Origins of *Amaranthus tuberculatus* (Waterhemp) in Central Valley Agroecosystems: A Population Genetics Approach Using Genotyping-By Sequencing

Alexander J. Lopez*[1], Dr. Anil Shrestha[2], Dr. Lynn M. Sysnoskie[3], Dr. Katherine E. Waselkov[1]

[1]Department of Biology, College of Science and Mathematics, California State University Fresno, Fresno, CA, USA, [2]Department of Viticulture and Enology, College of Science and Mathematics, California State University Fresno, Fresno, CA, USA, [3]School of Integrative Plant Science, College of agriculture and Life sciences, Cornell University, Ithaca, New York, USA. *Corresponding author: (alexsteeler17@mail.fresnostate.edu).

Abstract

Agriculturally invasive weed species are notoriously capable of rapidly evolving adaptations, such as herbicide resistance, in response to intense selection pressures posed by the agroecosystem environments they invade. This capability to evolve on contemporary timescales presents many challenges, however, it also presents some of the best opportunities for population geneticists to study the underlying genetic mechanisms facilitating adaptive evolution. Nevertheless, there remains very little known about many of the fundamentals of the genetic mechanisms facilitating rapid evolutionary adaptation in agroecosystem invasions, due to a noticeable lack of genetic and evolutionary perspectives in weed science research. In this study, we will attempt to address some of these gaps by investigating the origin of recently established infestations of the invasive agricultural weed *Amaranthus tuberculatus* (waterhemp) within Merced County agroecosystems, using population genetics approaches. Waterhemp is known to be widely problematic throughout Midwest agroecosystems, however, it has not previously been reported within California agroecosystems and it is unknown where these infestations originate from. By determining the origins of these recently arisen waterhemp populations in Merced, we hope to shed some light on the relative importance of preadaptation and evolutionary change in overcoming selection pressures in agroecosystem invasions, as well as the role and importance of multiple introductions. To accomplish this, we will collect DNA samples from known Merced populations and generate genome-wide short-read sequence data for individual genotyping and SNP discovery, using the genotyping-by-sequencing (GBS) next generation sequencing technique on the Illumina Hiseq4000 platform. This data will then be used to assess allelic variation within Merced populations, and be compared to a broad survey of potential Midwestern source populations for source identification using both model- and non-model-based clustering methods.

Amaranthus tuberculatus (Waterhemp)

- Dioecious annual herb native to Midwestern riparian environments
- Two varieties: *rudis* (west) and *tuberculatus* (east)
- *Rudis* variety began invading agroecosystems in 1950's
  - Invasion initially facilitated by weedy life-history traits, continued success due to rapid and repeated evolution of herbicide resistance
    - Multiply-resistant biotypes not uncommon
    - Glyphosate resistance first reported in 2005, since confirmed in 19 states and Ontario, Canada
  - Designated in the top 10 worst weeds in America (WSSA, 2017)

