

Rothamsted Repository Download

A - Papers appearing in refereed journals

Cai, L., Comont, D., Macgregor, D., Lowe, C., Beffa, R., Sasaki, C. A. and Neve, P. 2022. The blackgrass genome reveals patterns of non-parallel evolution of polygenic herbicide resistance. *New Phytologist*.
<https://doi.org/10.1111/nph.18655>

The publisher's version can be accessed at:

- <https://doi.org/10.1111/nph.18655>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/98769/the-blackgrass-genome-reveals-patterns-of-non-parallel-evolution-of-polygenic-herbicide-resistance>.

© 1 December 2022, Please contact library@rothamsted.ac.uk for copyright queries.

The blackgrass genome reveals patterns of divergent evolution of non-target site resistance to herbicides.

Lichun Cai^{1†}, David Comont^{2†}, Dana MacGregor², Claudia Lowe², Roland Beffa^{3,5*}, Paul Neve^{2,4*}, Christopher Saski^{1*}

¹ Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, USA.

² Department of Biointeractions & Crop Protection, Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom.

³ Königsteiner Weg 4, 65835 Liederbach, Germany.

⁴ Department of Plant and Environmental Sciences, University of Copenhagen, Højbakkegård Allé 13, Tåstrup 2630, Denmark.

⁵ Previous address: Bayer Crop Sciences, Industriepark Höchst, Frankfurt am Main, Germany.

†Co-contributing authors – these authors contributed equally to the work.

*Co-corresponding senior authors – these authors contributed equally to the work.

Paul Neve (pbneve@plen.ku.dk)

Christopher Saski (saski@clemson.edu)

Roland Beffa (roland.beffa@t-online.de)

Table S1. Genome size estimation based on flow cytometry.

Sample ID	Population	Phenotype	DNA content (pg/2C)	Mbp/2C	Gbp/1C
BG1	Rothamsted	S	6.90	6750	3.37
BG2	Rothamsted	S	6.74	6595	3.30
BG3	Rothamsted	S	6.81	6656	3.33
BG4	PELD	R	6.77	6625	3.31
BG5	PELD	R	7.54	7373	3.69
BG6	PELD	R	6.77	6625	3.31
BG7	LOLA 81	S	6.84	6687	3.34
BG8	LOLA 81	S	6.74	6595	3.30
BG9	LOLA 81	S	7.00	6846	3.42
BG10	LOLA 91	R	6.90	6750	3.37
BG11	LOLA 91	R	6.77	6625	3.31
BG12	LOLA 91	R	6.87	6718	3.36
Average	NA	NA	6.89	6737	3.37

Table S2. Genome size estimation based on k-mer analysis.

<i>k</i> -mer length	Total number of <i>k</i> -mer	<i>k</i> -mer coverage	Estimated genome size
17	237,986,849,712	70	3400Mb
19	234,416,794,205	68	3447Mb
21	230,846,540,235	66	3498Mb
31	212,995,740,384	60	3550Mb

Table S3. Statistics of the assembled seven chromosomes of *A. myosuroides*.

Chromosome	Anchored scaffold number	Length of chromosome
Chr1	627	781,714,917
Chr2	416	571,710,514
Chr3	350	522,029,352
Chr4	321	443,909,942
Chr5	278	369,436,349
Chr6	264	359,679,614
Chr7	238	351,570,514
Total	2494	3,400,051,202

Table S4. BUSCO analysis of genome completeness.

Description	Gene number	Percentage (%)
Complete BUSCOs (C)	1563	96.9
Complete and single-copy BUSCOs (S)	1497	92.8
Complete and duplicated BUSCOs (D)	66	4.1
Fragmented BUSCOs (F)	14	0.9
Missing BUSCOs (M)	37	2.2
Total BUSCO groups searched	1614	100

Table S5. Summary of cytogenic and assembly length of each blackgrass chromosome (Johnsson, 1944).

Chromosome	Cytogenic length (μm)	Assembly length (bp)
1	8.4	781,714,917
2	6.4	571,710,514
3	5.4	522,029,352
4	5.0	443,909,942
5	4.4	369,436,349
6	3.8	359,679,614
7	3.0	351,570,514

Table S6. Statistics of the annotated transposon elements (TEs).

Class	Superfamily	Number	Total size (bp)	Percentage (%)
Class 1	LTR-retrotransposons			
	<i>Gypsy</i> (RLG)	1922808	1,367,269,327	39.16
	<i>Copia</i> (RLC)	445775	298,678,755	8.56
	Unclassified (RLX)	1574609	623,678,411	17.86
	Non-LTR retrotransposons			
	Long interspersed nuclear elements (RIX)	20527	12,053,133	0.35
	Short interspersed nuclear elements (SIX)	2807	797,889	0.02
Class 2	DNA transposons			
	hAT (DTA)	41471	15,852,349	0.45
	CACTA (DTC)	335408	169,435,014	4.85
	Harbinger (DTH)	124608	61,410,666	1.76
	Mutator (DTM)	181317	79,309,130	2.27
	Mariner (DTT)	189514	60,919,894	1.74
	Unclassified (DXX)	24210	11,107,295	0.32
	Helitron (DHH)	248002	109,086,060	3.12
Others	XXX	176175	41,788,046	1.20
Total	TEs	5287231	2,851,385,969	81.68

Table S7. Summary of identified QTLs from CC2 and CC5 populations.

