


# Chromosome-Scale Genome Assembly of the Hexaploid Taiwanese Goosefoot “Djulis” (*Chenopodium formosanum*)

David E. Jarvis <sup>1,\*</sup>, John S. Sproul<sup>2</sup>, Beatriz Navarro-Domínguez<sup>3</sup>, Karol Krak<sup>4</sup>, Kate Jaggi<sup>1</sup>, Yung-Fen Huang<sup>5</sup>, Tzu-Yun Huang<sup>6</sup>, Tzu Che Lin<sup>7</sup>, Eric N. Jellen<sup>1</sup>, and Peter J. Maughan<sup>1</sup>

<sup>1</sup>Department of Plant and Wildlife Sciences, Brigham Young University, Provo, Utah 84602, USA

<sup>2</sup>Department of Biology, University of Nebraska Omaha, Omaha, Nebraska 68182, USA

<sup>3</sup>Departamento de Genética, Facultad de Ciencias, Universidad de Granada, Granada 18071, Spain

<sup>4</sup>Department of Ecology, Czech University of Life Sciences, Prague 16500, Czech Republic

<sup>5</sup>Department of Agronomy, National Taiwan University, Taipei 10617, Taiwan

<sup>6</sup>Department of Crop Improvement, Taitung District Agricultural Research and Extension Station, Taitung City 950244, Taiwan

<sup>7</sup>Department of Plant Industry, National Pingtung University of Science and Technology, Neipu 91201, Taiwan

\*Corresponding author: E-mail: david\_jarvis@byu.edu.

Accepted: 21 July 2022

## Abstract

Djulis (*Chenopodium formosanum* Koidz.) is a crop grown since antiquity in Taiwan. It is a BCD-genome hexaploid ( $2n = 6x = 54$ ) domesticated form of lambsquarters (*C. album* L.) and a relative of the allotetraploid (AABB) *C. quinoa*. As with quinoa, djulis seed contains a complete protein profile and many nutritionally important vitamins and minerals. While still sold locally in Taiwanese markets, its traditional culinary uses are being lost as diets of younger generations change. Moreover, indigenous Taiwanese peoples who have long safeguarded djulis are losing their traditional farmlands. We used PacBio sequencing and Hi-C-based scaffolding to produce a chromosome-scale, reference-quality assembly of djulis. The final genome assembly spans 1.63 Gb in 798 scaffolds, with 97.8% of the sequence contained in 27 scaffolds representing the nine haploid chromosomes of each sub-genome of the species. Benchmarking of universal, single-copy orthologs indicated that 98.5% of the conserved orthologous genes for Viridiplantae are complete within the assembled genome, with 92.9% duplicated, as expected for a polyploid. A total of 67.8% of the assembly is repetitive, with the most common repeat being *Gypsy* long terminal repeat retrotransposons, which had significantly expanded in the B sub-genome. Gene annotation using Iso-Seq data from multiple tissues identified 75,056 putative gene models. Comparisons to quinoa showed strong patterns of synteny which allowed for the identification of homoeologous chromosomes, and sub-genome-specific sequences were used to assign homoeologs to each sub-genome. These results represent the first hexaploid genome assembly and the first assemblies of the C and D genomes of the Chenopodioideae subfamily.

**Key words:** long-read sequencing, chromosome-scale, polyploid evolution, retrotransposon, LTR, *Gypsy*.

## Significance

The high-quality genome assembly of djulis is the first sequenced hexaploid in the genus *Chenopodium*. The genome is a valuable resource for studying genome evolution in polyploids and will facilitate efforts to characterize, conserve, and improve this nutritious, understudied crop.

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

## Introduction

The *Chenopodium album* L. aggregate represents a taxonomically complex group whose evolutionary history has largely been shaped by hybridization and polyploidization (Mandák et al. 2018; Walsh et al. 2015; Štorchová et al. 2015). The variation in ploidy level includes taxa with diploid ( $2n = 18$ ), allotetraploid ( $2n = 36$ ), allohexaploid ( $2n = 54$ ), and decaploid ( $2n = 90$ ) chromosome counts (Mandák et al. 2016). Plants of this group have been domesticated as seed and/or vegetable crops at various times and places, mostly from weeds that are notorious colonizers of human-disturbed environments, like *C. album* L.. Fortuitously, these plants tend to be highly nutritious, with favorable levels of mineral nutrients and vitamins along with excellent amino acid profiles for their seed and foliar proteins (Yadav and Sehgal 2002; Repo-Carrasco et al. 2003; Vega-Gálvez et al. 2010; Pandey and Gupta 2014). Even as free-living weeds, *Chenopodium* species have been gathered as potherbs throughout human history (Huai and Pei 2000). The increasing popularity of the South American-native pseudocereal quinoa (*C. quinoa* Willd.,  $2n = 36$ , AABB sub-genomes) has aroused interest in its cultivated cousins, among them Andean cañahua or kañiwa [*C. pallidicaule* Aellen,  $2n = 18$ , AA (Mangelson et al. 2019)], Mesoamerican huauzontle [*C. berlandieri* ssp. *nuttalliae* (Safford) H.D. Wilson and Heiser,  $2n = 36$ , AABB; (Wilson and Jr 1979)], Himalayan bathua [*C. giganteum* D. Don,  $2n = 54$ , BBCCDD; (Partap and Kapoor 1985)], and Taiwanese djulis or hangli [*C. formosanum* Koidz.,  $2n = 54$ , BBCCDD; (Chu et al. 2016)].

