Received: 13 August 2024 | Accepted: 2 February 2025

DOI: 10.1002/tpg2.70010

The Plant Genome

ORIGINAL ARTICLE

Special Section: Tribute to Ron Phillips: Crop Genetics, Genomics and Biotechnology

A pangenome reveals LTR repeat dynamics as a major driver of genome evolution in *Chenopodium*

Kate E. Jaggi[1] | **Karol Krak**[2,3] | **Helena Štorchová**[4] | **Bohumil Mandák**[2,3] | **Ashley Marcheschi**[1] | **Alexander Belyayev**[2] | **Eric N. Jellen**[1] | **John Sproul**[5] | **David Jarvis**[1] | **Peter J. Maughan**[1]

[1]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, Utah, USA

[2]Institute of Botany of the Czech Academy of Sciences, Průhonice, Czech Republic

[3]Faculty of Environmental Sciences, Czech University of Life Sciences Prague, Prague, Czech Republic

[4]Institute of Experimental Botany, Czech Academy of Sciences, Prague, Czech Republic

[5]Department of Biology, Brigham Young University, Provo, Utah, USA

Correspondence
Peter J. Maughan, Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT, USA.
Email: Jeff_Maughan@byu.edu

Assigned to Associate Editor David Edwards.

Funding information
TowArds Next 997 GENeration Crops, Grant/Award Number: CZ.02.01.01/00/22_008/0004581; Czech Science Foundation, Grant/Award Numbers: RVO67985939, 20-20286S; National Institute of Food and Agriculture, Grant/Award Number: 1022158

Abstract

The genus *Chenopodium* L. is characterized by its wide geographic distribution and ecological adaptability. Species such as quinoa (*Chenopodium quinoa* Willd.) have served as domesticated staple crops for centuries. Wild *Chenopodium* species exhibit diverse niche adaptations and are important genetic reservoirs for beneficial agronomic traits, including disease resistance and climate hardiness. To harness the potential of the wild taxa for crop improvement, we developed a *Chenopodium* pangenome through the assembly and comparative analyses of 12 *Chenopodium* species that encompass the eight known genome types (A–H). Six of the species are new chromosome-scale assemblies, and many are polyploids; thus, a total of 20 genomes were included in the pangenome analyses. We show that the genomes vary dramatically in size with the D genome being the smallest (~370 Mb) and the B genome being the largest (~700 Mb) and that genome size was correlated with independent expansions of the Copia and Gypsy LTR retrotransposon families, suggesting that transposable elements have played a critical role in the evolution of the *Chenopodium* genomes. We annotated a total of 33,457 pan-*Chenopodium* gene families, of which ~65% were classified as shell (2% private). Phylogenetic analysis clarified the evolutionary relationships among the genome lineages, notably

Abbreviations: ATGC, allotetraploid goosefoot complex; BUSCO, Benchmarking Universal Single-Copy Orthologs; cDNA, complementary DNA; COG, conserved orthologous gene; cp, chloroplast; *FTL*, flowering locus T-like; Hi-C, high-throughput chromosome conformation capture; IR, inverted repeat; ITS, internal transcribed spacer; LCS, large single copy; LTR, long terminal repeat retrotransposon; MYA, million years ago; ONT, Oxford Nanopore Technology; QV, quality value; satDNA, satellite DNA; SSC, small single copy; TE, transposable element.

resolving the taxonomic placement of the F genome while highlighting the uniqueness of the A genome in the Western Hemisphere. These genomic resources are particularly important for understanding the secondary and tertiary gene pools available for the improvement of the domesticated chenopods while furthering our understanding of the evolution and complexity within the genus.

Plain Language Summary

The genus *Chenopodium* consists of about 150 plant species found worldwide which are known for their climate adaptability and use as regionally important food crops, like quinoa. In this study, we analyzed the genomes of 12 *Chenopodium* species to better understand what forces shaped their evolution. We showed that species differ in genome sizes and in their number of unique gene families, some of which could help improve crops by adding traits such as disease resistance. The study also looked at how the expansion of repetitive DNA sequences and structural changes in the DNA could have led to new species. The findings are important for crop improvement, especially for plants like quinoa, while also providing valuable insights into the diversity of the genus. This work helps scientists understand how these plants evolved and how they can be used to develop better crops.

1 | INTRODUCTION

The genus *Chenopodium* L. *sensu stricto* encompasses ~150 diverse species distributed worldwide, with a concentration in temperate regions (Abugoch James, 2009; Krak et al., 2016). As members of the family Amaranthaceae and subfamily Chenopodioideae, *Chenopodium* species are known for their global distribution, with some taxa occupying distinct ecological niches, while others are invasive weeds adapted to landscape disturbances. They exhibit a range of ploidy levels, from diploid to hexaploid, with genome types labeled A–H (Mandák et al., 2018). Both wild and domesticated *Chenopodium* species have been historically significant as regionally important food crops and medicinal herbs across multiple ancient American and Eurasian civilizations. Notable among the domesticated species are *Chenopodium quinoa* Willd. and *Chenopodium pallidicaule* Aellen in the Andes, *Chenopodium berlandieri* ssp. *nuttalliae* (Saff.) H.D. Wilson & Heiser in Mesoamerica, *C. berlandieri* ssp. *jonesianum* Smith & Funk in pre-Columbian eastern North America, and domesticated forms of *Chenopodium album* L. in Formosa and the Himalayas (Jarvis et al., 2022; Kistler & Shapiro, 2011; Partap & Kapoor, 1985; Repo-Carrasco et al., 2003; Smith & Yarnell, 2009; Wilson & Heiser, 1979). The pseudocereal and vegetable chenopods are universally valued for the nutritional richness of their seed and foliar protein, tolerance to abiotic stress (salt and cold), and economic importance, especially as food security crops for subsistence farmers on the Andean altiplano (Repo-Carrasco et al., 2003; Vega-Gálvez et al., 2010).

Quinoa (*C. quinoa*) exemplifies the expanding interest in chenopod food crops. It alone occupies over 191,000 ha of land in Peru, Bolivia, and Ecuador (Basantes-Morales et al., 2019). Its seed protein content varies from 12% to 23% and includes a complete profile of essential amino acids; it is gluten-free; and it possesses the additional benefits of dietary fiber, lipids, carbohydrates, vitamins, and minerals (Dakhili et al., 2019). Quinoa is also a drought- and cold-resistant halophyte that has attracted significant production interest in ~150 countries, extending well beyond its traditional South American cultivation range (Alandia et al., 2020; Vega-Gálvez et al., 2010). In contrast, wild *Chenopodium* species, including *C. berlandieri* Moq., demonstrate markedly broader geographic distributions and a wider range of niche adaptations, thereby serving as unique secondary and tertiary genetic reservoirs for advantageous traits. These traits, particularly disease resistance, are essential for adapting the domesticated species for cultivation beyond the Andes (Poonia & Upadhayay, 2015). For example, *C. berlandieri* var. *bushianum* (Aellen) Cronquist, which produces large seeds (up to 2 mm in diameter), thrives along river bottoms and agricultural fields in eastern North America. In contrast, *C. berlandieri* var. *macrocalycium* (Aellen) Cronquist is a bushy plant that inhabits the high-tide line where seaweed accumulates on New England beaches (Maughan et al., 2024). While wild *Chenopodium* can function as genetic reserves for

valuable traits, one species in particular, *C. album* or lambsquarters, is a competitive weed causing yield reductions of 23%–36% in barley, 16%–40% in wheat, 15%–61% in soybeans, and 12%–70% in maize (Bajwa et al., 2019). Wild *Chenopodium* species thus represent a dual role, serving as valuable untapped genetic resources for enhancing domesticated crops while also presenting challenges as troublesome weeds in agriculture for other crops. Genomic characterization of weeds like lambsquarters can reveal the biological mechanisms by which these plants evolve resistance to herbicides, monopolize resources in their competitive interactions with crops, exert allelopathic influences through the rhizosphere, and otherwise respond with phenotypic plasticity to producers' cultural weed control strategies.

Taxonomically, *Chenopodium* presents challenges due to its high phenotypic plasticity, varied ploidy levels, and complex evolutionary history marked by hybridization events (Bhargava & Srivastava, 2019; Mandák et al., 2012). Recent sequence-based analyses have revealed polyphyletic origins for taxa formerly included in the broad-sense genus and emphasized the importance of genomic resource collection and conservation in elucidating genetic relationships within the genus (Belyayev et al., 2019; Jarvis et al., 2022). Speciation in Eurasian *Chenopodium* is characterized by relatively high diversity for allopolyploids and low taxonomic diversity at the diploid level, with a notable absence of diploids composed of subgenomes C, F, and G that are found in the polyploids (Mandák et al., 2018). In contrast, New World *Chenopodium* consists of a single allotetraploid complex (AABB, which includes quinoa) with a profusion of diploid taxa carrying variants of the A genome (Maughan et al., 2024; Young et al., 2023). The other continent with a unique assemblage of *Chenopodium* taxa is Australia; however, these taxa are dominated by woody perennials that have not yet been characterized genetically (Mosyakin & Iamonico, 2017).

Emerging genomic studies are helping illuminate evolutionary processes involved in *Chenopodium* genome/subgenome divergence. For example, subgenome-specific repeat sequences, specifically *Gypsy* long terminal repeat retrotransposons (LTRs), have been implicated in the dramatic increase in size of the B subgenome in *Chenopodium formosanum* (Jarvis et al., 2022), while structural variants have been linked to the genetic separation of Mesoamerican, lowland Chilean, and highland Andean quinoa ecotypes (Rey et al., 2023). A fuller understanding of the genomic landscape across the genus is essential for understanding the evolution processes at work within the family and identifying which wild taxa might serve as valuable resources for crop improvement, especially for adapting quinoa production to regions outside of the Andes (Heitkam et al., 2020; Jarvis et al., 2022). Maughan et al. (2024) recently showed that interspecific crosses between *C. quinoa* and *C. berlandieri* var. *boscianum* (Moquin-Tandon) Wahl produced viable, highly fertile hybrids with normal chromosome pairing, while crosses with other wild taxa (*C. berlandieri* var. *zschackei* (Murr) Murr ex Graebner failed to produce hybrids and the hybrids exhibited reduced fecundity, suggesting potential genomic incompatibility with some of the wild species (R. Jellen, personal communication, 2024).

The development of a pangenome for *Chenopodium* will help clarify the genomic relationships among the species and their subgenomes, including the identification of structural variants (inversions, translocations, fissions, and fusions) that can propel speciation by disrupting meiotic pairing of homologous chromosomes. From a pre-breeding standpoint, structural rearrangements are of particular interest, as they represent potential barriers to the introgression of targeted traits from wild taxa into the elite breeding lines. Indeed, a recent pangenome analysis of oat (*Avena sativa* L.) revealed multiple hidden barriers for oat improvement, including the presence of a large, unbalanced reciprocal translocation between chromosomes 1C and 1A and large pericentric inversions on chromosomes 3C and 7D (Tinker et al., 2022). Similarly, a barley pangenome study identified large inversion polymorphisms in barley (*Hordeum vulgare* L.) on chromosomes 2H and 7H and underscores the reduction of genetic recombination in inversion heterozygotes as a cause of unexplained linkage patterns and difficulties in trait segregation between breeding lines (Jayakodi et al., 2020).

In this study, we present a pangenome analysis of the *Chenopodium* genus, including the development of six new chromosome-scale assemblies and pangenome-level analyses across a set of 12 *Chenopodium* species, which encompasses all eight currently identified genome types

Core Ideas

- We report a pangenome of *Chenopodium* that encompass the eight known subgenomes (A–H) for the genus.
- Long terminal repeat retrotransposons (LTRs) are a major source of structural variation, and their independent expansions are correlated with genome size.
- Orthogroup gene phylogenies confirmed the evolutionary relationships among the eight major subgenome lineages.
- Flowering time genes show dynamic evolution, likely contributing to chenopods' success across diverse environments.
- A standardized system for chromosome naming and orientation is provided to facilitate interlaboratory research.

(A–H) representing the major evolutionary lineages in *Chenopodium*. The selected species represent diverse geographic locations throughout Eurasia and the Americas, ensuring broad coverage of taxon diversity and distribution range. From our pangenus comparative analysis, we identified core/softcore-gene clusters, shell gene clusters and cloud/private gene clusters. Private genes may be particularly useful for identifying candidate genes in the wild and weedy species that could contribute desirable and novel traits (e.g., disease resistance) for targeted introgression into elite breeding lines. Our analyses also link chromosomal structural variation and genome evolution to the expansion of LTRs in specific genome types within the genus. Additionally, we clarify the evolutionary relationships among genome types A–H through an orthogroup gene family phylogeny and hypothesize on the potential hybridization events that led to polyploids. We also present a standardized convention for naming and orienting chromosomes in *Chenopodium* genome assemblies to promote uniformity among genomic resources for *Chenopodium* and to better facilitate cross-genome comparisons. Our work provides invaluable genomic resources for further basic research, breeding efforts, and a deeper understanding of the *Chenopodium* genus.

2 | MATERIALS AND METHODS

2.1 | Species selection

Chenopodium species were selected which collectively contained genomes A–H. These species, collected in diverse locations throughout Eurasia, include *Chenopodium strictum* Roth (CD-genome tetraploid, $2n = 4x = 36$); *Chenopodium acuminatum* Willd. (D-genome diploid, $2n = 2x = 18$); *Chenopodium opulifolium* Schrad. Ex W.D.J. Koch & Ziz (BCF-genome hexaploid, $2n = 6x = 54$); *Chenopodium sosnowskyi* Kapeller (AG-genome tetraploid, $2n = 4x = 36$); *Chenopodium vulvaria* L. (H-genome diploid, $2n = 2x = 18$); and *Chenopodium pamiricum* Iljin (E-genome diploid, $2n = 2x = 18$). Additional species for which whole-genome assemblies were already available were also included to better inform pangenome analyses. These species, collected in North and South America, included *Chenopodium watsonii* A. Nelson (A-genome diploid, $2n = 2x = 18$) (Young et al., 2023); *C. pallidicaule* Aellen (A-genome diploid, $2n = 2x = 18$; Mangelson et al., 2019); *C. ficifolium* Sm. (B-genome diploid, $2n = 2x = 18$; Ludwig et al., 2025); *C. quinoa* (AB-genome tetraploid, $2n = 4x = 36$; Jarvis et al., 2017; Rey et al., 2023); *C. berlandieri* ssp. *nuttalliae* (AB-genome tetraploid, $2n = 4x = 36$; Maughan et al., 2024); and the Taiwanese species *C. formosanum* Koidz. (BCD-genome hexaploid, $2n = 6x = 54$; Jarvis et al., 2022) (Table S1; Figure S1). Genome types A–D are represented by multiple species, with five containing the A genome, five having the B genome, three species containing the C genome, and three having the D genome. The 12 species are composed of six diploids, four tetraploids, and two hexaploids. Although allotetraploid and allohexaploid species present greater complexity for genome assembly and annotation, *C. opulifolium* (BCF-hexaploid) and *C. sosnowskyi* (AG-tetraploid) were selected since no extant C, F, and G genome diploids are known.

2.2 | Tissue collection, DNA extraction, and PacBio HiFi sequencing

Seeds from *C. acuminatum*, *C. strictum*, *C. vulvaria*, *C. sosnowskyi*, and *C. opulifolium* were manually scarified and sterilized with a 10% bleach solution. The seeds were treated with 500 µL of 30 µM potassium nitrate, 500 µL of 100 ppm gibberellic acid, and Hi-Yield Captan Fungicide 50 W solution and placed on wet filter paper in petri dishes to germinate. They were kept at 4˚C for 14 days and then allowed to germinate at room temperature for 3 days. Germinating seeds were transplanted into hydroponic systems in a growth chamber at BYU (Provo, Utah) with a photoperiod of 11 h and broad-spectrum lighting. Temperatures in the growth chamber were set between 18˚C and 20˚C. The hydroponics solution consisted of 18.9 g of General Hydroponics MaxiGro Hydroponics Plant Food (10-5-14) (General Hydroponics) dissolved in 13.25 L of deionized water and was replaced every 2 weeks.

High molecular weight (HMW) DNA was extracted from fresh leaf tissue of *C. acuminatum*, *C. strictum*, *C. vulvaria*, *C. sosnowskyi*, and *C. opulifolium* using a CTAB-Genomic-tip protocol as described by Vaillancourt and Buell (2019). The extracted HMW DNA samples were quantified with the Nanodrop One/OneC Microvolume UV-Vis Spectrophotometer, and the samples were screened for quality control parameters including DNA concentration (<800 µg/mL) and contamination (260/280 and 260/230 ≅ 2.0). HMW DNA from an E-genome diploid (*C. pamiricum*) was extracted from a plant grown in a greenhouse at the Czech University of Life Sciences (Prague, Czech Republic) using a low-input extraction protocol (Russo et al., 2022). All HMW DNA samples were sheared to 17 Kb on a Diagenode Megaruptor and then made into SMRTbell adapted libraries using a SMRTbell Express Template Prep Kit 2.0 (Pacific BioSciences). Size selection was performed using a Sage Science BluePippin to select fragments greater than 10 Kb and then sequenced at the Brigham Young University DNA Sequencing Center (Provo, UT, USA) using Sequel II Sequencing Kit 2.0 with Sequencing Primer v5 and Sequel Binding kit 2.2 for 30 h with adaptive loading using PacBio SMRT Link recommendations.

2.3 | Primary contig assembly and Hi-C scaffolding

A primary contig assembly for each species was generated using hifiasm v0.16.1 (Cheng et al., 2021) with default parameters for inbred species (-l0). Hi-C (high-throughput chromosome conformation capture) was used to scaffold the primary contig assembly into pseudo-molecules. In brief, fresh leaf tissue of each species was sent to Phase Genomics for Hi-C library preparation and sequencing at the industry standard of 100 million read pairs per genome Gb. Hi-C reads were then aligned to their respective primary contig assemblies using the Burrows-Wheeler Aligner (H. Li & Durbin, 2010). Only uniquely aligned paired-end reads were retained for downstream analyses. Contigs were then clustered, ordered, and oriented using Proximo, an adapted proximity-guided assembly platform based on the LACHESIS method with proprietary parameters developed at Phase Genomics (Bickhart et al., 2017; Burton et al., 2013; Peichel et al., 2017). Gaps between contigs within the scaffolded assembly were filled with 100 Ns. Completeness of the assembled genomes was evaluated with Benchmarking Universal Single-Copy Orthologs (BUSCO) v5 (Manni et al., 2021) using the Embryophyta (embryophyta_odb10) orthologous plant gene dataset.

