our entire data reinforced the importance of a potential multi-genetic response to overcome plant defense and generates new insights into complex genetic and regulatory mechanisms involved in insect-plant interactions.

Keywords— insect biotype, virulence, rapid adaptation, genome scans, SNPs, soybean aphid

Introduction

Agroecosystems impose intense selective pressures on biodiversity. These pressures present significant barriers to population stability and persistence for some species, ultimately leading to low species and genetic diversity (Altieri, 1999; Emmerson et al., 2016; Morales et al., 2013; Ovčarenko et al., 2014). As some species perish, others flourish in these constantly changing landscapes. Rapid adaptation in agroecosystems facilitates the evolution of insect pests which negatively impact crop production. To counter these negative impacts, management practices decrease pest populations but also affect non-pests and promote further pest adaptation (*e.g.* resistance or virulence evolution). Identifying the factors that facilitate pest adaptation and persistence is critically important to improve the sustainability of agroecosystems and agricultural crop production.

Insect crop pests cause significant losses in terms of direct damage (*e.g.* yield loss) and increase input costs used for control. In conventional agroecosystems, common insect control methods include synthetic chemicals and plant resistance (both transgenic and natural, *e.g.* host-plant resistance). These methods typically have initial success with declining efficacy over time as selection pressure leads to the evolution of resistance (Hawkins et al., 2019; Smith, 2021; Tabashnik & Carrière, 2017). No matter what the control technology, insect adaptation to control strategies is a constant occurrence and challenge in agroecosystems.

In insect crop pests, virulence refers to the ability of an insect to overcome a host plant's natural resistance mechanisms, allowing them to successfully survive, reproduce and feed on the resistant plant variety. This is conceptually parallel to insecticide resistance, which refers to an insect's reduced susceptibility to a chemical control agent. Both insect resistance and virulence can be based on simple genetic polymorphisms in one or a few genes. These polymorphisms may alter the protein sequences of particular toxin receptors (Tabashnik, 2015; Bass and Nauen, 2023), or modulate insect-plant interactions (*i.e.* effectors, Hogenhout & Bos, 2011; Yates & Michel, 2018). Given the intense selection pressure of insecticides and resistant crops over a short timescale, polymorphisms and their respective signature of selective sweeps should be detectable in insect populations using genome scans (Coates et al., 2020; Péliissié et al., 2022; Taylor et al., 2021). Any polymorphisms that significantly differ in frequency among adapted and non-adapted populations (*e.g.* outliers) might be either responsible for, or linked with, the resistant or virulent phenotype. When combined with genome annotations and whole genomic sequences, outliers can be further filtered to identify those within genes or linked to candidate resistance genes such as toxin receptors or effectors (Coates et al., 2020).

Alternatively, insect resistance or virulence could result from changes in gene expression and regulation. For example, resistant insects can overexpress certain detoxification genes enabling faster insecticide degradation (Wang et al., 2022) or lack expression of certain effectors that plants recognize to mount an immune response (Du et al., 2020; Zhao et al., 2015). Changes in gene expression are not always dependent on DNA polymorphisms and are sometimes mediated by other mechanisms including structural variation caused by transposable elements (TEs) and epigenetic regulation (Li et al., 2019; Brevik et al., 2021). Interspersed throughout a genome, TEs can change gene expression by inserting themselves into a gene or by influencing the expression of nearby genes (among other mechanisms) (Hirsch & Springer, 2017; Petersen et al., 2019; Slotkin & Martienssen, 2007). In insects, TE expression can be controlled by differential DNA methylation (Lewis et al., 2020), or through repressive marks on histones that package DNA (Glastad et al., 2019; Ninova et al., 2020; Ramat & Simonelig, 2021). Such repressive histone marks (*e.g.* H3K9 methylation) lead to transcriptional silencing of TEs as well as nearby

genes (Glastad et al., 2019). Long non-coding RNAs (lncRNAs) are a type of regulatory non-coding RNA (ncRNA) that may aid adaptation, including TE regulation (Soares de Melo & Wallau 2022; Li et al., 2019). For instance, differential expression of a lncRNA leads to greater survival of the green peach aphid, *Myzus persicae* when feeding on different host plants (Chen et al., 2020), and a lncRNA in *Aphis citricidus* is predicted to target a TE protein and regulate wing formation (Shang et al., 2021). Additional research is needed on how these mechanisms influence insect adaptation, especially in crop pests responding to selection pressures within agroecosystems.

Aphids are emerging as an important model to understand rapid adaptation (Mathers et al., 2017; Wooley et al., 2020; Srinivasan & Brisson, 2012). Many aphids have resistance to several insecticides (Singh et al., 2021), as well as can overcome host plant resistance in the form of virulent biotypes (*i.e.* virulence evolution) (Smith, 2021). Several aphid species have whole genome sequences available which facilitate integrated investigations of genetic and molecular mechanisms for regulation and adaptation (Mathers et al., 2019; Preez et al., 2020; Srinivasan & Brisson, 2012; Sattar et al., 2012). In this research, we investigated adaptation to host plant resistance in the soybean aphid, *Aphis glycines*. Native to east Asia, this aphid invaded North America ca. 2000 and has spread to become one of the most important pests of soybean. Soybean aphids are extremely host-specific and alternate between overwintering hosts (mostly *Rhamnus* sp.) and soybean during the summer and autumn. Asexual reproduction occurs on soybean and, prior to soybean senescence and harvest, sexual forms are produced, migrating to *Rhamnus* where mating and egg laying occurs. The recent invasion and reduced genetic diversity (Michel et al. 2009), the restricted distribution of *Rhamnus* (Heimpel et al. 2010), and the interplay between dispersal and reproduction modes (Orantes et al. 2012) provide a complex population genetic structure and demographic history.

Yet soybean aphid populations have overcome host plant resistance in soybean varieties that carry gene(s) for resistance to *Aphis glycines* (*i.e.* *Rag* genes). The exact mechanism for *Rag* based resistance is not known (Natukunda & MacIntosh 2020), but likely involves secondary metabolites (antibiosis, Meng et al. 2011; Joshi et al. 2021). Host plant resistant soybean can contain different numbers of *Rag* genes (Natukunda & MacIntosh 2020), but most research has focused on *Rag1*, *Rag2* and their combination (*Rag1/Rag2*). Like other host-plant resistance systems, insect populations that overcome resistance are usually termed virulent biotypes (as opposed to avirulent biotypes which do not survive on host-plant resistance). Soybean aphid biotypes are defined by their differential survival and reproduction on *Rag* soybean compared to non-*Rag* soybean or to established laboratory biotype colonies. Currently there are 4 known soybean aphid biotypes: biotype 1 which is avirulent (cannot survive) on any *Rag* soybean; and 3 virulent biotypes: biotype 2 (surviving on *Rag1*); biotype 3 (surviving on *Rag2*), and biotype 4 (surviving on *Rag1/Rag2*) (note: Hesler and Beckendorf 2020 lists an additional biotype, provisionally named Accrue, which needs additional characterization). Biotype 4 constitutively overexpresses more genes than biotype 1 aphids, including TEs, but also has decreased expression of putative effectors (Yates-Stewart et al., 2020). Furthermore, single nucleotide polymorphisms (SNPs) with differential frequencies and selection histories occur among different types of virulent soybean aphid biotypes (Coates et al., 2020). However, most studies included only laboratory colonies of soybean aphids, which, because of asexual reproduction, have been in isolation for several years. Since natural populations of the soybean aphid have an obligatory sexual reproduction phase, recombination has likely altered genotypes present in the field compared to laboratory colonies and changed linkage between genetic variation and the virulence phenotype (however, at the time of collection in our study, a minimum of 13 sexual

generations occurred, suggesting that genomic linkage might still be widespread). In fact, additional research has identified a variety of virulence responses from field-collected soybean aphids (Lagos-Kutz et al., 2020). Analyzing the soybean aphid genome in natural populations would help connect virulence patterns to laboratory colonies and deepen our understanding on how reproductive modes, migration, and selection influence genetic variation for virulence.

We examined the connection between genetic variation and virulence by scanning the soybean aphid genome for signatures of selection from field collections in the North Central region of the U.S. Using a pool-sequencing approach, we identified SNPs with significantly different frequencies (*i.e.* outliers) among virulent biotype 4 and avirulent biotype 1 soybean aphids. We then integrated the SNP data with a previous transcriptomic study from biotype 4 and biotype 1 laboratory colonies (Yates-Stewart et al., 2020) to identify outlier SNPs related to candidate genes and regulatory elements (lncRNAs) for virulence. These SNPs were used to genotype field-collected soybean aphids that were screened for virulence to *Rag1/Rag2* and a soybean variety with three *Rag* genes: *Rag1/Rag2/Rag3*. Combining all data, we uncovered a field-collected clone with virulence to both *Rag1/Rag2* as well as *Rag1/Rag2/Rag3*. This virulent field colony, provisionally named 3R virulent, showed significant, and sometimes fixed, genotypic differences from established laboratory colonies of both biotype 1 and biotype 4. Furthermore, field screening of additional soybean aphid populations supported outlier status of 3 SNPs with individuals matching the 3R virulent phenotype. Our data set provides further evidence of a complicated basis for virulence adaptation.

Material and Methods

Information on our field collections (geographical coordinates, biotype designation, sample sizes for sequencing, the associated Sequence Read Archive (SRA) number, etc.) can be found in Supplementary Table S1. The SRA archive for the whole-genome sequencing data analyzed in this work is also part of the first release of the soybean aphid biotype 4 genome (accession number PRJNA312843, Wenger et al., 2020).

Biotype collection and identification for sequencing

Soybean aphids were collected at six field sites during the late summer of 2013 (likely during or after large-scale dispersal). Previous data has indicated the timing of our collections to be during a period of large-scale dispersal and panmixia (Orantes et al. 2013) for the soybean aphid, which may lead to lower baseline level of F_{ST} (from a neutral evolution perspective). This period of panmixia occurs prior to a more restricted migration to their primary host for sexual reproduction and overwintering—any genetic variation spread through summer dispersal would likely be transferred via sexual reproduction (a hypothesis which remains untested). Our sites were distributed across an East-West transect of the species' North American range (Figure 1A). In each location, a soybean aphid-infested leaf was removed from a soybean plant and placed in an individual plastic bag. Only one leaf per soybean plant was removed and we collected a total of 50 infested leaves per location from a: 1) a soybean aphid susceptible variety (SD-01-76R with no known *Rag* genes), 2) and a soybean variety containing both *Rag1* and *Rag2* genes for soybean aphid resistance (from Wiarda et al., 2012, which used LD05-16521 for *Rag1* and LD08-089051a for *Rag2*). Bags were shipped overnight to Michel's Laboratory. Upon receipt, a single aphid per sampled leaf was placed on a susceptible leaf in a parafilm-sealed Petri dish to establish a clonal lineage (*i.e.* an aphid population originating from 1 asexually reproducing female, Figure 1B). After establishment of clonal lineages (~ 50 total aphids in a Petri dish), we

transferred an age-synchronized adult to a whole plant (~ 14 days after planting) of each soybean cultivar (described above) for phenotyping. We defined virulence as any aphid doubling its population on a resistant plant (*Rag1/Rag2*) within seven days (phenotyping data available upon request). For this initial screening, we only included aphids for which we had high confidence of soybean aphids being either biotype 1 (*i.e.*, cannot survive on any *Rag*) or biotype 4 (*i.e.*, can survive and reproduce on *Rag1/Rag2*). After phenotyping, all aphids from the same clonal lineage were stored in a 1.5 mL micro-centrifuge tube at -20°C until DNA extraction.

DNA extraction and pool-sequencing

We pooled ~ 20 aphids from a single clonal lineage prior to DNA extraction. DNA was then extracted using the E.Z.N.A DNA isolation kit (Omega Bio-Tek, Norcross GA) following the manufacturer's directions, and quantified fluorometrically (Qubit – Life Technologies, Carlsbad CA). After standardizing the DNA concentration of each individual clonal lineage, five clonal lineages of the same biotype/locality/sample were combined to produce 21 DNA pools (35 to 550ng of total DNA). Illumina sequencing libraries were constructed from the pooled DNA using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England BioSciences, Ipswich, MA). Libraries of each sample pool were sequenced four times (with libraries of pools MN-Av.1, NW-V.1, and WI-Av.2 sequenced 5 times, see Supplementary Table S1) on an Illumina MiSeq platform in a 300bp paired-end configuration at the CFAES Research and Analytical Support Core (RASC) at The Ohio State University. The sequencing replication of each library provided technical replications, which were assessed to ensure quality before further analyses. We sequenced 21 pooled samples: eight from biotype 1 pools and thirteen from biotype 4 pools (some collection sites had more than one pool per biotype due to variability in colony establishment and ambiguity for biotype designation, see Figure 1A).

Sequencing, quality control, and read mapping

With a few modifications, we used the pipeline developed by Kapun et al., (2020) for *Drosophila melanogaster* whole-genome pool-sequencing data (Pool-seq) (<https://dest.bio>). We used the updated reference genome of the virulent soybean aphid biotype 4 version 2.1 (Mathers, 2020) (accessed on March 10th, 2021; the initial version of the biotype 4 genome included genomic data from this study, Wenger et al. 2020). The specific, modified pipeline for our sequencing analysis is available at <https://github.com/vitorpavinato/AglyPoolseq>. The pipeline used 1) cutadapt v.2.3 (Martin, 2011) to remove sequencing adapters and to keep reads with a minimum length of 75bp; 2) FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) to check read quality before and after adapter trimming; and 3) bbmerge (Bushnell et al., 2017) to merge overlapping reads (after trimming). The pipeline maps the trimmed reads against a hologenome containing the biotype 4 *A. glycines* genome (including the mtDNA), and genomes of associated bacterial symbionts (*Buchnera aphidicola*, *Arsenophonus* spp., *Regiella insecticola*, and *Wolbachia* spp.). We included these additional genomes to remove potential contamination that might be present in the raw sequencing reads. Read mapping was performed with bwa mem v.0.7.15 (Li, 2013) with default parameters. Read mapping only retained reads that uniquely mapped (PHRED score >20) to the biotype 4 genome v.2.1 for further analysis. Duplicated reads from PCR were removed with Picard MarkDuplicates (v.1.109; <http://broadinstitute.github.io/picard>). With a list of indels produced with GATK v3.4-46 (McKenna et al., 2010), RealignerTargetCreator was used by GATK v3.4-46 IndelRealigner to improve alignment around indels.

236 SNP calling and allele frequencies

237 We used the heuristic approach from PoolSNP (Kapun et al., 2020) to call SNPs both within a
 238 sample's technical replicates (see Supplementary Material, Methods 1) and globally (*i.e.*,
 239 combining all technical replicates for each pooled sample). The aggregation of technical
 240 replicates included all sequencing runs from each pooled sample as they did not show significant
 241 differences (see Supplementary Material, Methods 1, Supplementary Figures S1, S2, and S3,
 242 Supplementary Table S1). We first obtained the read counts for every nucleotide position of the
 243 genome for each sampled pool alignment file with *samtools mpileup*. On the raw read counts for
 244 each sample (MPILEUPs), we applied quality control parameters to: 1) remove reads with a
 245 mapping quality score (PHRED) below 25 ($\text{PHRED} < 25$); 2) mask positions that did not meet
 246 the required read depth (minimum = 4 and maximum = 99% of the total read depth distribution);
 247 3) mask SNPs located within transposable elements (TE) and repetitive regions; and, 4) mask
 248 SNPs located 5bp upstream and downstream of an indel (see Supplementary Material, Methods
 249 1). Each sample's masked mpileup files were used in PoolSNP. PoolSNP used a heuristic
 250 approach to identify SNPs using the following criteria: minor allele count (MAC) must be >5
 251 reads and minor allele frequency (MAF) must be >5% across all pooled samples (Rodrigues et
 252 al., 2021). These values were determined based on evaluations of different MAC and MAF
 253 parameters (see, Supplementary Material, Methods 1). SNPs were annotated with SnpEff version
 254 4.3 (<https://pcingola.github.io/SnpEff/>) using the GFF file from Mathers (2020) accessed on
 255 March 2021 to identify SNPs causing synonymous and non-synonymous amino acid changes.

256 We calculated the reference allele frequencies (based on the biotype 4 reference genome) with
 257 the R package Poolfstat (Hivert et al., 2018). To overcome any read count bias (Smadja et al.,
 258 2012), we implemented a maximum-likelihood (ML) approach (Leblois et al., 2017; Nouhaud et
 259 al., 2018; Smadja et al., 2012). Since we only used bi-allelic SNPs, the frequency of the
 260 alternative allele (*i.e.* biotype 1) is 1 minus the frequency of the reference allele.

261 Genetic diversity and population structure

262 We calculated the total heterozygosity H_T , which is defined as

$$263 \quad H_T = 2\bar{p}_i(1 - \bar{p}_i)$$

264 where $\bar{p}_i$ is the average reference allele frequency of SNP i across all pools. The genome-wide
 265 H_T was calculated by averaging H_T across all SNPs. Within-pool estimates of each SNP
 266 expected heterozygosity (H_E) were obtained with the within-pool allele frequencies and using the
 267 following estimate:

$$268 \quad H_E = \frac{N}{N-1} 2\bar{p}_i(1 - \bar{p}_i)$$

269 where N represents the number of haploid genomes in the pool (that is 2x the number of
 270 individuals in the pool) and $\bar{p}_i$ is the maximum-likelihood allele frequency of allele i within each
 271 pool. We also calculated the synonymous (S) and non-synonymous (N) SNP H_E and their
 272 proportions (p_S ; p_N). Within-pools H_E , synonymous and non-synonymous H_E , and p_S and p_N
 273 estimates were obtained across scaffolds and, for each population pool, averaged across
 274 scaffolds. Comparisons among pools, phenotypes, localities, and geographical regions (*e.g.* East:
 275 Ohio and Pennsylvania; West: Wisconsin, Minnesota and North Dakota) were obtained with
 276 ANOVA performed in R 4.4.3 (Team R 2025).

We estimated the global F_{ST} among pools with the R package Poolfstat (Hivert et al., 2018). To produce permutations and calculate the 95% confidence interval, we obtained a block-jackknife distribution of global F_{ST} 's by generating genomic windows with at least ten SNPs. Pairwise F_{ST} 's among the 21 sampled pools were obtained similarly using the R package Poolfstat.

Genome scan for outliers associated with virulence

Given the difference in phenotype among virulent and avirulent aphids, we expected to identify SNPs co-segregating with candidate genes associated with *Rag1/Rag2* virulence. (Full methods are included in Supplementary Material, Methods, 2). We used a comprehensive and conservative approach to determine outliers from soybean aphid populations due to its complex population genetic and demographic history. Since we collected our samples in 2013, only ~13 sexual generations had occurred assuming its initial invasion occurred in 2000. In addition, evidence exists for both a founder effect (Michel et al. 2009) as well as seasonal bottlenecks interrupted by asexual reproduction (i.e. clonal amplification) and widespread dispersal (Orantes et al. 2012). These factors may retain larger sections of linkage disequilibrium and artificially increase the number of false positives, which is already a known occurrence in genome scan studies. To limit false positives, our overall conservative approach implemented 5 genome scan tests and integrated data with gene expression from laboratory colonies to identify areas of congruence between natural and laboratory populations and increase the confidence of SNP outliers and genes related to virulence.

We first scanned the soybean aphid genome for signals of allele frequency differences among pools independently of the pooled sample phenotype. We applied the classical Lewontin and Krakauer single-locus F_{ST} (S- F_{ST}) scan (Lewontin & Krakauer, 1973) specific for pool-seq data (Hivert et al., 2018). This F_{ST} scan accounts for the sampling variance involved in pooling samples for sequencing and combines the signal of multiple SNPs in a sliding-window approach. Poolfstat (Hivert et al., 2018) then calculates a multi-locus F_{ST} (M- F_{ST}) from non-fixed sliding-windows. With Poolfstat, we restricted the estimation of S- F_{ST} for SNPs with no missing data (Poolfstat calculated F_{ST} for 19,842 SNPs). For the M- F_{ST} , we identified outliers if their M- F_{ST} values were higher than 0.25, which is 32% higher than the 99th% percentile of the empirical M- F_{ST} distribution (99% M- F_{ST} quantile = 0.19). We transformed the genome coordinates of each significant M- F_{ST} back to the individual SNPs coordinates to compare among methods.

To strengthen and expand outlier detection, we employed four additional methods developed with different assumptions: a) two methods that identified outlier SNPs correlated with a covariable (i.e. using the assigned phenotype of each pool – biotype, as a covariable), and b) two methods that searched for SNPs with significant genetic differentiation among populations. Genome scan methods that associated SNPs with a covariable were: 1) BAYPASS auxiliary covariate model, which computes a Bayes factor for each SNP from the posterior distributions (BAY_{aux} BFmc), and 2), BAYPASS XtX statistics, calculated by considering the covariate model (BAY_{aux} XtX) (Gautier, 2015). Genome scans that looked for differentiation among populations (that were covariate-free) were: 3) SelEstim (SEL) (Vitalis et al., 2014), and 4) BAYPASS core model based on the XtX statistic (BAY_{core} XtX) (Coop et al., 2010; Günther & Coop, 2013). All three BAYPASS methods share a similar approach to control for shared ancestry (i.e. population structure) and reduce false positives. All four methods increase the outlier detection power by accounting for biases imposed from demography. For BAYPASS, outliers were defined as those that had XtX or BFmc statistics higher than the 99% quantile of distribution obtained with simulated data. For SelEstim, we selected candidate SNPs with values of the Kullback–Leibler

Divergence (KLD) statistics above the 95% quantile of the distribution of estimated values in each of the three independent runs. We then cross-referenced outlier SNPs from all 5 methods, grouping them by the covariate or covariate-free model. A candidate SNP for virulence was defined as an outlier if it was included in at least 3 tests. We do acknowledge that 2 tests (BAY_{core} XtX and BAY_{aux} XtX) may not be truly independent. BAY_{core} XtX identifies loci with high levels of differentiation regardless of phenotype, while BAY_{aux} XtX recalculates the same statistic after conditioning on the phenotype (in our case, virulence). However, treating these as two different tests provides a more stringent definition of outlier status.

Transcriptomics and differential gene expression

We re-analyzed previous whole transcriptome data from biotype 4 and biotype 1 soybean aphids (from laboratory colonies) reared on both susceptible (IA3027) and *Rag1/Rag2* (IA3027RA12) soybean (NCBI BioProject accession number: PRJNA598670; Yates-Stewart et al., 2020). Data was generated from soybean aphid nymphs (less than 24 hours old) feeding on a soybean trifoliolate for 12 hours. Three replicates of each biotype-soybean plant combination (with individual replicates containing a pool of 24-60 nymphs) were collected, followed by RNA sequencing (see Yates-Stewart et al. 2020). We aligned transcriptome data and quantified gene expression using the biotype 4 genome (version 2.1 of the assembly and gene models accessed on March 2021, Mathers et al. (2020)), following the same methodology as previously described (Yates-Stewart et al., 2020). We identified differentially expressed genes (DEG) with the R package DEseq2 (Anders et al., 2015). For comparisons between biotypes on the same host plant, we determined significant genes using adjusted p-value (P_{adj}) cutoff of 0.01. Next, we distinguished two types of gene expression: 1) genes with constitutive expression, and 2) genes with biotype-host-specific expression. Genes with constitutive expression showed significant up or down regulation in biotype 4 regardless of the host plant they were exposed to (*i.e.* *Rag1/Rag2* or susceptible, $P_{adj} > 0.05$ and the absolute $\log_2$ fold change $< |1|$). In contrast, genes with biotype-host-specific expression showed significant regulation depending on the interaction between the biotype and the type of plant. Both patterns were included since biotype 4 showed an overall increase and constitutive increased gene expression compared to biotype 1 (Yates-Stewart et al. 2020).

We performed independent Gene Ontology (GO) enrichment analysis with topGO (Alexa and Rahnenfuhrer, 2023) for significant differentially and constitutively expressed genes (up- or down-regulated) in biotype 4 compared to biotype 1 aphids. For this analysis, gene ontology annotations for the biotype 4 genome v.2.1 were accessed from AphidBase on July 12th, 2021. For each analysis, we performed Fisher's exact test with a false discovery rate correction of 1%.

For each outlier SNP, we searched if it was within 10Kbp of an annotated gene (5Kbp up-and downstream) by mapping back to the soybean aphid genome using SNP coordinates with *bedtools window* (Quinlan & Hall, 2010) on the biotype 4 genome v.2.1 gene model. From this list, we focused on genes with significant ($P_{adj} < 0.01$) and constitutive differential expression ($\log_2$ fold change $> |2|$) that were also found near outlier SNPs (within 10Kbp) identified in at least three genome scan tests.

Long non-coding RNAs annotation and expression

We downloaded the coding sequences of all predicted soybean aphid lnc-RNAs from InsectBase 2.0 (accessed November 2022, Mei et al., 2021) and aligned these sequences to the biotype 4 genome with BLAT v.35 (Kent, 2002). With a custom script (available upon request), we

retained the best alignment hits and converted the alignment into genome coordinates. Using UCSC-biotools, we converted the BED file into a GFF file to use with featureCounts. Combining the soybean aphid biotype 4 virulent genome version 2.1, the genomics coordinates of the lncRNAs, and the same tools and parameters as in Yates-Stewart et al., (2020), we identified differentially expressed lncRNAs with the DESeq2 R package (Anders et al., 2015) and a false discovery rate (FDR) corrected P -value of 1% ($P_{adj} < 0.01$). We specifically looked for members of the lncRNA family *Ya*, previously found to be involved in host feeding in *M. persicae* (Chen et al., 2020). We aligned the *Ya* coding sequences from the soybean aphid biotype 4 genome v1 (accessed December 2022) to the soybean aphid genome v2.1 with BLAT and retained only the most related sequences (custom script available upon request).

SNP validation in laboratory colonies and field samples

After identifying SNPs in candidate genes, we genotyped additional samples to validate their influence on virulence. During the summer of 2022, field collected aphids from Wooster demonstrated survival on a soybean line with 3 *Rag* genes (*Rag1/Rag2/Rag3*, i.e. LD14-8003) and are currently maintained as a colony in the Michel laboratory (hereafter referred to the 3R virulent colony, corresponding to the 3 *Rag* genes (*Rag1/Rag2/Rag3*). No-choice tests were performed in 2024 to compare the fecundity of females of this colony to those of the biotype 4 colony. Five age-synchronized adult aphids were transferred to 14-day old whole plants grown from single seeds in 4-inch pots filled with Promix potting soil using soybean cultivars LD14-8007 (no *Rag* genes), LD14-8001 (contains *Rag1/Rag2*) and LD14-8003 (contains *Rag1/Rag2/Rag3*, all provided by B. Diers, University of Illinois). Each plant was contained within a 7.5 cm by 30 cm plastic cylindrical cage with side windows sealed with silk fabric mesh (0.1 mm aperture). Initial plant growth and subsequent no choice tests were performed in growth chambers at 24°C and 50% humidity, with 300 $\mu\text{mol m}^{-2}\text{s}^{-1}$ light for 16 h daily. Aphids were counted at 14 days post infestation (DPI). Each soybean cultivar vs biotype was replicated 4 times for a total of 16 plant-aphid treatment combinations. Data were tested for normality and transformed using a Tukey transformation, followed by an analysis of variance (ANOVA) performed in the R 4.4.3 package (Team R 2025).

Field collected soybean aphids came from three locations during the summer of 2024: Columbus-Ohio (approx. 40.014, -83.039), Wooster-Ohio (approx. 40.763, -81.908), and Brookings-South Dakota (approx. 44.352, -96.807) (Figure 1A). A total of 56 field collected individuals were allowed to reproduce on susceptible leaves in Petri dishes and then assessed for virulence using the different resistant soybean varieties as previously described (see above, full phenotyping comparisons in Prischmann et al. accepted). Based on phenotypic data we chose a similar number of aphids from each biotype for genotyping, totaling 15 soybean aphids (1 from Columbus, 11 from Wooster, and 3 from Brookings). DNA was extracted from single individuals from these field collected samples as well as from laboratory-maintained soybean aphid colonies: 2 individuals from each of biotype 1, biotype 4, and the 3R virulent colony using the DNeasy Blood and Tissue kit (Qiagen, Germany) following the manufacturer's directions.

Primers were designed within 15 regions encompassing the positions of SNPs near candidate genes and lncRNAs (Table 1, Supplementary Table S2) using Primer Blast (NCBI) (Primers for genes on scaffold 255 and 1 SNP on scaffold 615 could not be designed or did not provide suitable sequencing results). Subsequent PCR amplifications were performed in 25 μL reactions using 12.5 μL of 2X Go Taq G2 Green Master Mix (Promega, Wisconsin), 0.4 μM of each primer, and 5 ng of gDNA. The thermocycling profile consisted of one cycle of 2 min at 94°C, 35 cycles of 30 sec at 94°C, 30 sec at 55-59°C and 1 min at 72°C, with a final extension of 5 min

at 72°C (see Supplementary Table S2). Following visualization on 1.0% (w/v) agarose gels all PCR products were Sanger sequenced (Functional Biosciences, Wisconsin). Sequences were edited in Geneious (<http://www.geneious.com>) and genotyped for the predicted SNP allele. Allele frequencies and Fisher's exact probability test (genic differentiation) for all populations were estimated in GenePop (Raymond and Rousset, 1995; Rousset, 2008); locus-by-locus F_{st} was estimated in Arlequin (Excoffier and Lischer, 2010). For each validated SNP, we tested for an association between individual reproductive success and genotype using an additive linear regression model. Reproductive success was measured as the population size on *Rag1/Rag2/Rag3* plants. Genotypes were coded additively as 0, 1, or 2, using corresponding alleles present in the biotype 4 genome: homozygous reference, heterozygous, and homozygous alternate genotypes, respectively. Each SNP was analyzed independently. Statistical significance of regression slopes was assessed using permutation tests (10,000 permutations) to account for small sample size and potential deviations from normality. All analyses were performed in R.

Results

Sequencing and SNP identification

The combined paired-end sequencing runs generated 237 million reads. On average, 1.3 million high-quality reads with a minimum length of 75bp and a maximum length of 300bp were uniquely mapped to the biotype 4 reference genome (Supplementary Material, Supplementary Table S3). No differences were found among the technical replicate runs based on PCA (Supplementary Material, Methods 1). After combining all reads from each pooled sample, we produced alignment files with 5.17M reads, resulting in an average sequencing coverage of 14x (ranging from 8x to 26x, Supplementary Material, Supplementary Table S4). Based on evaluations of different minor allele count (MAC) and minor allele frequency (MAF) parameters (Supplementary Material, Figure S4), we used $MAC > 5$ and $MAF > 0.05$ to retain 262,866 high-quality, bi-allelic SNPs from all sampled pools (Supplementary Material-Methods).

Population genetic diversity and structure

The average total expected heterozygosity among 262,866 SNPs was low: 0.342 (95% C.I. 0.336 - 0.349). The within-pool genetic diversity reflected a low average, ranging from 0.291 (WO-V.1) to 0.367 (NW-V.1). Genetic diversity was significantly different among pools ($P < 0.001$), biotypes ($P < 0.001$), and localities ($P < 0.001$), with the pools of biotype 4 aphids showing significantly higher genetic diversity ($0.336 \pm SE = 0.001$, $P = 0.001$) than biotype 1 aphids ($0.323 \pm SE = 0.002$, $P = 0.002$), and pools from Minnesota having the highest diversity ($0.346 \pm SE = 0.003$). There was no difference in the levels of genetic diversity between the western and eastern regions ($P = 0.558$) (Supplementary Material, Figure S5). The genetic diversity of synonymous and non-synonymous mutations differed among pools, biotypes, and localities but not between geographic regions (Supplementary Material, Figures S6A and B). The proportion of synonymous and non-synonymous mutations were significantly different among pools, biotypes, localities, and regions ($P < 0.01$) (Supplementary Material, Figures S6C and D).