Taiwanese djulis and Himalayan bathua are hypothesized to have arisen as independent Asian domestications from the cosmopolitan allohexaploid BBCCDD weed, lambsquarters (*C. album* L. [Krak et al. 2016; Mandák et al. 2018; Kolano et al. 2019]). Djulis was traditionally used as a fermentation starter for small grain winemaking by Taiwanese aboriginals (Tsai et al. 2010). Djulis is of particular interest not only as a high-quality food source, but also as a biological pigment source and for its documented medicinal properties, high levels and diversity of secondary metabolites—among them phytosterols and triterpenes (Tsai et al. 2010, 2011; Huang et al. 2019)—and for its potential as a bioenergy source (Yang et al. 2014). Djulis extracts and secondary metabolites demonstrated beneficial effects on cardiovascular diseases like hypertension (Chen et al. 2019); in improving symptoms for Type 2 diabetes patients (Li et al. 2021); in inhibiting formation of colon cancer (Lee et al. 2019); and for curtailing adipogenesis (Chyau et al. 2018). Djulis extract has also traditionally been used as an insecticide (Chio et al. 2013; Chuang et al. 2018). To facilitate characterization and preservation of djulis genetic diversity and to begin to characterize the genetic

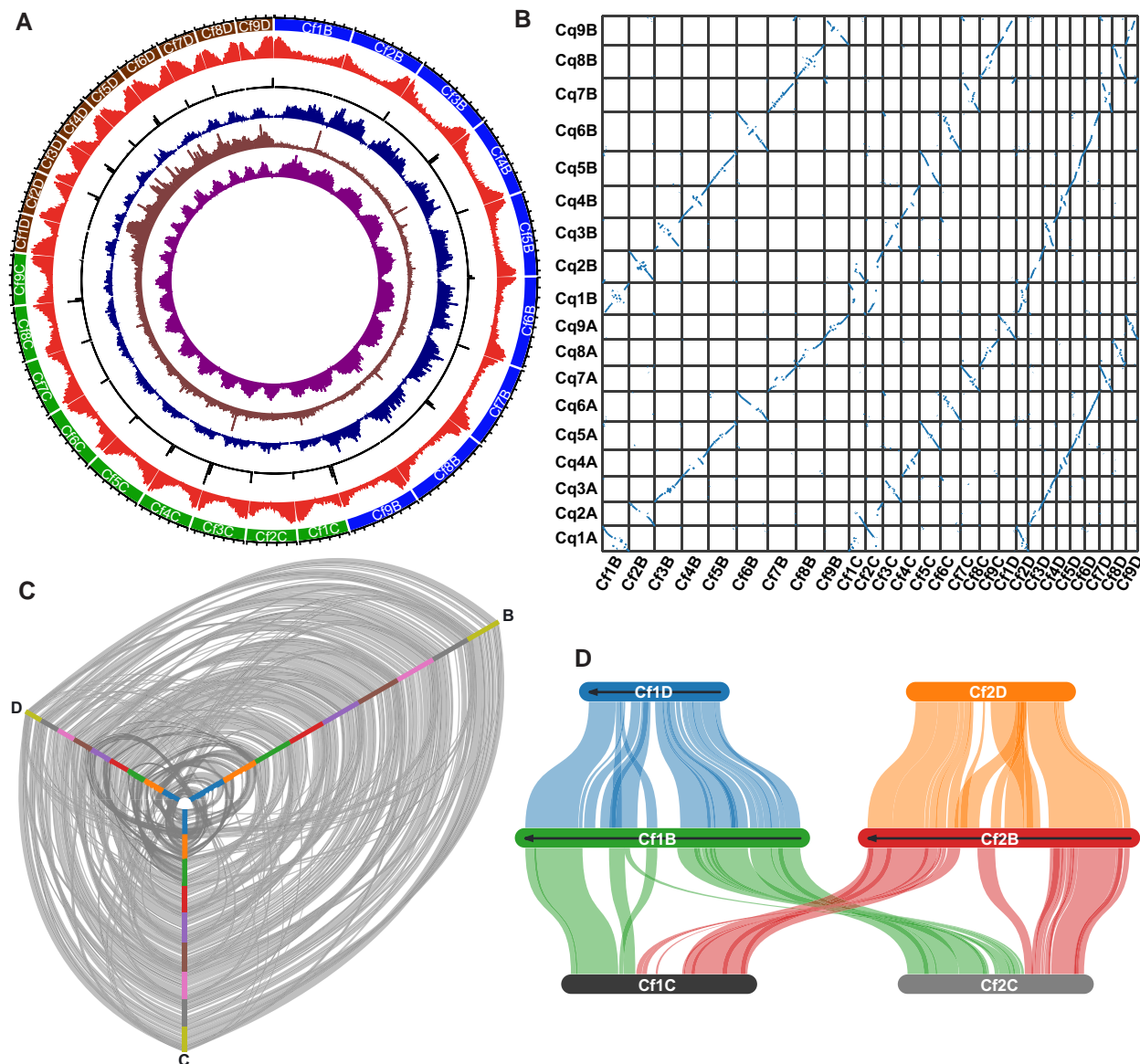
mechanisms underlying its desirable agronomic and medicinal traits, we present a reference-quality, whole-genome assembly of *C. formosanum* and compare its constituent B, C, and D sub-genomes for genic and repetitive sequence compositions.

