This is a “preproof” accepted article for Weed Science. This version may be subject to change in the production process, *and does not include access to supplementary material*.
DOI: 10.1017/wsc.2024.79

Structured, Inbred and Plastic: the Genome and Population Genetics of the weed False cleavers (Galium spurium)

Short Title: *Weed Genomics: Galium spurium*

Sara L. Martin[1], Elizabeth Sears[2], Tracey James[3], Shaun Sharpe[4] and Breanne D. Tidemann[5]

[1]Research Scientist, Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Canada.

[2]Research Technician, Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Canada.

[3]Research Technician, Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, Canada.

[4]Research Scientist, (ORCID: 0000-0002-7683-703X), Saskatoon Research and Development Centre, Agriculture and Agri-Food Canada, Saskatoon, Canada.

[5]Research Scientist, (ORCID: 0000-0002-1070-2044), Lacombe Research and Development Centre, Agriculture and Agri-Food Canada, Lacombe, Canada.

Author for correspondence: Sara L. Martin, Research Scientist, (ORCID: 0000-0003-2055-6498) Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, 960 Carling Ave. Ottawa, Ontario, K1A 0C6 Email: sara.martin@agr.gc.ca

Abstract

False cleavers (*Galium spurium* L.) is an aggressive weed from the Rubiaceae. Here we assemble a chromosome scale draft of its genome, laying the foundations for determining the genetic basis of auxinic herbicide resistance and for systematic research into its polyphyletic genus. We use the genome to examine the population genetics in material from the Canadian Prairies and, in concert with a common greenhouse experiment, to examine whether the phenotypic variation observed in the field results primarily from genetic or environmental factors. The genome assembly covers approximately 85% of *G. spurium*'s expected 360Mbp genome size with 94% of BUSCO genes complete and most single copy (89%). Approximately 37% of the genome is repetitive elements and 35,540 genes were annotated using RNA-Seq data, including 100 homologs for genes involved or, potentially involved in, herbicide resistance. The genome shows strong synteny with other members of the Rubiaceae including smooth bedstraw (*Cruciata laevipes* Opiz) and robusta coffee [*Coffea canephora* (Pierre ex Froehner]. Double digested RADseq data for the 19 populations from the Canadian Prairies indicated that *G. spurium* has high levels of population structure (F_{ST} = 0.54) and inbreeding (F_{IS} =0.86) with low levels of hetrozygosity (H_O = 0.02) and nucleotide diversity (π = 0.0003). Variation in flowering time and seed weight largely overlapped among populations grown in the greenhouse. A redundancy analysis investigating genotype-phenotype associations showed few associations between SNP variation and these characteristics. In contrast, the majority of SNPs under selection were associated with mericarp hook density. This suggests that for most traits, environmental variation rather than genetic variation likely underlies phenotypic differences observed in the field. Several genes of interest including several homologs involved in the assembly of the Skp1-Culliun-F-Box IR1/AFB E3 ubiquitin ligase complex (e.g. CAND1, ECR1) are located in areas of the genome with evidence of selection and are targets for further investigation.

Key words: *Galium*, Rubiaceae, Weed Genome, ddRAD tags, Population Genetics, Auxin Resistance, False Cleavers

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

Introduction

The Rubiaceae is the fourth largest plant family and contains several economically important crop plants – including one essential for scientific research – coffee (*Coffea arabica* L.). *Galium* is the largest and widest spread genus within the tribe Rubieae with 600-700 species and a cosmopolitan distribution that primarily centers on temperate regions (Chen and Ehrendorfer 2001; Soza and Olmstead 2010). Although *Galium* species have been used as a coffee substitute (Turner and Szczawinski 1978), the main economic impact of the genus is as weeds, with 18 listed by Holm as weeds in jurisdictions around the world (Holm et al. 1991).

A particularly problematic species is false cleavers (*Galium spurium* L.) (2n = 20, 2C = 0.75 pg, 370 Mb), which can cause significant crop losses in cereals, canola (*Brassica napus* L.), and sugar beet (*Beta vulgaris* L.) (Malik and Vanden Born 1988). The large variability of *G. spurium* has resulted in many synonyms and misidentifications with other taxa of *Galium* and, adding to confusion and likely to morphological diversity, the species is considered to be a parental linage within the polymorphic polyploid complex with *Galium aparine* L. (Chen and Ehrendorfer 2001). This occasionally results in *G. spurium* not being recognized in keys for the genus, which exacerbates difficulties in distinguishing the species by limiting the number of locally relevant keys providing for this separation (Gleason and Cronquist 1991; Looman and Best 1979; Voss and Reznicek 2012). Differentiation of *Galium* species based on incompletely represented morphology can be extremely difficult. However, *G. aparine* is found in temperate zones around the world and at higher elevations in the tropics, often in non-agricultural settings such as moist deciduous woods. *Galium spurium* has a similarly broad distribution, but typically occurs in sunnier, more open and disturbed environments such as fence rows, roadsides, and waste grounds. It is considered to be the weedier and more aggressive of the two species (Malik and Vanden Born 1988). As the species climbs and then grows over top of crop species, it can cause lodging and difficulties with harvest while lowering yield. The similar seed size and shape also make it a serious contaminant of canola (Hall et al. 1998). The abundance of the species has rapidly increased across the Canadian Prairies since the 1970s and it is now one of the top 10 most abundant weeds (Leeson et al. 2005).

The evolution of herbicide resistance has added complications to the control of *G. spurium* with resistance to acetolactate synthesis inhibitors (ALS) and synthetic auxins such as

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

quinclorac reported (Beckie et al. 2012; Hall et al. 1998; Heap 2023). The genetic basis of ALS resistance is a point mutation at the target site. However, while the genetic basis of quinclorac resistance has been shown to be controlled by a locus distinct from the ALS point mutation and to be a single, recessive nuclear trait in material from Alberta, the locus involved has not yet been identified (Van Eerd 2004; Van Eerd et al. 2004). The spread of ALS resistance has been documented in the Canadian Prairie herbicide resistance surveys since 2007 (Beckie et al. 2013) but, the spread and extent of the auxin resistant biotype is currently unknown.

Here we assemble a chromosome level draft for the genome of *Galium spurium* and situate it in its evolutionary context by comparing its genome structure to that of its nearest relatives with available chromosome level genome assemblies: smooth bedstraw (*Cruciata laevipes* Opiz), blue field madder (*Sherardia arvensis* L.), baby lilac shrub (*Leptodermis oblonga* Bunge) and robusta coffee (*Coffea canephora* Pierre ex A. Froehner). We then use the genome to facilitate investigation of the species' population biology. Specifically, we examine the diversity and structure of 19 populations collected from Alberta and Saskatchewan, Canada using reduced genome representation (ddRADtags) (Peterson et al. 2012). We complement this genetic analysis with data from a common garden experiment conducted in the greenhouse to examine whether phenotypic variation observed in the field is the result of genetic or environmental variation. Finally, we examine homologs of genes known or suspected to be involved in the evolution of herbicide resistance and examine their proximity to SNPs showing evidence of selection. This work provides insights into the population genetics and genomics of this tenacious weed species while providing a foundation for future work examining the genetic basis of herbicide resistance and, more broadly, into the systematics of this complicated, currently polyphyletic, genus (Yang et al. 2018).

Methods

Plant Material and Phenotypic Measurements

Seed was collected in bulk from 19 locations, 11 in Alberta, Canada and 8 from Saskatchewan, Canada (Figure 1) and sent to the Ottawa Research and Development Centre in Ottawa, Ontario, Canada for growth and analysis. While initial plans were to use a random 'W' sample across fields selected for collection, drought conditions in the year of collection severely limited the availability of cleavers populations. As a result, the cleavers populations were

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

collected in bulk with no set protocol, aside from collecting enough seed from the population at each location to allow conductance of various planned experiments with the material. A voucher for the individual sequenced to chromosome level is available in DAO with the bar code: 01-01667380. The map of locations was produced in R (R Core Team 2021) (v. 4.1.1 "Kick Things") using the packages: maps (Brownrigg et al. 2022), rnaturalearth (South 2017), sf (Pebesma 2018), tidyverse (Wickham et al. 2019), and units (Pebesma et al. 2016).

Seeds were sown into Promix soil (soil, peat, and sand; 1:2:1; Promix, Rivière-du-Loup, Québec, Canada) in 10 cm diameter pots and placed in the greenhouse with day and night temperatures set to 20°C days and 18°C nights and a 16-hour light cycle. In the first experiment germination rates were tested, with twenty seeds per population placed on potting soil, covered lightly, and then scored for emergence weekly for one month. In the second experiment, a common garden experiment, 19-22 plants per population were grown and randomly assigned a location in the greenhouse and were re-randomized two weeks after planting, but were too large at four weeks for an additional round of randomization. These plants were grown in a common environment to evaluate their morphological and phenotypic characteristics while minimizing the effect of environmental variation. Plant height was measured using a measuring tape each week for three weeks starting when the first plants began flowering until the majority of the plants were too large and had to be folded over and tied to stakes. Plant height was measured from the soil surface to the most distal end of the longest branch. Flowering status was evaluated in three categories – 1) vegetative, 2) flowering but no seeds, and 3) seeds forming, and was recorded weekly for the three weeks over the same period. Flower and seed size were measured using a dissecting scope and digital calipers for three flowers or seeds of each of three individuals from each population. Plants were harvested after flowering by clipping the stems at the soil surface and placing them in a large paper bag. These bags were then placed in a drying oven at 30°C for 7 days prior to weighing until seed cleaning to keep the material dry. Total plant (Ohaus Voyager Pro Precision, Parsippany, NJ, U.S.) and total seed weight (Mettler-Toledo AT200, Columbus, OH, U.S.) as well as three replicates of 100 seed weight were measured. The total seed weight and average 100 seed weight were used to estimate seed production, unless fewer than 300 seeds were produced. In these cases, the seed were counted exactly. Seed trichome density was scored using a dissecting scope as naked, slightly hooked or very hooked (Figure 2).

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

Flow cytometry was conducted to verify the DNA content of each individual included in the greenhouse experiment analysis according to the procedure followed in the Martin Laboratory (Martin et al. 2017) except that individuals were quickly screened with a single run. Radish was used as the internal standard. The flowPloidy package (Smith et al. 2018) in R was used to analyze data from the flow cytometer.

Statistical Analyses of Phenotypic Data

The statistical analysis of phenotypic data was conducted with R. Specifically, phenotypic data were evaluated following box-cox transformation from the MASS package (Venables and Ripley 2002) using one-way analyses with the function oneway.test which uses Welch test's that do not assume equal variances across populations as fixed variable followed by Tukey's post-hoc test. The proportions of plants germinating, flowering, or with seed at different time points were analyzed with Pearson's chi-squared test for binary variables (chisq.bintest) with a Benjamini & Hochberg correction for the p-values to control family-wise error rates from the package RVAideMemoire (Hervé 2022). Exploration of the data and analyses were aided by functions in the car package (Fox and Weisberg 2019). Confidence intervals for proportions were estimated using the score method (Newcombe 1998). The function mulcompLetters from the multcompView package (Graves et al. 2019) was used to simplify the results of post-hoc pair wise comparisons for plotting. Additionally, functions from Hmisc (Harrell 2022) and plotrix (Lemon 2006) were used for visualization.

DNA and RNA Sampling and Sequencing

One individual was chosen for genome assembly and leaf tissue was collected on dry ice before extraction using the protocol published by Workman et al. (2018). High Molecular Weight DNA was extracted, following kit directions, using a Nanobind Plant Nuclei Big DNA kit (Circulomics Inc., Baltimore, MD, U.S.) followed by a Short Read Eliminator Kit (Circulomics Inc., Baltimore, MD, U.S.). DNA was first quantified with an Invitrogen™ Qubit™ (Thermo Scientific™, Waltham, MA, U.S.) using the dsDNA HS (High Sensitivity) assay according to manufacturer's instructions. Following quantification, DNA was then assessed for quality (impurities and fragment size) using a DropSense (Trinean, Pleasanton, CA, USA) and the Agilent TapeStation 4200 (Agilent., Santa Clara, CA, U.S.) on the Genomic DNA Screentape. Sequencing was completed using 4 Flow Cells on the Oxford Nanopore Technology

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

MinION system (ONT; Oxford Nanopore Technologies, Oxford Science Park, UK). DNA was also sent to Genome Quebec for sequencing of Illumina 150bp paired-end reads.

To facilitate annotation of the genome, RNA was extracted from rosette leaves. Tissue was first flash frozen in liquid Nitrogen and stored at -80˚C until ready for extraction. RNA extraction was performed on 50mg of rosette tissue using the Qiagen RNeasy Plant Mini Kit (Qiagen, Germantown, MD, U.S.) according to manufacturer's instructions, adding an extra wash step to ensure a thorough flushing of the column. RNA was first quantified using a NanoDrop™ One (Thermo Scientific™, Waltham, MA U.S.), then assessed for quality using a DropSense (Trinean, Pleasanton, CA, U.S.) and the Agilent TapeStation 4200 (Agilent, Santa Clara, CA, U.S.) on the RNA Screentape following manufacturer's instructions. After ensuring they met the required parameters for sequencing, three RNA samples were sent to BGI (BGI, San Jose, CA, U.S.) for mRNA library preparation and sequencing of 24 million 150 bp paired-ends per sample.

DNA was extracted using NucleoSpin-96 Plant II kit (Macherey-Nagel, Germany (supplied by D-Mark Biosciences) for all individuals grown in greenhouse experiment and sent to the University of Laval for ddRAD tag library creation and to Genome Quebec for sequencing (Peterson et al. 2012).

Packages in R were then used to analyze or visualize the data including functions from ape (Paradis and Schliep 2019), apex (Schliep et al. 2020), Biostrings (Pagès et al. 2021), geosphere (Hijmans 2022), ggplot2 (Wickham 2016), graph4lg (Savary et al. 2020), IRanges (Lawrence et al. 2013), msa (Bodenhofer et al. 2015), phangorn (Schliep 2011), plyr (Wickham et al. 2022), Rsamtools (Morgan et al. 2021), seqinr (Charif and Lobry 2007) and stringr (Wickham 2022).

Genome Assembly and Annotation

Long read ONT data (116X coverage) was called with Guppy (v 5.0) and assembled using CANU (v2.2) (Koren et al. 2017). Contigs that represented the chloroplast were removed based on identity scores and length from Mummer (v 4.0) (Marçais et al. 2018). The remaining contigs were checked to determine if they represented contaminates using Blobtools (v 1.1.1) (Laetsch and Blaxter 2017). However, no appreciable contamination was detected. This assembly was then polished three times using the Illumina data (113X coverage) using Pilon

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

(v1.23) (Walker et al. 2014) after alignment using bwa (0.7.17) (Li and Durbin 2009). The polished assembly was scaffolded into a chromosome level assembly by Phase Genomics (Seattle, WA, U.S.) using chromosomal conformation data (HiC) after processing DNA with a Proximo Hi-C 2.0 Kit in the Martin Laboratory using their propriety pipeline. Mummer and the Whole-Genome Duplication Integrated analysis tool kit (WGDI v 0.6.1) (Sun et al. 2022) were used to compare *G. spurium*'s genome to that of *C. laevipes* (GCA_963678965.1), *S. arvensis* (GCA_948330725.1), *L. oblonga* (Guo et al. 2021), *C. canephora* (AUK_PRJEB4211_v1) and *Asclepias syriaca* L. (GCA_027405835.1). The genome assembly of *G. spurium* was assessed using QUAST v5.1.0rc1 (Gurevich et al. 2013), BUSCO (5.4.2 - with the embryophyte_odb10 dataset of 2192 genes) (Simao et al. 2015), and Samtools v1.9 (Li et al. 2009). MCScanx (Wang et al. 2012) was used to produce files for analysis with WGDI. The genome assembly is available on NCBI (bioproject PRJNA1143567), on the International Weed Genomics Consortium online database WeedPedia (https://www.weedgenomics.org/species/galium-spurium/), and from the corresponding author on reasonable request.

