G3, 2024, 14(6), jkae080
https://doi.org/10.1093/g3journal/jkae080
Advance Access Publication Date: 12 April 2024
Genome Report

A de novo genome assembly of *Solanum bulbocastanum* Dun., a Mexican diploid species reproductively isolated from the A-genome species, including cultivated potatoes

Awie J. Hosaka [1,2], Rena Sanetomo [3], Kazuyoshi Hosaka [3,*]

[1]Nihon BioData Corporation, Takatsu, Kawasaki, Kanagawa 213-0012, Japan
[2]Kihara Institute for Biological Research, Yokohama City University, Yokohama 244-0813, Japan
[3]Potato Germplasm Enhancement Laboratory, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan

*Corresponding author: Potato Germplasm Enhancement Laboratory, Obihiro University of Agriculture and Veterinary Medicine, Obihiro, Hokkaido 080-8555, Japan.
Email: spudman@obihiro.ac.jp
The Hawkes (1990) classification system is tentatively adopted throughout the text.

Potato and its wild relatives are distributed mainly in the Mexican highlands and central Andes of South America. The South American A-genome species, including cultivated potatoes, are reproductively isolated from Mexican diploid species. Whole-genome sequencing has disclosed genome structure and similarity, mostly in cultivated potatoes and their closely related species. In this study, we generated a chromosome-scale assembly of the genome of a Mexican diploid species, *Solanum bulbocastanum* Dun., using PacBio long-read sequencing, optical mapping, and Hi-C scaffolding technologies. The final sequence assembly consisted of 737.9 Mb, among which 647.0 Mb were anchored to the 12 chromosomes. Compared with chromosome-scale assemblies of *S. lycopersicum* (tomato), *S. etuberosum* (non-tuber-bearing species with E-genome), *S. verrucosum*, *S. chacoense*, *S. multidissectum*, and *S. phureja* (all four are A-genome species), the *S. bulbocastnum* genome was the shortest. It contained fewer transposable elements (56.2%) than A-genome species. A cluster analysis was performed based on pairwise ratios of syntenic regions among the seven chromosome-scale assemblies, showing that the A-genome species were first clustered as a distinct group. Then, this group was clustered with *S. bulbocastanum*. Sequence similarity in 1,624 single-copy orthologous gene groups among 36 *Solanum* species and clones separated *S. bulbocastanum* as a specific group, including other Mexican diploid species, from the A-genome species. Therefore, the *S. bulbocastanum* genome differs in genome structure and gene sequences from the A-genome species. These findings provide important insights into understanding and utilizing the genetic diversity of *S. bulbocastanum* and the other Mexican diploid species in potato breeding.

Keywords: *Solanum bulbocastanum*; Mexican wild diploid potato; de novo genome assembly; comparative genomics

Introduction

Modern potato cultivars (*Solanum tuberosum* L., $2n = 4x = 48$), together with Andean native cultivars and over 100 closely related wild species, form section *Petota* in the genus *Solanum* (Hawkes 1990; Spooner *et al.* 2014). Hawkes (1990) further separated non-tuber-bearing and tuber-bearing species in the section *Petota* into subsections *Estolonifera* and *Potatoe*, respectively. Recent molecular phylogenetics supported the two subsections as Etuberosum and Petota clades, respectively, and, together with Tomato and some other minor clades, formed the large Potato clade (Gagnon *et al.* 2022). Their centers of diversity are found in the Mexican highlands and the central Andes (Hawkes 1990; Spooner *et al.* 2014). The diploid species of the two centers of diversity are strictly isolated from each other by reproductive barriers (Hawkes 1958). Based on the meiotic chromosome pairing of interspecific hybrids, the A-genome was assigned to the South American species, including cultivated potatoes (Matsubayashi 1991). These species are referred to as A-genome species hereafter. Since sexual hybrids between Mexican diploid species and A-genome species are extremely difficult to obtain, the genome affinity of Mexican diploid species has long been argued (Matsubayashi 1991; Pendinen *et al.* 2008).