2.4 | Additional genomes

In addition to the six de novo assembled genomes described above, genome assemblies of six previously published *Chenopodium* species were included in our pangenome analysis (Table S1). These include the A-genome diploids *C. watsonii* and *C. pallidicaule* ($2n = 2x = 18$; Mangelson et al., 2019; Young et al., 2023), a B-genome diploid, *C. ficifolium* ($2n = 2x = 18$; Ludwig et al., 2025), the AB-allotetraploids, *C. quinoa* and *C. berlandieri* ssp. *nuttalliae* ($2n = 4x = 36$) (Maughan et al., 2024; Rey et al., 2023), and BCD-allohexaploid *C. formosanum* ($2n = 6x = 54$; Jarvis et al., 2022). Plant tissue of *C. ficifolium* was collected from the location of a naturalized, not native, population of the species. Different sequencing technologies and assemblers were used to assemble these previously published genomes. In brief, the genome of *C. watsonii* was sequenced using Oxford Nanopore Technologies MinION long reads (R9 flow cells) assembled with Canu (Koren et al., 2017) and scaffolded with Hi-C technology. The genome of *C. pallidicaule* was sequenced with short Illumina reads assembled with the ALLPATHS-LG assembler (Gnerre et al., 2011), gap-filled with long PacBio CLR reads and scaffolded with Hi-C technology. The genome of *C. ficifolium* was sequenced using PacBio HiFi reads and assembled with hifiasm (Cheng et al., 2021). *C. berlandieri* ssp. *nuttalliae* was sequenced with PacBio CLR reads and assembled with Canu, polished with Illumina short reads, and scaffolded with Hi-C technology. The genome of *C. quinoa* was sequenced with PacBio CLR reads and assembled with the smrtmake assembly pipeline, gap filled and polished with the same PacBio CLR reads used to produce the assembly, and ultimately scaffolded with Hi-C technology. Finally, the genome of *C. formosanum* was sequenced with PacBio HiFi technology, assembled with Canu, and scaffolded with Hi-C techniques (Jarvis et al., 2022). The completeness of the six additional genomes was also evaluated with BUSCO v5 (Manni et al., 2021) using the Embryophyta orthologous plant gene dataset.

2.5 | Polyploid subgenome assignment

Classifications (A–H) for each subgenome in the polyploid species were determined by mapping subgenome-specific reads from known reference *Chenopodium* assemblies onto each assembly with Minimap2 (L. Li, 2018) using the default settings for PacBio HiFi long reads (-ax map-hifi) followed by filtering of reads with a quality values <45 using Samtools view (-q 45) (Danecek et al., 2021) to ensure only high-quality reading mapping was utilized for subgenome assignments. Specifically, reads from the D-genome diploid *C. acuminatum* were mapped against the scaffolded assembly of *C. strictum* using Minimap2 (L. Li, 2018) to identify the nine chromosomes comprising the D-subgenome of *C. strictum*. The remaining nine chromosomes were subsequently assigned as belonging to the C subgenome. For AG-tetraploid *C. sosnowskyi*, reads from A-genome diploid *C. watsonii* were mapped against *C. sosnowskyi* to identify the nine A-subgenome chromosomes, with the remaining nine chromosomes being assigned to the G-subgenome. To identify the B- and C-subgenome chromosomes of *C. opulifolium* (BCF-hexaploid), Illumina short reads from B-genome diploid *C. suecicum* (Jarvis et al., 2017) and C-subgenome specific reads from the *C. strictum* reference assembly were mapped against the *C. opulifolium* scaffolded assembly, with the remaining nine chromosomes assigned as F-subgenome chromosomes. *Chenopodium acuminatum*, *C. vulvaria*, and *C. pamiricum* are D-, H-, and E-subgenome diploids, respectively, and therefore did not need subgenome classification. The chromosomes of the remaining polyploids (*C. quinoa*, *C. berlandieri* ssp. *nuttalliae*, and *C. formosanum*) had already been assigned subgenomes, and these designations were preserved in the current study.

2.6 | Genome quality assessment

Assemblies were evaluated using Inspector (Chen et al., 2021). The inspector option "–datatype hifi" was used for

evaluating the assemblies of *C. acuminatum*, *C. strictum*, *C. vulvaria*, *C. sosnowskyi*, *C. opulifolium*, *C. pamiricum*, *C. ficifolium*, and *C.* formosanum; the option "–datatype nanopore" for the *C. watsonii* assembly, and the option "–datatype CLR" for the assemblies of *C. pallidicaule*, *C. quinoa*, and *C. berlandieri* ssp. *nuttalliae*. For error correction, the script "inspector-correct" was employed with default parameters for all of the PacBio HiFi-based de novo assembled genomes.

2.7 | Chloroplast assembly and non-species contaminant identification

Included in the final genome assemblies are complete chloroplast (cp) genomes for each species. cp assemblies were generated de novo for eight of the species following the methods previously described by Jarvis et al. (2022) using the previously assembled *C. formosanum* cp genome (Jarvis et al., 2022) as a bait genome. PacBio HiFi sequencing reads from each species were mapped using Minimap2 (L. Li, 2018) to the *C. formosanum* cp assembly, and mapped reads were filtered based on the following parameters: mlen = 8000, qval = 60, and a guanine-cytosine (GC) content range of 31%–52%. The resulting selected reads were subsequently assembled with HiCanu v.2.1 (Nurk et al., 2020) with default parameters with an expected genome size of ~150 Kb. Because the cp genome of *C. watsonii* was assembled using short paired-end reads, the program GetOrganelle (Jin et al., 2020) was utilized to assembly the cp genome using k-mer sizes of 21, 45, 65, 85, 105, and 115 bp; a minimum read length of 30 bp; and the embplant database with the cp assembly of *C. formosanum* as a reference. cp assemblies for *C. quinoa* and *C. berlandieri* ssp. *nuttalliae* were used as previously reported (Jarvis et al., 2017; Maughan et al., 2019). All cp assemblies were oriented relative to the published *C. quinoa* cp assembly (Jarvis et al., 2022; MK159176). All 12 resulting cp assemblies were annotated using GeSeq (Tillich et al., 2017). A phylogeny based on the ~80 Kb large single-copy region of the cps was produced by first aligning the sequences with Muscle v3.8.1551 (Edgar, 2004) using the options "-maxiters 2" and "-diags." IQ-TREE v2.3.4 (Nguyen et al., 2015) was then used to generate a phylogenetic tree from the aligned sequences using the model finder plus option, with an ultra-fast bootstrap and SH-like approximate likelihood ratio test, both at $n = 1000$. The "-bnni" option was employed to reduce the risk of overestimating branch support.

While Hi-C scaffolding incorporated the vast majority of the primary contig sequence into pseudo-chromosome scaffolds, some small primary contigs remained unincorporated. BlobTools v.1.1.1 (Laetsch & Blaxter, 2017) was used to identify if any of the remaining, unincorporated contigs were off-species contaminants. Taxonomic classifications were assigned at the order level to these unincorporated contigs using the "view" command of BlobTools with options "–taxrule bestsumorder." Any contigs identified as non-Caryophyllales were subsequently removed from the genome assembly. Primary contigs with a cp origin were identified with a BLASTn search against the species cp genome assembly using default parameters (Ye et al., 2006). Similarly, primary contigs with a mitochondrial origin were identified using the previously published mitochondria from *C. quinoa* (GenBank accession MK182703; Maughan et al., 2019). All contigs that exhibited ≥99% sequence identity across ≥99% of the total contig length with the assembled cp or *C. quinoa* mitochondria were removed from the final genome assembly for each species.

2.8 | IsoSeq sequencing

The IsoSeq data were derived from root, leaf, stem, and whole flower tissues from *C. acuminatum*, *C. strictum*, *C. ficifolium*, *C. berlandieri* ssp. *nuttalliae*, and *C. formosanum*. Each species was processed and sequenced individually. RNA samples were extracted using the Zymo Research Direct-zol RNA MiniPrep Plus kit. The quantity and quality of extracted RNA were first tested using a Nanodrop spectrophotometer and further evaluated with a Bioanalyzer. After quality check, RNA from each of the different tissues for each species was pooled in equal molar ratios to synthesize full-length complementary DNA (cDNA) using a NEBNext single cell/low input cDNA synthesis and amplification kit (E6421L) which uses a template switching method to generate full-length cDNAs (New England BioLabs). IsoSeq libraries were prepared from the cDNA of each species according to standard protocols using the SMRTbell v3.0 library prep kit and sequenced on a single SMRT cell 8 M for each species using a PacBio Sequel II at the DNA sequencing center at Brigham Young University (Provo, Utah, USA).

2.9 | De novo annotation

Prior to annotation, repetitive elements were identified and masked using RepeatModeler2 v2.0.1 (Flynn et al., 2020) and RepeatMasker v.4.1.2 (Smit et al., 2013–2015) using methods detailed more fully in Section 2.12. To standardize the annotation across the genomes, we de novo annotated each soft-masked genome with Braker3 v.3.0.7 (Gabriel et al., 2024) using a genus-level IsoSeq dataset consisting of the combined raw IsoSeq data (see Section 2.8) as primary evidence. The Viridiplantae OrthoDB v.11 (Kuznetsov et al., 2023), Caryophyllales UniProtKB (www.uniprot.org/uniprotkb), and the NCBI Caryophyllales RefSeq (www.ncbi.nlm.nih.gov/protein) databases served as training evidence for AUGUSTUS. Braker3 initially trains GeneMark ETP with the

OrthoDB proteins and IsoSeq hints and then subsequently trains AUGUSTUS based on GeneMark ETP predictions. This process results in a set of gene models predicted by AUGUSTUS, and these AUGUSTUS and GeneMark ETP gene models are then combined using TSEBRA. Functional annotation and gene ontology terms (PANTHER-15.0, Pfam-33.1, PIRSF-3.10, PIRSR-2021_02, PRINTS-42.0, and SUPERFAMILY-1.75) were added to the gene models using BLASTp (*e*-value < 1.0e-6; Ye et al., 2006) by querying predicted protein models against the Swiss-Prot (downloaded 7/2023) and InterProScan (Jones et al., 2014) databases, respectively. The completeness of the annotations for each species was evaluated with BUSCO v5 (Manni et al., 2021) using the Embryophyta protein dataset.

2.10 | Orthogroup analysis, gene ontology enrichment, and phylogeny

A gene-based pangenome was produced using a gene family clustering strategy starting with the longest protein from each gene model. Nonredundant proteins, identified with CD-HIT v4.8.1 (-c 1 -aS 1) (W. Li & Godzik, 2006), were then clustered into gene family orthogroups using OrthoFinder2 v.2.5.4 (Emms & Kelly, 2019). Gene families that were present in all 12 species, 10–11 species, 3–9 species, 1–2 species were categorized as core, softcore, shell, and cloud, respectively. Those gene families that are found in a single species were subcategorized as private. Trapid v2.0, using default parameters and the PLAZA 4.5 dicot database, was used to investigate functional enrichment (Bucchini et al., 2021) associated with the private gene families in each species. Each enriched functional annotation is given *q*-value that is corrected for multiple hypothesis testing using the Benjamini–Hochberg correction. Summary visualization of the gene ontology (GO) terms was performed with GO-Figure (Reijnders & Waterhouse, 2021) at a maximum *p*-value < 1.0e-3.

A rooted subgenome phylogeny was produced with the inclusion of the *Dysphania ambrosioides* (L.) Mosyakin & Clemants (epazote) genome (CoGe ID #68555) using the multiple sequence alignment approach found in OrthoFinder2 and elicited with the "-M msa" argument to produce bootstrap values. *Dysphania* R. Br. is a sister genus of *Chenopodium* in *Amaranthaceae*. Orthologous and syntenic relationships across the species were visualized using riparian plots generated with GENESPACE v1.2.3 (Lovell et al., 2022). Since OrthoFinder2 uses a single representative sequence from each species for each orthogroup to build a phylogeny, we also generated a subgenome-level phylogenetic tree based on single-copy orthologs identified with Compleasm (Huang & Li, 2023) based on the eudicots_odb10 database, with *B. vulgaris* vEL10.1 (McGrath et al., 2023) included as an outgroup. The single copy orthologs were aligned using MAFFT v.7.471 (Katoh & Standley, 2013). During partition finding in ModelFinder (Kalyaanamoorthy et al., 2017) within IQ-TREE (Minh et al., 2020), protein models were selected for each newly chosen subset, and the phylogenetic tree that fit the model was estimated. A total of 1000 ultrafast bootstrap replicates were generated with the "-bb" option to assess node support, and the resulting subgenome tree was inferred using Astral v.5.7.1 (Mirarab et al., 2014).

2.11 | Genome visualization

The positions of genes (see *De novo* annotation section) were mapped across the 12 species and the positions of *Gypsy* and *Copia* LTRs were determined using RepeatModeler2 v2.0.1 (Flynn et al., 2020) and RepeatMasker v.4.1.2 (Smit et al., 2013–2015) as described in the transposable element (TE) and satellite DNA landscapes section and then mapped across the chromosomes of the 12 *Chenopodium* species. A 39-bp putative centromeric satellite subrepeat was identified using TRASH (Wlodzimierz et al., 2023) and used in conjunction with a 294-bp (12-13p; HM641822.1) centromeric repeat sequence derived from *C. quinoa* to identify centromeric regions across the 12 *Chenopodium* species (Kolano et al., 2011; Orzechowska et al., 2018). The locations of these centromeric sequences were determined across all the chromosomes of the *Chenopodium* species. The positions of telomeric repeats were determined using BLAST (Ye et al., 2006) as previously described (Jarvis et al., 2017). Subgenomes A, B, C, and D were visualized and confirmed in the seven polyploid chenopods as described in Section 2.5. The densities of genes, *Gypsy* and *Copia* LTRs, centromeric repeat sequences, telomeric repeats, and subgenome-specific reads as well as mean GC content were visualized in 500-Kb windows across all chromosomes using Circa (http://omgenomics.com/circa). GENESPACE (Lovell et al., 2022) was utilized to visualize syntenic relationships across the *Chenopodium* genomes based on the syntenic blocks identified with OrthoFinder2 (Emms & Kelly, 2019).

2.12 | Transposable element and satellite DNA landscapes

RepeatModeler2 v2.0.1 (Flynn et al., 2020) was run on each genome assembly to identify species-specific repetitive elements as described previously. The identified elements (consensi.fa.classified) were then combined with the Viridiplantae repetitive element library, which includes DNA sequences from both the Dfam and Repbase databases. RepeatMasker v4.1.2 (Smit et al., 2013–2015) was used with options "-a, -xsmall, -gff" to produce a soft-masked genome and an alignment file of the repetitive elements for all 12 species. Repeat

sequence libraries for each species have been submitted to the Dfam database (www.dfam.org). Repeat landscapes were produced for each species as well as each constituent subgenome of each species separately using divergence values calculated with RepeatMasker's calcDivergenceFromAlign.pl script. To approximate the time scale of TE evolution, divergence values of the *x*-axis in the repeat landscapes were converted into time using the equation: time = divergence/2*r*, where $r = 1.3 \times 10^{-8}$ mutations per site per year, a rate that has been previously estimated for LTR element sequence evolution in rice (Ma & Bennetzen, 2004). *Copia* and *Gypsy* TE family sequences were also isolated and used to construct separate repeat landscapes to visualize these most abundant classified TE families across *Chenopodium* species.

Previously established satellite DNA (satDNA) libraries from Belyayev, Josefiová, et al. (2020) were utilized to identify satDNA sequences in the 12 *Chenopodium* genomes based on the *C. formosanum* assembly (Jarvis et al., 2022). The satellite repeats were then masked in the 12 assemblies and their respective subgenomes using RepeatMasker (Smit et al., 2013–2015). Tandem Repeat Finder (Benson, 1999) was then run on the masked assemblies to identify novel satDNAs. SatDNA landscapes were produced, showing the abundance of each satDNA family versus the genetic divergence of each copy. This analysis specifically targeted genomes that lacked prior characterization, notably F and G.

2.13 | FTL gene analysis

FTL genes were identified in genomes of all 12 species, using *FTL*s found previously in *C. quinoa* and *C. ficifolium* as the queries by BLASTn (*e*-value < 1.0e-5). A translation-informed alignment of 94 *Chenopodium FTL* coding sequences of complete genes (excluding pseudogenes) was produced by MUSCLE (Edgar, 2004) with default parameters, as implemented in Geneious 7.1.5, and edited manually. The alignment was analyzed by the maximum-likelihood (ML) method using RAxML (Stamatakis, 2014) implemented in CIPRES (Miller et al., 2010). A gamma model of rate heterogeneity was applied on three distinct data partitions corresponding to three codon positions. Bootstrap support of the majority rule consensus tree was calculated from 1000 pseudoreplicates. Gaps were treated as missing characters.

3 | RESULTS AND DISCUSSION

3.1 | Primary contig assembly and Hi-C scaffolding

PacBio HiFi sequencing was conducted on *C. acuminatum*, *C. strictum*, *C. vulvaria*, *C. sosnowskyi*, *C. opulifolium*, and *C. pamiricum*. Across the six species, we achieved between 35.37x and 70.49x coverage (in *C. acuminatum* and *C. vulvaria*, respectively). While average sequencing read lengths ranged between 14,869 and 16,864 bp (*C. acuminatum* and *C. sosnowskyi*, respectively) for the other five species, the average read length for *C. pamiricum* was only 7064 bp. The significantly shorter average read length of *C. pamiricum* likely reflects its unique DNA extraction method (see Section 2) compared to the DNA extraction protocol used for the other species. We note that the shorter average read length did not affect the final quality of the assembled genome as seen by the genome quality value (QV) value (57.51; see Table 1 and Section 3.4). Primary contig assemblies were then generated for these species using the hifiasm assembler (Cheng et al., 2021). The total size of the assembled contigs varied among the species, ranging from a low of 421,860,573 bp (*C. vulvaria*, H-genome diploid) to a high of 1,852,144,078 bp (*C. opulifolium*, BCF hexaploid), consistent with expectations for increasing ploidy level (Table S2). The number of contigs in the primary assembly ranged from 494 to 1024 for species *C. acuminatum*, *C. sosnowskyi*, *C. opulifolium*, *C. strictum*, and *C. vulvaria*. Not unexpectedly, the contig number for the E-genome diploid, *C. pamiricum*, exhibited a higher total contig count of 3215, likely resulting from the shorter mean HiFi read length achieved for this species. Alternatively, PacBio HiFi reads are known to show an AT bias, similar to the GC bias of Illumina reads, which can lead to less continuous contig assemblies in AT-rich genomes (McCartney et al., 2022; H. Wang, Xu, et al., 2023); however, this is unlikely the case here, since the GC contents of the *Chenopodium* genomes reported here are all very similar, ranging from 35.7% to 37.0%. The six species had N50 statistics ranging from 16,704,986 bp (*C. pamiricum*) to 63,806,435 bp (*C. opulifolium*). Not unexpectedly, the lowest contig L50 statistics, 6 and 7, were observed in the diploids *C. acuminatum* and *C. vulvaria*, respectively. The size of the largest contig assembled also varied among these six species and ranged from 49,699,371 bp (*C. acuminatum*) to 98,156,632 bp (*C. opulifolium*).

