RESEARCH ARTICLE

Ecology and Evolution
Open Access WILEY

Disease resistance gene count increases with rainfall in *Silphium integrifolium*

Kyle Keepers[1] (ORCID) | Kelsey Peterson[1] | Andrew Raduski[1] | Kathryn M. Turner[2] |
David Van Tassel[2] | Kevin Smith[3] | Alex Harkess[4] | James D. Bever[5] | Yaniv Brandvain[1]

[1]Department of Plant and Microbial Biology, University of Minnesota, St Paul, Minnesota, USA

[2]Crop Protection Ecology, The Land Institute, Salina, Kansas, USA

[3]Department of Agronomy and Plant Genetics, University of Minnesota, St Paul, Minnesota, USA

[4]HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, USA

[5]Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA

Correspondence
Kyle Keepers, Department of Plant and Microbial Biology, University of Minnesota, 1500 Gortner Ave, St Paul, MN 55108, USA.
Email: kylekeepers@gmail.com

Funding information
National Science Foundation, Grant/Award Number: 1738041

Abstract

Intracellular plant defense against pathogens is mediated by a class of disease resistance genes known as NB-LRRs or NLRs (R genes). Many of the diseases these genes protect against are more prevalent in regions of higher rainfall, which provide better growth conditions for the pathogens. As such, we expect a higher selective pressure for the maintenance and proliferation of R genes in plants adapted to wetter conditions. In this study, we enriched libraries for R genes using RenSeq from baits primarily developed from the common sunflower (*Helianthus annuus*) reference genome. We sequenced the R gene libraries of *Silphium integrifolium* Michx, a perennial relative of sunflower, from 12 prairie remnants across a rainfall gradient in the Central Plains of the United States, with both Illumina short-read ($n=99$) and PacBio long-read ($n=10$) approaches. We found a positive relationship between the mean effective annual precipitation of a plant's source prairie remnant and the number of R genes in its genome, consistent with intensity of plant pathogen coevolution increasing with precipitation. We show that RenSeq can be applied to the study of ecological hypotheses in non-model relatives of model organisms.

KEYWORDS
common gardens, host-pathogen dynamics, prairies, R genes, RenSeq, *Silphium integrifolium*

TAXONOMY CLASSIFICATION
Evolutionary ecology, Genetics, Genomics

1 | INTRODUCTION

Silphium integrifolium Michx (commonly known as rosinweed or silflower) is a native perennial prairie species being domesticated for production as a forage and oilseed species (Van Tassel et al., 2017; Vilela et al., 2018). The plant's perennial nature allows it to provide novel ecosystem services in agronomic production systems (Glover et al., 2010; Van Tassel et al., 2017) while producing oilseeds with similar nutritional composition to annual sunflower (*Helianthus annuus* L. (Asterales: Asteraceae)) seeds (Reinert et al., 2018). However, *S. integrifolium* suffers infections from diverse pathogen enemies across its native range. The stems, leaves, and flowers can be infected with multiple strains (Turner, unpublished observations) of *Puccinia silphii* Schwein (1832), or *Silphium* rust (Turner et al., 2018), *Colletotrichum silphii* Davis (1919) (Horst, 2008) syn. and *Colletotrichum dematium* (Pers.) Grove (Cybernome; Farr et al., 1989)

called leaf blotch hereafter. Recently, _Silphium_ clear vein (SCV) leaf and stem disease has been described and hypothesized to be caused by a virus based on the identification of viral sequences in _Silphium integrifolium_ that are similar to the Dahlia common Mosaic Virus (DCMV) and Dahlia endogenous plant pararetroviral sequence (DvEPRS, formerly DMV-D10) (Almeyda et al., 2014; Cassetta et al., 2023). Understanding how _S. integrifolium_ defends against these pathogens is crucial for progress in breeding programs and will elucidate evolutionary questions posed in the species (Cassetta et al., 2023; Van Tassel et al., 2017). In this study, plants from 12 wild populations of _S. integrifolium_ were planted in a reciprocal transplant design to better understand adaptation to disease pressures along a rainfall gradient. Phenotypic data revealed that these wild populations have distinctive morphologies and showed variable levels of tolerance to biotic stresses, such as pests and pathogens (Cassetta et al., 2023). Individuals with resistance to diseases of _S. integrifolium_ were visually identified in the study, and have been previously identified in breeding populations (Turner et al., 2018), but the genetic basis of disease resistance has not been investigated. Previous studies of crop pathogen resistance have investigated the phenotypic effects conferred by a specific gene (Rouse et al., 2011) or a small set of genes (Zhu et al., 2013). Here, we look at the diversity of disease resistance genes in wild populations of _S. integrifolium_ to answer evolutionary questions regarding adaptation of wild populations to their pathogen pressures and suggest applications for germplasm enhancement.

Environments with high precipitation are conducive to increased pathogen pressures (Clarkson et al., 2014; Granke & Hausbeck, 2010; Islam & Toyota, 2004; Magarey et al., 2005; Rowlandson et al., 2015). Several studies have found patterns of resistance alleles consistent with pathogen selection intensity increasing with precipitation (Abbott et al., 1991; Dong et al., 2009; Wahl, 1970). We hypothesize that _S. integrifolium_ populations evolved in areas with high precipitation will have a more diverse complement of disease resistance genes to deal with this challenge. To test this hypothesis, we collected seeds from _S. integrifolium_ stands from four prairie remnants each in three geographic regions in the Central Plains of the United States. The regions are situated along a gradient of effective precipitation, or the amount of rainfall that is not lost to potential evapotranspiration, referred to in this study as the climate moisture index (CMI). We refer to these regions as "West," from West-Central Kansas, "Central", from Eastern Kansas, and "East," from Central Illinois (Figure 1). Prairie remnants in the West were drier, receiving around $-55\,kg/m^2$/month of CMI per year, whereas sites in the East were wetter, receiving closer to $-15\,kg/m^2$/month (Table 1). We note that plant defensive gene diversity increased along this gradient in _Andropogon gerardii_, a native perennial grass that co-occurs with _S. integrifolium_ (Rouse et al., 2011) and soil pathogen diversity also increased along this same gradient (Delavaux et al., 2021).

A major class of disease resistance genes in plants are the cytoplasmic nucleotide binding and leucine-rich repeat (NLR) genes (Caplan et al., 2008; Eitas & Dangl, 2010). These NLRs (hereafter,
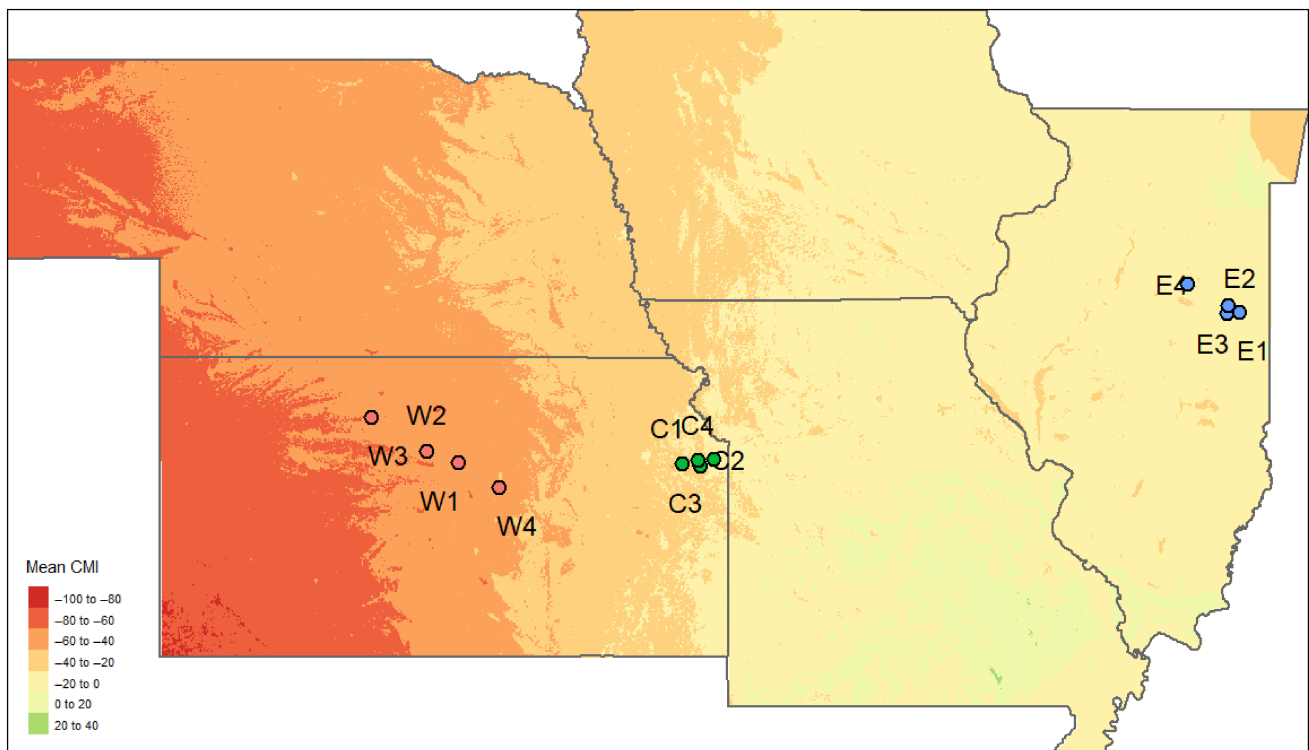

FIGURE 1 Prairie locations. Western prairies were distributed across Central Kansas. Central prairies were located in Eastern Kansas, which has a much wetter climate than Central Kansas. Eastern prairies were in Illinois, which has a similar level of precipitation to Eastern Kansas, but has higher Climate Moisture Index (precipitation minus evapotranspiration), which allows more water to be utilized by plants. Circles are filled according to the region-wise color scheme used throughout this manuscript.

