

Title: Dataset associated with “Genomic-based epidemiology reveals independent origins and gene flow of glyphosate resistance in *Bassia scoparia* populations across North America”

Abstract: Genomic-based epidemiology can provide insight into the origins and spread of herbicide resistance mechanisms in weeds. We used kochia (*Bassia scoparia*) populations resistant to the herbicide glyphosate from across western North America to test the alternative hypotheses that 1) a single *EPSPS* gene duplication event occurred initially in the Central Great Plains and then subsequently spread to all other geographical areas now exhibiting glyphosate-resistant kochia populations or that 2) gene duplication occurred multiple times in independent events in a case of parallel evolution. We used qPCR markers previously developed for measuring the structure of the *EPSPS* tandem duplication to investigate whether all glyphosate-resistant individuals had the same *EPSPS* repeat structure. We also investigated population structure using simple sequence repeat (SSR) markers to determine the relatedness of kochia populations from across the Central Great Plains, Northern Plains, and the Pacific Northwest. We found that the original *EPSPS* duplication genotype was predominant in the Central Great Plains where glyphosate resistance was first reported. We identified two additional *EPSPS*-duplication genotypes, one having geographic associations with the Northern Plains and the other to the Pacific Northwest. The *EPSPS* duplication genotype from the Pacific Northwest seems likely to represent a second, independent evolutionary origin of a resistance allele. We found evidence of gene flow across populations and a general lack of population structure. The results support at least two independent evolutionary origins of glyphosate resistance in kochia, followed by substantial and mostly geographically localized gene flow to spread the resistance alleles into diverse genetic backgrounds.

Contact: Todd Gaines, todd.gaines@colostate.edu

Recommended data citation: Ravet, K., C. Sparks, A. Dixon, A. Küpper, E. Westra, D. Pettinga, P. Tranel, J. Felix, D. Morishita, P. Jha, A. Kniss, P. Stahlman, P. Neve, E. Patterson, P. Westra, and T. Gaines. 2020. Dataset associated with “Genomic-based epidemiology reveals independent origins and gene flow of glyphosate resistance in *Bassia scoparia* populations across North America.” Colorado State University Libraries. <https://doi.org/10.25675/10217/216180>

Associated publication: Ravet, K., Sparks, C.D., Dixon, A.L., Küpper, A., Westra, E.P., Pettinga, D.J., Tranel, P.J., Felix, J., Morishita, D.W., Jha, P., Kniss, A., Stahlman, P.W., Neve, P., Patterson, E.L., Westra, P. and Gaines, T.A. (2021), Genomic-based epidemiology reveals independent origins and gene flow of glyphosate resistance in *Bassia scoparia* populations across North America. Mol Ecol. Accepted Author Manuscript. <https://doi.org/10.1111/mec.16215>

Location where data were collected:

| Population | Country | State/Province | City or County | Year | n (plant) | Resistance |
|------------|---------|----------------|----------------|------|-----------|------------|
| CO1R | USA | Colorado | Akron* | 2012 | 18 | R |
| CO2R | USA | Colorado | Brush | 2012 | 18 | R |
| CO3R | USA | Colorado | Cope | 2012 | 18 | R |
| CO4R | USA | Colorado | Julesburg | 2011 | 18 | R |
| CO5R | USA | Colorado | Kit Carson | 2013 | 18 | R |
| CO6R | USA | Colorado | Otis | 2012 | 18 | R |
| CO7R | USA | Colorado | Strasburg | 2012 | 18 | R |