In the Hi-C scaffolded assemblies of these six species, the total assembly size remained similar to those of the primary contig assemblies, reflecting only slight changes where 100 Ns were added to bridge contigs in assembled scaffolds (Table S3). As expected, the N50 values for the scaffolded assemblies were in each case higher compared to the primary contig assemblies due to improved chromosome contiguity at the scaffold level. The *C. opulifolium* scaffold assembly had the highest N50 value (63,806,435 bp) among the species. The L50 of the scaffolded assemblies for each species were consistent with the expected number of chromosomes for each ploidy level. The diploid species *C. acuminatum* and *C. vulvaria* ($n = 9$) each displayed L50 values of 5, the tetraploid species *C. strictum* and *C. sosnowskyi* ($n = 18$) displayed L50

TABLE 1 Final assembly statistics after correcting with Inspector.

Species	Diploids						Tetraploids				Hexaploids	
	C. pallidicaule[a]	*C. watsonii*[a]	*C. ficifolium*[a]	*C. acuminatum*	*C. pamiricum*	*C. vulvaria*	*C. quinoa*[a]	*C. berlandieri* ssp. *nuttalliae*[a]	*C. sosnowskyi*	*C. strictum*	*C. opulifolium*	*C. formosanum*[a]
Ploidy (genomes)	2x (A)	2x (A)	2x (B)	2x (D)	2x (E)	2x (H)	4x (AB)	4x (AB)	4x (AG)	4x (CD)	6x (BCF)	6x (BCD)
Previous genome size estimate (bp)[g]	362,500,000	547,760,000	692,000,000	403,000,000	459,000,000	392,000,000	1,390,000,000	1,296,000,000	864,000,000	822,000,000	1,732,000,000	1,630,000,000
Sequencing technology	Illumina	ONT	PacBio HiFi	PacBio HiFi	PacBio HiFi	PacBio HiFi	PacBio CLR	PacBio CLR	PacBio HiFi	PacBio HiFi	PacBio HiFi	PacBio HiFi
Assembler	ALLPATHS-LG[b]	Canu	Hifiasm	Hifiasm	Hifiasm	Hifiasm	Canu[c]	Canu[c]	Hifiasm	Hifiasm	Hifiasm	Canu
Scaffolding technology	Hi-C	Hi-C	Linkage Map	Hi-C	Hi-C	Hi-C	Hi-C	Hi-C	Hi-C	Hi-C	Hi-C	Hi-C
Total length (bp)	354,275,569	547,847,464	745,574,147	430,761,830	495,315,900	399,332,675	1,326,010,624	1,295,344,866	912,083,985	849,085,852	1,838,432,416	1,630,221,110
Genome BUSCO[d]	1565 (1%)	1558 (3.3%)	1565 (1.4%)	1578 (1.4%)	1573 (1.4%)	1580 (1.1%)	1586 (80.5%)	1526 (72.4%)	1577 (84.7%)	1582 (85.6%)	1569 (90.2%)	1573 (90.9%)
No. of scaffolds	247	1375	329	191	956	381	2927	335	277	274	341	800
N50 (bp)	38,120,756	55,144,696	79,863,818	45,290,957	51,596,856	41,253,042	66,924,138	70,051,864	53,423,613	51,621,315	70,761,739	61,076,155
L50	5	5	5	5	5	5	9	9	8	8	12	11
Largest scaffold (bp)	45,403,159	64,479,649	92,986,964	50,028,913	58,193,823	53,616,385	87,284,847	91,041,477	62,931,823	62,289,035	98,161,049	92,862,483
N count	356,369	187,907	100	800	4600	1700	1,280,475	87,300	1000	647	100	4405
Gaps	5075	5338	1	8	46	17	1470	873	10	7	1	46
Cp genome (bp)[e]	151,282	152,098	151,826	152,067	152,069	152,309	152,079	152,125	152,241	152,194	152,194	152,090
Gene models[f]	18,900	21,804	24,816	22,740	21,677	19,211	48,376	48,620	41,186	44,880	68,671	67,266
Protein BUSCO[d]	1549 (6.9%)	1542 (10.4%)	1541 (7.2%)	1557 (8.3%)	1573 (8.9%)	1552 (7.2%)	1560 (89%)	1559 (90%)	1568 (92%)	1569 (91.7%)	1570 (95.2%)	1571 (94.7%)
Genome QV	27.82	25.07	76.93	42.82	57.51	75.82	39.78	43.34	61.63	48.13	56.88	62.08

Note: Contaminant scaffolds identified with BlobTools as well as any scaffolds identified as derived from the chloroplasts and mitochondria were removed. Primary contig and scaffolding statistics prior to error correction can be found in Tables S2 and S3. The genomes of *C. acuminatum*, *C. strictum*, *C. vulvaria*, *C. sosnowskyi*, *C. opulifolium*, and *C. pamiricum* were assembled de novo, while the genomes of *C. watsonii*, *C. pallidicaule*, *C. ficifolium*, *C. quinoa*, *C. berlandieri* ssp. *nuttalliae*, and *C. formosanum* were previously reported as indicated.

Abbreviations: cp, chloroplast; Hi-C, high-throughput chromosome conformation capture; ONT, Oxford Nanopore Technology; QV, quality value.

[a]Genome assemblies for species from previous publications: *C. pallidicaule*—Mangelson et al., 2019; *C. watsonii*—Young et al., 2023; *C. ficifolium*—Ludwig et al., 2025; *C. quinoa*—Rey et al., 2023; *C. berlandieri* ssp. *nuttalliae*—Maughan et al., 2024; Jarvis et al., 2022.

[b]Gap filled with long PacBio CLR reads.

[c]Polished with Illumina reads.

[d]Number of complete COGs identified with BUSCO in the embryophyta_odb10 dataset ($n = 1614$) for the genome and annotated proteins. Percent duplicated COGs is provided in parenthesis. See also Figure S6.

[e]128 genes, 8 rRNAs, and 31 tRNAs were identified by GeSeq (Tillich et al., 2017) for all chloroplasts.

[f]Gene models identified by Braker3 in the chromosomes for each species include alternative splicing variants.

[g]Previous genome size estimates for *C. pallidicaule* (Mangelson et al., 2019), *C. watsonii* (Young et al., 2023), *C. quinoa* (Jarvis et al., 2017), *C. berlandieri* (Maughan et al., 2024), *C. formosanum* (Jarvis et al., 2022), and all remaining species (Vít et al., 2016).

values of 8, and the hexaploid *C. opulifolium* ($n = 27$) displayed an L50 value of 12. The number of gaps among the scaffolded assemblies ranged from 1 to 46, with a mean of 14.83, and although *C. pamiricum* had a similar scaffold N50 and L50 statistics to the other species, it had significantly more gaps (46).

3.2 | Additional genomes

We note that our de novo assembled genomes generally showed similar genome statistics with those of the previously assembled *Chenopodium* genomes (*C. pallidicaule*, *C. watsonii*, *C. formosanum*, *C. quinoa*, *C. berlandieri* ssp. *nuttalliae*, and *C. ficifolium*) which are included in our pangenome analyses. Although some differences did exist, they are likely a reflection of older, more error-prone sequencing technology and assembly methodologies (Table 1) used to generate the assemblies in addition to biological differences of the genomes themselves. For instance, the total assembly sizes varied across these species as expected relative to their ploidy levels and predicted genome sizes (1C values; Mandák et al., 2016), ranging from 362,509,632 bp for the A-genome diploid *C. pallidicaule* to 1,630,219,857 bp for the BCD-hexaploid *C. formosanum*. However, the *C. pallidicaule* A-genome assembly (Mangelson et al., 2019), which is based primarily on Illumina short reads, was substantially smaller than the Oxford Nanopore Technology (ONT)-based genome assembly of *C. watsonii* (547,762,155 bp), another A-genome diploid. This difference is likely due to the inability of short reads to reliably span repetitive regions of the genome leading to a significantly reduced assembly size. The *C. pallidicaule* assembly also exhibited a substantially higher number of contigs compared to the other species (4641) as well as the lowest scaffold N50 statistic (38,119,892 bp), indicating a more fragmented and less contiguous assembly. While the *C. watsonii* assembly is more contiguous, we note that it has the highest number of gaps ($n = 5{,}338$) among all assemblies—likely stemming from high error rate of the ONT reads. Perhaps not unexpectedly, the assembly of *C. ficifolium* (Ludwig et al., 2025), which used HiFi reads and the hifiasm assembler, exhibited genome statistics similar to our de novo assemblies and showed the highest contiguity among all genomes. Indeed, the primary contig assembly of *C. ficifolium* was so contiguous (all but one chromosome was telomere-to-telomere) that the authors forewent Hi-C scaffolding and used linkage maps to scaffold the assembly to chromosome scale. The *C. berlandieri* ssp. *nuttalliae* assembly was a PacBio continuous long read (CLR) assembly that had 873 gaps (1 gap per 1.5 Mb), which is better than the short-read and ONT assemblies, but still far more than the PacBio HiFi-based assemblies that averaged 17.25 gaps (1 gap per 402 Mb). Although these additional genome assemblies are generally less contiguous than our six de novo assemblies, they are similarly complete with regard to gene coding regions. For example, across the six de novo assemblies between 97.2% and 98% of all conserved orthologous genes (COGs) were found as complete copies, which is similar to the statistics (94.5%–98.2%) observed for these additional assemblies (Table 1)—indicating that the subsequent pangenome analyses should not be compromised, but enhanced, by the inclusion of these additional *Chenopodium* assemblies.

3.3 | Subgenome assignment, chromosome naming, and orientation

After Hi-C scaffolding, the de novo assembled genomes displayed large pseudo-molecules that corresponded to the expected number of haploid chromosomes for each species. Classifications of the chromosomes to specific subgenomes in the polyploid species were determined by mapping subgenome-specific reads from known reference *Chenopodium* assemblies onto each assembly with minimap2 (L. Li, 2018). Mapping reads from the D-genome diploid *C. acuminatum* against the CD-tetraploid *C. strictum* allowed for the identification of the nine chromosomes comprising the D-subgenome of *C. strictum*, which all had significantly more mapping hits than the remaining nine C-subgenome chromosomes (Figure S2). Similarly, A-subgenome specific reads from the A-genome diploid *C. watsonii* were mapped against *C. sosnowskyi* (AG-tetraploid) to identify the nine A-subgenome chromosomes with the remaining nine chromosomes being assigned to the G subgenome (Figure S3). The chromosomes of the remaining polyploids (*C. quinoa*, *C. berlandieri* ssp. *nuttalliae*, and *C. formosanum*) were previously assigned to specific subgenome types, and these designations were preserved in the current study. Using this approach clearly delineated subgenome assignments. For instance, of the 27 chromosomes identified in the hexaploid *C. opulifolium* assembly, the nine chromosomes identified as belonging to the B subgenome had, on average, 102.8 times more read mapping hits from the *C. suecicum* (B genome) reads relative to read mapping hits on the other subgenome chromosomes (Figure S4). Similarly, the nine chromosomes identified as belonging to the C subgenome had an average of 14 times more read mapping hits from C genome-specific *C. strictum* reads compared to the other 18 chromosomes in the assembly. The remaining nine chromosomes, with very low read mapping from both the B genome- and C genome-specific reads, were by default assigned to the F subgenome of *C. opulifolium*. The read mapping to specific subgenome chromosomes, and thus the subgenome assignment, can clearly be seen in the circular genome plots (Figure 1). *Chenopodium acuminatum*,

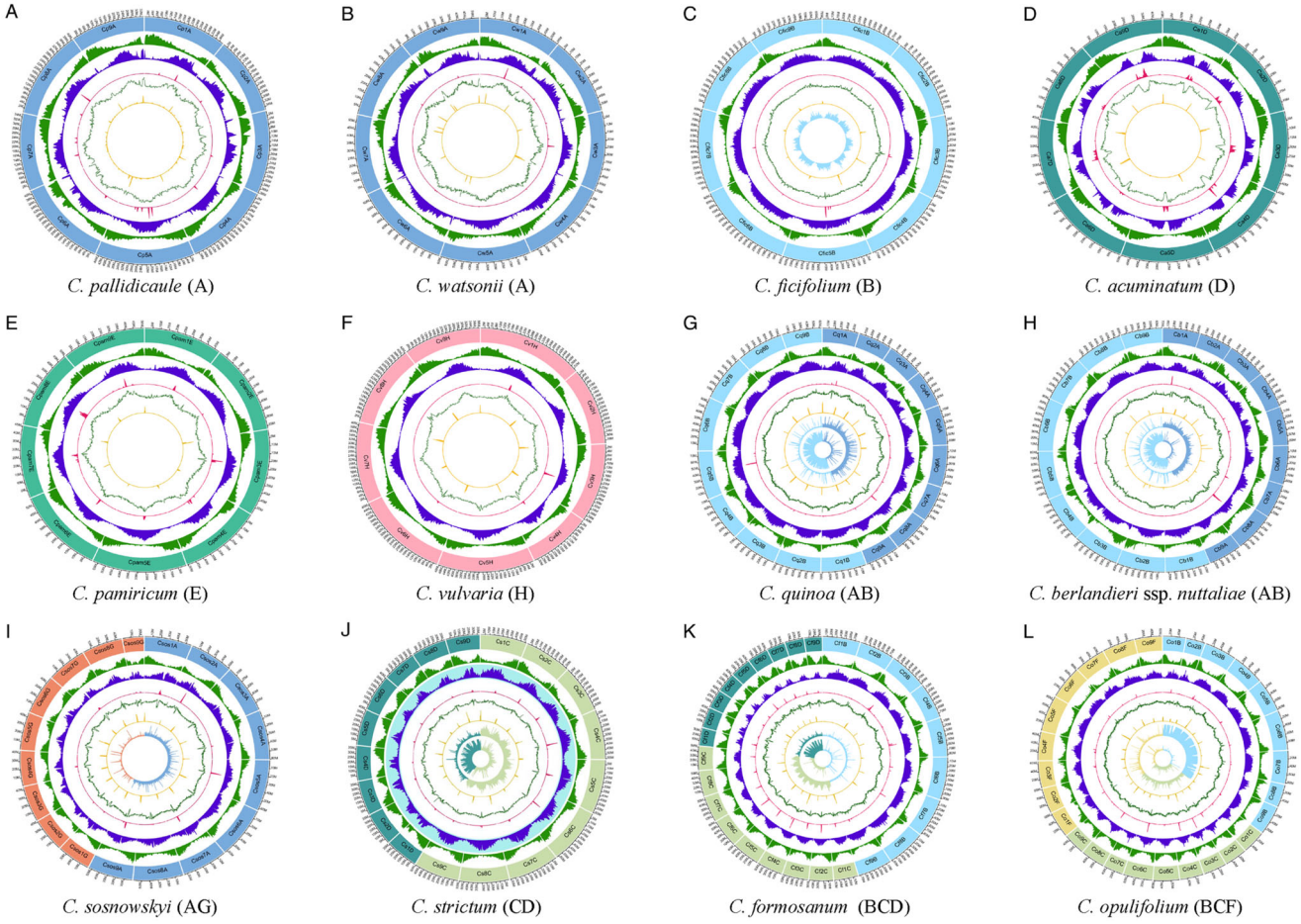

FIGURE 1 Genome features of 12 *Chenopodium* species with genome types in paratheses. Tick marks represent 5 Mb. From outside to inside, tracks represent chromosomes, gene density, LTR *Copia* and long terminal repeat retrotransposon (LTR) *Gypsy* repetitive element density, centromeric repeat sequence, guanine-cytosine (GC) content, and telomeric repeat density. Diploid species ($n = 9$) are shown in panels (A) *C. pallidicaule*, (B) *C. watsonii*, (C) *C. ficifolium*, (D) *C. acuminatum*, (E) *C. pamiricum*, and (F) *C. vulvaria*. Tetraploid species ($n = 18$) are in panels (G) *C. quinoa*, (H) *C. berlandieri* ssp. *nuttalliae*, (I) *C. sosnowskyi*, and (J) *C. strictum*. Hexaploid species ($n = 27$) are in panels (K) *C. formosanum* and (L) *C. opulifolium*. Subgenomes in the polyploids are identified by genome-specific read mapping in tracks 7 and 8: A-genome-specific reads are from *C. pallidicaule*, B-genome reads from *C. ficifolium*, C-genome reads from the C subgenome of *C. opulifolium*, and D-genome reads from *C. acuminatum*.

C. vulvaria, and *C. pamiricum* are D-, H-, and E-genome diploids, respectively, and therefore did not need subgenome classification.

Once chromosomes were assigned to specific subgenomes, we chose to standardize their naming across assemblies relative to chromosomes of *Beta vulgaris* L. vEL10.1, a distantly related but well-studied species within *Amaranthaceae* with the same haploid chromosome number ($x = 9$) (McGrath et al., 2023). Syntenic dot plots between species, including *B. vulgaris*, were used to identify and standardize the names of chromosomes. Where naming was inconclusive against *B. vulgaris* and conflicting against the chromosomes of different *Chenopodium* subgenomes, we prioritized synteny with other chromosomes of the same subgenome. Where chromosomal rearrangements confounded the identification of orthologous chromosomal relationships and to verify our chromosome name assignment, sequencing reads specific to each chromosome were mapped to other *Chenopodium* subgenomes, and the relative mapping frequency of the reads was used to verify and/or clarify the naming assignment of each chromosome in question. For example, syntenic dot plots identified three chromosomes of *C. vulvaria* with significant chromosomal rearrangements relative to *B. vulgaris*; thus, synteny alone was insufficient to name these chromosomes. Reads specific to each *C. vulvaria* chromosome were subsequently mapped against the chromosomes of *C. berlandieri* ssp. *nuttalliae* (subgenomes A and B) and those of *C. formosanum* (subgenomes C and D) to assign chromosome names with more accuracy based on read mapping rates (Figure S5). Once each chromosome was named, the orientation (relative to the *B. vulgaris* reference genome) of each chromosome was standardized to facilitate subsequent interspecies/genome

comparisons and to achieve uniformity across all assemblies in the dataset. We encourage other *Chenopodium* researchers to employ the same naming and orientation conventions, thus promoting effective inter-laboratory research.