The global F_{ST} was low but significant (0.004; 95% C.I. 0.002 - 0.006, $P < 0.01$). The pairwise F_{ST} among the 21 pooled samples ranged from -0.053 (NW-V.1 & WI-V.3) to 0.110 (WI-Av.1 & WO-V.1). To evaluate genetic differentiation between biotype 4 and biotype 1 we used a framework that accounted for demography in the form of isolation-by-distance (IBD) (Wang & Bradburd, 2014). F_{ST} between biotype 4 and biotype 1 pools (Supplementary Material, Figure

S7) was not significantly different when IBD was accounted for all populations ($P = 0.172$), nor the eastern populations (Ohio and Pennsylvania, $P = 0.327$). The F_{ST} among biotype 4 and biotype 1 was significantly different for populations in the western region (North Dakota, Minnesota, and Wisconsin populations, $P = 0.013$); the pairwise F_{ST} between North Dakota biotype 4 pool-1 (ND-V.1) and Wisconsin biotype 1 pool-1 (WI-Av.1) had the highest $F_{ST} = 0.077$ (Supplementary Material, Figure S7B). The results suggest that drift and IBD are not solely responsible for genetic differentiation, at least in the Western populations, and that the virulence phenotype may also be an important factor.

Genome scans for virulence

Using the classical Lewontin and Krakauer F_{ST} -based test (Lewontin & Krakauer, 1973) we identified 31 genomic regions from 16 scaffolds with significantly high multi-locus F_{ST} s (Supplementary Figure S8A). Within these regions, we then identified a total of 102 individual SNPs that had a single-locus F_{ST} higher than the within-scaffold quantile threshold of 95% (Supplementary Figure S8B). Many scaffolds (62.5%) contained several outlier SNPs. Several of these scaffolds had lower genetic diversity (as measured by H_E , Supplementary Figure S9) or lower SNP density around outlier SNPs, consistent with a recent selective sweep. We complemented the traditional F_{ST} outlier detection with genome scan methods developed with different assumptions, attempting to reduce the false discovery of candidate genes in potentially increased areas of linkage disequilibrium due to demography and limited opportunities for recombination. The number of identified outlier SNPs varied among methods (see Supplementary Material: Figures S10-S12), but together they identified 10,418 unique outlier SNPs (3.96% of all SNPs) that showed signatures of selection. From this list of outliers, 6,866 (2.61% of all SNPs) were significantly associated with biotypes (Figure 2A), and 8,071 (3.07%) were significantly differentiated among geography (Figure 2B). Outlier SNPs were in close proximity (within 10Kbp) to 4,238 genes (out of 19,452; 21.79% of effective genes). A total of 3,217 (16.54%) genes were significantly associated with differences between biotypes 1 and 4 (Figure 2C) and 3,021 genes (15.53%) were significantly differentiated among geographical locations (Figure 2D). Most outlier SNPs were unique to each method; there were no outlier SNPs common in all 5 methods, 37 SNPs common in 4 methods and 2,107 SNPs common in 3 methods. Genes near outlier SNPs (from at least three outlier tests) totaled 1,004 genes, corresponding to 5.16% of the effective number of genes. We focused on genes near outliers identified from at least 3 genome scan methods for further investigation.

Integration of genome scans with differential expression analysis

To understand how the genes identified with our genome scan may influence virulence adaptation, we compared our data to a previous differential gene expression study between biotype 1 and 4. We re-analyzed these data that were generated using nymphs fed susceptible and *Rag1/Rag2* soybean (3 replicates of each biotype-plant combination, see Yates-Stewart et al. 2020). Our reanalysis identified a total of 3,160 differentially expressed genes: 1,791 (56.68%) up- and 1,369 (43.32%) down-regulated among all comparisons between avirulent vs. virulent aphids. Several up-regulated genes were linked to major ontology classifications related to gene regulation including histone acetylation, inhibition of cell proliferation, protein ubiquitylation, RNA silencing, and small RNA regulation (Supplementary Material, Figure S13). There were 642 DEGs out of 4,238 (10.29%) genes near a SNP outlier. Since most differences in gene expression were specific to biotype and not dependent on plant type (Yates-Stewart et al., 2020), we focused on DEGs with constitutive expression. Based on our criteria of minimum $\pm 2 \log_2$

fold change and near an outlier SNP in at least three genome scans, we found 14 genes with significant difference in expression ($\pm 2 \log_2$ fold change and $P_{\text{adj}} < 0.01$) and near an outlier SNP (Table 1; Supplementary Table S5 list of associated SNPs, outlier test results; a list containing all constitutively expressed outlier genes regardless of the $\log_2\text{FC}$ threshold that were identified in at least 3 outlier tests can be found in Supplementary Table S6). Outlier SNPs on scaffold 133 were associated with Agly004871, a gene that had constitutive significant gene expression differences among biotype 1 and biotype 4 aphids. However, additional sequence characterization showed that these SNPs were near or within the 5'-untranslated region of the nearby gene Agly004838. This gene was significantly overexpressed in biotype 4 aphids on *Rag1/Rag2* resistant plants (5.97, $P_{\text{adj}} < 0.001$); biotype 4 aphids also had higher expression of this gene on susceptible plants (2.58) but was marginally significant ($P_{\text{adj}} < 0.053$, Table 1). Based on the sequence characterization and gene expression pattern, Agly004838 was included as an additional candidate gene. We also added Agly000983 as a candidate gene as it had significant gene expression differences (although with a $\log_2$ fold change < 2) and was partner gene for 2 differentially expressed lncRNAs (see below). Therefore, our 16 total candidate genes were distributed on 11 scaffolds, with scaffolds 114, 129, 237 having multiple genes. There is an overrepresentation of genes ($N=10$) located on the X chromosome, with three on chromosome 3, two on chromosome 1 and one on chromosome 2.

Our analysis also identified a total of 147 differentially expressed lnc-RNAs: 136 (4.75% of all lnc-RNAs) up- and 11 (0.38% of all lnc-RNAs) down-regulated among all comparisons between biotype 4 vs. biotype 1 aphids (Supplemental Material Supplementary Table S7); 5 lncRNAs were also near an outlier SNP and distributed on 3 different scaffolds and 3 different chromosomes (Table 1). None of the lncRNAs similar to the *Ya*-type were differentially expressed at a level $\pm 2 \log_2$ fold change although *Ya4* was located near an outlier SNP. Two lncRNAs (Agly-lnc-000191 and Agly-lnc-000554) on scaffold 114 had Agly000983 as their predicted partner gene (see above).

SNP genotyping of laboratory and field collected soybean aphids

Having identified SNPs in candidate genes (Table 1), we designed molecular assays to genotype both laboratory and additional field collected soybean aphid colonies. For these tests, we included bioassays to describe a unique soybean aphid laboratory colony with virulence to *Rag1/Rag2/Rag3* (3R virulent). Using no-choice tests, there was a significant effect of both soybean aphid biotype ($F = 27.02$, $P = 0.002$) and soybean genotype ($F = 5.36$, $P = 0.039$) but not their interaction ($F = 0.73$, $P = 0.41$). Overall, the 3R virulent colony had higher reproduction on both resistant plant types (*Rag1/Rag2* and *Rag1/Rag2/Rag3*) than the biotype 4 colony (Figure 3). The 3R virulent colony had the highest reproduction on *Rag1/Rag2* (273 ± 6.22 aphids), which was significantly different than all other treatments (Figure 3). Reproduction of this colony on *Rag1/Rag2/Rag3* (102.25 ± 9.06) was significantly larger than reproduction of the biotype 4 on *Rag1/Rag2/Rag3* (24.25 ± 6.22) and on *Rag1/Rag2* (48.75 ± 12.16), (Figure 3) and demonstrated a stronger virulence phenotype than biotype 4.

The outlier SNPs identified through genome scans and differential gene expression (Table 1) were used for genotype comparisons of biotype 1, biotype 4 and 3R virulent soybean aphids. By analyzing the flanking sequences, we identified 8 additional SNPs and a 3bp insertion (scaffold 279) (see Supplementary Table S8 for all SNPs); however all new SNPs were in complete linkage disequilibrium with allelic variation at the predicted SNPs. To limit bias, we retained the original outlier allele from each gene for population genetic analyses. Surprisingly, both biotype 1 and biotype 4 laboratory colonies shared the same genotypes at all analyzed loci.

The 3R virulent colony had fixed differences at 9 SNPs when compared to both biotype 1 and biotype 4 laboratory colonies (Supplementary Table S8). Screening of field collected aphids revealed nine biotype 1 aphids, and six 3R virulent aphids. When comparing biotype 1 to 3R virulent aphids, there was a significant difference in allele frequencies based on Fisher's exact test over all SNPs ($P=0.003$) and with SNPs associated with 4 genes: Agly004871 & Agly004838 ($P = 0.006$), Agly001847 ($P = 0.006$), and Agly001843 ($P = 0.008$, Table 2). F_{ST} analyses showed similar differentiation with an overall value of 0.162 ($P=0.022$) and significant F_{ST} with the same genes: Agly004871 & Agly004838 (0.477, $P = 0.014$), Agly001847 (0.412, $P = 0.021$) and Agly001843 (0.412, $P = 0.012$, Table 2). No significant genotyping differences existed with field collected biotype 4 aphids. Single-locus genotype-fitness association analyses found a significant correlation between reproductive output and the allele copy number for the same three SNPs with frequency differences. SNPs associated with Agly004871 & Agly004838 and Agly001843 showed positive additive effects on reproductive output (regression slope $\beta = 7.10$ and $\beta = 7.29$ respectively) indicating that the presence of the alternative allele was associated with higher reproductive output (Figure 4). Another SNP near Agly001847 showed a significant, negative additive effect ($\beta = -7.29$) indicating that the presence of alternative alleles was associated with a decrease in reproduction (Figure 4). The genotyping and phenotyping of additional aphids provided strong associations between allelic variation and reproductive success when exposed to *Rag* resistance.

Discussion

The adaptation of insects continues to impact the sustainability and productivity of agroecosystems. In some cases, the mechanisms of adaptation are well-known. Yet we lack a solid understanding in other insect pest systems, especially those that solely depend on natural plant defense against insects. We investigated genomic and molecular mechanisms associated with the soybean aphid rapidly overcoming natural resistance in soybean. Critical to our analysis was combining signatures of selection in field collected samples with differences in gene expression in laboratory colonies. This combination showed concurring evidence that virulence adaptation is associated with changes in DNA sequence and expression of genes and gene regulatory pathways. By integrating these data sets, we improved our understanding of soybean aphid virulence and provided additional data that support a complex genetic and regulatory basis for facilitating rapid adaptation in agroecosystems (Pélissié et al., 2018).

Population genomics of field-collected soybean aphids

For most insect crop pests, large populations usually generate high levels of genetic diversity (Singh et al., 2021; Zhan et al., 2021). However, aphids tend to be clonal for most, if not all, of their life cycle, which can decrease genetic diversity (Harrison & Mondor, 2011). Populations in our study showed an overall low level of genetic diversity yet were consistent with previous studies that used less molecular marker data (Kim et al., 2016; Orantes et al., 2012; Wenger & Michel, 2013). Low levels of genetic diversity likely reflect; 1) the recent invasion of North America ca. 2000; 2) seasonal bottlenecks while migrating to and from their winter host (buckthorn trees, *i.e.* *Rhamnus* spp.) and soybeans (Orantes et al., 2012); 3) impacts from pest control. Low levels in genetic diversity can also result in low levels of genetic differentiation among populations, which was apparent in our study. However, in western populations, we detected a slight but significant signal of differentiation between biotype 1 and biotype 4. Reduced genome-wide diversity driven by demographic processes may limit power to detect

locus-specific signals of divergence selection, particularly for variants under recent and/or weak selection. Furthermore, linkage disequilibrium may be artificially large in natural soybean aphid populations due to the founder effect, repeated seasonal bottlenecks, and clonal amplification. Therefore, our genomic scan approach utilized multiple methods to identify outliers with strong allele frequency differentiation between biotypes, consistent with locus-specific directional selection. These outliers then were integrated with differential gene expression and validated with additional field collections. Consequently, this approach may have missed loci with small fitness differences and where selection acted on standing genetic variation that produces more diffuse genomic signals not well captured by differentiation-based genome scans. Although our approach may have been strictly conservative in identifying outliers, it reduced the likelihood of including false positives. This bias likely explains the absence of outlier signals near genes traditionally implicated in aphid adaptation to host plant resistance (e.g. effector proteins) or those involved with detoxification (e.g. P450) (Bansal et al., 2014; Coates et al., 2020). Future studies investigating signatures of selection in the soybean aphid would benefit by incorporating and modeling the complexities of the soybean aphid's life history, demographics and recombination using, for example, simulation-based approaches (e.g. Approximate Bayesian Computation, deep learning frameworks, Flagel et al., 2019; Pavinato et al., 2022) to provide different balances between false positives and true outliers. Until such time, additional analyses are needed to strengthen inference from genome scans. For this study we integrated outliers with gene expression changes among soybean aphid biotypes to add confidence in candidate genes potentially involved with virulence.

Candidate genes and lncRNAs for soybean aphid virulence

Previous research suggests that aphids have adapted to host plants through multiple mechanisms such as odor reception, detoxification, or effector proteins. In some cases, polymorphisms have been found (e.g. odor reception, detoxification), although additional functional testing has been lacking. In other cases, gene expression patterns differentiate adapted or virulent phenotypes (such as effectors or cytochrome P450s), but have limited polymorphisms that may change protein expression, function or recognition by plants especially in natural populations. Many of our 16 candidate genes did not fit with the typical pattern, with a single cytochrome P450 (Agly002329) and no predicted effectors (Table 1). Instead, many of our candidate genes had roles involved in expression and regulation processes, with five scaffolds (114, 129, 237, 255 and 615) having multiple candidates. Interestingly, 10 out of 16 candidate genes mapped to the X-chromosome, which has been known to exhibit faster evolution compared to autosomes (Jaquiéry et al. 2012). All our candidate genes may represent regulatory control points through which selection acts on downstream virulence pathways. For example, Scaffold 114 had 3 candidates: two uncharacterized proteins, and a gene related to histone modification and epigenetic activation. Scaffold 255 contained a *PIWI* gene (Agly017739) (Supplementary Material, Figure S14) that was recently confirmed to be differentially expressed among biotype 1 and biotype 4 aphids (Haller et al., 2025). These genes indicate a possible role for epigenetic regulation in virulence.

Based on genotyping of laboratory and field collections, a few SNPs showed significant allele frequency differences (Table 2) which may suggest their associated genes have a specific role in virulence and plant defense interaction. Two genes were related to *lachesin*, which belong to the immunoglobulin superfamily and contribute to cell adhesion and nervous system development. Although a role in aphid-plant interactions is not known, the comma butterfly (*Polygonia c-album*) differentially expresses a *lachesin*-like gene as part of a plastic transcriptional response

upon plant host switching (Schneider et al., 2024) and could activate similar pathways as plant defense. Another gene was *ATPase ASNA1* homolog (Agly004871) which has functions related to transporting tail-anchored proteins to the endoplasmic reticulum (Hedge and Keenan 2011; Xing et al., 2017). Effectors from both plant pathogens and insects can have C-terminal tail anchor motifs that interact with the plant endoplasmic reticulum (Breeze et al., 2023, Yan et al., 2023) via the ATP-dependent guided entry of tail (GET) proteins pathway (Breeze et al., 2023); these pathways are similar to those used by *ASNA1*. There could be an unidentified soybean aphid effector that interacts with this *ASNA1* to alter soybean interactions. *ASNA1* has also been involved with detoxification in *C. elegans* (Raj et al., 2021) and the parasitic protist *Eimeria tenella* (Yu et al., 2021). *ASNA1* is sensitive to reactive oxygen species (ROS) and oxidative stress (Raj et al., 2021), which are common mechanisms of plant resistance to insects, especially aphids (Goggin and Fischer 2022). A redox sensitive *ASNA1* could help a virulent soybean aphid tolerate a ROS response in *Rag* soybean which has been reported (Natukunda et al., 2021). Located close to *ASNA1* was a *ZIP1* zinc transporter (Agly004838) which had significant gene expression differences among aphids on *Rag1/Rag2* plants and marginally significant differences on susceptible plants (Table 1). Zinc influences insect-plant-microbe interactions (Khan et al., 2023), and higher zinc availability can increase virulence in plant pathogens (Cabot et al., 2019; Kahn et al., 2023). Two zinc transporters were differentially expressed among biotypes of the cereal aphid, *Sitobion avenae* as part of its transcriptional plasticity in responding to defense in different host plants (Wang et al., 2020). Previous research indicated that soybean aphids have less reproductive success on soybean grown under zinc deficiency or excess (Helfenstein et al., 2015) and may require differential expression of zinc transporters (Cabot et al., 2019). These genes are undergoing further characterization in our laboratory.

The role of lncRNAs in aphids has only recently been investigated (Chen et al., 2020; Shang et al., 2021), but research does support a role for plant host feeding, adaptation to specific host-plant defenses and insecticide resistance (Meng et al., 2021; Olivares-Castro et al., 2021; Peng et al., 2021; Shi et al., 2022). In our analysis of lncRNAs in the soybean aphid, we found that most of our lncRNAs were upregulated in the biotype 4 colony, including 5 that were located near an outlier SNP. Partner genes of 5 lncRNAs were also differentially expressed and near outliers (Table 1). Whether lncRNAs are involved with virulence through direct interactions (Chen et al., 2020) or indirectly via regulation and partner genes is unknown. Additionally, we were unable to examine expression differences in field collected soybean aphids (since only DNA was preserved). However, future investigations could add to the evidence of lncRNA's role in insect-plant interactions.

670 Soybean aphids with virulence to *Rag1/Rag2/Rag3*

Currently, only 4 soybean aphid biotypes have been officially described; however, some data show varied responses to soybean varieties with aphid resistance (Lagos-Kutz et al., 2020). We discovered aphids with virulence to *Rag1/Rag2* and *Rag1/Rag2/Rag3* and confirmed their presence by screening field collected soybean aphids. Our 3R virulent laboratory colony had significantly higher reproduction compared to biotype 4 on *Rag1/Rag2/Rag3* (Figure 3). Solely based on laboratory colonies, the 3R virulent laboratory colony had fixed differences at 9 SNPs compared to biotype 1 and 4 (Supplementary Table S8). This pattern was partially supported with genotypes from field samples; especially with SNPs in genes Agly004871 & Agly004838, Agly001847 and Agly001843 (Table 2, Figure 4). The complexity of connecting genotype to phenotype for virulence could suggest a continuous basis for virulence rather than discrete, and

one based on many genes with minor effect (Simon and Peccoud 2018; Singh et al., 2020). This is further supported by the variation in genotypes and phenotypes seen from our field screening. Many alleles among our laboratory colony comparisons were fixed but then were variable in soybean aphids collected from the field. The opposite pattern occurred with the SNPs near Agly004871 & Agly004838: the 3R virulent colony was homozygous along with 5 out of the 6 field collected 3R virulent aphids supporting the strong correlation (Figure 4), with fixed differences between the biotype 1 and biotype 4 soybean aphids (Table 2, Supplementary Table S8, and Figure 4). In some cases, the allele in 3R virulent aphids was the alternative allele and did not match the biotype 4 reference genome, suggesting that the biotype 4 reference genome may contain a mixture of biotype 4 and 3R virulent aphids (since our initial screening was based on survival and reproduction on only *Rag1/Rag2*). Additionally, heterozygosity is not represented in reference genome sequences, yet our Sanger sequences showed heterozygosity within biotype 4 at some of the SNPs; the allele represented in the genome sequence may not be associated with virulence. A further complication would be if virulence is partially recessive or involve allele-specific expression (potentially via differential regulatory pathways or lncRNA expression). For example, as a homozygote, the 3R virulent aphid would express both copies of a gene whereas biotype 4, as a heterozygote, would express 1 copy (or none if there was allele specific expression). Our data indicates the importance of field collections; while research with laboratory colonies should continue, these genomes do not exist in natural populations due to obligatory sexual reproduction. Selection in the soybean aphid occurs in agroecosystems and must be incorporated to further refine the basis of virulence. Our list of candidate genes and lncRNAs provides a solid foundation for further field screening, genotyping, gene expression and epigenetic differentiation.

Conclusion

Our study provided evidence that virulence in the soybean aphid may be multi-genic or plastic and depend on many interacting pathways. While previous research on virulence and host-plant adaptation in aphids and pathogens has focused on effectors, this study highlighted the importance of genes involved in regulation, lncRNAs and zinc transport. Fully unraveling virulence adaptation requires additional research that confirms differential expression of genes, lncRNAs, and potentially other epigenetic factors along with corresponding evaluation of gene function and the virulence phenotype (*e.g.* RNA interference). Soybean aphid virulence likely prevents the wide-scale adoption of aphid-resistant soybean (Tilmon et al., 2021) and, without a full understanding on virulence, these soybean varieties may see limited deployment. The occurrence of yet another biotype suggests that soybean aphids are highly adaptable, despite a complex basis for virulence. Other strategies should be considered to extend the durability of host-plant resistance against the soybean aphid, such as including a susceptible refuge (Wenger et al. 2014; Vahrenhorst and O'Neal 2015). As evidence for insecticide resistance in soybean aphid increases and spreads (Koch et al. 2018), soybean producers will need durable management tools like host-plant resistance. Improving our understanding of virulence will help plant resistance remain a viable and effective tool to protect against crop damage and ensure sustainable crop production.

Data availability

The pipeline developed to process the *Aphis glycines* pooled-seq data can be found at <https://github.com/vitorpavinato/AglyPoolseq>. All scripts necessary to reproduce the analysis presented here can also be found at the pipeline's page. Whole genome and transcriptome data were already deposited in NCBI under the accession numbers: PRJNA312843 (Wenger et al., 2020) and PRJNA598670 (Yates-Stewart et al., 2020). Annotated VCF file, the annotation of Ya genes and lncRNAs on the soybean aphid virulent genome v2.1 and read counts tables for lncRNA expression analyses can be found at zenodo (<https://zenodo.org/records/8341943>).

Acknowledgements

This project received funding from the North Central Soybean Research Program, The Ohio Soybean Council, and the Center for Applied Plant Sciences at OSU. S. Shahid was funded by a fellowship from the Life Sciences Research Foundation and Simons Foundation, and Keith Slotkin was funded by grant MCB-1904326 from the US National Science Foundation. We are thankful to the Ohio Supercomputer Center (OSC, 1987) for providing computing resources. L. Hesler and E. Beckendorf (USDA-ARS) provided samples from South Dakota. C. Wallace, M. Lewis, E. Lang and T. Elek assisted with rearing of the 3R virulent aphid. We received aphid collections from D. Prischmann-Voldseth (ND), L. Hesler (SD), B. Potter (MN), M. Crossley (WI), J. Tooker (PA).

Contributions

V.A.C.P. wrote drafts of the manuscript, reanalyzed the population genomics, transcriptomics and lncRNA data, and integrated the datasets. E. P-M. generated and analyzed sequencing data, characterized SNP locations and gene annotations, performed and analyzed bioassays, contributed to conceptualization and manuscript drafts; J.A.W. and A.M. conceptualized the initial work on field-samples population genomics and designed the collections, generated the data set and wrote initial drafts. S.S. and R.K.S contributed to manuscript drafts, data analysis and integration with regulation and pathways.

Conflict of Interest Statement

The authors declare no conflicts of interest.

Ethics Statement

All the research complies with applicable laws on sampling from natural populations and animal experimentation (including the ARRIVE guidelines). All insects were collected from university research farms, and sent with USDA-APHIS-PPQ permits. Data are publicly available.

Figure Legends

Figure 1. Aphid collection and phenotyping. A) Aphids were collected at multiple localities across an East-West transect of the species North American range. Two cultivars: a susceptible (SD-01-76R) and a resistant were planted in each locality to collect avirulent and virulent biotypes. B) Aphids from all cultivars were placed individually in Petri dishes with leaves from susceptible plants for colony establishment in the laboratory. Adults were then placed in different cultivars for phenotyping.

Figure 2. Candidate SNPs and genes under selection. Venn diagrams showing A) the number of SNPs identified by genome scan methods that associate SNP frequencies with the biotype as covariate, B) SNP markers identified by genome scan methods that detect signatures of geographic differentiation among pools, C) Genes close to outlier SNPs identified with an association to biotype, and D) Genes close to outlier SNPs identified with geographic differentiation. All genes close to significant SNPs were subject to the gene ontology enrichment analysis. Names of tests are listed near the diagrams (see Methods).

Figure 3. Reproduction on no-choice tests using whole plants and virulent soybean aphids. Soybean cultivars possessed resistance at *Rag1/Rag2* (LD14-8001) and *Rag1/Rag2/Rag3* (LD14-8003). Aphids were counted 10 days post-infestation. Solid bars represent biotype 4 and stippled bars represent 3R virulent soybean aphids. Large dots represent data point. Bars labeled with different letters are significantly different ($P < 0.05$) based on ANOVA.

Figure 4. Single-locus additive association tests for targeted outlier SNPs in new field samples. Box plots showing the distributions of individual reproductive output for the three genotypes (Ref/Ref, Ref/Alt, Alt/Alt) for each SNP exhibiting a significant additive genotype-fitness association. Effect sizes (β) correspond to per-allele slope from additive linear regressions. SNPs near Agly004871 and Agly001843 showed significant positive effects on reproductive output (regression slopes $\beta = 7.10$ and $\beta = 7.29$ respectively), whereas the SNP near Agly001847 showed a significant negative effect ($\beta = -7.29$).

References

- Alexa, A., & Rahnenfuhrer, J. (2023). *topGO: Enrichment analysis for gene ontology* (Version 2.54.0). Bioconductor. <https://bioconductor.org/packages/topGO>
- Altieri, M. A. (1999). The ecological role of biodiversity in agroecosystems. *Agriculture, Ecosystems & Environment*, 74(1–3), 19–31.
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq: A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169.
- Bansal, R., Mian, M. A., Mittapalli, O., & Michel, A. P. (2014). RNA-seq reveals a xenobiotic stress response in the soybean aphid, *Aphis glycines*, when fed aphid-resistant soybean. *BMC Genomics*, 15(1), 972.
- Bass, C., & Nauen, R. (2023). The molecular mechanisms of insecticide resistance in aphid crop pests. *Insect Biochemistry and Molecular Biology*, 156, e103937.
- Breeze, E., Vale, V., McLellan, H., Pecrix, Y., Godiard, L., Grant, M., & Frigerio, L. (2023). A tell-tail sign: A conserved C-terminal tail-anchor domain targets a subset of pathogen effectors to the plant endoplasmic reticulum. *Journal of Experimental Botany*, 74(10), 3188–3202.
- Brevik, K., Bueno, E. M., McKay, S., Schoville, S. D., & Chen, Y. H. (2021). Insecticide exposure affects intergenerational patterns of DNA methylation in the Colorado potato beetle, *Leptinotarsa decemlineata*. *Evolutionary Applications*, 14(3), 746–757.
- Bushnell, B., Rood, J., & Singer, E. (2017). BBMerge: Accurate paired shotgun read merging via overlap. *PLoS Genetics*, 12(10), e0185056.
- Cabot, C., Martos, S., Llugany, M., Gallego, B., Tolrà, R., & Poschenrieder, C. (2019). A role for zinc in plant defense against pathogens and herbivores. *Frontiers in Plant Science*, 10, 1171.
- Chen, Y., Singh, A., Kaithakottil, G. G., Mathers, T. C., Gravino, M., Mugford, S. T., van Oosterhout, C., Swarbreck, D., and Hogenhout, S., A. (2020). An aphid RNA transcript migrates systemically within plants and is a virulence factor. *Proceedings of the National Academy of Sciences*, 117(23), 12763–12771.
- Coates, B. S., Hohenstein, J. D., Giordano, R., Donthu, R. K., Michel, A. P., Hodgson, E. W., & O’Neal, M. E. (2020). Genome scan detection of selective sweeps among biotypes of the soybean aphid, *Aphis glycines*, with differing virulence to resistance to *A. glycines* (Rag) traits in soybean, *Glycine max*. *Insect Biochemistry and Molecular Biology*, 124, 103364.
- Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, 185(4), 1411–1423.
- Du, B., Chen, R., Guo, J., & He, G. (2020). Current understanding of the genomic, genetic, and molecular control of insect resistance in rice. *Molecular Breeding*, 40(2), 24.
- Emmerson, M., Morales, M. B., Oñate, J. J., Batáry, P., Berendse, F., Liira, J., Aavik, T., Guerrero, I., Bommarco, R., Eggers, S., Pärt, T., Tschardtke, T., Weisser, W., Clement, L., Bengtsson, J. (2016). Chapter two: How agricultural intensification affects biodiversity and ecosystem services. *Advances in Ecological Research*, 55, 43–97.

- 819 Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to
820 perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*,
821 10(3), 564–567.
- 822 Flagel, L., Brandvain, Y., & Schrider, D. R. (2019). The unreasonable effectiveness of
823 convolutional neural networks in population genetic inference. *Molecular Biology and Evolution*,
824 36, 220–238.
- 825 Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-
826 specific covariates. *Genetics*, 201(4), 1555–1579.
- 827 Glastad, K. M., Hunt, B. G., & Goodisman, M. A. D. (2019). Epigenetics in insects: Genome
828 regulation and the generation of phenotypic diversity. *Annual Review of Entomology*, 64, 185–
829 203.
- 830 Goggin, F. L., & Fischer, H. D. (2022). Reactive oxygen species in plant interactions with
831 aphids. *Frontiers in Plant Science*, 12, 811105.
- 832 Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele
833 frequencies. *Genetics*, 195(1), 205–220.
- 834 Haller, A., Poelstra, J. W., Pipatpongpinoy, W., Kreuter, N., Wilson, J. R., & Michel, A. (2025).
835 Investigating a role for piRNA-associated piwi genes in overcoming host-plant resistance in the
836 soybean aphid, *Aphis glycines*. *Journal of Insect Science*, 25(3), 19.
- 837 Harrison, J. S., & Mondor, E. B. (2011). Evidence for an invasive aphid “superclone”: Extremely
838 low genetic diversity in oleander aphid (*Aphis nerii*) populations in the southern United States.
839 *PLoS ONE*, 6(3), e17524.
- 840 Hawkins, N. J., Bass, C., Dixon, A., & Neve, P. (2019). The evolutionary origins of pesticide
841 resistance. *Biological Reviews*, 94(1), 135–155.
- 842 Hegde, R. S., & Keenan, R. J. (2011). Tail-anchored membrane protein insertion into the
843 endoplasmic reticulum. *Nature Reviews Molecular Cell Biology*, 12(12), 787–798.
- 844 Heimpel, G. E., L. E. Frelich, D. A. Landis, K. R. Hopper, K. A. Hoelmer, Z. Sezen, M. K.
845 Asplen, and K. Wu. (2010). European buckthorn and Asian soybean aphid as components of an
846 extensive invasional meltdown in North America. *Biological Invasions* 12:2913-2931.
- 847 Hesler, L.S., Beckendorf, E.A. (2020). Dataset: Soybean aphids per plant among soybean lines
848 containing various *Rag* genes. Ag Data Commons.
- 849 Helfenstein, J., Pawlowski, M. L., Hill, C. B., Stewart, J., Lagos-Kutz, D., Bowen, C. R.,
850 Frossard, E., & Hartman, G. L. (2015). Zinc deficiency alters soybean susceptibility to pathogens
851 and pests. *Journal of Plant Nutrition and Soil Science*, 178(6), 896–903.
- 852 Hirsch, C. D., & Springer, N. M. (2017). Transposable element influences on gene expression in
853 plants. *Biochimica et Biophysica Acta (BBA) – Gene Regulatory Mechanisms*, 1860(1), 157–165.
- 854 Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic
855 differentiation from pool-seq data. *Genetics*, 210(1), 315–330.
- 856 Hogenhout, S. A., & Bos, J. I. (2011). Effector proteins that modulate plant–insect interactions.
857 *Current Opinion in Plant Biology*, 14(4), 422–428.