## Results and Discussion

### Genome Assembly and Annotation

Cytogenetic analysis confirmed that *C. formosanum* is a hexaploid with 54 chromosomes (supplementary fig. S1, Supplementary Material online). DNA sequencing of the *C. formosanum* genome using three PacBio HiFi cells produced 67.906 Gb in 3.982 million reads with minimum quality scores of Q20 and an average read length of 16.923 kb. Using this data, we estimated a genome size of 1.693 Gb (supplementary fig. S2, Supplementary Material online), indicating that the sequencing data represented approximately 41X coverage. Sequencing reads were assembled into 1,914 contigs with a total length of 1.668 Gb (98.5% of predicted genome size) and a contig N50 of 35.895 Mb (L50 of 17) (supplementary table S1, Supplementary Material online). Contigs were scaffolded with Hi-C into a preliminary scaffold assembly containing 1,872 scaffolds spanning 1.668 Gb (supplementary table S1, Supplementary Material online). Scaffold PGA\_823\_1\_5040, the shortest sequence in the assembly, could not be classified by either Kraken or BlobTools (supplementary table S2, Supplementary Material online) and was therefore removed. We also assembled the complete *C. formosanum* chloroplast sequence, spanning 152,194 bp (supplementary fig. S3, Supplementary Material online). We removed 1,052 scaffolds that shared  $\geq 99\%$  sequence identity with the assembled chloroplast over  $\geq 99\%$  of the total scaffold length. We also removed 21 scaffolds totaling 652,248 bp that were identified as mitochondrial sequence by NCBI, resulting in a final, cleaned genome assembly of 798 scaffolds spanning 1.630 Gb (supplementary table S1, Supplementary Material online).

The 27 largest chromosome-scale scaffolds, representing the haploid chromosome number, contain 97.8% of the total assembly length (supplementary fig. S4, Supplementary Material online). Nine of these scaffolds consist of a single contig (supplementary fig. S4, Supplementary Material online). The chromosomes appear to be largely complete, with 24 chromosomes showing an abundance of telomeric repeats (10 at both ends and 14 at one end, fig. 1A). A Benchmarking of universal, single-copy orthologs (BUSCO) analysis likewise indicated that the genome assembly is largely complete, with 97.7% and 98.5% of BUSCO genes identified as complete in the Embryophyta and Viridiplantae datasets, respectively (supplementary fig. S5, Supplementary Material online). Not surprisingly for a polyploid genome, 91.1% and 92.9% of the genes in the Embryophyta and



**Fig. 1.**—(A) Circular representation of the nine chromosomes in each of the *B* (outer track, blue), *C* (green), and *D* (brown) sub-genomes of *C. formosum*. Tick marks represent 10 Mb. From inside to outside, tracks represent the density of *Gypsy* LTR elements, mapped *C. acuminatum* reads, 18–24J repeats, telomeric repeats, and genes. (B) Dotplot visualization of syntenic genes between the genomes of quinoa (*y*-axis) and *C. formosum* (*x*-axis). (C) Hive plot visualization of syntenic genes among the *B*, *C*, and *D* sub-genomes of *C. formosum*. For the inside to the outside, chromosomes of each sub-genome are arranged in order from 1–9. (D) Detailed view of the syntenic relationships between chromosomes 1 and 2 of the *B*, *C*, and *D* sub-genomes of *C. formosum*.