Mexican diploid species are grouped into four taxonomic series (*Morelliformia*, *Bulbocastana*, *Pinnatisecta*, and *Polyadenia*) (Hawkes 1990). Meiotic chromosome behaviors in interspecific hybrids between Mexican diploid species from different taxonomic series indicated that their genomes are somewhat more diverged from each other than those between A-genome species from the South American series (Magoon *et al.* 1958; Marks 1968; Matsubayashi and Misoo 1977). Molecular analyses revealed clear genome differentiation of Mexican diploid species from the A-genome species (Hosaka *et al.* 1984; Spooner *et al.* 1991, 2008; Pendinen *et al.* 2008; Rodríguez and Spooner 2009). Recent advancement in nucleotide sequencing technologies allows comparing whole-genome sequences of the tuber-bearing *Solanum* species (Aversano *et al.* 2015; Leisner *et al.* 2018; Zhou *et al.* 2020; Tiwari *et al.* 2021; Yan *et al.* 2021; Hoopes *et al.* 2022; Hosaka *et al.* 2022; Sun *et al.* 2022), which revealed a monophyletic origin of Mexican diploid species sharing the common ancestor with A-genome species (Li *et al.* 2018; Huang *et al.* 2019; Gagnon *et al.* 2022; Tang *et al.* 2022).

Received on 04 February 2024; accepted on 06 April 2024

Solanum bulbocastanum Dun. ($2n = 2x = 24$), a representative species of the series *Bulbocastana*, is morphologically distinct with simple leaves, without lateral leaflets, from most of the other *Petota* species and well known for its resistance to late blight (caused by *Phytophthora infestans*) and Columbia root-knot nematode (*Meloidogyne chitwoodi*) (Graham *et al.* 1959; Toxopeus 1964; Bali *et al.* 2022). The late blight resistance genes *Rpi-blb1/RB*, *Rpi-blb2*, *Rpi-blb3*, and *Rpi-bt1* have been cloned from this species (Song *et al.* 2003; Lokossou *et al.* 2010). Although direct crosses with cultivated potatoes are impossible, breeders transferred these resistance genes by bridging crosses via *S. verrucosum* Schlechtendal and by somatic hybridization (Hermsen and Ramanna 1976; Austin *et al.* 1993; Brown *et al.* 1995; Jansky and Hamernik 2009; Tiwari *et al.* 2018). Thus, *S. bulbocastanum* is an important species to investigate from genome evolution and potato breeding standpoint.

The *S. bulbocastanum* PG6241 genome has been sequenced to the level of contigs using a PacBio long-read sequencing technology (Tang *et al.* 2022). In this study, we generated a chromosome-scale assembly of the *S. bulbocastanum* genome using PacBio long-read sequencing and, in addition, Optical mapping and Hi-C scaffolding technologies. A naturally chromosome-doubled monohaploid clone of *S. bulbocastanum* ($2n = 2x = 24$) was used to reduce the complexity caused by the heterozygous nature of the species. This is the Mexican diploid species' first de novo chromosome-scale assembly except for *S. verrucosum*, the only diploid A-genome species distributed in Mexico (Hosaka *et al.* 2022). The chromosome-scale assemblies facilitated to reveal structural differences between the *S. bulbocastanum* genome and the previously reported *S. phureja*, *S. multidissectum*, *S. verrucosum*, *S. chacoense*, *S. etuberosum*, and *S. lycopersicum* genomes (Leisner *et al.* 2018; Hosaka *et al.* 2022; Tang *et al.* 2022; Zhou *et al.* 2022; Yang *et al.* 2023) in the Potato, Etuberosum, and Tomato clades.

Materials and methods

Plant material

A naturally doubled monohaploid clone of *S. bulbocastanum* (11H21, available as PI 666967 from the US Potato Genebank at Sturgeon Bay, WI, USA) that was derived from anther culture (Irikura and Sakaguchi 1972) and maintained in vitro in our laboratory (Sanetomo and Hosaka 2021), was used for sequencing.