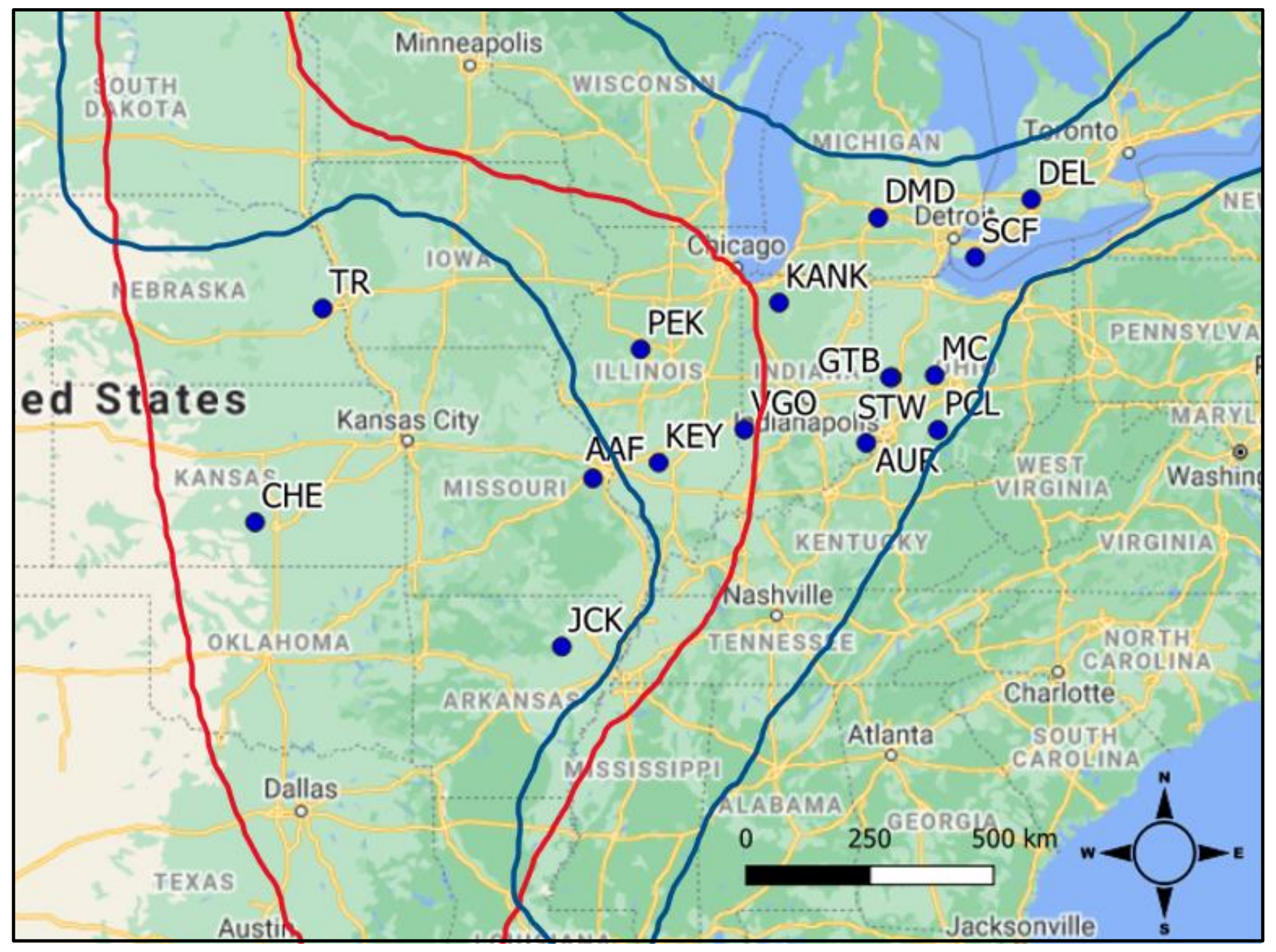

Figure 1 — Historical ranges of the two *A. tuberculatus varieties* in the Midwestern U.S. (Waselkov & Olsen, 2014). Red line corresponds to the *rudis* variety (focus of this study), and blue line corresponds to the non-weedy *tuberculatus* variety. Blue circles represent *A. tuberculatus* var. *rudis* populations included in this study for comparison. Geographic coordinates were plotted using QGIS software (QGIS, 2020).

Waterhemp in the Central Valley

- Non-native to California
  - Increasingly observed invading Merced County agroecosystems since 2010 (Figure 3)
- Origin of invasion unknown
  - Speculated to be introduced via seed-contaminated agricultural machinery or products
- Potential ability to spread further in Central Valley agroecosystems unknown
  - Local agroecosystems are much drier and more saline than in the Midwest
- Incidence of herbicide resistance unknown
  - Several populations are growing in fields planted with glyphosate-resistant corn varieties

Figure 2 — *A. tuberculatus* var. *rudis* growing in Merced County agroecosystems (2020).

Experimental Design

Specific Aim: To determine the geographic origins and routes of invasion of recently arisen infestations of waterhemp in Merced county agroecosystems.