- 858 Jaquiéry, J. Stoeckel, S., Rispe, C., Mieuze, L., Legeai, F., Simon, J.-C. (2012). Accelerated
859 evolution of sex chromosomes in aphids, an XO System. *Molecular Biology and Evolution*,
860 29(2), 837–847.
- 861 Joshi, K., Baumgardner, J.L., MacPhail, M., Acharya, S.R., Blotevogel, E., Dayan, F.E.,
862 Nachappa, P. and Nalam, V.J. (2021). The source of *Rag5*-mediated resistance to soybean aphids
863 is located in the stem. *Frontiers in Plant Science*, 12:689986.
- 864 Kapun, M., Barrón, M. G., Staubach, F., Obbard, D. J., Wiberg, R. A. W., Vieira, J., ...
865 Gonzalez, J. (2020). Genomic analysis of European *Drosophila melanogaster* populations
866 reveals longitudinal structure, continent-wide selection, and previously unknown DNA viruses.
867 *Molecular Biology and Evolution*, 37(9), 2661–2678.
- 868 Kent, W. J. (2002). BLAT: The BLAST-like alignment tool. *Genome Research*, 12(4), 656–664.
- 869 Kim, H., Hoelmer, K. A., & Lee, S. (2016). Population genetics of the soybean aphid in North
870 America and East Asia: Test for introduction between native and introduced populations.
871 *Biological Invasions*, 19(2), 597–614.
- 872 Khan, S., & Lang, M. (2023). A comprehensive review on the roles of metals mediating insect–
873 microbial pathogen interactions. *Metabolites*, 13(7), 839.
- 874 Koch, R.L., Hodgson, E. W., Knodel, J. J., Varenhorst, A. J., Potter, B. D. (2018). Management
875 of insecticide-resistant soybean aphids in the Upper Midwest of the United States. *Journal of*
876 *Integrated Pest Management*, 9, 23.
- 877 Lagos-Kutz, D., Pawlowski, M. L., Diers, B. W., Purandare, S. R., Tilmon, K. J., & Hartman, G.
878 L. (2020). Virulence of soybean aphid, *Aphis glycines* (Hemiptera: Aphididae) clones on
879 detached leaves and whole plants. *Journal of the Kansas Entomological Society*, 92, 497–511.
- 880 Leblois, R., Gautier, M., Rohfritsch, A., Foucaud, J., Burban, C., Galan, M., Loiseau, A., Sauné,
881 L., Branco, M., Gharbi, K., Vitalis, R., Kerdelhué, C. (2017). Deciphering the demographic
882 history of allochronic differentiation in the pine processionary moth *Thaumetopoea pityocampa*.
883 *Molecular Ecology*, 27(1), 264–278.
- 884 Lewis, S. H., Ross, L., Bain, S. A., Pahita, E., Smith, S. A., Cordaux, R., Miska, E. A., Lenhard,
885 B., Jiggins, F. M., & Sarkies, P. (2020). Widespread conservation and lineage-specific
886 diversification of genome-wide DNA methylation patterns across arthropods. *PLoS Genetics*,
887 16(6), e1008864.
- 888 Lewontin, R. C., & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of
889 selective neutrality of polymorphisms. *Genetics*, 74(1), 175–195.
- 890 Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
891 *arXiv Preprint arXiv:1303.3997*.
- 892 Li, M., Xiao, H., He, K., & Li, F. (2019). Progress and prospects of noncoding RNAs in insects.
893 *Journal of Integrative Agriculture*, 18(4), 729–747.
- 894 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads.
895 *EMBnet journal*, 17(1), 10–12.

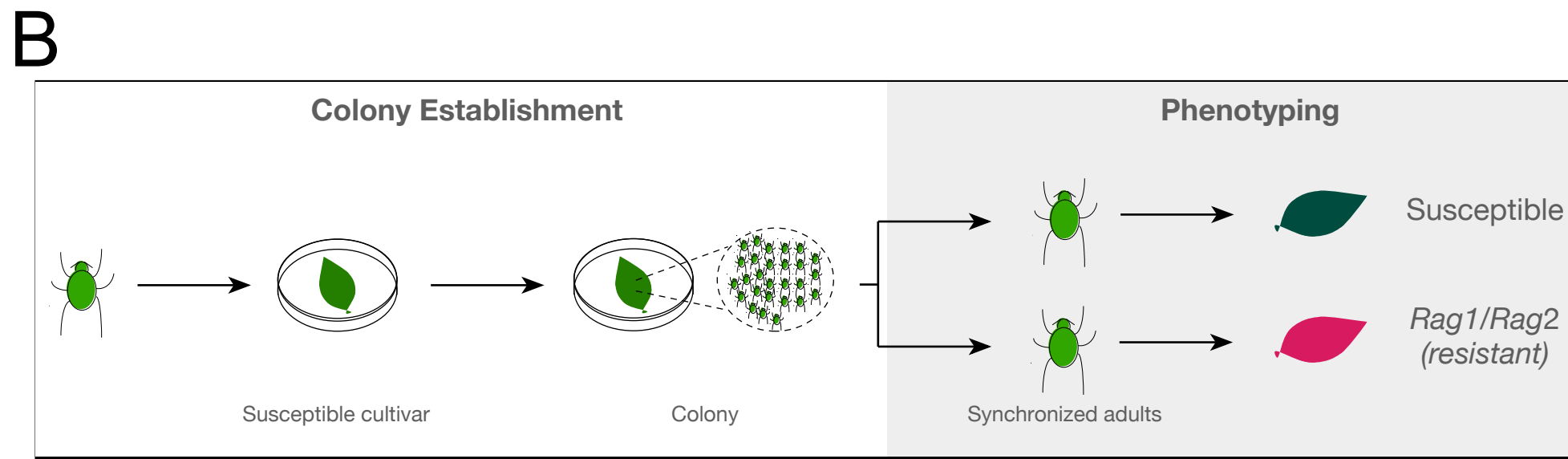
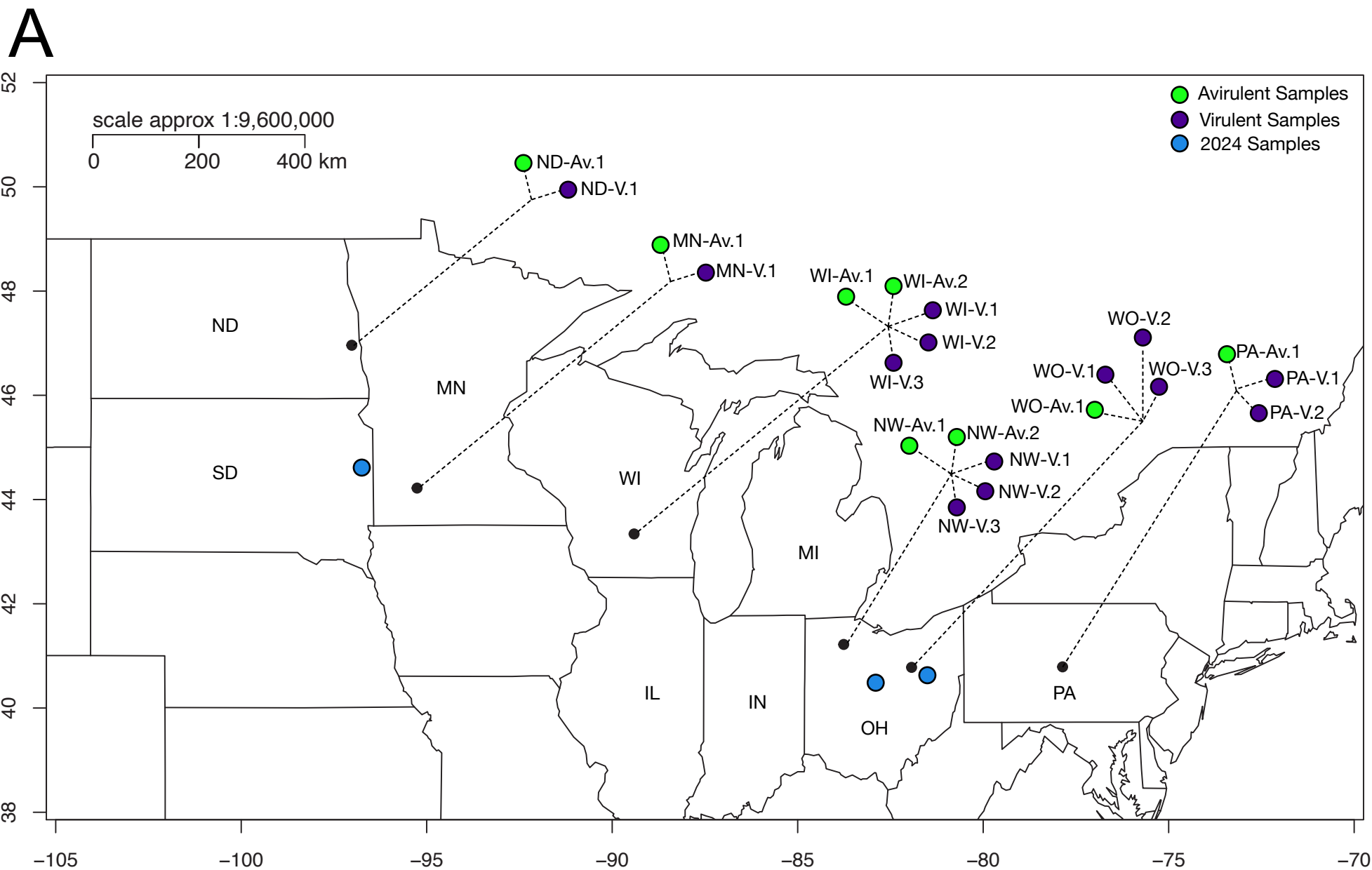
- Mathers, T. C., Chen, Y., Kaithakottil, G., Legeai, F., Mugford, S. T., Baa-Puyoulet, P., ... Hogenhout, S. A. (2017). Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. *Genome Biology*, 18(1), 27.
- Mathers, T. C., Mugford, S. T., Percival-Alwyn, L., Chen, Y., Kaithakottil, G., Swarbreck, Hogenhout, S., A., van Oosterhout, C. (2019). Sex-specific changes in the aphid DNA methylation landscape. *Molecular Ecology*, 28(18), 4228–4241.
- Mathers, T. C. (2020). Improved genome assembly and annotation of the soybean aphid (*Aphis glycines* Matsumura). *G3: Genes, Genomes, Genetics*, 10, 899–906.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.
- Michel, A.P., Zhang, W., Kyo Jung, J., Kang, S.T., Mian, M.A. (2009). Population genetic structure of *Aphis glycines*. *Environmental Entomology*. 38(4):1301-11.
- Mei, Y., Jing, D., Tang, S., Chen, X., Chen, H., Duanmu, H. Cong, Y., Chen, M., Ye, X., Zhou, H., He, K., Li, F. (2021). InsectBase 2: A comprehensive gene resource for insects. *Nucleic Acids Research*, 50(1), 1040–1045.
- Meng F, Han Y, Teng W, Li Y, Li W. (2011). QTL underlying the resistance to soybean aphid (*Aphis glycines* Matsumura) through isoflavone-mediated antibiosis in soybean cultivar ‘Zhongdou 27’. *Theoretical and Applied Genetics*, 23(8):1459–1465.
- Meng, L., Yuan, G., Chen, M., Dou, W., Jing, T., Zheng, L., Peng, M., Bai, W., Wang, J. (2021). Genome-wide identification of long non-coding RNAs (lncRNAs) associated with malathion resistance in *Bactrocera dorsalis*. *Pest Management Science*, 77(5), 2292–2301.
- Morales, A. C., Lavagnini, T. C., & Freitas, S. (2013). Loss of genetic variability induced by agroecosystems: *Chrysoperla externa* (Hagen) (Neuroptera: Chrysopidae) as a case study. *Neotropical Entomology*, 42(1), 32–38.
- Natukunda, M. I., Hohenstein, J. D., McCabe, C. E., Graham, M. A., Qi, Y., Singh, A. K., & MacIntosh, G. C. (2021). Interaction between Rag genes results in a unique synergistic transcriptional response that enhances soybean resistance to soybean aphids. *BMC Genomics*, 22(1), 887.
- Natukunda, M. I., & MacIntosh, G. C. (2020). The resistant soybean–*Aphis glycines* interaction: Current knowledge and prospects. *Frontiers in Plant Science*, 11, 1223.
- Ninova, M., Godneeva, B., Chen, Y. C. A., Luo, Y., Prakash, S. J., Jankovics, F., Erdélyi, M., Aravin, A., A., Tóth, K., F., (2020). The SUMO ligase Su(var)2-10 controls hetero- and euchromatic gene expression via establishing H3K9 trimethylation and negative feedback regulation. *Molecular Cell*, 77(3), 571–585.e4.
- Ninova, M., Griffiths-Jones, S., & Ronshaugen, M. (2017). Abundant expression of somatic transposon-derived piRNAs throughout *Tribolium castaneum* embryogenesis. *Genome Biology*, 18(1), 184.

- 935 Nouhaud, P., Gautier, M., Gouin, A., Jaquiéry, J., Peccoud, J., Legeai, F., Mieuzet, L., Smadja,
936 C., M., Lemaitre, C., Vitalis, R., Simon, J. (2018). Identifying genomic hotspots of
937 differentiation and candidate genes involved in the adaptive divergence of pea aphid host races.
938 *Molecular Ecology*, 27(16), 3287–3300.
- 939 Olivares-Castro, G., Cáceres-Jensen, L., Guerrero-Bosagna, C., & Villagra, C. (2021). Insect
940 epigenetic mechanisms facing anthropogenic-derived contamination: An overview. *Insects*,
941 12(9), 780.
- 942 Orantes, L. C., Zhang, W., Mian, M. A. R., & Michel, A. P. (2012). Maintaining genetic
943 diversity and population panmixia through dispersal and not gene flow in a holocyclic
944 heteroecious aphid species. *Heredity*, 109(2), 127–134.
- 945 Ovčarenko, I., Kapantaidaki, D. E., Lindström, L., Gauthier, N., Tsagkarakou, A., Knott, K. E.,
946 & Vänninen, I. (2014). Agroecosystems shape population genetic structure of the greenhouse
947 whitefly in northern and southern Europe. *BMC Evolutionary Biology*, 14(1), 165.
- 948 Pavinato, V. A. C., Di Mita, S., Marin, J.-M., & Navascués, M. de. (2022). Joint inference of
949 adaptive and demographic history from temporal population genomic data. *Peer Community*
950 *Journal*, 2, e78.
- 951 Péliissié, B., Crossley, M. S., Cohen, Z. P., & Schoville, S. D. (2018). Rapid evolution in insect
952 pests: the importance of space and time in population genomics studies. *Current Opinion in*
953 *Insect Science*, 26, 8–16
- 954 Péliissié, B., Chen, Y. H., Cohen, Z. P., Crossley, M. S., Hawthorne, D. J., Izzo, V., & Schoville,
955 S. D. (2022). Genome resequencing reveals rapid, repeated evolution in the Colorado potato
956 beetle. *Molecular Biology and Evolution*, 39(2), msac016
- 957 Peng, T., Pan, Y., Tian, F., Xu, H., Yang, F., Chen, X., Gao, X., Li, J., Wang, H., Shang, Q.
958 (2021). Identification and the potential roles of long non-coding RNAs in regulating acetyl-CoA
959 carboxylase ACC transcription in spirotetramat-resistant *Aphis gossypii*. *Pesticide Biochemistry*
960 *and Physiology*, 179, 104972.
- 961 Petersen, M., Armisen, D., Gibbs, R. A., Hering, L., Khila, A., Mayer, G., Richards, S., Niehuis,
962 O., Misof, B. Diversity and evolution of the transposable element repertoire in arthropods with
963 particular reference to insects. *BMC Ecology and Evolution*, 19(1), 11.
- 964 Preez, P. H. d., Breeds, K., Burger, N. F. V., Swiegers, H. W., Truter, J. C., & Botha, A.-M.
965 (2020). DNA methylation and demethylation are regulated by functional DNA
966 methyltransferases and DnTET enzymes in *Diuraphis noxia*. *Frontiers in Genetics*, 11, 452.
- 967 Prischmann-Voldseth, D. A., Helser, L., Beckendorf, E., Penton-Michel, E.H., Michel A. No
968 benefit of a foliar insecticide on soybean aphid-resistant soybeans when non-aphid pest pressure
969 and aphid-virulence frequency are low. *Journal of Economic Entomology*. In press.
- 970 Qui, S., Wu, N., Sun, X., Xue, Y., Xia, J. (2025). Chromosome-level genome assembly of
971 soybean aphid. *Scientific Data*, 12, 386.
- 972 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing
973 genomic features. *Bioinformatics*, 26(6), 841–842.

- 974 Raj, D., Billing, O., Podraza-Farhanieh, A., Kraish, B., Hemmingsson, O., Kao, G., & Naredi, P.
975 (2021). Alternative redox forms of ASNA-1 separate insulin signaling from tail-anchored protein
976 targeting and cisplatin resistance in *C. elegans*. *Scientific Reports*, 11(1), 8678.
- 977 Ramat, A., & Simonelig, M. (2021). Functions of PIWI proteins in gene regulation: New arrows
978 added to the piRNA quiver. *Trends in Genetics*, 37(2), 188–200.
- 979 Raymond, M., & Rousset, F. (1995). GENEPOP (Version 1.2): Population genetics software for
980 exact tests and ecumenicism. *Journal of Heredity*, 86(3), 248–249.
- 981 Rodrigues, M. F., Vibranovski, M. D., & Cogni, R. (2021). Clinal and seasonal changes are
982 correlated in *Drosophila melanogaster* natural populations. *Evolution*, 75(8), 2042–2054.
- 983 Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for
984 Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106.
- 985 Soares de Melo, E., & Wallau, G. L. (2022). Mosquito long non-coding RNAs are enriched with
986 transposable elements. *Genetics and Molecular Biology*, 45(1), e20210215.
- 987 Sattar, S., Addo-Quaye, C., Song, Y., Anstead, J. A., Sunkar, R., & Thompson, G. A. (2012).
988 Expression of small RNA in *Aphis gossypii* and its potential role in the resistance interaction
989 with melon. *PLoS ONE*, 7(11), e48579.
- 990 Schneider, K., Steward, R. A., Celorio-Mancera, M. d l. P., Janz, N., Moberg, D., Wheat, C. W.,
991 & Nylin, S. (2024). Plasticity for the win: Flexible transcriptional response to host plant switches
992 in the comma butterfly (*Polygonia c-album*). *Molecular Ecology*, 33(16), e17479.
- 993 Shang, F., Ding, B.-Y., Zhang, Y.-T., Wu, J.-J., Pan, S.-T., & Wang, J.-J. (2021). Genome-wide
994 analysis of long non-coding RNAs and their association with wing development in *Aphis*
995 *citricidus* (Hemiptera: Aphididae). *Insect Biochemistry and Molecular Biology*, 139, 103666.
- 996 Shi, L., Li, W., Zeng, H., Shi, Y., & Liao, X. (2022). Systematic identification and functional
997 analysis of long noncoding RNAs involved in indoxacarb resistance in *Spodoptera litura*. *Insect*
998 *Science*, 29(6), 1721–1736.
- 999 Simon, J.-C., & Peccoud, J. (2018). Rapid evolution of aphid pests in agricultural environments.
1000 *Current Opinion in Insect Science*, 26, 17–24.
- 1001 Singh, K. S., Cordeiro, E. M. G., Troczka, B. J., Pym, A., Mackisack, J., Mathers, T. C., ... Bass,
1002 C. (2021). Global patterns in genomic diversity underpinning the evolution of insecticide
1003 resistance in the aphid crop pest *Myzus persicae*. *Communications Biology*, 4(1), 847.
- 1004 Singh, K. S., Troczka, B. J., Duarte, A., Balabanidou, V., Trissi, N., Paladino, ... Bass, C.
1005 (2020). The genetic architecture of a host shift: An adaptive walk protected an aphid and its
1006 endosymbiont from plant chemical defenses. *Science Advances*, 6(19), eaba1070.
- 1007 Slotkin, R. K., & Martienssen, R. (2007). Transposable elements and the epigenetic regulation of
1008 the genome. *Nature Reviews Genetics*, 8(4), 272–285.
- 1009 Smadja, C. M., Canbäck, B., Vitalis, R., Gautier, M., Ferrari, J., Zhou, J. J., & Butlin, R. K.
1010 (2012). Large-scale candidate gene scan reveals the role of chemoreceptor genes in host plant
1011 specialization and speciation in the pea aphid. *Evolution*, 66(9), 2723–2738.

- 1012 Smith, C. M. (2021). Conventional breeding of insect-resistant crop plants: Still the best way to
1013 feed the world population. *Current Opinion in Insect Science*, 45, 7–13.
- 1014 Srinivasan, D. G., & Brisson, J. A. (2012). Aphids: A model for polyphenism and epigenetics.
1015 *Genetics Research International*, 2012, 431531.
- 1016 Tabashnik, B. E. (2015). ABCs of insect resistance to Bt. *PLoS Genetics*, 11(11), e1005646.
- 1017 Tabashnik, B. E., & Carrière, Y. (2017). Surge in insect resistance to transgenic crops and
1018 prospects for sustainability. *Nature Biotechnology*, 35(10), 926–935.
- 1019 Taylor, K. L., Hamby, K. A., DeYonke, A. M., Gould, F., & Fritz, M. L. (2021). Genome
1020 evolution in an agricultural pest following adoption of transgenic crops. *Proceedings of the*
1021 *National Academy of Sciences*, 118(52), e2020853118.
- 1022 Team R. (2025). *R: A language and environment for statistical computing*.
- 1023 Tilmon, K. J., Michel, A., & O’Neal, M. E. (2021). Aphid resistance is the future for soybean
1024 production, and has been since 2004: Efforts towards a wider use of host plant resistance in
1025 soybean. *Current Opinion in Insect Science*, 45, 53–58.
- 1026 Varenhorst, A. J., and O’Neal, M. E. (2015). The effect on an interspersed refuge on *Aphis*
1027 *glycines* (Hemiptera: Aphididae), their natural enemies, and biological control. *Journal of*
1028 *Economic Entomology*, 109, 406–415.
- 1029 Vitalis, R., Gautier, M., Dawson, K. J., & Beaumont, M. A. (2014). Detecting and measuring
1030 selection from gene frequency data. *Genetics*, 196(3), 799–817.
- 1031 Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23(23),
1032 5649–5662.
- 1033 Wang, D., Liu, D., Shi, X., Yang, Y., Zhang, N., & Shang, Z. (2020). Transcriptome profiling
1034 revealed potentially important roles of defensive gene expression in the divergence of insect
1035 biotypes: A case study with the cereal aphid *Sitobion avenae*. *BMC Genomics*, 21, 546.
- 1036 Wang, Y., Wilson, A. E., & Liu, N. (2022). A new method to address the importance of
1037 detoxified enzyme in insecticide resistance: Meta-analysis. *Frontiers in Physiology*, 13, 818531.
- 1038 Wenger, J. A., Cassone, B. J., Legeai, F., Johnston, J. S., Bansal, R., Yates, A. D., Coates, B., S.,
1039 Michel, A., P. (2020). Whole genome sequence of the soybean aphid, *Aphis glycines*. *Insect*
1040 *Biochemistry and Molecular Biology*, 123, 102917.
- 1041 Wenger, J. A., & Michel, A. P. (2013). Implementing an evolutionary framework for
1042 understanding genetic relationships of phenotypically defined insect biotypes in the invasive
1043 soybean aphid (*Aphis glycines*). *Evolutionary Applications*, 6(7), 1041–1053.
- 1044 Wenger, J. A., Ramstad M., Mian, M. A. R., Michel A. (2014). The use of refuge in host plant
1045 resistance systems for the control of virulent biotype adaptation in the soybean aphid
1046 (Hemiptera:Aphididae). *Journal of Economic Entomology*, 107: 1599-1609.
- 1047 Wiarda, S. L., Fehr, W. R., & O’Neal, M. E. (2012). Soybean aphid (Hemiptera: Aphididae)
1048 development on soybean with *Rag1* alone, *Rag2* alone, and both genes combined. *Journal of*
1049 *Economic Entomology*, 105, 252–258.

- 1050 Wooley, S. C., Smith, D. S., Lonsdorf, E. V., Brown, S. C., Whitham, T. G., Shuster, S. M., &
1051 Lindroth, R. L. (2020). Local adaptation and rapid evolution of aphids in response to genetic
1052 interactions with their cottonwood hosts. *Ecology and Evolution*, 10(19), 10532–10542.
- 1053 Xing, S., Mehlhorn, D. G., Wallmeroth, N., Asseck, L. Y., Kar, R., Voss, A., Denninger, P.,
1054 Schmidt, V. A. F., Schwarzländer, M., Stierhof, Y.-D., Grossmann, G., & Grefen, C. (2017).
1055 Loss of GET pathway orthologs in *Arabidopsis thaliana* causes root hair growth defects and
1056 affects SNARE abundance. *Proceedings of the National Academy of Sciences*, 114(8), E1544–
1057 E1553.
- 1058 Yan, Z.-W., Chen, F.-Y., Zhang, X., Cai, W.-J., Chen, C.-Y., Liu, J., Wu, M.-N., Liu, N.-J., Ma,
1059 B., Wang, M.-Y., Chao, D.-Y., Gao, C.-J., & Mao, Y.-B. (2023). Endocytosis-mediated entry of
1060 a caterpillar effector into plants is countered by jasmonate. *Nature Communications*, 14(1), 6551.
- 1061 Yates, A. D., & Michel, A. (2018). Mechanisms of aphid adaptation to host plant resistance.
1062 *Current Opinion in Insect Science*, 26, 41–49.
- 1063 Yates-Stewart, A. D., Daron, J., Wijeratne, S., Shahid, S., Edgington, H. A., Slotkin, R. K., &
1064 Michel, A. (2020). Soybean aphids adapted to host-plant resistance by down-regulating putative
1065 effectors and up-regulating transposable elements. *Insect Biochemistry and Molecular Biology*,
1066 121, 103363.
- 1067 Yu, Y., Dong, H., Zhao, Q., Zhu, S., Liang, S., Wang, Q., Wang, H., Yu, S., Huang, B., & Han,
1068 H. (2021). Molecular characterization and analysis of the ATPase ASNA1 homolog gene of
1069 *Eimeria tenella* in a drug-sensitive strain and drug-resistant strains. *International Journal for*
1070 *Parasitology: Drugs and Drug Resistance*, 15, 115–125.
- 1071 Zhan, H. X., Hao, Z. P., Tang, R., Zhu, L. N., Zhou, J. J., & Hou, S. M. (2021). High genetic
1072 diversity and strong genetic structure of *Strongyloides variegatus* populations in oilseed rape
1073 production areas of China. *BMC Ecology and Evolution*, 21, 18.
- 1074 Zhao, C., Escalante, L. N., Chen, H., Benatti, T. R., Qu, J., Chellapilla, S., ..., Richards, S.R.
1075 (2015). A massive expansion of effector genes underlies gall formation in the wheat pest
1076 *Mayetiola destructor*. *Current Biology*, 25(5), 613–620.