Population	Chromosome	Name	Start(bp)	End(bp)	Length(bp)	Peak delta SNP index	Peak position (bp)	Average delta SNP index
CC2	Chr2	<i>qtl-cc2-2-1</i>	78,155,610	80,414,345	2,258,735	0.35	79,286,939	0.31
	Chr2	<i>qtl-cc2-2-2</i>	84,968,723	86,789,597	1,820,874	0.33	85,985,573	0.31
	Chr2	<i>qtl-cc2-2-3</i>	155,616,708	158,148,961	2,532,253	0.41	156,325,090	0.39
	Chr3	<i>qtl-cc2-3-1</i>	314,531,189	315,950,139	1,418,950	0.30	315,049,343	0.27
	Chr5	<i>qtl-cc2-5-1</i>	190,968,113	192,166,507	1,198,394	0.29	191,929,219	0.27
	Chr5	<i>qtl-cc2-5-2</i>	210,023,549	211,392,868	1,369,319	-0.29	210,693,320	-0.26
	Chr6	<i>qtl-cc2-6-1</i>	139,536,063	142,546,311	3,010,248	0.28	140,494,464	0.27
CC5	Chr2	<i>qtl-cc5-2-1</i>	28,124,368	30,262,000	2,137,632	-0.33	29,036,282	-0.30
	Chr3	<i>qtl-cc5-3-1</i>	197,701,391	198,108,389	406,998	-0.30	197,792,990	-0.27
	Chr3	<i>qtl-cc5-3-2</i>	224,118,449	226,070,109	1,951,660	0.30	224,317,480	0.29
	Chr3	<i>qtl-cc5-3-3</i>	229,745,161	230,705,090	959,929	0.29	230,421,617	0.28
	Chr3	<i>qtl-cc5-3-4</i>	248,724,538	251,387,988	2,663,450	-0.35	249,793,371	-0.33
	Chr3	<i>qtl-cc5-3-5</i>	281,782,841	288,136,846	6,354,005	-0.39	283,456,062	-0.35
	Chr3	<i>qtl-cc5-3-6</i>	308,719,007	313,553,107	4,834,100	-0.41	309,948,215	-0.37
	Chr3	<i>qtl-cc5-3-7</i>	444,582,369	447,919,982	3,337,613	-0.41	446,569,526	-0.37

Table S8. Statistical assessment of over-representation of differentially expressed genes (DEGs) per chromosome, calculated using a one-sided Fishers exact test. Rows in bold represent chromosomes with a significantly greater number of DEGs than expected.

Chromosome	Total	Observed DEG	Expected DEG	Odds ratio	Pval	p.adjust
CC2 seed family						
Chr0	2111	7	19.08	0.37	0.995	1.000
Chr1	9278	64	83.84	0.76	0.959	1.000
Chr2	7080	124	63.98	1.95	<0.001	<0.001
Chr3	7041	62	63.62	0.97	0.606	1.000
Chr4	5563	27	50.27	0.54	0.997	1.000
Chr5	5217	40	47.14	0.85	0.805	1.000
Chr6	4383	55	39.61	1.38	0.074	0.593
Chr7	4590	30	41.48	0.73	0.924	1.000
CC5 seed family						
Chr0	2111	12	14.08	0.86	0.722	1.000
Chr1	9278	37	61.90	0.60	0.996	1.000
Chr2	7080	83	47.24	1.77	<0.001	0.008
Chr3	7041	84	46.98	1.80	<0.001	0.006
Chr4	5563	14	37.12	0.38	1.000	1.000
Chr5	5217	34	34.81	0.97	0.595	1.000
Chr6	4383	18	29.24	0.62	0.961	1.000
Chr7	4590	20	30.63	0.64	0.954	1.000

Table S9. Summary of gene number for NTSR-related gene families. The percentage was presented in the parenthesis.

Species	P450	GST	ABC	UGT	AKR	Total NTSR	Total genes
<i>A.thaliana</i>	245 (0.89%)	55 (0.20%)	130 (0.47%)	107 (0.39%)	22 (0.08)	559 (2.04)	27420
<i>O.sativa</i>	356 (1.00%)	79 (0.22%)	133 (0.37%)	180 (0.50%)	29 (0.08%)	777 (2.18%)	35662
<i>A.myosuroides</i>	506 (1.17%)	93 (0.22%)	146 (0.34%)	278 (0.64)	46 (0.11%)	1069 (2.48%)	43152

Table S10. Summary of differentially expressed NTSR-related gene number in each blackgrass chromosome in CC2 and CC5 families.

Count of NTSR-related genes in each blackgrass chromosome						
Chr	P450	GST	ABC	UGT	AKR	Total
1	131	11	20	78	4	244
2	80	34	41	55	4	214
3	115	10	16	45	16	202
4	34	2	23	18	2	79
5	51	21	10	37	10	129
6	46	12	16	24	1	99
7	49	3	20	21	8	101

Count of differentially expressed NTSR-related gene in CC2						
Chr	P450	GST	ABC	UGT	AKR	Total
1	5	1	2	3	0	11
2	6	1	0	6	2	15
3	0	5	0	3	2	10
4	1	0	0	0	0	1
5	2	1	0	0	0	3
6	3	1	1	2	0	7
7	1	0	1	1	0	3

Count of differentially expressed NTSR-related gene in CC5						
Chr	P450	GST	ABC	UGT	AKR	Total
1	4	0	1	0	0	5
2	3	1	0	3	1	8
3	1	4	0	1	0	6
4	1	0	0	0	0	1
5	1	0	0	1	0	2
6	0	1	1	2	0	4
7	0	0	1	1	0	2

**Pacific bioscience
long reads**
~42 million reads
~513Gb (144x)

**BioNano
(optical mapping)**
~3,685,283
molecules
~860Gb total length
(241x)

Hi-C
~421 million
reads
~126Gb(35x)

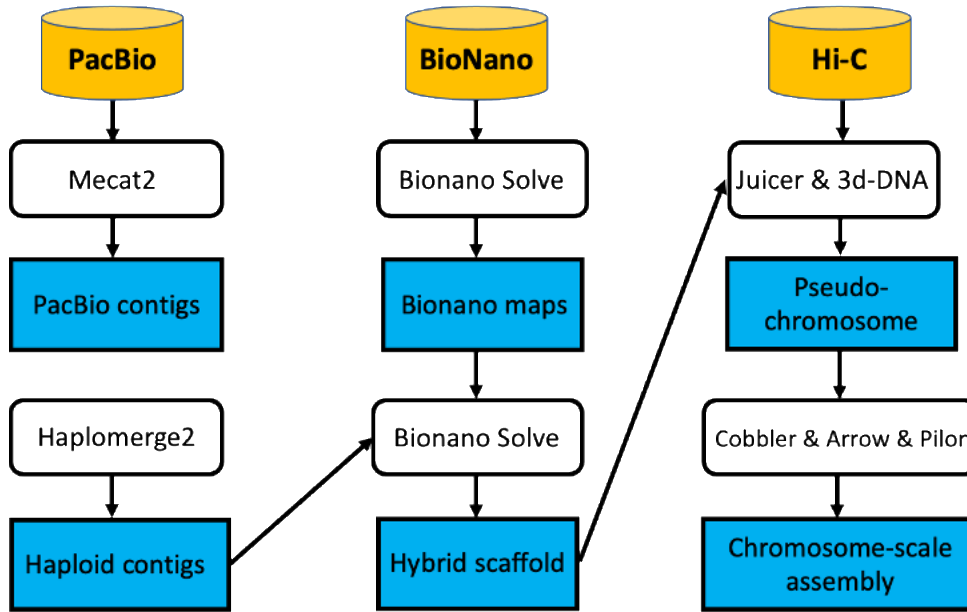


Fig. S1. Pipeline of genome assembly for the blackgrass.