RNA sequence data was trimmed, cleaned and filtered using SOAPnuke (v2.1.5) (Chen et al. 2018) and aligned to the genome using HISAT2 (v2.2.1) (Kim et al. 2019). Ninety-nine percent of the 72 million read pairs mapped to the genome. RepeatModeler (v2.0.3) (Smit and Hubley 2008-2015) and RepeatMasker (v4.1.2.p1) (Smit et al. 2013-2010) were used to generate a masked version of the genome. The BRAKER pipeline (v 1.9) (Brůna et al. 2021; Hoff et al. 2019) was used to produce annotations based on this masked genome and the hints files generated by HISAT2. This pipeline relied on AUGUSTUS (Stanke et al. 2008), GeneMark-ET (Lomsadze et al. 2014), and Samtools. This procedure was also followed to add annotations to the genome of *L. oblonga* using RNA-seq data downloaded from the National Center for Biotechnology Information's Sequence Read Archive (SRR9839476) to allow comparative analysis with *G. spurium* using WGDI and GENESPACE (Lovell 2023; Lovell et al. 2022). GENESPACE uses other tools including OrthoFinder (Emms and Kelly 2018; Emms and Kelly 2019), which constructed the phylogeny of the species included here using the STAG method. Following masking, AUGUSTUS was used to annotate the genomes of *C. laevipes*, *S. arvensis* and common milkweed (*Asclepias syriaca* L.) with tomato (*Solanum lycopersicum* L.) as the model.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

Repetitive DNA was annotated using the Extensive *de novo* TE Annotator (EDTA v 1.8.3) (Ou et al. 2019), which relies on LTRHarvest (Ellinghaus et al. 2008), LTR_Finder (Xu and Wang 2007), LTR_retriever (Ou and Jiang 2018), TIR-Learner (Su et al. 2019), Generic Repeat Finder (Shi and Liang 2019), HelitronScanner (Xiong et al. 2014) and TEsorter (Zhang et al. 2019). LTRHarvest, LTR_Finder and LTR_Retriever were also used to calculate the LTR Assembly Index (LAI) (Ou et al. 2018). The genome sequence was also analyzed to detect Helitrons with EAHelitron (Hu et al. 2019).

Genes that could be involved in the evolution of herbicide resistance were identified through homology using blastp, the amino acid sequences of the genes identified during the BRAKER pipeline, and amino acid sequences of the genes of interest downloaded from GenBank or The Arabidopsis Information Resource (tair) (Supplementary Table 1). A cutoff 1-e50 was used to identify the best hits from these searches of the annotated proteins with score considered as a secondary indication to eliminate hits which had much smaller scores (<50%) than the best hit for the reference protein. BLAST was run on the coding sequence for the protein (Supplementary Table 2) to cross check that the best hits (Supplementary Table 3) with identifications indicated that the protein of interest had been likely been identified.

SNP Calling and Population Genetics Analyses

SNP calling was completed for the RAD Seq data using the reference-based pipeline in STACKS (V 2.6) (Catchen et al. 2013) after alignment to the genome using bwa with the parameter (-L) for reducing soft clipping set to 500 (Li and Durbin 2009). Following the creation of the variant call file (vcf) by STACKS, data were further filtered using VCFtools (0.1.16) (Danecek et al. 2011) using the recommendations of O'Leary et al. (2018) to create two datasets for further analysis. STACKS initially identified 25K loci covering 6.7 million bp of the genome with 86K variants. The first filtered data set included only genotypes with a read depth of greater than 5, loci with a mean read depth over 15, SNPS with a quality greater than 20, and SNPs with a minimum allele count over 3. These data contained just over 40K SNPs for 365 individuals with less than one percent missing and were used for examining genome wide nucleotide diversity. This 40K set was further filtered for missingness following the O'Leary protocol (O'Leary et al. 2018) and thinned so that SNPs were at least 500 bp apart to reduce linkage. This second data set contained just under four thousand SNPs with less than one percent missing for

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

337 individuals and was used for examining population structure, the association between genetic and morphological variation and evidence of selection. VCFtools was also used to calculate nucleotide diversity with the remaining population summary statistics calculated with hierfstat (Goudet and Jombart 2022). A Mantel test (mantel.randtest::ade4) (Thioulouse et al. 2018) was used to determine if there was a relationship between genetic and geographic distance and redundancy analysis (vegan::rda) (Oksanen et al. 2022) was used to determine whether the morphological variation observed in the greenhouse was associated with variation within the SNPs. The R package pcadapt (v 4.3.5) was used to examine linkage disequilibrium and to look for evidence of SNPs in regions under selection with the final analysis using K=5, linkage disequilibrium parameters set to size =200 and threshold = 0.1, and a minimum minor allele frequency of 0.02 (Luu et al. 2017; Privé et al. 2020). This method examines association between population structure and genetic variants with the assumption that outlying variants are indicative of local adaptation (Privé et al. 2020). Tools for population genetics and genomics in R were used to analyze and visualize the SNP data including functions from the following packages: adegenet (Jombart 2008), circlize (Gu et al. 2014), colorspace (Zeileis et al. 2020), dartR (Mijangos et al. 2022), Gviz (Hahne and Ivanek 2016), gwscaR (Flanagan 2023) , magrittr (Bache and Wickham 2022), pegas (v1.1) (Paradis 2010), phytools (v1.2-0) (Revell 2012), png (Urbanek 2022), poppr (Kamvar et al. 2014), pinfsc50 (Knaus and Grünwald 2016), psych (Revelle 2023), qvalue (Storey et al. 2022), VariantAnnotation (v1.42.1) and vcfR (Knaus and Grünwald 2016).

Results and Discussion

With the evolution of herbicide resistance, it is becoming increasingly important to be able to correctly identify weed species, understand their biology well enough to enable integrated management techniques, and understand the genetic basis of herbicide resistance so that individuals with these resistances can be identified. As a result, having the genetic tools to accomplish these goals is key. *Galium spurium* is a weedy species that grows up and over crops, causes lodging, reduces yield quantity and quality, and creates difficulties with harvest (Hall et al. 1998). It has developed herbicide resistance and belongs to a large genus of difficult to identify species. Here we produce a chromosome level assembly of *G. spurium*, annotate it with RNA-Seq data, then use the genome in investigations of population genetics and to examine

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

patterns of selection across the genome. This represents a foundational step towards identifying the genetic basis of auxinic herbicide resistance and towards clarifying evolutionary relationships within *Galium*.

Genome Assembly and Annotation

The flow cytometry estimates for 2C DNA content for *G. spurium* accessions grown in the greenhouse ranged from 0.71 – 0.74 pg with an average 0.72pg (sd ± 0.02) and indicated a haploid genome size of approximately 360Mbp. The chromosome level draft sequence produced for *Galium spurium* covers 85% this expected genome size and includes 94% of the expected core eukaryotic genes. The initial assembly of the *G. spurium's* genome from the long read ONT data (116X coverage) using CANU resulted in a draft assembly with 290 contigs with an NG50, the sequence length of the shortest contig at half the expected genome size when combined with the larger contigs, of 10.7 Mb and total assembly size of 318.5 Mb. A total of 128 of these contigs were placed on ten scaffolds by Phase Genomics using Hi-C data resulting in a chromosome level assembly (Table 1). After polishing the assembly with PILON, analysis with BUSCO indicated 94% of the expected genes were complete with the vast majority (89%) represented in a single copy and a minority (5%) found to be duplicated. Just under 150 million RNA data reads were used to annotate the genome with 98% of these aligning to the genome. The BRAKER pipeline identified and annotated 35,549 genes based on this data with 32,269 genes annotated using the same pipeline and downloaded data for *L. oblonga*. Repeat annotation using EDTA indicated that 37% of the genome was comprised of repetitive elements with the largest proportion of these identified as class 1 retrotransposons of the "Gypsy" type (11.5%) (Figure 3). Helitron density was assessed by EAHelitron as 6.7 and the LTR Assembly Index (LAI) was 18.27 indicating that the genome meets the level of completeness for a reference level genome (Ou et al. 2018).

Comparative Genomics

The genome of *G. spurium* has ten chromosomes, but for the genus, other genera in the Rubiaceae, and in sister families such as the Apocynaceae, the more common base chromosome number is n = 11. The Rubiaceae is divided into three subfamilies: Rubioideae, which includes *Galium*; Ixoroideae, which includes *Coffea*; and Cinchonoideae, which includes *Cinchona*

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

(sources of quinine) (Bremer 2009; Stevens 2017). *Galium*, *Cruciata* and *Sherardia*, are all within the tribe Rubieae (Rubioideae), while *Leptodermis* is in a different tribe - the Paederieae (Rubioideae). As a result, *Cruciata laevipes* (n=11) and *Sherardia arvensis* (n=11) are the currently available chromosomal level assemblies that are most closely related to *Galium spurium* with this chromosome number followed by *Leptodermis oblonga* (n=11) and then, more distantly by *Coffea canephora* (n=11) and *Asclepias syriaca* (n=11), which is an out group from the Apocynaceae (Figure 4) (Stevens 2017).

The genome of *G. spurium* shows strong synteny with the genomes of the other species from the Rubiaceae (Figure 5, 6) including *C. canephora*, despite the divergence of their genera approximately 68 MYA (Kumar et al. 2022) (Figure 6B). Three chromosomes are largely syntenic when the genomes of *C. laevipes* and *G. spurium* are compared; *G. spurium's* chromosomes 6, 7, and 10 show strong synteny with *C. laevipes 's* chromosomes 5, 7, and 9 respectively. However, there is no clearly identifiable mechanism, neither end to end joining nor nested chromosome fusion (Lysak 2022), that contributed to chromosome number reduction in *G. spurium* (Figure 6A). Rather the genes from the chromosomes of this closest assembled relative are distributed among multiple chromosomes in *G. spurium*. For example, genes from *C. laevipes*'s chromosome 3 (cl03) are distributed among five of *G. spurium*'s chromosomes with multiple reciprocal translocations required explain the current structure. This indicates, that though the genome have strong synteny across the entirety of some chromosomes (e.g. cl05 (or cc01) and gs06), more information is needed to reconstruct the evolution of *G. spurium*'s current chromosome structure. Comparison to other genomes in *Galium* or *Asperula*, would likely be informative for understanding chromosome evolution of *G. spurium*.

Population Genetics

The overall nucleotide diversity (π), the average number of pairwise differences between individuals, averaged 3×10^{-4}. Across all populations 83% of individuals were homozygous for the major allele and 16% were homozygous for the minor allele (Figure 7A). Populations had low observed heterozygosity (H_O) with a value of 0.02 overall (Figure 7B). The F_{ST} for the populations was 0.54 with pairwise values ranging from 0.15 to 0.79 (Figure 7C) indicating strong population differentiation (Conner and Hartl 2004) and there were many more homozygotes ($F_{IS} = 0.86$) in the population than expected given random mating. Overall,

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

Provesti's genetic distance between individuals averaged 0.23. The number of alleles that were identical by state and shared between two individuals within a population averaged 88% and the values only a little lower, at 75% when individuals from different populations were compared. This level of genetic diversity and homozygosity is most similar to that of the germplasm of genetically depauperate crop species. For example, nucleotide diversity was found to be $\pi = 2.3 \text{ X } 10^{-4}$ for 736 accessions of coffee (*Coffea arabica* L.) (compare with to lower ploidy ancestors: C. *canephora* Pierre ex A. Froehner $\pi = 2.6 \text{ X } 10^{-3}$ and *C. eugenioides* S. Moore $\pi = 1.1 \text{ X } 10^{-3}$), major allele frequences that were largely greater than 95% and a deficiency of heterozygotes - likely as a result of reductions in diversity following an origin involving a single allopolyploidization event (Scalabrin et al. 2020). The levels of heterozygosity in *G. spurium* is similar to that of 376 tea accessions (*Camellia sinensis* (L.) O. Kuntze) from the Rwebitaba Tea Research Centre where H_O was 0.06 (Tadeo et al. 2024) and 31 accessions of peanut (*Arachis hypogaea* L.) from Taiwan had major allele frequences of 0.87 and an average genetic distance of 0.17. This suggests a high rate of self-fertilization, inbreeding, low gene flow among populations and low genetic diversity.

This high level of population structure and inbreeding has likely contributed to the relatively slow spread of ALS and quinclorac resistance across the Canadian Prairies. For example, the rate of spread of ALS, glyphosate, or auxinic herbicide resistance in the wind pollinated, outcrossing species Kochia (*Bassia scoparia* (L.) A. J. Scott) with low population structure ($F_{ST} = 0.01$) (Martin et al. 2020) has been rapid, increasing from 4% of sampled sites (Hall et al. 2014) to 78% of sites in ten-years (Geddes et al. 2023). In contrast, ALS resistant *Galium* was first documented in 1996 in Alberta, Canada (Hall et al. 1998), field surveys from 2007 to 2011 and then 2012 to 2017 found increases in Alberta, Canada from 17% to 44%, but lower, more constant levels in Saskatchewan (20%) and Manitoba (17%) (Beckie et al. 2020; Beckie et al. 2013). The structured nature of these population is also reflected in 1) a significant Mantel test indicating a correlation ($p < 0.001$) between geographic distance and Provesti's genetic distance (dist.genpop::adegenet; Figure 8), 2) in 55% of the variation being explained by population (44% by individual) in an AMOVA, and 3) the clumping of populations as visualized in the PCA of molecular variation (Figure 9). The low genetic diversity of these populations may also limit the potential of the populations to evolve resistances in response to selection from

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

herbicides without recruiting additional variation from processes such as mutation, hybridization, and polyploidization.

Phenotype

A striking aspect of *G. spurium* populations in the field is the level of morphological variation observed. In the common greenhouse environment, while the characteristics we measured showed significant variation, the majority of the populations were statistically equivalent (Figure 10, 11). A total of 384 plants were included in the common garden greenhouse experiment with 19 – 22 individuals representing each of the 19 populations. All accessions had the morphological characteristics expected for *G. spurium*, specifically green flowers and small seeds (mericarps), rather than the expected morphological characteristics for *G. aparine* of white flowers and larger mericarps. Some individuals from one accession were shorter with little distance between internodes and narrower leaves giving them a shorter "mossy" habit, however the flowers, mericarps, and DNA content were all consistent with the other individuals grown here and other members of the same accession had more typical morphology suggesting this is part of the variation for the species.

