California State University, Fresno / Department of Plant Science

Combining large genetic datasets with small sample sizes to identify adaptive evolution in a pest species

Jacob Wenger[1*] & Andy Michel[2]

1. Department of Plant Science - California State University, Fresno. 2. Department of Entomology - The Ohio State University OARDC * Contact author: jawenger@csufresno.edu - 559-278-7754

INTRODUCTION

Genome Wide Association Studies (GWAS)

GWAS are used to identify gene alleles associated with adaptive traits. GWAS compares the allele frequency of 10's or 100's of genetic markers between phenotypically distinct populations, to identify variants that are associated with the trait of interest. To have high confidence in these associations a large number of individuals (40+) must be sampled and sequenced for each phenotype/population. Large sample sizes can be difficult to collect when the studied trait is rare or difficult to measure (resistance for example) and expensive to sequence. Recent models suggest that high confidence results can be produced from small sample sizes if a large number of markers are sequenced (100's or 1,000's) per sample (Willing et al. 2012, Flagel et al. 2014) .

Here we use a high-density GWAS approach with small populations to identify alleles associated with virulence in the soybean aphid.

Soybean aphid is a significant pest of soybean in the northern Midwest. Aphid resistant soybean varieties expressing *Rag* (resistance to *Aphis glycines)* genes were bred Virulent biotypes capable of overcoming the plant's defenses evolved quickly for both commercialized *Rag* genes (*Rag1* & *Rag2*) (Table 1). Virulence is influenced by multiple interacting genes, environmental conditions (Hough *et al* 2016), and obviation by other aphids (Varenhorst et al 2015). As such, it's extremely difficult to gather a large sample of a single biotype.

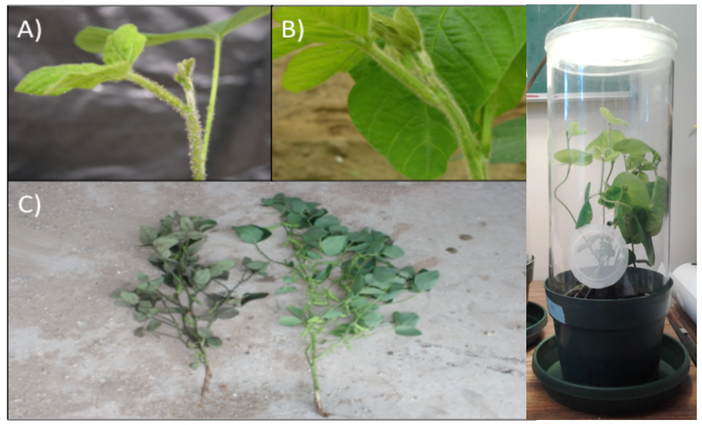

Figure 1: Comparative aphid population growth on A) HPR and B) susceptible soybean. C) Susceptible (left) and HPR (right) soybean after aphid exposure. D) Live plant assay in the lab. Photo: K. Tilmon, 2011.

Table 1: Relative virulence of biotypes compared in this research

	Aphid response to resistance gene		
Biotype	*Rag1*	*Rag2*	*Rag1 & Rag2*
1	avirulent	avirulent	avirulent
4	**Virulent**	**Virulent**	**Virulent**

MATERIALS AND METHODS

Collection

Aphids were collected on susceptible (Dowling) and pyramid resistant plants (*Rag1/Rag2*) in 8 states across the Midwest (Figure 2). A live plant assay was used to identify the collected aphids to biotype (Figure 1D). DNA was extracted from aphids belonging to biotype 1 (fully avirulent) and biotype 4 (fully virulent) (Table 1).

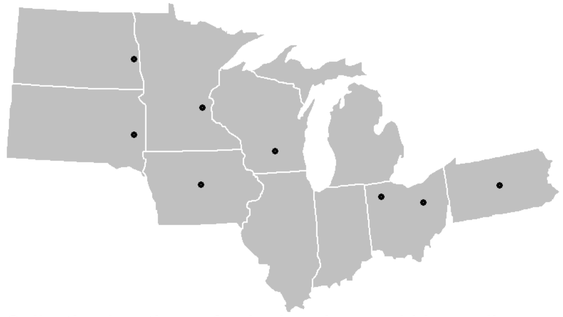

Figure 2: Collection locations of paired soybean aphid sampling

DNA pooling and analysis

The DNA from five aphids from each biotype (1 and 4) and geographic location combination was pooled and then sequenced on an Illumina MiSeq (300bp paired-end reads). Sequence reads were cleaned/assembled using CLC Genomics Workbench. Then Popoolation2 was used to isolate single nucleotide polymorphisms (SNP) markers, calculate pool-adjusted allele frequencies and F_{ST} (a measure of genetic differentiation between populations) values for each SNP marker.