1. Sample Collection (leaf tissue)

Merced – 7 populations
- 20 individuals/population (10 male + 10 female)

Midwest – 16 populations
- 4-10 individuals/population

2. DNA Extraction

Omega Bio-Tek Plant DNA extraction kit
- Dry tissue, LN2, mortar & pestle
- High MW pure DNA at 20 ng/μL: quantity/quality assessed with Qubit fluorometer, nanodrop, and gel electrophoresis

3. DNA Sequencing

Genotyping-by-sequencing (GBS)
- Digested using *ApeKI* RE
- Illumina HiSeq 100-bp single end reads (10x coverage) at UC Davis
- Generates genome-wide short-read sequence data

4. Raw Data Processing

STACKS v1.44 pipeline
- Demultiplex, trim, filter reads
- Align reads to reference genome for individual genotyping & SNP discovery
- Population-level statistics

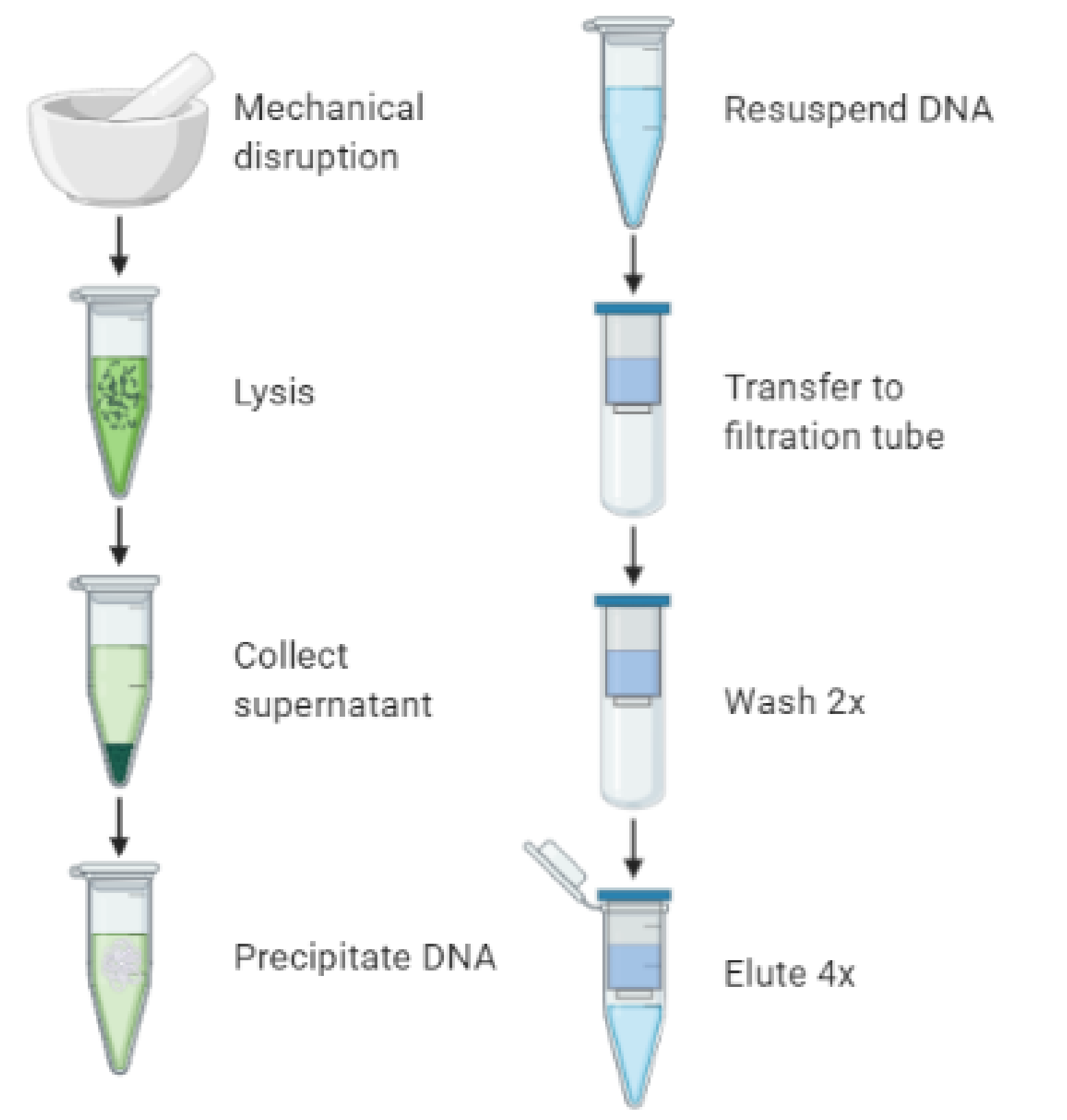

Figure 4 — General protocol followed to accomplish Merced waterhemp DNA extractions. Image created in BioRender.com (2020).