In the germination experiment, on average 51% of the seeds (20 per population) had emerged after seven days (Figure 10A). Variation was seen by population with AB-06 showing faster germination (90%) than ten other populations including the slowest to germinate SK-08 (15%), but most populations showed overlapping confidence intervals and were not statistically different because of high variability. Six weeks after planting an average of 43% of the plants had begun flowering (Figure 10B). Again, there was statistically significant variation among populations with AB-02 and AB-06 having significantly more flowering individuals than thirteen populations including AB-09, AB-10 or SK-08, but most of the populations were not statistically different. Similarly, after 8 weeks, at the end of the greenhouse experiment, the majority of populations had a high proportion of plants that had set seed (79%) and, while population AB-10 had statistically fewer plants with seed than most other populations, most populations were not statistically distinct (Figure 10C).

The morphological characteristics measured showed less variation than the phenological characteristics, but still showed few statistical differences overall. The average height of the plants at eight weeks was 107.2 cm with the tallest population, AB-04, averaging 132.0 cm and

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

the shortest, AB-10, which averaged less than half of this at 53.8 cm (Figure 11A). Total estimated seed production averaged 3,158 with AB-01 and SK-01 producing the most seed with over 3,500 seeds each, and AB-10 the least with half of this amount (1,720, Figure 11B). Average diameter of a mericarp was 1.8 mm with the populations with largest such as AB-04 (2.04 mm) statistically different from the smallest SK-06 (1.54 mm; Figure 11C). However, as with germination, flowering, and seed production, the majority of populations were not statistically different from each other for height, total seed production or seed size. Flower size did not differ statistically across populations and averaged 1.9 mm (0.18 mm standard deviation).

A redundancy analysis (RDA) was used to investigate the association between genotype and the phenotype observed in the greenhouse. Five characteristics measured in the greenhouse were retained for the analysis after eliminating continuous variables that had strong correlation to each other and factors that had too few individuals in each category. The retained characteristics were height and flowering status at four weeks, total plant weight, seed count, and the density of hooks on the mericarp. The RDA indicated these predictors explained (R^2) 5.5% of the variation in the data after adjustment and the model was significant according to a permutation test (1000; $p < 0.001$). A total of 34 candidate SNPs with RDA loadings 3 standard deviations or more from the center of their distributions were examined. This minimized false positive rates for SNPs of interest by setting the two-tailed p-value at 0.0027. Of these 34 SNPs the majority (24) were associated with the density of hooks of the mericarp (Figure 2, 12). Among the remaining SNPs with high loadings, 9 were associated with total plant weight at harvest and one with early flowering time. This indicates that the majority of these SNPs are likely to be under selection as a result of association with the density of hooks on their mericarps. Individuals that were scored as without hooks or with a low density of hooks occurred sporadically across the populations, but were most frequent in populations SK06, SK07, and SK08 (Figure 1). Variation in the density of spines on the mericarps of *G. spurium*, has been previously described and used to separate individuals into two forms: *G. spurium* f. *spurium* with smooth fruit, and *G. spurium* f. *vaillantii* with hooked spines (Moore 1975). Moore (1975) indicates that Linnaeus's specimen of the species was the smooth form, but that the smooth form was less common in the material examined from the Canadian Prairies. Further, he noted that one of the two sites with smooth form specimens was near Melfort, SK, close to where our populations with the most smooth fruited individuals were collected. Malik and Vanden Born (1988) noted that they observed an

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

intermediate variant in material from Alberta and Saskatchewan and that the presence of more than one form in a field was common. The presence of the hooked spines on the mericarp are inferred to provide an advantage for dispersal by animals (Malik and Vanden Born 1988). More work is needed to understand the genetic basis of hooks on these fruits and how the character influences the fitness of these populations, but the species may offer an excellent opportunity for this work as the development of nearly isogenic lines that differ for this character should be possible.

The data from the common environment, combined with the result of the population genetics above suggests that some of the disjunct morphological variation observed within populations is likely the result of a combination of the inbred nature of groups of individuals observed within the same field and environmentally induced variation. Differences among individuals in a field such as a lack of hooks on the mericarp, likely reflects homozygosity for an alternative allele. However, given the minimal amount of genetic variation observed overall and the strong genetic similarity among samples even from different populations, much of the variation in characteristics such as emergence and flowering time is likely attributable to environmental variation given differing conditions rather than genetic differences.

Potential and Selection Herbicide Resistance Genes

Management of *G. spurium* in western Canadian crops has been primarily reliant on herbicides from the following modes of action: ALS inhibitors (Group 2), synthetic auxins (Group 4), 5-nolpyruvylshikimate-3-phosphate synthase (EPSPS) inhibitors (Group 9), glutamine synthetase inhibitors (Group 10), and more recently on PPO inhibitors (Group 14), and 4-hydroxyphenylpyruvate dioxygenase (HPPD) inhibitors (Group 27) (Anonymous 2022). In some crops, such as some pulse crops, the ALS inhibiting herbicides are nearly the only option. Resistance to ALS inhibitors has spread through the populations of *G. spurium* on the Canadian Prairies though it has not yet reached fixation (Beckie et al. 2020). The acetolactate synthase (ALS) proteins from *G. aparine* and *Coffea arabica* mapped to *G. spurium*'s chromosome gs06. The amino acid sequence captured in this genome assembly has one of the point mutations, Tryptophan 574 to Leucine, reported to confer ALS resistance in *G. aparine* by Deng et al. (2019). No changes are observed at residue Pro197 or Asp376, which are the two other sites Deng et al. (2019) reported to confer ALS resistance in *G. aparine*. Two other changes in the

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

amino acid sequence are present compared to the susceptible sequence in GenBank (Sun et al. 2011): Thr59Ala and Ala380Thr, but these sites do not correspond to any of the other sites (Ala122, Ala205, Arg377, Ser653 or Gly654) reported to confer resistance in other weed species (Tranel et al. 2023). We used the R package PCAdapt to detect SNPs under selection and compared them to the location of the genes of interest (Luu et al. 2017; Privé et al. 2020). Five principal components were chosen based on the scree plot and alpha was set to 0.01 for a false discovery rate of 1:100 after correction with the Benjamini-Hochberg procedure. The SNPs were thinned to reduce the effect of linkage disequilibrium with size set to 200, threshold set to 0.1, and the minimum minor allele frequency set to 2%. A total of 192 SNPs were identified as outliers and their position was compared to the position of the genes of interest for involvement in herbicide resistance (Figure 13; Table 2; Supplementary Table 4). Two outlying SNPs were found within 0.5 Mbp of the ALS gene on gs06, a potential indication of the selection on the locus that is occurring with the ongoing spread of the ALS resistance.

In contrast to ALS resistance, resistance to quinclorac, a synthetic auxin, has not yet become widespread in these populations (Beckie et al. 2020; Van Eerd 2004; Van Eerd et al. 2004), though surveying has been limited. There has been a lower global incidence of synthetic auxin herbicide resistance compared to ALS inhibitors (Heap 2023). This is likely a consequence the importance of auxin, which should be viewed, not as a hormone, but as an ancient, complex, connective, impetus signal with self-organizing transport streams of indole-3-acetic acid (IAA) (Zažímalová et al. 2014). The target site of synthetic auxins such as quinclorac is still being elucidated with the complexity, flexibility, and redundancy of the auxin pathway complicating our understanding of both its normal and abnormal functioning (Todd et al. 2020; Zazimalova et al. 2010). Synthetic auxins in susceptible plants are highly stable IAA mimics that lead to an "auxin overdose" (Grossmann 2010). The cause of demise in *G. spurium* is suspected to be the production and accumulation of reactive oxygen species including hydrogen peroxide (H_2O_2) and resistance, controlled by a recessive nuclear gene, has been hypothesized to be the result of a mutation at a target site or along the signal transduction pathway (Van Eerd et al. 2005).

One critical auxin receptor is TIR1, an F-box protein found in the nucleus (Dharmasiri et al. 2005a; Kepinski and Leyser 2005). Auxins regulates gene expression by acting as "molecular glue", filling the cavity between TIR1 and the substrate by forming a continuous hydrophobic

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

core among the three (Tan et al. 2007). Multiple homologues of TIR1 have been identified (Dharmasiri et al. 2005b) and Arabidopsis mutants for auxin signaling F-box protein 5 (AFB5) have demonstrated differential resistance to picolinic acids (picloram, clopyralid) but no other synthetic auxins (Walsh et al. 2006). TIR1/AFB are the substrate recognition subunit of $SCF^{TIR1/AFB}$, an E3 ubiquitin ligase complex which catalyses the conjugation of ubiquitin to Aux/IAA repressor proteins, which then is targeted by the 26S proteasome for degradation (Quint and Gray 2006). Once the AUX/IAA repressor proteins are degraded, auxin response factor (ARF) transcription factors can bind to auxin responsive genes in the DNA (Li et al. 2016).

The herbicide 2,4-D was found to similarly influence the binding pocket of TIR1 as IAA, which promotes SCF ubiquitin-ligase-catalyzed degradation of AUX/IAA (Tan et al. 2007). Transcription of ACC synthase (*ACS*) genes are upregulated with auxins, shown with an antisense construct in transgenic tomato (Grossmann and Schmülling 1995). In *G. aparine,* this was shown to promote an upregulation of *NCED* gene expression (Kraft et al. 2007), ACC synthase activity, ACC, and ethylene production, where ethylene also upregulated ABA levels (Hansen and Grossmann 2000). It has been suggested that the ethylene-induced increase in ABA promotes growth inhibition and senescence (Grossmann 2010; Grossmann and Hansen 2001). In peas (*Pisum sativum* L.) treated with 2,4-D, it was shown that reactive oxygen species were accumulated, leading to protein and lipid membrane damage (Pazmiño et al. 2011). The authors also found that there was differential tolerance based on staging, where mature tissue was more tolerant than reproductively immature tissue.

In their recent review Todd et al. (2020) suggest that target-site mutations in auxin perception and signaling could involve three major groups of auxin signaling proteins: 1) auxin efflux and influx transporters (PIN, ABCB, AUX/LAX), 2) auxin perception and signaling proteins: auxin response factors (ARF), transcriptional repressors (AUX/IAA), and the Skp1-Culliun-F-Box TIR1/AFB ubiquitin ligase complex (AXR1, ECR1, RCF1, HSP90/SGT1, RB, NEDD8, CAND1, COP9, CUL1) and 3) the transmembrane kinase family (TMK). The ARF proteins compete for promoter target sites of auxin response genes and dimerize with AUX/IAA repressor proteins at low auxin concentrations while, when auxin concentrations are high, AUX/IAA repressor proteins interact with TIR1/AFB, the AUX/IAA proteins become

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

ubiquitinated and ultimately degraded via the 26S proteasome pathway (reviewed in Luo et al. (2018)). In Arabidopsis mutants, loss of TMK activity affected auxin signal transduction in root and potentially shoot development, cell expansion in roots, hypocotyls, and stamen filaments, and leaf cell expansion and cell proliferation (Dai et al. 2013). Additionally, a mutation in the auxin receptor auxin binding protein 1 (ABP1) was identified in Group 4, synthetic auxin, resistant wild mustard from Manitoba (Grossmann 2010; Zheng and Hall 2001).

Multiple homologs for these auxin efflux and influx transports were identified in the *G. spurium* genome: four of the PIN-FORMED auxin efflux carrier protein (PIN1), three of the AUXIN1/LIKE-AUX1 (AUX/LAX) auxin influx carrier proteins, and 13 of the ATP-Binding Cassette B proteins (ABCB) (Supplementary Table 5). Nine proteins with homology to the transcriptional repressors, indole-3-acetic acid inducible (IAA), nine with homology to auxin response factors (ARFs) and six with homology to both were also found (IAA/ARF). Five homologs of TIR1/AFB were found and at least one homolog of most of the genes involved in the formation of the Skp1-Cullin-F-Box TIR1/AFB E3 ubiquitin ligase complex were located, as was one homolog of ABP1 (Supplementary Table 5). Four homologs of TMK were also located. We note that the annotation of the genome will not be complete as there will be genes that were not expressed in the rosette at the stage or which have not have been annotated correctly by the pipeline. Additional RNA work could improve these annotations and may indicate additional homologs.

Fifteen genes associated with synthetic auxin resistance (Group 4) (Table 2; Figure 13) are among the top twenty gene homologs with known or potential roles in herbicide resistance that contain SNPs identified as likely under selection within 1 Mbp. This includes nine genes associated with auxin perception: 1) five outlying SNPs of a AUX/IAA repressor protein homolog on gs06, 2) four SNPs each near CAND1 and ERC1 (components for Skp1-Cullin-F-Box TIR1/AFB E3 ubiquitin ligase complex) on gs02 and gs05, respectively, 3) three SNPs close to an ARF homolog on gs08, 4) 1 SNP near ARF on gs02, 5) 1 SNP near IAA/ARF on gs08, 6) 1 SNP near TIR/AFB on gs06, and 7) 1 SNP near AUX/IAA on gs01 (Table 2). For auxin transport genes, five homologs were identified: 1) four SNPs were found close to the auxin influx transporter (AUX/LAX) gene on gs08, 2) 2 SNPs were found close to AUX/LAX on gs04, 3) 2 SNPs were located near ABCB on gs06, 4) 1 SNP near a PIN homolog on gs01, and 5)

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

another SNP near a second PIN homolog on gs01. Lastly, 2 SNPs were located near the TMK gene on gs08 and 2 SNPs were located near ABP1 on gs06 (Table 2). SNPs showing evidence of selection are not necessarily contained in a gene under selection, but rather can be linked to nearby genes under selection. This means that, while this analysis indicates regions of the genome where selection is occurring, the region contains numerous genes making conclusive determinations that selection is occurring on a specific gene impossible, nor can we determine what is causing that selection. More specifically, a SNP showing strong selection near known herbicide resistant genes could be indicative of selection on that gene or on other nearby genes and may or may not be related to the evolution of herbicide resistance. As a result, while these SNPs cannot be conclusively attributed to selection on these genes, they can motivate further, more targeted research, such as differential expression analysis in herbicide resistant and susceptible individuals. This chromosome level assembly provides a fundamental resource for this analysis.

One or two homologs of the genes targeted by the remaining herbicides used to control *G. spurium* (EPSPS inhibitors, s glutamine synthetase inhibitors, PPO inhibitors, and HPPD inhibitors (Anonymous 2022)) were also found in the assembly. The predicted EPSPS protein was found as a single copy on gs04 and, as expected from a lack of reported resistance, did not show evidence of carrying a point mutation at Thr102, Ala103, or Pro106 (Gaines and Heap 2023). A single homologs of protoporphyrinogen oxidase 2 (PPO2) was identified on gs02. The amino acids where changes are reported to potentially confer PPO2 resistance (Asn98, Arg128, Gly383, Leu385, Leu398, Gly399, Tyr400, Leu401, and Phe421) are all conserved in this genome sequence (Rangani et al. 2019). Two glutamine synthetase homologs were identified on gs03 and gs10 with each single candidate SNPs within 1Mbp of each of their positions. A single homolog of HPPD was identified on gs02, however, while *Amaranthus* species and wild radish (*Raphanus raphanistrum* L.) have evolved HPPD resistance the mechanism for this resistance appears to be a change in the plant's metabolism or the gene's expression pattern rather than point mutations in the target site (Lu et al. 2020; Nakka et al. 2017).