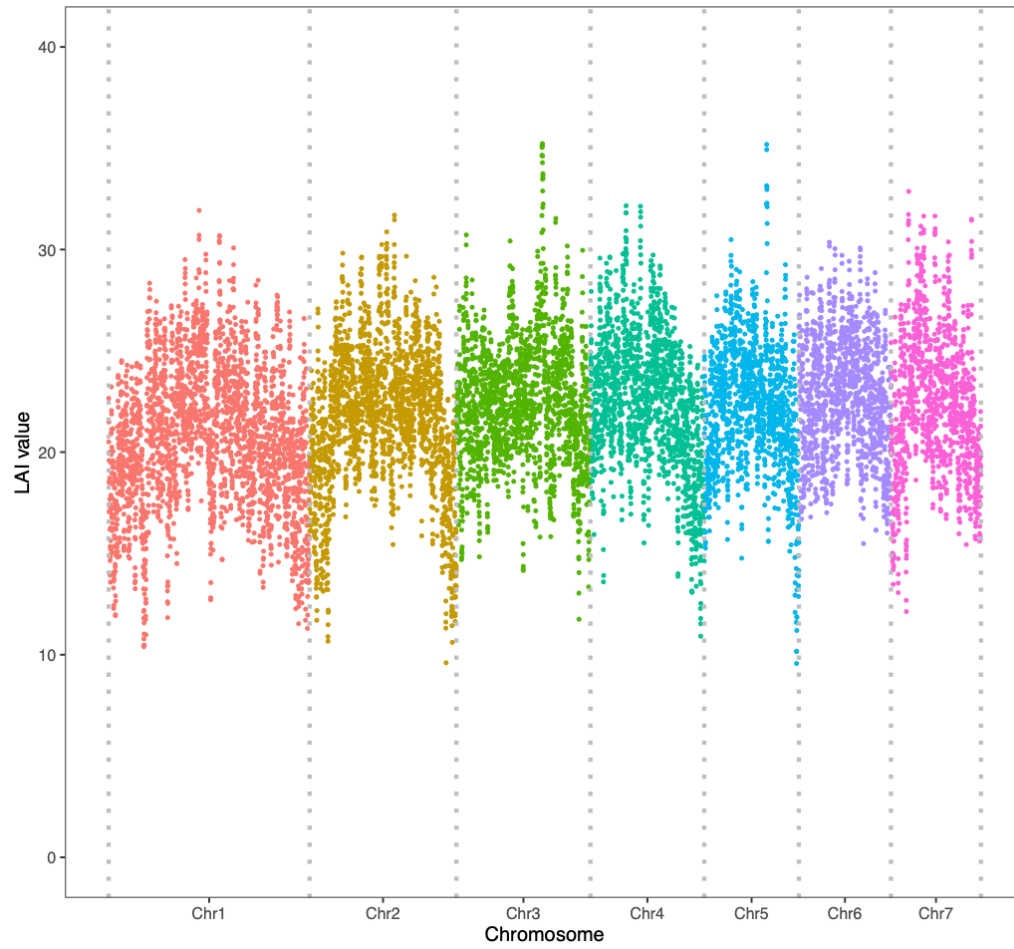


Fig. S2. Long terminal repeat assembly index (LAI) values across seven assembled blackgrass chromosomes.

RNA samples
from root,
stem, leaf,
flower (3
stages)

Related
species used:
Arabidopsis,
Barley, Wheat,
Rice, Maize,
Sorghum,
Brachypodium

Ab initio gene
prediction
software used:
SNAP,
Augustus,
GeneMark

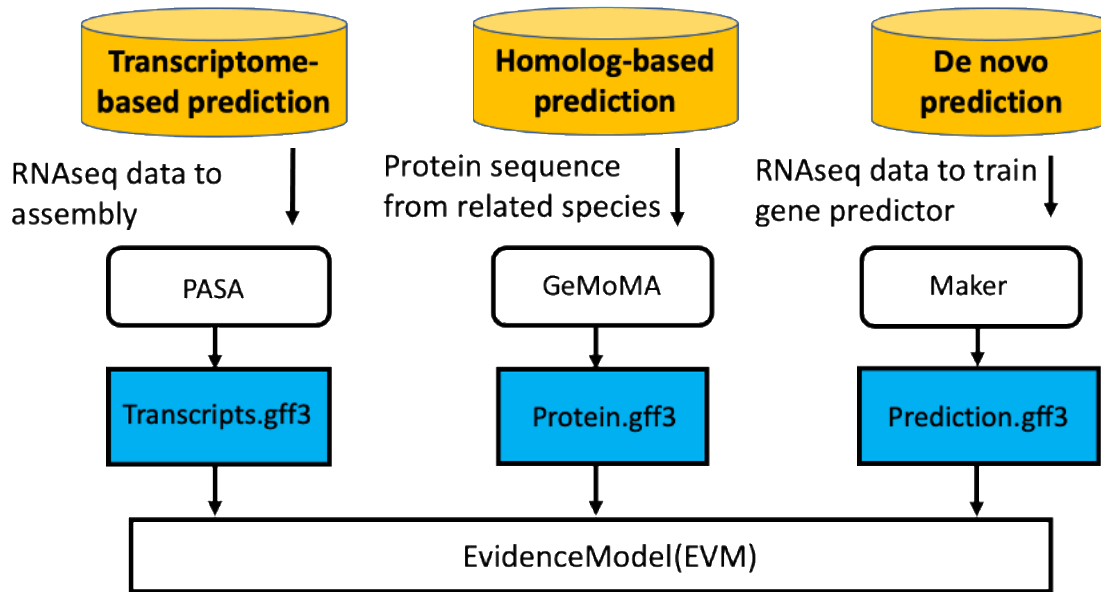


Fig. S3. Pipeline of gene annotation for the blackgrass.