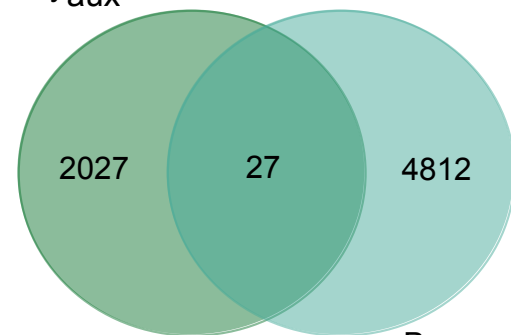


SNPs

Genes of 107

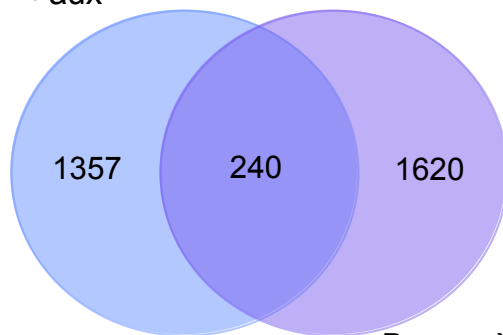
A

Bay_{aux}BFmc



C

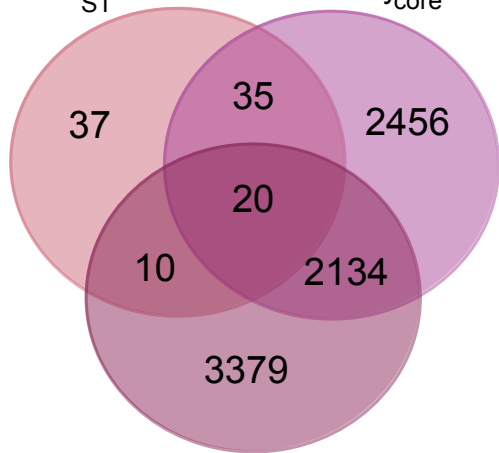
Bay_{aux}BFmc



B

M- F_{ST}

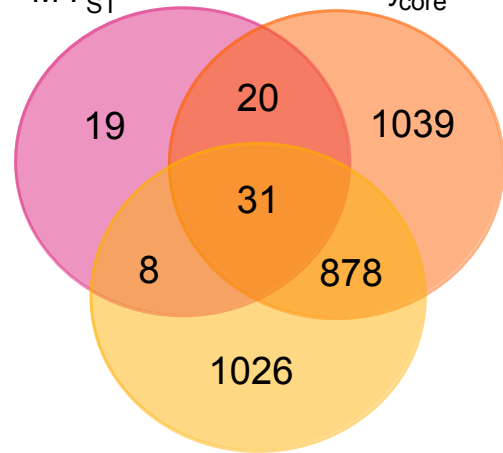
Bay_{core}XtX



D

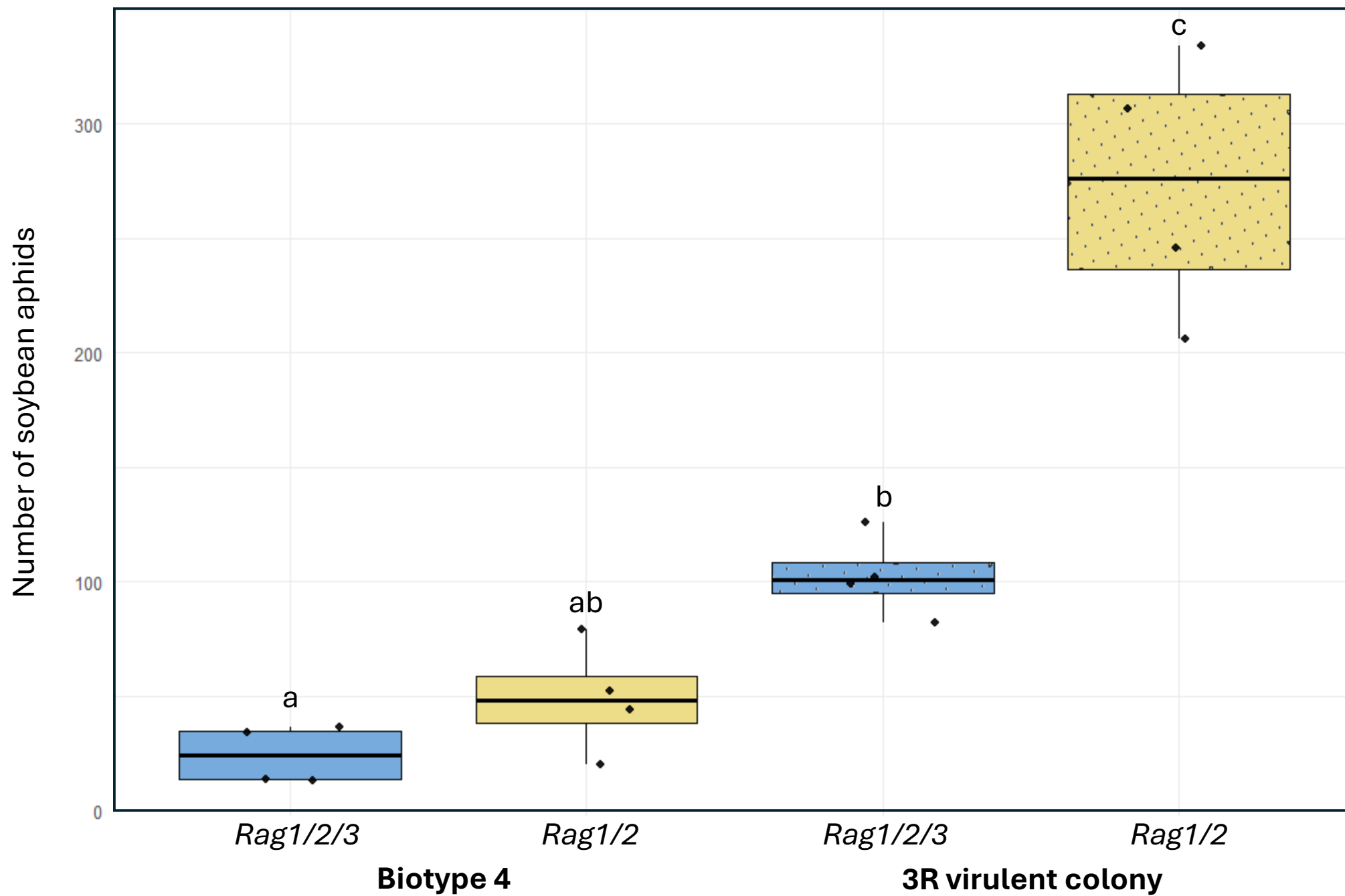
M- F_{ST}

Bay_{core}XtX

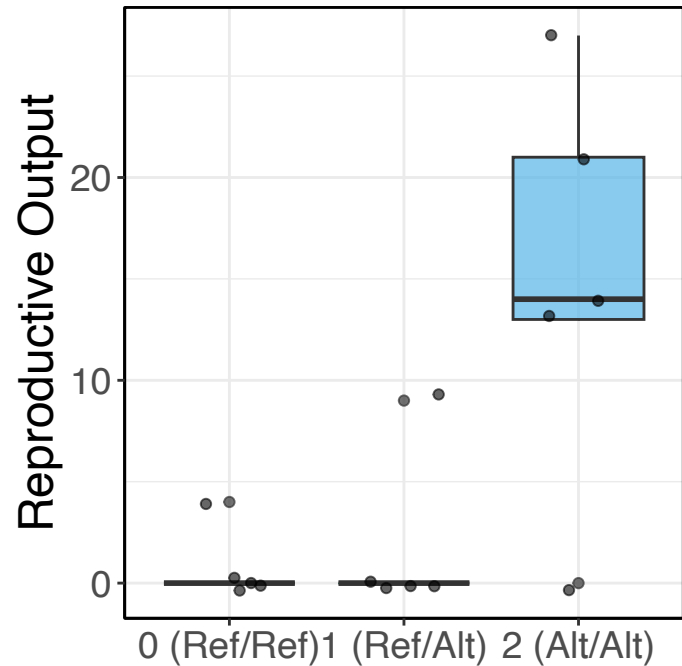


Biotype

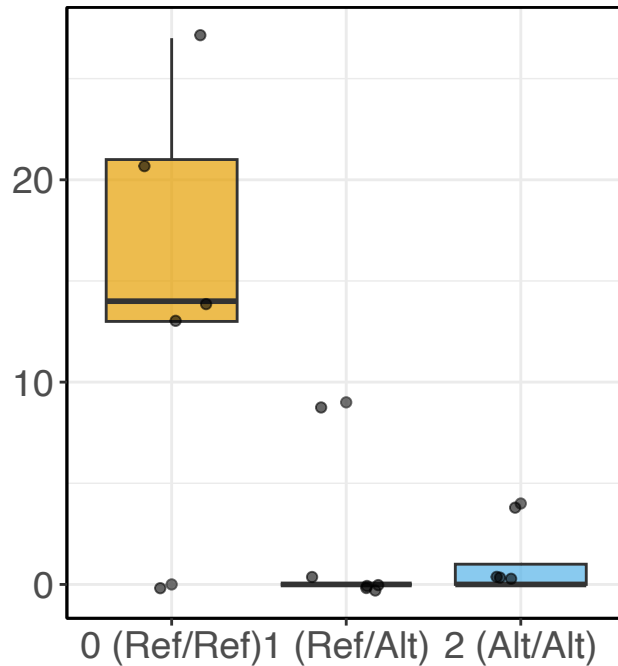
Geography



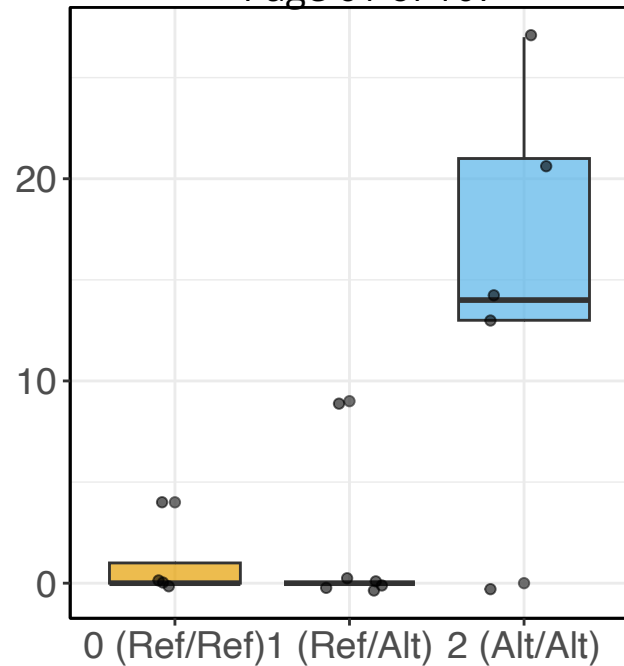
Agly004871 ($P = 0.006$, $\beta = 7.10$)



Agly001847 ($P = 0.0084$, $\beta = -7.29$)



Agly001843 ($P=0.0084$, $\beta=7.29$)



Genotype

Table 1. Differentially expressed genes associated with SNP outliers

Gene ^a	Scaffold ^a	Chromosome ^b	Location ^c	log ₂ FC: S/R ^d	Description ^e
Agly018942	25	1	Exon (S)	5.65/9.67	Uncharacterized protein, Leucine rich repeat
Agly002329	66	1	5'-UTR Exon (S)	2.39/2.52	Cytochrome P450
Agly017739	255	2	5'-adjacent	4.15/5.01	P-element induced wimpy testis (PIWI) Partner active lncRNA Agly-lnc-000004 (3' antisense)
Agly-lnc-000004	255	2	3'-adjacent	6.44/6.00	Putative TE-PIWI pathway coordinator and regulatory hub, Partner genes Agly017739 (3' antisense) and Agly017737 (overlapping sense)
Agly000978	114	3	5'-adjacent	3.68/4.01	PCAF N-terminal domain-containing protein, Histone acetyltransferase activity
Agly000981	114	3	5'-adjacent Intron	1.95/2.97	Uncharacterized protein
Agly000983	114	3	Intron	*1.11/1.30	Uncharacterized protein, Mortality factor 4-like
Agly-lnc-000191	114	3	3'-adjacent	6.44/6.00	Putative intragenic antisense lncRNA; putative transcriptional/post-transcriptional regulator; Partner gene Agly000983
Agly-lnc-000554	114	3	3'-adjacent	8.63/9.11	Putative 3' antisense regulatory lncRNA; putative transcriptional/post-transcriptional regulator; Partner gene Agly000983
Agly010975	54	X	5'-adjacent	6.11/6.49	Uncharacterized protein
Agly010068	105	X	5'-UTR	6.24/4.51	PiggyBac transposable element-derived protein
Agly002471	129	X	3'-adjacent	6.72/9.12	MI domain-containing protein, CWC22 homolog
Agly002450	129	X	Intron	9.05/8.32	Fibronectin type III domain containing protein
Agly004871	133	X	3'-adjacent	7.96/7.67	ATPase ASNA1 homolog
Agly004838	133	X	5'-UTR	†2.58/5.97	Zinc transporter ZIP1
Agly001847	237	X	Exon (NS) Intron	6.15/6.07	Lachesin-like, Immunoglobulin domain containing protein
Agly001843	237	X	5'-adjacent	8.36/5.60	Lachesin-like, Immunoglobulin domain containing protein

Agly014114	279	X	3'-adjacent	3.78/4.28	MADF domain-containing protein
Agly008028	615	X	5'-adjacent	5.85/7.00	Uncharacterized protein Partner active lncRNAs: Agly-lnc-001286 (5' sense) and Agly-lnc-001286-2 (sense overlapping)
Agly-lnc-001286 Agly-lnc-001286-2	615	X	Internal	5.82/5.77 5.83/5.90	Putative sense-oriented regulatory lncRNA; potentially Co-transcriptional activator/enhancer-like; Coordinately upregulated with target gene, Partner gene Agly008028

^aBased on Mathers 2020 genome. Names correspond to the biotype 4 genome only and do not directly translate to the biotype 1 genome. See Supplemental Table S5 for g-numbers; ^bChromosomal position of the SNP and gene based on Qiu et al., 2025; ^cposition description of SNP(s), multiple SNPs are listed in rows sequentially based on position number; Untranslated region (UTR) based on homology with other aphid species, adjacent indicates unconfirmed but likely UTR; Internal is within the predicted lncRNA sequence; ^dlog₂ fold change in virulent aphid on susceptible soybean (S) and log₂ fold change in virulent aphid on resistant soybean (R)—note: only DEGs with significant log₂ fold change ($P_{\text{adj}} < 0.001$) and greater than 2 or less than -2 in both host plant comparisons listed, see Supplemental Table S9 for full list. ^eFunctional annotations from AphidBase (bipaa.genouest.org/is/aphidbase/) accessed April 2026; In bold: constitutively expressed long non-coding RNAs (lncRNAs) found near constitutively expressed genes identified in at least three outlier tests. For the full list of constitutively expressed lncRNAs see Supplemental Table S10; DEG=differentially expressed gene, S=synonymous, NS=non-synonymous. * Agly000983: included as a candidate gene due to significant gene expression and being a partner gene of two differentially expressed lncRNAs (although not log₂ fold change < 2; >log₂ [†]Agly004838: included as a candidate gene as SNPs are more proximate and have significant genetic differentiation in field samples (see Table 2). Gene expression was significantly different for resistant plants (5.97, $P_{\text{adj}} < 0.001$) and marginally significant for susceptible (2.58, $P_{\text{adj}} < 0.053$).

Table 2. Allele frequency comparisons with avirulent and 3R virulent field collected aphids. Fishers exact test across all loci: $\chi^2=55.8136$, $df=30$, $P=0.003$; F_{st} over all loci 0.162, $P=0.022$. Genes in bold have significant F_{ST} values.

Gene ID	Scaffold	Avirulent		3R Virulent		Fisher's Exact Test P value	F_{st} , P -value
		Alternative allele ^a	Biotype 4 reference allele ^a	Alternative allele ^a	Biotype 4 reference allele ^a		
Agly018942	25	0.444	0.556	0.500	0.500	1.000	-0.078, 1.000
Agly010975	54	0.350	0.650	0.400	0.600	1.000	-0.075, 1.000
Agly002329	66	0.750	0.250	0.700	0.300	1.000	-0.074, 1.000
Agly010068	105	0.250	0.750	0.600	0.400	0.105	0.169, 0.177
Agly000978	114	0.300	0.700	0.500	0.500	0.424	0.009, 0.457
Agly000981	114	0.350	0.650	0.500	0.500	0.462	-0.031, 0.763
Agly-lnc-000191	114	0.700	0.300	0.500	0.500	0.428	0.009, 0.457
Agly-lnc-000554	114	0.700	0.300	0.500	0.500	0.426	0.009, 0.456
Agly002471	129	0.350	0.650	0.700	0.300	0.124	0.155, 0.167
Agly002450	129	0.350	0.650	0.700	0.300	0.120	0.155, 0.163
Agly004871	133	0.300	0.700	0.900	0.100	0.006	0.477, 0.014
Agly001847	237	0.650	0.350	0.100	0.900	0.008	0.412, 0.012
Agly001843	237	0.350	0.650	0.900	0.100	0.006	0.412, 0.021
Agly014114	279	0.300	0.700	0.600	0.400	0.140	0.104, 0.288
Agly008028	615	0.300	0.700	0.500	0.500	0.402	-0.004, 0.255

^aalleles predicted based on biotype 4 virulent soybean aphid genome version 2.1 (Mathers 2020)

**Adaptation to Plant Defense in an Agricultural Insect Pest:
Integrating Genome Scans and Gene Expression in the Soybean Aphid Reveals Multi-
genic Pathways**

AUTHORS:

Vitor A. C. Pavinato^{1,6}, Erin H. Penton-Michel¹, Saima Shahid^{2,3,7}, Jacob A. Wenger^{1,8}, R. Keith Slotkin^{2,3,4}, Andy Michel^{1,5*}

¹Dept. of Entomology, College of Food, Agricultural and Environmental Sciences, The Ohio State University, Wooster, OH, USA

²Dept. of Molecular Genetics, College of Arts & Science, The Ohio State University, Columbus, OH USA

³Donald Danforth Plant Science Center, St. Louis, MO, USA & Division of Biological Sciences, University of Missouri-Columbia, MO, USA

⁴Division of Biological Sciences, University of Missouri-Columbia, MO, USA

⁵The Center for Applied Plant Sciences, The Ohio State University, Columbus, OH USA

⁶Current Address: Department of Biology, Temple University, Philadelphia, PA, USA

⁷Current Address: University of Sheffield, UK

⁸Current Address: Dept. of Plant Science, California State University, Fresno, CA, USA

*corresponding author: michel.70@osu.edu

keywords: insect biotype, virulence, rapid adaptation, genome scans, SNPs, soybean aphid

Supplementary Material, Method 1: Pipeline for read processing and variant discovery

Section 1: Pool-seq variant calling

The pipeline we developed to process soybean aphid pool-seq data

(<https://github.com/vitorpavinato/AglyPoolseq>) first generates an alignment file (BAM) for each raw sequencing data (provided as a fastq file) (see main text for the description of how we processed raw reads and mapped then to the soybean aphid virulent reference genome).

The pipeline generates a MPILEUP-formatted file from each sequencing alignment file (BAM) with samtools mpileup (this file type contains variants discovered in an BAM file, and we called it a raw variant file). Variant filtering was proceeded with samtools

mpileup function with the parameter “-B” to suppress the recalculation of the per-base alignment qualities and the parameter “-Q 25” to filter out bases with low mapping quality.

Another filtering was applied during the conversion of MPILEUP files to gSYNC files.

In each gSYNC file, the pipeline masked bases that: 1) had minimum coverage of x ($\text{min_cov} < x$, x being 3 for the sequencing technical replicates and 4 for the pools' aggregated files, see below we created the aggregated files from the technical replicates files), and 2) had maximum coverage higher than 99% of the sequencing depth percentile given the scaffold/file ($\text{max_cov} > 99\%$). The pipeline also masked SNPs located within repetitive regions as identified by RepeatMasker (Smit et al. 2013) (see main text) and SNPs located 5bp upstream and downstream of an indel identified by GATK IndelRealigner (McKenna et al., 2010)(see main text), in each file. Masked SNPs were then imputed with "." within each file, representing missing data for the sample allele count.

Section 2: Sequencing quality assessment

We assessed the quality of the whole-genome pool sequencing by checking the quality and consistency among the technical replicates' files. We first performed a multi-sample variant calling which resulted in a VCF-type file by combining all masked gSYNC files from all

technical replicates into one, 2-dimensional matrix; rows contained SNPs and columns included their respective variant counts. SNPs were called using this matrix by PoolSNP, resulting in a VCF file by setting a global Minor Allele Count (MAC) of 50 and a global Minor Allele Frequency (MAF) of 0.05. SNPs in this file were retained if they contained less than 50% missing data across replicates, minimum and maximum coverage of 3 reads (see above), and 99% quantile of the read coverage distribution.

All SNPs in the technical replicates allele count matrix were used to assess the sequencing quality. We visually checked for errors by conducting a principal component analysis (PCA) on the matrix of reference allele frequency obtained with the ML approach (described in the main text). We scaled and centered the allele frequency matrix, imputed missing data with zeros, and performed the single value decomposition on the transposed allele frequency matrix, where rows represent technical replicates and columns represented SNPs. The first two PCAs were added to a scatterplot, and any presence of abnormal clustering was evaluated (Figure S1). We also assessed the correlation between the amount of missing data and the mean technical replicates' sequencing coverage (Figure S2), the average percentage of the scaffold that was sequenced, the mean site depth, the mean base quality from reading alignments, and the mean base mapping quality (Figure S3). These data were obtained from the alignment files in BAM format with samtools coverage with correlations performed in R (Team R). As expected, the percentage of missing data reduces with the sequencing coverage. Other than that, no abnormal sequencing technical replicates were found, giving us the confidence to aggregate the replicated alignment files of each sampled pool.

Section 3: Assessing MAC and MAF parameters for SNP filtering

We assessed how different combinations of MAC and MAF might impact the quality of SNP filtering in our Pool-seq experiment. First, we aggregated the technical replicates alignment

files (BAM) for each sampled pool. Then we determined the appropriate combination of MAC and MAF that could produce the highest number of SNPs with the lowest genome wide ratio of the proportion of nonsynonymous (p_N) to synonymous (p_S) mutations p_N/p_S .

In neutral-equilibrium conditions, we expect similar proportions of synonymous and non-synonymous mutations (Kapun et al., 2021) with some regions with more synonymous variants than non-synonymous variants because of the action of purifying selection (Messer & Petrov, 2013). A higher number of nonsynonymous mutations relative to synonymous mutations might suggest errors in the sequencing or the downstream processing of sequencing reads. Finding MAC and MAF values that give a neutrality level of p_N/p_S would remove sequencing errors that might inflate the identification of outliers by genome scans. This would allow the genome-wide estimates to be close to the neutral expectation but with some genes differing non-randomly in the proportion of synonymous and non-synonymous mutations due to locus-specific factors such as purifying or positive selection.

To generate VCFs files for each MAF and MAC combination we combined all masked gSYNC files from the aggregated pools into one, 2-dimensional matrix as described above. We sampled 15% of the total genomic regions (~ 1Mbp total) from this matrix and produced VCF files for each combination of MAC at 5, 10, 15, 20, 50, 100, and MAF ranging from 0 to 0.2 by parsing different values of MAC and MAF to the PoolSNP framework. To calculate p_N/p_S , we first annotated the VCF generated from each combination of the MAC and MAF thresholds with SNPeff version 4.3 (Cingolani et al., 2012), and counted the number of synonymous and nonsynonymous SNPs in each scaffold/pool simply by defining each missense SNP as non-synonymous, and as synonymous variant synonymous SNP.

The best combination of MAC and MAF was 5 and 0.05 (Figure S4). The VCF used for all subsequent analyses was obtained with these global filters. This VCF contained SNPs with less than 50% missing data across pools, minimum and maximum coverage of 4 reads, and 99% quantile of the read coverage distribution. This VCF was then annotated one more time with SNPeff. [Click or tap here to enter text.](#)

For the analyses of the aggregated data, most of the pools showed the negative correlation between of sequencing coverage and percentage of missing data (see above), but some had more missing data than expected. Together with the p_N/p_S and MAC-MAF relationship, we decided to keep all pools (including those with more than 30% of missing data, see Figure S4) as with all pools allowed us to cover most of the soybean aphid geographical distribution.

Supplementary Material, Method 2, Genome Scans

Section 4: Genome Scan Methods for Searching for virulence genes

With population genomics scans, we expected to identify SNPs in candidate genes that might be associated with virulence in the soybean aphid. We employed five genome scan methods: 1) Poolfstat multi-locus F_{ST} ($M-F_{ST}$), 2) SelEstim (SEL) (Vitalis et al., 2014), 3) BAYPASS core model based on the XtX statistic (BAY_{core} XtX) (Coop et al., 2010; Günther & Coop, 2013), 4) BAYPASS auxiliary covariate model (BAY_{aux} BFmc), and 5) BAYPASS XtX controlled by the covariate model (BAY_{aux} XtX) (Gautier, 2015). The first three tests are covariate-free and use the among-pools allele frequency variance of each SNP to identify candidate SNPs outliers. Higher SNPs allele frequency variance than the baseline expected might signal the presence of selection acting on the SNPs but more importantly on nearby genes because of linked selection. The other two methods implemented in BAYPASS auxiliary models used the high covariance between allele frequencies the biotypes to find evidence of association between the marker and the phenotype. BAYPASS auxiliary models

are implementation of genome-association methods for discovery of variants that co-vary with phenotypes.

For the multi-locus F_{ST} , we used a sliding window approach implemented in the R package (Poolfstat) (Hivert et al., 2018). The function *computFST()* first calculated all SNP's single-locus F_{ST} ($S-F_{ST}$ when possible due to missing populations), then averaged all these values within windows. Window size is not fixed by the number of base pairs but by the number of SNPs (we set each window to have atleast ten SNPs, and the average window size in our analysis was 9.9 Kbp). We determined outlier's multi-locus SNPs with a $M-F_{ST} > 0.25$, which is 32% higher than the 99th % percentile of the empirical $M-F_{ST}$ distribution (99% $M-F_{ST}$ quantile = 0.19). Because the $M-F_{ST}$ gives genome coordinates for the window, and we wanted to compare the outliers identified with this method to other methods that are SNP-based (Figure S8A), we transformed the coordinates of each significant $M-F_{ST}$ back to the SNPs coordinates. We did it by taking the coordinates of SNPs that had the $S-F_{ST}$ higher than the 95% quantile of the scaffold distribution of $S-F_{ST}$ s. This approach worked as a two-layers filter: first, significant scaffolds were identified with the windowed version of the F_{ST} ($M-F_{ST}$) (Figure S8B); then candidate SNPs and genes were then identified within those scaffolds (Figure S9).

SelEstim (SEL) is a hierarchical Bayesian method based on the diffusion approximation of allele frequencies in a subdivided population. The method in SEL explicitly accounts for selection by assuming that every locus is targeted by selection to some extent. This allows the estimation of the selection strength that might have determined the allele frequencies of each locus and each locus/subpopulation and accounts for the effects of gene flow among populations. We performed three independent runs of SelEstim using the same parameters but with different seeds to check chain convergence and to account for stochastic differences

among runs. For each run, we defined twenty pilot runs of 500 iterations (used to adjust the proposal distribution for each model parameter), 120,000 burn-in iterations, followed by 50,000 iterations that updated the model parameters every 50 steps (thinning interval). We used the coda package (Plummer et al., 2006) to check the MCMC chain's convergence. We selected candidate markers above the 95% quantile of KLD from the distribution of estimated values from each independent run. We retained only candidate SNP outliers jointly identified in the three independent runs (Figure S10).

Unlike the other two methods, BAY_{core} first estimates the covariance structure among the pool's allele frequency (that works as a kinship matrix) and uses it to account for the shared evolutionary history of the sampled pools. It accounts for population genetic structure and shared demography, reducing false positives due to factors other than genotype x phenotype association. Baypass employs mainly two approaches: 1) identify markers with higher variance among the populations (the core model of Günther and Coop, 2013; Gautier et al. 2015) ($BAY_{core} XtX$ and $BAY_{aux} XtX$). 2) identify outliers associated with covariable(s) (the covariable model Gautier et al., 2015) ($BAY_{aux} BFmc$); BAY_{aux} method, differed the previous outliers tests mentioned before as it performs an association analysis between each SNP allele frequency and a covariate; in our analysis, we only use the biotype of origin as a covariate. $BAY_{aux} XtX$ is similar to $BAY_{core} XtX$, but it corrects the XtX statistics for both the genetic structure (shared history of the populations) and the biotype effect (covariate effect).

We ran the core model ($BAY_{core} XtX$) with 20 pilot runs (used to adjust the parameters for the MCMC), each with 1,000 iterations. After the pilot runs, we run an additional 50,000 burn-in iterations followed by 25,000 final iterations sampled every 25. Because we had not well-behaved p-values, we calibrated the core's model XtX statistics with pseudo-observed data. We simulated 100,000 SNPs using the Baypass function *simulate.baypass()* with the

matrix estimated with the data. We ran the analysis of the PODs data with the same number of pilot runs, burn-in iterations, and final iterations lengths for analyzing the observed data. Since the POD analysis was performed by re-estimating the matrix, we check its convergence with the observed data by comparing the two matrices. We defined outliers for the BAY_{core} SNPs with XtX statistics higher than the 99 % quantile of the PODs XtX distribution (Figure S11A). We ran Baypass auxiliary covariate model similarly as before, but this time we used: 1) the matrix estimated with the first analysis and 2) a covariate file, which defines the value for the covariate for each pool, being zero if it was an avirulent pool, one if it was a virulent pool. Outliers were identified using the similar calibration employed on the core model for both the XtX corrected model (BAY_{aux} XtX) (Figure S11B) and the covariate (BAY_{aux} BFmc) (Figure S11C).

Supplementary Material, Tables

- *Supplementary Material, Table S1. Sampled pools collection information and associated meta-data.*
- *Supplementary Material, Table S2. PCR primers for genotyping candidate SNPs.*
- *Supplementary Material, Table S3. Sequencing summary statistics for the technical replicated data.*
- *Supplementary Material, Table S4. Sequencing summary statistics for the aggregated data.*
- *Supplementary Material, Table S5. Candidate genes with constitutive expression identified in at least three outlier tests.*
- *Supplementary Material, Table S6. Long non-coding RNAs (lncRNAs) constitutively expressed.*

- *Supplementary Material, Table S7. Field-collected samples and laboratory colonies genotypes for candidate SNPs.*

Supplementary Material, Figures

- *Supplementary Material, Figure S1. Quality Assessment 1: technical replicate samples - PCA.*
- *Supplementary Material, Figure S2. Quality Assessment 2: technical replicate samples - % of missing data and mean read coverage.*
- *Supplementary Material, Figure S3. Quality Assessment 3: technical replicate samples - % missing data and SNPs quality.*
- *Supplementary Material, Figure S4. Quality Assessment 4: pN/pS as a function of MAC and MAF for aggregated SNP dataset.*
- *Supplementary Material, Figure S5. Overall Genetic diversity.*
- *Supplementary Material, Figure S6. Genetic diversity and proportion of synonymous and non-synonymous mutations.*
- *Supplementary Material, Figure S7. Heatmap of pairwise F_{ST} 's.*
- *Supplementary Material, Figure S8. Manhattan plots of the Poolfstat F_{ST} genome scan test*
- *Supplementary Material, Figure S9. Single-locus- F_{ST} and genetic diversity for scaffolds that harbor significant multi-locus $-F_{ST}$ loci.*
- *Supplementary Material, Figure S10. Manhattan plots for SelEstim runs.*
- *Supplementary Material, Figure S11. Manhattan plots of BAYPASS genome-scans tests.*
- *Supplementary Material, Figure S12. Distribution of the number of outliers per scaffold.*
- *Supplementary Material, Figure S13. Plots of the GO enrichment analysis of significantly DEG with constitutive up and down-regulation.*

- *Supplementary Material, Figure S14. IGV view of the entire scaffold_255.*

References

- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Coop, G., Witonsky, D., Rienzo, A. D., & Pritchard, J. K. (2010). Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, 185(4), 1411–1423. <https://doi.org/10.1534/genetics.110.114819>
- Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* 201(4), 1555–1579. <https://doi.org/10.1534/genetics.115.181453>
- Günther, T., & Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, 195(1), 205–220. <https://doi.org/10.1534/genetics.113.152462>
- Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring Genetic Differentiation from Pool-seq Data. *Genetics*, 210(1), 315–330. <https://doi.org/10.1534/genetics.118.300900>
- Kapun, M., Nunez, J. C. B., Bogaerts-Márquez, M., Murga-Moreno, J., Paris, M., Outten, J., Coronado-Zamora, M., Tern, C., Rota-Stabelli, O., Guerreiro, M. P. G., Casillas, S., Orengo, D. J., Puerma, E., Kankare, M., Ometto, L., Loeschcke, V., Onder, B. S., Abbott, J. K., Schaeffer, S. W., ... Bergland, A. O. (2021). Drosophila Evolution over Space and Time (DEST): A New Population Genomics Resource. *Molecular Biology and Evolution*, 38(12), 5782–5805, <https://doi.org/10.1093/molbev/msab259>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Messer, P. W., & Petrov, D. A. (2013). Frequent adaptation and the McDonald–Kreitman test. *Proceedings of the National Academy of Sciences*, 110(21), 8615–8620. <https://doi.org/10.1073/pnas.1220835110>
- Plummer M, Best N, Cowles K, Vines K (2006). “CODA: Convergence Diagnosis and Output Analysis for MCMC.” *R News*, 6(1), 7–11.
- Smit, AFA, Hubley, R & Green, P. *RepeatMasker Open-4.0*. 2013-2015
<<http://www.repeatmasker.org>>.
- Team R. (2025) R: A Language and Environment for Statistical Computing.
- Vitalis, R., Gautier, M., Dawson, K. J., & Beaumont, M. A. (2014). Detecting and measuring selection from gene frequency data. *Genetics*, 196(3), 799–817. <https://doi.org/10.1534/genetics.113.152991>

Supplementary Figures

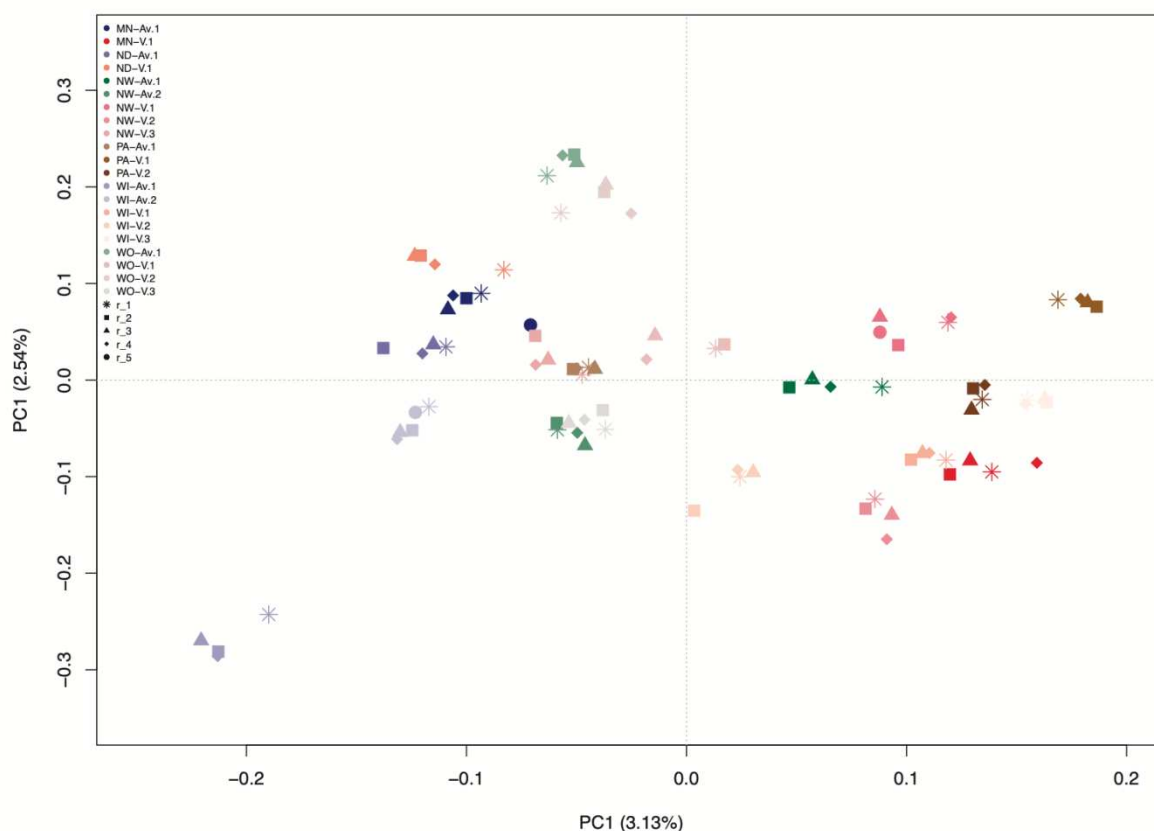


Figure S1. *Quality Assessment 1: consistency among technical replicated sequencing runs.* We ran a principal component analysis (PCA) on the maximum-likelihood imputed reference allele frequency to identify problematic sequencing replicates. All sequencing libraries were sequenced together four times (except for MN-Av.1, NW-V.1, and WI-Av.2 pools that were sequenced for an additional run) on different days to assess sequencing consistency and to increase sequencing coverage per each sampled pool. All libraries were consistent as different runs from the same library fell in the same cluster.