Two of genes with the highest number of SNPs in the vicinity are the two of the four homologs of alpha-tubulin genes TOR2 and TUA5 located on gs08 and gs10. Alpha-tubulins are the target of Group 3, dinitroaniline herbicides such as trifluralin and ethalfluralin. While

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

trifluralin does not control *G. spurium* and ethalfluralin only provides suppression (Anonymous 2022), it is possible the species is exposed to these microtubule inhibitors when other species are the primary targets for control. Point mutations at six sites in the alpha-tubulin genes have been determined to be the source of Group 3 resistance in grass species: goosegrass [*Eleusine indica* (L.) Gaertn.] (Anthony et al. 1998; Yamamoto et al. 1998), green foxtail [*Setaria viridis* (L.) P. Beauv.] (Délye et al. 2004), shotawn foxtail (*Alopecurus aequalis* Sobol.) (Hashim et al. 2012), and rigid ryegrass (*Lolium rigidum* Gaudin) (Chu et al. 2018). None of these point mutations are represented in our draft genome and these genes are involved in fundamentally processes such as successful mitosis and movement of organelles (Ludwig et al. 1987), so it is entirely possible that selection on these genes, if it is occurring, is unrelated to herbicide resistance.

Here we have produced a reference quality, chromosome level assembly for *G. spurium* and compared the genome with its nearest relatives with similar information available. We note that the currently sequenced genomes with n = 11 in the family show relative stability, but there is no clear mechanism that caused the reduction to n=10 in *G. spurium* given the available genomes. The ddRADtags data for the 19 Canadian Prairie populations indicate very low heterozygosity and high inbreeding rates, which result in structured populations with limited gene flow. This likely explains the relatively slow movement of herbicide resistance genes in these populations. This work also indicates some genes candidates of interest that may be involved in the evolution of auxinic resistance in this species. These can be targeted for further investigation and include the IAA gene on chromosome 6 and the AUX gene on chromosome 8. The availability of the *G. spurium* genome will facilitate future research to understand the genetic basis of auxin resistance in the species and into the systematics research into of the polyphyletic genus *Galium*.

Acknowledgements

We would like to thank the ORDC greenhouse staff for their help in growing the plants for this work, Talyor Withey for counting, measuring and photographing the seeds produced during the greenhouse trial and Shefali Vishwakarma for initial work on the RADseq files in STACKS. We would also like to thank Dr. Laforest for suggestions on which genes to include in the list of interest for the evolution of herbicide resistance.

Funding

This work was funded by Agriculture and Agri-Food Canada through the Agricultural Plant Systematics Project (J-002275).

Competing interests: The authors declare none.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

References

Anonymous (2022) 2022 guide to crop protection. Regina, Saskatchewan: Saskatchewan Ministry of Agriculture.

Anthony RG, Waldin TR, Ray JA, Bright SWJ, Hussey PJ (1998) Herbicide resistance caused by spontaneous mutation of the cytoskeletal protein tubulin. Nature 393:260-263

Bache SM, Wickham H (2022) magrittr: a forward-pipe operator for R. R package version 2.0.3

Beckie H, Shirriff S, Leeson J, Hall L, Harker K, Dokken-Bouchard FB, CA (2020) Herbicide-resistant weeds in the Canadian prairies: 2012 to 2017. Weed Technol 34:461-474

Beckie HJ, Lozinski C, Shirriff S, Brenzil CA (2013) Herbicide-resistant weeds in the Canadian prairies: 2007 to 2011. Weed Technol 27:171-183

Beckie HJ, Warwick SI, Sauder CA, Kelln GM, Lozinski C (2012) Acetolactate synthase inhibitor–resistant *Galium spurium* in western Canada. Weed Technol 26:151-155

Bodenhofer U, Bonatesta E, Horejs-Kainrath C, Hochreiter S (2015) msa: an R package for multiple sequence alignment. Bioinformatics 31:3997-3999

Bremer B (2009) A review of molecular phylogenetic studies of Rubiaceae. Ann Mo Bot Gard 96:4-26

Brownrigg R, Minka TP, Deckman A (2022) maps: Draw geographical maps. R package version 3.4.1

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M (2021) BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics Bioinform 3. doi:10.1093/nargab/lqaa108

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124-40

Charif D, Lobry JR (2007) Seqin{R} 1.0-2: a contributed package to the {R} project for statistical computing devoted to biological sequences retrieval and analysis. Pages 207-232 in

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

Bastolla U, Porto M, Roman HE, Vendruscolo M, eds. Structural approaches to sequence evolution: molecules networks populations. New York: Springer Verlag

Chen T, Ehrendorfer F (2001) *Galium.* Pages 104-141 in Committee FoCE, ed. Flora of China

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q (2018) SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience 7:1-6

Chu Z, Chen J, Nyporko A, Han H, Yu Q, Powles S (2018) Novel alpha-tubulin mutations conferring resistance to dinitroaniline herbicides in *Lolium rigidum*. Front Plant Sci 9:97

Conner JK, Hartl DL (2004) A primer of ecological genetics. USA: Sinauer.

Dai N, Wang W, Patterson SE, Bleecker AB (2013) The TMK subfamily of receptor-like kinases in *Arabidopsis* display an essential role in growth and a reduced sensitivity to auxin. PLoS One 8:e60990

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G (2011) The variant call format and VCFtools. Bioinformatics 27:2156-8

Délye C, Menchari Y, Michel S, Darmency H (2004) Molecular bases for sensitivity to tubulin-binding herbicides in green foxtail. Plant Physiol 136:3920-32

Deng W, Di Y, Cai J, Chen Y, Yuan S (2019) Target site resistance mechanisms to tribenuron-methyl and cross-resistance patterns to ALS-inhibiting herbicides of catchweed bedstraw (*Galium aparine*) with different ALS mutations. Weed Sci 67:183-188

Dharmasiri N, Dharmasiri S, Estelle M (2005a) The F-box protein TIR1 is an auxin receptor. Nature 435:441-445

Dharmasiri N, Dharmasiri S, Weijers D, Lechner E, Yamada M, Hobbie L, Ehrismann JS, Jürgens G, Estelle M (2005b) Plant development is regulated by a family of auxin receptor F-box proteins. Dev Cell 9:109-119

Ellinghaus D, Kurtz S, Willhoeft U (2008) LTRharvest: an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9:18

Emms DM, Kelly S (2018) STAG: species tree inference from all genes. bioRxiv doi:10.1101/267914

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238

Flanagan SP (2023) gwscaR: genome-wide selection components analysis in R. R package version 0.1.0

Fox J, Weisberg S (2019) An R companion to applied regression. 3rd edn. Thousand Oaks, CA, USA: Sage

Gaines TA, Heap IM (2023) Mutations in herbicide-resistant weeds to inhibition of enolpyruvyl shikimate phosphate synthase. www.weedscience.com. Accessed Aug 9, 2023

Gleason HA, Cronquist A (1991) Manual of vascular plants of northeastern United States and adjacent Canada. 2nd edn. USA: The New York Botanical Garden

Goudet J, Jombart T (2022) hierfstat: estimation and tests of hierarchical F-statistics. R package version 0.5-11

Graves S, Piepho H-P, Selzer L, Dorai-Raj S (2019) multcompView: visualizations of paired comparisons. R package version 0.1-8

Grossmann K (2010) Auxin herbicides: current status of mechanism and mode of action. Pest Manag Sci 66:113-120

Grossmann K, Hansen H (2001) Ethylene-triggered abscisic acid: a principle in plant growth regulation? Physiol Plant 113:9-14

Grossmann K, Schmülling T (1995) The effects of the herbicide quinclorac on shoot growth in tomato is alleviated by inhibitors of ethylene biosynthesis and by the presence of an antisense construct to the 1-aminocyclopropane-1-carboxylic acid (ACC) synthase gene in transgenic plants. Plant Growth Regul 16:183-188

Gu Z, Gu L, Eils R, Schlesner M, Brors B (2014) circlize: implements and enhances circular visualization in R. Bioinformatics 30:2811-2812

Guo X-M, Wang Z-F, Zhang Y, Wang R-J (2021) Chromosomal-level assembly of the *Leptodermis oblonga* (Rubiaceae) genome and its phylogenetic implications. Genomics 113:3072-3082

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072-1075

Hahne F, Ivanek R (2016) Visualizing genomics data using gviz and bioconductor. Pages 335-351 *in* Mathe E, Davis S, eds. Statistical genomics: methods and protocols. New York, NY, USA: Springer New York

Hall LM, Stromme KM, Horsman GP, Devine MD (1998) Resistance to acetolactate synthase inhibitors and quinclorac in a biotype of false cleavers (*Galium spurium*). Weed Sci 46:390-396

Hansen H, Grossmann K (2000) Auxin-induced ethylene triggers abscisic acid biosynthesis and growth inhibition. Plant Physiol 124:1437-1448

Harrell FE (2022) Hmisc: Harrell miscellaneous. CRAN

Hashim S, Jan A, Sunohara Y, Hachinohe M, Ohdan H, Matsumoto H (2012) Mutation of alpha-tubulin genes in trifluralin-resistant water foxtail (*Alopecurus aequalis*). Pest Manag Sci 68:422-429

Heap I (2023) The international herbicide-resistant weed database. weedscience.org/Home.aspx. Accessed Jan 8, 2022

Hervé M (2022) RVAideMemoire: testing and plotting procedures for biostatistics. R package version 0.9-81-2

Hijmans RJ (2022) geosphere: spherical trigonometry. R package version 1.5-18

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-genome annotation with BRAKER. Methods Mol Biol 1962:65-95

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

Holm LG, Pancho JV, Herberger JP, Plucknett DL (1991) A geographical atlas of world weeds. Malabar, Florida, USA: Krieger Publishing Company, 731 p

Hu K, Xu K, Wen J, Yi B, Shen J, Ma C, Fu T, Ouyang Y, Tu J (2019) Helitron distribution in Brassicaceae and whole genome helitron density as a character for distinguishing plant species. BMC Bioinformatics 20:354

Jombart T (2008) adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403-1405

Kamvar Z, Tabima J, Grünwald N (2014) Poppr: an R package for genetic analysis of populations with clonal partially clonal and/or sexual reproduction. PeerJ 2. doi:10.7717/peerj.281

Kepinski S, Leyser O (2005) The *Arabidopsis* F-box protein TIR1 is an auxin receptor. Nature 435:446-451

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37:907-915

Knaus BJ, Grünwald NJ (2016) vcfR: a package to manipulate and visualize (VCF) data in R. bioRxiv doi:10.1101/041277

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722-736

Kraft M, Kuglitsch R, Kwiatkowski J, Frank M, Grossmann K (2007) Indole-3-acetic acid and auxin herbicides up-regulate 9-cis-epoxycarotenoid dioxygenase gene expression and abscisic acid accumulation in cleavers (*Galium aparine*): interaction with ethylene. J Exp Bot 58:1497-1503

Kumar S, Suleski M, Craig JE, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB (2022) TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol 39

Laetsch DR, Blaxter ML (2017) BlobTools: interrogation of genome assemblies. F1000Research 6

Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, Morgan M, Carey V (2013) Software for computing and annotating genomic ranges. PLoS Comput Biol 9

Leeson JY, Thomas AG, Hall LM, Brenzil CA, Andrews T, Brown KR, Van Acker RC (2005) Prairie weed surveys of cereal, oilseed, and pulse crops from the 1970s to the 2000s. Saskatoon, Canada: Agriculture and Agri-Food Canada

Lemon J (2006) plotrix: a package in the red light district of R. R News 6:8-12

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-1760

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078-2079

Li S-B, Xie Z-Z, Hu C-G, Zhang J-Z (2016) A review of auxin response factors (ARFs) in plants. Front Plant Sci 7. doi:10.3389/fpls.2016.00047

Lomsadze A, Burns PD, Borodovsky M (2014) Integration of mapped RNA-seq reads into automatic training of eukaryotic gene-finding algorithms. Nucleic Acids Res 42. doi:10.1093/nar/gku557

Looman J, Best KF (1979) Budd's flora of the Canadian prairie provinces. Canada: Agriculture Canada.

Lovell JT (2023) GENESPACE: synteny- and orthology-constrained comparative genomics. R package version 1.0.13

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson J, Harkess A, Emms D, Goodstein DM, Schmutz J (2022) GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife 11. doi:10.7554/eLife.78526

Lu H, Yu Q, Han H, Owen MJ, Powles SB (2020) Evolution of resistance to HPPD-inhibiting herbicides in a wild radish population via enhanced herbicide metabolism. Pest Manag Sci 76:1929-1937

Ludwig SR, Oppenheim DG, Silflow CD, Snustad DP (1987) Characterization of the a-tubulin gene family of *Arabidopsis thaliana*. Proc Natl Acad Sci U S A 84:5833-5837

Luo J, Zhou JJ, Zhang JZ (2018) Aux/IAA gene family in plants: molecular structure, regulation, and function. Int J Mol Sci 19. doi:10.3390/ijms19010259

Luu K, Bazin E, Blum MGB (2017) pcadapt: an R package to perform genome scans for selection based on principal component analysis. Mol Eco Res 17:67-77

Lysak MA (2022) Celebrating Mendel, McClintock, and Darlington: On end-to-end chromosome fusions and nested chromosome fusions. Plant Cell 34:2475-2491

Malik N, Vanden Born WH (1988) The biology of Canadian weeds 86. *Galium aparine* L. and *Galium spurium* L. Can J Plant Sci 68:481-499

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A (2018) MUMmer4: a fast and versatile genome alignment system. PLoS Comput Biol 14. doi:10.1371/journal.pcbi.1005944

Martin SL, Benedict L, Wei W, Sauder CA, Beckie HJ, Hall LM (2020) High gene flow maintains genetic diversity following selection for high EPSPS copy number in the weed kochia (Amaranthaceae). Sci Rep 10:18864. doi:10.1038/s41598-020-75345-6

Martin SL, Smith TW, James T, Shalabi F, Kron P, Sauder CA (2017) An update to the Canadian range, abundance, and ploidy of *Camelina* spp. (Brassicaceae) east of the Rocky Mountains. Botany 95:405-417

Mijangos JL, Gruber B, Berry O, Pacioni C, Georges A (2022) dartR v2: an accessible genetic analysis platform for conservation ecology and agriculture. Methods Ecol Evol. doi:10.1111/2041-210X.13918

Moore RJ (1975) The *Galium aparine* complex in Canada. Can J Bot 53:877-893

Morgan M, Pagès H, Obenchain V, Hayden N (2021) Rsamtools: binary alignment (BAM), FASTA, variant call (BCF), and tabix file import. R package version 2.10.0. bioconductor.org/packages/Rsamtools

Nakka S, Godar AS, Wani PS, Thompson CR, Peterson DE, Roelofs J, Jugulam M (2017) Physiological and molecular characterization of hydroxyphenylpyruvate dioxygenase (HPPD)-inhibitor resistance in *Amaranthus palmeri*. Front Plant Sci 8:555. doi:10.3389/fpls.2017.00555

Newcombe RG (1998) Two-sided confidence intervals for the single proportion: comparison of seven methods. Stat Med 17:857-872

O'Leary SJ, Puritz JB, Willis SC, Hollenbeck CM, Portnoy DS (2018) These aren't the loci you're looking for: principles of effective SNP filtering for molecular ecologists. Mol Ecol. doi:10.1111/mec.14792

Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Solymos P, Stevens MHH, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista HBA, FitzJohn R, Friendly M, Furneaux B, Hannigan G, Hill MO, Lahti L, McGlinn D, Ouellette M-H, Cunha ER, Smith T, Stier A, Ter Braak CJF, Weedon J (2022) vegan: community ecology package. R package version 2.6-4

OpenAI (2024) ChatGPT (GPT-4). www.openai.com

Ou S, Chen J, Jiang N (2018) Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res 46. doi:10.1093/nar/gky730

Ou S, Jiang N (2018) LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol 176:1410-1422. doi:10.1104/pp.17.01310

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, Jiang N, Hirsch CN, Hufford MB (2019) Benchmarking transposable element annotation methods for creation of a streamlined comprehensive pipeline. Genome Biol 20:275. doi:10.1186/s13059-019-1905-y

Pagès H, Aboyoun P, Gentleman R, DebRoy S (2021) Biostrings: efficient manipulation of biological strings. R package version 2.62.0

Paradis E (2010) pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26:419-420

Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526-528

Pazmiño DM, Rodriguez-Serrano M, Romero-Puertas MC, Archilla-Ruiz A, Del Río LA, Sandalio LM (2011) Differential response of young and adult leaves to herbicide 2,4-dichlorophenoxyacetic acid in pea plants: role of reactive oxygen species. Plant Cell Environ 34:1874-1889

Pebesma E (2018) Simple features for R: standardized support for spatial vector data. R J 10:439-446

Pebesma E, Mailund T, Heibert J (2016) Measurement units in R. R J 8:486-494

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS One 7. doi:10.1371/journal.pone.0037135

Pividori M (2024) Chatbots in science: what can ChatGPT do for you? Nature. Aug 14 2024. doi: 10.1038/d41586-024-02630-z

Privé F, Luu K, Vilhjalmsson BJ, Blum MGB (2020) Performing highly efficient genome scans for local adaptation with R package pcadapt. Mol Biol Evol 37:2153-2154

Quint M, Gray WM (2006) Auxin signaling. Curr Opin Plant Biol 9:448-453

Rangani G, Salas-Perez RA, Aponte RA, Knapp M, Craig IR, Mietzner T, Langaro AC, Noguera MM, Porri A, Roma-Burgos N (2019) A novel single-site mutation in the catalytic domain of protoporphyrinogen oxidase IX (PPO) confers resistance to PPO-inhibiting herbicides. Front Plant Sci 10:568. doi:10.3389/fpls.2019.00568

Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol Evol 3:217-223

Revelle W (2023) psych: procedures for psychological psychometric and personality research. R package version 2.3.6

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

Savary P, Foltête J-C, Moal H, Vuidel G, Garnier S (2020) graph4lg: a package for constructing and analyzing graphs for landscape genetics in R. Methods Ecol Evol. 1:1-9

Scalabrin S, Toniutti L, Di Gaspero G, Scaglione D, Magris G, Vidotto M, Pinosio S, Cattonaro F, Magni F, Jurman I, Cerutti M, Suggi Liverani F, Navarini L, Del Terra L, Pellegrino G, Ruosi MR, Vitulo N, Valle G, Pallavicini A, Graziosi G, Klein PE, Bentley N, Murray S, Solano W, Al Hakimi A, Schilling T, Montagnon C, Morgante M, Bertrand B (2020) A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. Sci Rep 10:4642

Schliep K, Jombart T, Kamvar ZN, Archer E, Harris R (2020) apex: phylogenetic methods for multiple gene data. R package version 1.0.4

Schliep KP (2011) phangorn: phylogenetic analysis in R. Bioinformatics 27:592-593

Shi J, Liang C (2019) Generic repeat finder: a high-sensitivity tool for genome-wide de novo repeat detection. Plant Physiol 180:1803-1815

Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210-3212

Smit A, Hubley R (2008-2015) RepeatModeler Open-1.0. www.repeatmasker.org

Smit A, Hubley R, Green P (2013-2010) RepeatMasker Open-4.0. www.repeatmasker.org

Smith T, Kron P, Martin S (2018) flowPloidy: an R package for genome size and ploidy assessment of flow cytometry data. Appl Plant Sci 6. doi:10.1002/aps3.1164

South A (2017) rnaturalearth: world map data from natural earth. R package version 0.1.0

Soza VL, Olmstead RG (2010) Molecular systematics of tribe Rubieae (Rubiaceae): evolution of major clades, development of leaf-like whorls, and biogeography. Taxon 59:755-771

Stanke M, Diekhans M, Baertsch R, Haussler D (2008) Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637-644

Stevens PF (2017) Angiosperm phylogeny website v14. www.mobot.org/mobot/research/apweb/welcom.html Accessed Feb 6, 2024

Storey JD, Bass AJ, Dabney A, Robinson D (2022) qvalue: Q-value estimation for false discovery rate control. R package version 2.28.0

Su W, Gu X, Peterson T (2019) TIR-Learner: a new ensemble method for TIR transposable element annotation provides evidence for abundant new transposable elements in the maize genome. Mol Plant 12:447-460

Sun J, Wang J-x, Zhang H-j, Liu J-l, Bian S-n (2011) Study on mutations in ALS for resistance to tribenuron-methyl in *Galium aparine* L. Agric Sci China 10:86-91

Sun P, Jiao B, Yang Y, Shan L, Li T, Li X, Xi Z, Wang X, Liu J (2022) WGDI: a user-friendly toolkit for evolutionary analyses of whole genome duplications and ancestral karyotypes. Mol Plant 15:1841-1851

Tadeo K, Ronald K, Vereriano T, Robooni T (2024) Genetic diversity and structure of Ugandan tea (*Camellia sinensis* (L.) O. Kuntze) germplasm and its implication in breeding. Genet Resour Crop Evol 71:481-496

Tan X, Calderon-Villalobos LIA, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. Nature 446:640-645

R Core Team (2021) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Version 4.1.1. www.R-project.org

Thioulouse J, Dray S, Dufour A-B, Siberchicot A, Jombart T, Pavoine S (2018) Multivariate analysis of ecological data with ade4. Springer. doi: 10.1007/978-1-4939-8850-1

Todd OE, Figueiredo MRA, Morran S, Soni N, Preston C, Kubes MF, Napier R, Gaines TA (2020) Synthetic auxin herbicides: finding the lock and key to weed resistance. Plant Sci 300:110631. doi:10.1016/j.plantsci.2020.110631

Tranel PJ, Wright TR, Heap IM (2023) Mutations in herbicide-resistant weeds to inhibition of acetolactate synthase. www.weedscience.com. Accessed Aug 9, 2023

Turner NJ, Szczawinski AF (1978) Wild coffee and tea substitutes of Canada. Ottawa: National Museum of Natural Sciences

Urbanek S (2022) png: read and write PNG images. R package version 0.1-8.

Van Eerd LL (2004) Characterization and inheritance of quinclorac resistance in a false cleavers (*Galium spurium* L.) biotype. Doctor of Philosophy: University of Guelph

Van Eerd LL, McLean MD, Stephenson GR, Hall JC (2004) Resistance to quinclorac and ALS-inhibitor herbicides in *Galium spurium* is conferred by two distinct genes. Weed Res 44:355-365

Van Eerd LL, Stephenson GR, Kwiatkowski J, Grossmann K, Hall JC (2005) Physiological and biochemical characterization of quinclorac resistance in a false cleavers (*Galium spurium* L.) biotype. Pest Manag Sci 53:1144-1151

Venables WN, Ripley BD (2002) Modern applied statistics with S. 4th edn. New York: Springer

Voss EG, Reznicek AA (2012) Field manual of Michigan flora. USA: The University of Michigan Press

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9. doi:10.1371/journal.pone.0112963

Walsh TA, Neal R, Merlo AO, Honma M, Hicks GR, Wolff K, Matsumura W, Davies JP (2006) Mutations in an auxin receptor homolog AFB5 and in SGT1b confer resistance to synthetic picolinate auxins and not to 24-dichlorophenoxyacetic acid or indole-3-acetic acid in *Arabidopsis*. Plant Physiol 142:542-552

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40. doi:10.1093/nar/gkr1293

Wickham H (2016) ggplot2: elegant graphics for data analysis. New York: Springer-Verlag

Wickham H (2022) stringr: simple consistent wrappers for common string operations. R package version 1.5.0

Wickham H, Averick M, Bryan J, Chang W, McGowan LDA, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019) Welcome to the tidyverse. J Open Source Softw 4:1696. doi:10.21105/joss.01686

Wickham H, François R, Müller HL (2022) dplyr: a grammar of data manipulation. R package version 1.0.10

Workman R, Renee F, Kilburn D, Hao S, Liu K, Timp W (2018) Recalcitrant plant species for third generation sequencing. Protocol Exchange

Xiong W, He L, Lai J, Dooner HK, Du C (2014) HelitronScanner uncovers a large overlooked cache of helitron transposons in many plant genomes. Proc Natl Acad Sci U S A 111:10263-10268

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35. doi:10.1093/nar/gkm286

Yamamoto E, Zeng L, Baird WV (1998) Alpha-tubulin missense mutations correlate with antimicrotubule drug resistance in *Eleusine indica*. Plant Cell 10:297-308

Yang LE, Meng Y, Peng DL, Nie ZL, Sun H (2018) Molecular phylogeny of *Galium* L. of the tribe Rubieae (Rubiaceae) - emphasis on Chinese species and recognition of a new genus *Pseudogalium*. Mol Phylogenet Evol 126:221-232

Zazimalova E, Murphy AS, Yang H, Hoyerova K, Hosek P (2010) Auxin transporters--why so many? Cold Spring Harb Perspect Biol 2. doi:10.1101/cshperspect.a001552

Zažímalová E, Petrášek J, Benková E, Mueller-Roeber B, Balazadeh S (2014) Auxin and its role in plant development. Springer Vienna 33:21-33. doi:10.1007/978-3-7091-1526-8

Zeileis A, Fisher J, Hornik K, Ihaka R, McWhite C, Murrell P, Stauffer R, Wilke C (2020) colorspace: a toolbox for manipulating and assessing colors and palettes. J Stat Softw 96:1-49

Zhang R-G, Wang Z-X, Ou S, Li G-Y (2019) TEsorter: lineage-level classification of transposable elements using conserved protein domains. bioRxiv doi:10.1101/800177

Zheng H-g, Hall JC (2001) Understanding auxinic herbicide resistance in wild mustard: physiological, biochemical, and molecular genetic approaches. Weed Sci 49:276-281

Table 1. Genome assembly statistics for *Galium spurium*.

Metric	Main Assembly
contigs (>= 50000 bp)	10
Total length	313,817,154
N50	31,727,068
NG50	30,373,747
N's per 100kbp*	3.9
BUSCO (eudicots_odb10)	
Complete (C)	2192 (94%)
Complete Single Copy (S)	2081 (89%)
Complete Duplicated (D)	111 (5%)
Fragmented (F)	44 (2%)
Missing (M)	90 (4%)

* Number of base pairs represented with N rather than a sequenced nucleotide usually as a result of Ns placed between contigs during scaffolding.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

Table 2. Top twenty homologs of genes with known or potential roles in herbicide resistance ranked by proximity to single nucleotide polymorphisms (SNPs) under selection. Columns contain information on the homolog, the gene's position in the genome, the distance to the nearest SNP indicating selection and the number of these SNPs within a 1 Mbp window centred on the of the gene's midpoint. See Supplementary Table 4 for all results. SNPs were identified as under selection using ordination with almost four thousand SNPs across the genome included. The Benjamini-Hochberg procedure was used to adjust significance value to control false discovery which was set to 1%. SNPs were thinned to reduce the effects of linkage disequilibrium.

Homolog	Putative Role	Chromosome	Gene Name	Nearest SNP Distance (bp)	SNPs within 1 Mbp
AT	Alpha tubulin	gs08	GS3201901	19,783	6
IAA	Auxin induced gene	gs06	GS2558401	11,751	5
AUX/LAX	Auxin influx transporter	gs08	GS3259201	26,589	4
AT	Alpha tubulin	gs10	GS0705701	212,913	4
CAND1	Cullin-associated and neddylation-dissociated	gs02	GS1201501	79,948	4
ECR1	RUB-activating enzyme	gs05	GS1976901	151,288	4
ARF	Auxin response factor	gs08	GS2988101	43,412	3
AUX/LAX	Auxin influx carrier	gs04	GS1619901	218,631	2
GLN	Glutamine Synthetase	gs05	GS1972001	61,194	2
TMK	Transmembrane Kinase	gs08	GS3026001	18,963	2
ABCB	ATP-binding cassette B	gs06	GS2393501	93,165	2
ALS	Acetolactate synthase	gs06	GS2530401	312,834	2
ABP1	Auxin Binding Protein	gs06	GS2594801	85,962	2
ARF	Auxin response factor	gs02	GS1185401	491,484	1
IAA/ARF	Auxin induced gene	gs08	GS3157301	497,879	1
TIR/AFB	Auxin receptor	gs06	GS2589801	381,117	1
PIN	Auxin efflux carrier	gs01	GS0071401	316,058	1
IAA	Auxin induced gene	gs01	GS0066501	174,697	1
GLN	Glutamine Synthetase	gs01	GS0176201	66,200	1
PIN	Auxin efflux carrier	gs01	GS0181001	485,054	1

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

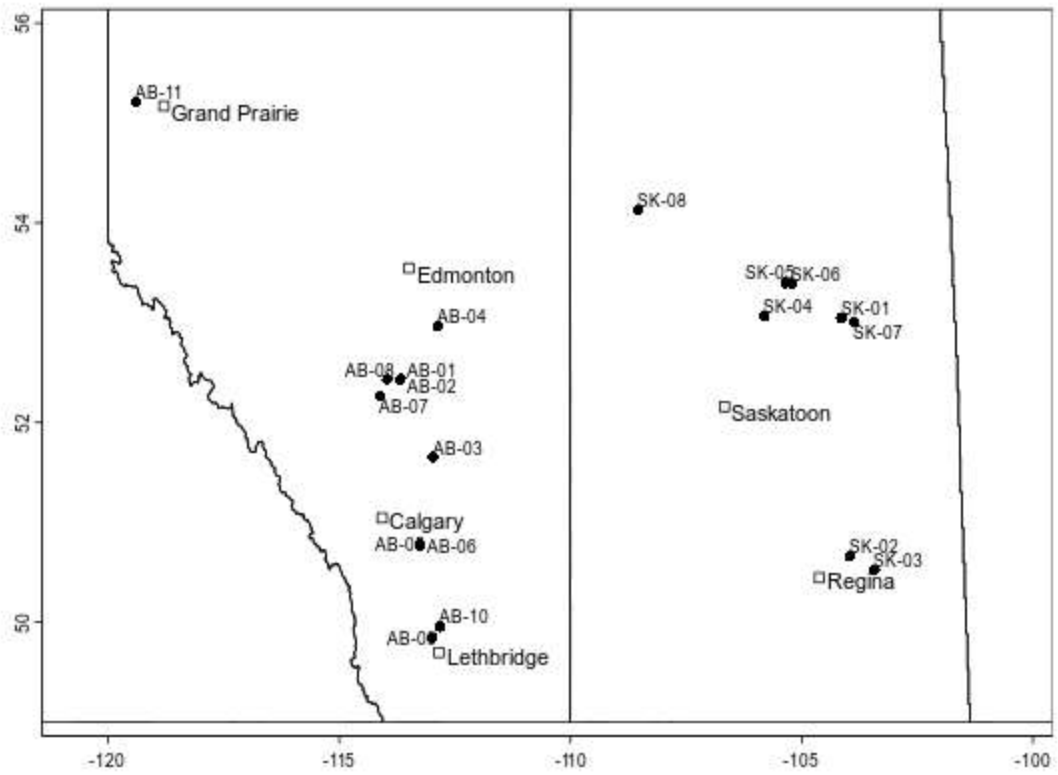

Figure 1. Map of the Canadian Prairie Provinces Albert and Saskatchewan indicating sample locations for populations collected for use in this research with latitude indicated on the left and longitude indicated along the bottom.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