Viridiplantae datasets, respectively, were identified as duplicates in the *C. formosum* genome.

A total of 67.8% of the *C. formosum* genome assembly was annotated as repetitive by RepeatMasker (supplementary table S3, Supplementary Material online), similar to the repetitive fraction predicted by GenomeScope (65.6%, supplementary fig. S2, Supplementary Material online). The most abundant categorized repeats are *Gypsy* (22.28%) and *Copia* (7.04%) long terminal repeats (LTRs), with 28.90% unknown. To facilitate gene annotation, 3,021

million PacBio Iso-Seq reads with a minimum quality score of Q20 were produced from stems, leaves, petioles, and inflorescences, and reads were assembled into 207,192 high-quality, full-length, non-concatemer transcripts (supplementary table S4, Supplementary Material online). Gene annotation using the assembled transcripts as evidence identified 75,056 protein-coding genes. Gene density is low in centromeric regions and greater toward the chromosome ends (fig. 1A).

Comparison of gene synteny with *C. quinoa* enabled the identification of homoeologous sets of *C. formosum*

chromosomes, with three chromosomes showing clear evidence of synteny with each *C. quinoa* chromosome (fig. 1B); To assign chromosomes within each homoeologous set to the B, C, or D sub-genome, we first identified and plotted the position of the 18–24J microsatellite repeat, which was previously shown to be overrepresented in the *Chenopodium* B sub-genomes (Kolano et al. 2011). Nine *C. formosanum* chromosomes showed a clear overrepresentation of the 18–24J repeat (fig. 1A) and were assigned to the B sub-genome. To identify chromosomes belonging to the D sub-genome, we mapped and visualized the location of sequencing reads from the D-genome diploid *C. acuminatum*. Nine chromosomes showed a clear overabundance of mapped *C. acuminatum* reads (fig. 1A) and were assigned to the D sub-genome. The remaining nine chromosomes were assigned to C sub-genome. Using this approach, each homoeologous set of three chromosomes was unambiguously assigned a B, C, and D chromosome. The three sub-genomes displayed a high degree of synteny (fig. 1C), although there is clear evidence of large structural rearrangements, including a translocation between Cf1C and Cf2C (fig. 1D).

### Repeat Composition

We noted substantial differences in sub-genome size, with all B-genome chromosomes larger than all C chromosomes, which are in turn all larger than all D chromosomes (fig. 1A, [supplementary table S5, Supplementary Material](#) online). B sub-genome chromosomes comprise 45.95% of the total chromosome length, whereas C and D subgenomes comprise 31.22% and 22.82%, respectively. Despite this more than two-fold difference in sub-genome size, the number of non-repetitive bases was similar across sub-genomes (210 Mb, 197 Mb, and 174 Mb in B, C, and D, respectively), suggesting repetitive elements account for most of the size variation across sub-genomes. Repetitive elements comprised 71.3%, 60.3%, and 52.2% of the B, C, and D sub-genomes, respectively (fig. 2A). LTRs *Gypsy* elements were the most abundant classified repeats in all three sub-genomes (fig. 2A) and were particularly abundant in B sub-genome chromosomes (fig. 1A). Repeat landscape plots revealed an apparent historical expansion of *Gypsy* elements in the B subgenome centered at  $K=7$  sequence divergence (fig. 2B). The abundance of *Gypsy* elements in the B subgenome amount to ~137 Mb of additional LTR sequences compared to the next largest sub-genome (C), and accounts for ~58% of the difference in total assembly length between the two.

An equivalent analysis in the B sub-genomes of *C. quinoa* and *C. berlandieri* also showed major *Gypsy* expansions centered at  $K=7$  (figs. 2B–D), suggesting the expansion occurred in the ancestral B sub-genome prior to the origin of the allopolyploids. The *Gypsy* element peak in the B sub-genomes from all three species dates from 2.3–3 million years ago (mya) (figs. 2B–D). This timing roughly corresponds with the

estimated origin of the B genome (Mandák et al. 2018). We also identified 18 satellite DNA (satDNA) families in the *C. formosanum* assembly, 11 of which are novel, with the three sub-genomes differing in both overall satDNA abundance and family composition, in accordance with previous analyses of the B and D sub-genomes (Belyayev et al. 2020) (fig. 2E).