3.4 | Genome evaluation and correction

The genome assemblies of the *Chenopodium* species were evaluated using Inspector (Chen et al., 2021). Inspector provides not only basic genome statistics but also an overall QV for each genome based on the number of structural and small-scale errors detected scaled by the total base pairs of the assemblies. The assemblies for *C. watsonii*, *C. quinoa*, and *C. berlandieri* ssp. *nuttaliae* were not error corrected with Inspector, as they were error corrected in their original publications, but were evaluated for overall genome quality and reported in Table S4. Among the Inspector-corrected genomes, *C. berlandieri* ssp. *nuttalliae* displayed the lowest read mapping rate (86.7%), while *C. pallidicuale* showed the highest split-read rate (24.0%), or proportion of aligned reads with split alignments, while the other 12 genomes had significantly higher mapping rates, ranging from 96.3% (*C. quinoa*) to 100% (*C. vulvaria* and *C. ficifolium*), and much lower split read rates ranging from 0.47% (*C. ficifolium*) to 10.97% (*C. quinoa*). Inspector also evaluated the assemblies for structural errors including expansions, collapses, haplotype switches, and inversions. Notably, *C. watsonii* exhibited a significantly higher structural error count (6414), primarily attributed to collapses (6325) and expansions (54), while the next highest structural error count among the species was only 409 in *C. pallidicaule*. All other genomes averaged only 23.2 structural errors. The lower quality of the *C. watsonii* and *C. pallidicaule* assemblies was not unexpected, being that they are unique among the set of genomes included in the study. The *C. pallidicaule* assembly was primarily based on jumping libraries generated from Illumina short reads and thus is expected to suffer from issues associated with short-read assembly technologies, including genome fragmentation (numerous N-gaps), repeat miss-assemblies (expansion/collapse), and the inability to correctly resolve complex structural variations such as large duplications and satellite DNA. Similarly, the *C. watsonii* assembly was developed from early generation ONT flow cells (R9), which also suffered from high sequencing error rates. Notably, no structural errors were detected in the assembly of *C. ficifolium*. Not unexpectedly, small-scale assembly error rates (base substitution or collapse/expansion < 50 bp) mirrored the structural error rates, with the *C. watsonii* (547) and *C. pallidicaule* (43) showing the highest rate of small-scale assembly errors per megabase and *C. ficifolium* (0.01) and *C. formosanum* (0.02) showing the lowest rate. Not unexpectedly, the genome QV score was the lowest for the *C. watsonii* and *C. pallidicaule* assemblies (25.07 and 27.82, respectively), while the other genomes were quite high, ranging from 39.78 (*C. quinoa*) to 76.9 (*C. ficifolium*). For comparison purposes, the QV score of the HiFi/hifiasm assembly of the human genome reference HG002, from the genome in a Bottle project conducted by the National Institute of Standards and Technology, was scored at 53.6 (Chen et al., 2021).

BlobTools (Laetsch & Blaxter, 2017) was then used to determine if any unincorporated contigs were off-species contaminants. On average, 3.2 contigs were removed from each of the final genome assemblies. These contigs were generally small, averaging only 44,460.7 bp and ranged in size from 2200 to 268,814 bp. The largest number of contaminants was identified in the *C. quinoa* assembly ($n = 13$), while zero contaminants were identified in four of the assemblies (*C. ficifolium*, *C. acuminatum*, *C. vulvaria*, and *C. strictum*). The very low proportion of the sequence contamination in the final genome assemblies highlights the importance of growing the target plants in disease-free, hydroponic growth chambers prior to HMW DNA extraction.

3.5 | cps

We constructed the cp genomes de novo for each species and then oriented them against the previously published *C. quinoa* cp (MK159176; Maughan et al., 2019). The quadripartite structure of the cps consisted of a large single-copy (LSC) region, a small single copy (SSC) region and a pair of inverted repeat (IR) regions. The cp assemblies were annotated and visualized with GeSeq (Tillich et al., 2017; Figure 2). Despite differences in ploidy levels and genome sizes among species, the cp assemblies were all similar in size, ranging from 151,282 to 152,309 bp for *C. pallidicaule* and *C. vulvaria*, respectively, and overall gene content (128 genes/31 tRNAs; Table 1) like that reported for other *Amaranthaceae* (She et al., 2022; Xu et al., 2022). In many of the species, we found heteroplasmy for an inversion of the SSC as was originally reported by Palmer (1983) who showed that the plastome often exists in two inversion states that differ in the relative orientation of the SSC region. The inversion is believed to occur either through intramolecular recombination between the IRs or perhaps through recombination-dependent DNA replication of the cp genome in its linear conformation (Oldenburg & Bendich, 2004). In three of the species studied (*C. strictum*, *C. pamiricum*, and *C. acuminatum*), only the inverted version relative to the reference *C. quinoa* cp, of the plastome, was detected (Figure 2). To investigate the maternal inheritance of the cp in the polyploids, a plastome phylogeny, with *Dysphania ambrosioides* (CoGe ID #68555) as the outgroup, was constructed using the LSC (80 Kb) to avoid the heteroplasmic SSC inversion (Figure 2). Confirming previous reports (Mandák et al., 2018; Maughan et al., 2019),

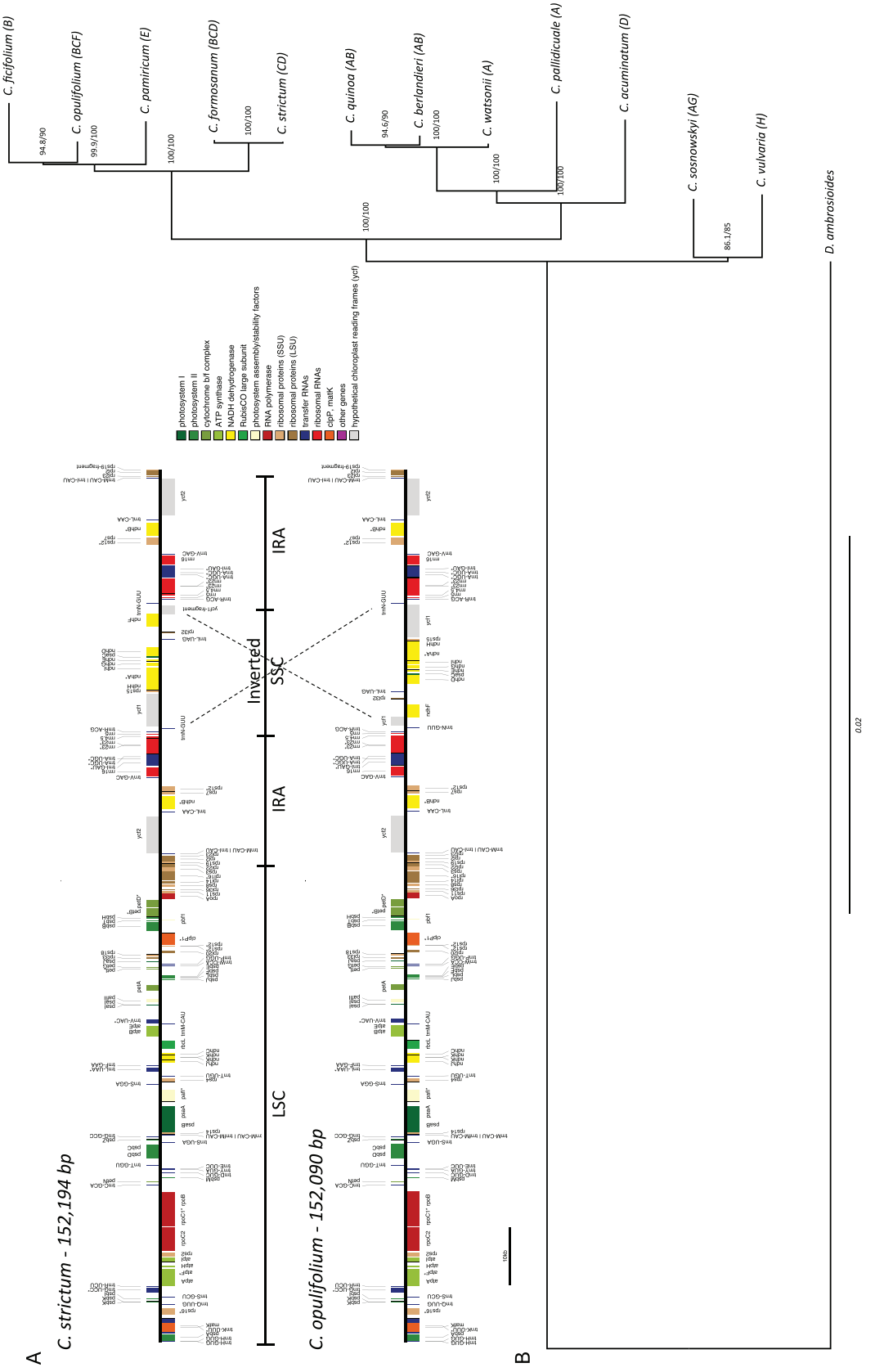

FIGURE 2 The structure of the *C. strictum* and *C. opulifolium* chloroplast genome. (A) Genes on the top of the annotation are transcribed in the forward direction, while genes on the bottom are transcribed in the reverse direction. The locations of the inverted repeats (IRA and IRB) separate the chloroplast genomes into large single-copy (LSC) and small single-copy (SSC) regions. The gene functional groups are color-coded. The two inversion states of the SSC region are shown with dotted lines. (B) Phylogenetic relationships among the *Chenopodium* species based on SNP variants in the LSC region of the chloroplast. *Dysphania ambrosioides* is included as an outgroup species. Scale bar, 0.02 substitutions per site.

the A-genome diploids (*C. watsonii and C. pallidicaule*) are within the same clade as the AB tetraploids (*C. quinoa* and *C. berlandieri* ssp. *nuttalliae*), whereas the B-genome diploid (*C. ficifolium*) is in a distal clade—suggesting that the A genome parent was the putative maternal parent in the original AB polyploidization event. While no extant C diploid species are known, the observation that the D diploid (*C. acuminatum*) is distal to the CD tetraploid (*C. strictum*) hints that the C diploid was the likely maternal parent in the initial CD polyploidization event. Furthermore, the close placement of the BCD hexaploid *C. formosanum* with *C. strictum* suggests that they share a common plastome ancestor—potentially beginning with the hybridization of C (maternal parent) and D diploids to form a CD tetraploid, followed by the CD tetraploid (maternal) hybridizing with a B diploid to form the BCD hexaploid. We note that the B plastome from *C. ficifolium* is in a separate subclade relative to *C. strictum* and *C. formosanum*. Similarly, while no known G diploids exist for comparison, the placement of the AG tetraploid (*C. sosnowskyi*) distally to the A diploid (*C. watsonii* and *C. pallidicaule*) implies that the G genome parent was the likely maternal parent in the hybridization event that formed the initial AG tetraploid. The apparent sequence of events leading to the BCF hexaploid *C. opulifolium* is less clear, since neither C nor F diploids are available for comparison—highlighting the importance for additional germplasm collections within the genus.

3.6 | De novo annotation

Braker3 (Gabriel et al., 2024) identified an average of 21,525, 45,766, and 67,969 gene models in the chromosomes for the diploid, tetraploid, and hexaploid species, respectively (Table 1, Table S5). For comparison, Rey et al. (2024) reported a similar gene number for eight diverse *C. quinoa* (tetraploids) genomes, ranging from 43,733 – 48,564. The number of mRNAs (splice variants) was 24,229, 51,167, 75,535, or about 1.2 splice variants per gene model for all ploidy levels. The average gene length across all species was 3,848 bp, while the longest coding sequence was 16,405, 16,434, and 16,497 bp for the di-, tetra-, and hexaploids, respectively. The total gene space (total gene length/total chromosome length) averaged 18.9%, 16.9%, and 15.1% for the di-, tetra-, and hexaploid species, respectively. This decrease in total gene space was also reflected in the total coding sequence (CDS) space, which decreased from 6.8% for the diploids, 6.1% for the tetraploids, to 5.5% for the hexaploids. This loss of gene space was not unexpected. While polyploidization initially increases the number of genes in the new organism, over evolutionary time, some of these duplicated genes (orthologs) undergo neo- or subfunctionalization, gaining new functions. Meanwhile, many others are lost or silenced due to the lack of selective pressure to maintain multiple copies, often leading to a decrease in gene number from the initial count at the chromosome doubling event. In hexaploid wheat (*Triticum aestivum*), Zhang et al. (2020) showed that polyploidization and the subsequent genomic rearrangements that followed due to non-homologous exchange were the major contributors to gene loss in modern wheat. Likewise, many authors have shown that polyploidization activates TEs, which are not considered plant genes, and leads to an increase in the abundance of TEs and other repetitive elements, contributing to genome expansion, complexity, and restructuring (Madlung & Wendel, 2013).

Annotation completeness was assessed using BUSCO on the predicted gene models with the COG set from Embryophyta (Figure S6). The annotation all showed high BUSCO scores, with an average of 96.6% of the COGs being found as complete copies across all species. The annotations showed complete COG identification ranging from 95.5% to 97.5%. There was only a slight difference in the percentage of complete COGs identified between di-, tetra-, and hexaploid assemblies, specifically 96.2%, 96.9%, and 97.3%, respectively. The small increase across the ploidy levels was not unexpected as the probability of annotating a COG should increase as more genome copies are present in the assembly. Similarly, duplication of the COGs, as detected by BUSCO, is also expected to increase by ploidy level. Indeed, the average percent COG duplication for the diploids was only 8.15%, whereas this percentage increased to 90.68% and 94.95% for tetra- and hexaploid species, respectively.

3.7 | Orthogroup analysis

Using OrthoFinder2 (Emms & Kelly, 2019), we identified a total of 33,457 orthologous gene families across the 12 *Chenopodium* species. The average pangenome size derived from any two *Chenopodium* species constitutes only 66.7% of the total pangenome size when all 12 species are considered together, confirming that multiple species are needed to capture the full genetic diversity present within *Chenopodium* (Figure 3A). As expected, the total number of shared (pan) gene families increases as genomes (species) are sequentially added to the analysis, reaching a plateau when $n = 11$, while the number of core genes decreases. The plateau observed in shared genes suggests that the species we included are representative of the genus and sufficient for pangenome analyses, at least in respect to the Eurasian and Western hemispheric species represented in this study. We found that 64.8% of gene families were found to be in the shell, which lies between the number of shell gene families identified in the *Glycine soja*, the wild progenitor of soybean (51.4%; Y. Li et al., 2014), and that reported for 503 maize samples (82.7%; Hirsch et al., 2014) supporting the prediction that an outbred species should have a larger shell genome than highly autogamous

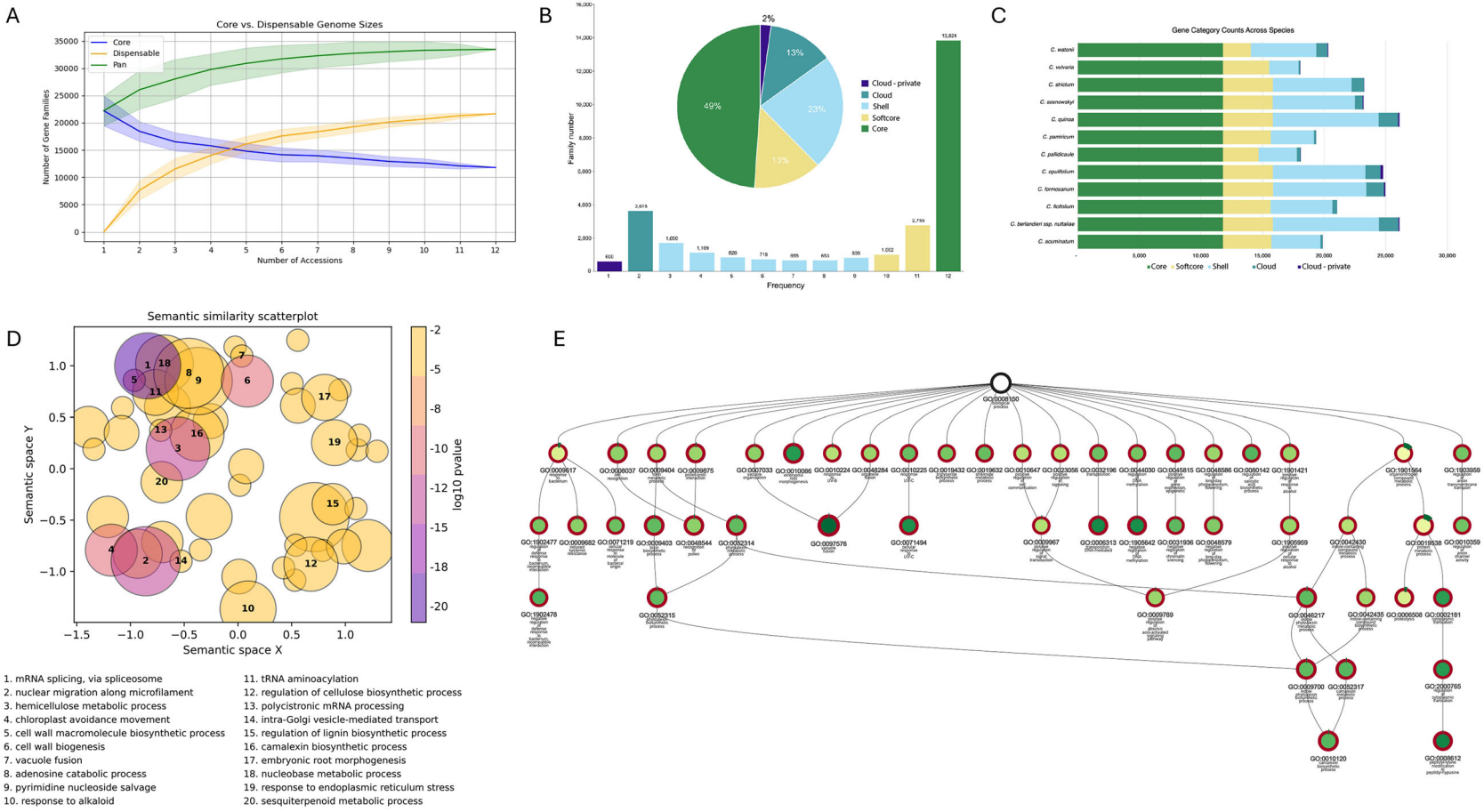

FIGURE 3 Pangenome analysis in *Chenopodium*. (A) The number of pan, core, and dispensable gene families with increasing number of accessions. The dispensable category consists of the softcore, shell, and cloud gene families. Each number was randomly sampled 100 times. (B) The composition of the *Chenopodium* pangenome in private, cloud, shell, softcore, and core gene families, and the frequency of core gene family numbers. (C) The number of pangenome gene family categories in each *Chenopodium* species. (D) Semantic similarity of enriched gene ontology (GO) terms for biological process across all species scaled by *p*-value and number of GO terms. (E) Hierarchical GO enrichment for camalexin biosynthesis for *C. formosanum* (Trapid 2.0; Bucchini et al., 2021). Each node represents an enriched GO term. Node size and color are scaled by the *q*-value and by the enrichment fold of the GO term, respectively. Node outer band is determined by the percentage of the subset's transcripts that are annotated with the enriched GO term.

species like soybean. Many chenopods are gynomonoecious showing facultative autogamous reproduction and thus are expected to have shell gene numbers intermediate to outcrossing and inbreeding species. In *C. quinoa*, outcrossing rates are reported to vary across accessions, ranging between 0.5% and 17.4% (Gandarillas et al., 1979; Silvestri & Gil, 2000), although heat-induced male sterility is well known in the genus and has been successfully employed in making hybrids (Maughan et al., 2024). Of the orthologous gene families identified, 35.3% (11,800) were detected in all species, 16.3% (5447) in 10–11 species, 32.0% (12,075) in 3–10 species, 16.4% (5503) in 1–2 species and were defined as core, softcore, shell, and cloud, respectively. Within the cloud, a total of 744 (2.2%) gene families were identified as private (Figure 3B). Unsurprisingly, polyploid species contained substantially more shell and cloud orthologous gene families compared to the diploid *Chenopodium* species (Figure 3C). These findings are similar to other pangenome results from wild soybean (*Glycine soja*; Y. Li et al., 2014), maize (*Zea mays*; B. Wang, Hou, et al., 2023), and apple (*Malus*; T. Wang et al., 2023), where the percentage of core genes declined asymptotically with the inclusion of each additional genome, indicating that each species (genome) possesses only a subset of the total genes and that each possessed a substantial number of private gene families (Figure 3C).