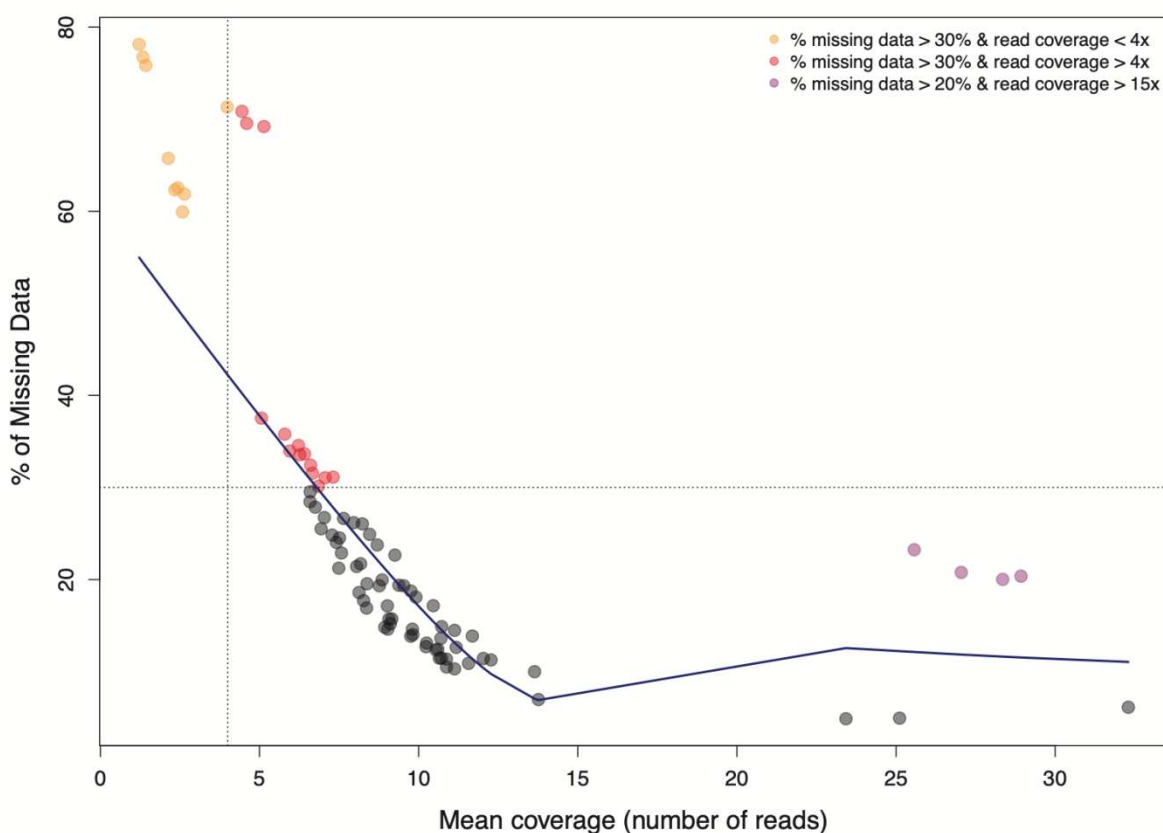


Figure S2. *Quality Assessment 2: percentage of missing data within a technical replicate as a function of the read coverage.* The mean average represents the mean across all positions of the total reads obtained for each technical replicate's run. The amount of missing data is expected to decrease with the read coverage. Some replicates (in orange) had a high amount of missing data, as expected by their low average read coverage. Still, other replicates of a pool (in red and purple), even with higher coverage, had a higher amount of missing data. All pools were kept for population genomics analysis.

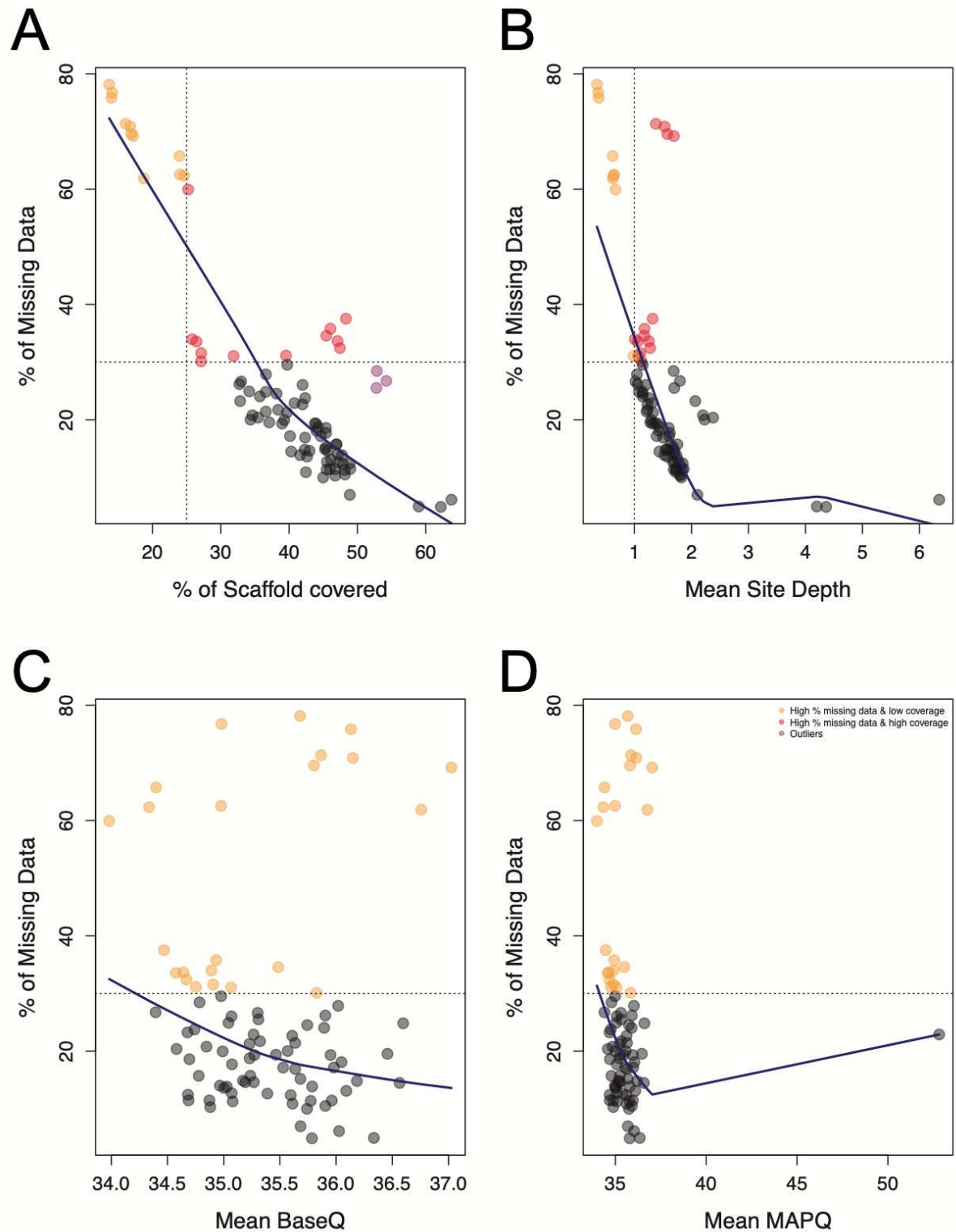


Figure S3. *Quality Assessment 3: the amount of missing data and its relationship with mapping statistics.* A) Average percentage of scaffold coverage (from the BAM files), B) Mean site depth, C) Mean base quality (BaseQ), and D) mean mapping quality (MAPQ). The

percentage of missing data is only correlated with the percentage of scaffold coverage and the mean site depth. All pools were kept for population genomics analysis.

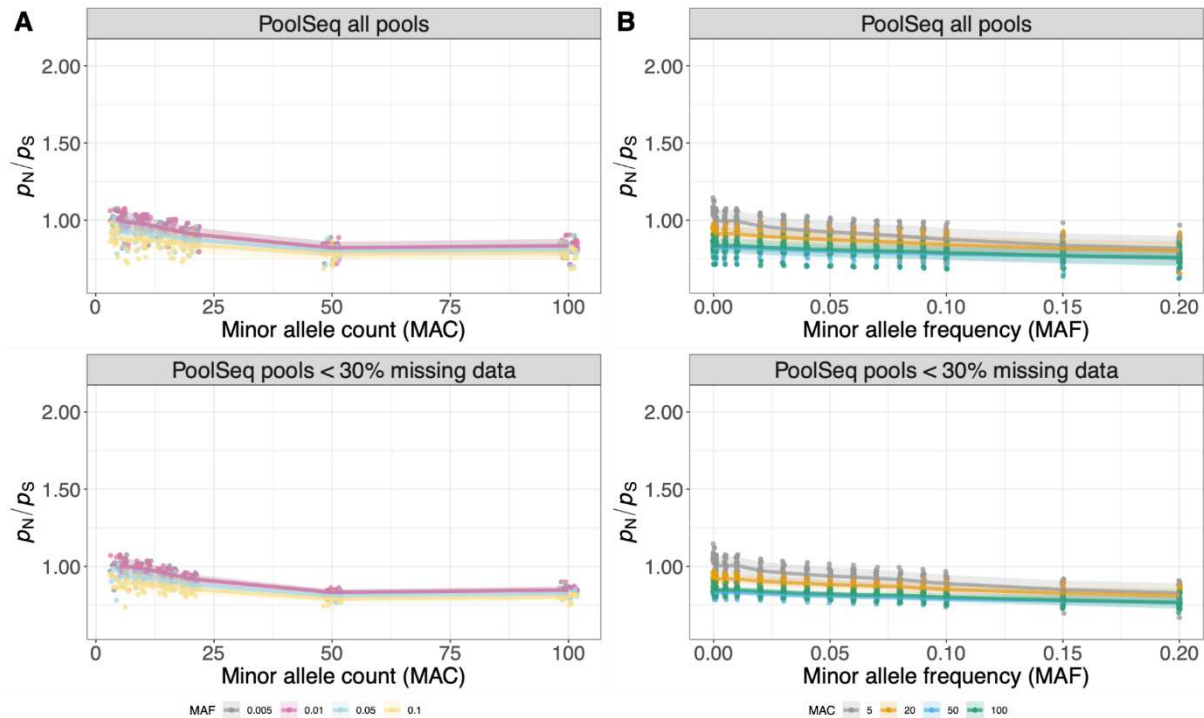


Figure S4. *Quality Assessment 4: quality control of SNPs called with Pool-SNP using different combinations of the heuristic thresholds.* Panels A and B show the effect of the heuristic thresholds minor allele count (MAC) and minor allele frequency (MAF) on the proportion of synonymous to nonsynonymous mutations (p_N/p_S) ratios of SNPs called with the Pool-SNP method. Each MAC plot (panel A) shows points corresponding to genome-wide values of the p_N/p_S ratio of each sampled pool in each value of the MAC (x-axis). Different MAF values are depicted in different colors; the strong line is the average value, and the lighter ribbon is the standard deviation. The top plot of panel A shows the relationship of p_N/p_S when MAC increases when all pools are included; the bottom plot shows the same relationship but without pools with more than 30% of missing genotypes. Panel B contains the same information but shows the relationship between p_N/p_S and MAF, with color representing different MAC values. MAC thresholds affect the p_N/p_S ratios until $MAC = 50$, but MAF values have little influence on p_N/p_S ratios regardless of the MAC value. We decided to keep all pools, because removing pools with more than 30% of missing data did not change the behavior of the p_N/p_S .

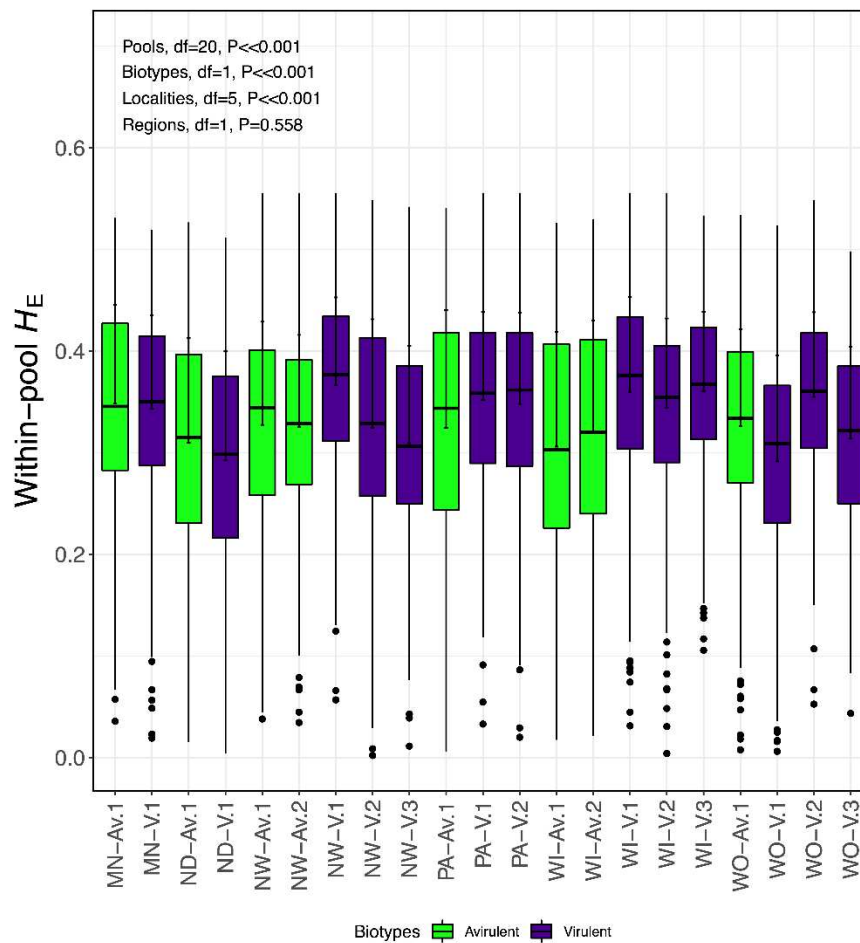


Figure S5. Population genetic diversity of *Aphis glycines* across a West-East transect of the species North America's' distribution. Within-pool genetic diversity as the expected heterozygosity shows significant differences among pools, biotypes, localities, and regions on the levels of heterozygosity.

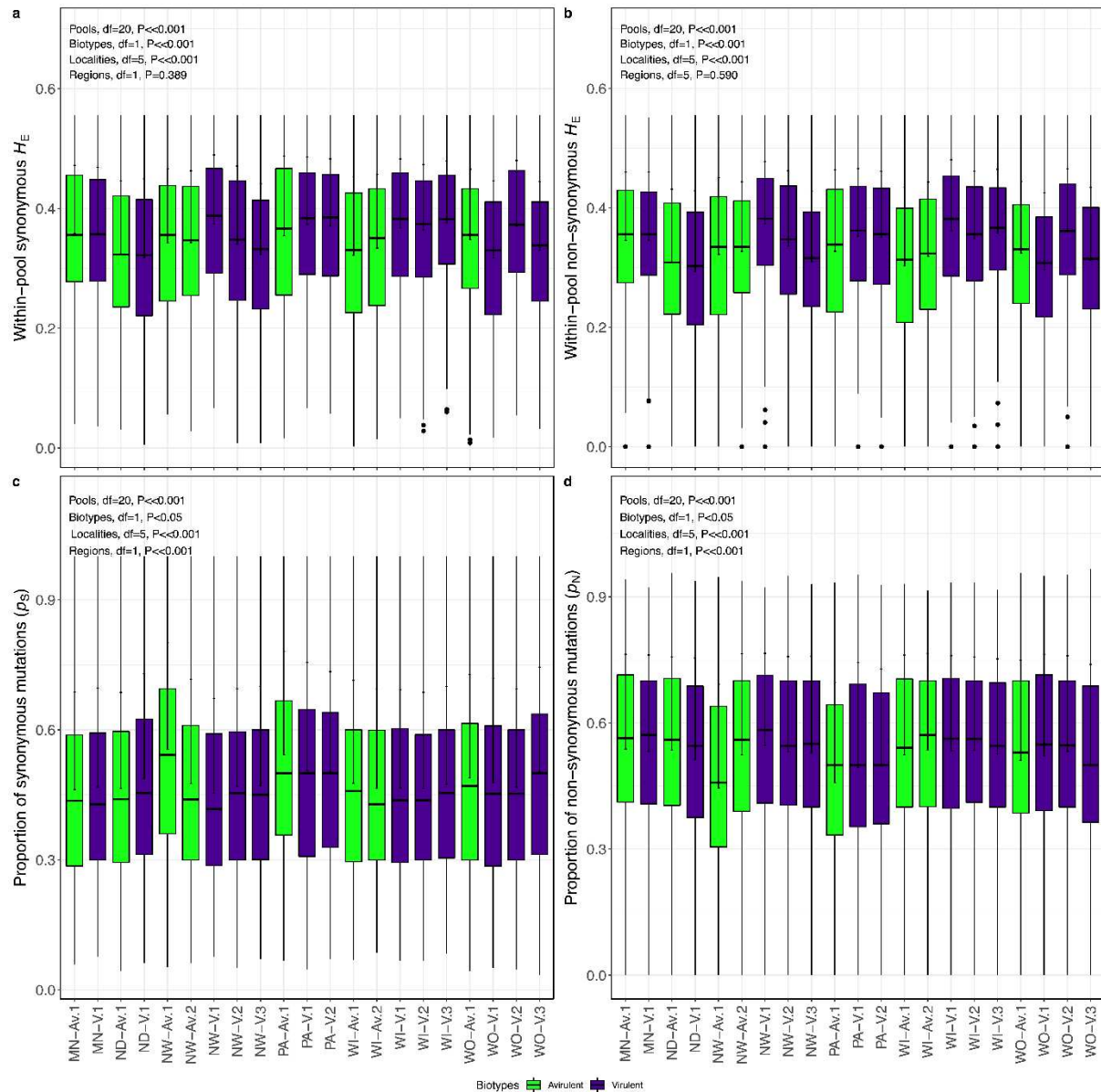


Figure S6. Genetic diversity and proportion of synonymous and non-synonymous mutations. a) within-pool synonymous H_E , and b) within-pool non-synonymous H_E , c) Proportion of synonymous mutations (p_S), d) proportion of non-synonymous mutations (p_N). The proportion of synonymous and non-synonymous mutations and their respective diversity significantly differed among pools, biotypes, localities, and regions (west vs. east). The proportion of non-synonymous was higher than the synonymous mutation overall, but the levels of diversity for these two types of mutations were, on average, similar.

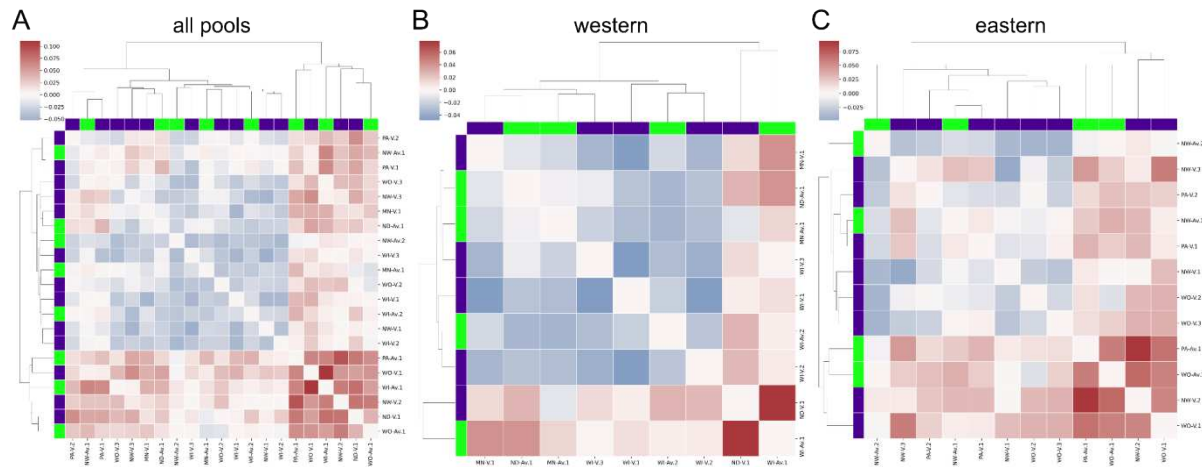


Figure S7. Heatmap of pairwise F_{ST} 's. Biotypes are shown with different colors in the bar below or in front the dendrogram. A) all 21 pools, B) west (WI, MN, and ND), and C) east NW, WO, and PA) pools.

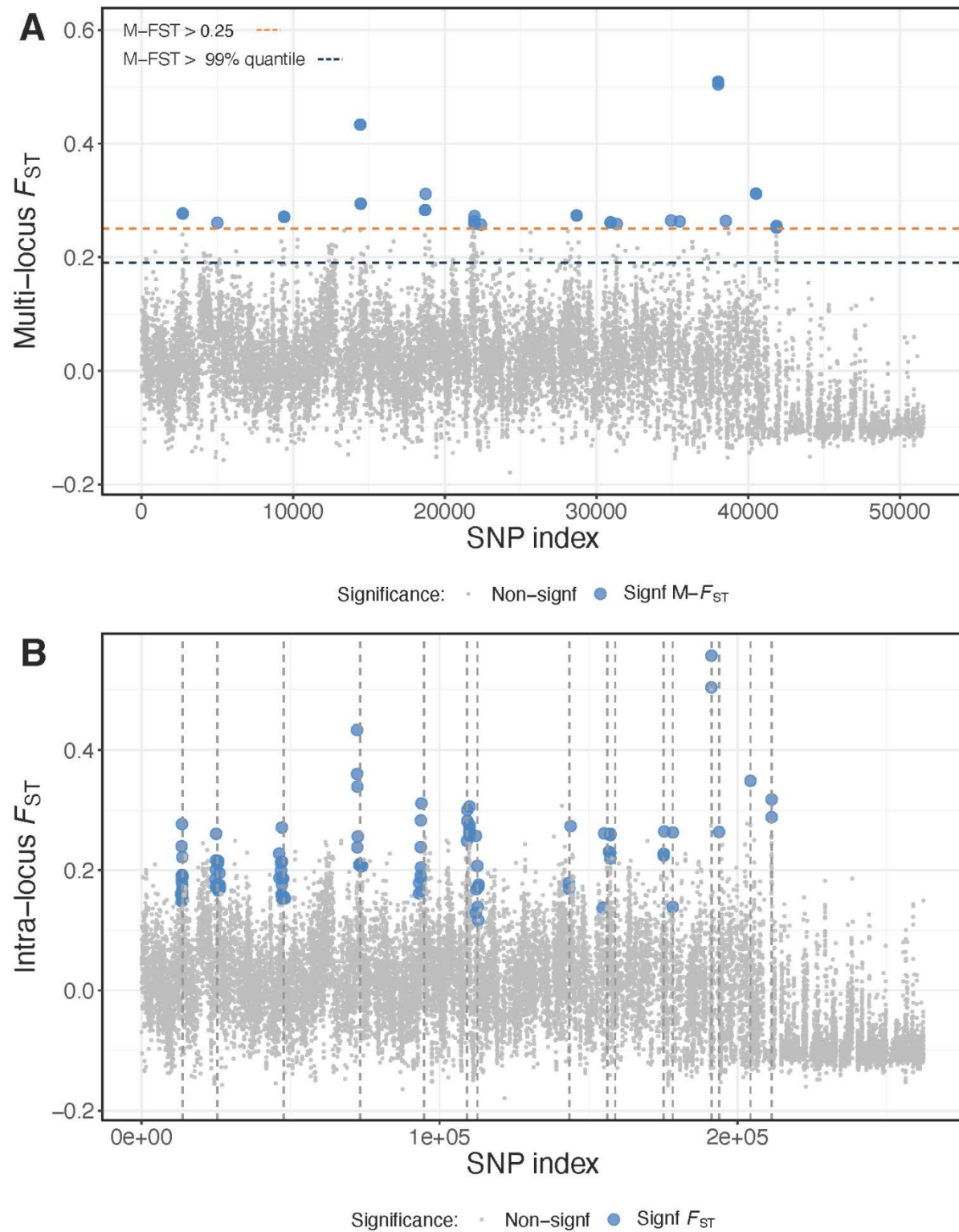


Figure S8. Manhattan plots of the Poolfstat F_{ST} genome scan tests A) Shows the multi-locus F_{ST} values calculated in non- fixed windows of at least 10 SNPs. SNPs associated with candidate genes from Table 1 are indicated by the text box. B) Shows the single-locus F_{ST} values. $M-F_{ST}$

values were converted to $S-F_{ST}$ values by taking the single-locus F_{ST} values higher than the 95% quantile of intra-scaffold distribution of F_{ST} values. SNPs highlighted in Figure A and B include those SNPs higher than each test threshold ($M-F_{ST} > 0.25$; $S-F_{ST}$ higher than the 95% quantile of the intra-scaffold distribution of F_{ST} s).

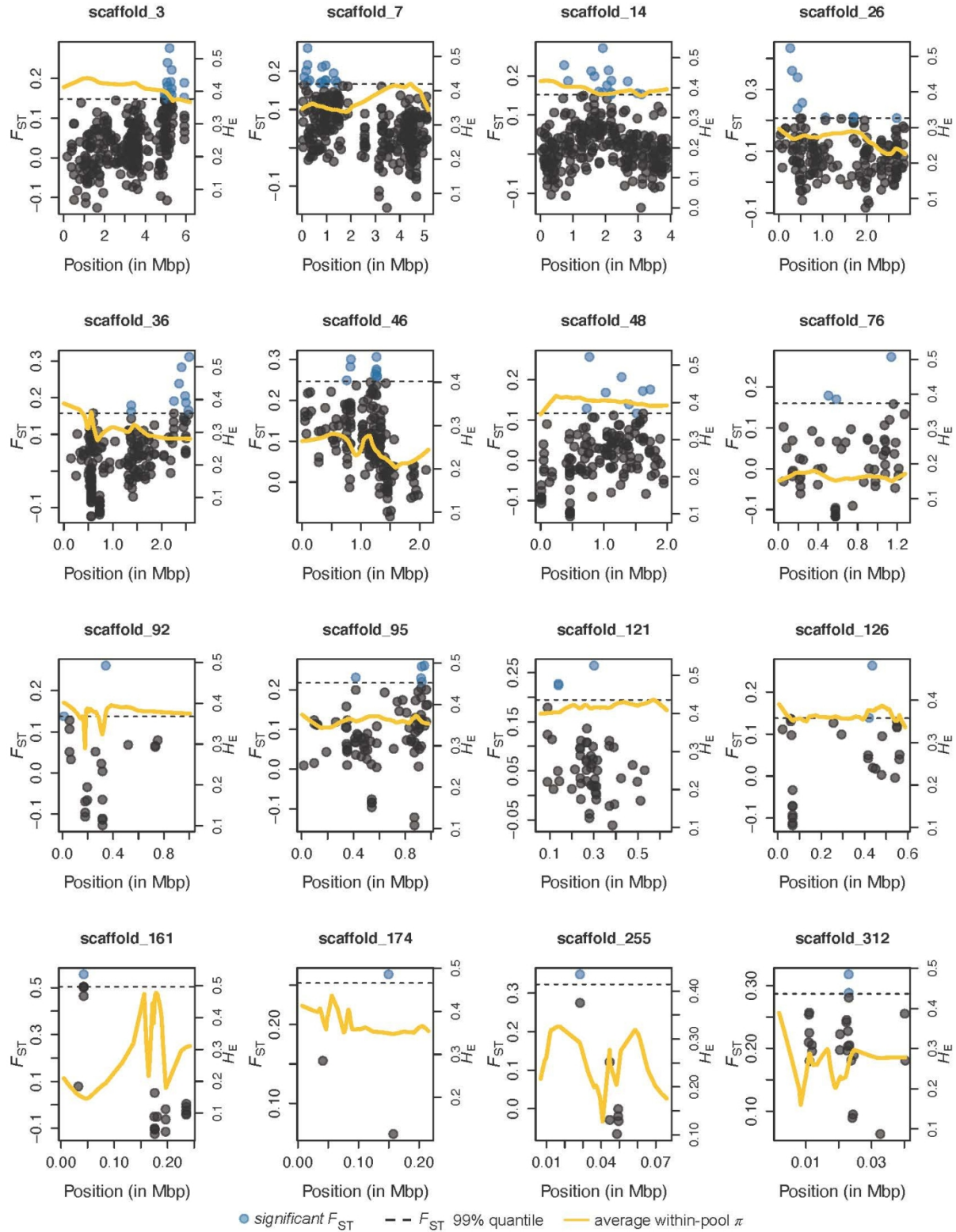


Figure S9. Single-locus- F_{ST} and genetic diversity for significant scaffolds identified with multi-locus F_{ST} scan. Each black circle corresponds to S- F_{ST} estimates and the yellow line corresponds to a smoothed line calculated from single-locus values of the average within-pool π

genetic diversity. Dashed line corresponds to the within-scaffold F_{ST} 95% quantile and blue circles correspond to the significant S- F_{ST} .

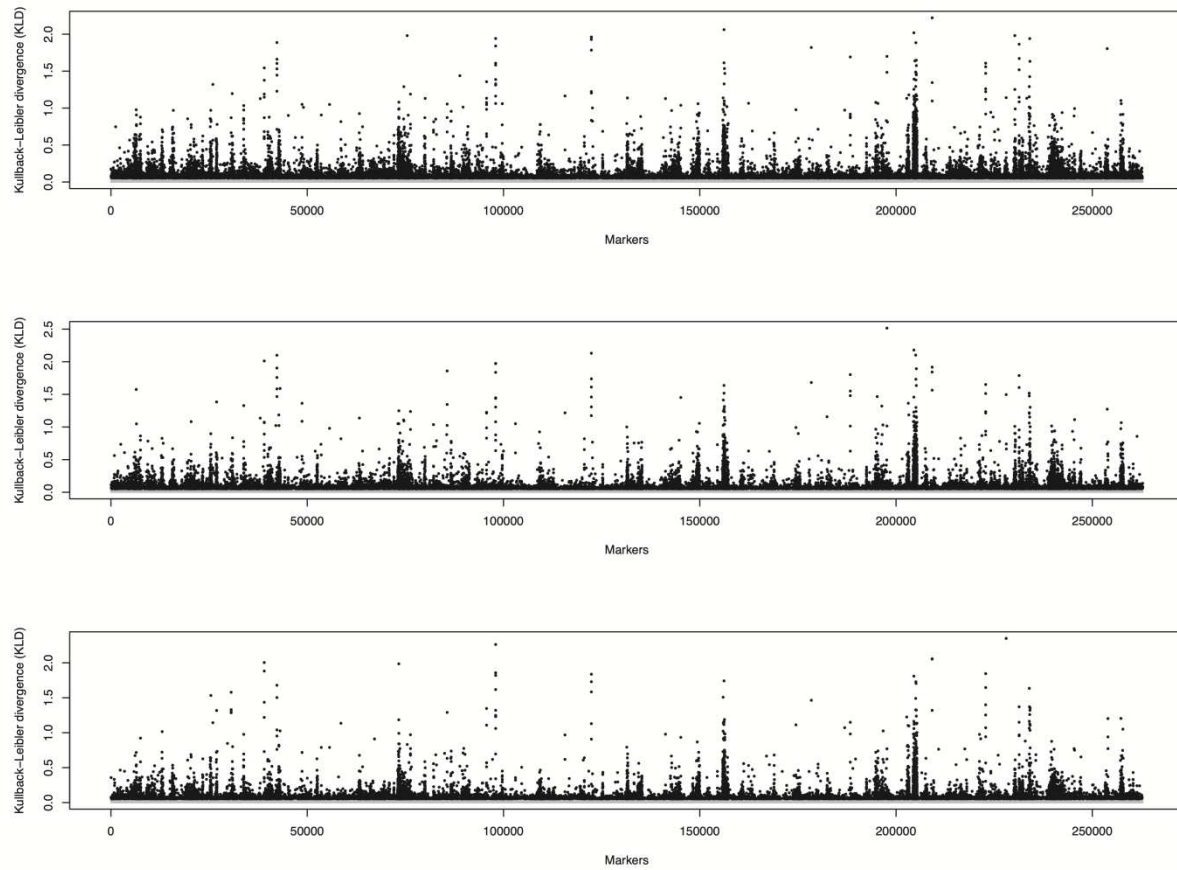


Figure S10. Manhattan plots for SelEstim runs. A total of 5,543 SNPs was identified in common in the three independent SelEstim runs. Each run Manhattan plots had estimated KLD forming peaks approximately on the same genomic regions.