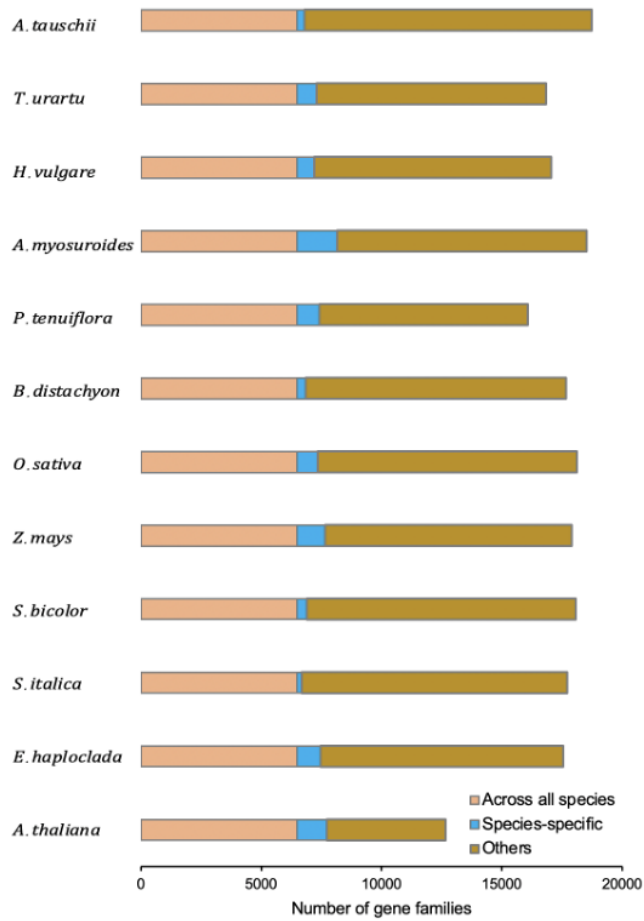


Fig. S4. The distribution of different types of gene family, including the ones presented across all the species, species-specific and all other types.

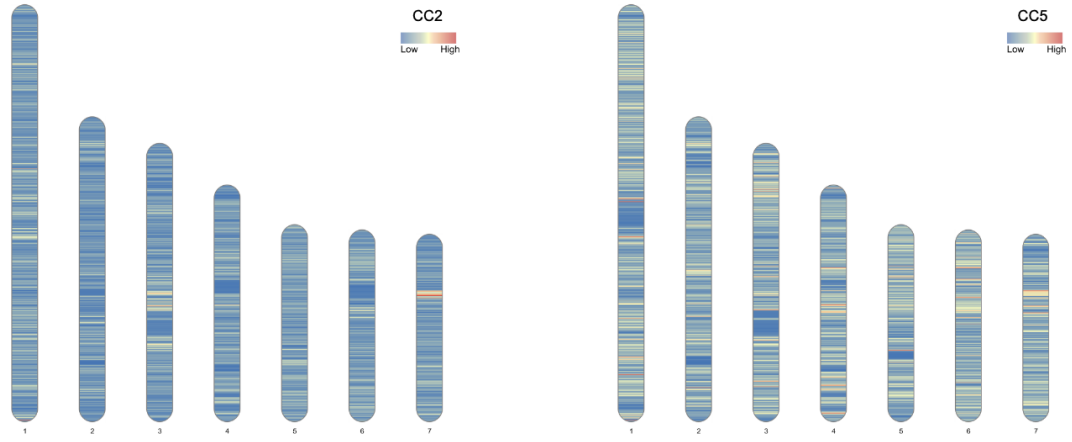


Fig. S5. SNP marker density for CC2 and CC5 population, respectively.