TABLE 1 Location information of the 12 pristine prairie remnants from which the *Silphium integrifolium* plants were sourced.

Source prairie	Site name	Latitude	Longitude	Climate moisture index (kg/m^2/month)
W1	W1	38.9509	−98.151	−55.2
W2	W2	39.0668	−98.5651	−55.4
W3	W3	39.4016	−99.2865	−57.3
W4	W4	38.6963	−97.6129	−48.7
C1	Prairie Park	38.9331	−95.2152	−24.5
C2	Ogg Prairie	38.9897	−94.7989	−21.1
C3	Kill Creek	38.9173	−94.9745	−26.4
C4	Lexington Lake	38.9684	−95.0102	−16.8
E1	Pellisville	40.461	−87.9242	−12.9
E2	Prospect	40.4451	−88.0973	−13.3
E3	Loda	40.5273	−88.0765	−14.1
E4	Weston	40.7467	−88.6143	−20.7

Note: Climate Moisture Index is calculated as the difference between precipitation and potential evapotranspiration, which is an indicator for the availability of moisture for plants to uptake.

R genes), often contain an N-terminal domain that is either of a coiled-coil type, called CNLs, or a toll-interleukin-like type (TNLs), whose divergence predates the split between monocots and eudicots (Meyers et al., 1999), estimated to be about 160 Mya by TimeTree (Kumar et al., 2017). R genes are highly diverse, numbering in the hundreds in many plant species (Jupe et al., 2012; Toda et al., 2020). These genes evolve rapidly and are enriched in presence/absence variation even at the intraspecies level, limiting the utility of a single reference genome to elucidate the diversity of R genes within a species. This study employed R-gene enrichment sequencing (RenSeq; Jupe et al., 2013), which uses DNA baits to reduce complexity of genomic DNA libraries. The baits we designed come from the common sunflower, *Helianthus annuus* L., which is separated from *S. integrifolium* by between 22 (Meireles et al., 2020) and 33 million years ago (Zhang et al., 2021). We sequenced the RenSeq libraries from *S. integrifolium* plants across the rainfall gradient using both Illumina short-read and PacBio long-read technologies.

RenSeq is not a new technology—having been developed in 2013. It has been used to address basic problems in plant immunity (Jupe et al., 2013) and applied questions in plant breeding in numerous crop species and their wild relatives for crop improvement (e.g., Arora et al., 2019) and to understand mechanisms of plant immunity (e.g., Witek et al., 2021). Despite the utility of RenSeq in plant breeding projects, and in helping us better understand plant disease resistance, this technology has not been used to address basic evolutionary and biogeographical questions.

Using both simple linear models as well as mixed effect models, we found a significant positive correlation between the effective precipitation of a plant's host prairie and the number of R genes detected in both the Illumina and PacBio datasets. While this observation does not itself prove the broader theory that disease resistance correlates with rainfall in plants, it is certainly consistent with theoretical expectations, and should motivate additional evaluation of these ideas in other taxa. Because we demonstrate the economic utility of the RenSeq approach in non-model systems, our work not only points to data consistent with our hypothesis but paves the way toward evaluating the generality of this finding.

2 | METHODS

2.1 | Plant materials

Silphium integrifolium seeds were collected from 12 prairies in 2017 and 2018 (Figure 1; Table 1). The prairie remnants collected spanned from central Kansas to central Illinois. The prairie remnants have variable historical, geographic, and climatic characteristics described in Cassetta et al. (2023). From each of the 12 prairies, seeds were collected from approximately 40 mother plants randomly chosen from throughout the prairie remnant. As *S. integrifolium* is an obligate outcrosser, seeds collected from a single mother were a half-sibling family.

Silphium integrifolium seeds were germinated at room temperature together in a heated greenhouse at The Land Institute, Salina KS in 2019. A total of 96 genets comprising eight individuals from each prairie remnant from different mothers were selected from the seedlings. Selection of the 96 individuals was randomized although the seedlings chosen had to be moderately vigorous. The pots containing the seedlings were randomized and placed on greenhouse benches. Pots were re-randomized several times during the season. The plants grew rapidly but remained as vegetative rosettes. Each crown (genet) was divided into four pieces (ramets) with approximately equal numbers of dormant buds and roots, although a few crowns were too small and only produced 2–3 pieces. The ramets were repotted into 10-gallon fabric planters (Smart Pot) using Pro-Mix BX.

2.2 | Tissue collection

One ramet from each genet was grown in The Land Institute Greenhouse until young leaves had emerged. On March 12, 2020, black plastic pots were placed over one clone of each of the 99 plants emerging rosette and left for 48h to slightly etiolate the young leaves, which improves the quality of extracted DNA in photosynthetic active tissue containing phenolic compounds (Dabo et al., 1993). On March 14, 3–5 unexpanded leaves were excised from each crown. The collected tissue was placed in a labeled envelope and immediately placed on dry ice in a cooler. Within 60min of collection, the sampled leaf tissue was placed in a lyophilizer and lyophilized for 48h.

2.3 | Clone placement across the gradient and phenotyping

After young leaf tissue was collected from plants in the greenhouse, the plants remained in the greenhouse under ideal conditions allowing them to grow rapidly. One cloned ramet from each of the available genets was then moved to each of the following locations within the regions the seed was sourced from: (1) breeding plots in Salina, KS at The Land Institute (West), (2) Prairie Park Nature Center, Lawrence, KS (Central), (3) Research plots at the ISU horticulture center in Normal, Illinois (East), and (4) Land Institute Crop Protection Ecology greenhouse (backup site to conserve genets in a disease- and insect-free state). The pots remained undisturbed at each site for 2years, 2020 and 2021 except for the Central site cloned plants, which, in May 2021, were moved from the Prairie Park Nature Center to the Perennial Agriculture Project field station, approximately, 6miles away. The clones received occasional watering in each site to promote disease growth and avoid plant death due to the above ground pots drying out.

In 2020 and 2021, disease scores and insect herbivory damage were recorded on each of the clones in all three sites. Data were collected during their flowering period, between the months of July and September. Diseases were recorded according to methods in Cassetta et al. (2023). Briefly, leaf rust severity was quantified by visually assessing the total percentage of leaf area affected (0%–100%) using the rust scale described by Turner et al. (2018). Similarly, stem rust severity was recorded as the total percentage of the stems covered in rust (0%–100%). Leaf blotch was quantified by assessing the total percentage of leaf area affected (0%–100%). *Silphium* clear vein was rated on a 0–10 scale based on symptoms observed on the entire plant including leaf distortion (narrow width, curled, toothed, twisted, or pinched tips), and twisted meristem symptoms. Leaf herbivory scores were collected by quantifying the percentage of total leaves with evidence of insect herbivore (mining, sucking, chewing) damage (0%–100%).

2.4 | Climate moisture index

A BIOCLIM+ raster map of the continental United States containing 30-year climate moisture index averages (1981–2010) was downloaded from the CHELSA (Climatologies at high resolution for the Earth's land surface areas) database (Brun et al., 2022). The climate moisture index is calculated as the difference between mean annual precipitation and mean potential evapotranspiration. Smaller CMI values (and in the case of this study, more negative), represent drier climates, and places where abiotic stress is likely higher.

2.5 | Sequencing/RenSeq/QC

DNA was extracted from lyophilized leaf tissue using the CTAB method by the University of Minnesota Genomics Center. DNA was then shipped on dry ice to Arbor Biosciences for sequencing, library preparation, and RenSeq data acquisition. Extracted DNA was prepared for sequencing using Illumina library prep kits. Sequencing libraries were enriched for R genes using the RenSeq protocol described in Jupe et al. (2013) using Arbor's myBaits systems. In short, biotinylated single-stranded RNA oligo baits were designed based on annotated R genes from *Helianthus annuus*, a relative of *Silphium integrifolium*, within the same subtribe, estimated to be between 22.5 (Meireles et al., 2020) and 33.5 million years divergent (Zhang et al., 2021). Hybridized oligos were pulled down using streptavidin-coated magnetic beads. Libraries were then sequenced on an Illumina NovaSeq S4 sequencer using paired-end 150bp reads by Arbor Biotechnologies in Cambridge, MA.