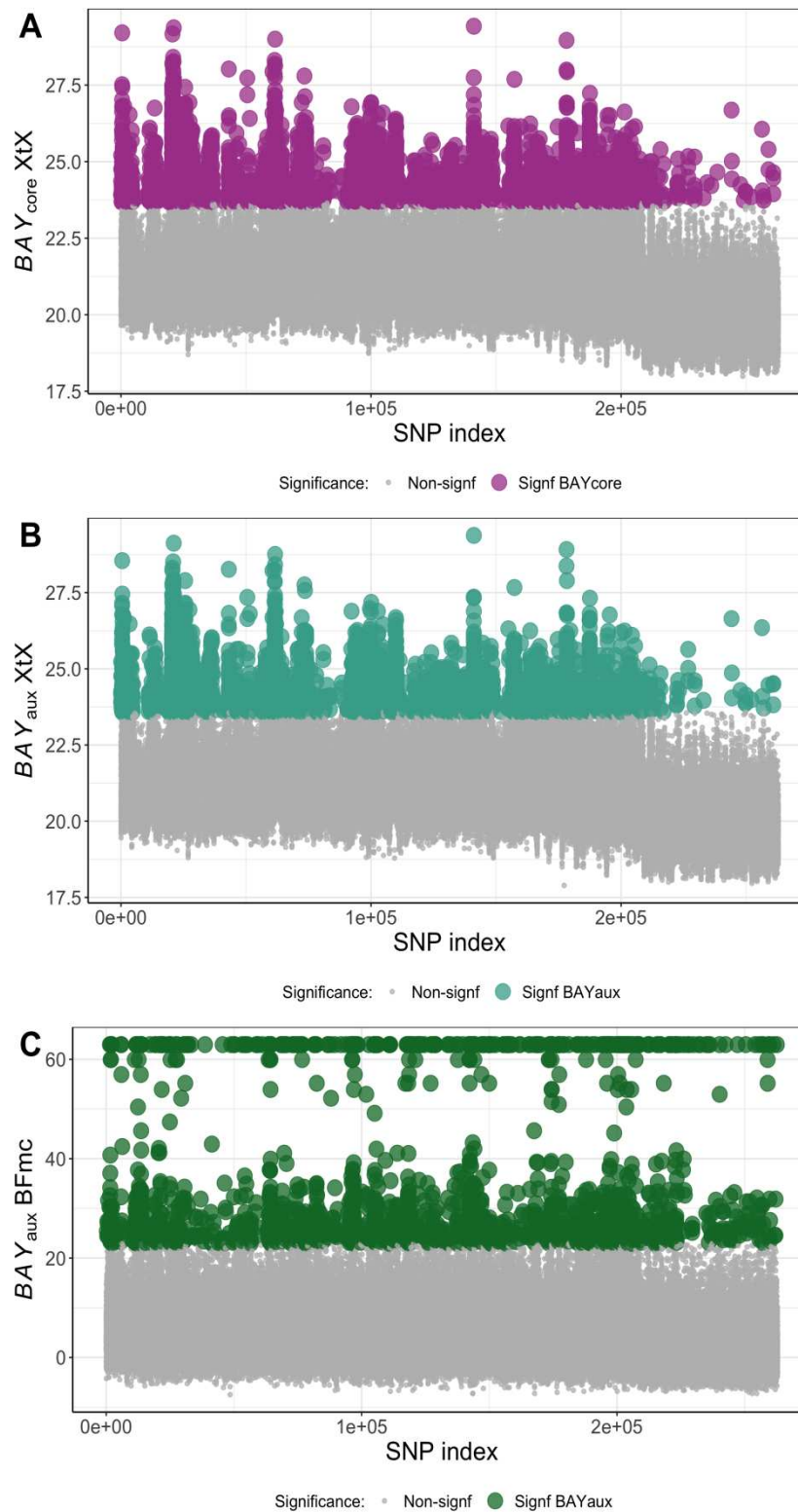


Figure S11. Manhattan plots for BAYPASS genome-scans tests. Significant SNPs were highlighted in each scatterplot. A) Highlighted SNPs corresponds to those with XtX statistics higher than the 99 % quantile of the PODs XtX distribution. B) Highlighted SNPs corresponds to those with XtX statistics higher than the 99 % quantile of the PODs XtX_{aux} distribution, and in

C) highlighted SNPs corresponds to those with BFmc statistics higher than the 99 % quantile of the PODs BFmc distribution.

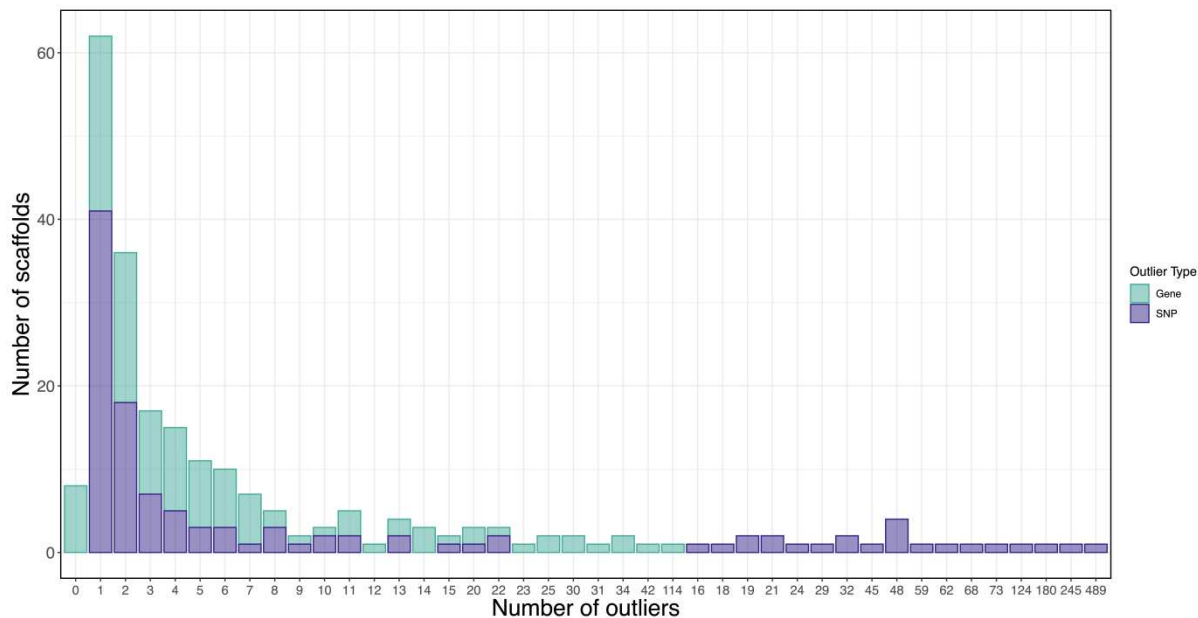


Figure S12. Distribution of the number of outliers SNP and genes per scaffold. The majority of scaffolds had fewer SNPs and genes outliers. The maximum number of outliers in one scaffold were 469 SNPs and 114 genes.

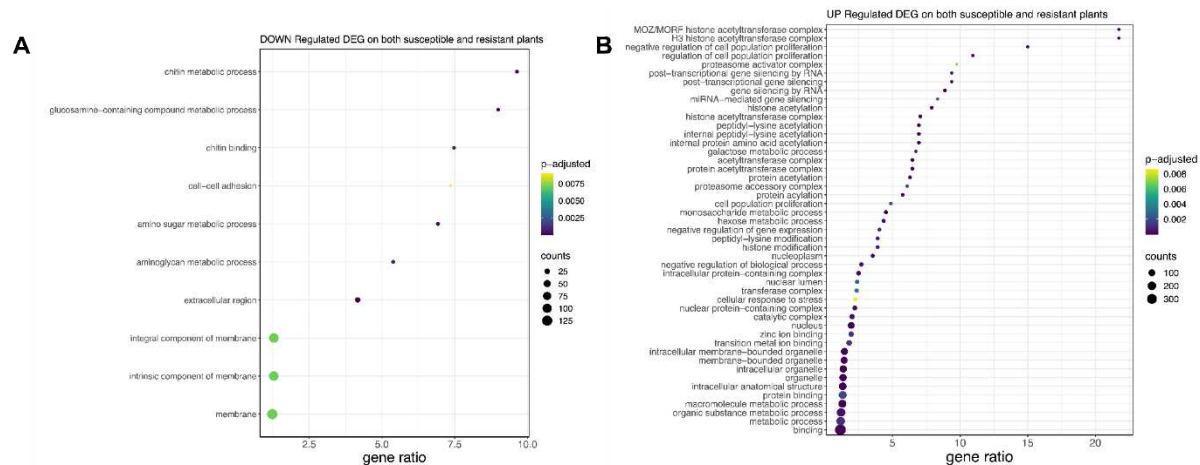


Figure S13. Gene ontology enrichment analysis of different gene sets. A) Significantly down-regulated genes (in the virulent biotype) when aphids were fed on susceptible and resistant soybean cultivars (constitutive down-regulation); B) Significantly up-regulated genes (in the virulent biotype) when aphids were fed on susceptible and resistant soybean cultivars (constitutive up-regulation); The gene set of up-regulated genes was enriched for genes of pathways associated with gene expression regulation (*e.g.*, RNA silencing and DNA methylation), and regulation of transposable elements (TE).

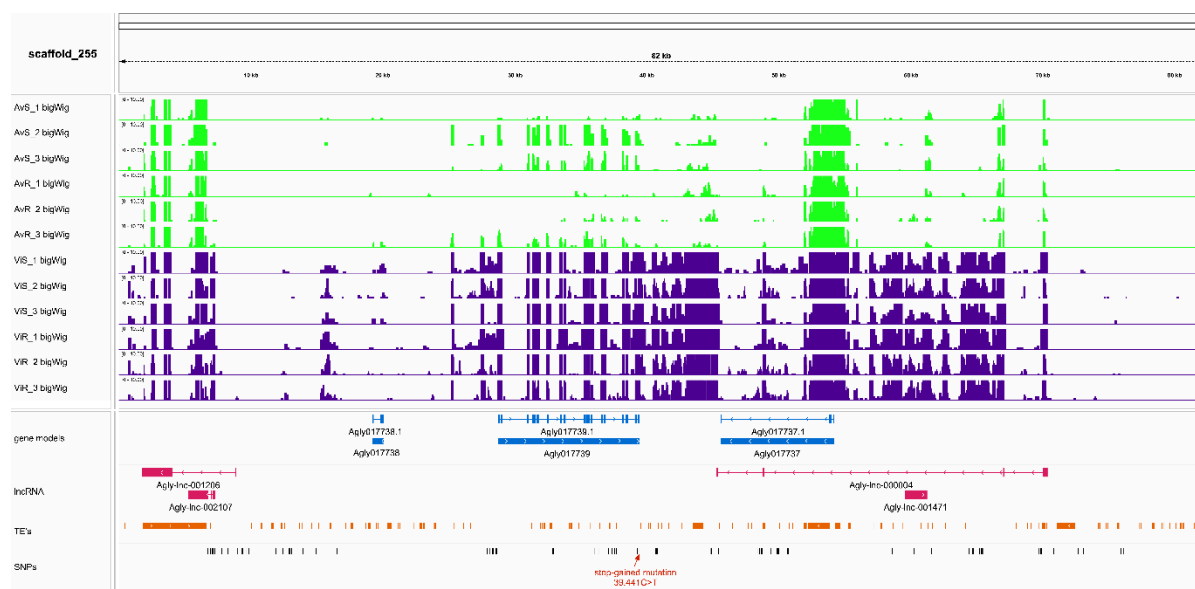


Figure S14. IGV view of the entire scaffold_255. Top panel shows the histograms of read counts of each biological replicate: Green corresponds to RNA pools from the avirulent biotype and purple from the virulent biotype. Bottom panel contain the genomic features: light-blue corresponds annotated gene models, with the gene Agly017739 that corresponds to the predicted PIWI gene; pink corresponds to annotated long non-coding RNAs; orange to annotated transposable elements, and black to the identified SNPs. Note the highlighted SNP that cause a premature stop-codon on Agly017739.

Table S1-Collection Metadata

Table S1. Details of the samples, their collection day, locality, coordinates, and number of individual aphids pooled, prefix file name, and associated SRA numbers. Each rows represents the technical replicated pool names because each library that consisted of each sample pool were sequenced four to five times.

SampleID	Continent	Country	City	Latitude	Longitude	CollectionDate	Season	Locality	Phenotype	Set	Pool size	Filename prefix	SRA accession	SRA experiment	Seq Platform	Station ID	Collector	Year	Year days
MN-Av.1.140711	North America	US	Lamberton	44.22	-95.26	2013-07-07	Summer	Minnesota	Avirulent	AndyLab_poolseq	5	Pool1_140711	SRR3190066	SRX1602077	illumina_MiSeq	USLMN0001	Bruce Potter	2013	188
MN-Av.1.140722	North America	US	Lamberton	44.22	-95.26	2013-07-07	Summer	Minnesota	Avirulent	AndyLab_poolseq	5	Pool1_140722	SRR3190066	SRX1602077	illumina_MiSeq	USLMN0001	Bruce Potter	2013	188
MN-Av.1.140725	North America	US	Lamberton	44.22	-95.26	2013-07-07	Summer	Minnesota	Avirulent	AndyLab_poolseq	5	Pool1_140725	SRR3190066	SRX1602077	illumina_MiSeq	USLMN0001	Bruce Potter	2013	188
MN-Av.1.140730	North America	US	Lamberton	44.22	-95.26	2013-07-07	Summer	Minnesota	Avirulent	AndyLab_poolseq	5	Pool1_140730	SRR3190066	SRX1602077	illumina_MiSeq	USLMN0001	Bruce Potter	2013	188
MN-Av.1.AddRun	North America	US	Lamberton	44.22	-95.26	2013-07-07	Summer	Minnesota	Avirulent	AndyLab_poolseq	5	Pool1_AddRun	SRR3190066	SRX1602077	illumina_MiSeq	USLMN0001	Bruce Potter	2013	188
MN-V.1.140711	North America	US	Lamberton	44.22	-95.26	2013-07-10	Summer	Minnesota	Virulent	AndyLab_poolseq	5	Pool2_140711	SRR3190067	SRX1602078	illumina_MiSeq	USLMN0001	Bruce Potter	2013	191
MN-V.1.140722	North America	US	Lamberton	44.22	-95.26	2013-07-10	Summer	Minnesota	Virulent	AndyLab_poolseq	5	Pool2_140722	SRR3190067	SRX1602078	illumina_MiSeq	USLMN0001	Bruce Potter	2013	191
MN-V.1.140725	North America	US	Lamberton	44.22	-95.26	2013-07-10	Summer	Minnesota	Virulent	AndyLab_poolseq	5	Pool2_140725	SRR3190067	SRX1602078	illumina_MiSeq	USLMN0001	Bruce Potter	2013	191
MN-V.1.140730	North America	US	Lamberton	44.22	-95.26	2013-07-10	Summer	Minnesota	Virulent	AndyLab_poolseq	5	Pool2_140730	SRR3190067	SRX1602078	illumina_MiSeq	USLMN0001	Bruce Potter	2013	191
ND-Av.1.140711	North America	US	Prosper	46.96	-97.02	2013-08-08	Summer	Nort_Dakota	Avirulent	AndyLab_poolseq	5	Pool3_140711	SRR3190068	SRX1602079	illumina_MiSeq	USPND0001	Deirdre Prischm	2013	220
ND-Av.1.140722	North America	US	Prosper	46.96	-97.02	2013-08-08	Summer	Nort_Dakota	Avirulent	AndyLab_poolseq	5	Pool3_140722	SRR3190068	SRX1602079	illumina_MiSeq	USPND0001	Deirdre Prischm	2013	220
ND-Av.1.140725	North America	US	Prosper	46.96	-97.02	2013-08-08	Summer	Nort_Dakota	Avirulent	AndyLab_poolseq	5	Pool3_140725	SRR3190068	SRX1602079	illumina_MiSeq	USPND0001	Deirdre Prischm	2013	220
ND-Av.1.140730	North America	US	Prosper	46.96	-97.02	2013-08-08	Summer	Nort_Dakota	Avirulent	AndyLab_poolseq	5	Pool3_140730	SRR3190068	SRX1602079	illumina_MiSeq	USPND0001	Deirdre Prischm	2013	220
ND-V.1.140711	North America	US	Prosper	46.96	-97.02	2013-08-11	Summer	Nort_Dakota	Virulent	AndyLab_poolseq	5	Pool4_140711	SRR3190069	SRX1602080	illumina_MiSeq	USPND0001	Deirdre Prischm	2013	223
ND-V.1.140722	North America	US	Prosper	46.96	-97.02	2013-08-11	Summer	Nort_Dakota	Virulent	AndyLab_poolseq	5	Pool4_140722	SRR3190069	SRX1602080	illumina_MiSeq	USPND0001	Deirdre Prischm	2013	223
ND-V.1.140725	North America	US	Prosper	46.96	-97.02	2013-08-11	Summer	Nort_Dakota	Virulent	AndyLab_poolseq	5	Pool4_140725	SRR3190069	SRX1602080	illumina_MiSeq	USPND0001	Deirdre Prischm	2013	223
ND-V.1.140730	North America	US	Prosper	46.96	-97.02	2013-08-11	Summer	Nort_Dakota	Virulent	AndyLab_poolseq	5	Pool4_140730	SRR3190069	SRX1602080	illumina_MiSeq	USPND0001	Deirdre Prischm	2013	223
NW-Av.1.140711	North America	US	Hoytville	41.22	-83.76	2013-06-07	Summer	Nort_Western	Avirulent	AndyLab_poolseq	5	Pool5_140711	SRR3190070	SRX1602081	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	158
NW-Av.1.140722	North America	US	Hoytville	41.22	-83.76	2013-06-07	Summer	Nort_Western	Avirulent	AndyLab_poolseq	5	Pool5_140722	SRR3190070	SRX1602081	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	158
NW-Av.1.140725	North America	US	Hoytville	41.22	-83.76	2013-06-07	Summer	Nort_Western	Avirulent	AndyLab_poolseq	5	Pool5_140725	SRR3190070	SRX1602081	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	158
NW-Av.1.140730	North America	US	Hoytville	41.22	-83.76	2013-06-07	Summer	Nort_Western	Avirulent	AndyLab_poolseq	5	Pool5_140730	SRR3190070	SRX1602081	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	158
NW-Av.2.140711	North America	US	Hoytville	41.22	-83.76	2013-06-15	Summer	Nort_Western	Avirulent	AndyLab_poolseq	5	Pool6_140711	SRR3190071	SRX1602082	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	166
NW-Av.2.140722	North America	US	Hoytville	41.22	-83.76	2013-06-15	Summer	Nort_Western	Avirulent	AndyLab_poolseq	5	Pool6_140722	SRR3190071	SRX1602082	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	166
NW-Av.2.140725	North America	US	Hoytville	41.22	-83.76	2013-06-15	Summer	Nort_Western	Avirulent	AndyLab_poolseq	5	Pool6_140725	SRR3190071	SRX1602082	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	166
NW-Av.2.140730	North America	US	Hoytville	41.22	-83.76	2013-06-15	Summer	Nort_Western	Avirulent	AndyLab_poolseq	5	Pool6_140730	SRR3190071	SRX1602082	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	166
NW-V.1.140711	North America	US	Hoytville	41.22	-83.76	2013-06-19	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool7_140711	SRR3190072	SRX1602083	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	170
NW-V.1.140722	North America	US	Hoytville	41.22	-83.76	2013-06-19	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool7_140722	SRR3190072	SRX1602083	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	170
NW-V.1.140725	North America	US	Hoytville	41.22	-83.76	2013-06-19	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool7_140725	SRR3190072	SRX1602083	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	170
NW-V.1.140730	North America	US	Hoytville	41.22	-83.76	2013-06-19	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool7_140730	SRR3190072	SRX1602083	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	170
NW-V.1.AddRun	North America	US	Hoytville	41.22	-83.76	2013-06-19	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool7_AddRun	SRR3190072	SRX1602083	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	170
NW-V.2.140711	North America	US	Hoytville	41.22	-83.76	2013-06-20	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool8_140711	SRR3190073	SRX1602084	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	171
NW-V.2.140722	North America	US	Hoytville	41.22	-83.76	2013-06-20	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool8_140722	SRR3190073	SRX1602084	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	171
NW-V.2.140725	North America	US	Hoytville	41.22	-83.76	2013-06-20	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool8_140725	SRR3190073	SRX1602084	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	171
NW-V.2.140730	North America	US	Hoytville	41.22	-83.76	2013-06-20	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool8_140730	SRR3190073	SRX1602084	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	171
NW-V.3.140711	North America	US	Hoytville	41.22	-83.76	2013-06-24	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool9_140711	SRR3190074	SRX1602085	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	175
NW-V.3.140722	North America	US	Hoytville	41.22	-83.76	2013-06-24	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool9_140722	SRR3190074	SRX1602085	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	175
NW-V.3.140725	North America	US	Hoytville	41.22	-83.76	2013-06-24	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool9_140725	SRR3190074	SRX1602085	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	175
NW-V.3.140730	North America	US	Hoytville	41.22	-83.76	2013-06-24	Summer	Nort_Western	Virulent	AndyLab_poolseq	5	Pool9_140730	SRR3190074	SRX1602085	illumina_MiSeq	USHOH0001	Jacob Wenger	2013	175
PA-Av.1.140711	North America	US	Rock Springs	40.79	-77.86	2013-06-08	Summer	Pennsylvania	Avirulent	AndyLab_poolseq	5	Pool10_140711	SRR3190075	SRX1602086	illumina_MiSeq	USRPA0001	Jacob Wenger	2013	159
PA-Av.1.140722	North America	US	Rock Springs	40.79	-77.86	2013-06-08	Summer	Pennsylvania	Avirulent	AndyLab_poolseq	5	Pool10_140722	SRR3190075	SRX1602086	illumina_MiSeq	USRPA0001	Andy Michel	2013	159
PA-Av.1.140725	North America	US	Rock Springs	40.79	-77.86	2013-06-08	Summer	Pennsylvania	Avirulent	AndyLab_poolseq	5	Pool10_140725	SRR3190075	SRX1602086	illumina_MiSeq	USRPA0001	Andy Michel	2013	159
PA-Av.1.140730	North America	US	Rock Springs	40.79	-77.86	2013-06-08	Summer	Pennsylvania	Avirulent	AndyLab_poolseq	5	Pool10_140730	SRR3190075	SRX1602086	illumina_MiSeq	USRPA0001	Andy Michel	2013	159
PA-V.1.140711	North America	US	Rock Springs	40.79	-77.86	2013-06-10	Summer	Pennsylvania	Virulent	AndyLab_poolseq	5	Pool11_140711	SRR3190076	SRX1602087	illumina_MiSeq	USRPA0001	Andy Michel	2013	161
PA-V.1.140722	North America	US	Rock Springs	40.79	-77.86	2013-06-10	Summer	Pennsylvania	Virulent	AndyLab_poolseq	5	Pool11_140722	SRR3190076	SRX1602087	illumina_MiSeq	USRPA0001	Andy Michel	2013	161
PA-V.1.140725	North America	US	Rock Springs	40.79	-77.86	2013-06-10	Summer	Pennsylvania	Virulent	AndyLab_poolseq	5	Pool11_140725	SRR3190076	SRX1602087	illumina_MiSeq	USRPA0001	Andy Michel	2013	161
PA-V.1.140730	North America	US	Rock Springs	40.79	-77.86	2013-06-10	Summer	Pennsylvania	Virulent	AndyLab_poolseq	5	Pool11_140730	SRR3190076	SRX1602087	illumina_MiSeq	USRPA0001	Andy Michel	2013	161
PA-V.2.140711	North America	US	Rock Springs	40.79	-77.86	2013-06-11	Summer	Pennsylvania	Virulent	AndyLab_poolseq	5	Pool12_140711	SRR3190077	SRX1602088	illumina_MiSeq	USRPA0001	Andy Michel	2013	162
PA-V.2.140722	North America	US	Rock Springs	40.79	-77.86	2013-06-11	Summer	Pennsylvania	Virulent	AndyLab_poolseq	5	Pool12_140722	SRR3190077	SRX1602088	illumina_MiSeq	USRPA0001	Andy Michel	2013	162
PA-V.2.140725	North America	US	Rock Springs	40.79	-77.86	2013-06-11	Summer	Pennsylvania	Virulent	AndyLab_poolseq	5	Pool12_140725	SRR3190077	SRX1602088	illumina_MiSeq	USRPA0001	Andy Michel	2013	162
PA-V.2.140730	North America	US	Rock Springs	40.79	-77.86	2013-06-11	Summer	Pennsylvania	Virulent	AndyLab_poolseq	5	Pool12_140730	SRR3190077	SRX1602088	illumina_MiSeq	USRPA0001	Andy Michel	2013	162
WI-Av.1.140711	North America	US	Arlington	43.34	-89.41	2013-08-25	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool17_140711	SRR3190078	SRX1602089	illumina_MiSeq	USAWI0001	Michael Crossley	2013	237
WI-Av.1.140722	North America	US	Arlington	43.34	-89.41	2013-08-25	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool17_140722	SRR3190078	SRX1602089	illumina_MiSeq	USAWI0001	Michael Crossley	2013	237
WI-Av.1.140725	North America	US	Arlington	43.34	-89.41	2013-08-25	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool17_140725	SRR3190078	SRX1602089	illumina_MiSeq	USAWI0001	Michael Crossley	2013	237
WI-Av.1.140730	North America	US	Arlington	43.34	-89.41	2013-08-25	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool17_140730	SRR3190078	SRX1602089	illumina_MiSeq	USAWI0001	Michael Crossley	2013	237
WI-Av.2.140711	North America	US	Arlington	43.34	-89.41	2013-08-26	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool18_140711	SRR3190079	SRX1602090	illumina_MiSeq	USAWI0001	Michael Crossley	2013	238
WI-Av.2.140722	North America	US	Arlington	43.34	-89.41	2013-08-26	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool18_140722	SRR3190079	SRX1602090	illumina_MiSeq	USAWI0001	Michael Crossley	2013	238
WI-Av.2.140725	North America	US	Arlington	43.34	-89.41	2013-08-26	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool18_140725	SRR3190079	SRX1602090	illumina_MiSeq	USAWI0001	Michael Crossley	2013	238
WI-Av.2.140730	North America	US	Arlington	43.34	-89.41	2013-08-26	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool18_140730	SRR3190079	SRX1602090	illumina_MiSeq	USAWI0001	Michael Crossley	2013	238
WI-Av.2.AddRun	North America	US	Arlington	43.34	-89.41	2013-08-26	Summer	Wisconsin	Avirulent	AndyLab_poolseq	5	Pool18_AddRun	SRR3190079	SRX1602090	illumina_MiSeq	USAWI0001	Michael Crossley	2013	238
WI-V.1.140711	North America	US	Arlington	43.34	-89.41	2013-08-30	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool19_140711	SRR3190080	SRX1602091	illumina_MiSeq	USAWI0001	Michael Crossley	2013	242
WI-V.1.140722	North America	US	Arlington	43.34	-89.41	2013-08-30	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool19_140722	SRR3190080	SRX1602091	illumina_MiSeq	USAWI0001	Michael Crossley	2013	242
WI-V.1.140725	North America	US	Arlington	43.34	-89.41	2013-08-30	Summer	Wisconsin	Virulent										

Table S1-Collection Metadata

WI-V.2.140722	North America	US	Arlington	43.34	-89.41	2013-08-31	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool20_140722	SRR3190081	SRX1602092	Illumina_MiSeq	USAWI0001	Michael Crossley	2013	243
WI-V.2.140725	North America	US	Arlington	43.34	-89.41	2013-08-31	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool20_140725	SRR3190081	SRX1602092	Illumina_MiSeq	USAWI0001	Michael Crossley	2013	243
WI-V.2.140730	North America	US	Arlington	43.34	-89.41	2013-08-31	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool20_140730	SRR3190081	SRX1602092	Illumina_MiSeq	USAWI0001	Michael Crossley	2013	243
WI-V.3.140711	North America	US	Arlington	43.34	-89.41	2013-09-03	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool21_140711	SRR3190082	SRX1602093	Illumina_MiSeq	USAWI0001	Michael Crossley	2013	246
WI-V.3.140722	North America	US	Arlington	43.34	-89.41	2013-09-03	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool21_140722	SRR3190082	SRX1602093	Illumina_MiSeq	USAWI0001	Michael Crossley	2013	246
WI-V.3.140725	North America	US	Arlington	43.34	-89.41	2013-09-03	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool21_140725	SRR3190082	SRX1602093	Illumina_MiSeq	USAWI0001	Michael Crossley	2013	246
WI-V.3.140730	North America	US	Arlington	43.34	-89.41	2013-09-03	Summer	Wisconsin	Virulent	AndyLab_poolseq	5	Pool21_140730	SRR3190082	SRX1602093	Illumina_MiSeq	USAWI0001	Michael Crossley	2013	246
WO-Av.1.140711	North America	US	Wooster	40.78	-81.93	2013-06-26	Summer	Wooster	Avirulent	AndyLab_poolseq	5	Pool13_140711	SRR3190083	SRX1602094	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	177
WO-Av.1.140722	North America	US	Wooster	40.78	-81.93	2013-06-26	Summer	Wooster	Avirulent	AndyLab_poolseq	5	Pool13_140722	SRR3190083	SRX1602094	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	177
WO-Av.1.140725	North America	US	Wooster	40.78	-81.93	2013-06-26	Summer	Wooster	Avirulent	AndyLab_poolseq	5	Pool13_140725	SRR3190083	SRX1602094	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	177
WO-Av.1.140730	North America	US	Wooster	40.78	-81.93	2013-06-26	Summer	Wooster	Avirulent	AndyLab_poolseq	5	Pool13_140730	SRR3190083	SRX1602094	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	177
WO-V.1.140711	North America	US	Wooster	40.78	-81.93	2013-07-01	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool14_140711	SRR3190084	SRX1602095	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	182
WO-V.1.140722	North America	US	Wooster	40.78	-81.93	2013-07-01	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool14_140722	SRR3190084	SRX1602095	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	182
WO-V.1.140725	North America	US	Wooster	40.78	-81.93	2013-07-01	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool14_140725	SRR3190084	SRX1602095	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	182
WO-V.1.140730	North America	US	Wooster	40.78	-81.93	2013-07-01	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool14_140730	SRR3190084	SRX1602095	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	182
WO-V.2.140711	North America	US	Wooster	40.78	-81.93	2013-07-04	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool15_140711	SRR3190085	SRX1602096	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	185
WO-V.2.140722	North America	US	Wooster	40.78	-81.93	2013-07-04	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool15_140722	SRR3190085	SRX1602096	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	185
WO-V.2.140725	North America	US	Wooster	40.78	-81.93	2013-07-04	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool15_140725	SRR3190085	SRX1602096	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	185
WO-V.2.140730	North America	US	Wooster	40.78	-81.93	2013-07-04	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool15_140730	SRR3190085	SRX1602096	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	185
WO-V.3.140711	North America	US	Wooster	40.78	-81.93	2013-07-06	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool16_140711	SRR3190086	SRX1602097	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	187
WO-V.3.140722	North America	US	Wooster	40.78	-81.93	2013-07-06	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool16_140722	SRR3190086	SRX1602097	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	187
WO-V.3.140725	North America	US	Wooster	40.78	-81.93	2013-07-06	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool16_140725	SRR3190086	SRX1602097	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	187
WO-V.3.140730	North America	US	Wooster	40.78	-81.93	2013-07-06	Summer	Wooster	Virulent	AndyLab_poolseq	5	Pool16_140730	SRR3190086	SRX1602097	Illumina_MiSeq	USWOH0002	Jacob Wenger	2013	187

Table S2. List of primer sequences and associated annealing temperatures used to PCR amplify scaffold regions encompassing SNPs

Gene	SNP(s)	Primer sequence (5'-3')	T _{annealing} (°C)	Product length (bp)
Agly018942	[T/C]	AGs25F: TAGTCGGATGCCGTACGTTG	57	525
		AGs25R: TGGCACTATATGACTGTATGCTT		
Agly010975	[T/G][G/A][G/A]	AGs54F: CGACAGGTAATTTTCTGGGGAT	57	489
		AGs54R: GTTGCAGTCACCA GTGGAGT		
Agly002329	[C/T][G/A][C/A]	AGs66F: TGTCAATACCGTGGTCGGTG	59	698
		AGs66R: GGAGACCATTACTCCCCAATGA		
Agly010068	[A/T][A/T][A/T]	AGs105F: AGTGGCAGACTATCGGTGC	57	442
		AGs105R: ATGTACACGTGGCCTACCAG		
Agly000978	[A/G]	AGs114dF: TGGCAACGGATATACGCACT	57	460
		AGs114dR: GCGATTATAATCTGTCGTGTT CAT		
Agly000981	[T/A]	AGs114cF: TTTATACCTCCCATCCTCCCATCCT	57	632
		AGs114cR: CTCCATTGGATCACCGGGC		
Agly-lnc-000191	[A/C][T/A][A/C]	AGs114aF: TGCTGCCATACA ACTAACCCA	55	358
		AGs114aR2: CGGCGTCATCAATATCACTTGG		
		AGs114aF2: CCATCTAATCAGCCGCTAGAAC		297
		AGs114aR: AAGGAGTTGAGGGGTGGGTT		
Agly-lnc-000554	[A/T][C/T][C/G]	AGs114bF: CGTAGCTTTTACCACGGTCCT	57	451
		AGs114bR: GGTCTTCGGAGAACAACGAGT		
Δslv002471	T A/GTTC/GTTC/TT	AGs129bF: AGTAGGTGGTAAAGGCGTGC	57	593

Agly002471	[A/G][C/G][C/T]	AGs129bR: TGAATATGTTCTGGCGACACT	57	393
Agly002450	[C/A]	AGs129aF: CTCTGCCCTGCTACTGTGAA	57	457
		AGs129aR: CCGAGTTGCCGCGTTTATTT		
Agly004871	[A/T][T/C]	AGs133F: CGGCTATTGGAGTTGTTTGAGA	57	557
		AGs133R: CCGCCATTAGATCGCGTTTG		
Agly001847	[T/C][C/A][A/T][C/T]	AGs237bF2: GTCAGTGAAACTGCAATACCTCAA	55	574
		AGs237bR: ACACCGAGTTATTGGTGCGT		505
		AGs237cF: TGGTTGTTTCAGTAGTTCCCAA		
		AGs237cR: GGTATTGCAGTTTCACTGACGA		
Agly001843	[G/T]	AGs237aF: CAGGCTAGTTGTGTCGTCTATG	57	333
		AGs237aR: CTCGATACCATATTGCGCGT		
Agly014114	[C/T][AAG][C/T]	AGs279F: ATGTGAAAATGTTAGCACAACAAAA	57	218
		AGs279R: ACCTCACATGTCTAGATGTTCTCC		
Agly008028	[A/G]	AGs615F: GTGCATTGCCGTCCAAAAA	55	389
		AGs615R: TGCTGTTCCAGTTGTGTCGT		

Table S3-Sequencing Statistics

Table S3. Sequencing statistics for the technical replicates data. Total number of sequenced reads, number of mapped reads (in millions), proportion of PCR duplicates, average read depth, proportion of missing data, and the effective read depth (from 40,105 SNPs). The table also shows the the percentage of mapped reads to each of the other genomes in the reference hologenome.