DNA extraction, genome sequencing, and initial assembly

High-molecular-weight DNA was extracted, qualified, and used for HiFi read sequencing using the PacBio Sequel lle system (PacBio) as previously described (Hosaka *et al.* 2022). The bam files were converted to a FASTQ file using BAM2fastx 1.3.1 (PacBio) and used for genome assembly with the Hifiasm 0.16.1-r375 assembler (Cheng *et al.* 2021). The "-l 0" option was used to turn off purge haplotigs since the plant was a completely homozygous diploid clone.

Optical mapping

Genomic DNA was extracted from young leaves using the Plant DNA Isolation Kit (Bionano Genomics, San Diego, CA, USA). The isolated DNA was labeled with Direct Labeling Enzyme 1 using the DLS DNA Labeling Kit (Bionano Genomics, San Diego, CA). The labeled DNA was loaded onto a Saphyr chip and run on the Saphyr Optical Genome Mapping (OGM) Instrument (Bionano Genomics). The output molecules were assembled and then merged with the contigs to generate hybrid scaffold sequences using OGM-specific pipelines Bionano Access and Solve (versions 1.7.1.1 and 3.7_03302022_283, respectively) with default parameters. AS ONE Corp. (Osaka, Japan) collected and analyzed data.

Hi-C sequencing and scaffolding

To construct the Hi-C library, proximity ligation and library amplification were performed by using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA, USA) and the Kapa Hyper Prep kit (KAPA Biosystems, Cape Town, South Africa), respectively. The sampled leaves were ground in liquid nitrogen, resuspended with 5 ml of PBS buffer with 1% formaldehyde, and rotated for 10 min. The sample was centrifuged at 5000×*g* for 5 min, and the supernatant was discarded. The precipitate was resuspended with 5 ml of Nuclei Isolation Buffer (10 mM HEPES, 5 mM KCl, 5 mM $MgCl_2$, 5 mM EDTA, 1 M sucrose, and 0.2% Triton X-100). The sample was filtrated through a 40-μm cell strainer. The filtrated sample was centrifuged at 5000×*g* for 5 min, and the supernatant was discarded. The precipitate was resuspended with 1 ml of Nuclei Isolation Buffer, centrifuged at 5000×*g* for 5 min, and the supernatant was discarded. The residue was resuspended with 500 μl Nuclei Isolation Buffer and layered on 500 μl Nuclei Separation Buffer (10 mM HEPES, 5 mM KCl, 5 mM $MgCl_2$, 5 mM EDTA, 1 M sucrose, and 15% Percoll). The sample was centrifuged at 3000×*g* for 5 min, and the supernatant was discarded. The precipitate was used for DNase treatment and proximity ligation steps. The ligation product was eluted with 50 μl of TE buffer (10 mM Tris–Cl buffer pH 8.0 and 1 mM EDTA pH 8.0). Ten microliter of Dynabeads MyOne Streptavidine C1 beads was rinsed in 150 μl of TWB buffer and resuspended with 100 μl of 2× NTB buffer. Fifty microliter of the ligation product and 50 μl of TE buffer were mixed with the beads and rotated for 15 min. The beads were washed once with 500 μl of LWB, twice with 500 μl of NWB, once with 200 μl of Wash Buffer, and resuspended with 25 μl of sterile water. The resuspended beads were mixed with 3.5 μl of End Repair & A-tailing Buffer and 1.5 μl of End Repair & A-tailing Enzyme Mix and incubated at 20 °C for 30 min and then, at 65 °C for 30 min. The reaction solution was mixed with 15 μl of Ligation Buffer, 5 μl of DNA Ligase, and 2.5 μl of Dual-index adapter and incubated at 20 °C for 15 min. The beads were washed once with 500 μl of LWB, twice with 500 μl of NWB, once with 200 μl of Wash Buffer, and resuspended with 10 μl of TE buffer. Then, 12 μl of 2× KAPA HiFi Hot Start Ready Mix and 2.5 μl of 2× Library Amplification Primer Mix were added to the solution. Polymerase chain reaction (PCR) was performed with the following conditions: 98 °C for 45 s, then 16 cycles of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s, and a final extension at 72 °C for 1 min. The supernatant of the PCR product was purified and dual size-selected using 0.5× and 0.3× volumes of AMPure XP beads (Beckman, Indianapolis, IN, USA) and eluted with 12 μl of TE buffer.