A subsample of 10 individuals was selected for a complementary R gene analysis using Pacific Biosciences (PacBio) long-read sequencing. High-molecular-weight enriched DNA from three individuals from each of the West, Central, and East prairies, as well as one individual derived from a cultivated lineage, were prepared for long read sequencing and then sequenced on a Pacific Biosciences Sequel II using circular consensus sequencing (CCS) by Arbor Biotechnologies.

A total of 99 Illumina libraries (96 from prairie remnants + 3 cultivated "elite" lineages) and 10 PacBio libraries were sequenced. Both library types were filtered for sequence quality using Trimmomatic v0.39 (Bolger et al., 2014) and then assessed using fastqc v0.11.7 (Andrews, 2010). The Illumina libraries were subsampled to a standardized assembly input size of 1Gb to avoid biasing R gene counts by library input size. The PacBio libraries were not subsampled to a standard input size, but a correlation between input size and total R gene count in this data type reveals no dependence of final assembly size on initial input size (Figure S3; $p = .414$). Six Illumina libraries failed to meet the input threshold and were excluded from analysis. The libraries that met the input threshold were then de novo assembled using SPAdes v3.15.3 (Bankevich et al., 2012) with -k 21,33,45,55,65,81.

2.6 | Draft genome assembly

A chromosome-scale genomic assembly of an F1 *Silphium integrifolium* × *S. perfoliatum* is being prepared at the time of writing. The *S. integrifolium* mother was a leaf blotch resistant, large-headed selection from a breeding population that had undergone approximately six cycles of recurrent selection for increased number of fertile ray florets ("feminization") per head and disease resistance. The founders of the breeding population were sourced from wild central Kansas populations, that is, located within the geographic region we call "West" in this study. Flash-frozen tissue was sent to Arizona Genomics Institute for high-molecular-weight DNA extraction, which was then prepared for PacBio Sequel-II CCS sequencing at the Genome Sequencing Center at HudsonAlpha Institute for Biotechnology. The libraries were sequenced at roughly 21.5× depth assuming a 10 Gb genome size. Reads were assembled de novo using hifiasm assembler (Cheng et al., 2021) using default parameters.

2.7 | R gene annotations

We annotated the Illumina RenSeq libraries, PacBio RenSeq libraries, and the draft genome for their R gene content. Each assembly was annotated for R genes using NLR-Annotator (Steuernagel et al., 2020), which searches for common R gene peptide motifs in protein-coding sequence. The annotations were performed with the "-a" flag, which outputs the amino acid sequence of the nucleotide-binding domain (NBARC domain). We used the NBARC sequences of the draft genomic assembly to infer a maximum likelihood tree in RAxML v.8.2.11 (Stamatakis, 2006, 2014) using the peptide substitution model "PROTGAMMAAUTO" with 100 rapid bootstraps.

The two broad, non-overlapping classes of R genes that we analyzed in this study, TNLs (containing a toll-interleukin-like type N-terminal domain) and CNLs (containing a coiled-coil type N-terminal domain), were identified from the annotations based on diagnostic motifs. To identify TNLs, we parsed annotations for the motifs 13, 15, and 18 (as defined in Table S1 of Jupe et al., 2013). For CNLs, we searched for motifs 2, 6, 16, and 17. These two sets of motifs occur only in their respective class of N-terminal domain.

2.8 | Principal coordinates analysis

To test for associations between R genes and the phenotypes collected on the cloned plants across the gradient, we performed simple linear modeling as well as mixed effect modeling, which treats CMI as a fixed effect, and source prairie as a random effect. We found one significant correlation between R Genes of the various populations and leaf blotch. To interrogate which R genes were responsible for the association between leaf blotch and R gene counts, we performed a census of which R gene NBARC sequences occurred in which individuals (Table S5). The census was then used to generate a principal coordinate analysis in the R package "vegan"

(Oksanen, 2015) against each of the five phenotypic covariates included in the study. *p*-values were corrected for multiple tests using the Benjamini and Hochberg method (1995).

2.9 | Library enrichment calculations

The fasta files containing the annotated R genes for each Illumina library were concatenated into a fasta containing every observed R gene from the Illumina dataset (hence, a "Pan-NLRome"). Reads from each Illumina RenSeq library, as well as 11 1×-depth WGS libraries from seven *S. integrifolium*, two *S. perfoliatum*, one *S. laciniatum*, and one *S. terebinthinaceum*, were mapped to the Pan-NLRome using *bwa mem* (Li, 2013). The fraction out of the total reads in the library that mapped to the Pan-NLRome with a SAM CIGAR mapping score of between 148 and 151 M was calculated, representing reads that mapped to an R gene in the Pan-NLRome with between 98.01% and 100% identity.

3 | RESULTS

3.1 | Reference genome assembly

The draft genome used in this study was 9.9 Gb in length and totalled 2584 contigs. The longest contig was 182 Mb long and the mean contig size was 3.8 Mb. The L50 (which is the number of contigs, ordered longest-to-smallest, to comprise 50% of the assembly size) was 77, and the N50 (the size of the contig that passes the 50% assembly size threshold) was 40.2 Mb.

The maximum likelihood tree constructed from the R genes in the draft assembly was annotated according to peptide motifs that are diagnostic for either CNL or TNL identity (Figure 2). A deep evolutionary divergence between the two broad classes is seen, although we observe isolated instances of R genes falling incongruently within the wrong clade.

3.2 | Library enrichment

The libraries of seven *S. integrifolium* that were sequenced using WGS, and thus were unenriched for R genes, contained a median of 1.78% of reads that mapped with high fidelity (98% identity or higher) to contigs in the Pan-NLRome. Libraries that had been enriched for R genes using oligo baits designed from *Helianthus annuus* contained a median of 63.4% of high-fidelity reads mapping to R genes, representing a 36-fold enrichment for R genes (Figure 3).

3.3 | R gene counts by geographic region

The Illumina R gene assemblies ranged from 2.2 to 51 Mb, and averaged 15 Mb in size. We detected an average of 386.3, 410.0, and 414.0 R genes in plants sourced from the Western, Central, and
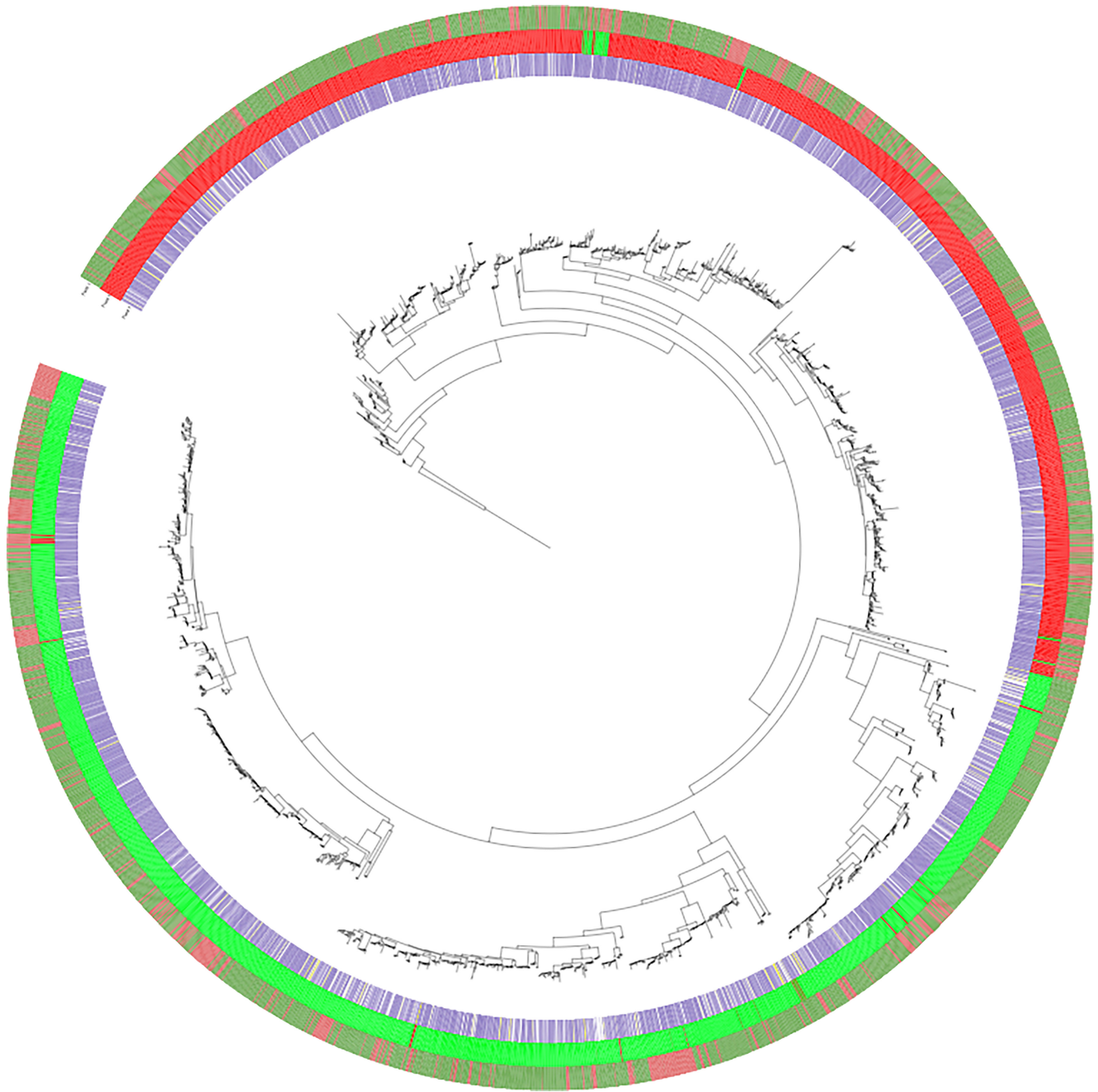
FIGURE 2 ML tree constructed from the peptide sequence of the nucleotide-binding domain (NB-ARC) of 590 R genes in the genome of the draft *Silphium integrifolium* genome. Color bar represents the probability that a sequence belongs to a TNL (green) or not (red).