A functional enrichment analysis showed that the private gene families were enriched for a wide variety of biological processes (Figure 3D), including many processes involved with environmental adaptability. For example, in *C. formosanum*, camalexin biosynthesis (GO:0010120), a complex metabolic pathway that produces a potent antimicrobial compound crucial for plant defense against pathogens, was enriched (q-value 5.6e-4) (Koprivova et al., 2019) (Figure 3E). In *C. opulifolium* cp avoidance movement (GO:0009903), an adaptation that helps plants thrive in varying light conditions was enriched (2.54e-11). In *C. acuminatum*, regulation of defense response to oomycetes (GO:1902288) was enriched (3.79e-3), while in *C. quinoa*, the TOR signaling (GO:0031929) pathway was enriched (1.48e-2). TOR signaling allows plants to dynamically adjust their growth and metabolism in response to changing environmental conditions, such as light, temperature, and water availability (Shi et al., 2018). In *C. pallidicaule*, the glucuronoxylan biosynthetic process (GO:0010417), an important component of the cell wall, was enriched (5.97e-13) and has been implicated in modulating aluminum sensitivity (Zhu et al., 2017). Private gene families provide unique insight into potential gene targets for important niche adaptive alleles which can be potentially exploited through hybridization and introgression. This is particularly important as the global demand for quinoa has increased so dramatically that many nations are now attempting to produce quinoa outside of its traditional production range. Unfortunately, because quinoa is narrowly adapted to the high plains of the Andean plateau, where disease/pest stress is low and summer temperatures rarely exceed 20°C, production outside of the Andes has proven difficult. Maughan et al. (2024) recently showed that quinoa can readily cross with other tetraploid AB species, many of which exhibit unique and beneficial biotic (disease) and abiotic (heat and salinity tolerance) adaptive characteristics and thus represent important secondary gene pools for future improvement of quinoa. In addition, they identified a significant genetic bottleneck separating wild Andean and cultivated quinoas from wild lowland accessions of *C. hircinum* and, more generally, the highly diverse allotetraploid goosefoot complex (ATGC) in North America.

3.8 | Orthogroup inferred phylogeny

Mandák et al. (2018) used sequence information from the third intron of the FLOWERING LOCUS T LIKE paralogs (*FTL1* and *FTL2*), the internal transcribed spacer (ITS) of the nuclear ribosomal DNA and the intergenic spacer regions of three cp genes to study *Chenopodium* genome evolution and speciation. They concluded that there are eight major monophyletic evolutionary lineages (A–H) in *Chenopodium*; however, their analysis revealed several incongruencies, specifically with the number of clades identified between trees of the cpDNA/*FTL* ($n = 8$) and the ITS ($n = 5$) as well as the inconsistent placement of the F genome, which was basal in the *FTL* tree, but formed a sister clade with the B genome in the cpDNA tree. An orthogroup-inferred phylogeny ($n = 20{,}662$) allowed us to further investigate these evolutionary relationships across the entire set of orthologous gene families. Indeed, we specifically sequenced species that represented each of the eight genome variants found in Eurasian and American chenopods to facilitate this in-depth analysis. We note that since no extant diploid species exist (or as of yet they remain undescribed) for genome types C, F and G, polyploid species with subgenomes of these genomes were included in the sequencing panel. To investigate the phylogenetic relationships among the genomes, we extracted subgenomes from the polyploids and treated each subgenome as separate taxa for the orthogroup gene tree analysis. Thus, along with the sequenced diploid genomes, we analyzed a set of 20 genomes, which included five A genomes; five B genomes; three C genomes; three D genomes; and one each of the E, F, G, and H genomes. *Dysphania ambrosioides* (CoGe ID #68555), which is a sister genus to *Chenopodium*, was included as an outgroup. Our analysis clearly confirms the presence of eight monophyletic lineages corresponding to the previous genome designations, all with a bootstrap value of 100 (Figure 4). The closely related B and C genome lineages form a clade with a sister relationship to the E genome. Within the B-genome clade, the New World species (*C. quinoa* and

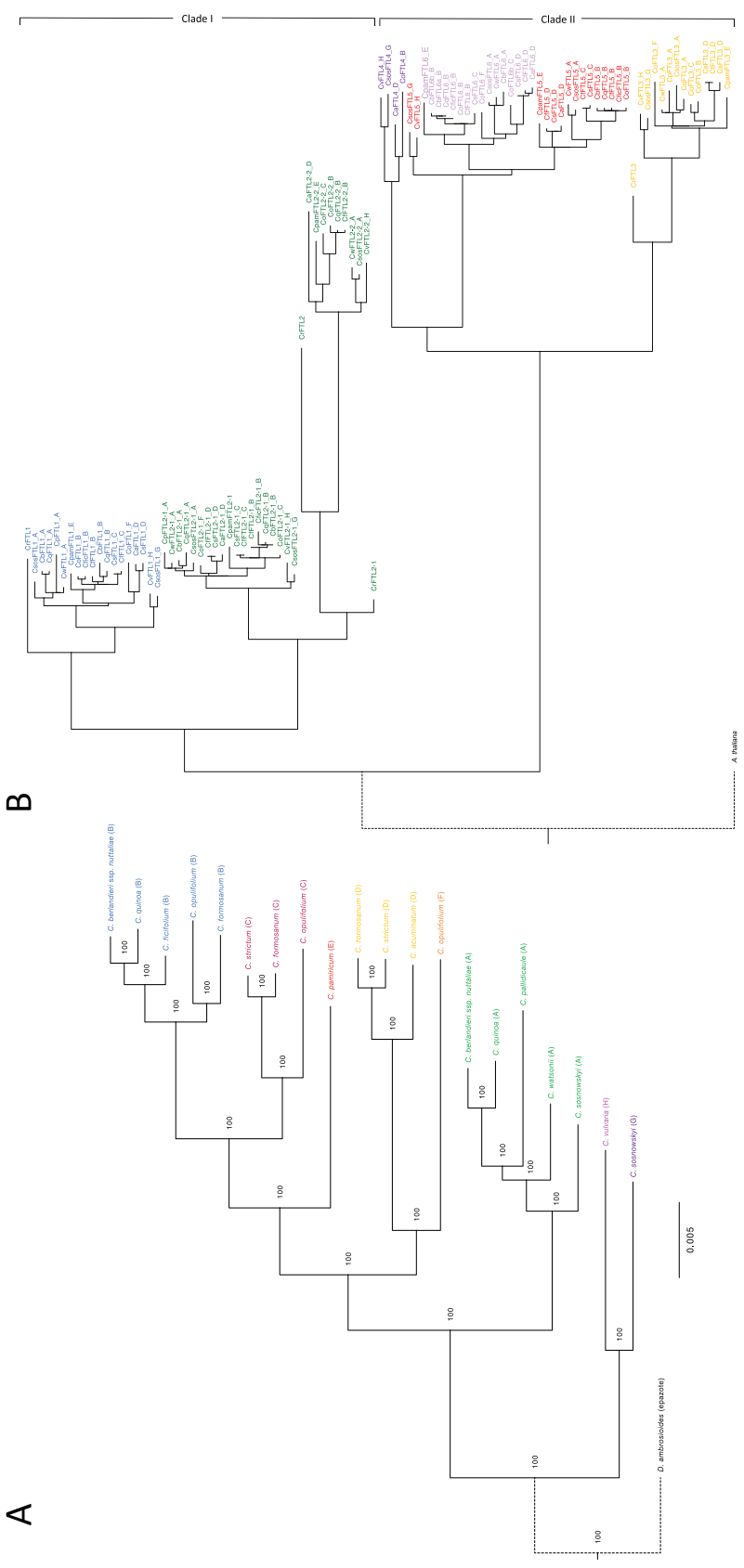

FIGURE 4 Subgenome and *FTL* phylogeny analysis. (A) Subgenome phylogenetic tree based on OrthoFinder2 using the multiple sequence alignment argument with *Dysphania ambrosioides* (epazote) as the outgroup. (B) Phylogeny of intact *FTL* genes from 12 *Chenopodium* species using the Arabidopsis *AtFT* gene as the outgroup. The *FTL* genes of *C. rubrum* (EU128013 and EF445636) were employed to root the clades of the *FTL1*, *FTL2-1*, *FTL2-2*, and *FTL3* genes. The *CrFTL2-1* gene was placed at the base of both *FTL2-1* and *FTL2-2* in this tree. Bootstrap values are shown at nodes in both trees.

C. berlandieri ssp. *nuttalliae*) were the most closely related of the group, which is not surprising as they are likely to have arisen from the same ancestral AB tetraploid species (Maughan et al., 2024). Notably, the F genome of *C. opulifolium* (BCF) is placed as the sister group to the D genome clade, indicating a previously unrecognized evolutionary relationship between the F and D genomes and resolving inconsistencies in prior analyses. We also note that both the C and D genomes of the hexaploid *C. formosanum* (BCD) are closely related with the subgenomes of tetraploid *C. strictum* (CD), supporting the conclusion that the polyploidization sequence for the BCD polyploids began with a CCDD tetraploid, followed by subsequent hybridization with a B diploid (Mandák et al., 2018). All five New World A genomes form a clade, again with *C. quinoa* and *C. berlandieri* ssp. *nuttalliae* being closely related. The A subgenome of *C. sosnowskyi* is sister to the remaining subgenomes in the A clade, while its G genome and the H genome of *C. vulvaria* form a clade sister to all other genomes, suggesting that they may represent the descendants of some of the earliest divergent lineages within the genus. To confirm the phylogeny, we also produced a maximum likelihood estimation phylogeny (Figure S7) based on 1596 single-copy COGs identified by Compleasm (Huang & Li, 2023) in the *Chenopodium* genomes with the sister genus *B. vulgaris* (McGrath et al., 2023) as an outgroup. This phylogeny showed identical placement of all taxa as observed in the orthogroup tree, thereby supporting the accuracy of our phylogenetic reconstruction.

3.9 | Genome features

Analysis of gene density, based on gene models identified by Braker3, revealed a consistent pattern across all species (Figure 1). Gene density increased toward the distal ends of all chromosomes and decreased toward the centromeric regions. In contrast, the density of *Gypsy* and *Copia* TEs, which represents on average 44% of the TEs identified across the species, exhibited peaks around the centromeric region, declining toward the telomeres with one particular exception. In the analysis of *C. acuminatum* (D-genome diploid), the TE densities peaked in pericentromeric regions and declined sharply through the putative centromere across all chromosomes (Figure 1D). This decrease was mirrored by a decrease in GC content, which likely reflects the high AT content (72%) of the 39-bp putative centromeric satellite subrepeat identified using TRASH (Wlodzimierz et al., 2023). Thus, we speculate that the centromeric region in the D genome of the *C. acuminatum* assembly is longer and more intact than the assemblies of other genomes. Indeed, we see similar characteristics, although more subtly, in the D subgenomes of *C. strictum* (CD tetraploid) and *C. formosanum* (BCD hexaploid). Alternatively, it could be that a novel non-*Gypsy/Copia* TE element is uniquely occupying the D-genome centromeres or, although unlikely, the centromeres in the other genomes have collapsed in assemblies, obscuring this characteristic.

A 294-bp (12-13p; HM641822.1) centromeric repeat sequence derived from *C. quinoa* was used to identify the location of the centromeric regions across all the *Chenopodium* species (Kolano et al., 2011; Orzechowska et al., 2018). We note that the 39-bp subrepeat identified by TRASH is repeated multiple times within the 12–13p sequence. Centromeric regions were clearly identifiable on the majority of chromosomes across all species, with most being located in either metacentric or submetacentric positions. As noted previously, the centromere is a complicated repeat region, subject to collapse during the assembly process; thus, it is not surprising that the centromeric regions were not identified for all chromosomes, especially for those assemblies that employed short reads or first-generation/error-prone long reads. Mean GC content (~35%) across all chromosomes and species showed a consistent trend with a slight increase from the telomeres toward gene-rich pericentromeric regions, with sharp decreases at centromeric regions themselves (Figure 1). Increased GC content is believed to provide structural stability to DNA and is often associated with gene regulatory elements as well as contributing to the formation of open chromatin structures (Romiguier & Roux, 2017), while the AT-rich composition of the centromere is implicated in assembly of the kinetochore, which is essential for proper chromosome segregation during cell division (Kursel & Malik, 2016).

Telomeric repeats were observed on at least one end of 172 of the 180 chromosomes assembled across all 12 species, and on both ends of 117 of the 180 total chromosomes. Of the 117 chromosomes with telomeric repeats observed on both ends, the number of Ngaps (>100 Ns) ranged from a low of 0 to a high of 457 (*C. palidicaule*), with 71 chromosomes classified as telomere-to-telomere (gapless). As expected, the non-HiFi-based assemblies showed the highest number of Ngaps per chromosome. The telomere, which consists of a conserved repeat of seven bases, is notoriously difficult to fully resolve due to the lack of complexity of the repeat, resulting in telomeres often being left unresolved (un/misassembled or severely collapsed). Such is the case for several chromosomes of *C. watsonii* and *C. pallidicaule*, both A-genome diploids, as well as for two chromosomes in *C. berlandieri* ssp. *nuttalliae*, where telomeric signals are found internal to the chromosome, indicating potential mis-assemblies, structural rearrangements including telomeric inversions (potentially unique to the A subgenome), or the presence of telomere-like satellite DNA sequences that are being erroneously captured by our BLAST analyses using the telomeric repeat sequence. We note that all three of these species were sequenced with short-read or first-generation/error-prone long-read (PacBio CLR, ONT) technologies.

GENESPACE (Lovell et al., 2022) was used to visualize syntenic relationships across genomes based on syntenic blocks identified by OrthoFinder2 (Figure 5). Given that these are interspecific comparisons, it is not surprising that large genomic rearrangements including inversions (e.g., Cq5A:Cb5A; Figure 5A) and translocations (Co9B:C7B; Figure 5B) are present within comparisons of the same genome types (Figure 5A–D), nor is it surprising that syntenic relationships are more complicated across genome types (Figure 5E). As anticipated based on their phylogenetic position as descendants of some of the earliest diverging lineages within the genus, the G and H genomes were differentiated by large and complex rearrangements, whereas the B and C genomes, which share a more recent ancestor, are more syntenically conserved. Rey et al. (2023) recently showed that a large inversion on chromosome 3B in *C. quinoa* differentiated coastal ecotypes from altiplano ecotypes. Interestingly, chromosome 3B is the most highly rearranged of all chromosomes in the within-genomes comparisons (Figure 5B), suggesting that it is potentially more susceptible to rearrangement. TEs are believed to generate genomic instability through non-allelic homologous recombination (Raskina et al., 2008) and, indeed, the B-genome chromosomes have increased dramatically in size due to a dramatic accumulation of *Gypsy* LTRs (see below) which might explain this increased level of genomic rearrangement—although it is interesting that some chromosomes (e.g., 1B, 3B, 4B; Figure 5B) appear to show more rearrangements than others, suggesting that TE accumulation is not likely the sole factor leading to chromosomal dysgenesis. In contrast, the small D genome, which did not experience a similar expansion of LTRs, maintains a high level of syntenic continuity (Figure 5D). Structural rearrangements can disrupt meiotic chromosome pairing and hinder recombination, complicating the transfer of targeted traits. Thus, an understanding of the extent and nature of structural rearrangements between related *Chenopodium* taxa, as reported here, is essential for developing successful pre-breeding strategies.