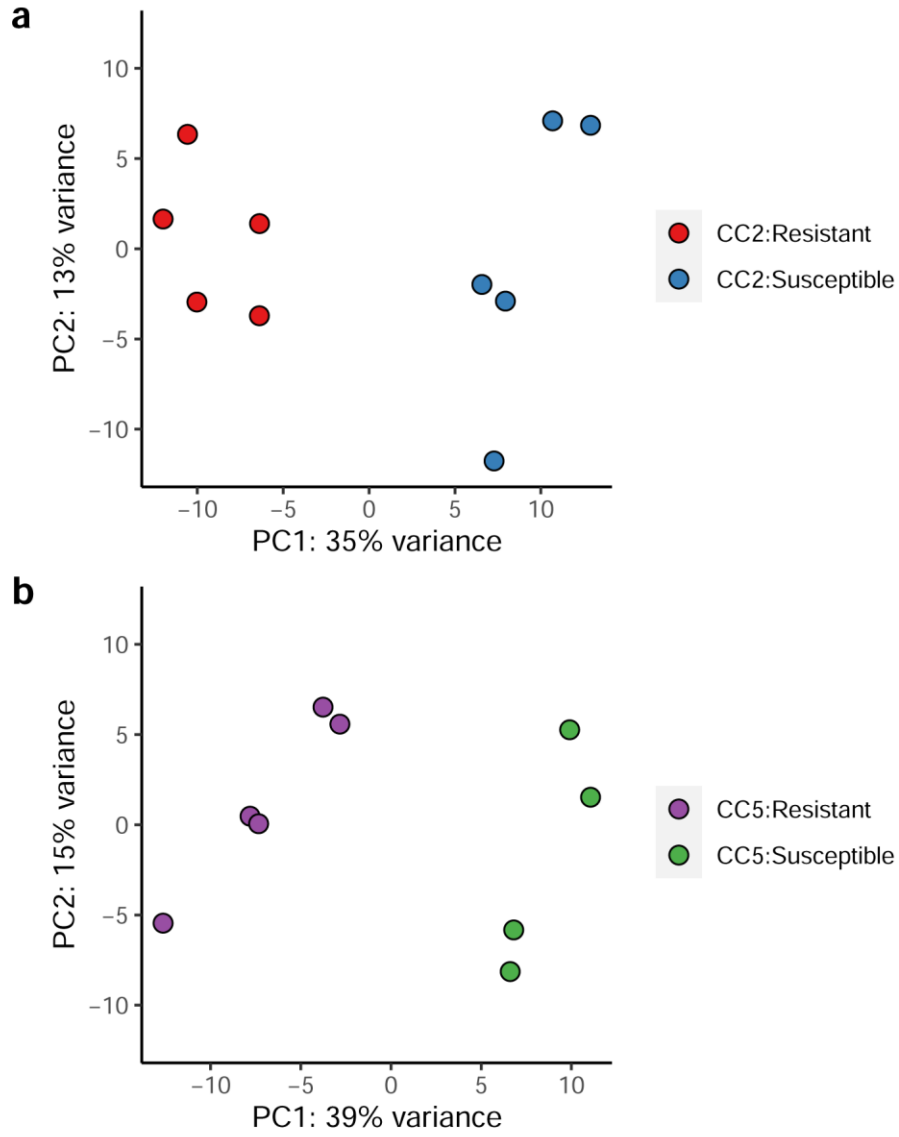


Fig. S6. Principal components analysis of all gene expression data for (a) the CC2 family and (b) the CC5 family alone.

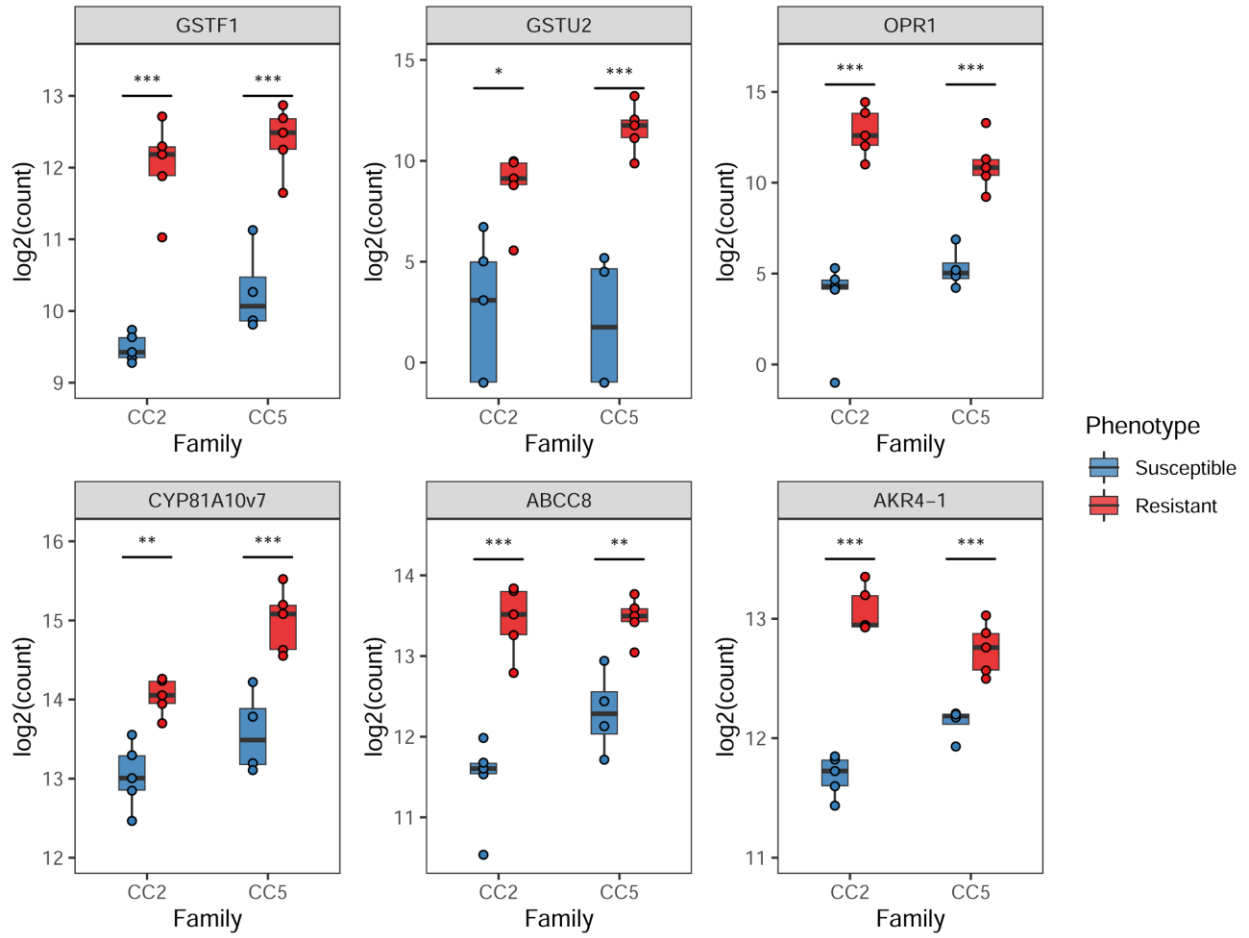


Fig. S7. Differential expression of previously reported NTSR candidate genes.

