

Title: Dataset associated with “Investigating the origins and evolution of a glyphosate-resistant weed invasion in South America”

Abstract: The global invasion, and subsequent spread and evolution of weeds provides unique opportunities to address fundamental questions in evolutionary and invasion ecology. *Amaranthus palmeri* is a widespread glyphosate-resistant (GR) weed in the USA. Since 2015, GR populations of *A. palmeri* have been confirmed in South America, raising questions about introduction pathways and the importance of pre- versus post-invasion evolution of GR traits. We used RAD-Seq genotyping to characterize genetic structure of populations from Brazil, Argentina, Uruguay and the USA. We also quantified gene copy number of the glyphosate target, 5-enolpyruvyl-3-shikimate phosphate synthase (*EPSPS*) and the presence of an extra-chromosomal circular DNA (eccDNA) replicon known to confer GR in USA populations. Populations in Brazil, Argentina, and Uruguay were only weakly differentiated (pairwise $F_{ST} \leq 0.043$) in comparison to USA populations (mean pairwise $F_{ST} = 0.161$, range = 0.068–0.258), suggesting a single major invasion event. However, elevated *EPSPS* copy number and the *EPSPS* replicon were identified in all populations from Brazil and Uruguay, but only in a single Argentinean population. These observations are consistent with independent *in situ* evolution of glyphosate resistance in Argentina, followed by some limited recent migration of the eccDNA based mechanism from Brazil to Argentina. Taken together, our results are consistent with an initial introduction of *A. palmeri* into South America sometime before the 1980s, and local evolution of GR in Argentina, followed by a secondary invasion of GR *A. palmeri* with the unique eccDNA based mechanism from the USA into Brazil and Uruguay during the 2010’s.

Contact: Paul Neve, pbneve@plen.ku.dk

Recommended data citation: Gaines, T., G. Slavov, D. Hughes, A. Kuepper, C. Sparks, J. Oliva, M. Vila-Aiub, M. A. Garcia, A. Merotto, and P. Neve. 2021. Dataset associated with " Investigating the origins and evolution of a glyphosate-resistant weed invasion in South America." Colorado State University. Libraries. <http://dx.doi.org/10.25675/10217/232637>

Associated publication: Gaines, T. A., Slavov, G. T., Hughes, D., Küpper, A., Sparks, C. D., Oliva, J., Vila-Aiub, M. M., Garcia, M. A., Merotto, A., & Neve, P. (2021). Investigating the origins and evolution of a glyphosate-resistant weed invasion in South America. *Molecular Ecology*, 30, 5360– 5372. <https://doi.org/10.1111/mec.16221>

Location where data were collected:

Code	Country	# Plants	Location	Collection site	Year
ARG-P1	Argentina	11	West Rio Cuarto, Cordoba	Soybeans	2016
ARG-P2	Argentina	8	Sampacho, Cordoba	Soybeans	2016
ARG-P3	Argentina	15	Vizcacheras, San Luis	Roadside	2016
ARG-P4	Argentina	9	Justo Daract, San Luis	Corn	2016

ARG-P5	Argentina	12	Justo Daract, San Luis	Grain elevator	2016
ARG-P6	Argentina	13	Pizarro, Cordoba	Soybeans	2016
ARG-P7	Argentina	8	Pizarro/Valeria, Cordoba	Sorghum	2016
ARG-P8	Argentina	8	Las Lomas, Villa Valeria, Cordoba	Corn	2016
ARG-P9	Argentina	8	Melideo de La Serna, Cordoba	Soybeans	2016
ARG-P10	Argentina	18	Rio Quinto, Cordoba	Soybeans	2016
BRZ-P1	Brazil	21	Tapurah, Mato Grosso	Soybeans / cotton	2016
BRZ-P2	Brazil	18	Ipiranga do Norte, Mato Grosso	Soybeans / cotton	2016
BRZ-P3	Brazil	21	Ipiranga do Norte, Mato Grosso	Soybeans / cotton	2016
BRZ-P4	Brazil	28	Campos de Julio, Mato Grosso	Soybeans / cotton	2016
URU-P1	Uruguay	19	Colonia Valdense, Colonia	Corn	2017
URU-P2	Uruguay	17	Porvenir, Paysandú	Soybeans	2017
URU-P3	Uruguay	16	Colonia Tomas Berreta, Rio Negro	Soybeans	2017
AZ-R	USA	17	Buckeye, Arizona	Cotton	2012
AZ-S	USA	17	Sahuarita, Arizona	Desert	2012
CO-R	USA	14	Yuma County, Colorado	Sugar beet	2015
GA-R	USA	16	Macon, Georgia	Cotton	2006
GA-S	USA	17	Worth County, Georgia	Cotton	2004
KS-S	USA	13	Ottawa, Kansas	Soybean	2005
NC-R	USA	2	North Carolina	Cotton	2006
TN-R	USA	17	Jackson, Tennessee	Soybean	2007
AZS-2	USA	17	Tucson, Arizona	Desert	1981

Time period during which data were collected: 2017 to 2019

Format of data files:

.vcf

File Information:

HapMap.vcf – VCF file from UNEAK on RAD-Seq data containing 4,659 SNPs that can be used to reproduce all analyses, with a bit of formatting for STRUCTURE, smartpca, PLINK, and BA3

About the datasets:

The commands, options, and software versions to run the analyses are described in the materials and methods section of the manuscript.

The Illumina short reads used to generate the VCF file in UNEAK are available in NCBI SRA under BioProject PRJNA672995.