Figure 5 — Infestations of *A. tuberculatus* var. *rudis* growing within Merced County agroecosystems (2020).

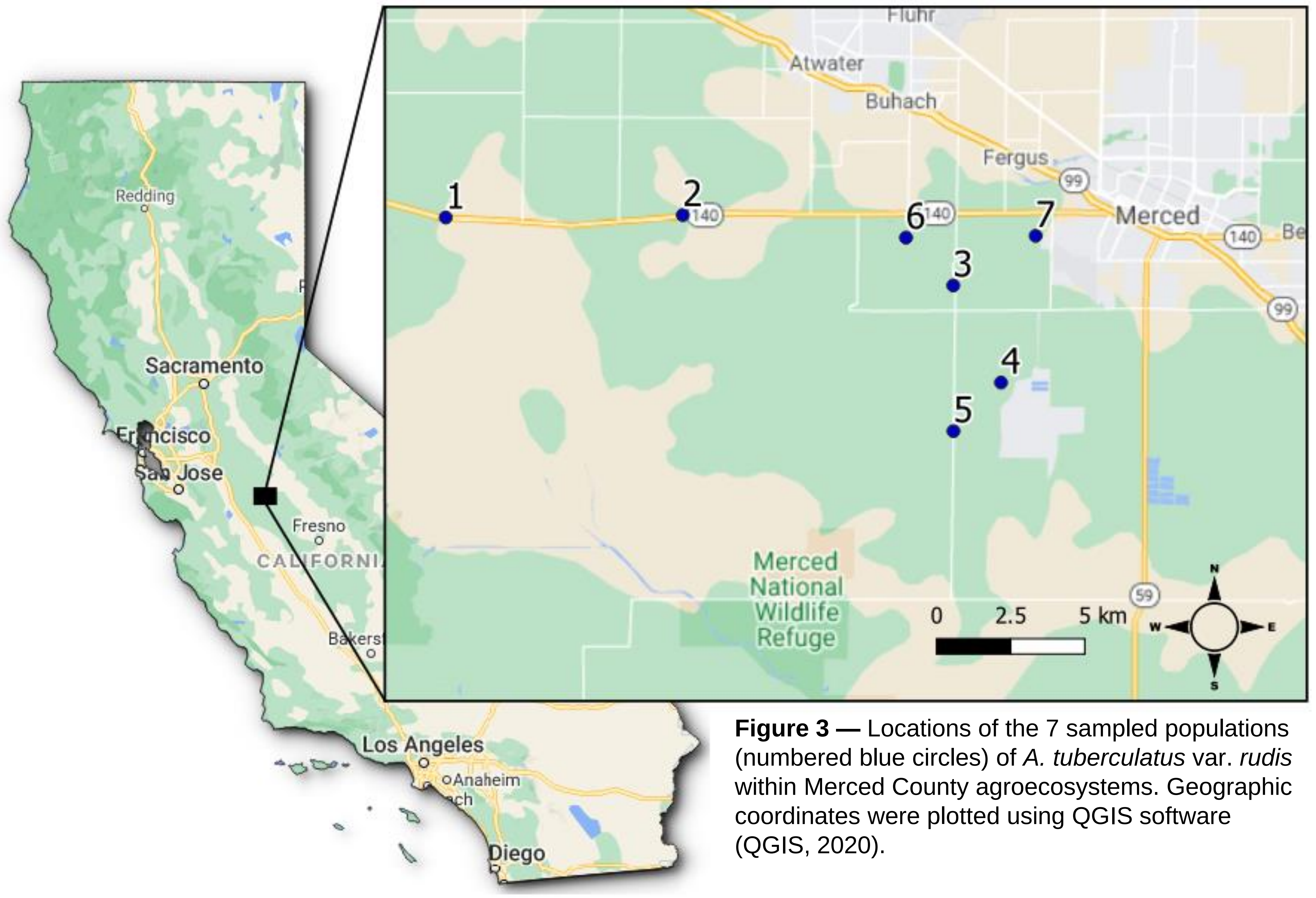

Figure 3 — Locations of the 7 sampled populations (numbered blue circles) of *A. tuberculatus* var. *rudis* within Merced County agroecosystems. Geographic coordinates were plotted using QGIS software (QGIS, 2020).