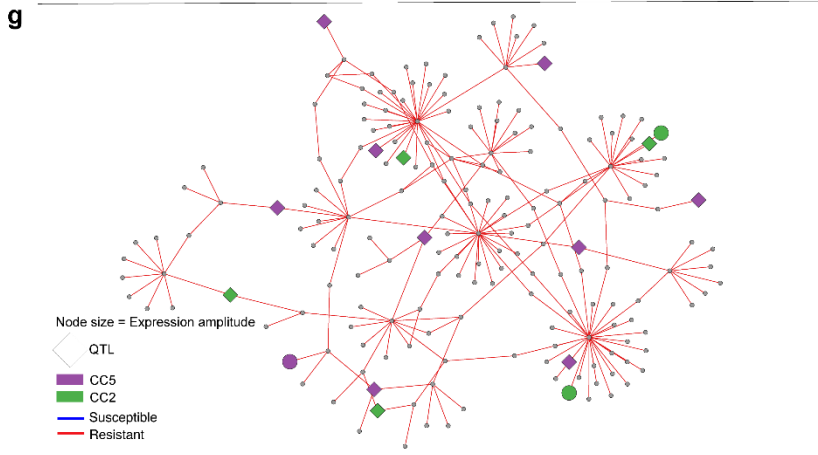
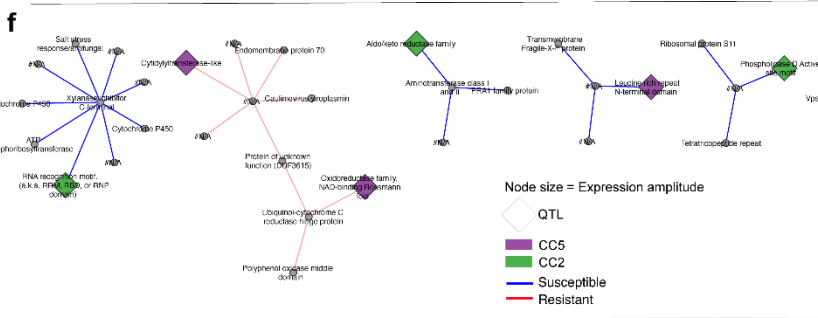
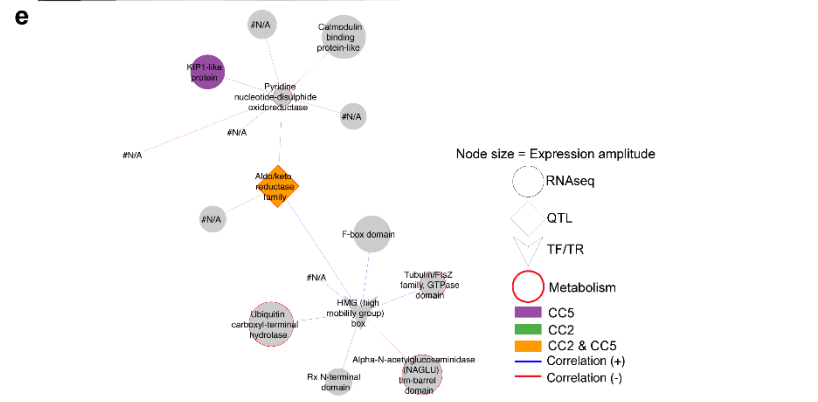
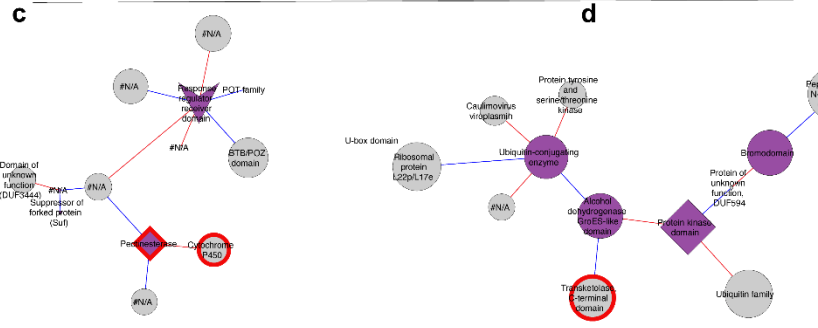
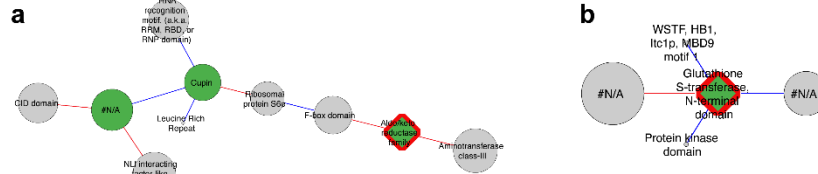


Fig. S8. Parallel and overlapping sub-graphs in CC2 and CC5. (a-b) Hub genes unique to CC2 subset from the traditional network discovered in the bulk-segregant analysis. (c-d) Hub genes unique to CC5 subset from the CC5 traditional network. (e) A hub gene that was discovered in both the CC2 bulk segregant analysis and is differentially expressed in resistant individuals in both families. (f) Representative condition-specific sub-graphs that display both independent and common gene-interaction pairs between CC2 and CC5 families. (g) A representative sub-graph from the condition specific network analysis that shows a large subgraph of genes identified in the bulk segregant analysis.