| | | | | | | |
|-------|--------|----------|-----------|------|----|----------------|
| CO8R | USA | Colorado | Strasburg | 2014 | 18 | R |
| ID1R | USA | Idaho | Ada | 2014 | 9 | R |
| ID2R | USA | Idaho | Ada | 2014 | 9 | R |
| KS1S | USA | Kansas | Barton | 2012 | 9 | S ⁺ |
| KS2S | USA | Kansas | Finney | 2012 | 9 | S ⁺ |
| KS3R | USA | Kansas | Gray | 2012 | 9 | R |
| KS4R | USA | Kansas | Greeley | 2012 | 9 | R |
| KS5S | USA | Kansas | Meade | 2012 | 9 | S ⁺ |
| KS6S | USA | Kansas | Ness | 2012 | 9 | S ⁺ |
| KS7S | USA | Kansas | Philip | 2012 | 9 | S |
| KS8S | USA | Kansas | Pratt | 2012 | 9 | S ⁺ |
| KS9R | USA | Kansas | Scott | 2012 | 9 | R |
| KS10R | USA | Kansas | Scott | 2012 | 9 | R |
| KS11R | USA | Kansas | Stevens | 2012 | 9 | R |
| KS12R | USA | Kansas | Thomas | 2012 | 9 | R |
| KS13R | USA | Kansas | Thomas | 2007 | 9 | R |
| KS14R | USA | Kansas | Wallace | 2012 | 9 | R |
| KS15R | USA | Kansas | Wichita | 2012 | 9 | R |
| MT1R | USA | Montana | Chester | 2012 | 9 | R |
| MT2R | USA | Montana | Gilford | 2012 | 9 | R |
| MT3R | USA | Montana | Joplin | 2012 | 9 | R |
| OK1R | USA | Oklahoma | Cimarron | 2012 | 9 | R |
| OR1R | USA | Oregon | Malheur | 2014 | 9 | R |
| OR2R | USA | Oregon | Malheur | 2014 | 9 | R |
| OR3R | USA | Oregon | Malheur | 2014 | 9 | R |
| OR4R | USA | Oregon | Malheur | 2014 | 9 | R |
| OR5R | USA | Oregon | Malheur | 2014 | 9 | R |
| OR6R | USA | Oregon | Malheur | 2014 | 9 | R |
| OR7R | USA | Oregon | Malheur | 2014 | 9 | R |
| OR9S | USA | Oregon | Malheur | 2015 | 9 | S |
| TX1R | USA | Texas | Hartley | | 12 | R |
| TX2R | USA | Texas | Hartley | | 18 | R |
| TX3R | USA | Texas | Hartley | | 18 | R |
| TX4R | USA | Texas | Hartley | | 18 | R |
| TX5R | USA | Texas | Hartley | | 18 | R |
| WY1R | USA | Wyoming | Powell | 2015 | 9 | R |
| AB1R | CANADA | Alberta | | | 18 | R |

Time period during which data were collected: 2007 to 2015

Format of data files:

.txt
.Rmd
.xlsx
.csv

File Information:

SSR Kochia_Ravet et al_Molecular Ecology.xlsx – Microsoft Excel file containing fragment lengths for 11 SSR markers and allele bins (processed data); data used to construct Figures 3 and 4, and Table 4

Allele length dataset.csv – Comma Delimited Values file containing the same data as the “Allele length dataset” sheet in the above Excel file (saved in .csv format for preservation purposes)

SNP database.csv – Comma Delimited Values file containing the same data as the “SNP database” sheet in the above Excel file (saved in .csv format for preservation purposes)

Kochia_info_popinfo_OCT232020.csv – Comma Delimited Values file containing data on the geographical region and year of collection for each population in the SSR data set

Kochia_Ravet et al_pop gen R code_April2021.Rmd – text file containing code to run population genetics analysis in R, used to generate figure 3 and table 4

STRUCTURE_analysis_data.txt – text file containing SSR genotype data in correct format to run STRUCTURE analysis, used to generate figure 4

kochia_oct2020_genepop.txt – text file formatted as a single space delimited text file, SSR data modified by hand to create genepop format; use for subsequent analysis following line 79 in the .Rmd file; used to generate figure 3 and table 4

EPSPS and MGE database kochia.csv – Comma Delimited Values file containing individual-level raw data for EPSPS, Type I, Type II, and MGE copy number, used to assign genotype for Figure 1, and plotted in Figure 2; and to generate the population means shown in Table 2.

Abbreviations:

Ind = individual

Pop = population

About the datasets:

Allele length determination and SNP call (binning) method are explained in Materials and Methods in the associated article.

Both datasets consist of 11 loci.

The data in the EPSPS and MGE database are quantitative PCR results for EPSPS, Type I, Type I, and MGE copy number for all individuals included in the study.

Sequences:

A full list of primer sequences is provided in Supplementary Material for the associated article.