The prepared library was sequenced on the NovaSeq 6000 (Illumina, San Diego, CA) platform. The reads were aligned to the scaffolded contigs using Juicer 1.6 (Durand *et al.* 2016). Since DNase I was used to digest fixed nucleosomes, the "-s none -y none" option was specified. The generated contact maps were then used for scaffolding with a 3D-DNA pipeline (Dudchenko *et al.* 2017) with the "–rounds 0" parameter. The scaffolds were reviewed and manually corrected using JuiceBox 1.11.08 (https://github.com/aidenlab/Juicebox). Identities and directions of the corrected scaffolds were determined based on the alignment to the *S. phureja* Juz. & Bukasov (described as *S. tuberosum* Andigenum group in Spooner *et al.* 2014) DM v8.1 reference genome (Yang *et al.* 2023) using RagTag (Alonge *et al.* 2022).

Annotation

To construct homology-based gene models, annotation results of *S. bulbocastanum* PG6241 (Tang *et al.* 2022) were lifted to our assembly using Liftoff (Shumate and Salzberg 2021) with the "-polish -copies" option. To construct RNA-seq-based gene models, publicly available datasets of *S. bulbocastanum* PG6241 in Sequence Read Archive (SRA) database (SRR15560132, SRR15560142, SRR15560143, SRR15560145, SRR15560146, SRR15560147, SRR15560148, SRR15560149, SRR15560150, SRR15560151, SRR15560152, and SRR15560153) were obtained using Sratools (https://github.com/ncbi/sra-tools). After filtering the low-quality reads using Fastp 0.23.2 (Chen *et al.* 2018), the reads were aligned to the assembled genome using Hisat2 2.2.1 (Kim *et al.* 2019). Putative transcripts were identified in each *S. bulbocastanum* dataset, and the resulting General Transfer Format (GTF) files were merged using Stringtie 2.2.1 (Kovaka *et al.* 2019) and used to search putative open-reading-frames by TransDecoder.LongOrfs (Haas BJ, https://github.com/TransDecoder/TransDecoder) with default parameters. To improve the prediction accuracy of protein-coding regions, we performed Blastp search of the putative amino acids against *S. bulbocastanum* PG6241 protein sequences with the "-max_target_seqs 1 -outfmt 6 -evalue 1e-5" parameter. Protein-coding regions were predicted using TransDecoder.Predict with the Blastp result and the "–single_best_only" option. To construct the annotation file against the assembled genome, the "cdna_alignment_orf_to_genome_orf.pl" command implemented in TransDecoder (Haas BJ, https://github.com/TransDecoder/TransDecoder) was used. The two sets of gene models were merged and provided to the MAKER annotation pipeline (Cantarel *et al.* 2008). Gene models of tomato pre-trained in AUGUSTUS (Stanke *et al.* 2004) were also included. Transposable elements (TEs) were identified using EDTA 2.0 (Ou *et al.* 2019). To filter the gene-related sequences during the detection of TEs, the coding sequences (CDS) were provided. The distribution of genes and TEs was visualized as a circular heatmap generated by Circos (Krzywinski *et al.* 2009).

Assessment of assembly and annotation quality

The contig number and length distribution of the initial assembly, the assembly after Optical mapping, and the final assembly were obtained by the "stat" command of Seqkit 0.15.0 (Shen *et al.* 2016). L50 statistics and gap content were calculated by the "fx2tab" command of Seqkit with "-nlH -B N" options. The assembly and annotation completeness were assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO; Simão *et al.* 2015) against the genomic sequences and annotated protein sequences available in the Solanales odb10 database (Kriventseva *et al.* 2019).