Eastern prairies, respectively (left facet of Figure 4a). The plants averaged 47.38 (West), 48.69 (Central), and 45.14 (East) TNL genes, and 166.1 (West), 171.3 (Central), and 175.3 (East) CNL genes (Table S2).

The PacBio R gene assemblies ranged from 4.9 to 13 Mb, but averaged 8 Mb. Within this dataset, we only had three plants per region, so the inferential power was lower, but our confidence in the contigs assembled was higher due to longer reads. We found an average of 355.0, 563.3, and 592.7 R genes in the Western, Central, and Eastern prairies, respectively (right facet of Figure 4a). Within this dataset, we found an average of 54.00 (West), 81.67 (Central), and 81.33 (East) TNLs, and 121.3 (West), 158.7 (Central) and 159 (East) CNLs (Table S2).

As compared to the Illumina data, PacBio assemblies resulted in more "complete," as compared to "partial" genes for both intact and pseudogenized R genes (Figure 4b), indicating that many of the R genes predicted as pseudogenes in the Illumina data likely represent complete genes that are mis-categorized due to assembly limitations.

3.4 | Comparing to reference genome

The reference genome, which was sequenced at 21.5× coverage using PacBio long reads, contained a greater number of R genes
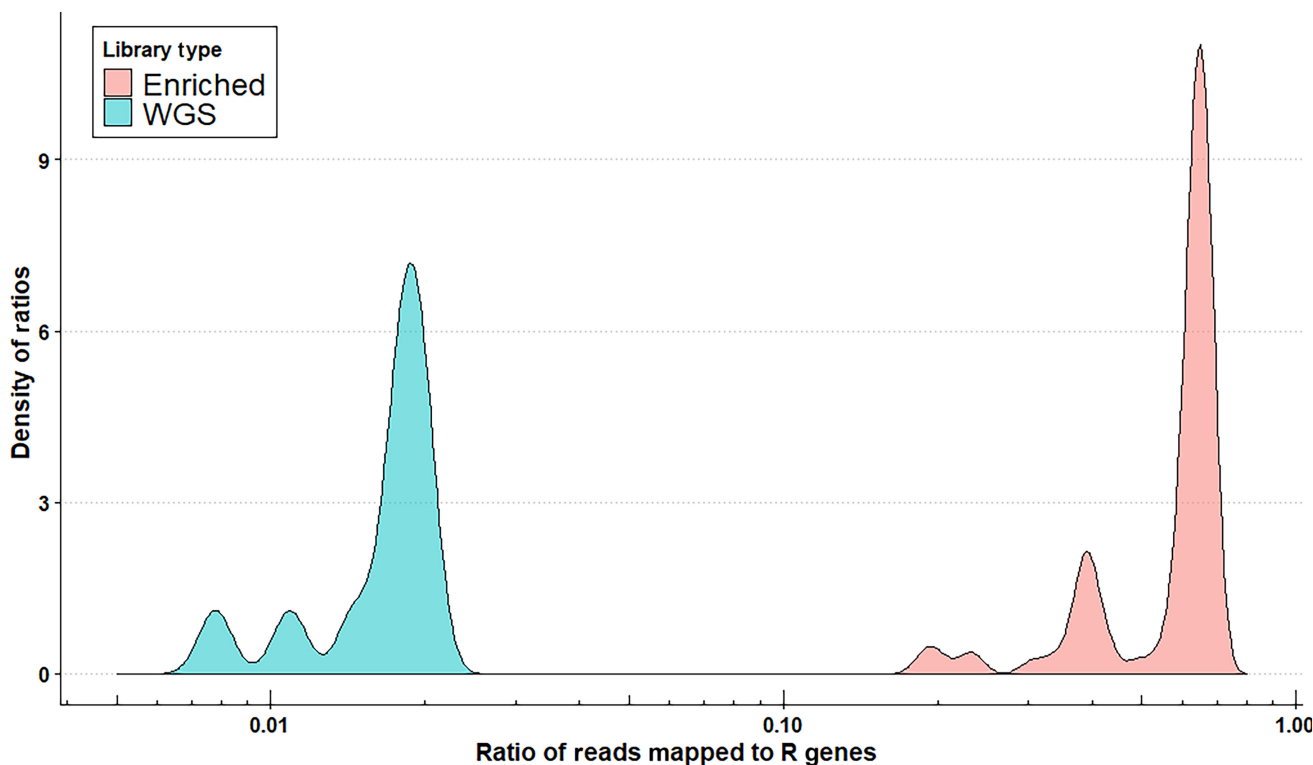
FIGURE 3 All R genes among all samples in the dataset were concatenated into a single fasta file (Pan-NLRome). The raw libraries from each sample were mapped to the Pan-NLRome. The ratios of the number of mapped reads to the total read count in each library were calculated. A median of 63% of reads in the enriched libraries were putatively sourced from R genes and 1.8% of reads in the un-enriched WGS libraries. Note the log scale of the x-axis.

in total (873; Table S2), of which 281 and 147 were annotated as CNLs and TNLs, respectively. This compares to the 603 candidates found by NLR-Annotator in the closely related sunflower genome (Toda et al., 2020). It contained a comparable percentage of complete R genes relative to its total number of R genes (50.7%) as the enriched PacBio libraries (58.2%, 54.2%, and 57.3% in the West, Central, and East, respectively). Eight hundred and thirty-seven R genes in the reference (as well as the 281 and 147 CNLs and TNLs, respectively) exceeded counts from both the lower-coverage, R-gene-enriched PacBio assemblies or the Illumina short-read assemblies, below, suggesting either that the reference had more R genes (which is plausible as it an interspecific F1) than the rest of the samples and/or that some genes were missed in the enrichment process.

3.5 | R gene counts increase with increasing climate moisture index

We found a strong positive association between R gene count and CMI in both the Illumina and PacBio datasets (Figure 5a,b; Table S3) from the simple linear model. We also found a significant positive correlation between Illumina R gene counts and CMI using the mixed effect model ($MS_{CMI} = 9493$, $MS_{error} = 1059.6$, $F = 8.96$; Degrees of freedom from Satterthwaite's method [Satterthwaite, 1946]:

$df_{CMI} = 1$, $df_{error} = 9.73$, $p = .0139$, Figure 5a; Table S4), on top of substantial variation among source prairies. Specifically, the model predicted an additional 0.707 ± 0.236 R genes for every additional unit of CMI. As average CMI ranged from −57.3 units, in the most arid prairies sampled, to −12.9 units in the highest-rainfall locations, the model predicts that plants from prairies with the highest annual CMI will have 31.4 more R genes than those from prairies with the least CMI. Because most variation in CMI is associated with source region (West, Central, or East), we also modeled source region, rather than CMI as a fixed effect. This analysis revealed a similar, but nominally insignificant negative association between source region and the number of R genes ($MS_{region} = 8753$, $MS_{error} = 4376.5$, $F = 4.13$; Degrees of freedom from Satterthwaite's method: $df_{region} = 2$, $df_{error} = 8.82$, $p = .0542$, see descriptive statistics in the section above and in Table S4).