3.10 | Transposable element and satellite DNA composition

The TE repeat composition of all 12 *Chenopodium* species and their individual subgenomes were identified and annotated using RepeatModeler (Figure 6A; Figure S8). On average, ~71% of the DNA of each species was classified as repetitive, with the most common repeat element identified being classified as LTRs (~32%). LTRs are the most abundant repeat element found in flowering plants (Du et al., 2010; Galindo-González et al., 2017), and their abundance is strongly correlated with genome size (Tenaillon et al., 2010). To investigate the effect of TE elements content across all subgenomes, we analyzed each genome separately, including each subgenome from the polyploids (Figure 6A; Figure S8). LTR *Gypsy* elements showed the largest range in abundance across species, from 12.6% to 35.5% (mean = 23.4%), with LTR *Copia* elements ranging from 4.8% to 15.2% (mean = 8.1%). A strong correlation ($R^2 = 0.91$) was seen between subgenome size and LTR *Gypsy* content, while only weak correlations were identified with LTR *Copia* ($R^2 = 0.14$) and DNA transposons ($R^2 = 0.29$), the second and third most common repeats identified, respectively (Figure 6B). Jarvis et al. (2022) previously implicated *Gypsy* element expansion in the substantial size differences in subgenomes in *C. formosanum* (BCD hexaploid), with the B subgenome larger than both the C and D subgenomes. Indeed, the five B genomes included in our study were, on average, 30% larger than the other genomes, with an average genome size of 699 Mb and exhibited the highest average *Gypsy* content (34%; Figure S8).

Not only has the B subgenome of *C. formosanum* increased in size due to *Gypsy* LTR expansion, but all B genomes, including those found in C. *opulifolium*, *C. ficifolium*, *C. quinoa*, and *C. berlandieri* ssp. *nuttalliae*, have experienced a *Gypsy* expansion leading to an increase in genome size. This observation suggests that the expansion of *Gypsy* likely occurred in an ancestral B genome prior to the polyploidization event that led to the tetraploids (AB) and hexaploids (BCF and BCD) carrying the B subgenome. The repeat landscape plots show that the expansion of the *Gypsy* element in the B subgenome is centered at $K = 6$ (i.e., the Kimura value which roughly corresponds to 6% sequence divergence), dating the expansion to ~2.31 million years ago (MYA) (Figure 6D). Interestingly, we also see an increase in *Gypsy* LTR content in the F subgenome of *C. opulifolium*. Similar to the B genome, the F genome is large (653 Mb), but curiously the increase in size is not solely due to an expansion of *Gypsy* elements (which also peaks at ~2.31 MYA) but is also due to an expansion of *Copia* elements that have increased in content from a genus average 8.1%–13.2% at nearly the same time (Figure 6D; Figure S8). The intermediately sized A genomes (mean = 530 Mb; excluding *C. pallidicaule* short-read assembly) have the largest *Copia* content, averaging 13.7%, substantially more than the genus average (8.1%), while the *Gypsy* content of the A genomes is near the average for all species, suggesting that the size in the A genome has been primarily influenced by *Copia* LTR expansion. The repeat landscape plot for the A genomes shows the peak of this expansion to have occurred at approximately $K = 9$ or between 3.1 and 3.9 MYA (Figure 6D). Lastly, the D genomes, averaging only 371 Mb, are the smallest of the genomes. Notably, and perhaps not unexpectedly, these genomes show the smallest content of both *Copia* (5.5%) and *Gypsy* (13.9%) elements.

We note that there is a large fraction of unclassified repeats for all genomes (Figure 6A; Figure S8), which is not

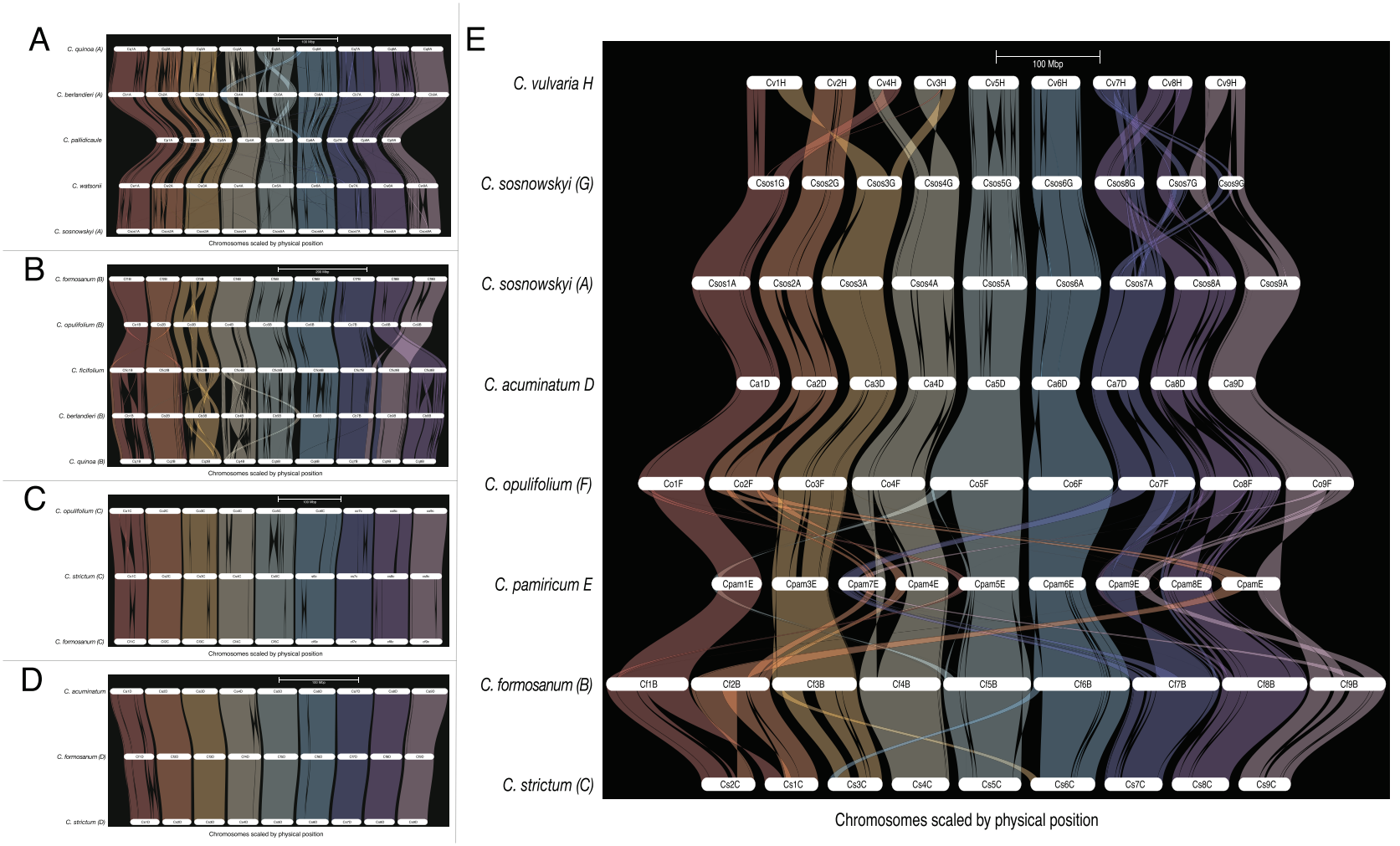

FIGURE 5 Syntenic relationships of orthologous regions among *Chenopodium* genomes. Chromosomes are ordered horizontally (1–9), and ribbons are color-coded to show syntenic relationships among chromosomes. Genomes are ordered vertically by general phylogenetic positions for all A genomes (A), B genomes (B), C genomes (C), and D genomes (D) included in the study. To facilitate the visualization of syntenic relations across genome types (A–H), a riparian synteny plot was generated for a single representative from each genome type (E).

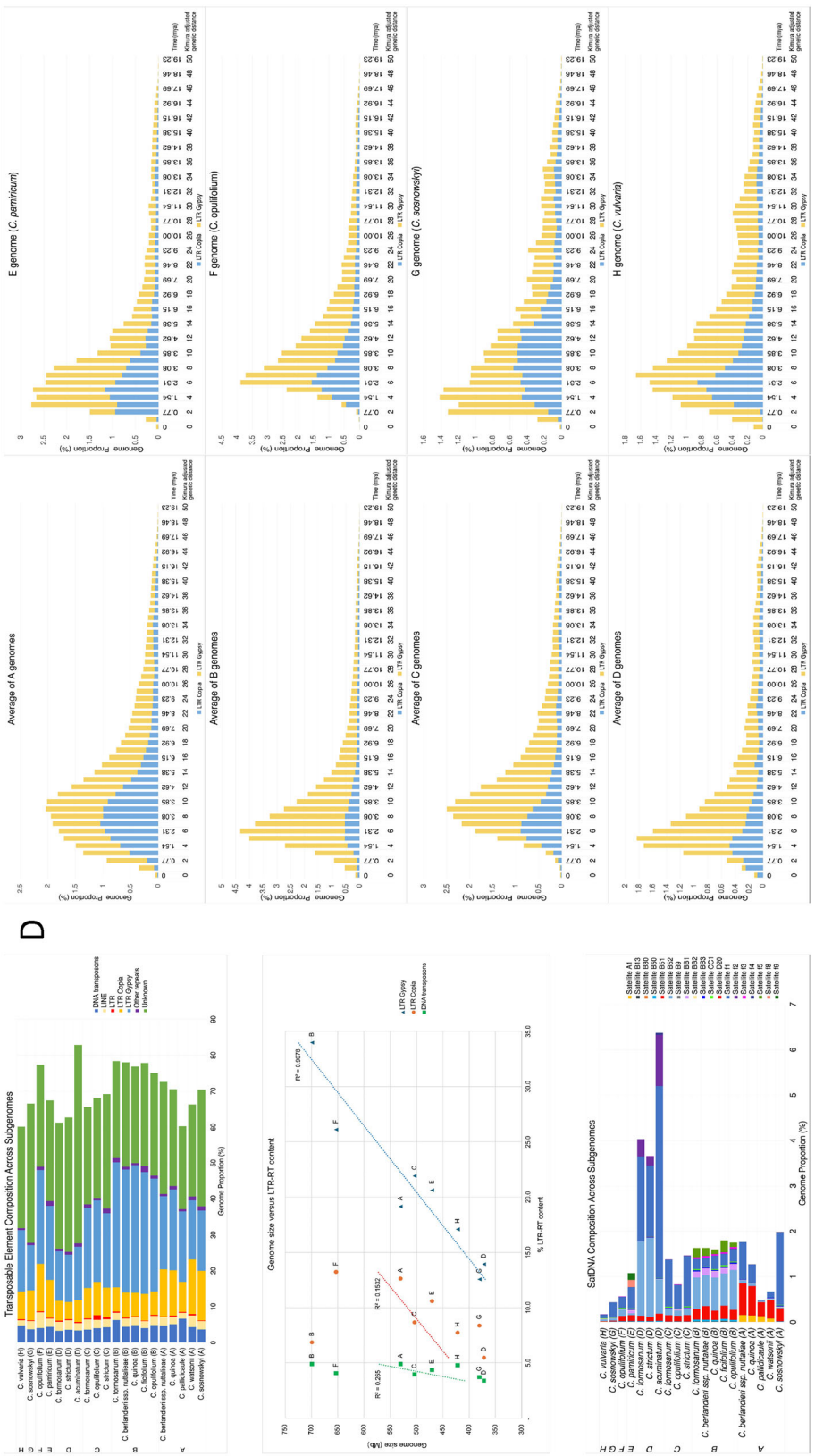

FIGURE 6 Repeat composition. (A) Percent transposable elements (TEs) composition across the *Chenopodium* genomes. (B) Correlation between average genome size versus TE content (%). The *C. pallidicaule* genome was not included in the analysis due to its truncated genome size resulting from its short-read-based assembly. (C) Percent satellite DNA (satDNA) family composition across the *Chenopodium* genomes. (D) TE landscapes of *Copia* and *Gypsy* long terminal repeat retrotransposon (LTR) elements across *Chenopodium* genomes. The average composition is plotted when multiple copies of the same genome type were available. The *y*-axis shows TE abundance as a percentage of the genome. The *x*-axis shows both sequence divergence (CpG-adjusted Kimura distance) relative to consensus sequences for TE superfamilies and an estimate of evolutionary time calculated using LTR sequence evolution rate estimates in rice.

surprising given that there have been relatively few studies of repetitive elements in *Chenopodium* and those have been primarily focused on rDNA, peri-centric repetitive, subgenome-specific repeat, and satellite DNA sequences (Belyayev, Jandová, et al., 2020; Heitkam et al., 2020; Maughan et al., 2006; Orzechowska et al., 2018). These unclassified TE elements have a weak negative correlation with genome size ($R^2 = 0.30$), suggesting that, as a group, they are not significantly involved in genome size expansion across all genomes. However, it is notable that *C. acuminatum* (D diploid) has a substantially increased percentage (55.3%) of unclassified elements indicating the presence of a unique TE expansion that could not be classified by available repeat libraries, but which partly accounts for its larger than usual genome (407 Mb) relative to the D subgenomes of the polyploids *C. strictum* (343 Mb) and *C. formosanum* (364 Mb) (Figure S9). Repeat landscape plots of the D subgenome also suggest that *C. acuminatum* has experienced a very recent expansion of LTR *Copia* elements within the last million years ($K = 0$) (Figure S10). This recent expansion comprises over 2% of the genome of *C. acuminatum* and is unique relative to the other *Chenopodium* D subgenomes as well as all other genomes included in this study. These repeat analyses, especially for the *Gypsy* and *Copia* LTR elements, indicate that TEs have and are continuing to play a crucial role in shaping the evolutionary histories of the various *Chenopodium* genomes.

satDNA differs from TEs in that they typically consist of tandem repeats. In plants, they are static and serve mainly in structural roles and are often localized to specific regions of the genome such as at the centromere (Garrido-Ramos, 2017). Our satDNA analysis revealed the presence of 19 satDNA families (Belyayev, Jandová, et al., 2020; Jarvis et al., 2022). Remarkably, at a broad scale, minimal differences in prevalence and composition were observed within genome types. However, there are unique differences among genomes. For instance, there is an increase in the proportion of satDNA, specifically the f1, f2, and B52 families, in all D genomes studied (*C. acuminatum*, *C. formosanum*, and *C. strictum*) relative to the other genomes (Figure 6C). This increase is not correlated with increased genome size, as the D genomes are among the smallest *Chenopodium* genomes analyzed. Other specific satDNA families are also strongly associated with specific genome lineages (Figure 6C). For example, the f5 family is prevalent in the B genome, f2 family in the D genome, and f8 and f9 families in the E genome (Figure S11). Notable is a satDNA family tentatively named "A1" that was particularly abundant in the North American A genomes (*C. quinoa*, *C. berlandieri* ssp. *nuttalliae*, and *C. pallidicaule*), but nearly undetectable in the Eurasian genomes including the A genome of *C. sosnowskyi*. An analysis of whole-genome short-read sequencing data from *C. bryoniifolium*, a Siberian A-genome diploid species ($2n = 2x = 18$), also did not reveal its presence, while an analysis of short-read data for other North American A-genome species did confirm it as one of the major components of the satellitome (K. Krak, personal communication, 2024), suggesting that the "A1" satDNA family potentially amplified independently in the species from the Western Hemisphere. We note that *C. pallidicaule* exhibits a single distinct peak of sequence divergence for this satDNA, whereas other American A genomes display two peaks, indicating that this family is dynamically evolving in terms of both rate and timing of sequence accumulation (Figure S11). This suggests that satDNA in general, and the "A1" family in particular, may be tightly associated with the unique, radiative pattern of speciation in the Western Hemispheric A-genome clade that has resulted in at least 40 distinct taxa.

3.11 | *FTL* gene evolution

The transition from vegetative growth to reproduction is regulated by various signaling pathways that respond to environmental cues, such as temperature, day length, and plant age. At the core of the flowering pathway is the FLOWERING LOCUS T (*FTL*) gene (Kardailsky et al., 1999). Paralogs of *FTL*, created through duplications, control distinct and diverse functions in plants, including tuber development in potatoes (Navarro et al., 2011) and acting as floral inhibitors in sugar beets (Pin et al., 2010); thus, they play a crucial role in the fine-tuning of plant development and reproduction. A current understanding of the *FTL* evolution in *Chenopodium* is reviewed by Štorchová (2020); here, we extend the phylogenetic analysis of *FTL* genes across 12 *Chenopodium* species revealing two main clades. Clade I is composed of *FTL1*, *FTL2-1*, and *FTL2-2* genes (Figure 4B). The expression of clade I *FTL* genes has been observed during floral induction in species like *C. ficifolium* (Gutierrez-Larruscain et al., 2022) and *C. quinoa* (Patiranage et al., 2021). The ectopic expression the *FTL1* and *FTL2-1* genes of *C. ficifolium* accelerated flowering in Arabidopsis (Abeyawardana et al., 2023). In contrast, clade II includes *FTL3*, *FTL4*, *FTL5*, and *FTL6* genes (Figure 4B). Interestingly, the expression of clade II *FTL* genes has been not reported in transcriptomic studies of floral induction (Gutierrez-Larruscain et al., 2022; Patiranage et al., 2021) and are therefore not likely to contribute to the regulation of flowering. However, Drabešová (2016) reported the expression of *CrFTL3* in seeds and germinating seedlings of *Chenopodium rubrum*, suggesting a potential role in seed development or germination.

Each of the seven *FTL* genes form well-supported subclade in the phylogenetic tree (Figure 4B; but see details of *FTL5* below). The *FTL* genes from the same genome tend to form clades, regardless of whether they originate from diploids or polyploids. The *FTL* genes from the G and H genomes of *C. sosnowskyi* and *C. vulvaria*, respectively, are located at

basal positions for each *FTL* subclade, diverging just after the *FTL* gene from the outgroup species (C. *rubrum*) supporting their early ancestry within the genus. The *FTL5* genes of the H and G subgenomes comprise a sister clade to all other *FTL5* and *FTL6* genes, indicating that the duplication event responsible for the creation of the *FTL5* and *FTL6* paralogs occurred after the divergence of the ancestor of the H and G genomes. *FTL6* underwent additional duplication in the B subgenome of *C. berlandieri* ssp. *nuttalliae* and in the C subgenome of hexaploid *C. opulifolium*. The *FTL6a* copy in the C subgenome of *C. opulifolium* is a pseudogene, carrying a large deletion in exon 1. Interestingly, the functional *FTL6b* copy from the C subgenome is more similar to the D-genome *FTL6* genes, suggesting that this copy is evolving dynamically (Figure 4B, *FTL6/5* clade). Indeed, many *FTL* genes have been often lost or pseudogenized. *Chenopodium acuminatum* (D) possesses the complete set of seven *FTL* genes, while *C. vulvaria*, which speciated prior to the duplication event that produced *FTL5* and *FTL6*, contains the complete set of six *FTLs* (Table S6). The extent of gene loss differed between the clades, with clade I genes less prone to loss or pseudogenization. *FTL2-1* is the least dispensable gene, present in an intact form in all the genomes, except for the B genome of *C. opulifolium*. It is followed by *FTL1*, which is pseudogenized in only three subgenomes in two hexaploids (*C. opulifolium* C and *C. formosanum* D) and one tetraploid (*C. berlandieri* B). At least one intact copy of *FTL1* and *FTL2-1* is present in each *Chenopodium* species, reflecting their critical role in flowering time regulation. In contrast, clade II exhibits numerous cases of both pseudogenization and gene losses. The most dispensable gene is *FTL4*, existing as an intact gene in only four genomes. It is missing in the diploids *C. pamiricum* (E) and *C. pallidicaule* (A), which do not harbor any intact copy of the gene. Similarly, *FTL3* and *FTL5* have been lost or pseudogenized in seven and five of the 20 genomes, including in the diploids *C. ficifolium* (B) and *C. pallidicaule* (A), respectively, suggesting both are dispensable (Table S6).