Data Analysis

Source identification: Genetic clustering
- *STRUCTURE* v2.3.4 (Pritchard et al., 2000)
  - Model-based Bayesian genetic clustering method
  - Relies on comparison of potential clustering likelihoods
- Principal Component Analysis (PCA)
  - Non-model-based genetic clustering method
  - Relies on matrix feature extraction to reduce genetic matrix dimensionality
  - Will be conducted in R v3.6.1 using the package *adegenet* v2.1.3

Invasion route identification: Demographic modeling
- Approximate Bayesian computation (ABC)
  - Model-based Bayesian method that infers complex demographic history based on posterior probabilities of different invasion scenarios
  - DIY ABC V1.0 software (Cornuet et al., 2010)

Implications

- May allow us to predict the incidence of herbicide resistance based on source
  - Many populations in Midwest have already been characterized
- Further understanding of the role of multiple introductions and preadaptation
  - Little is known about the colonization dynamics of agroecosystem invasions
- Potential to help inform the development of management strategies
  - Waterhemp is widely problematic in the Midwest, effective strategies will be needed to prevent it from establishing as a major weed throughout CA

Project Timeline

Time Period	Task
June 2019 ✓	• Collect genetic specimens from the 7 identified waterhemp populations in Merced, Ca.
January - October 2020 ✓	• Extract high quality DNA from samples. • Pre-process samples for sequencing platform.
In progress	• Genotype-by-Sequencing (GBS) via the High-Throughput Sequencing Illumina HiSeq4000 platform.
Spring 2021	• Raw data processing, Genotyping and SNP discovery. • Data analysis: *STRUCTURE*, PCA, ABC.
May - August 2021	• Thesis submission (M.S. in Biology). • Exit seminar and article publication.

Future Work in Waselkov and Shrestha Labs

- Determine incidence of herbicide resistance in Merced waterhemp
  - Genetic resistance and greenhouse dose response screening
- Assess the ability of Merced waterhemp to survive under Central Valley conditions
  - Growth potential under various levels of moisture and salt stress

References

1. Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22: 3124-3140.
2.Cornuet, J. M., A. Estoup, and V. Ravigné. 2010. Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). BMC Bioinformatics 11: 401.
3. Earl, D., and B. vonHoldt. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4: 359-361.
4. Heap, I. 2019. International survey of herbicide resistant weeds. Website www.weedscience.org.
5. Jakobsson, M., and N. A. Rosenberg. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23: 1801-1806.
6. Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers., 1403-1405. Bioinformatics
7. Pimentel, D., R. Zuniga, and D. Morrison. 2005. Update on the environmental and economic costs.
8. Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of Population Structure Using Multilocus Genotype Data. *Genetics* 155: 945.
9. QGIS. 2020. QGIS Graphic Information System. Open Source Geospatial Foundation Project. https://qgis.osgeo.org.
10. Rosenberg, N. A. 2004. distruct : a program for the graphical display of population structure. Molecular Ecology Notes 4: 137-138.
11. Sauer, J. 1957. Recent Migration and Evolution of the Dioecious Amaranths. Evolution 11: 11-31.
12. Tranel, P. J., and F. Trucco. 2009. 21st-Century Weed Science: A Call for Amaranthus Genomics, Weedy and Invasive Plant Genomics, 53-81.
13. Waselkov, K. E., and K. M. Olsen. 2014. Population genetics and origin of the native North American agricultural weed waterhemp (Amaranthus tuberculatus; Amaranthaceae. *American Journal of Botany* 101: 1726-1736.
14. WSSA. 2017. WSSA Survey Ranks Most Common and Most Troublesome Weeds in Broadleaf Crops, Fruits and Vegetables. WSSA: Weed Science Society of America. Website.

Acknowledgements

We thank the CSU Agricultural Research Institute (Project #20-02-121) for funding this research, and the USDA National Institute of Food and Agriculture Award 2019-67013-29355 (subaward 095689-17507) for matching funding that made the ARI grant possible.