Our PacBio sequencing recovered more R genes per individual and had relatively fewer R genes annotated as "partial" (relative to "complete") than the Illumina data (Table S2). Despite much lower inferential power, we still see a significant positive association between CMI and R gene count in the PacBio dataset ($MS_{CMI} = 92,006$, $MS_{error} = 9961$, $F = 7.193$, $df_{CMI} = 1$, $df_{error} = 7$, $p = .0315$, Figure 5b). Like the case for the Illumina data, modeling R gene count from the PacBio data revealed a similarly consistent but nominally insignificant negative association ($MS_{region} = 53,074$ $MS_{error} = 10,164$, $F = 4.96$; Degrees of freedom $df_{region} = 2$, $df_{error} = 6$, $p = .054$, see descriptive statistics in the section above and in Table S4).
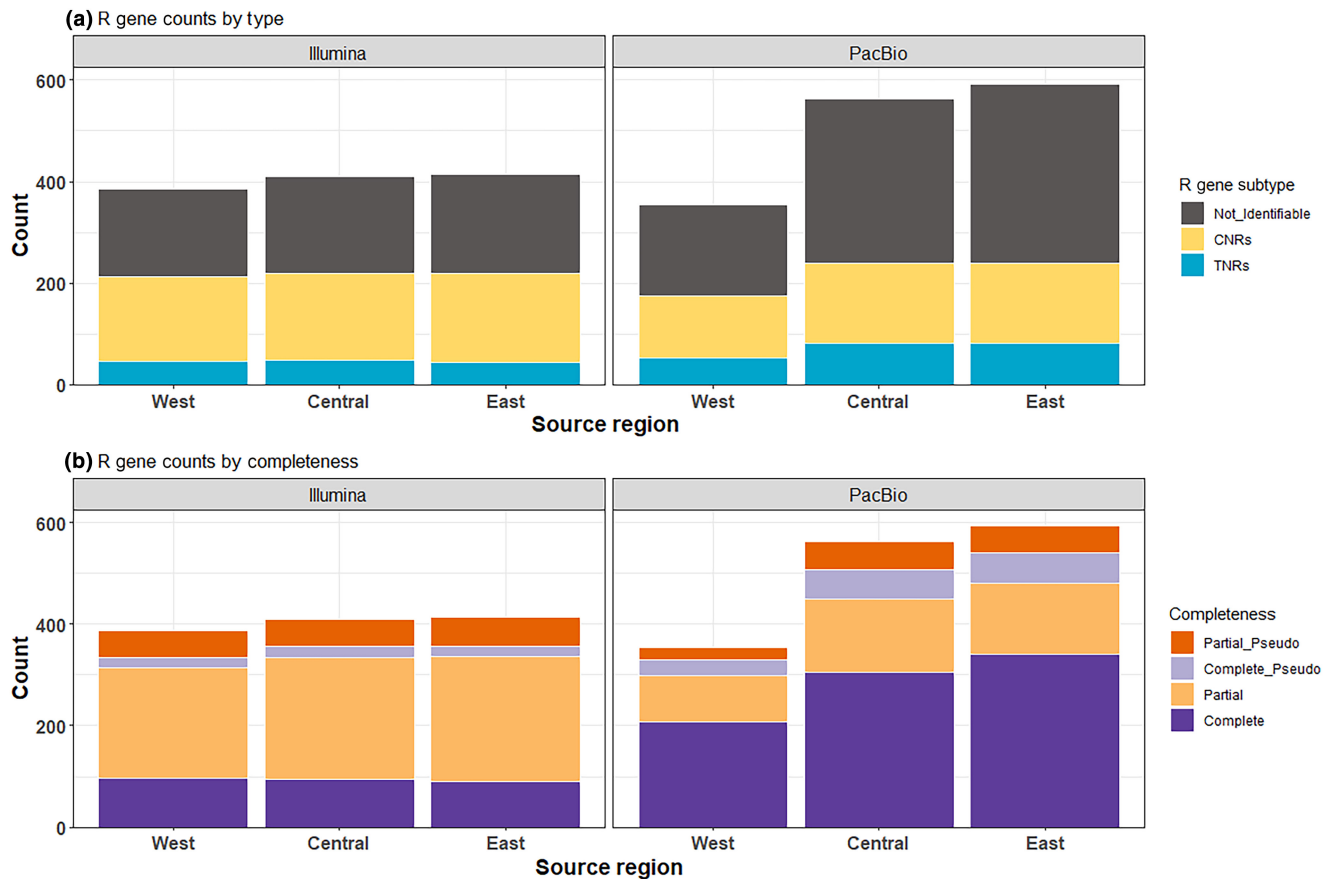
FIGURE 4 (a) Breakdown of R genes annotated by whether they contained a coiled-coil N-terminal domain, a toll-interleukin-like N-terminal domain, or neither. Those that were not identifiable either lacked a diagnostic N-terminal domain by their nature, or due to incomplete assembly. (b) R genes broken down by their completeness and inferred functionality. Genes labeled "pseudo" are likely pseudogenes due to the presence of internal stop codons. The higher number of "Complete" genes in the PacBio assemblies compared to the Illumina libraries are likely due to more reliable assembly from the long reads.

3.6 | CNLs and TNLs

We did not find a significant relationship between CMI and TNL counts ($p = .695$, Figure 5d, $MS_{region} = 14.072$, $MS_{error} = 86.559$, $F = 0.1626$; Degrees of freedom from Satterthwaite's method: $df_{region} = 1$, $df_{error} = 10.215$), and only a marginal significance with CNL counts (Figure 5c; $p = .0634$, $MS_{region} = 651.19$, $MS_{error} = 150.22$, $F = 4.3349$; Degrees of freedom from Satterthwaite's method: $df_{region} = 1$, $df_{error} = 10.22$).

Due to having only one plant sampled per prairie, we report the results of a fixed-effect linear regression model for the PacBio dataset. Despite a much lower inferential power ($n = 9$), we still found a positive correlation between CMI and the PacBio R genes (Figure 5b; $p = .019$). There was a significant positive relationship between CMI and PacBio TNL counts (Figure S1B; $p = .018$), but not CNLs (Figure S1A; $p = .14$).

3.7 | Correlations between phenotypes and R gene counts

There were no significant correlations between Illumina R gene counts and leaf rust severity (Figure 6b; $p = .63$), rust response

(Figure 6c; $p = .73$), stem rust severity (Figure 6a; $p = .51$), or insect leaf herbivory (not pictured; $p = .23$). We found that R gene counts were negatively correlated with leaf blotch scores (Figure 6d; $p = .0043$). When source prairie was modeled as a random effect, leaf blotch scores were still highly, negatively correlated with CMI ($p = 8.94e\text{-}5$, slope $= -0.0082$).

3.8 | Principal coordinate correlations among phenotypes

The PCoA analysis (Figure S4) of the unique NBARC sequences (Supplemental file "AllNBARCs.fasta"; Table S5) against the disease phenotypes yielded no significant candidates for genes responsible for disease resistance after correcting for multiple tests.

4 | DISCUSSION

We enriched *Silphium integrifolium* libraries for R genes using baits developed from the related *Helianthus annuus* genome to characterize the diversity across a rainfall gradient. We found that the
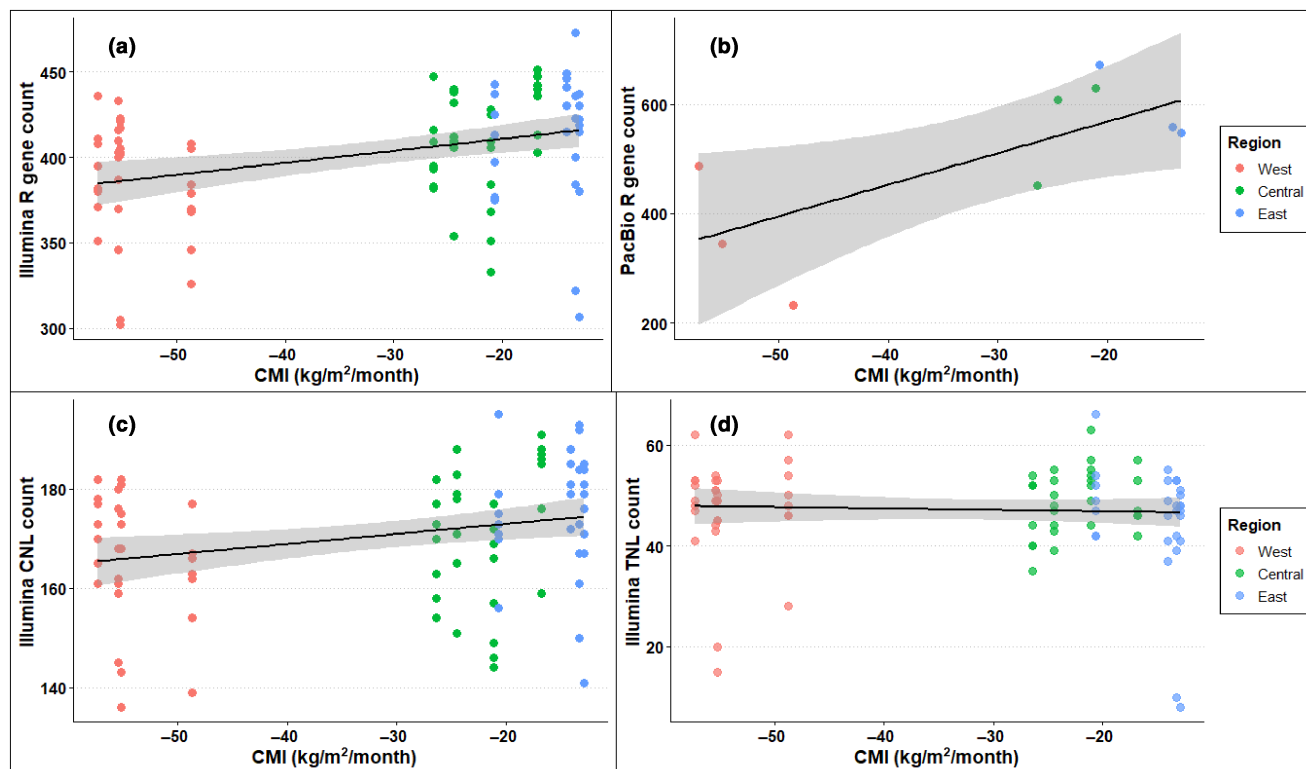
FIGURE 5 (a) Total R gene counts recovered from Illumina short-read libraries plotted against the yearly CMI, or the amount of precipitation that is stored in the soil and not lost to evaporation. (b) Total R gene counts recovered from the PacBio long-read libraries. (c) Counts of R genes from the short-read libraries that were identified as type coiled-coil (CNLs). (d) Counts of R genes from the short-read libraries of type toll-interleukin-like receptor (TNLs).

baits recovered a substantial portion of R genes in the reference. We found positive geographic associations between R gene count and rainfall. Here, we discuss the *Silphium* R gene content across the rainfall gradient, as well as the performance of the RenSeq enrichment method, and the biological implications of the geographic pattern we uncovered.