## Materials and Methods

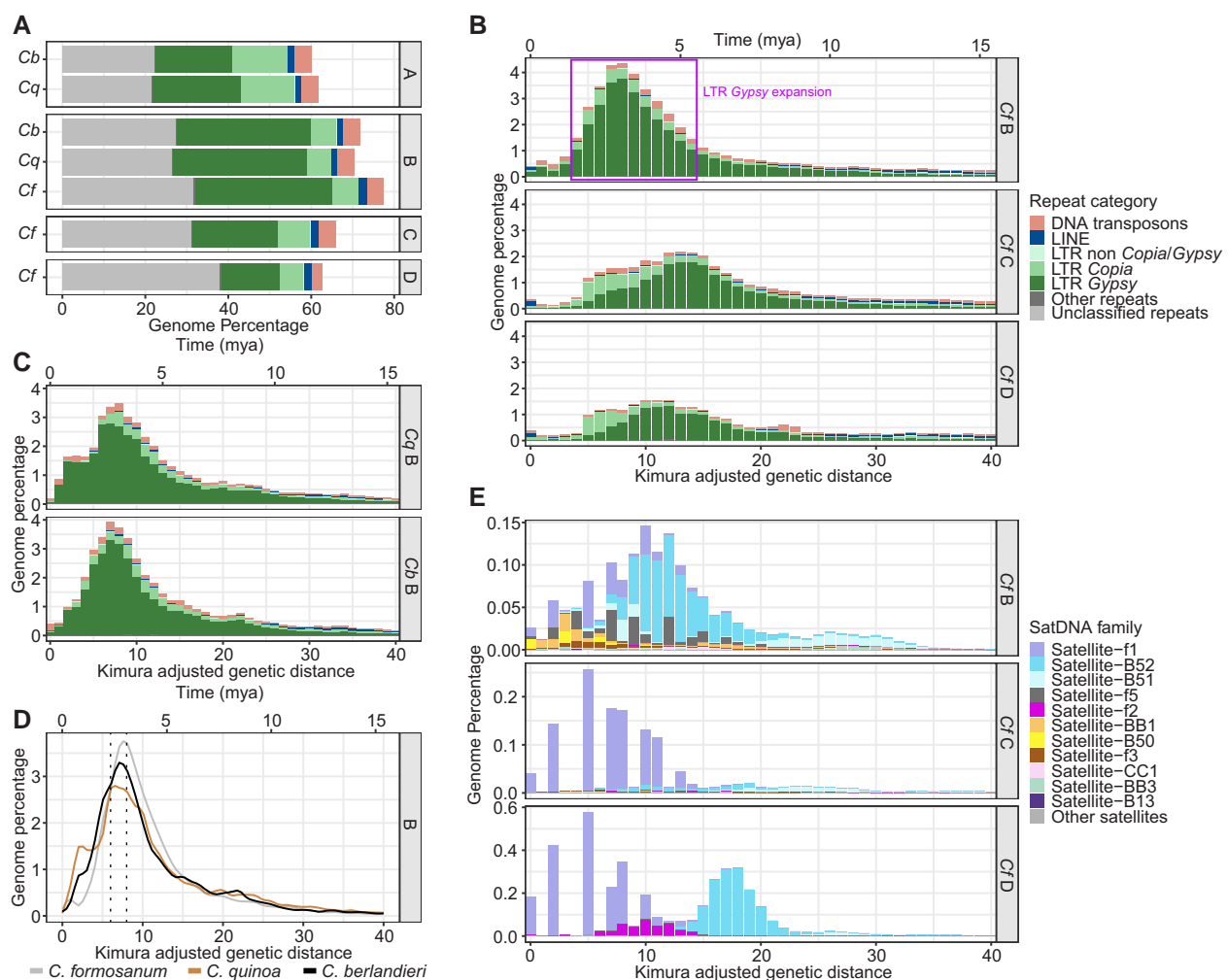
### Genome Assembly and Annotation

DNA extraction, PacBio sequencing, whole-genome and transcriptome assembly, and gene and repeat annotation were performed as previously described (Jarvis et al. 2022), with the exception that EST and protein evidence came from *C. quinoa* (NCBI BioProject PRJNA394242) and *Beta vulgaris* (BioProject PRJNA413079). All protein-coding genes shorter than 50 amino acids and without a hit to the uniprot-sprot database were removed. Hi-C scaffolding was performed by Phase Genomics. Genome size was estimated using GenomeScope (Ranallo-Benavidez et al. 2020). Genome, transcriptome, and proteome completeness was assessed using BUSCO (Simão et al. 2015). Putative contaminant sequences were identified using Kraken 2 (Wood et al. 2019) and BlobTools (Laetsch and Blaxter 2017).