Unlike the clade I *FTL* genes, which regulate flowering, clade II *FTL* genes are not involved in floral induction or inhibition, as they are not expressed in leaves or flowers at any developmental stage. Nonetheless, at least one clade II gene has been identified in each sequenced *Amaranthaceae* species, including sugar beet and spinach. Due to the absence of clade II transcripts in floral transcriptomic studies, these genes often go unnoticed. However, the low but significant expression of *FTL3* in *C. rubrum* seeds suggests a role in seed development and/or germination (Drabešová et al., 2016). Clade II genes have rapidly expanded in *Chenopodium*, with up to four paralogs per genome in species like *C. acuminatum* and *C. opulifolium*. Moreover, clade II *FTL* genes have experienced more pseudogenizations and losses than clade I genes, making their evolution recent and highly dynamic. This dynamic evolution may contribute to the diversification and success of chenopods across a wide variety of niche environments, making the study of clade II gene functions potentially insightful.

4 | CONCLUSIONS

The present effort to survey pangenome sequences of Eurasian and American *Chenopodium* revealed a series of genomes we now know to be highly divergent in terms of overall size, due to changes driven by and repeatome sequences, as well as for synteny and collinearity due to structural variation. The greatest amount of genome diversity appears to be within a Eurasian center of origin, where speciation has been primarily driven by major genome divergence into groups A–H, followed by wholesale allopolyploidization. In contrast, New World *Chenopodium* evolution has been characterized by adaptive radiation of AA diploids into 40+ species (Young et al., 2023) and the formation of a single, widespread allotetraploid AABB complex. This pangenome resource provides essential data to identify genes underlying common and divergent domestication traits of AA, AABB, and BBCCDD taxa in such diverse locations as the Andean, Mesoamerican, and Himalayan highlands; the woodlands of eastern North America; and the subtropical island of Taiwan (Jarvis et al., 2022; Maughan et al., 2024; Partap & Kapoor, 1985). These sequenced species may harbor diverse strategies for dealing with salinity, drought, extreme cold, heat, and biotic stressors. This set of notoriously resilient, and phenotypically plastic, species should also provide clues regarding genes underlying weediness. We note that further studies are needed to address the diversity of *Chenopodium* and close relatives within Australia, where the genus has notably diverged into multiple perennial, woody shrubs, some with fleshy fruits (reviewed by Mosyakin & Iamonico, 2017).

AUTHOR CONTRIBUTIONS

Kate E. Jaggi: Formal analysis; investigation; methodology; writing—original draft; writing—review and editing. **Karol Krak**: Conceptualization; formal analysis; investigation; methodology; resources; validation; writing—original draft; writing—review and editing. **Helena Štorchová**: Conceptualization; formal analysis; investigation; methodology; writing—original draft; writing—review and editing. **Bohumil Mandák**: Conceptualization; methodology; project administration; resources; validation; writing—review and editing. **Ashley Marcheschi**: Data curation; formal analysis; investigation. **Alexander Belyayev**: Methodology; resources; validation; writing—review and editing. **Eric N. Jellen**: Conceptualization; data curation; funding acquisition; investigation; methodology; resources; supervision; validation; writing—review and editing. **John Sproul**: Con-

ceptualization; data curation; funding acquisition; investigation; methodology; visualization; writing—review and editing. **David Jarvis**: Conceptualization; data curation; funding acquisition; investigation; methodology; visualization; writing—review and editing. **Peter J. Maughan**: Conceptualization; data curation; funding acquisition; investigation; methodology; project administration; resources; supervision; writing—original draft; writing—review and editing.

ACKNOWLEDGMENTS

This research was supported in part by the U.S. Department of Agriculture, National Institute of Food and Agriculture, accession number 1022158 (David Jarvis, Peter J. Maughan, Eric N. Jellen). Karol Krak and Bohumil Mandák were supported with funding from the Czech Science Foundation, grant number 20–20286S and long-term research development project RVO67985939. Helena Štorchová was supported from the project TowArds Next GENeration Crops, reg. no. CZ.02.01.01/00/22_008/0004581, within the ERDF Programme Johanes Amos Comenius. M. Lomonosova provided the seeds of *C. pamiricum*. We express our gratitude to Matthew Gregg, Brigham Young University, for his assistance in developing and maintaining quinoaDB (http://quinoadb.org/).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Raw PacBio (genome assembly), Hi-C (scaffolding), and IsoSeq (annotation) reads have been deposited in GenBank under BioProject ID PRJNA1132190 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA1132190). The six de novo assembled genomes (*C. acuminatum*, *C. pamiricum*, *C. strictum*, *C. vulvaria*, *C. sosnowskyi*, and *C. opulifolium*) as well as *C. ficifolium* have also been deposited at DDBJ/ENA/GenBank under the accessions JBGFFP000000000- JBGFFV000000000. The genomes and annotations are also available on the Comparative Genomics (CoGe) platform of CyVerse https://genomevolution.org/coge/) and on quinoaDB (http://quinoadb.org/). The CoGe and quinoaDB platforms allow users to download bulk data (genome+annotation), view the genome using JBrowse, and perform a variety of comparative genome analyses, including BLAST analyses and multispecies synteny comparisons. A complete listing of raw data and whole genome assemblies deposited in NCBI, CoGe, and quinoaDB is provided in Table S7. Repeat sequence libraries for each species have been submitted to the Dfam database (www.dfam.org).

ORCID

Eric N. Jellen https://orcid.org/0000-0002-7906-4845
Peter J. Maughan https://orcid.org/0000-0003-3714-3411

REFERENCES

Abeyawardana, O. A. J., Moravec, T., Krüger, M., Belz, C., Gutierrez-Larruscain, D., Vondráková, Z., Eliášová, K., & Štorchová, H. (2023). The FLOWERING LOCUS T LIKE 2-1 gene of *Chenopodium* triggers precocious flowering in *Arabidopsis* seedlings. *Plant Signaling & Behavior*, *18*(1), 2239420. https://doi.org/10.1080/15592324.2023.2239420

Abugoch James, L. E. (2009). Quinoa (*Chenopodium quinoa* Willd.). In S. L. Taylor (Ed.), *Advances in food and nutrition research* (pp. 1–31). Elsevier. https://doi.org/10.1016/s1043-4526(09)58001-1

Alandia, G., Rodriguez, J. P., Jacobsen, S.-E., Bazile, D., & Condori, B. (2020). Global expansion of quinoa and challenges for the Andean region. *Global Food Security*, *26*, 100429. https://doi.org/10.1016/j.gfs.2020.100429

Bajwa, A. A., Zulfiqar, U., Sadia, S., Bhowmik, P., & Chauhan, B. S. (2019). A global perspective on the biology, impact and management of *Chenopodium album* and *Chenopodium murale*: Two troublesome agricultural and environmental weeds. *Environmental Science and Pollution Research*, *26*, 5357–5371. https://doi.org/10.1007/s11356-018-04104-y

Basantes-Morales, E. R., Alconada, M. M., & Pantoja, J. L. (2019). Quinoa (*Chenopodium quinoa* Willd.) production in the Andean region: Challenges and potentials. *Journal of Experimental Agriculture International*, *36*(6), 1–18. https://doi.org/10.9734/jeai/2019/v36i630251

Belyayev, A., Jandová, M., Josefiová, J., Kalendar, R., Mahelka, V., Mandák, B., & Krak, K. (2020). The major satellite DNA families of the diploid *Chenopodium album* aggregate species: Arguments for and against the "library hypothesis". *PLoS ONE*, *15*, e0241206. https://doi.org/10.1371/journal.pone.0241206

Belyayev, A., Josefiová, J., Jandová, M., Kalendar, R., Krak, K., & Mandák, B. (2019). Natural history of a satellite DNA family: From the ancestral genome component to species-specific sequences, concerted and non-concerted evolution. *International Journal of Molecular Sciences*, *20*, 1201. https://doi.org/10.3390/ijms20051201

Belyayev, A., Josefiová, J., Jandová, M., Mahelka, V., Krak, K., & Mandák, B. (2020). Transposons and satellite DNA: On the origin of the major satellite DNA family in the *Chenopodium* genome. *Mobile DNA*, *11*, 20. https://doi.org/10.1186/s13100-020-00219-7

Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, *27*, 573–580. https://doi.org/10.1093/nar/27.2.573

Bhargava, A., & Srivastava, S. (2019). Advantages and cost of participatory plant breeding. *Participatory plant breeding: Concept and applications* (pp. 87–107). Springer.

Bickhart, D. M., Rosen, B. D., Koren, S., Sayre, B. L., Hastie, A. R., Chan, S., Lee, J., Lam, E. T., Liachko, I., Sullivan, S. T., Burton, J. N., Huson, H. J., Nystrom, J. C., Kelley, C. M., Hutchison, J. L., Zhou, Y., Sun, J., Crisà, A., Ponce de León, F. A., … Smith, T. P. L. (2017). Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nature Genetics*, *49*, 643–650. https://doi.org/10.1038/ng.3802

Bucchini, F., Del Cortona, A., Kreft, Ł., Botzki, A., Van Bel, M., & Vandepoele, K. (2021). TRAPID 2.0: A web application for taxonomic and functional analysis of *de novo* transcriptomes. *Nucleic Acids Research*, *49*, e101–e101. https://doi.org/10.1093/nar/gkab565

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., & Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nature Biotechnology*, *31*, 1119–1125. https://doi.org/10.1038/nbt.2727

Chen, Y., Zhang, Y., Wang, A. Y., Gao, M., & Chong, Z. (2021). Accurate long-read *de novo* assembly evaluation with Inspector. *Genome Biology*, *22*, 312. https://doi.org/10.1186/s13059-021-02527-4

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*, 170–175. https://doi.org/10.1038/s41592-020-01056-5

Dakhili, S., Abdolalizadeh, L., Hosseini, S. M., Shojaee-Aliabadi, S., & Mirmoghtadaie, L. (2019). Quinoa protein: Composition, structure and functional properties. *Food Chemistry*, *299*, 125161. https://doi.org/10.1016/j.foodchem.2019.125161

Danecek, P., Bonfield, J., Liddle, J., Marshall, J., Ohan, V., Pollard, M., Whitman, A., Keane, T., McCarthy, S., Davies, R., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Drabešová, J., Černá, L., Mašterová, H., Koloušková, P., Potocký, M., & Štorchová, H. (2016). The evolution of the FT/TFL1 genes in *Amaranthaceae* and their expression patterns in the course of vegetative growth and flowering in *Chenopodium rubrum*. *G3 Genes|Genomes|Genetics*, *6*, 3065–3076. https://doi.org/10.1534/g3.116.028639

Du, J., Tian, Z., Hans, C. S., Laten, H. M., Cannon, S. B., Jackson, S. A., Shoemaker, R. C., & Ma, J. (2010). Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: Insights from genome-wide analysis and multi-specific comparison. *The Plant Journal*, *63*, 584–598. https://doi.org/10.1111/j.1365-313X.2010.04263.x

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*, 1792–1797. https://doi.org/10.1093/nar/gkh340

Emms, D. M., & Kelly, S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biology*, *20*, 238. https://doi.org/10.1186/s13059-019-1832-y

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, *117*, 9451–9457. https://doi.org/10.1073/pnas.1921046117

Gabriel, L., Brůna, T., Hoff, K. J., Ebel, M., Lomsadze, A., Borodovsky, M., & Stanke, M. (2024). *BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA*. https://doi.org/10.1101/2023.06.10.544449

Galindo-González, L., Mhiri, C., Deyholos, M. K., & Grandbastien, M.-A. (2017). LTR-retrotransposons in plants: Engines of evolution. *Gene*, *626*, 14–25. https://doi.org/10.1016/j.gene.2017.04.051

Gandarillas, H., Alandia, S., Cardozo, A., & Mujica, A. (1979). "*Mejoramiento genético'' in Qinua y Kaniwa cultivos Andinos*. Instituto Interamericano de Ciencias Agrícolas.

Garrido-Ramos, M. A. (2017). Satellite DNA: An evolving topic. *Genes*, *8*(9), 230. https://doi.org/10.3390/genes8090230

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., Sharpe, T., Hall, G., Shea, T. P., Sykes, S., Berlin, A. M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E. S., & Jaffe, D. B. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, *108*, 1513–1518. https://doi.org/10.1073/pnas.1017351108

Gutierrez-Larruscain, D., Krüger, M., Abeyawardana, O. A. J., Belz, C., Dobrev, P. I., Vaňková, R., Eliášová, K., Vondráková, Z., Juříček, M., & Štorchová, H. (2022). The high concentrations of abscisic, jasmonic, and salicylic acids produced under long days do not accelerate flowering in *Chenopodium ficifolium* 459. *Plant Science*, *320*, 111279. https://doi.org/10.1016/j.plantsci.2022.111279

Heitkam, T., Weber, B., Walter, I., Liedtke, S., Ost, C., & Schmidt, T. (2020). Satellite DNA landscapes after allotetraploidization of quinoa (*Chenopodium quinoa*) reveal unique A and B subgenomes. *The Plant Journal*, *103*, 32–52. https://doi.org/10.1111/tpj.14705

Hirsch, C. N., Foerster, J. M., Johnson, J. M., Sekhon, R. S., Muttoni, G., Vaillancourt, B., Peñagaricano, F., Lindquist, E., Pedraza, M. A., Barry, K., de Leon, N., Kaeppler, S. M., & Buell, C. R. (2014). Insights into the maize pan-genome and pan-transcriptome. *The Plant Cell*, *26*, 121–135. https://doi.org/10.1105/tpc.113.119982

Huang, N., & Li, H. (2023). Compleasm: A faster and more accurate reimplementation of BUSCO. *Bioinformatics*, *39*, btad595. https://doi.org/10.1093/bioinformatics/btad595

Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmöckel, S. M., Li, B., Borm, T. J. A., Ohyanagi, H., Mineta, K., Michell, C. T., Saber, N., Kharbatia, N. M., Rupper, R. R., Sharp, A. R., Dally, N., Boughton, B. A., Woo, Y. H., Gao, G., Schijlen, E. G. W. M., Guo, X., … Tester, M. (2017). The genome of *Chenopodium quinoa*. *Nature*, *542*, 307–312. https://doi.org/10.1038/nature21370

Jarvis, D. E., Sproul, J. S., Navarro-Domínguez, B., Krak, K., Jaggi, K., Huang, Y.-F., Huang, T.-Y., Lin, T. C., Jellen, E. N., & Maughan, P. J. (2022). Chromosome-scale genome assembly of the hexaploid Taiwanese goosefoot "Djulis" (*Chenopodium formosanum*). *Genome Biology and Evolution*, *14*, evac120. https://doi.org/10.1093/gbe/evac120

Jayakodi, M., Padmarasu, S., Haberer, G., Bonthala, V. S., Gundlach, H., Monat, C., Lux, T., Kamal, N., Lang, D., Himmelbach, A., Ens, J., Zhang, X.-Q., Angessa, T. T., Zhou, G., Tan, C., Hill, C., Wang, P., Schreiber, M., Boston, L. B., … Stein, N. (2020). The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature*, *588*, 284–289. https://doi.org/10.1038/s41586-020-2947-8

Jin, J.-J., Yu, W.-B., Yang, J.-B., Song, Y., dePamphilis, C. W., Yi, T.-S., & Li, D.-Z. (2020). GetOrganelle: A fast and versatile toolkit for accurate *de novo* assembly of organelle genomes. *Genome Biology*, *21*, 241. https://doi.org/10.1186/s13059-020-02154-5

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics (Oxford, England)*, *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*, 587–589. https://doi.org/10.1038/nmeth.4285

Kardailsky, I., Shukla, V. K., Ahn, J. H., Dagenais, N., Christensen, S. K., Nguyen, J. T., Chory, J., Harrison, M. J., & Weigel, D. (1999). Activation tagging of the floral inducer *FT*. *Science*, *286*, 1962–1965. https://doi.org/10.1126/science.286.5446.1962

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability.