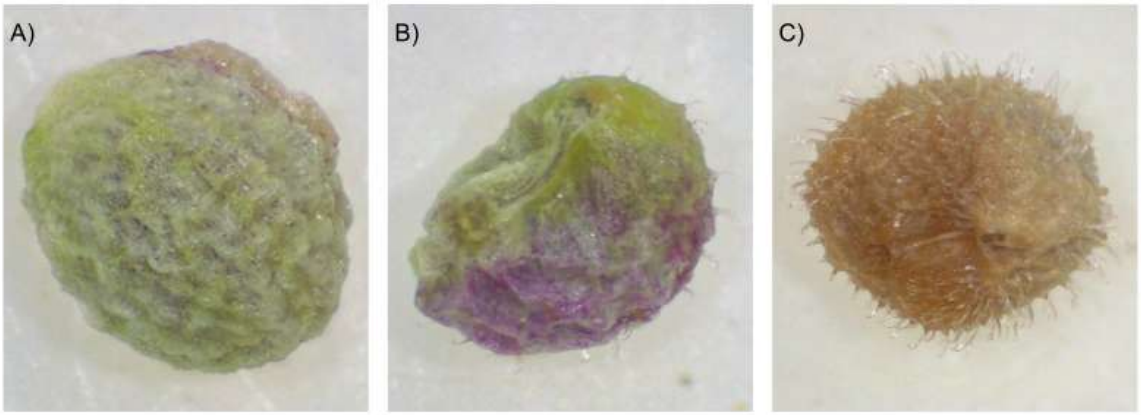

Figure 2. The three visual categories scored for hook density on the mericarps of *Galium spurium* A) naked, B) sparsely hooked, C) densely hooked.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

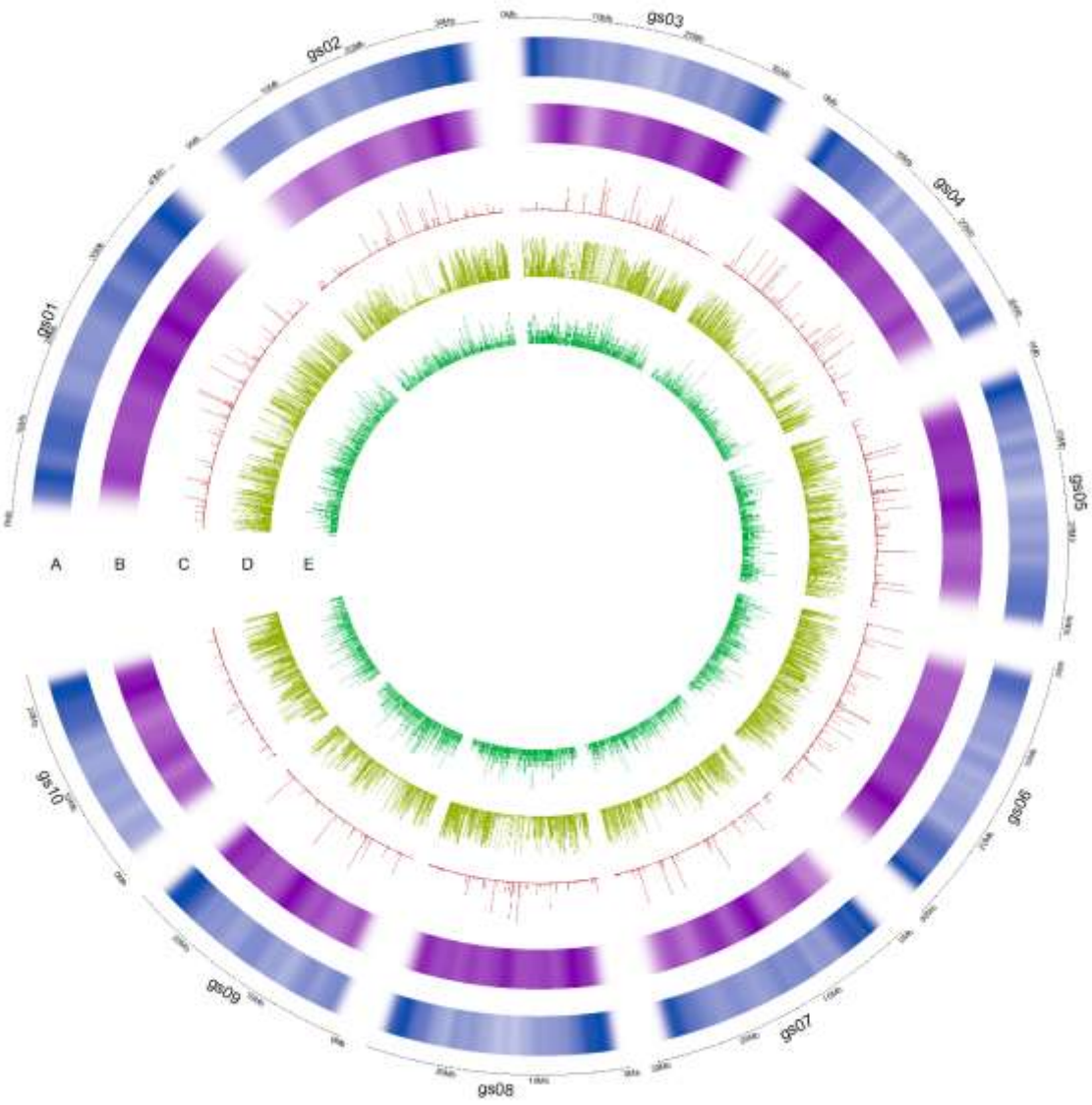

Figure 3. Circle plot of the *Galium spurium* genome with A) gene density, B) SNP density, C) heterozygosity, D) alternative allele frequency, and E) long terminal repeat frequency. An indication of chromosome length is plotted along the outside edge.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

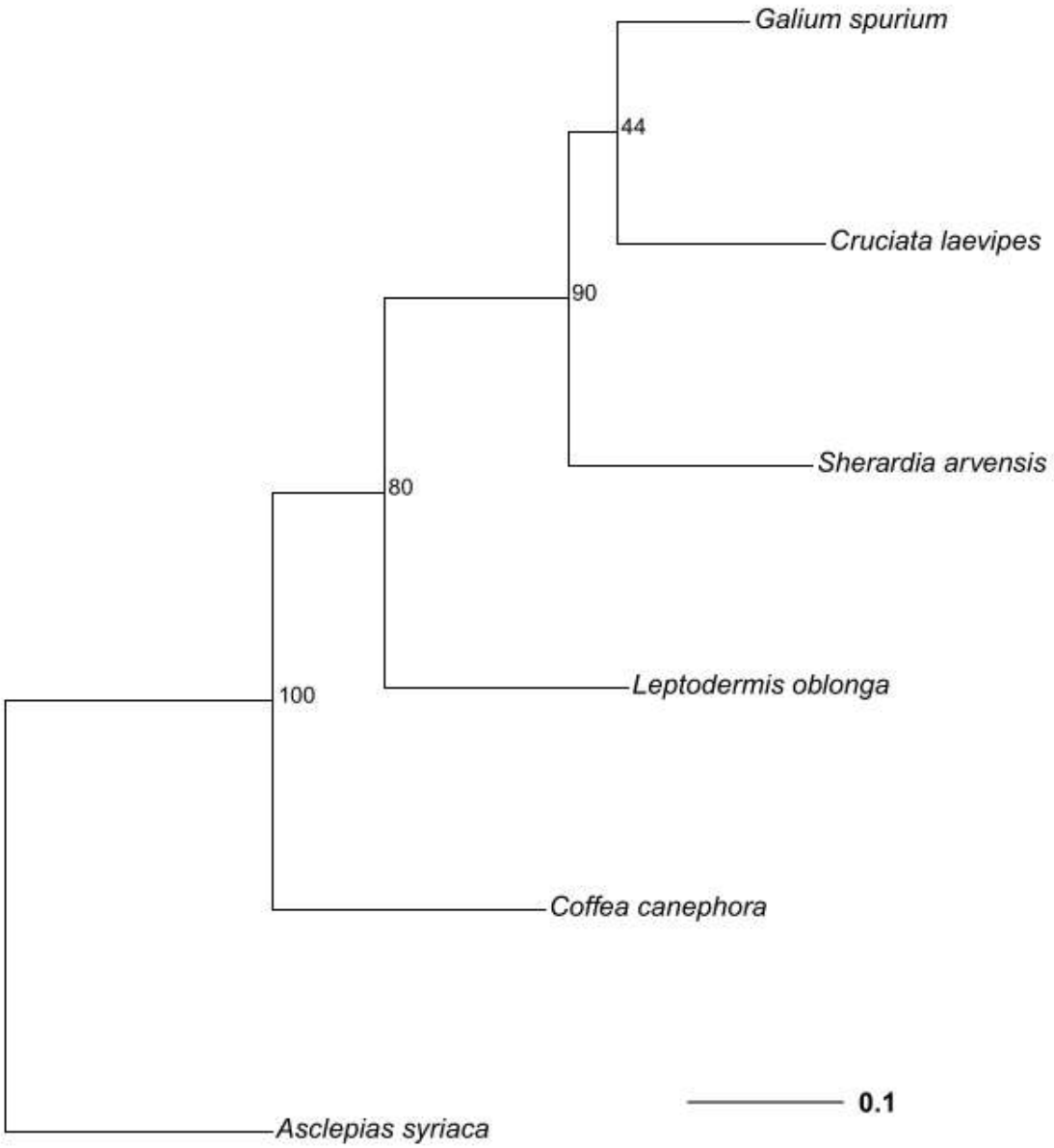

Figure 4. A phylogenetic tree indicating the evolutionary relationship among species studied here produced by OrthoFinder. The values at the nodes indicate support calculated by OrthoFinder using the Species Tree Inference from All Genes (STAG) method and indicates the proportion of orthogroup trees that had the bipartition. *Galium*, *Cruciata* and *Sherardia*, are all within the tribe Rubieae (Rubioideae), while *Leptodermis* is in the tribe Paederieae (Rubioideae) and *Coffea canephora* is in a different sub-family Ixoroideae. *Asclepias syriaca* is an out group from the Apocynaceae.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

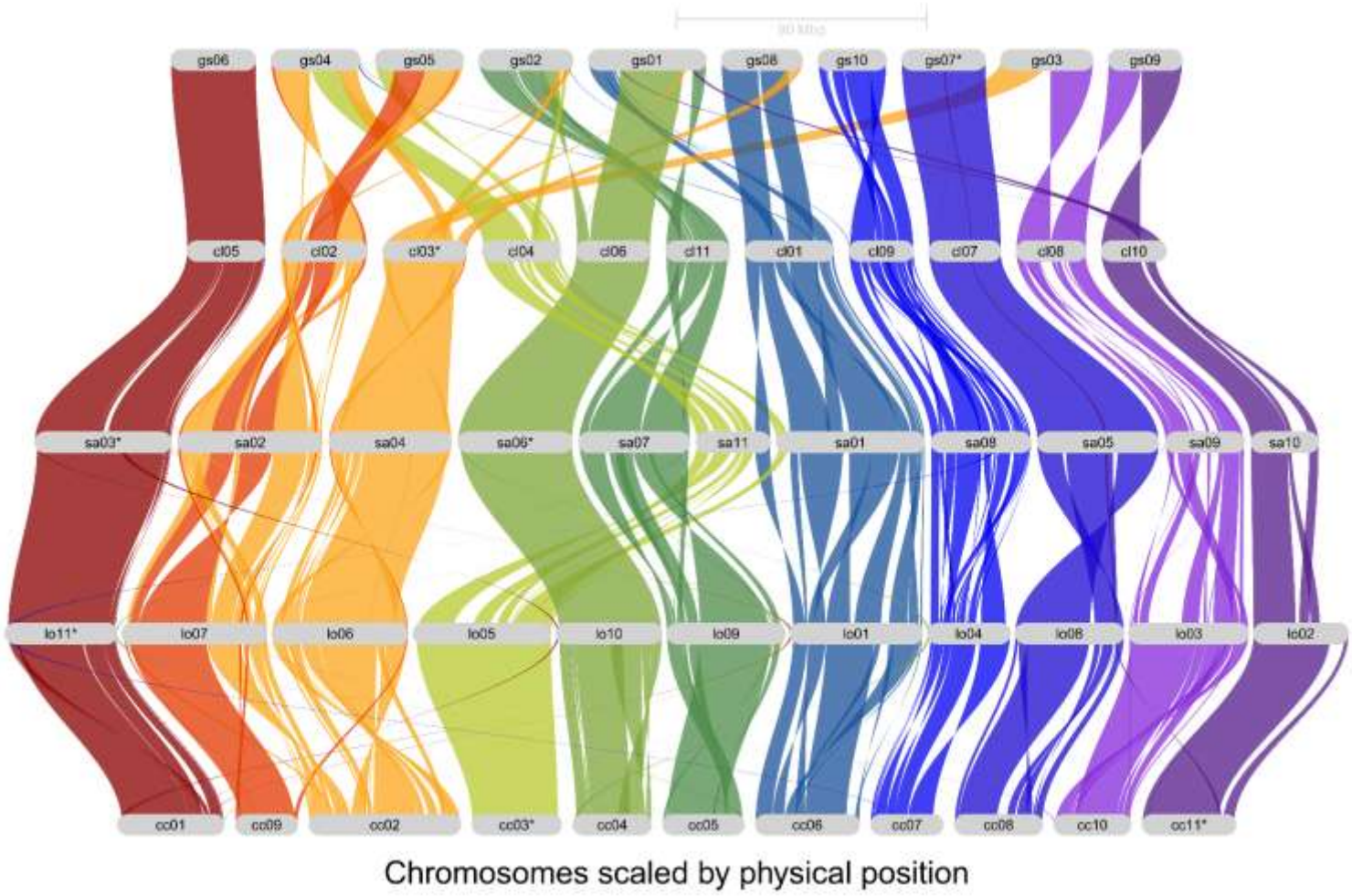

Figure 5. A riparian plot of syntenic relationships among species from the Rubiaceae generated by GENESPACE showing genome structure across the family with *Coffea canephora* at the bottom and working upwards: *Leptodermis oblonga*, *Sherardia arvensis*, *Cruciata laevipes* and finally *Galium spurium* at the top of the plot. Asterisks indicate that the chromosome was inverted to clarify the plot. Note that *Coffea canephora's* cc09 has been shifted to better indicate its position in the chromosomes of the Rubiaceae species.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