To assemble the chloroplast, PacBio reads were mapped to the *C. quinoa* chloroplast genome (GenBank acc: MK159176) using minimap2 (v.2.24) (Li 2018) with the “map-hifi” argument. Samtools v1.9 (Danecek et al. 2021) was used to identify mapped reads with a minimum query length (mlen) >8000, query value (qval) >40, and GC content between 32–40%. These reads were included in a final assembly of the chloroplast genome using HiCanu v2.1 (Nurk et al. 2020) using default parameters, and the assembly was annotated using GeSeq (Tillich et al. 2017). Scaffolds from the genome assembly with sequence homology to the assembled chloroplast were identified by performing a nucleotide BLAST search of the genome assembly against the chloroplast sequencing using default parameters. All scaffolds that shared ≥99% sequence identity with the assembled chloroplast over ≥99% of the total scaffold length were removed from the genome assembly.

### Genome Visualization and Comparison

The positions of *Gypsy* LTRs were determined using RepeatModeler v2.0.1 (Flynn et al. 2020) and RepeatMasker v4.1.0 (Smit et al. 2013), as described below. The positions of 18–24J repeats and telomeric repeats were determined using BLAST, as previously described (Jarvis et al. 2017). The positions of mapped sequencing reads from *C. acuminatum* were determined by generating PacBio Hifi reads from one cell using DNA extracted from roots of a single *C. acuminatum* plant and then mapping the reads to the *C. formosanum* genome assembly using minimap2 (Li 2018) with the



**FIG. 2.**—(A) Abundance of repeats in the major repetitive element categories for the A, B, C, and D sub-genomes in *C. quinoa* (Cq), *C. berlandieri* (Cb), and *C. formosanum* (Cf). (B) Repetitive element landscape plots for *C. formosanum* sub-genomes. The y-axis shows TE abundance as a percentage of the genome for major repeat categories. The x-axis shows both sequence divergence (CpG adjusted Kimura distance) relative to consensus sequences for TE superfamilies (below) and an estimate of evolutionary time calculated using LTR sequence evolution rate estimates in rice. (C) Repeat landscapes for the B sub-genome in quinoa and *C. berlandieri*. (D) Repeat landscape isolating LTR *Gypsy* elements in *C. formosanum*, *C. quinoa*, and *C. berlandieri*. (E) Repeat landscape of satellite DNAs in sub-genomes of *C. formosanum*.

“-ax map-pb” option. The densities of *Gypsy* LTRs, mapped *C. acuminatum* reads, 18–24J repeats, telomeric repeats, and genes were visualized in 500 kb windows using Circa (<http://omgenomics.com/circa>).

Syntenic genes within the *C. formosanum* genome and between the *C. formosanum* and *C. quinoa* genomes were identified using the CoGe SynMap (Lyons et al. 2008) tool and MCScanX (Wang et al. 2012) and visualized using SynVisio (Bandi and Gutwin).

### Repetitive Element Identification and Annotation

Repetitive elements were identified *de novo* in the *C. formosanum* assembly using RepeatModeler v2.0.1 (Flynn et al. 2020) and annotated in each *C. formosanum* sub-genome separately with RepeatMasker v4.1.0 (Smit et al.

2013). Prior to sub-genome annotation we merged the custom repeat library from RepeatModeler with RepeatMasker’s internal “Viridiplantae” library and used the resulting combined library for annotation of sub-genomes with the search engine set to “ncbi” and using the -xsmall option. For comparative analyses that included *C. berlandieri* (genomeevolution.org, CoGe id62441) and *C. quinoa* (CoGe id60716) assemblies, the same repeat annotation methods were used. We summarized repeat abundance in each sub-genome by parsing output from repeat annotation software using custom scripts and generating plots in R v3.5.1 (Team 2018).

### satDNA Identification and Annotation

To identify novel satDNA families, we used RepeatMasker to mask previously described (Belyayev et al. 2020) satDNAs in

the *C. formosanum* assembly. The repeat masked *C. formosanum* sequence was then used as input for TRF v 4.09.1 (Benson 1999). TRF output files were parsed using custom shell scripts. We selected candidate satDNA sequences based on their abundance (“Copy\_number” >50) and monomer length (“Consensus\_size” >12 bp). Monomer consensus sequences of the selected satDNA candidates were checked and apparent microsatellite sequences were excluded from further analyses. The remaining candidates were sorted according to the size of monomer consensus sequences. Monomers with similar sizes were aligned together using MAFFT (Kato et al. 2002) as implemented in Geneious 11.1.5 (Biomatters Ltd, Auckland, New Zealand), to search for sequence homology among them. In case of significant homology, the monomers consensus were grouped together and a new consensus sequence was created. We then aligned the newly described satDNA families to each other and to those previously described (Belyayev et al. 2020). Both the newly described as well as the previously identified satDNA monomer consensus were mapped onto the *C. formosanum* assembly using Geneious’ “Live Annotate and Predict” function. To account for sequence variation between the different arrays of the same satDNA family in the genome we performed three mappings with similarity thresholds of 90%, 80% and 75%.