Comparison of genome structures

The genome structure of *S. bulbocastanum* was compared with those of six chromosome-scale assemblies: *S. lycopersicum* L. (cv. Heinz 1706 SL5.0; Zhou *et al.* 2022), *S. etuberosum* Lindl. (PG0019 [PI 558302]; Tang *et al.* 2022), *S. verrucosum* (11H23 [PI 666968]; Hosaka *et al.* 2022), *S. chacoense* Bitter (M6 v5; Leisner *et al.* 2018, http://spuddb.uga.edu/M6_v5_0_download.shtml), *S. multidissectum* Hawkes [PG5068, (PI 458379, described as *S. candolleanum* Berthault in Spooner *et al.* 2014); Tang *et al.* 2022], and *S. phureja* (DM v8.1, Yang *et al.* 2023). Chromosomal sequences of these seven genomes were used for pairwise alignment by Minimap2 (Li 2018) and outputted as SAM and PAF files by defining "-ax asm5 -eqx" and "-x ams5" options, respectively. The SAM files were converted to the bam files using SAMtools 1.61.1 (Li *et al.* 2009). Syntenic regions and structural variations were detected using SyRI (Goel *et al.* 2019) with the "-k -F B" option. The results were visualized by Plotsr 0.5.4 (Goel and Schneeberger 2022). Pairwise ratios of syntenic regions were calculated by dividing the length of syntenic areas by the total chromosome length of each species. The PAF files were used for generating dot-plots by D-GENIES (Cabanettes and Klopp 2018).

Phylogenetic inference

Detection of orthogroups and phylogenetic inference among tuber-bearing *Solanum* species were assessed using OrthoFinder (Emms and Kelly 2015, 2019). In addition to the seven species used for the genome structure comparison, 21 wild species and seven clones of three cultivated species (Tang *et al.* 2022) and *S. melongena* Wall. (Barchi *et al.* 2021) as an outgroup species were used for the analysis. The longest protein sequences of their annotated gene models, located on the chromosomes or primary contigs, were extracted and used as input datasets for OrthoFinder. The phylogenetic relationship was first inferred by the Species Tree Inference from All Genes (STAG) method (Emms and Kelly 2018) with default parameters. Secondly, a multiple sequence alignment (MSA) procedure was performed with "-M msa" and "-A mafft" options, and the single-copy ortholog groups were detected and aligned using MAFFT (Katoh and Standley 2013). The tree was inferred by FastTree 2 (Price *et al.* 2010) implemented in the Orthofinder pipeline. IQ-TREE 2 (Minh *et al.* 2020) was also used for maximum likelihood inference of phylogenetic relationships using "-m MFP" and "-bb 1000" options to automatically calculate the best-fit amino-acid substitution model and to replicate 1,000 times for an ultrafast bootstrap approximation.

Characterization of unanchored contigs

Unanchored contigs were aligned with each other and with the assembled genome using Minimap2 with the "-x asm5 -P" option. HiFi reads were aligned to the assembled genome using Minimap2 with the "-ax map-hifi" option. Together with genes and TEs, the coverages of unanchored contigs and HiFi reads were visualized in a circular heatmap generated by Circos. TEs in the regions where unanchored contigs were densely aligned were counted by BEDTools 2.30.0 (Quinlan and Hall 2010) with the intersect command. These regions were aligned to the contigs of 23 wild species and six clones of three cultivated species (Tang *et al.* 2022) and to the seven chromosome-scale assemblies using Minimap2 with the "-ax map-hifi" option. The regions with >80% homology were counted.