SampleID	Total Number of Reads (M)	Number of Mapped Reads (M)	Proportion of PCR duplicates	Average Read Depth	Proportion of Missing Data	Effective Read Depth	Targeted genome (%)	mitochondrion genome (%)	<i>B.aphidicola</i> genome (%)	<i>Arsenophonus</i> genome (%)	<i>CR insecticola</i> genome (%)	<i>Wolbachia</i> genome (%)
MN-Av.1.140711	1.0	0.941295	0.005514	9.82174152	0.228849271	4.955034607	97.58654289	0.005029127	1.055488035	0.504169984	0.011944177	0.836825783
MN-Av.1.140722	1.2	1.195218	0.006532	10.88233878	0.171400075	5.211264358	97.63179606	0.005197591	1.011715205	0.508456398	0.013860243	0.828974507
MN-Av.1.140725	1.3	1.242156	0.006994	11.48018805	0.146066575	5.344547274	97.59843998	0.00651066	1.039005997	0.507196264	0.013021319	0.835825778
MN-Av.1.140730	1.3	1.26203	0.007448	11.79477762	0.131031044	5.41174488	97.64660343	0.005317527	1.032929566	0.499691114	0.013137419	0.802320944
MN-Av.1.AddRun	5.0	4.72507	0.048616	34.40671483	0.061264182	7.748088316	97.65171839	0.005427236	0.98168266	0.511454391	0.01193992	0.837777405
MN-V.1.140711	1.2	1.325917	0.007418	9.982082146	0.185911981	4.99551652	97.70963758	0.00423852	0.848745063	0.498806894	0.022679801	0.915892145
MN-V.1.140722	1.2	1.386017	0.007213	10.04932885	0.177085151	5.01230187	97.6938285	0.00405365	0.852262221	0.521498571	0.021263886	0.907093177
MN-V.1.140725	1.3	1.460293	0.007767	10.7564937	0.157162449	5.182230609	97.69898348	0.005468805	0.846584495	0.515350463	0.023022993	0.910589767
MN-V.1.140730	1.3	1.376763	0.007852	10.74334774	0.151951128	5.17917738	97.70630573	0.003580764	0.827657822	0.526085867	0.026139578	0.910230243
ND-Av.1.140711	1.3	1.393514	0.009736	12.02942667	0.114524374	5.460617224	98.21350474	0.004487429	0.594833616	0.345745704	0.00876117	0.832667342
ND-Av.1.140722	1.2	1.370698	0.009298	11.72062268	0.127041516	5.396080422	98.19502891	0.003834822	0.604599471	0.357072569	0.010636204	0.828828021
ND-Av.1.140725	1.3	1.46191	0.009889	12.40460307	0.102954744	5.536631482	98.18302483	0.002646476	0.613982349	0.340716774	0.010246611	0.849382964
ND-Av.1.140730	1.3	1.357519	0.009878	12.26791943	0.11362673	5.509234695	98.23805265	0.003290626	0.591654534	0.347782593	0.011992503	0.807227092
ND-V.1.140711	1.0	0.834296	0.006053	10.61694964	0.311233013	5.14962195	98.30871321	0.009660407	0.649632547	0.338949091	0.023256535	0.66978821
ND-V.1.140722	1.2	0.962659	0.0065	11.12864945	0.260391472	5.267089823	98.30008926	0.007851305	0.63223663	0.349486359	0.020764634	0.689571815
ND-V.1.140725	1.2	1.004077	0.006746	11.40992936	0.237601297	5.329269969	98.31029505	0.008024513	0.640574119	0.341685762	0.019813613	0.679606938
ND-V.1.140730	1.2	0.971671	0.007101	11.96437552	0.226580227	5.447173088	98.33056477	0.010138893	0.653190474	0.337246195	0.022018807	0.646840864
NW-Av.1.140711	2.0	1.002635	0.043785	6.929248676	0.618675976	4.093063318	98.32161825	0.00138562	0.11837154	1.288032796	0.026029862	0.244561936
NW-Av.1.140722	1.1	0.550152	0.025694	5.575695394	0.781274155	3.579740906	98.2691715	0.001441906	0.12057943	1.340252081	0.030099798	0.238455286
NW-Av.1.140725	1.1	0.580944	0.026782	5.752197213	0.767360678	3.651679277	98.28230915	0.00221968	0.117130807	1.326856421	0.025782437	0.245701504
NW-Av.1.140730	1.2	0.595792	0.028496	5.907627206	0.758409176	3.713707349	98.31501661	0.001332529	0.129421915	1.289555302	0.025984323	0.238689324
NW-Av.2.140711	1.3	1.22141	0.013696	12.38286143	0.136092757	5.53229598	98.33668213	0.008555324	0.706832689	0.324043067	0.039761885	0.584124904
NW-Av.2.140722	1.0	0.958949	0.010768	10.44502484	0.217154968	5.108834508	98.35862043	0.007364167	0.682067111	0.330142886	0.044185004	0.577620399
NW-Av.2.140725	1.0	0.99901	0.011579	11.05956015	0.199551178	5.251562745	98.3649087	0.008269107	0.683645899	0.330365754	0.037260794	0.575549746
NW-Av.2.140730	1.0	0.971818	0.011745	10.86290023	0.193217803	5.20680256	98.33531254	0.008087467	0.701255093	0.33404312	0.043201409	0.578100366
NW-V.1.140711	2.1	1.941919	0.01531	14.79379288	0.069642189	5.966732461	98.36329332	0.006701734	0.577269969	0.379236289	0.035350368	0.638148316
NW-V.1.140722	1.0	0.943063	0.008612	9.360280648	0.278568757	4.83478562	98.37723023	0.006005576	0.556200599	0.391942835	0.035085205	0.633535556
NW-V.1.140725	1.1	1.013697	0.009319	9.953751115	0.245206333	4.98841098	98.35146522	0.00646984	0.587481056	0.374956623	0.035682147	0.643945112
NW-V.1.140730	1.1	0.964763	0.009137	9.687286477	0.248248348	4.920579831	98.40010632	0.005872493	0.55191129	0.38325742	0.035441009	0.623411465
NW-V.1.AddRun	4.3	3.883605	0.049382	26.41702954	0.049445206	7.254031939	98.34989413	0.005879723	0.547120994	0.404652753	0.03389788	0.658554524
NW-V.2.140711	1.6	1.489762	0.014172	12.98116745	0.108939035	5.648610967	98.79208867	0.003486229	0.495044527	0.250941449	0.013475616	0.444963505
NW-V.2.140722	1.0	0.983443	0.00976	9.762375263	0.240393966	4.93987951	98.79669617	0.005075946	0.484651354	0.266487181	0.012791385	0.434297966
NW-V.2.140725	1.1	1.024317	0.010539	10.24419343	0.214162823	5.060311967	98.78562486	0.003997863	0.500122861	0.267954303	0.013163694	0.429136423
NW-V.2.140730	1.2	1.051663	0.011378	10.41257593	0.195412043	5.101059251	98.81725935	0.003514408	0.476344712	0.265575237	0.014057631	0.423248661
NW-V.3.140711	1.8	1.572876	0.017673	15.15671228	0.099912729	6.024917769	98.34139494	0.006455194	0.733297345	0.318519508	0.029617947	0.570715065
NW-V.3.140722	1.3	1.187266	0.014196	12.62427763	0.171574617	5.579969375	98.34355026	0.006287769	0.725105467	0.331071972	0.030851984	0.56313255
NW-V.3.140725	1.4	1.283243	0.01513	13.56775221	0.13863608	5.756913976	98.330564	0.005895504	0.733835072	0.328364044	0.028391505	0.572949877
NW-V.3.140730	1.4	1.213309	0.015054	13.01370382	0.144819848	5.654762885	98.336754	0.005331968	0.711366608	0.339113187	0.032894143	0.574540097
PA-Av.1.140711	1.2	1.028777	0.05298	13.89695652	0.713252712	5.815366701	97.94295171	9.59892E-05	0.113555209	1.376292974	0.137552506	0.429551615
PA-Av.1.140722	1.3	1.16155	0.055059	15.26007357	0.708540082	6.041183344	97.91795743	0.000339953	0.112949407	1.40043667	0.13683111	0.431485433
PA-Av.1.140725	1.4	1.186515	0.058341	15.09301564	0.69547438	6.014827335	97.95714711	8.32289E-05	0.113274507	1.364870421	0.130336428	0.434288302
PA-Av.1.140730	1.5	1.237922	0.066404	16.68577443	0.692033412	6.252685105	97.9719966	0.000718527	0.106900896	1.363445482	0.13428477	0.422653729
PA-V.1.140711	1.1	1.054713	0.041981	33.3119275	0.232365042	7.691167173	98.90042168	0.007979802	0.091527373	0.731353701	0.004037972	0.264679472
PA-V.1.140722	1.2	1.203852	0.040328	34.14213326	0.207754644	7.734590682	98.87526644	0.005543042	0.10405801	0.741591794	0.004031303	0.269509407
PA-V.1.140725	1.3	1.257274	0.044239	36.32520274	0.203665378	7.841347818	98.89517978	0.007569468	0.100657819	0.719099459	0.004026313	0.273467163
PA-V.1.140730	1.2	1.186273	0.04368	35.44557017	0.200149607	7.799565511	98.87702541	0.006396043	0.097731537	0.741514583	0.002387856	0.274944568

Table S3-Sequencing Statistics

PA-V.2.140711	1.3	1.185427	0.027912	9.73546554	0.315496821	4.932979929	98.24285522	0.001422722	0.106955183	0.813378269	0.002259617	0.833128991
PA-V.2.140722	1.2	1.09215	0.025545	9.011591905	0.339633462	4.740051201	98.1573331	0.001451848	0.106801543	0.873740068	0.002268512	0.858404928
PA-V.2.140725	1.3	1.121053	0.026496	9.435229661	0.335544196	4.854704485	98.19378437	0.001768719	0.11151772	0.845624455	0.001945591	0.845359147
PA-V.2.140730	1.4	1.223853	0.030071	9.802704824	0.30123426	4.950184791	98.25987715	0.001783245	0.107967376	0.821832754	0.002512754	0.806026719
WI-Av.1.140711	1.0	1.267256	0.003472	8.103009259	0.375239995	4.476056518	95.90322163	0.009065354	1.437582369	1.665358919	0.009065354	0.975706374
WI-Av.1.140722	1.3	1.6665	0.004218	9.20715854	0.284553048	4.793607821	95.81792758	0.00948788	1.400561521	1.752018055	0.009719292	1.010285672
WI-Av.1.140725	1.3	1.758077	0.004547	9.60515927	0.267323276	4.899301831	95.87290454	0.008671637	1.392620986	1.709739456	0.010043731	1.006019653
WI-Av.1.140730	1.3	1.631838	0.004431	9.310560669	0.255080414	4.821486454	95.86660509	0.008514417	1.398020043	1.748825868	0.009756103	0.968278473
WI-Av.2.140711	0.9	0.985753	0.006898	9.357051872	0.295299838	4.833924057	98.27606167	0.006543416	0.633301958	0.343982323	0.011878816	0.728231818
WI-Av.2.140722	1.3	1.460393	0.009158	10.86492884	0.157287121	5.207268581	98.24607961	0.004818836	0.621358301	0.371389691	0.012216766	0.744136801
WI-Av.2.140725	1.4	1.499966	0.009475	11.41819869	0.140107219	5.331073287	98.27999154	0.005620727	0.617221906	0.350402709	0.012101094	0.734662029
WI-Av.2.140730	1.5	1.590976	0.010422	12.15213818	0.104974442	5.485763081	98.29007769	0.005922726	0.613906188	0.356797513	0.013404065	0.71989182
WI-Av.2.AddRun	3.7	3.880215	0.048615	24.62848979	0.048821843	7.112204413	98.24944138	0.00671769	0.600045568	0.36921755	0.012388136	0.762189671
WI-V.1.140711	1.2	1.304998	0.009274	12.59942583	0.14883431	5.57510882	98.68457962	0.006955448	0.550091835	0.238701792	0.006114679	0.513556626
WI-V.1.140722	1.3	1.449373	0.009539	12.79622706	0.126318414	5.613309179	98.66716629	0.005087939	0.530727025	0.268354396	0.008319468	0.52034488
WI-V.1.140725	1.4	1.545349	0.010315	13.83299143	0.112704152	5.804135612	98.65709303	0.004966563	0.545225392	0.258906273	0.00715959	0.526649157
WI-V.1.140730	1.3	1.440719	0.010127	13.58574968	0.114275028	5.760151728	98.69803761	0.005121749	0.537299136	0.253664992	0.0080979	0.497778614
WI-V.2.140711	1.3	1.589582	0.008655	12.10467268	0.12432365	5.476069633	98.76097945	0.007528226	0.564303285	0.237891946	0.000941028	0.428356068
WI-V.2.140722	1.2	1.515086	0.007933	11.3202709	0.138461538	5.309628077	98.75847622	0.006776374	0.540333569	0.258225893	0.001052641	0.4351353
WI-V.2.140725	1.3	1.601062	0.008674	12.10479624	0.114649046	5.47609492	98.75586825	0.005416848	0.55482224	0.248739182	0.001120727	0.43403275
WI-V.2.140730	1.3	1.532126	0.008538	12.0320635	0.123575614	5.461160504	98.79004837	0.006507565	0.539086715	0.251061872	0.000976135	0.412319342
WI-V.3.140711	1.2	1.172508	0.008395	11.81162108	0.193242738	5.415288041	98.26339105	0.014735685	0.752028062	0.336634183	0.030572312	0.602638704
WI-V.3.140722	1.1	1.150533	0.008049	11.63427458	0.193791298	5.377704964	98.23250985	0.011555041	0.76306388	0.344581677	0.031647016	0.616642536
WI-V.3.140725	1.2	1.206207	0.00843	12.00690502	0.187507792	5.455971663	98.26016461	0.010699588	0.736213992	0.33563786	0.033004115	0.624279835
WI-V.3.140730	1.2	1.176796	0.008922	12.103041	0.180875203	5.47573567	98.30406193	0.010972029	0.716304688	0.343846506	0.033169287	0.59164556
WO-Av.1.140711	1.1	0.769515	0.012858	10.23830018	0.310559781	5.058873566	98.45307141	0.006080592	0.085645791	0.723331751	0.058606561	0.673263894
WO-Av.1.140722	1.3	0.860157	0.013047	10.41479744	0.266325895	5.101592349	98.46458683	0.007518597	0.076342677	0.732889698	0.0608428	0.657819399
WO-Av.1.140725	1.3	0.914461	0.014253	11.26590517	0.249071188	5.297637264	98.40765579	0.005331024	0.076375078	0.726542407	0.06418988	0.719905826
WO-Av.1.140730	1.3	0.843663	0.01384	10.78086263	0.261762872	5.187880226	98.40838262	0.007195484	0.082807041	0.760715962	0.066174857	0.674724036
WO-V.1.140711	1.1	1.260959	0.005697	9.772931887	0.324223912	4.942581071	98.62346422	0.015619145	0.665943558	0.449878715	0.007888457	0.237205908
WO-V.1.140722	1.0	1.171908	0.005357	9.022833178	0.35788555	4.743159493	98.63019044	0.012894721	0.63421665	0.471675304	0.012216051	0.238806831
WO-V.1.140725	1.1	1.244141	0.00566	9.665664575	0.336466775	4.914995137	98.60817462	0.016146837	0.663219344	0.454749288	0.008792832	0.248917083
WO-V.1.140730	1.1	1.155718	0.005514	9.510517491	0.345667622	4.874559322	98.63218706	0.013253001	0.635455563	0.457658814	0.01273665	0.248708908
WO-V.2.140711	1.2	1.197614	0.013737	9.50872139	0.212342601	4.87408744	99.15333685	0.00359606	0.179217614	0.2599199	0.008279302	0.395650272
WO-V.2.140722	1.4	1.383357	0.015028	10.06547637	0.169056227	5.016315678	99.13244361	0.002459921	0.177910196	0.280069283	0.009767335	0.397349652
WO-V.2.140725	1.5	1.465563	0.016039	10.57529566	0.146116444	5.139802527	99.13653299	0.001707807	0.184579736	0.265393144	0.007172788	0.404613537
WO-V.2.140730	1.5	1.403063	0.016037	10.48962388	0.148111208	5.119480935	99.14711606	0.002640287	0.180110364	0.270736419	0.008563092	0.390833781
WO-V.3.140711	1.3	0.510601	0.012137	6.438001618	0.599227029	3.916535457	96.51323182	0.006295872	0.096727482	2.670785057	0.160449334	0.552510431
WO-V.3.140722	1.2	0.469269	0.011149	6.229470006	0.657499065	3.838369339	96.29138195	0.004347007	0.103500169	2.854327653	0.186714304	0.559728912
WO-V.3.140725	1.3	0.490466	0.012077	6.211714097	0.623238998	3.831620802	96.50918724	0.0027801	0.093133351	2.686172378	0.159260017	0.549466916
WO-V.3.140730	1.3	0.473279	0.011935	6.494542066	0.625383369	3.937388525	96.34216726	0.003697063	0.099204515	2.796211743	0.198819815	0.559899604
Average	1.4	1.283714	0.016647	12.50297173	0.264904625	5.329067307	98.16381936	0.005894141	0.526071673	0.677183632	0.031760948	0.59527025

118.7

Estimated genome size	3.17E+08
Read length	300
Total number of pair-end reads	237400000
Expected coverage	224.6687697
Expected coverage per sample	10.69851284

Table S4-Pooled Sequencing Stat

Table S4. Sequencing statistics for the aggregated data. Total number of sequenced reads, number of mapped reads (in millions), proportion of PCR duplicates, average read depth, proportion of missing data, and the effective read depth (from 262,866 SNPs). The table also shows the the percentage of mapped reads to each of the other genomes in the reference hologenome.

SampleID	Number of Mapped Reads	Proportion of PCR duplicates	Average Read Depth	Proportion of Missing Data	Effective Read Depth	Targeted genome (%)	mitochondrion genome (%)	<i>B.aphidicola</i> genome (%)	<i>Arsenophonus</i> genome (%)	<i>CR insecticola</i> genome (%)	<i>Wolbachia</i> genome (%)
MN-Av.1	9.161101	0.050103	23.61356033	0.031126125	7.025010174	97.63486847	0.005486836	1.007439875	0.508190497	0.012490186	0.831524134
MN-V.1	5.468019	0.022235	13.32676664	0.091818645	5.713079247	97.70205779	0.004352904	0.843823683	0.515596987	0.023274709	0.910893932
ND-Av.1	5.487891	0.026765	13.72541079	0.085967755	5.78510986	98.20695045	0.003554183	0.60147431	0.347705677	0.010395984	0.829919393
ND-V.1	3.730108	0.01791	11.20519626	0.190142506	5.284174747	98.31255928	0.008886402	0.643697207	0.341928708	0.021385377	0.671543025
NW-Av.1	2.542317	0.099036	9.392686343	0.698519398	4.843416831	98.30123316	0.001562937	0.120964034	1.307160322	0.026788009	0.242291541
NW-Av.2	4.065569	0.032655	11.95114071	0.196944451	5.444428091	98.34822231	0.008101649	0.694225058	0.329315267	0.040987634	0.57914808
NW-V.1	8.474024	0.059026	18.57957632	0.035109143	6.500997815	98.36153085	0.006143308	0.559994913	0.39184314	0.034725022	0.645762764
NW-V.2	4.449265	0.0334	12.26604916	0.193783144	5.508857486	98.79744754	0.003951721	0.489618172	0.261516082	0.013391942	0.434074544
NW-V.3	5.118118	0.041643	13.70400632	0.142837035	5.781303858	98.33816661	0.006021479	0.726515997	0.328511417	0.030353554	0.570430943
PA-Av.1	4.177778	0.146634	19.81115104	0.688731141	6.645550523	97.948094	0.000321028	0.11154664	1.375991576	0.134639277	0.429407478
PA-V.1	4.533738	0.076614	26.48292709	0.35526466	7.258991863	98.88666654	0.006845719	0.098746265	0.7332669	0.003616606	0.270857969
PA-V.2	4.340807	0.086171	13.59364243	0.46397404	5.761570081	98.21523768	0.001608957	0.10829353	0.837708802	0.00225254	0.834898491
WI-Av.1	6.27457	0.012411	11.76073615	0.090061096	5.404567229	95.86285253	0.008925177	1.405112707	1.722086166	0.009688012	0.991335408
WI-Av.2	9.163901	0.052094	19.05367302	0.029117497	6.558094396	98.26343064	0.006095996	0.611905386	0.361822132	0.012434148	0.744311693
WI-V.1	5.641261	0.026999	13.78012285	0.085248758	5.794807258	98.67615711	0.00548808	0.540680043	0.255386741	0.007450589	0.51483744
WI-V.2	6.137542	0.024496	13.75489288	0.07205192	5.790340941	98.76619868	0.006552911	0.549853191	0.248850803	0.001022893	0.427521524
WI-V.3	4.634561	0.023677	12.17535914	0.149144431	5.490490171	98.26517492	0.011982351	0.741745521	0.340125794	0.032107639	0.608863776
WO-Av.1	3.29668	0.040076	11.64103941	0.270495233	5.37914986	98.43260556	0.006521019	0.080073411	0.735935161	0.062566532	0.682298313
WO-V.1	4.782773	0.015825	10.08807383	0.159453866	5.021921919	98.62324549	0.014528327	0.65025583	0.458280241	0.01033034	0.243359772
WO-V.2	5.275007	0.046687	12.71645382	0.125695221	5.597904463	99.14191074	0.002553673	0.180557546	0.269293092	0.008432633	0.397252321
WO-V.3	1.894082	0.037851	7.911294825	0.611623413	4.416930715	96.41687224	0.004304959	0.098062969	2.749617559	0.175852582	0.555289694
Average	5.173767	0.046300	14.311131398	0.22700521323013	5.762223692	98.166737266	0.005894744	0.517361252	0.686673003	0.032104105	0.591229630

Table S5. G-numbers for candidate genes and genome scan test results for outlier SNPs (note * indicates outlier status based on the test listed in the column)

Gene ^a	g-number (GenBank)	Multilocus F _{ST}	SelEstim KLD	Outlier Test		BAYPASS _{aux} Bfmc
				BAYPASS _{core} XtX	BAYPASS _{aux} XtX	
Agly018942	g15475	*	*	*	*	
Agly002329	g2152		*	*	*	
Agly017739	g8271		*	*	*	
Agly-Inc-000004						
Agly000978	g5080		*	*	*	
Agly000981	g5076		*	*	*	
Agly000983	g5074		*	*	*	
Agly-Inc-000191						
Agly-Inc-000554						
Agly010975	g19619		*	*	*	
Agly010068	g4661		*	*	*	
Agly002471	g5632		*	*	*	
Agly002450	g5627		*	*	*	
Agly004871	g5778		*	*	*	
Agly004838	g5779					
Agly001847	g8030		*	*	*	
Agly001843	g8028		*	*	*	*
Agly014114	g8615		*	*	*	
Agly008028	g10321		*	*	*	
	g10321		*	*	*	
Agly-Inc-001286						
Agly-Inc-001286-						

Table S5. Details of the candidate gene outliers that showed constitutive expression and were identified close to SNPs detected in at least three outlier tests. Gene ID from InsectBase2 (<http://v2.insect-genome.com>, accessed on October 14th, 2022). Scaffold ID, log2FoldChange of the gene expression of the Virulent to the Avirulent strain, the log2FoldChange standard error, the p-value of the statistical test and its corresponding adjusted p-value for false discovery rate (FDR) of 10% for the experiment were the aphids were fed with a susceptible or a soybean resistant plant (Rag1/Rag2); how many outliers test the SNPs close (10kbp) to the gene were identified: all the tests, genotype-phenotype association tests, and geographic differentiation tests; The existence of an active transposable element near the gene as identified by Yates-Stewart et al. (2020), the functional gene description, gene ID of a homologous gene of *Drosophila melanogaster*, representative mutation effect as determined by SNPeff (inter-exon corresponds to intros or intergenic mutations), and additional information.