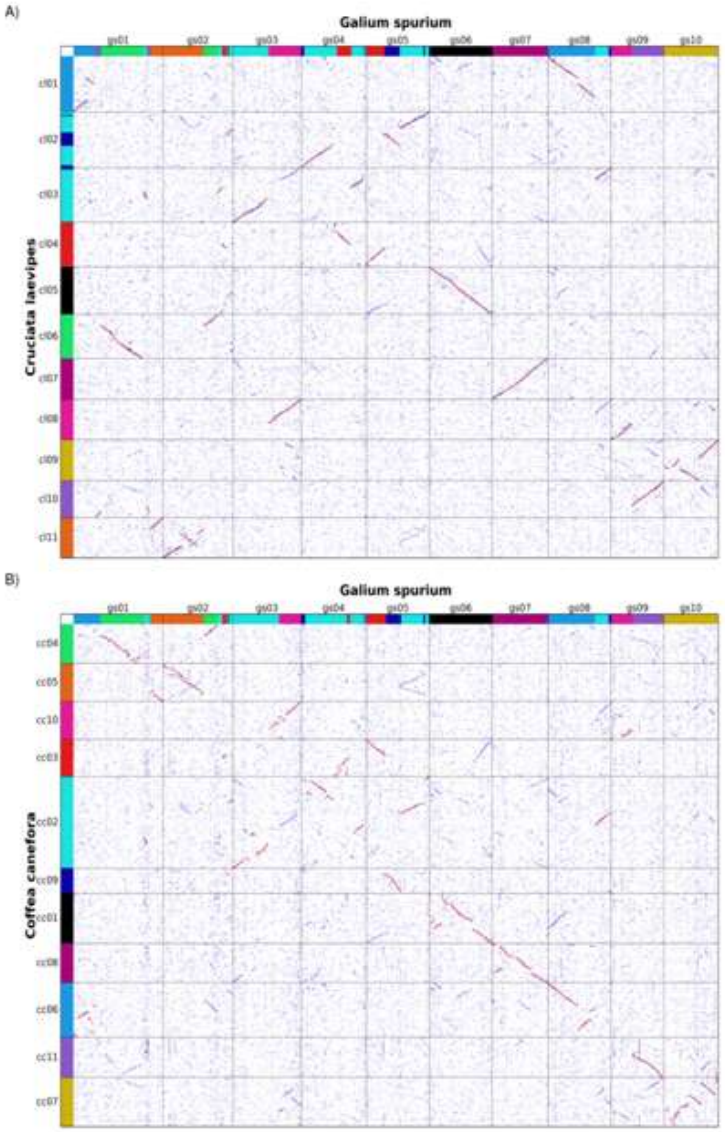

Figure 6. Plots of synteny between proteins of *Galium spurium* A) and *Cruciata laevipes* B) and *Coffea canefora*. Red dots indicate the best syntenic matches, blue dots indicate secondary matches with grey dots indicating lower order matches. For *Galium spurium* and *Cruciata laevipes*, chromosomes are coloured based on syntenic relationships with *Coffea canefora*. From these plots we see that several chromosomes are almost entirely syntenic (e.g. gs06, gs07 and gs10), but there is no clear pattern suggesting that end to end joining or nested chromosome fusion led to the reduction in chromosome number in *G. spurium* from n = 11 to n = 10.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

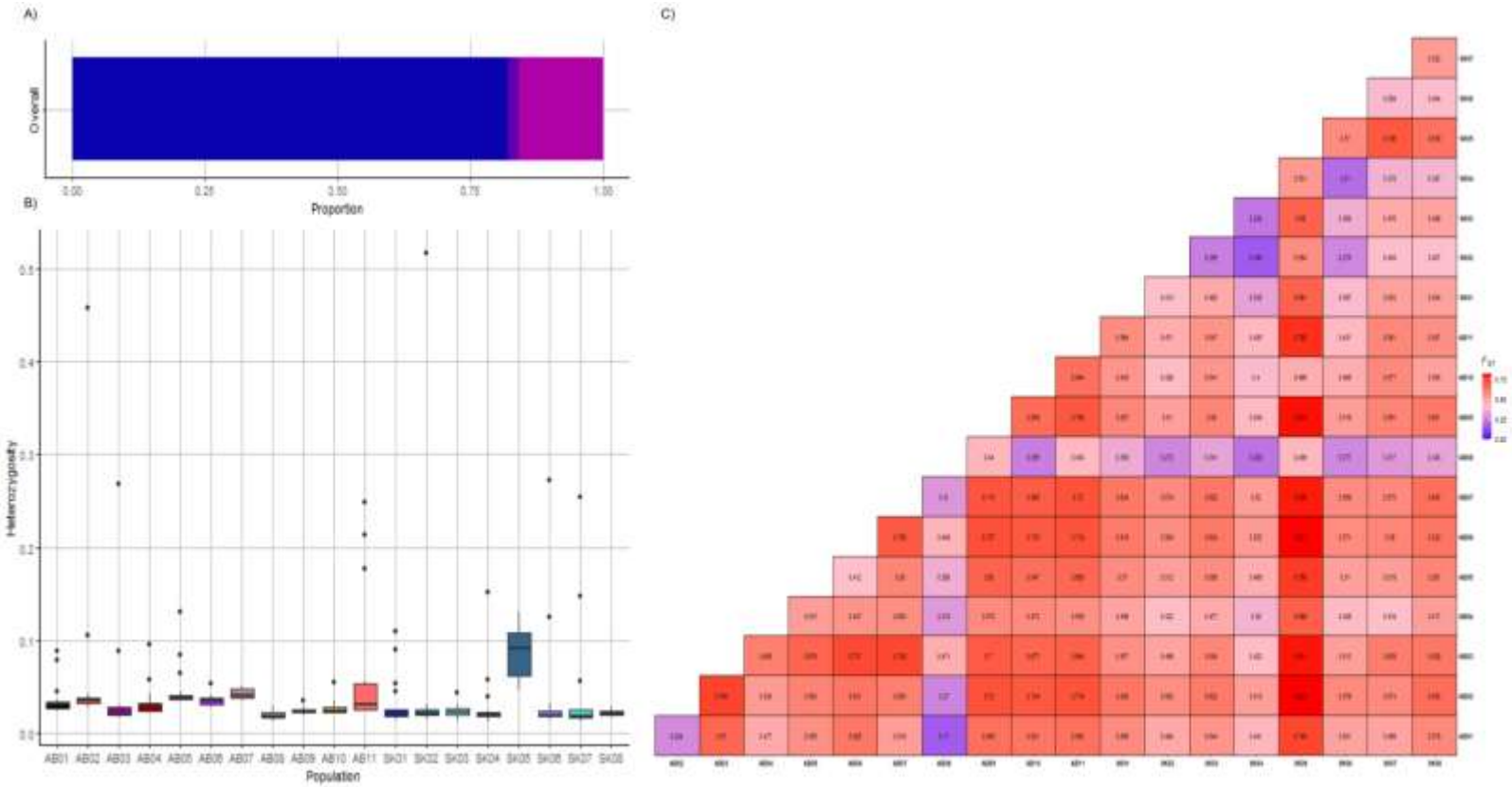

Figure 7. For the 19 populations studied here A) overall proportion of individuals showing homozygosity for the major allele (blue), heterozygosity (purple), and homozygosity (magenta) for the minor allele; B) levels of heterozygosity by population; and 3) a heat map of pairwise F_{ST} values for with cooler colours indicating lower F_{ST} values and warmer indicating higher. Heterozygosity was low across all populations with the majority of individuals (83%) homozygous for the reference allele with evidence of limited outcrossing. On average, the F_{ST} values were 0.54, but AB08 showed lower pairwise F_{ST} values across comparisons and SK05 showing higher values.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

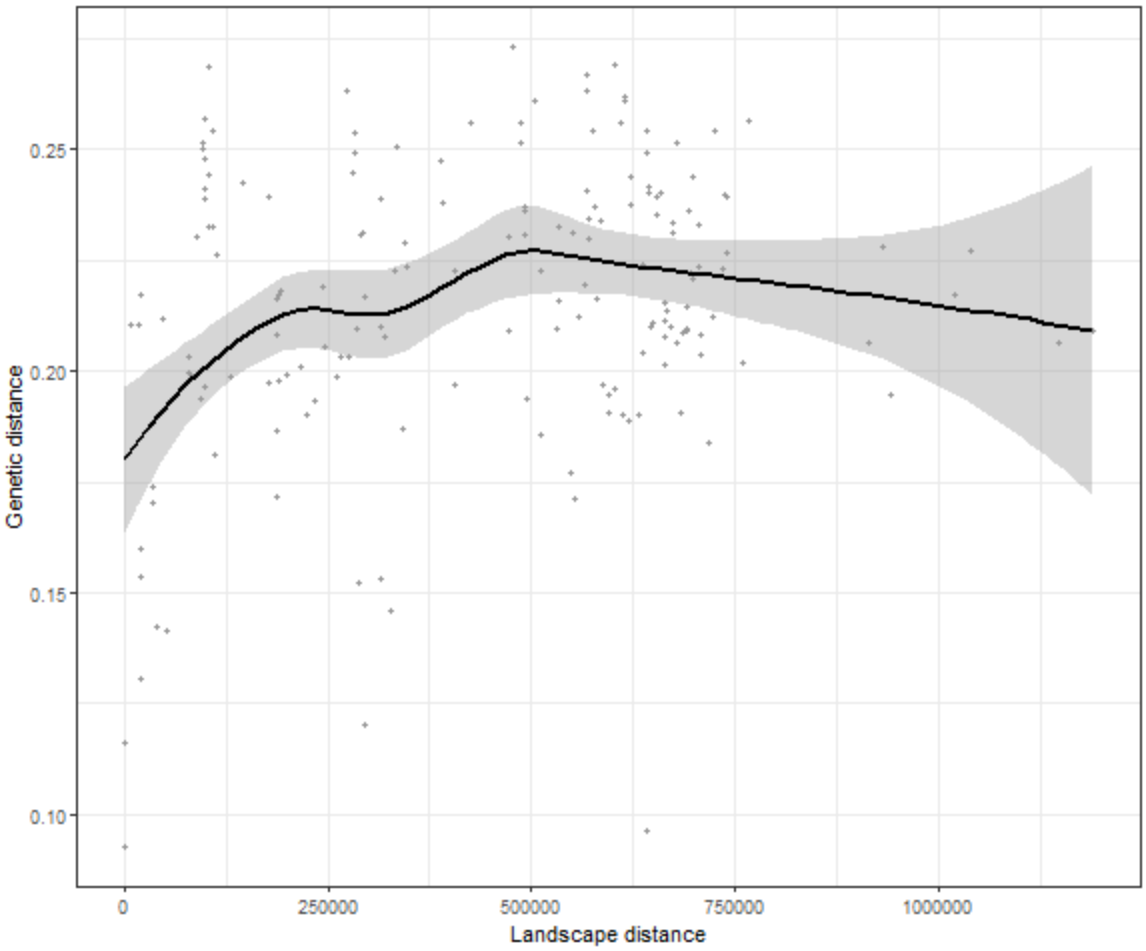

Figure 8. Provesti's genetic distance (y-axis) by geographic (landscape) distance (m) for pairings of the 19 populations. This graph (graph41g::scatter_dist) shows a smooth (loess) line with 95% confidence interval. A Mantel test indicates a significant ($p < 0.001$) positive relationship as indicated by Mantel test with 1000 Monte Carlo repetitions and provides evidence of isolation by distance.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

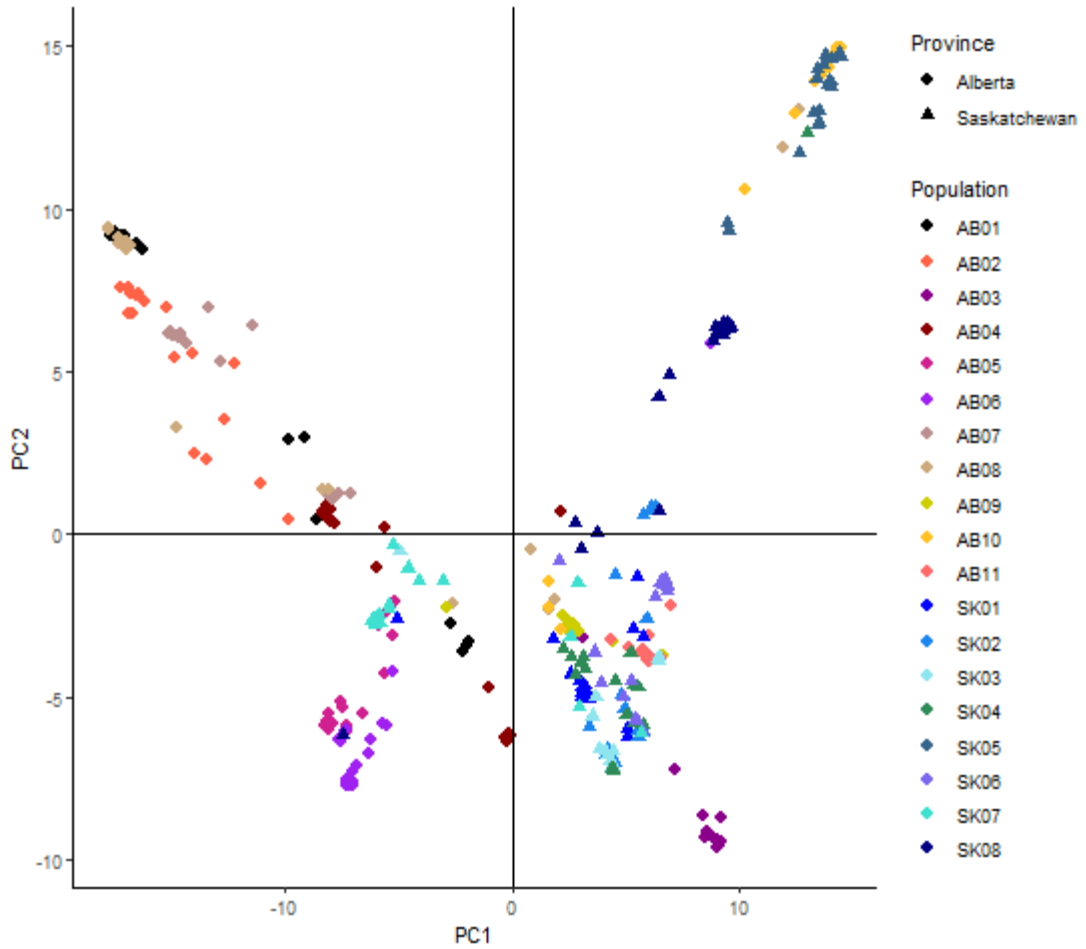

Figure 9. Principle component analysis of the genetic variation of SNP data. Circles in warm colours represent individuals from Alberta and triangles in cool colours those from Saskatchewan. Clumping of individuals from the same population is evident, but variation among populations provides spread. The 76% of the variation is explained by axis 1, 44% by axis 2.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