Results and discussion

Genome assembly

We obtained 72.7 Gb (90× coverage, assuming 800 Mb/genome) of HiFi reads using a PacBio Sequel IIe system with an N50 read size of 20.9 kb and an average read size of 21.1 kb. The initial Hifiasm assembly generated 1,586 contigs (N50 = 50.5 Mb) (Table 1). Chromosomes 2, 3, and 5–11 were already composed of single contigs. To improve contig contiguity, we performed Optical mapping. Of the 200.0 Gb (N50 = 345.9 kb) data generated, 155.0 Gb (N50 = 340.1 kb) were filtered and used for de novo assembly, resulting in 94 optical maps with a total length of 638.3 Mb (N50 = 23.3 Mb). By aligning these optical maps to the primary contigs, eight conflicts in the optical maps and 13 disputes in the primary contigs were found and corrected. The optical mapping might be ineffective since the primary contigs were highly contiguous. As a result, 22 hybrid scaffolds with a total length of 653.1 Mb

Table 1. Assembly statistics.

	Primary contigs with PacBio reads	Merged contigs after Optical mapping	Final contigs after Hi-C sequencing
Number of contigs	1,586	1,599	1,639
Total size, bp	737,832,402	737,833,609	737,908,709
Longest size, bp	61,881,511	61,881,511	72,187,831
Mean size, bp	465,215.9	461,434.4	450,218.9
N50 size, bp	50,507,639	44,207,125	52,534,229
L50, number	7	7	7
Gap (%)	-	0.000	0.010

(N50 = 50.5 Mb) were generated. These hybrid scaffolds and 1,577 unanchored contigs totaling 84.7 Mb were error-corrected and scaffolded with Omni-C read pairs using Juicer and a 3D-DNA pipeline (Supplementary Fig. 1, Table 1). Chromosome identities and directions were determined based on the reference sequences of DM v8.1. The final sequence assembly consisted of 737.9 Mb, among which 647.0 Mb were anchored to the 12 chromosomes; 527 and 17 contigs (28.8 and 1.0 Mb, respectively) were those of chloroplast and mitochondrial genomes. The Hifiasm assembly forms many small contigs for nonlinear genomes, such as chloroplast and mitochondrial genomes (Hosaka *et al.* 2022; Sharma *et al.* 2022; Sun *et al.* 2022). The remaining 1,083 contigs (61.0 Mb) were unanchored (Supplementary Table 1).

Genome annotation

Based on homology, 41,251 out of 56,032 gene models of *S. bulbocastanum* PG6241 were lifted using LiftOff. Using publicly available RNA-seq datasets of *S. bulbocastanum* PG6241, 22,832 protein-coding genes were predicted by TransDecoder. Both gene models and an ab initio pre-trained gene prediction model were used for final annotation by MAKER, and 35,523 genes were predicted. The number of predicted genes was slightly lower than that of DM v8.1 (40,155 genes). TEs were identified by EDTA using CDS of the predicted genes to filter gene-related sequences. As shown in Supplementary Table 2, TEs comprised 56.2% of the genome, which was lower than those of the pangenome of *Solanum* section *Petota* (75.5%; Bozan *et al.* 2023) and *S. phureja* DM v8.1 (60.3%; Yang *et al.* 2023). A significantly higher percentage of TEs was reported for the genome with in vitro propagation in its history (Bozan *et al.* 2023). The presently used *S. bulbocastanum* and previously used *S. verrucosum* clones have been maintained in vitro for almost half a century (Sanetomo and Hosaka 2021). Nevertheless, the *S. bulbocastanum* genome (56.2%) had a lower TE content than the *S. verrucosum* genome (61.8%; Hosaka *et al.* 2022). Among TEs, long terminal repeat (LTR)-type retrotransposons accounted for 26.4%, and terminal inverted repeat (TIR)-type transposons accounted for 18.2%, in which Gypsy (16.5%) and Mutator (8.8%) families were the most abundant, as previously reported in other *Solanum* species (Aversano *et al.* 2015; Gaiero *et al.* 2019; Hosmani *et al.* 2019; Bozan *et al.* 2023). However, the content of Gypsy was much lower in *S. bulbocastanum* than in *S. verrucosum* (24.2%), which is consistent with a finding that the Gypsy content in Mexican diploid species is lower than that in the A-genome species (Gaiero *et al.* 2019; Bozan *et al.* 2023). The genes were densely distributed in subtelomeric regions (Fig. 1a). Some class II transposons, such as Tc1_Mariner and Miniature Inverted-repeat Transposable Elements (MITEs), were distributed in a pattern similar to that of genes (Fig. 1b). In contrast, Gypsy and unknown LTR retrotransposons were densely distributed in pericentromeric regions. Similar distribution patterns have been reported in other species (Zavallo *et al.* 2020; Hosaka *et al.* 2022) except for hAT and PIF Harbinger, which were evenly distributed in the *S. verrucosum* chromosomes (Hosaka *et al.* 2022) but were densely distributed to specific regions in a few chromosomes of *S. bulbocastanum* (Fig. 1a).