### Repetitive Element Landscapes

To visualize the different composition of transposable elements and satDNAs among the three *C. formosanum* sub-genomes, we generated repeat landscape plots for *Chenopodium* species sub-genomes. We used the calcDivergenceFromAlign.pl built-in tool of RepeatMasker to obtain a histogram of the Kimura 2-Parameter divergence for each element. Next, we transformed the abundance values to express them as genome proportions by dividing the number of aligned nucleotides by the total number of nucleotides in the genome assembly. The resulting histograms (referred to as repeat landscapes) were plotted in R, using ggplot2. To get a rough time scale of transposable element evolution, we transformed to time the divergence values of the x-axis in the repeat landscapes using the equation: time = divergence/2r, where r =  $1.3 \times 10^{-8}$  mutations per site per year, which is the previously estimated rate of LTR element sequence evolution in rice (Ma and Bennetzen 2004).

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

### Acknowledgments

This work was supported by funding from a BYU Life Sciences start-up grant to D.E.J. Ed Wilcox and the BYU DNASC (RRID: SCR\_017781) provided sequencing support.

### Data Availability

The genome assembly of *C. formosanum* and sequence data for *C. formosanum* and *C. acuminatum* have been deposited at DDBJ/ENA/GenBank under the accession JAMXMD000000000, BioProject ID PRJNA840947 (<http://www.ncbi.nlm.nih.gov/bioproject/840947>); the version described in this paper is version JAMXMD010000000. The assembly and annotation for *C. formosanum* are also available at CoGe (<https://genomevolution.org/coge/>) id64369.

### Literature Cited

- Bandi V, Gutwin C. 2020. Interactive exploration of genomic conservation. In: *Proceedings of the 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020 (GI'20)*. Waterloo, CAN: Canadian Human-Computer Communications Society.
- Belyayev A, et al. 2020. The major satellite DNA families of the diploid *Chenopodium album* aggregate species: arguments for and against the “library hypothesis.” *PLoS ONE*. 15:e0241206.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27:573–580.
- Chen SY, Chu CC, Chyau CC, Yang JW, Duh PD. 2019. Djulis (*Chenopodium formosanum*) and its bioactive compounds affect vasodilation, angiotensin converting enzyme activity, and hypertension. *Food Biosci.* 32:100469.
- Chio EH, et al. 2013. Toxicity and repellence of Taiwanese indigenous djulis, *Chenopodium formosanum*, against *Aedes albopictus* (Diptera: Culicidae) and *Forcipomyia taiwana* (Diptera: Ceratopogonidae). *J Pest Sci.* 86:705–712.
- Chu CC, et al. 2016. Protective effect of djulis (*Chenopodium formosanum*) and its bioactive compounds against carbon tetrachloride-induced liver injury, *in vivo*. *J Funct Food.* 26:585–597.
- Chuang KJ, Chen ZJ, Cheng CL, Hong GB. 2018. Investigation of the antioxidant capacity, insecticidal ability and oxidation stability of *Chenopodium formosanum* seed extract. *Int J Mol Sci.* 19:2726.
- Chyau CC, Chu CC, Chen SY, Duh PD. 2018. The inhibitory effects of Djulis (*Chenopodium formosanum*) and its bioactive compounds on adipogenesis in 3T3-L1 adipocytes. *Molecules* 23:1780.
- Danecek P, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008.
- Flynn JM, et al. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc National Acad Sci.* 117: 9451–9457.
- Huai KHY, Pei SJ. 2000. Wild plants in the diet of arborchin mongol herdsmen in inner mongolia. *Econ Bot.* 54:528–536.
- Huang CY, Chu YL, Sridhar K, Tsai PJ. 2019. Analysis and determination of phytosterols and triterpenes in different inbred lines of Djulis (*Chenopodium formosanum* Koidz.) hull: a potential source of novel bioactive ingredients. *Food Chem.* 297:124948.
- Jarvis DE, et al. 2017. The genome of *Chenopodium quinoa*. *Nature* 542:307–312.
- Jarvis DE, et al. 2022. Chromosome-scale genome assembly of *Gilia yorkii* enables genetic mapping of floral traits in an interspecies cross. *Genome Biol Evol.* 14:evac017.
- Kato K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast fourier transform. *Nucleic Acids Res.* 30:3059–3066.
- Kolano B, et al. 2011. Chromosomal localization of two novel repetitive sequences isolated from the *Chenopodium quinoa* wild genome. *Genome* 54:710–717.