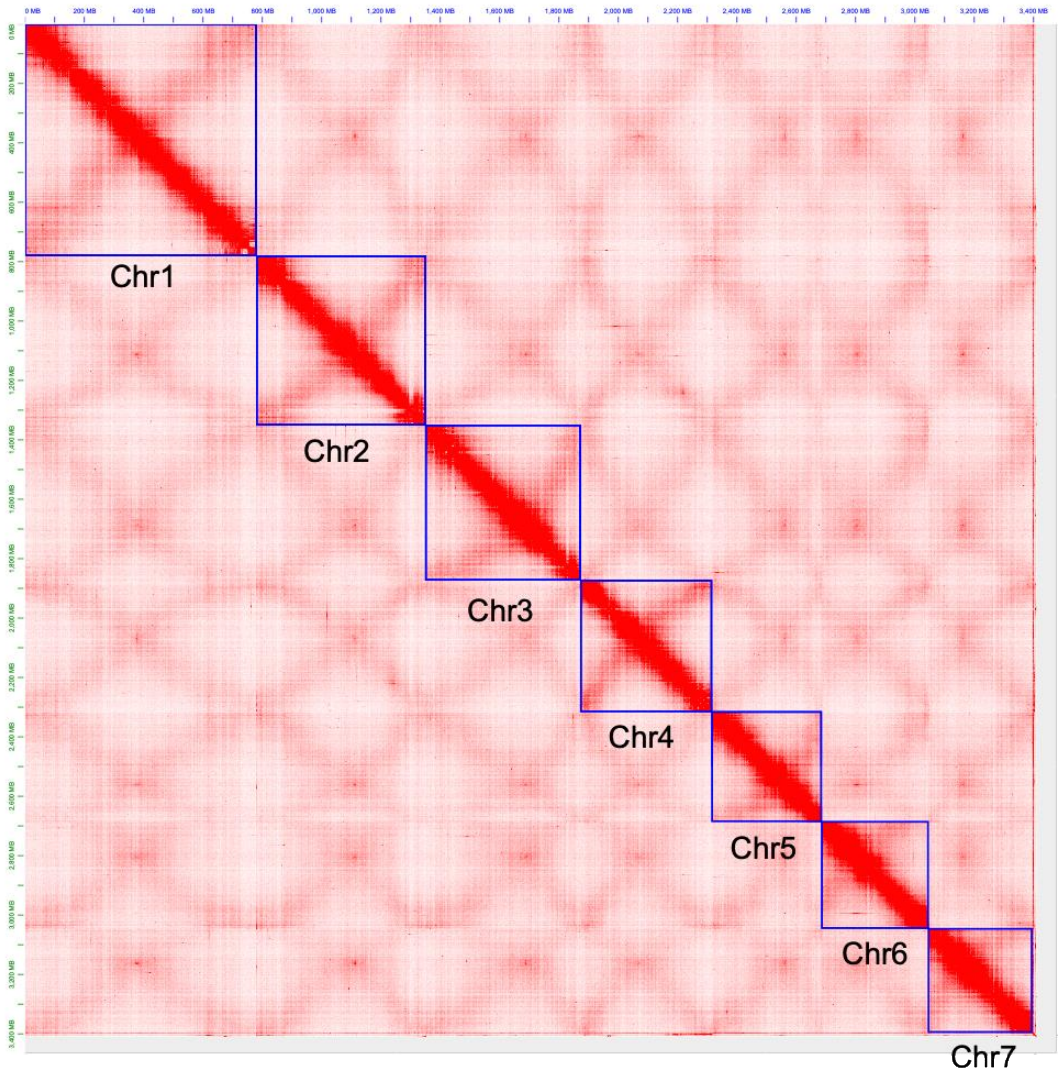


Fig. S9. Snapshot of Juicebox Assembly Tools output after manual correction.

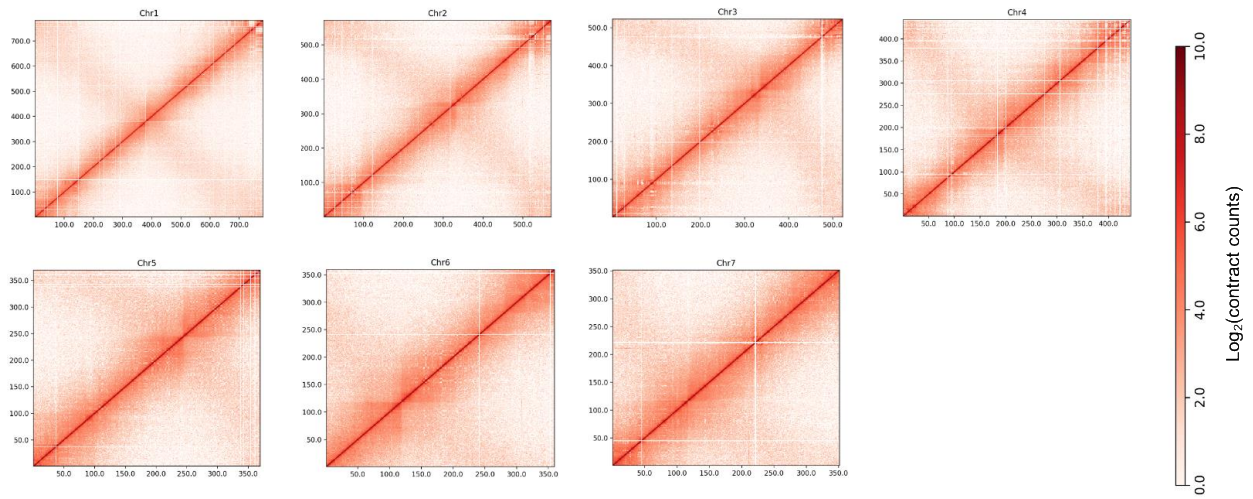


Fig. S10. Genome-wide analysis of chromatin interactions at 1-Mb resolution in the assembled blackgrass genome.