All SNPs with F_{ST} values >5x the mean were considered outliers, and were treated as candidate markers associated with biotypic virulence. The statistical significance of candidate markers was calculated via population randomization tests and Cochran-Mantel-Haenszel repeated tests of independence. BLAST2go was used to identify genes within 2000bp of candidate markers. Candidate markers are currently being verified via Sanger resequencing of field collected aphids

RESULTS

Outlier analysis

- F_{ST} analysis identified 3,007 outlier SNPs potentially associated with virulence
- Reduced to 84 SNPs after permutation and independence tests

Spatial and Functional Clustering of Outliers

- 84 SNPs grouped spatially within the genome into 59 'clusters'
  - 500-5000bp regions of genome
- 22 spatial clusters were within 2,000bp of a gene with known function
  - Most of these clustered within 3 major functional categories (Table 2)
  - 4 clusters - Transposable Elements
  - 8 clusters - Transcription factors / Epigenetic regulation
  - 5 clusters - Muscle structure

Field verification

- Completed resequencing has identified biotype specific sequence variants within Piggy-back transposons (Figure 3)
- Work is ongoing, and more sequencing is required to verify all functional clusters

Table 2: Primary gene function clusters of SNP outliers. Note: Outliers that produce non-synonymous mutations are marked with an X

SNP Cluster	Non-synonymous Mutation[X]	Associated Gene	Functional Annotation Groups
Transposable Element Clusters			
10	X	A. pisum piggybac transposable element protein 4	Transposable Element
11	X	A. gossypii transposon piggybac-like element	Transposable Element
15		A. pisum piggyBac transposable element protein 4	Transposable Element
16		A. pisum piggyBac transposable element protein 4	Transposable Element
Transcriptional Elements/Epigenetic Modifiers			
9		A. pisum craniofacial development protein 2-like	transcription regulation, histone modification
5		A. pisum histone H2A mRNA	histone subunit, DNA binding, nucleus, chromosome
17	X	A. pisum f-box-like wd protein ebi mrna	Gene expression regulation, histone/DNA binding
58		A. gossypii glyceraldehyde-3-phosphate dehydrog.	glycolysis, transcription/translation regulation
52		A. pisum probable nuclear hormone receptor hr38	Cuticular development, transcription factor
39		A. pisum homeobox protein hox-a1a-like mrna	Sequence specific DNA binding, transcrition reguation
18	X	A. pisum zinc finger mym-type protein 1-like	DNA/Protein Binding, gene expression, recombination
54	X	A. pisum myb-like protein aa mrna	Transcription factor, sequence specific binding
Muscle Structure			
46		A. pisum translation factor guf1 mitochondria	microtubule-based movement, axonemal dynein complex
47	X	A. pisum dynein heavy chain axonemal-like mrna	microtubule coding, ATP binding, metabolic processing
49		A. pisum myosin-I heavy chain-like transferase	Muscular structure, phosphorylation of myosin
50		A. pisum myosin-I heavy chain-like mRNA	Muscular structure, phosphyllation of myosin
55		A. pisum coiled-coil domain-containing protein	Functionally diverse, specific gene uncharacterized
Misc. Physiological			
14		A. pisum protocadherin-8 mrna	calcium ion binding, cell adhesion, embryo dev.
45		A. pisum glutamate receptor 1-like mrna	Glutamate receptor, neuron excitement

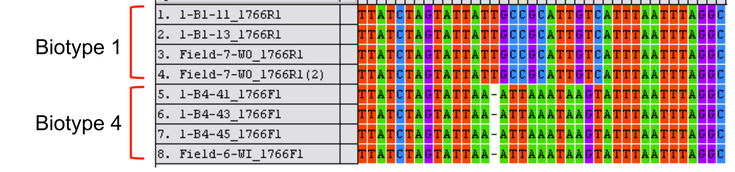

Figure 3: Biotype specific sequence variants found in aphid PiggyBac transposon

CONCLUSIONS

Soybean aphid

- Multiple spatial and functional sites of divergence suggest that virulence is a complex phenotype governed by multiple genes.
- A large proportion of outlier SNPs associated with genetic mechanisms that can be induced during periods of environmental stress.
  - Transposable elements, transcription factors, and epigenetic modifiers
  - Suggests virulence may be the product of changes in gene regulation induced by environmental stress.
- Resequencing has found outliers exist as verifiable genetic variants in the field.
  - Further work is required to isolate true variants from false positives.

Further applications

- Virulence associated genes were successfully identified using population samples as small as 5 individuals.
- Small sample sizes could be used to identify adaptive alleles in wild populations
  - Resistance, vector efficiency, dispersal, host preference, etc.

WORKS CITED

Flagel, L. E., R. Bansal, R. A. Kerstetter, M. Chen, M. Carroll, R. Flannagan, T. Clark, B. S. Goldman, and A. P. Michel. 2014. Western corn rootworm (Diabrotica virgifera virgifera) transcriptome assembly and genomic analysis of population structure. BMC Genomics. 15: 195.

Hough, A. R., Nechols, J., McCornack, B., Margolies, D., Sandercock, B., Donglin, Y., Murray, L. 2016. The Effect of Temperature and Host Plant Resistance on Population Growth of the Soybean Aphid Biotype 1 (Hemiptera: Aphididae). Environ Entomol 46 (1): 58-67.

Varenhorst, A. J., M. T. McCarville, and M. E. O'Neal. 2015. An Induced Susceptibility Response in Soybean Promotes Avirulent Aphis glycines (Hemiptera: Aphididae) Populations on Resistant Soybean. Environ. Entomol. nvv051.

Willing, E.-M., C. Dreyer, and C. van Oosterhout. 2012. Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. PLoS ONE. 7: e42649.

FRESNO STATE
Jordan College of Agricultural Sciences and Technology

The Ohio State University
College of Food, Agricultural, and Environmental Sciences

NCSRP North Central Soybean Research Program
SOY OHIO Ohio Soybean Council www.soyohio.org
NSF