Gene	g-number(GenBank)	Scaffold	Avirulent vs. Virulent fed Susceptible Plant				Avirulent vs. Virulent fed Resistant Plant				Outliers rank test	Biotypes	Geography	Gene near active TE	Description	Homology Dmel	Representative Mutation effect of nearby SNPs	Additional Information
			Log2FC	Log2FC SE	p-value	adj. p-value	Log2FC	Log2FC SE	p-value	adj. p-value								
Agly001892		scaffold_25	5.653451185	0.649548954	3.21E-18	1.51E-16	9.671962253	1.244934476	7.91E-15	2.69E-13	3	1	1	Y	KAF0767941.1 leucine-rich repeat and death domain-containing protein 1-like		synonymous	Maybe related to virulence
Agly002471		scaffold_129	6.724188293	0.559205464	2.64E-33	3.10E-31	9.119548603	1.059861002	7.67E-18	3.14E-16	3	1	1		XP_027847590.1 pre-mRNA-splicing factor CWC22 homolog		inter-exons	Enables RNA binding activity. Involved in mRNA splicing, via spliceosome
Agly002450		scaffold_129	9.052976558	1.051421869	7.29E-18	3.32E-16	8.316174725	0.768691339	2.81E-27	1.94E-25	3	1	1	Y	KAF0756761.1 Tyrosine-protein kinase receptor Tie-2		inter-exons	It mediates cell-to-cell communication and controls many complex biological functions, like cell growth, motility, differentiation, and metabolism.
Agly004871		scaffold_133	7.959423999	1.038594331	1.81E-14	6.23E-13	7.668560523	0.978841305	4.71E-15	1.64E-13	3	1	1		XP_027845315.1 ATPase ASNA1 homolog isoform X2		inter-exons	Maybe a virulence factor in <i>Eimeria tenella</i> .
Agly008028		scaffold_615	5.84532345	0.674399931	4.42E-18	2.05E-16	6.99774875	0.898015878	6.57E-15	2.26E-13	3	1	1		KAE9524253.1 hypothetical protein AGLY_015292		inter-exons	
Agly010935		scaffold_54	6.108866118	0.499441094	2.11E-34	2.66E-32	6.490267339	0.510103645	4.38E-37	4.87E-35	3	1	1		BAH72434.1 ACYP126172		inter-exons	
Agly001847		scaffold_237	6.152129172	1.251872828	8.91E-07	1.13E-05	6.072167964	1.244308849	1.06E-06	1.56E-05	3	1	1		XP_027850017.1 lachesin-like isoform X1		non-synonymous	
Agly001843		Scaffold_237	8.361034442	1.364169698	8.84E-10	1.80E-08	5.596710635	0.961302691	5.81E-09	1.17E-07	4	1	1		XP_027849989.1 lachesin-like		inter-exons	It encodes a cell surface protein that belongs to the Ig superfamily.
Agly017739		scaffold_255	4.14764426	0.962877932	1.65E-05	1.62E-04	5.006502659	0.97228105	2.62E-07	4.15E-06	3	1	1	Y	P-element induced wimpy testis (piwi)	piwi	inter-exons & stop-gained	Piwi has a role in epigenetic regulation, RNA interference, biogenesis of piRNA, and transposable elements silencing.
Agly010068		scaffold_105	6.240534079	1.438154506	1.43E-05	1.42E-04	4.509298833	1.139614995	7.59E-05	7.55E-04	3	1	1		XP_022161814.1 piggyBac transposable element-derived protein 4-like		inter-exons	Maybe associated with transcriptional plasticity.
Agly014114		scaffold_279	3.782168008	0.59550464	2.14E-10	4.68E-09	4.280724379	0.606257614	1.65E-12	4.63E-11	3	1	1		KAF0713924.1 MADF domain-containing protein		inter-exons	A putative transcription factor.
Agly000978		scaffold_114	3.677385195	0.237105217	2.99E-54	8.67E-52	4.012213259	0.242688548	2.15E-61	6.08E-59	3	1	1		XP_027851880.1 histone acetyltransferase KAT2A-like		inter-exons & non-synonymous	Transcriptional activator; repressor of NF-kappa-B, and acetylation.
Agly000981		scaffold_114	1.953009679	0.460611963	2.23E-05	2.13E-04	2.97260414	0.458838932	9.26E-11	2.21E-09	3	1	1	Y	Putative uncharacterized protein		inter-exons	
Agly002329		scaffold_66	2.387713966	0.470758331	3.94E-07	5.32E-06	2.517423139	0.475339495	1.18E-07	1.96E-06	3	1	1		QEE13708.1 cytochrome P450 6CY14		inter-exons	
Agly007635		scaffold_104	1.961161835	0.234800661	6.69E-17	2.82E-15	2.421770325	0.23435651	4.96E-25	3.10E-23	3	1	1		polyhomeotic proximal	Ph-p	inter-exons & non-synonymous	Part of a class of genes involved in detoxification; There is an active Inc upstream of the gene (Agly-Inc-000061).
Agly002452		scaffold_129	1.653379943	0.178420718	1.92E-20	1.06E-18	2.38816916	0.178304769	6.58E-41	8.16E-39	3	1	1		KAF0747863.1 bombyxin C-1-like		inter-exons	Gene products work by forming multiprotein complexes.
Agly003307		scaffold_115	1.841294843	0.226668267	4.54E-16	1.75E-14	2.006175133	0.225686493	6.15E-19	2.66E-17	3	1	1	Y	XP_027847596.1 THAP domain-containing protein 1-like		inter-exons	Secretory brain peptide that belongs to the insulin family.
Agly013824		scaffold_107	1.877769592	0.254818192	1.72E-13	5.31E-12	1.939492825	0.254147321	2.32E-14	7.47E-13	3	1	1		XP_027854375.1 alpha,alpha-trehalose-phosphate synthase [UDP-forming]-like		inter-exons & non-synonymous	
Agly010080		scaffold_105	1.300687946	0.287875377	6.24E-06	6.68E-05	1.763908726	0.282417552	4.22E-10	9.42E-09	3	1	1		XP_026822816.1 headcase protein homolog		inter-exons	
Agly004506		scaffold_72	1.441000831	0.220847016	6.81E-11	1.61E-09	1.675802615	0.217737347	1.40E-14	4.61E-13	3	1	1		XP_027851068.1 ring canal kelch homolog		inter-exons	
Agly016616		scaffold_36	1.880286608	0.343924947	4.57E-08	7.21E-07	1.519200456	0.336813886	6.47E-06	8.13E-05	3	1	1		KAE9544075.1hypothetical protein AGLY_001764		inter-exons	
Agly013930		scaffold_43	1.225380197	0.207386108	3.45E-09	6.34E-08	1.353570311	0.206318051	5.36E-11	1.31E-09	3	1	1		XP_027844233.1 transcription elongation factor SPT5-like		inter-exons	
Agly003289		scaffold_115	1.213353607	0.215285121	1.74E-08	2.93E-07	1.346215812	0.211857541	2.09E-10	4.80E-09	3	1	1		XP_026819812.1 homeobox protein 3-like isoform X2		inter-exons	
Agly005534		scaffold_3	1.160760533	0.221435793	1.59E-07	2.31E-06	1.324374908	0.218417126	1.33E-09	2.81E-08	3	1	1		KAF0766812.1 M-phase inducer phosphatase-like		inter-exons	There is an active Inc that overlaps the predicted gene model (Agly-Inc-001492).
Agly000983		scaffold_114	1.111188153	0.111096376	1.49E-23	1.01E-21	1.296427175	0.110952519	1.53E-31	1.32E-29	3	1	1	Y	XP_025193120.1 protein PFF0380w-like isoform X3		inter-exons	
Agly003640		scaffold_21	1.025157239	0.163526021	3.63E-10	7.79E-09	1.045092361	0.159523469	5.70E-11	1.40E-09	4	1	1		XP_027852418.1 pyruvate kinase-like isoform X2		inter-exons	There is an active Inc downstream the gene model (Agly-Inc-000250).
Agly007427		scaffold_13	0.717049856	0.114967641	4.46E-10	9.50E-09	0.896022713	0.114604113	5.35E-15	1.86E-13	3	1	1		KAF0772752.1 C2H2-type domain-containing protein		inter-exons	Protein is part of a transcription factor.
Agly002907		scaffold_6	0.539249521	0.167424565	1.28E-03	7.38E-03	0.700824795	0.166133245	2.46E-05	2.74E-04	3	1	1		KAF0755760.1zinc finger protein 1 isoform X1		inter-exons	
Agly012730		scaffold_77	0.628938088	0.115695957	5.45E-08	8.53E-07	0.655656051	0.11549858	1.37E-08	2.61E-07	3	1	1		XP_027852532.1 SUMO-activating enzyme subunit 2	Uba2	inter-exons	Gene products work as an E1 ligase for SUMO1 and SUMO2.
Agly010247		scaffold_76	0.550007377	0.14158056	1.02E-04	8.32E-04	0.598676902	0.140577697	2.06E-05	2.33E-04	4	1	1		XP_027846889.1 DNA polymerase eta-like isoform X1		non-synonymous	A Inc overlaps the 3rd exon (Agly-Inc-000717).
Agly004695		scaffold_95	1.008043214	0.181472275	2.78E-08	4.53E-07	0.597721965	0.180063738	9.02E-04	6.60E-03	3	1	1		VVC28026.1 MADF domain		inter-exons	
Agly014174		scaffold_109	0.458839199	0.127384063	3.16E-04	2.21E-03	0.548121272	0.126923974	1.57E-05	1.84E-04	3	1	1		XP_027841106.1 RNA helicase aquarius		inter-exons	
Agly005403		scaffold_3	0.456436141	0.095402533	1.72E-06	2.08E-05	0.531245487	0.094903697	2.17E-08	3.99E-07	3	1	1		XP_027851556.1 N-alpha-acetyltransferase 25, NatB auxiliary subunit		non-synonymous	
Agly014841		scaffold_10	0.558517252	0.142839604	9.23E-05	7.61E-04	0.495324812	0.139125564	3.70E-04	3.08E-03	3	1	1		XP_026814506.1 cubilin-like		inter-exons	
Agly013283		scaffold_1	0.638360774	0.088809313	6.58E-13	1.95E-11	0.421755682	0.087704208	1.52E-06	2.16E-05	3	1	1		XP_027851706.1 nuclear valosin-containing protein-like		inter-exons	
Agly003639		scaffold_21	0.306090914	0.086381994	3.95E-04	2.68E-03	0.312701616	0.086219875	2.87E-04	2.47E-03	3	1	1	Y	XP_026815765.1 acylamino-acid-releasing enzyme-like isoform X3		inter-exons	
Agly000889		scaffold_20	-0.285410534	0.08790884	1.17E-03	6.84E-03	-0.282030365	0.087774658	1.31E-03	9.07E-03	3	1	1		XP_026806799.1 microtubule-associated protein Jupiter-like isoform X3		inter-exons	
Agly004598		scaffold_149	-0.492480636	0.084709982	6.11E-09	1.09E-07	-0.360454671	0.08466879	2.07E-05	2.34E-04	3	1	1		XP_027836313.1 V-type proton ATPase catalytic subunit A		non-synonymous	
Agly004271		scaffold_37	-0.393902224	0.11793225	8.38E-04	5.15E-03	-0.378853941	0.117168177	1.22E-03	8.54E-03	3	1	1		KAF0758803.1 guanylate cyclase soluble subunit beta-1		inter-exons	
Agly003698		scaffold_21	-0.555268334	0.098983532	2.03E-08	3.39E-07	-0.399479686	0.098718593	5.20E-05	5.37E-04	4	1	1		XP_027844661.1 guanylate cyclase 32E isoform X1		inter-exons	
Agly013226		scaffold_1	-0.680149977	0.122307105	2.68E-08	4.39E-07	-0.420502819	0.122027863	5.69E-04	4.41E-03	3	1	1		XP_027843921.1 WD repeat-containing protein 47 isoform X2		inter-exons	
Agly004730		scaffold_95	-0.444274956	0.106969269	3.28E-05	3.00E-04	-0.431323168	0.106898678	5.46E-05	5.61E-04	3	1	1		XP_015374974.1 4-aminobutyrate aminotransferase, mitochondrial		inter-exons	

Agly013599	scaffold_40	-0.554763003	0.104337335	1.05E-07	1.58E-06	-0.470269148	0.104119153	6.28E-06	7.95E-05	3	1	1	XP_027848998.1 synaptic vesicle membrane protein VAT-1 homolog-like	inter-exons	
Agly009609	scaffold_14	-0.537953569	0.135958512	7.60E-05	6.40E-04	-0.471378772	0.135672699	5.12E-04	4.03E-03	3	1	1	Putative uncharacterized protein	non-synonymous	
Agly003988	scaffold_126	-0.535698924	0.109628031	1.03E-06	1.29E-05	-0.502541301	0.109252547	4.23E-06	5.52E-05	3	1	1	XP_027847443.1 LON peptidase N-terminal domain and RING finger protein 3 isoform X2	inter-exons	
Agly014753	scaffold_10	-1.145232808	0.126876546	1.77E-19	9.34E-18	-0.573554189	0.126473793	5.76E-06	7.34E-05	3	1	1	XP_027844768.1 NADPH oxidase 5	inter-exons	
Agly007649	scaffold_104	-0.856223118	0.13399035	1.66E-10	3.69E-09	-0.608293676	0.133779246	5.44E-06	6.98E-05	3	1	1	XP_027849324.1 scavenger receptor class B member 1	inter-exons	
Agly000179	scaffold_5	-0.528585926	0.092371313	1.05E-08	1.83E-07	-0.671986671	0.092058994	2.89E-13	8.57E-12	4	1	1	XP_027846745.1 proton-coupled amino acid transporter-like protein CG1139	inter-exons	
Agly002685	scaffold_174	-0.896481747	0.198785421	6.49E-06	6.92E-05	-0.693748959	0.196967677	4.28E-04	3.48E-03	4	1	1	VVC45035.1BTB/POZ domain,Potassium channel tetramerisation-type BTB domain,SKP1/BTB/POZ domain	inter-exons	
Agly004010	scaffold_126	-0.583230315	0.166830507	4.72E-04	3.12E-03	-0.693948384	0.164872817	2.57E-05	2.84E-04	3	1	1	XP_027847480.1 myb-like protein AA isoform X2	inter-exons	
Agly004265	scaffold_37	-1.228330532	0.162304653	3.79E-14	1.26E-12	-0.733105001	0.161467929	5.62E-06	7.18E-05	3	1	1	XP_026806316.1 glutaredoxin domain-containing cysteine-rich protein CG12206-like	inter-exons	
Agly004243	scaffold_37	-1.005272848	0.21900192	4.43E-06	4.92E-05	-0.75816999	0.218881303	5.33E-04	4.17E-03	3	1	1	XP_026806712.1 LIM domain-containing protein A	inter-exons	
Agly015126	scaffold_7	-0.681293674	0.13500353	4.50E-07	6.01E-06	-0.779982584	0.134824054	7.24E-09	1.44E-07	3	1	1	XP_008186802.1 CKLF-like MARVEL transmembrane domain-containing protein	inter-exons	
Agly014923	scaffold_7	-0.631621569	0.171875268	2.38E-04	1.72E-03	-0.922667356	0.172399121	8.70E-08	1.47E-06	4	1	1	XP_027851983.1 15-hydroxyprostaglandin dehydrogenase [NAD(+)]-like	inter-exons	
Agly013673	scaffold_40	-0.95677479	0.261466842	2.53E-04	1.81E-03	-0.95790391	0.260149051	2.31E-04	2.04E-03	3	1	1	XP_027851699.1 Usher syndrome type-1G protein homolog	inter-exons	
Agly010728	scaffold_9	-1.723726455	0.220793746	5.86E-15	2.10E-13	-0.961748425	0.220505955	1.29E-05	1.53E-04	3	1	1	VVC26798.1 PAN/Apple domain,Zona pellucida domain; Cuticlin-4	inter-exons	
Agly014088	scaffold_98	-1.89910799	0.296249345	1.45E-10	3.25E-09	-1.013779615	0.290684747	4.87E-04	3.88E-03	3	1	1	XP_026807280.1 extensin-like	inter-exons	
Agly005366	scaffold_3	-1.4113282	0.27361117	2.49E-07	3.47E-06	-1.044125893	0.272750629	1.29E-04	1.21E-03	5	1	1	Argonaute 3 (Ago3); serine proteinase stubble	inter-exons	AGO3
Agly015146	scaffold_7	-1.676331624	0.35333417	2.09E-06	2.50E-05	-1.096433085	0.333269777	1.00E-03	0.007218	3	1	1	KAE9526816.1 hypothetical protein AGLY_013464	inter-exons	
Agly004009	scaffold_126	-1.701367646	0.21940463	8.87E-15	3.13E-13	-1.100181894	0.219133919	5.15E-07	7.89E-06	3	1	1	Putative uncharacterized protein	inter-exons	
Agly014195	scaffold_109	-1.638116513	0.192376885	1.66E-17	7.34E-16	-1.346954299	0.191130968	1.82E-12	5.10E-11	3	1	1	XP_027840155.1 uncharacterized protein LOC114121871	inter-exons	
Agly010442	scaffold_42	-2.480483073	0.393457122	2.89E-10	6.27E-09	-1.364481686	0.390143798	4.70E-04	3.76E-03	3	1	1	XP_027854269.1 putative fatty acyl-CoA reductase CG5065	inter-exons	
Agly014767	scaffold_10	-2.199551948	0.313826221	2.40E-12	6.69E-11	-1.372503269	0.313781169	1.22E-05	1.46E-04	3	1	1	XP_027848136.1 protein obstructor-E isoform X1	inter-exons	Y
Agly015108	scaffold_7	-2.164621952	0.243436495	6.01E-19	2.99E-17	-1.410264807	0.242257229	5.84E-09	1.17E-07	3	1	1	XP_018909939.1 DEAD-box ATP-dependent RNA helicase rde-12-like	inter-exons	Y
Agly003719	scaffold_21	-1.364776967	0.257992148	1.22E-07	1.81E-06	-1.43623269	0.266229951	6.86E-08	1.18E-06	4	1	1	XP_026815778.1 membrane-bound alkaline phosphatase-like	inter-exons	
Agly015081	scaffold_7	-2.44553631	0.453624052	7.00E-08	1.08E-06	-1.527480064	0.452029647	7.27E-04	5.50E-03	4	1	1	KAE9526839.1 hypothetical protein AGLY_013487	inter-exons	
Agly004171	scaffold_37	-1.245257704	0.202313264	7.50E-10	1.54E-08	-1.705499047	0.203183797	4.70E-17	1.85E-15	3	1	1	XP_027840913.1 cytoplasmic polyadenylation element-binding protein	inter-exons	
Agly015057	scaffold_7	-1.852297452	0.467215207	7.35E-05	6.22E-04	-2.200703314	0.473766222	3.40E-06	4.50E-05	4	1	1	KAE9527004.1 hypothetical protein AGLY_013652	inter-exons	

AGO3

It is a repressor of transposable elements.

This gene has a Ya gene upstream that was detected by three outlier test (called Agly-Ya4); Ya4 might be a lnc-RNA (lnc-001198)

Table S7-lncRNA

Table S6. Details of long non-coding RNAs that showed constitutive expression (positive or negative log2FC regardless of host-plant the aphids were exposed to). LncRNA ID from InsectBase2 (<http://v2.insect-genome.com>, accessed on October 14th, 2022). Scaffold ID, log2FoldChange of the lncRNA expression of the Virulent to the Avirulent strain, the log2FoldChange standard error, the p-value of the statistical test and its corresponding adjusted p-value for false discovery rate (FDR) of 5% for the experiment where the aphids were fed with a susceptible or a soybean resistant plant (Rag1/Rag2).

lncRNA	Scaffold	Avirulent vs. Virulent fed Susceptible Plant				Avirulent vs. Virulent fed Resistant Plant			
		Log2FC	Log2FC SE	p-value	adj. p-value	Log2FC	Log2FC SE	p-value	adj. p-value
Agly-lnc-000167	Scaffold_202	11.61665228	0.518536639	3.70E-111	4.33E-108	12.02149675	0.43746311	3.04E-166	3.58E-163
Agly-lnc-002745	Scaffold_225	10.28035386	1.09919743	8.55E-21	9.89E-18	11.25117293	1.052896103	1.18E-26	1.37E-23
Agly-lnc-001218	Scaffold_30	10.47630781	1.102536035	2.06E-21	2.39E-18	10.7923649	1.062447676	3.05E-24	3.51E-21
Agly-lnc-000554	Scaffold_114	8.631559166	1.106218366	6.06E-15	6.87E-12	9.113006005	1.07518291	2.34E-17	2.64E-14
Agly-lnc-000013	Scaffold_118	7.151296621	1.268746717	1.74E-08	1.90E-05	8.332761622	1.245724781	2.25E-11	2.48E-08
Agly-lnc-002163	Scaffold_55	8.480899399	1.269065652	2.34E-11	2.63E-08	7.950034469	1.245594608	1.74E-10	1.92E-07
Agly-lnc-000482_3	Scaffold_383	6.727446538	0.718169802	7.43E-21	8.60E-18	7.747451208	0.881608109	1.52E-18	1.74E-15
Agly-lnc-000482_2	Scaffold_383	6.846709146	0.764590954	3.41E-19	3.93E-16	7.593187175	0.829746001	5.63E-20	6.43E-17
Agly-lnc-001735	Scaffold_89	6.638640814	0.955248732	3.66E-12	4.11E-09	7.562782646	1.086015913	3.31E-12	3.68E-09
Agly-lnc-000109	Scaffold_41	5.666451202	0.452720226	6.07E-36	7.08E-33	7.218735518	0.318727848	1.44E-113	1.69E-110
Agly-lnc-000392	Scaffold_117	6.146257608	0.677622517	1.19E-19	1.37E-16	7.166119255	0.731372957	1.15E-22	1.31E-19
Agly-lnc-000085	Scaffold_63	5.43509929	0.521527274	1.98E-25	2.30E-22	6.441564254	0.459568866	1.23E-44	1.44E-41
Agly-lnc-001966	Scaffold_24	6.875228082	1.331230046	2.41E-07	0.000262443	6.426610373	1.270011258	4.19E-07	0.000446694
Agly-lnc-002119	Scaffold_177	6.113036073	0.40521975	2.01E-51	2.35E-48	6.373418094	0.280129588	1.38E-114	1.62E-111
Agly-lnc-000191	Scaffold_114	6.443097907	1.360966732	2.20E-06	0.002364	6.003839817	1.314376004	4.93E-06	0.005169585
Agly-lnc-000482	Scaffold_383	5.303328976	0.759069851	2.82E-12	3.17E-09	5.995374415	0.870060386	5.55E-12	6.16E-09
Agly-lnc-001286_2	Scaffold_615	5.829837237	0.540182234	3.74E-27	4.35E-24	5.899553231	0.491335188	3.26E-33	3.78E-30
Agly-lnc-001103	Scaffold_112	6.118349738	1.365311373	7.42E-06	0.007909172	5.874068148	1.298753178	6.10E-06	0.006381851
Agly-lnc-001815	Scaffold_276	6.071717001	1.139187582	9.83E-08	0.000107622	5.859763163	1.151240082	3.58E-07	0.000382856
Agly-lnc-000630	Scaffold_694	5.294225148	0.860964165	7.79E-10	8.64E-07	5.821564376	0.801333167	3.73E-13	4.18E-10
Agly-lnc-001286	Scaffold_615	5.815825432	0.521879071	7.66E-29	8.92E-26	5.773508392	0.477208839	1.08E-33	1.25E-30
Agly-lnc-001162	Scaffold_149	4.769160681	0.915104745	1.87E-07	0.000204442	5.663116312	0.924259743	8.94E-10	9.81E-07
Agly-lnc-002075	Scaffold_134	4.5588796	0.498390873	5.84E-20	6.75E-17	5.513895432	0.450763116	2.09E-34	2.43E-31
Agly-lnc-002059	Scaffold_186	6.823282584	0.908807792	6.01E-14	6.80E-11	5.422579067	1.045087457	2.12E-07	0.000227304
Agly-lnc-000769	Scaffold_57	4.471241607	0.631326354	1.42E-12	1.60E-09	5.382460063	0.54648424	6.91E-23	7.92E-20
Agly-lnc-001836	Scaffold_387	4.959849568	0.806232242	7.66E-10	8.50E-07	5.273890975	0.682986058	1.15E-14	1.29E-11
Agly-lnc-000041	Scaffold_306	3.66728769	0.418757577	2.00E-18	2.30E-15	5.091338545	0.269988799	2.55E-79	2.98E-76
Agly-lnc-001207	Scaffold_160	4.810346986	0.657310024	2.51E-13	2.84E-10	5.0593587	0.575178915	1.42E-18	1.62E-15
Agly-lnc-000737	Scaffold_518	4.727622111	0.458939216	6.96E-25	8.08E-22	5.027013748	0.351624391	2.30E-46	2.68E-43
Agly-lnc-000286_2	Scaffold_467	3.932336138	0.553088668	1.16E-12	1.31E-09	5.022951248	0.672732841	8.24E-14	9.24E-11
Agly-lnc-001155	Scaffold_142	4.787009328	0.548881099	2.75E-18	3.17E-15	4.957544494	0.482420121	9.00E-25	1.04E-21
Agly-lnc-000632	Scaffold_44	4.589974055	0.569231486	7.42E-16	8.45E-13	4.9398657	0.542285587	8.29E-20	9.47E-17
Agly-lnc-000406	Scaffold_518	4.715806247	0.459938571	1.15E-24	1.33E-21	4.920943499	0.353945156	6.06E-44	7.06E-41
Agly-lnc-000327	Scaffold_75	4.824683129	0.602990039	1.23E-15	1.40E-12	4.851682282	0.470241081	5.88E-25	6.77E-22
Agly-lnc-001708	Scaffold_34	4.773904466	0.568131067	4.36E-17	5.00E-14	4.834801619	0.453276077	1.46E-26	1.69E-23
Agly-lnc-002160	Scaffold_740	4.742735766	0.614623574	1.20E-14	1.36E-11	4.812559055	0.479024093	9.51E-24	1.09E-20
Agly-lnc-000286	Scaffold_467	3.392547087	0.50445393	1.75E-11	1.97E-08	4.72795808	0.543110262	3.17E-18	3.61E-15
Agly-lnc-000102	Scaffold_276	4.773663642	0.440404923	2.24E-27	2.61E-24	4.673905785	0.319348717	1.66E-48	1.94E-45
Agly-lnc-000286_3	Scaffold_467	2.991959967	0.500735966	2.30E-09	2.55E-06	4.516147921	0.584130983	1.06E-14	1.20E-11

Table S7-lncRNA

Agly-lnc-000286_4	Scaffold_467	3.272841031	0.51608972	2.27E-10	2.54E-07	4.511420057	0.755793383	2.39E-09	2.61E-06
Agly-lnc-001391	Scaffold_57	4.257928147	0.817614274	1.91E-07	0.000208351	4.44453861	0.699861492	2.14E-10	2.36E-07
Agly-lnc-000390	Scaffold_467	4.265412966	0.532601708	1.16E-15	1.32E-12	4.427744147	0.425668439	2.43E-25	2.80E-22
Agly-lnc-000375	Scaffold_294	3.21790637	0.542165922	2.93E-09	3.24E-06	4.149921341	0.51045275	4.30E-16	4.85E-13
Agly-lnc-000004	Scaffold_255	3.850037484	0.467615431	1.82E-16	2.08E-13	4.120725584	0.362469201	6.00E-30	6.95E-27
Agly-lnc-000014_5	Scaffold_221	3.465127955	0.424476818	3.26E-16	3.72E-13	4.075085065	0.263625473	6.68E-54	7.82E-51
Agly-lnc-001663	Scaffold_170	4.06682522	0.50271407	5.98E-16	6.82E-13	4.066680558	0.539705598	4.88E-14	5.49E-11
Agly-lnc-000014_3	Scaffold_221	3.645229101	0.424390433	8.75E-18	1.01E-14	4.048753225	0.273660057	1.58E-49	1.85E-46
Agly-lnc-000014_4	Scaffold_221	3.461161351	0.410355708	3.33E-17	3.82E-14	3.98265578	0.259760057	4.67E-53	5.46E-50
Agly-lnc-000376	Scaffold_454	3.833819248	0.485349046	2.81E-15	3.19E-12	3.950848871	0.37868264	1.75E-25	2.02E-22
Agly-lnc-000070	Scaffold_511	3.380082075	0.451209846	6.83E-14	7.72E-11	3.924593298	0.388009233	4.76E-24	5.46E-21
Agly-lnc-000014_2	Scaffold_221	3.410009399	0.402486502	2.40E-17	2.76E-14	3.834799505	0.317420069	1.33E-33	1.54E-30
Agly-lnc-000014	Scaffold_221	3.387902442	0.405213976	6.23E-17	7.14E-14	3.829168315	0.253836015	2.03E-51	2.37E-48
Agly-lnc-000257	Scaffold_55	2.682504367	0.603111898	8.68E-06	0.009223	3.814738725	0.652293119	4.97E-09	5.42E-06
Agly-lnc-000008	Scaffold_94	3.455579955	0.778743468	9.11E-06	0.009662	3.804943608	0.670729335	1.40E-08	1.52E-05
Agly-lnc-000808	Scaffold_94	3.279268334	0.454202075	5.20E-13	5.88E-10	3.574837721	0.326627687	7.05E-28	8.15E-25
Agly-lnc-000629	Scaffold_540	2.892249645	0.489946293	3.57E-09	3.93E-06	3.077850353	0.362395324	2.01E-17	2.28E-14
Agly-lnc-000891	Scaffold_540	2.896196977	0.49055283	3.55E-09	3.92E-06	3.077850353	0.362395324	2.01E-17	2.28E-14
Agly-lnc-001153	Scaffold_540	2.886759766	0.491720035	4.34E-09	4.78E-06	3.066754162	0.364092904	3.67E-17	4.15E-14
Agly-lnc-000970	Scaffold_84	2.463817935	0.525309276	2.73E-06	0.002922831	2.97888759	0.414110968	6.32E-13	7.07E-10
Agly-lnc-000970_2	Scaffold_84	2.413611478	0.516056265	2.91E-06	0.003114314	2.885346454	0.402719093	7.80E-13	8.72E-10
Agly-lnc-000066	Scaffold_617	2.382618617	0.527547564	6.29E-06	0.006711	2.854877592	0.419472048	1.00E-11	1.11E-08
Agly-lnc-000340	Scaffold_617	2.427333726	0.517111407	2.68E-06	0.002871691	2.854838231	0.401175161	1.11E-12	1.24E-09
Agly-lnc-001065	Scaffold_617	2.427333726	0.517111407	2.68E-06	0.002871691	2.854838231	0.401175161	1.11E-12	1.24E-09
Agly-lnc-000418	Scaffold_104	2.808466836	0.433340858	9.11E-11	1.02E-07	2.806650782	0.327459985	1.03E-17	1.17E-14
Agly-lnc-001062	Scaffold_142	3.008080256	0.587513582	3.06E-07	0.000332391	2.801879019	0.555916135	4.65E-07	0.000495969
Agly-lnc-000970_3	Scaffold_84	2.449315085	0.510492566	1.60E-06	0.001729659	2.696810205	0.340700168	2.46E-15	2.78E-12
Agly-lnc-001876	Scaffold_617	3.169274317	0.659336884	1.53E-06	0.001656674	2.648780234	0.46333642	1.09E-08	1.18E-05
Agly-lnc-000672	Scaffold_276	2.127576713	0.444415229	1.69E-06	0.001819996	2.229163649	0.364125877	9.24E-10	1.01E-06
Agly-lnc-001342	Scaffold_83	1.933038851	0.390977859	7.65E-07	0.0008284	2.178020462	0.256098506	1.82E-17	2.07E-14
Agly-lnc-002664	Scaffold_166	2.631113189	0.472592353	2.59E-08	2.83E-05	2.120032367	0.404672028	1.62E-07	0.000173984
Agly-lnc-000456	Scaffold_83	2.42285692	0.390834598	5.68E-10	6.31E-07	2.064845662	0.291930367	1.52E-12	1.69E-09
Agly-lnc-002398	Scaffold_139	-1.95479589	0.412452848	2.14E-06	0.002305962	-1.49420718	0.304339848	9.12E-07	0.000967053

Table S7. SNP Genotypes from 2024 soybean aphids collected from the field. Note, only 1 SNP GeneID was used in population genetic analyses due to linkage disequilibrium. Heterozygotes are represented by IUPAC ambiguity codes. *+: AAG insertion present; -: AAG insertion absent.

Soybean aphids avirulent to Rag1/Rag2/Rag3												Soybean aphids virulent to Rag1/Rag2/Rag3						Laboratory Colonies		
GeneID	Scaffold	SNP	OH 1-AV	Woo 2-AV	Woo 5-AV	Woo 13-AV	Woo 25-AV	Woo 59-AV	SD 9-AV	SD 15-AV	SD 24-AV	Woo 27-V	Woo 29-V	Woo 34-V	Woo 42-V	Woo 49-V	Woo 52-V	B1-LAB	B4-LAB	3R Virulent LAB
Agly018942	25	[T/C]	C	T	Y	Y	C	Y	T	T		Y	Y	Y	C	T	Y	T	T	T
Agly010975	54	[T/G]	T	K	K	K	G	K	K	T	T	G	T	K	K	T	T	K	K	G
		[G/A]	G	R	R	R	A	R	R	G	G	A	G	R	R	G	G	R	R	A
		[G/A]	G	R	R	R	A	R	R	G	G	A	G	R	R	G	G	R	R	A
Agly002329	66	[C/T]	Y	T	T	T	T	C	T	Y	Y	Y	T	Y	Y	T	T	T	T	T
		[G/A]	R	A	A	A	A	G	A	R	R	R	A	R	R	A	A	A	A	A
		[C/A]	M	A	A	A	A	C	A	M	M	M	A	M	M	A	A	A	A	A
Agly010068	105	[A/T]	W	A	T	A	A	A	W	A	W	A	T	W	T	W	A	T	T	A
		[A/T]	W	A	T	A	A	A	W	A	W	A	T	W	T	W	A	T	T	A
		[A/T]	W	A	T	A	A	A	W	A	W	A	T	W	T	W	A	T	T	A
Agly000978	114	[A/G]	A	A	R	R	A	G	A	A	R	A	G	R	R	R	R	R	R	R
Agly000981	114	[T/A]	T	T	A	W	T	A	T	T	W	T	A	W	W	W	W	A	A	W
Agly-lnc-000191	114	[A/C]	C	C	M	M	C	A	C	C	M	C	A	M	M	M	M	A	A	M
		[T/A]	A	A	W	W	A	T	A	A	W	A	T	W	W	W	W	T	T	W
		[A/C]	C	C	M	M	C	A	C	C	M	C	A	M	M	M	M	A	A	M
Agly-lnc-000554	114	[T/A]	A	A	W	W	A	T	A	A	W	A	T	W	W	W	W	T	T	W
		[C/T]	T	T	Y	Y	T	C	T	T	Y	T	C	Y	Y	Y	Y	C	C	Y
		[C/G]	G	G	S	S	G	C	G	G	S	G	C	S	S	S	S	C	C	S
Agly002471	129	[A/G]	A	R	R	R	G	R	R	A	A	G	G	G	R	A	A	R	R	R
		[C/G]	C	S	S	S	G	S	S	C	C	G	G	G	S	C	C	S	S	S
		[C/T]	C	Y	Y	Y	T	Y	Y	C	C	T	T	T	Y	C	C	Y	Y	Y
Agly002450	129	[C/A]	C	M	M	M	A	M	M	C	C	A	A	A	M	C	C	M	M	M
Agly004871 & Agly004838	133	[A/T]	A	null	W	W	T	A	W	A	A	T	T	T	T	W	A	W	W	T
	237	[T/C]	T	Y	Y	Y	C	T	Y	T	T	C	C	C	C	Y	T	Y	Y	C
		[T/C]	C	Y	Y	Y	T	Y	Y	C	C	T	T	T	T	Y	C	Y	Y	Y
		[C/A]	A	M	M	M	C	M	M	A	A	C	C	C	C	M	A	M	M	M
Agly001847	237	[A/T]	T	W	W	W	A	W	W	T	T	A	A	A	A	W	T	W	W	W
Agly001843	237	[C/T]	T	Y	Y	Y	C	Y	Y	T	T	C	C	C	C	Y	T	Y	Y	Y
		[G/T]	G	K	K	K	T	K	K	G	G	T	T	T	T	K	G	K	K	K
		[C/T]	Y	C	Y	C	C	Y	T	C	Y	C	T	Y	T	Y	C	T	T	Y
Agly014114 & Agly014119	279	[AAG]	+ -	AAG	+ -	AAG	AAG	+ -	-	AAG	+ -	AAG	-	+ -	-	+ -	AAG	-	-	+ -
Agly008028	615	[C/T]	Y	C	Y	C	C	Y	T	C	Y	C	T	Y	T	Y	C	T	T	Y
		[A/G]	A	R	R	R	R	R	R	A	A	R	R	R	R	null	A	R	R	R