4.1 | Disease resistance and rainfall

4.1.1 | R gene counts correlate with precipitation

The significant correlation between net precipitation rate, as estimated by average CMI, and the number of R genes in a plant's genome, detected in both the Illumina short-read dataset as well as the underpowered PacBio long-read libraries ($n=9$), suggests evidence of possible genomic adaptation for higher pathogen load. In fact, the steeper relationship of R gene number and CMI observed with the PacBio long-read libraries suggests that the Illumina dataset underestimates the strength of this relationship. The trend does not appear to be driven by either TNLs or CNLs, for which the results conflict between the Illumina and PacBio datasets. Genes of type TNL, or those containing a toll-interleukin-like receptor in their N-terminal domain, do not differ significantly according to CMI. A stronger understanding of the relationship between climate factors

and the evolution of these two subtypes of R genes remains to be investigated.

4.1.2 | Mixed disease phenotype results

Despite the significant trend of the number of R genes increasing with CMI, leaf rust severity, leaf rust response, and stem rust severity were not significantly correlated with R gene count in the common gardens. These three traits are different phenotypic metrics to assess a specialist fungal pathogen, *Puccinia silphii*, commonly known as *Silphium* rust. These results are consistent with results from common gardens that found that eastern populations of *S. integrifolium* were not, on average, more resistant to *Silphium* rust (Cassetta et al., 2023). One possible explanation for why resistance to *Silphium* rust was not associated with variable R gene counts may lie in the concept of the "disease triangle" reviewed by Velásquez et al. (2018). The clones in this study were placed in environments where *S. integrifolium* pathogens were present, and experienced conditions that would be ideal for pathogenic growth: warm and wet. Even with effective immunity represented by the R genes sequenced in the plant populations, the level of resistance conferred by the resistance genes might have been obscured by the pathogen spread. Disease resistance in these wild populations is highly likely to be quantitative, meaning that resistance is not an all-or-nothing response, but
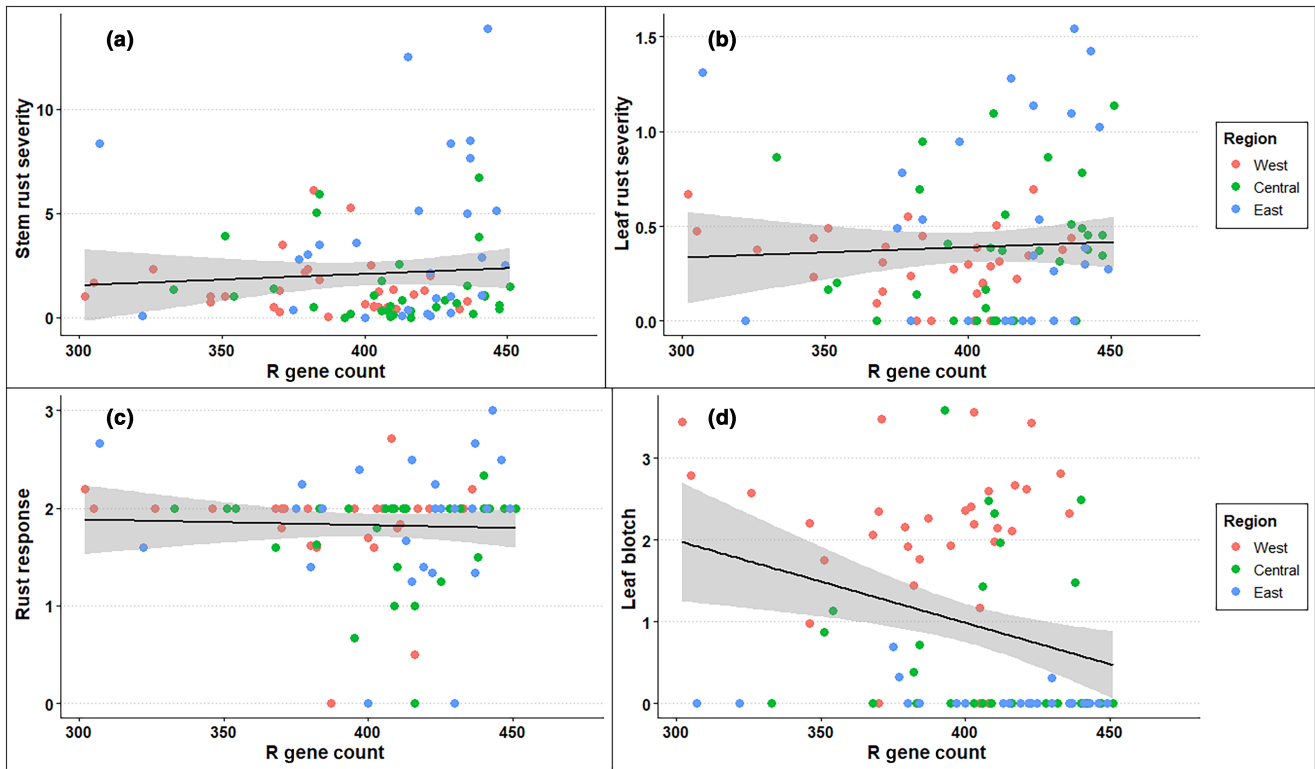
FIGURE 6 Comparisons against disease phenotypes. We found no relationship between R gene count in stem rust (a; $p = .5066$), leaf rust (b; $p = .6333$), or rust response (c; $p = .7287$). However, there was a strong negative relationship between R gene count and leaf blotch (d; $p = .004322$). This relationship is likely driven by broad differences between regions, as we find no significant relationships between R gene count and any disease phenotypes within regions (Figure S2).

rather a distribution of potential responses ranging from minor to major (French et al., 2016). Further studies of the wild populations with sequenced R genes, with controlled inoculations of plants with pathogens, would elucidate associations between R genes that may confer various levels of response to pathogens.

Further, the evolutionary dynamics of R genes may be muddled by seemingly competing forces of selection, with "arms race" dynamics expected to generate gradients in total host resistance, while oscillatory "red-queen" dynamics may not (Bergelson et al., 2002). It is possible, for example, that local co-evolution of *S. integrifolium* and *Silphium* rust has resulted in adaptation of the rust to its local host population, rather than variation in overall resistance. This red-queen dynamic is supported by observations from common gardens distributed along the gradient that showed much greater virulence of *Silphium* rust on their populations from their local region than from other regions (Cassetta et al., 2023). In the case of *Silphium* rust, R gene identity may be more important than R gene diversity.

We did observe that resistance allele diversity was associated with suppression of leaf blotch (Figure 6d). Our result affirms observations of a previous common garden study that found that eastern populations of *S. integrifolium* were more resistant to *Silphium* blotch, as well as to clear vein virus (Cassetta et al., 2023). These results are consistent with expectations of greater diversity of resistance alleles contributing to greater resistance to pathogens, and are qualitatively consistent with arms-race-type

dynamics. In *S. integrifolium*, leaf blotch is caused by the generalist fungal pathogen *Colletotrichum dematium* (Cybernome; Farr et al., 1989), as well as its more specialized congener, *Colletotrichum silphii* Davis (Horst, 2008). It is possible that arms race dynamics are more likely between hosts and generalist pathogens such as leaf blotch, while specialist pathogens such as the *Silphium* rust are more likely to generate red-queen dynamics.

We note that the observed relationship between resistance allele diversity and resistance to blotch may be spurious, as other heritable factors that covary with rainfall might contribute to disease resistance, such as plant secondary chemicals. We note, however, that insect herbivory may be an indicator of such overall resistance, and we did not observe any relationship between R gene diversity and insect leaf herbivory. This result is consistent with R genes predominantly acting as pathogen effector receptors, and not targeted toward resistance for insects (Chovelon et al., 2021).