Molecular Biology and Evolution, *30*, 772–780. https://doi.org/10.1093/molbev/mst010

Kistler, L., & Shapiro, B. (2011). Ancient DNA confirms a local origin of domesticated chenopod in eastern North America. *Journal of Archaeological Science*, *38*, 3549–3554. https://doi.org/10.1016/j.jas.2011.08.023

Kolano, B., Gardunia, B. W., Michalska, M., Bonifacio, A., Fairbanks, D., Maughan, P. J., Coleman, C. E., Stevens, M. R., Jellen, E. N., & Maluszynska, J. (2011). Chromosomal localization of two novel repetitive sequences isolated from the *Chenopodium quinoa* Willd. genome. *Genome*, *54*, 710–717. https://doi.org/10.1139/g11-035

Koprivova, A., Schuck, S., Jacoby, R. P., Klinkhammer, I., Welter, B., Leson, L., Martyn, A., Nauen, J., Grabenhorst, N., Mandelkow, J. F., Zuccaro, A., Zeier, J., & Kopriva, S. (2019). Root-specific camalexin biosynthesis controls the plant growth-promoting effects of multiple bacterial strains. *Proceedings of the National Academy of Sciences*, *116*, 15735–15744. https://doi.org/10.1073/pnas.1818604116

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Research*, *27*, 722–736. https://doi.org/10.1101/gr.215087.116

Krak, K., Vít, P., Belyayev, A., Douda, J., Hreusová, L., & Mandák, B. (2016). Allopolyploid origin of *Chenopodium album* s. str. (*Chenopodiaceae*): A molecular and cytogenetic insight. *PLoS ONE*, *11*, e0161063. https://doi.org/10.1371/journal.pone.0161063

Kursel, L. E., & Malik, H. S. (2016). Centromeres. *Current Biology*, *26*, R487–R490. https://doi.org/10.1016/j.cub.2016.05.031

Kuznetsov, D., Tegenfeldt, F., Manni, M., Seppey, M., Berkeley, M., Kriventseva, E. V., & Zdobnov, E. M. (2023). OrthoDB v11: Annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Research*, *51*, D445–D451. https://doi.org/10.1093/nar/gkac998

Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, *6*, 1287. https://doi.org/10.12688/f1000research.12232.1

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*, 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*, *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698

Li, W., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, *22*, 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Li, Y., Zhou, G., Ma, J., Jiang, W., Jin, L., Zhang, Z., Guo, Y., Zhang, J., Sui, Y., Zheng, L., Zhang, S., Zuo, Q., Shi, X., Li, Y., Zhang, W., Hu, Y., Kong, G., Hong, H., Tan, B., … Qiu, L. (2014). *De novo* assembly of soybean wild relatives for pan-genome analysis of diversity and agronomic traits. *Nature Biotechnology*, *32*, 1045–1052. https://doi.org/10.1038/nbt.2979

Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M., & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, *11*, e78526. https://doi.org/10.7554/eLife.78526

Ludwig, C. D., Maughan, P. J., Jellen, E. N., & Davis, T. M. (2025). The genome of *Chenopodium ficifolium*: Developing genetic resources and a diploid model system for allotetraploid quinoa. *BioRxiv*. https://doi.org/10.1101/2025.01.17.633571

Ma, J., & Bennetzen, J. L. (2004). Rapid recent growth and divergence of rice nuclear genomes. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 12404–12410. https://doi.org/10.1073/pnas.0403715101

Madlung, A., & Wendel, J. F. (2013). Genetic and epigenetic aspects of polyploid evolution in plants. *Cytogenetic and Genome Research*, *140*, 270–285. https://doi.org/10.1159/000351430

Mandák, B., Krak, K., Vít, P., Pavlíková, Z., Lomonosova, M. N., Habibi, F., Wang, L., Jellen, E. N., & Douda, J. (2016). How genome size variation is linked with evolution within Chenopodium sensu lato. *Perspectives in Plant Ecology, Evolution and Systematics*, *23*, 18–32.

Mandák, B., Krak, K., Vít, P., Lomonosova, M. N., Belyayev, A., Habibi, F., Wang, L., Douda, J., & Štorchová, H. (2018). Hybridization and polyploidization within the *Chenopodium album* aggregate analysed by means of cytological and molecular markers. *Molecular Phylogenetics and Evolution*, *129*, 189–201. https://doi.org/10.1016/j.ympev.2018.08.016

Mandák, B., Trávníček, P., Paštová, L., & Kořínková, D. (2012). Is hybridization involved in the evolution of the *Chenopodium album* aggregate? An analysis based on chromosome counts and genome size estimation. *Flora—Morphology, Distribution, Functional Ecology of Plants*, *207*, 530–540. https://doi.org/10.1016/j.flora.2012.03.010

Mangelson, H., Jarvis, D. E., Mollinedo, P., Rollano-Penaloza, O. M., Palma-Encinas, V. D., Gomez-Pando, L. R., Jellen, E. N., & Maughan, P. J. (2019). The genome of *Chenopodium pallidicaule*: An emerging Andean super grain. *Applications in Plant Sciences*, *7*, e11300. https://doi.org/10.1002/aps3.11300

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, *38*, 4647–4654. https://doi.org/10.1093/molbev/msab199

Maughan, P. J., Chaney, L., Lightfoot, D. J., Cox, B. J., Tester, M., Jellen, E. N., & Jarvis, D. E. (2019). Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (*Chenopodium quinoa* Willd.). *Scientific Reports*, *9*, 185. https://doi.org/10.1038/s41598-018-36693-6

Maughan, P. J., Jarvis, D. E., de la Cruz-Torres, E., Jaggi, K. E., Warner, H. C., Marcheschi, A. K., Bertero, H. D., Gomez-Pando, L., Fuentes, F., Mayta-Anco, M. E., Curti, R., Rey, E., Tester, M., & Jellen, E. N. (2024). North American pitseed goosefoot (*Chenopodium berlandieri*) is a genetic resource to improve *Andean quinoa* (*C. quinoa*). *Scientific Reports*, *14*(1), 12345. https://doi.org/10.1038/s41598-024-63106-8

Maughan, P. J., Kolano, B. A., Maluszynska, J., Coles, N. D., Bonifacio, A., Rojas, J., Coleman, C. E., Stevens, M. R., Fairbanks, D. J., Parkinson, S. E., & Jellen, E. N. (2006). Molecular and cytological characterization of ribosomal RNA genes in *Chenopodium quinoa* and *Chenopodium berlandieri*. *Genome*, *49*, 825–839. https://doi.org/10.1139/g06-033

McCartney, A. M., Shafin, K., Alonge, M., Bzikadze, A. V., Formenti, G., Fungtammasan, A., Howe, K., Jain, C., Koren, S., Logsdon, G. A., Miga, K. H., Mikheenko, A., Paten, B., Shumate, A., Soto, D. C., Sović, I., Wood, J. M. D., Zook, J. M., Phillippy, A. M., & Rhie, A. (2022). Chasing perfection: Validation and polishing strategies for telomere-to-telomere genome assemblies. *Nature Methods*, *19*, 687–695. https://doi.org/10.1038/s41592-022-01440-3

McGrath, J. M., Funk, A., Galewski, P., Ou, S., Townsend, B., Davenport, K., Daligault, H., Johnson, S., Lee, J., Hastie, A., Darracq, A., Willems, G., Barnes, S., Liachko, I., Sullivan, S., Koren, S., Phillippy, A., Wang, J., Liu, T., … Dorn, K. M. (2023). A contiguous *de novo* genome assembly of sugar beet EL10 (*Beta vulgaris* L.). *DNA Research*, *30*(1), dsac033. https://doi.org/10.1093/dnares/dsac033

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway computing environments workshop (GCE)* (pp. 1–8). IEEE.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mirarab, S., Reaz, R., Bayzid, M. S., Zimmermann, T., Swenson, M. S., & Warnow, T. (2014). ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics*, *30*, i541–i548. https://doi.org/10.1093/bioinformatics/btu462

Mosyakin, S. L., & Iamonico, D. (2017). Nomenclatural changes in *Chenopodium* (incl. *Rhagodia*) (*Chenopodiaceae*), with considerations on relationships of some Australian taxa and their possible Eurasian relatives. *Nuytsia—The Journal of the Western Australian Herbarium*, *28*, 255–271. https://doi.org/10.58828/nuy00843

Navarro, C., Abelenda, J. A., Cruz-Oró, E., Cuéllar, C. A., Tamaki, S., Silva, J., Shimamoto, K., & Prat, S. (2011). Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. *Nature*, *478*, 119–122. https://doi.org/10.1038/nature10431

Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology and Evolution*, *32*, 268–274. https://doi.org/10.1093/molbev/msu300

Nurk, S., Walenz, B. P., Rhie, A., Vollger, M. R., Logsdon, G. A., Grothe, R., Miga, K. H., Eichler, E. E., Phillippy, A. M., & Koren, S. (2020). HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Research*, *30*, 1291–1305. https://doi.org/10.1101/gr.263566.120

Oldenburg, D. J., & Bendich, A. J. (2004). Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. *Journal of Molecular Biology*, *335*, 953–970. https://doi.org/10.1016/j.jmb.2003.11.020

Orzechowska, M., Majka, M., Weiss-Schneeweiss, H., Kovařík, A., Borowska-Zuchowska, N., & Kolano, B. (2018). Organization and evolution of two repetitive sequences, 18–24J and 12–13P, in the genome of *Chenopodium* (*Amaranthaceae*). *Genome*, *61*, 643–652. https://doi.org/10.1139/gen-2018-0044

Palmer, J. D. (1983). Chloroplast DNA exists in two orientations. *Nature*, *301*, 92–93. https://doi.org/10.1038/301092a0

Partap, T., & Kapoor, P. (1985). The Himalayan grain chenopods. I. Distribution and ethnobotany. *Agriculture, Ecosystems & Environment*, *14*, 185–199. https://doi.org/10.1016/0167-8809(85)90035-0

Patiranage, D. S. R., Asare, E., Maldonado-Taipe, N., Rey, E., Emrani, N., Tester, M., & Jung, C. (2021). Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions. *Plant, Cell & Environment*, *44*, 2565–2579. https://doi.org/10.1111/pce.14071

Peichel, C. L., Sullivan, S. T., Liachko, I., & White, M. A. (2017). Improvement of the threespine stickleback genome using a Hi-C-based proximity-guided assembly. *Journal of Heredity*, *108*, 693–700. https://doi.org/10.1093/jhered/esx058

Pin, P. A., Benlloch, R., Bonnet, D., Wremerth-Weich, E., Kraft, T., Gielen, J. J. L., & Nilsson, O. (2010). An antagonistic pair of *FT* homologs mediates the control of flowering time in sugar beet. *Science*, *330*, 1397–1400. https://doi.org/10.1126/science.1197004

Poonia, A., & Upadhayay, A. (2015). *Chenopodium album* Linn: Review of nutritive value and biological properties. *Journal of Food Science and Technology*, *52*, 3977–3985. https://doi.org/10.1007/s13197-014-1553-x

Raskina, O., Barber, J. C., Nevo, E., & Belyayev, A. (2008). Repetitive DNA and chromosomal rearrangements: Speciation-related events in plant genomes. *Cytogenetic and Genome Research*, *120*, 351–357. https://doi.org/10.1159/000121084

Reijnders, M. J. M. F., & Waterhouse, R. M. (2021). Summary visualizations of gene ontology terms with GO-Figure! *Frontiers in Bioinformatics*, *1*, 638255. https://doi.org/10.3389/fbinf.2021.638255

Repo-Carrasco, R., Espinoza, C., & Jacobsen, S.-E. (2003). Nutritional value and use of the Andean crops quinoa (*Chenopodium quinoa*) and kañiwa (*Chenopodium pallidicaule*). *Food Reviews International*, *19*, 179–189. https://doi.org/10.1081/FRI-120018884

Rey, E., Abrouk, M., Dufau, I., Rodde, N., Saber, N., Cizkova, J., Fiene, G., Stanschewski, C., Jarvis, D., Jellen, E., Maughan, P., von Baer, I., Troukhan, M., Kravchuk, M., Hribova, E., Cauet, S., Krattinger, S., & Tester, M. (2024). *Genome assembly of a diversity panel of Chenopodium quinoa*. bioRxiv. https://doi.org/10.1101/2024.07.07.602379

Rey, E., Maughan, P. J., Maumus, F., Lewis, D., Wilson, L., Fuller, J., Schmöckel, S. M., Jellen, E. N., Tester, M., & Jarvis, D. E. (2023). A chromosome-scale assembly of the quinoa genome provides insights into the structure and dynamics of its subgenomes. *Communications Biology*, *6*, 1263. https://doi.org/10.1038/s42003-023-05613-4

Romiguier, J., & Roux, C. (2017). Analytical biases associated with GC content in molecular evolution. *Frontiers in Genetics*, *8*, 16. https://doi.org/10.3389/fgene.2017.00016

Russo, A., Mayjonade, B., Frei, D., Potente, G., Kellenberger, R. T., Frachon, L., Copetti, D., Studer, B., Frey, J. E., Grossniklaus, U., & Schlüter, P. M. (2022). Low-input high-molecular-weight DNA extraction for long-read sequencing from plants of diverse families. *Frontiers in Plant Science*, *13*, 883897. https://doi.org/10.3389/fpls.2022.883897

She, H., Liu, Z., Xu, Z., Zhang, H., Cheng, F., Wu, J., Wang, X., & Qian, W. (2022). Comparative chloroplast genome analyses of cultivated spinach and two wild progenitors shed light on the phylogenetic relationships and variation. *Scientific Reports*, *12*, 856. https://doi.org/10.1038/s41598-022-04918-4

Shi, L., Wu, Y., & Sheen, J. (2018). TOR signaling in plants: Conservation and innovation. *Development*, *145*(13), dev160887. https://doi.org/10.1242/dev.160887

Silvestri, V., & Gil, F. (2000). Alogamia en quinua. *Tasa en Mendoza (Argentina)* (pp. 71–76). Revista de la Facultad de Ciencias Agrarias, Universidad Nacional de Cuyo.

Smit, A. F. A., Hubley, R., & Green, P. (2013–2015). *RepeatMasker Open-4.0*. https://www.repeatmasker.org/

Smith, B. D., & Yarnell, R. A. (2009). Initial formation of an indigenous crop complex in eastern North America at 3800 B.P. *Proceedings of the National Academy of Sciences*, *106*, 6561–6566. https://doi.org/10.1073/pnas.0901846106

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Štorchová, H. (2020). The evolution of the FLOWERING LOCUS T-Like (FTL) genes in the goosefoot subfamily *Chenopodioideae*. *Evolutionary biology—A transdisciplinary approach* (pp. 325–335). Springer International Publishing.

Tenaillon, M. I., Hollister, J. D., & Gaut, B. S. (2010). A triptych of the evolution of plant transposable elements. *Trends in Plant Science*, *15*, 471–478. https://doi.org/10.1016/j.tplants.2010.05.003

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, *45*, W6–W11. https://doi.org/10.1093/nar/gkx391

Tinker, N. A., Wight, C. P., Bekele, W. A., Yan, W., Jellen, E. N., Renhuldt, N. T., Sirijovski, N., Lux, T., Spannagl, M., & Mascher, M. (2022). Genome analysis in *Avena sativa* reveals hidden breeding barriers and opportunities for oat improvement. *Communications Biology*, *5*, 474. https://doi.org/10.1038/s42003-022-03256-5

Vaillancourt, B., & Buell, C. R. (2019). *High molecular weight DNA isolation method from diverse plant species for use with Oxford Nanopore sequencing*. bioRxiv. https://doi.org/10.1101/783159

Vega-Gálvez, A., Miranda, M., Vergara, J., Uribe, E., Puente, L., & Martínez, E. A. (2010). Nutrition facts and functional potential of quinoa (*Chenopodium quinoa* Willd.), an ancient Andean grain: A review. *Journal of the Science of Food and Agriculture*, *90*, 2541–2547. https://doi.org/10.1002/jsfa.4158

Vít, P., Krak, K., Trávníček, P., Douda, J., Lomonosova, M. N., & Mandák, B. (2016). Genome size stability across Eurasian *Chenopodium* species (*Amaranthaceae*). *Botanical Journal of the Linnean Society*, *182*(3), 637–649. Portico. https://doi.org/10.1111/boj.12474

Wang, B., Hou, M., Shi, J., Ku, L., Song, W., Li, C., Ning, Q., Li, X., Li, C., Zhao, B., Zhang, R., Xu, H., Bai, Z., Xia, Z., Wang, H., Kong, D., Wei, H., Jing, Y., Dai, Z., … Wang, H. (2023). *De novo* genome assembly and analyses of 12 founder inbred lines provide insights into maize heterosis. *Nature Genetics*, *55*, 312–323. https://doi.org/10.1038/s41588-022-01283-w

Wang, H., Xu, D., Wang, S., Wang, A., Lei, L., Jiang, F., Yang, B., Yuan, L., Chen, R., Zhang, Y., & Fan, W. (2023). Chromosome-scale *Amaranthus tricolor* genome provides insights into the evolution of the genus *Amaranthus* and the mechanism of betalain biosynthesis. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, *30*(1), dsac050. https://doi.org/10.1093/dnares/dsac050

Wang, T., Duan, S., Xu, C., Wang, Y., Zhang, X., Xu, X., Chen, L., Han, Z., & Wu, T. (2023). Pan-genome analysis of 13 Malus accessions reveals structural and sequence variations associated with fruit traits. *Nature Communications*, *14*, 7377. https://doi.org/10.1038/s41467-023-43270-7

Wilson, H. D., & Heiser, C. B. (1979). The origin and evolutionary relationships of 'Huauzontle' (*Chenopodium nuttaliae* Safford), domesticated chenopod of Mexico. *American Journal of Botany*, *66*(2), 198–206. https://doi.org/10.2307/2442525

Wlodzimierz, P., Hong, M., & Henderson, I. R. (2023). TRASH: Tandem repeat annotation and structural hierarchy. *Bioinformatics*, *39*(5), btad308. https://doi.org/10.1093/bioinformatics/btad308.[CrossRef]

Xu, H., Xiang, N., Du, W., Zhang, J., & Zhang, Y. (2022). Genetic variation and structure of complete chloroplast genome in alien monoecious and dioecious *Amaranthus* weeds. *Scientific Reports*, *12*, 8255. https://doi.org/10.1038/s41598-022-11983-2

Ye, J., McGinnis, S., & Madden, T. L. (2006). BLAST: Improvements for better sequence analysis. *Nucleic Acids Research*, *34*, W6–W9. https://doi.org/10.1093/nar/gkl164

Young, L. A., Maughan, P. J., Jarvis, D. E., Hunt, S. P., Warner, H. C., Durrant, K. K., Kohlert, T., Curti, R. N., Bertero, D., Filippi, G. A., Pospíšilíková, T., Krak, K., Mandák, B., & Jellen, E. N. (2023). A chromosome-scale reference of *Chenopodium watsonii* helps elucidate relationships within the North American A-genome *Chenopodium* species and with quinoa. *The Plant Genome*, *16*(3), e20349. https://doi.org/10.1002/tpg270010.20349

Zhang, Z., Gou, X., Xun, H., Bian, Y., Ma, X., Li, J., Li, N., Gong, L., Feldman, M., Liu, B., & Levy, A. A. (2020). Homoeologous exchanges occur through intragenic recombination generating novel transcripts and proteins in wheat and other polyploids. *Proceedings of the National Academy of Sciences*, *117*, 14561–14571. https://doi.org/10.1073/pnas.2003505117

Zhu, X. F., Wan, J. X., Wu, Q., Zhao, X. S., Zheng, S. J., & Shen, R. F. (2017). *PARVUS* affects aluminium sensitivity by modulating the structure of glucuronoxylan in *Arabidopsis thaliana*. *Plant, Cell & Environment*, *40*, 1916–1925. https://doi.org/10.1111/pce.12999

SUPPORTING INFORMATION

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

How to cite this article: Jaggi, K. E., Krak, K., Štorchová, H., Mandák, B., Marcheschi, A., Belyayev, A., Jellen, E. N., Sproul, J., Jarvis, D., & Maughan, P. J. (2025). A pangenome reveals LTR repeat dynamics as a major driver of genome evolution in *Chenopodium*. *The Plant Genome*, *18*, e70010. https://doi.org/10.1002/tpg2.70010