- Kolano B, et al. 2019. Parental origin and genome evolution of several eurasian hexaploid species of *Chenopodium* (*Chenopodiaceae*). *Phytotaxa* 392:163–185.
- Krak K, et al. 2016. Allopolyploid origin of *Chenopodium album* s. str. (*Chenopodiaceae*): a molecular and cytogenetic insight. *PLoS ONE* 11:e0161063.
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000research* 6:1287.
- Lee CW, Chen HJ, Xie GR, Shih CK. 2019. Djulis (*Chenopodium formosanum*) prevents colon carcinogenesis via regulating antioxidative and apoptotic pathways in rats. *Nutrients* 11:2168.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.
- Li PH, et al. 2021. Functionality of djulis (*Chenopodium formosanum*) by-products and in vivo anti-diabetes effect in type 2 diabetes mellitus patients. *Biology* 10:160.
- Lyons E, Pedersen B, Kane J, Freeling M. 2008. The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Trop Plant Biol*. 1:181–190.
- Ma J, Bennetzen JL. 2004. Rapid recent growth and divergence of rice nuclear genomes. *P Natl Acad Sci U S A*. 101:12404–12410.
- Mandák B, et al. 2016. How genome size variation is linked with evolution within *Chenopodium* sensu lato. *Perspect Plant Ecol Evol Syst*. 23:18–32.
- Mandák B, et al. 2018. Hybridization and polyploidization within the *Chenopodium album* aggregate analysed by means of cytological and molecular markers. *Mol Phylogenet Evol*. 129:189–201.
- Mangelson H, et al. 2019. The genome of *Chenopodium pallidicaule*: an emerging andean super grain. *Appl Plant Sci*. 7:e11300.
- Nurk S, et al. 2020. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res*. 30:1291–1305.
- Pandey S, Gupta RK. 2014. Screening of nutritional, phytochemical, antioxidant and antibacterial activity of *Chenopodium album* (Bathua). *J Pharmacogn Phytochem*. 3:1–9.
- Partap T, Kapoor P. 1985. The himalayan grain chenopods. I. Distribution and ethnobotany. *Agric Ecosyst Environ*. 14:185–199.
- Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 11:1432.
- Repo-Carrasco R, Espinoza C, Jacobsen SE. 2003. Nutritional value and use of the andean crops quinoa (*Chenopodium quinoa*) and kañiwa (*Chenopodium pallidicaule*). *Food Rev Int*. 19:179–189.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212.
- Smit AFA, Hubley R, Green P. 2013. *RepeatMasker Open-4.0*. <http://www.repeatmasker.org>.
- Štorchová H, Drabešová J, Cháb D, Kolář J, Jellen EN. 2015. The introns in FLOWERING LOCUS T-LIKE (FTL) genes are useful markers for tracking paternity in tetraploid *Chenopodium quinoa* willd. *Genet Resour Crop Evol*. 62:913–925.
- Team RC. 2018. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Tillich M, et al. 2017. GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 45:W6–W11.
- Tsai PJ, Chen YS, Sheu CH, Chen CY. 2011. Effect of nanogrinding on the pigment and bioactivity of djulis (*Chenopodium formosanum* Koidz.). *J Agr Food Chem*. 59:1814–1820.
- Tsai PJ, Sheu CH, Wu PH, Sun YF. 2010. Thermal and pH stability of betacyanin pigment of djulis (*Chenopodium formosanum*) in Taiwan and their relation to antioxidant activity. *J Agr Food Chem*. 58:1020–1025.
- Vega-Gálvez A, et al. 2010. Nutrition facts and functional potential of quinoa (*Chenopodium quinoa* willd.), an ancient Andean grain: a review. *J Sci Food Agric*. 90:2541–2547.
- Walsh BM, Adhikary D, Maughan PJ, Emshwiller E, Jellen EN. 2015. *Chenopodium* polyploidy inferences from *Salt Overly Sensitive 1* (*SOS1*) data. *Am J Bot*. 102:533–543.
- Wang Y et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 40:e49–e49.
- Wilson HD, Heiser CB Jr. 1979. The origin and evolutionary relationships of “huauzontle” (*Chenopodium nuttalliae* Safford), domesticated chenopod of Mexico. *Am J Bot*. 66:198–206.
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 20:257.
- Yadav SK, Sehgal S. 2002. In vitro and in vivo availability of iron from Bathua (*Chenopodium album*) and spinach (*Spinacia oleracea*) leaves. *J Food Sci Technol*. 39:42–46.
- Yang BY, et al. 2014. Potential bioethanol production from Taiwanese chenopods (*Chenopodium formosanum*). *Energy* 76:59–65.

Associate editor: Dr. Maud Tenaillon