4.2 | A broader pattern of R gene evolution

4.2.1 | R gene genealogy recovers expected patterns

The maximum likelihood genealogy constructed from the NBARC domains of the R genes of the draft genomic assembly recovers

the deep evolutionary divergence between CNL-type and TNL-type R genes that has been reported in many other studies (Meyers et al., 2003; Mun et al., 2009; Neupane et al., 2018). The instances of individual R genes nesting incongruently within the wrong clade may be attributable to N-domain switching via recombination, which, to the best of our reckoning, has yet to be observed. More likely, it may be due to misidentification of upstream nucleotide sequence as incidentally translating to peptide sequence that meets the identity threshold for one of the diagnostic motifs that places those R genes in the wrong categorization.

4.3 | Performance of RenSeq for R gene enrichment

4.3.1 | Enrichment using distant reference

The R gene libraries presented in this study were developed by enriching *Silphium integrifolium* DNA for R genes using baits developed from *Helianthus annuus*, a model organism with well-developed genomic resources. Despite an estimated divergence age of between 22.5 (Meireles et al., 2020) and 33.5 million years ago (Zhang et al., 2021), the baits were able to enrich libraries to contain a median of 63% of reads originating from R genes, representing a 36-fold enrichment over WGS. For comparison, a previous study by Andolfo et al. (2013) found success enriching the *Solanum lycopersicum* (common tomato) with baits designed from *Solanum tuberosum* (common potato), a congener estimated to be 6.7 Ma divergent by TimeTree (Kumar et al., 2017). This study demonstrates the economical promise of RenSeq for studying the immune systems of non-model plants under a variety of ecological and evolutionary pressures. It also showcases the screening of crop wild relatives that are of agronomic interest, such as *S. integrifolium*, for disease resistance genes that might enable more robust response to pathogens that pose a challenge for the domestication of the plant.

The reference genome contained a much higher number of R genes in the draft genome assembly compared to the enriched libraries (873 compared to ~400–600), which can be accounted for by two factors. First, the draft genome is assembled from an F1 hybrid between two different species, *S. integrifolium* and *S. perfoliatum*. While we expect overlap in many of the R genes due to homology, our count is likely an overestimate of the true number contained within the haploid genome of *S. integrifolium* due to the nature of R genes as a rapidly diversifying gene family. Second, RenSeq enrichment likely captures a different (and smaller) subset of the R genes than the WGS PacBio sequencing we employed for the draft assembly.

4.3.2 | Caveats and future directions

Consistent with the prediction that more rainfall results in more pathogens, which result in more disease resistance—we find a positive correlation between R gene counts and precipitation. However, the distance between our observed correlation and our causal motivation is quite large. Here we suggest two routes to bridge this gap—independent replication and unraveling the causal chain.

Correlation does not imply causation. However, repeated independent replication "does waggle its eyebrows suggestively and gesture furtively while mouthing 'look over there'." (Munroe, 2009). Thus, a key step in establishing that the correlation uncovered reflects our biological hypothesis, rather than happenstance, would be to evaluate the generality of this pattern. Some such evidence already exists—similar results have been seen in RFLP-based studies of a few R genes in big bluestem and switchgrass (Rouse et al., 2011; Zhu et al., 2013). Testing for this pattern in other *Silphium* species represents a promising direction, as they would provide evolutionarily independent replication, while covering a similar precipitation gradient, and could use the same RenSeq baits developed here. Extending this study to more distant taxa would provide further evidence supporting this hypothesis.

Functional studies of these R-genes would provide more evidence for our motivating causal hypothesis. A complete, phased, chromosome-level assembly of *S. integrifolium* will both allow for better assessment of whether these genes are functional, and enable association studies to determine the loci of pathogen resistance for incorporation into breeding programs. Additional evaluation of the hypothesis that the number of pathogens affecting *Silphium* (and/or the variation in their ability to evade a specific R gene), as well as associating specific NLR alleles to resistance to specific pathogens would allow for a more mechanistic understanding of the association uncovered in this paper.

5 | CONCLUSION

We used RenSeq to sequence the R gene libraries of *S. integrifolium* plants from across a rainfall gradient to evaluate the hypothesis that CMI imposes a biotic stress whose evolutionary pressure should be evident within the genome. We detected a significant positive relationship between CMI and R gene count, supporting our hypothesis. We demonstrate that RenSeq can be used in the evaluation of ecological hypotheses and show promise for the economical interrogation of the specific R gene loci responsible for disease resistance.

AUTHOR CONTRIBUTIONS

Kyle Keepers: Formal analysis (lead); investigation (lead); writing – original draft (lead). **Kelsey Peterson:** Investigation (supporting); writing – original draft (supporting); writing – review and editing (supporting). **Andrew Raduski:** Data curation (supporting); writing – review and editing (supporting). **Kathryn M. Turner:** Conceptualization (equal); funding acquisition (equal); project administration (equal); writing – review and editing (supporting). **David Van Tassel:** Conceptualization (equal); funding acquisition (equal); project administration (equal). **Kevin Smith:** Conceptualization (equal); funding acquisition (equal); project administration (equal). **Alex Harkess:**

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The R gene libraries generated in this study are available on the NCBI SRA database at project accession PRJNA914292.

ORCID

Kyle Keepers https://orcid.org/0000-0002-2288-8018

REFERENCES

Abbott, D. C., Brown, A. H. D., & Burdon, J. J. (1991). Genes for scald resistance from wild barley (_Hordeum vulgare_ ssp _spontaneum_) and their linkage to isozyme markers. _Euphytica_, _61_, 225–231.

Almeyda, C. V., Eid, S. G., Saar, D., Samuitiene, M., & Pappu, H. R. (2014). Comparative analysis of endogenous plant pararetroviruses in cultivated and wild _Dahlia_ spp. _Virus Genes_, _48_(1), 140–152.

Andolfo, G., Sanseverino, W., Rombauts, S., Van de Peer, Y., Bradeen, J. M., Carputo, D., Frusciante, L., & Ercolano, M. R. (2013). Overview of tomato (_Solanum lycopersicum_) candidate pathogen recognition genes reveals important _Solanum_ R locus dynamics. _The New Phytologist_, _197_(1), 223–237. https://doi.org/10.1111/j.1469-8137.2012.04380.x

Andrews, S. (2010). _FastQC: A quality control tool for high throughput sequence data_. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arora, S., Steuernagel, B., Gaurav, K., Chandramohan, S., Long, Y., Matny, O., Johnson, R., Enk, J., Periyannan, S., Singh, N., Hatta, M. A., Athiyannan, N., Cheema, J., Yu, G., Kangara, N., Ghosh, S., Szabo, L. J., Poland, J., Bariana, H., … Wulff, B. B. H. (2019). Resistance gene cloning from a wild crop relative by sequence capture and association genetics. _Nature Biotechnology_, _37_, 139–143.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. _Journal of Computational Biology_, _19_(5), 455–477.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. _Journal of the Royal Statistical Society: Series B (Statistical Methodology)_, _57_(1), 289–300.

Bergelson, J., Kreitman, M., Stahl, E. A., & Tian, D. (2002). Evolutionary dynamics of plant R-genes. _Science_, _292_, 2281–2289. https://doi.org/10.1126/science.1061337

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. _Bioinformatics_, _30_(15), 2114–2120.

Brun, P., Zimmermann, N. E., Hari, C., Pellissier, L., & Karger, D. N. (2022). Global climate-related predictors at kilometre resolution for the past and future. _Earth System Science Data Discussions_, _14_, 5573–5603.

Caplan, J., Padmanabhan, M., & Dinesh-Kumar, S. P. (2008). Plant NB-LRR immune receptors: From recognition to transcriptional reprogramming. _Cell Host & Microbe_, _3_(3), 126–135.

Cassetta, E., Peterson, K., Bever, J. D., Brandvain, Y., VanTassel, D., Lubin, T. K., Alexander, H. M., Byers, D. L., Schiffner, S., & Turner, K. (2023). Adaptation of pathogens to their local plant host, _Silphium integrifolium_, along a precipitation gradient. _Ecosphere_, _14_(6), e4565.

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., & Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. _Nature Methods_, _18_(2), 170–175.

Chovelon, V., Feriche-Linares, R., Barreau, G., Chadoeuf, J., Callot, C., Gautier, V., Le Paslier, M. C., Berad, A., Faivre-Rampant, P., Lagnel, J., & Boissot, N. (2021). Building a cluster of NLR genes conferring resistance to pests and pathogens: The story of the Vat gene cluster in cucurbits. _Horticulture Research_, _8_, 72.

Clarkson, J. P., Fawcett, L., Anthony, S. G., & Young, C. (2014). A model for _Sclerotinia sclerotiorum_ infection and disease development in lettuce, based on the effects of temperature, relative humidity and ascospore density. _PLoS One_, _9_(4), e94049.

Dabo, S. M., Mitchell, E. D., & Melcher, U. (1993). A method for the isolation of nuclear DNA from cotton (_Gossypium_) leaves. _Analytical Biochemistry_, _210_(1), 34–38.

Delavaux, C. S., Schemanski, J. L., House, G. L., Tipton, A. G., Sikes, B., & Bever, J. D. (2021). Root pathogen diversity and composition varies with climate in undisturbed grasslands, but less so in anthropogenically disturbed grasslands. _The ISME Journal_, _15_(1), 304–317.