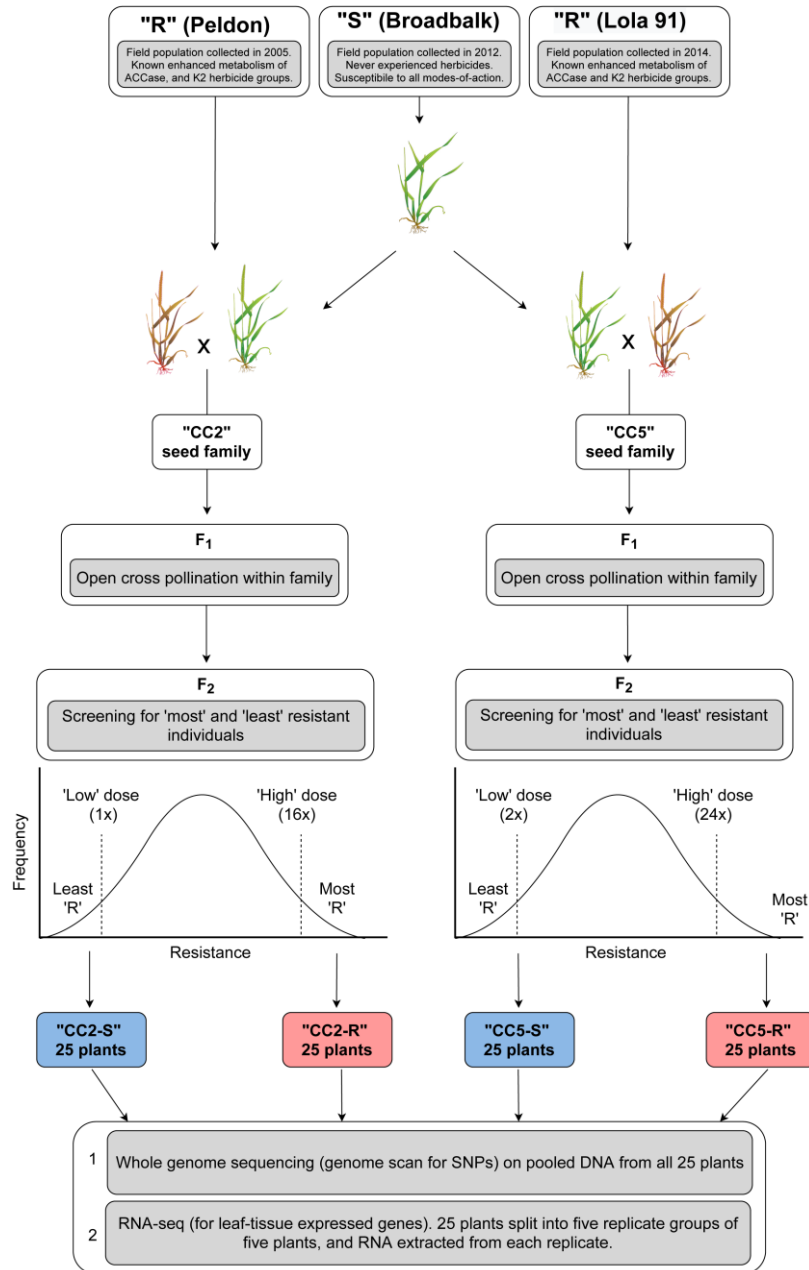


Fig. S11. Overview of the steps involved in creating the experimental seed families. Initially a single 'S' plant of the Broadbalk population was vegetatively cloned and cross pollinated with an individual from one of the two 'R' populations. Herbicide screening in the F₂ generation was performed using the ACCase inhibitor Fenoxaprop-P-ethyl. Tissue was collected from 25 plants identified as either 'R' or 'S' in each family, and used for whole genome sequencing and analysis of gene expression.

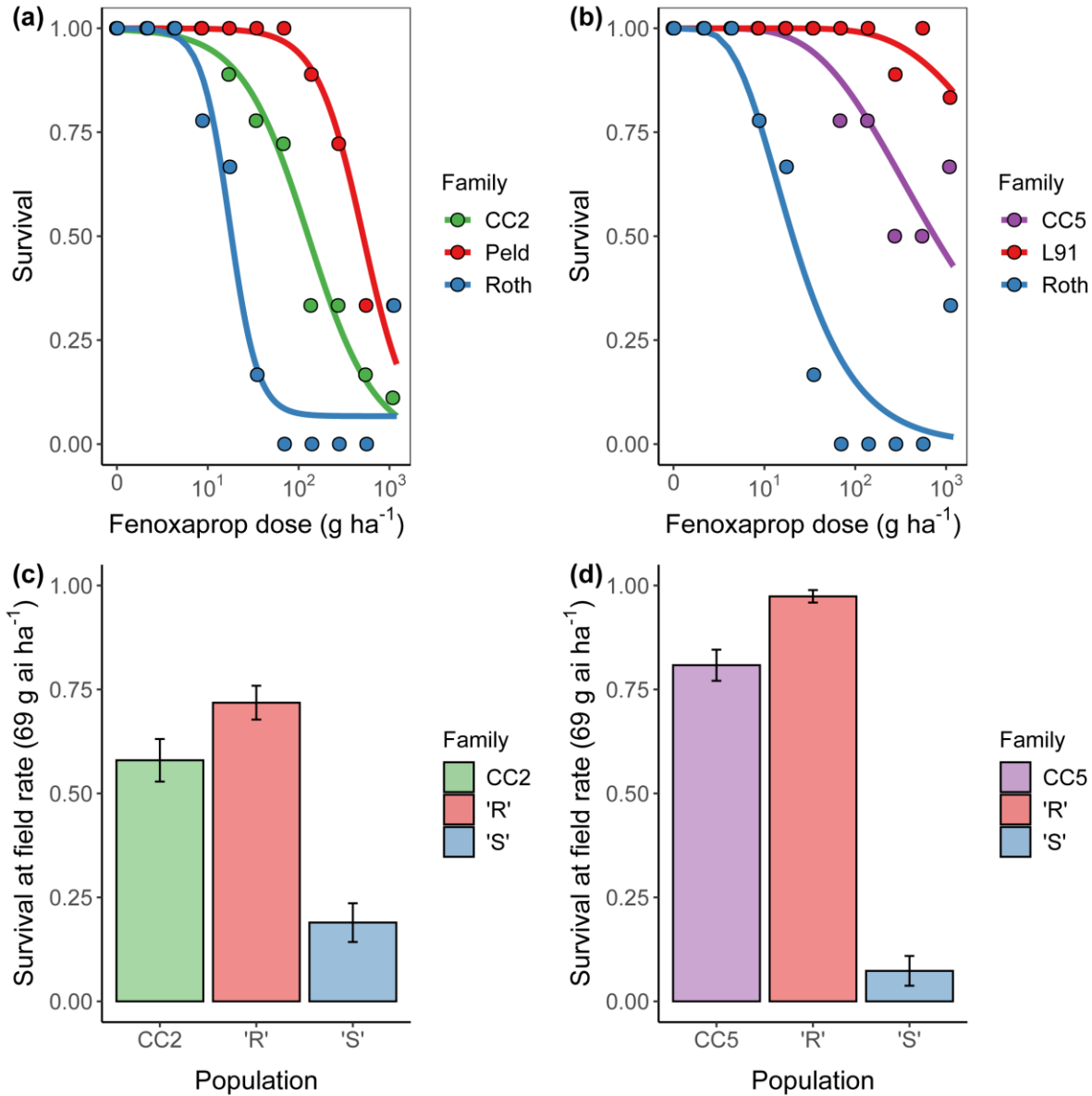


Fig. S12. Confirmation of resistance phenotype in the experimental seed families. **a** and **b** show the dose-response relationship of the two F₂ seed families (CC2 and CC5), relative to their respective parental populations. **c** and **d** show herbicide resistance of progeny from the 'R' and 'S' bulks, relative to the seed family from which they were derived. In all cases resistance was determined using a commercial formulation of the Acetyl CoA carboxylase inhibiting herbicide fenoxypop-P-ethyl.