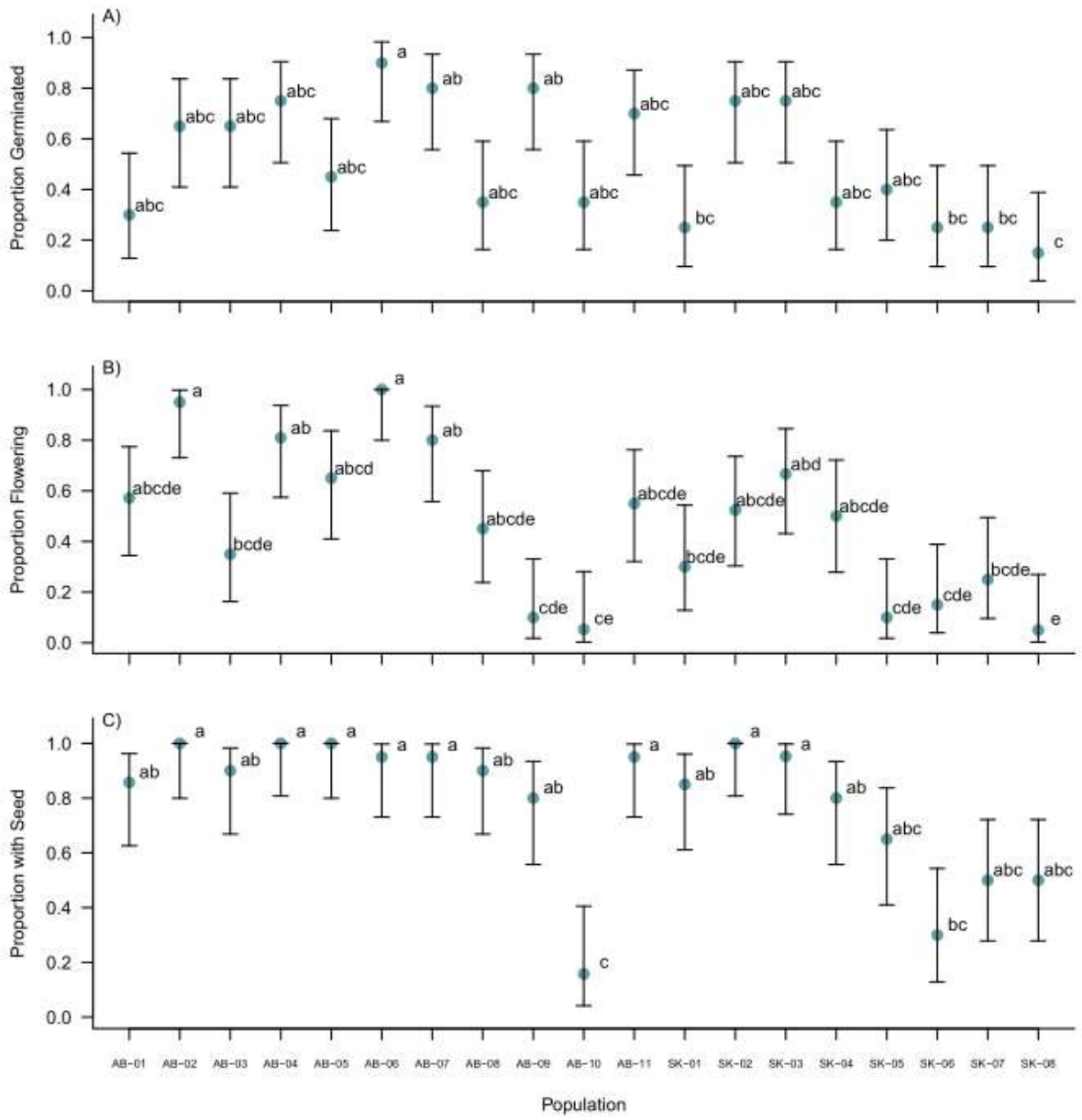

Figure 10. Phenological data from the 19 populations grown in the greenhouse including A) proportion of seed germinated after seven days, B) proportion flowering at six weeks, and C) proportion of plants with seed at eight weeks. In all cases statistical difference are observed, but the majority of populations are highly variable and show strong overlap with values from the other populations.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

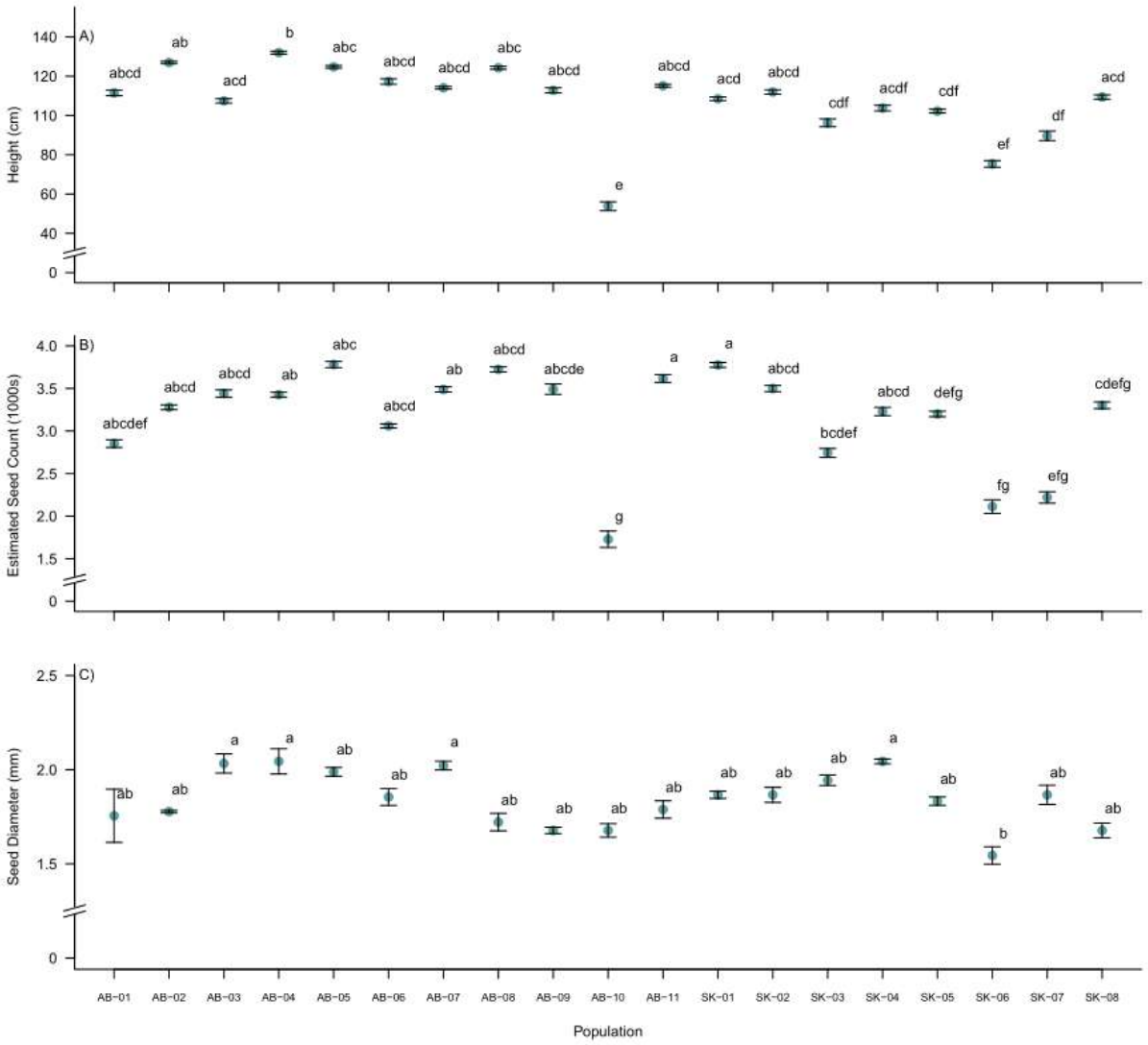

Figure 11. Morphological data from the 19 populations grown in the greenhouse for key characteristics including A) plant height at six weeks, B) total estimated seed count, and C) seed (mericarp) diameter.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

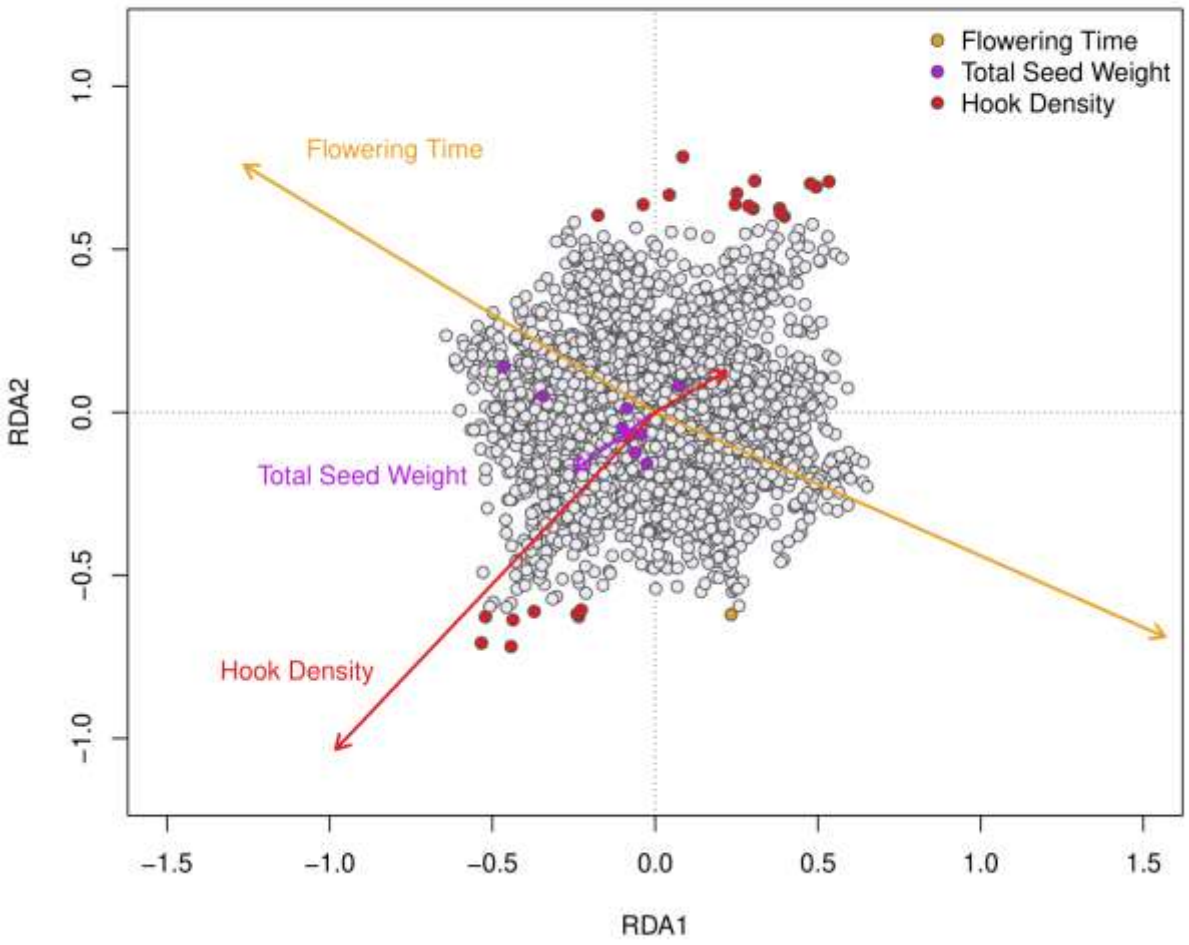

Figure 12. Redundancy analysis examining how variation in the SNP data is associated with variation in key traits. A total of 5.5% of the variation in the SNP data was explained by the included morphological characteristics. The majority of highly loading SNPs (70%), with a cut-off of 3 standard deviations from the mean, were associated with hook density and are highlighted in red. Most of the remaining SNPs are associated with total seed weight (26%) and highlighted in purple, while the one SNP associated with flowering time is in yellow.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press

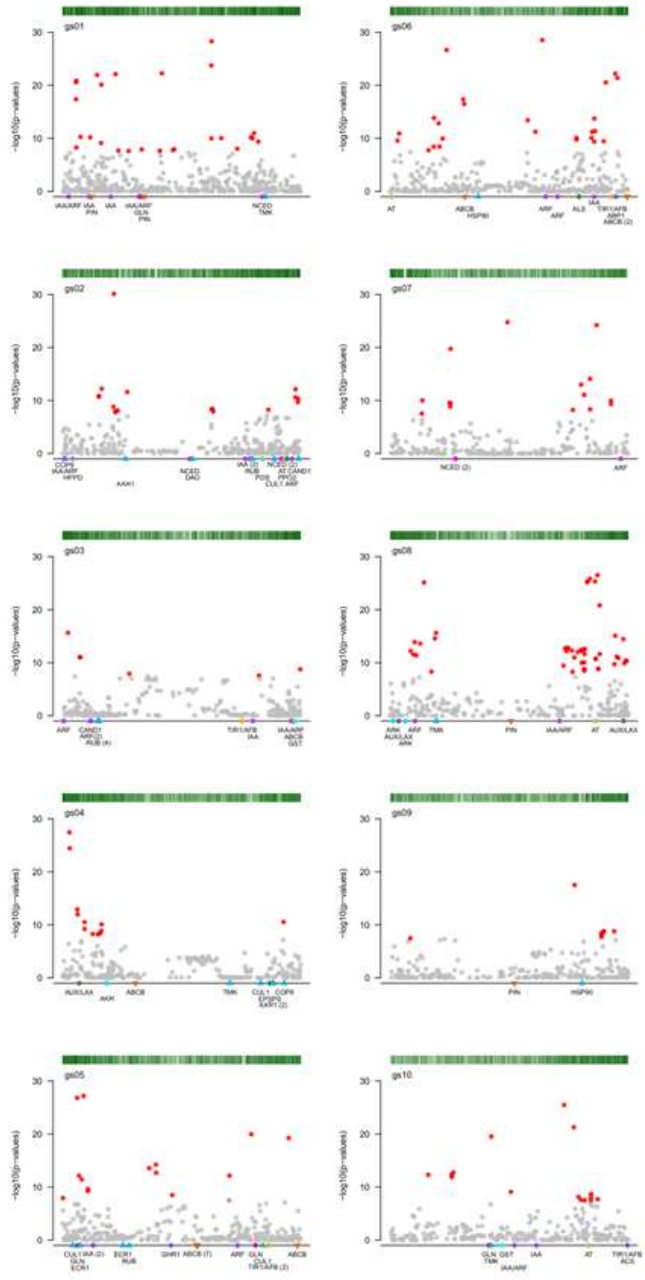

Figure 13. Manhattan plots of almost four thousand SNPs across the genome with the 192 SNPs showing evidence of selection after the Benjamini-Hochberg procedure was used to adjust significance value to control false discovery, the rate was set to 1% and SNPs were thinned to reduce the effects of linkage disequilibrium, highlighted in red. The green bar at the top of each chromosome plot indicates gene density with darker areas annotated with more genes. The location of genes with proteins that homologs to known or potential contributors to herbicide resistance.

https://doi.org/10.1017/wsc.2024.79 Published online by Cambridge University Press