Dong, P., Wei, Y. M., Chen, G. Y., Li, W., Nevo, E., & Zheng, Y. L. (2009). Resistance gene analog polymorphisms (RGAPs) in wild emmer wheat (_Triticum dicoccoides_) and their ecological associations. _Genetic Resources and Crop Evolution_, _56_(1), 121–136. https://doi.org/10.1007/s10722-008-9351-9

Eitas, T. K., & Dangl, J. L. (2010). NB-LRR proteins: Pairs, pieces, perception, partners, and pathways. _Current Opinion in Plant Biology_, _13_(4), 472–477.

Farr, D. F., Bills, G. F., Chamuris, G. P., & Rossman, A. Y. (1989). _Fungi on plants and plant products in the United States_. APS Press.

French, E., Kim, B. S., & Iyer-Pascuzzi, A. S. (2016). Mechanisms of quantitative disease resistance in plants. _Seminars in Cell & Developmental Biology_, _56_, 201–208.

Glover, J. D., Reganold, J. P., Bell, L. W., Borevitz, J., Brummer, E. C., Buckler, E. S., Cox, C. M., Cox, T. S., Crews, T. E., Culman, W., Dehaan, R., Eriksson, D., Gill, B. S., Holland, J., Hu, F., Hulke, B. S., Ibrahim, A. M. H., Jackson, W., Jones, S. S., … Xu, D. Y. (2010). Increased food and ecosystem security via perennial grains. _Science_, _328_(5986), 1638–1639.

Granke, L. L., & Hausbeck, M. K. (2010). Effects of temperature, humidity, and wounding on development of _Phytophthora_ rot of cucumber fruit. _Plant Disease_, _94_(12), 1417–1424.

Horst, R. (2008). _Westcott's plant disease handbook_. Springer.

Islam, T. M., & Toyota, K. (2004). Effect of moisture conditions and pre-incubation at low temperature on bacterial wilt of tomato

caused by _Ralstonia solanacearum_. _Microbes and Environments_, _19_(3), 244–247.

Jupe, F., Pritchard, L., Etherington, G. J., MacKenzie, K., Cock, P. J., Wright, F., Sharma, S. K., Bolser, D., Bryan, G. J., Jones, J. D. G., & Hein, I. (2012). Identification and localisation of the NB-LRR gene family within the potato genome. _BMC Genomics_, _13_(1), 1–14.

Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G. J., Maclean, D., Cock, P. J., Leggett, R. M., Bryan, G. J., Cardle, L., Hein, I., & Jones, J. D. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. _The Plant Journal_, _76_(3), 530–544.

Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. _Molecular Biology and Evolution_, _34_(7), 1812–1819.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. _arXiv_. https://doi.org/10.48550/arXiv.1303.3997

Magarey, R. D., Sutton, T. B., & Thayer, C. L. (2005). A simple generic infection model for foliar fungal plant pathogens. _Phytopathology_, _95_(1), 92–100.

Meireles, J. E., Cavender-Bares, J., Townsend, P. A., Ustin, S., Gamon, J. A., Schweiger, A. K., Schaepman, M. E., Asner, G. P., Martin, R. E., Singh, A., Schrodt, F., Chlus, A., & O'Meara, B. C. (2020). Leaf reflectance spectra capture the evolutionary history of seed plants. _New Phytologist_, _228_(2), 485–493.

Meyers, B. C., Dickerman, A. W., Michelmore, R. W., Sivaramakrishnan, S., Sobral, B. W., & Young, N. D. (1999). Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. _The Plant Journal_, _20_(3), 317–332.

Meyers, B. C., Kozik, A., Griego, A., Kuang, H., & Michelmore, R. W. (2003). Genome-wide analysis of NBS-LRR-encoding genes in _Arabidopsis_. _The Plant Cell_, _15_(4), 809–834.

Mun, J. H., Yu, H. J., Park, S., & Park, B. S. (2009). Genome-wide identification of NBS-encoding resistance genes in _Brassica rapa_. _Molecular Genetics and Genomics_, _282_(6), 617–631.

Munroe, R. (2009). _Correlation_ [Webcomic]. Retrieved from https://xkcd.com/552/

Neupane, S., Andersen, E. J., Neupane, A., & Nepal, M. P. (2018). Genome-wide identification of NBS-encoding resistance genes in sunflower (_Helianthus annuus_ L.). _Genes_, _9_(8), 384.

Oksanen, J. (2015). _Vegan: An introduction to ordination_. R Project. Retrieved from https://cran.r-project.org/web/packages/vegan/vignettes/intro-vegan.pdf

Reinert, S., Money, K. L., Rockstad, G. B., Kane, N. C., Van Tassel, D. L., & Hulke, B. S. (2018). Two contrasting laboratory methods improve _Silphium integrifolium_ Michx. germination rate to agronomically acceptable levels. _Euphytica_, _214_(9), 1–9.

Rouse, M. N., Saleh, A. A., Seck, A., Keeler, K. H., Travers, S. E., Hulbert, S. H., & Garrett, K. A. (2011). Genomic and resistance gene homolog diversity of the dominant tallgrass prairie species across the US Great Plains precipitation gradient. _PLoS One_, _6_(4), e17641.

Rowlandson, T., Gleason, M., Sentelhas, P., Gillespie, T., Thomas, C., & Hornbuckle, B. (2015). Reconsidering leaf wetness duration determination for plant disease management. _Plant Disease_, _99_(3), 310–319.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. _Biometrics Bulletin_, _2_(6), 110–114.

Stamatakis, A. (2006). RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. _Bioinformatics_, _22_(21), 2688–2690.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. _Bioinformatics_, _30_(9), 1312–1313.

Steuernagel, B., Witek, K., Krattinger, S. G., Ramirez-Gonzalez, R. H., Schoonbeek, H. J., Yu, G., Baggs, E., Witek, A. I., Yadav, I., Krasileva, K. V., Jones, J. D. G., Uauy, C., Keller, B., Ridout, C. J., & Wulff, B. B. (2020). The NLR-Annotator tool enables annotation of the intracellular immune receptor repertoire. _Plant Physiology_, _183_(2), 468–482.

Toda, N., Rustenholz, C., Baud, A., Le Paslier, M. C., Amselem, J., Merdinoglu, D., & Faivre-Rampant, P. (2020). NLGenomeSweeper: A tool for genome-wide NBS-LRR resistance gene identification. _Genes_, _11_(3), 333.

Turner, M. K., Ravetta, D., & Van Tassel, D. (2018). Effect of Puccinia silphii on yield components and leaf physiology in _Silphium integrifolium_: Lessons for the domestication of a perennial oilseed crop. _Sustainability_, _10_(3), 696.

Van Tassel, D. L., Albrecht, K. A., Bever, J. D., Boe, A. A., Brandvain, Y., Crews, T. E., Gansberger, M., Gerstberger, P., González-Paleo, L., Hulke, B. S., Kane, N. C., Johnson, P. J., Pestsova, E. G., Picasso-Risso, V. D., Prasifka, J. R., Ravetta, D. A., Schlautman, B., Sheaffer, C. C., Smith, K. P., … Wever, C. (2017). Accelerating silphium domestication: An opportunity to develop new crop ideotypes and breeding strategies informed by multiple disciplines. _Crop Science_, _57_(3), 1274–1284.

Velásquez, A. C., Castroverde, C. D. M., & He, S. Y. (2018). Plant–pathogen warfare under changing climate conditions. _Current Biology_, _28_(10), R619–R634.

Vilela, A., González-Paleo, L., Turner, K., Peterson, K., Ravetta, D., Crews, T. E., & Van Tassel, D. (2018). Progress and bottlenecks in the early domestication of the perennial oilseed _Silphium integrifolium_, a sunflower substitute. _Sustainability_, _10_(3), 638.

Wahl, I. (1970). Prevalence and geographic distribution of resistance to crown rust in Avena sterilis. _Phytopathology_, _60_, 746–749. https://doi.org/10.1094/Phyto-60-746

Witek, K., Lin, X., Karki, H. S., Jupe, F., Witek, A. I., Steuernagel, B., Stam, R., van Oosterhout, C., Fairhead, S., Heal, R., Cocker, J. M., Bhanvadia, S., Barrett, W., Wu, C. H., Adachi, H., Song, T., Kamoun, S., Vleeshouwers, V. G. A. A., Tomlinson, L., … Jones, J. D. G. (2021). A complex resistance locus in _Solanum americanum_ recognizes a conserved _Phytophthora_ effector. _Nature Plants_, _7_, 198–208.

Zhang, C., Huang, C. H., Liu, M., Hu, Y., Panero, J. L., Luebert, F., Gao, T., & Ma, H. (2021). Phylotranscriptomic insights into Asteraceae diversity, polyploidy, and morphological innovation. _Journal of Integrative Plant Biology_, _63_(7), 1273–1293.

Zhu, Q., Bennetzen, J. L., & Smith, S. M. (2013). Isolation and diversity analysis of resistance gene homologues from switchgrass. _G3: Genes, Genomes, Genetics_, _3_(6), 1031–1042.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.