Completeness of genome assembly and annotation

We calculated the LTR Assembly Index (LAI) scores to evaluate the completeness of the genome assembly. The LAI score of the present *S. bulbocastanum* genome was 9.63, slightly higher than that of the *S. bulbocastanum* PG6241 genome (LAI = 8.97). We also assessed the BUSCO scores of genome sequences in eight *Solanum* species, which were reported as either chromosome-scale assemblies or highly contiguous contigs (Supplementary Table 3). In a genome mode, 98.27% of BUSCO genes in Solanales were detected in the present *S. bulbocastanum* genome. The score was comparable to the other species, demonstrating the highly contiguous assembly. However, the BUSCO score of protein sequences was 91.76%, slightly lower than that of PG6241 protein sequences (95.04%). Since we used the RNA-seq datasets of PG6241, some genes were missed during annotation, possibly due to the sequence variations between the *S. bulbocastanum* strains (Supplementary Table 3).

Synteny and structural variation among *Solanum* species

Structural differences among seven chromosome-scale assemblies were compared. The chromosomal length of *S. bulbocastanum* was the shortest: 7.4% shorter than that of *S. phureja* DM v 8.1 (Supplementary Table 4 and Fig. 2). This is consistent with the finding that the genomes of Mexican diploid species are smaller in size and contained fewer TEs than those of A-genome species (Bozan *et al.* 2023). Dot-plots using D-Genies disclosed low sequence similarities and frequent losses of linearity in pericentromeric and centromere regions (Supplementary Figs. 2–4). Syntenic regions were further identified by SyRI (Fig. 2). Subtelomeric regions of chromosomes tended to be more conserved than pericentromeric regions, as previously described (Hosaka *et al.* 2022; Tang *et al.* 2022). Small but frequent pericentromeric inversions between the *S. bulbocastanum* and *S. verrucosum* genomes were found in chromosomes 1–4, 8–10, and 12, which might coincide with observations of minor structural chromosome differences at pachytene in the F_1 of *S. verrucosum* × *S. bulbocastanum* (Hermsen and Ramanna 1976). Furthermore, in addition to frequent small inversions, large centromeric inversions between the *S. bulbocastanum* and *S. etuberosum* genomes were observed in chromosomes 3, 6, 8, 9, and 12, which likely caused irregular meiosis as observed in F_1 hybrids between Mexican diploid species and *Etuberosa* series species (Ramanna and Hermsen 1981). To represent structural differences, pairwise ratios of syntenic regions were calculated among seven chromosome-scale assemblies (Supplementary Table 5) and visualized as a dendrogram constructed by an unweighted pair group method with arithmetic mean (UPGMA) (Fig. 3a). *S. chacoense* and *S. multidissectum* were clustered first, then with *S. phureja* and *S. verrucosum*, resulting in an A-genome species group. Since *S. multidissectum* is one of the putative ancestral species of cultivated potatoes, as shown in Fig. 3b (Spooner *et al.* 2005; Sukhotu and Hosaka 2006), there might be a possibility of misassembly in either the *S. chacoense* or *S. multidissectum* genomes. Highly repetitive sequences around a centromere might cause a large misassembled region (a centromeric inversion), as occurred in chromosome 12 of the *S. phureja* genome (DM v4.04), which was inverted in DM v6.1 (Pham *et al.* 2020). Consequently,

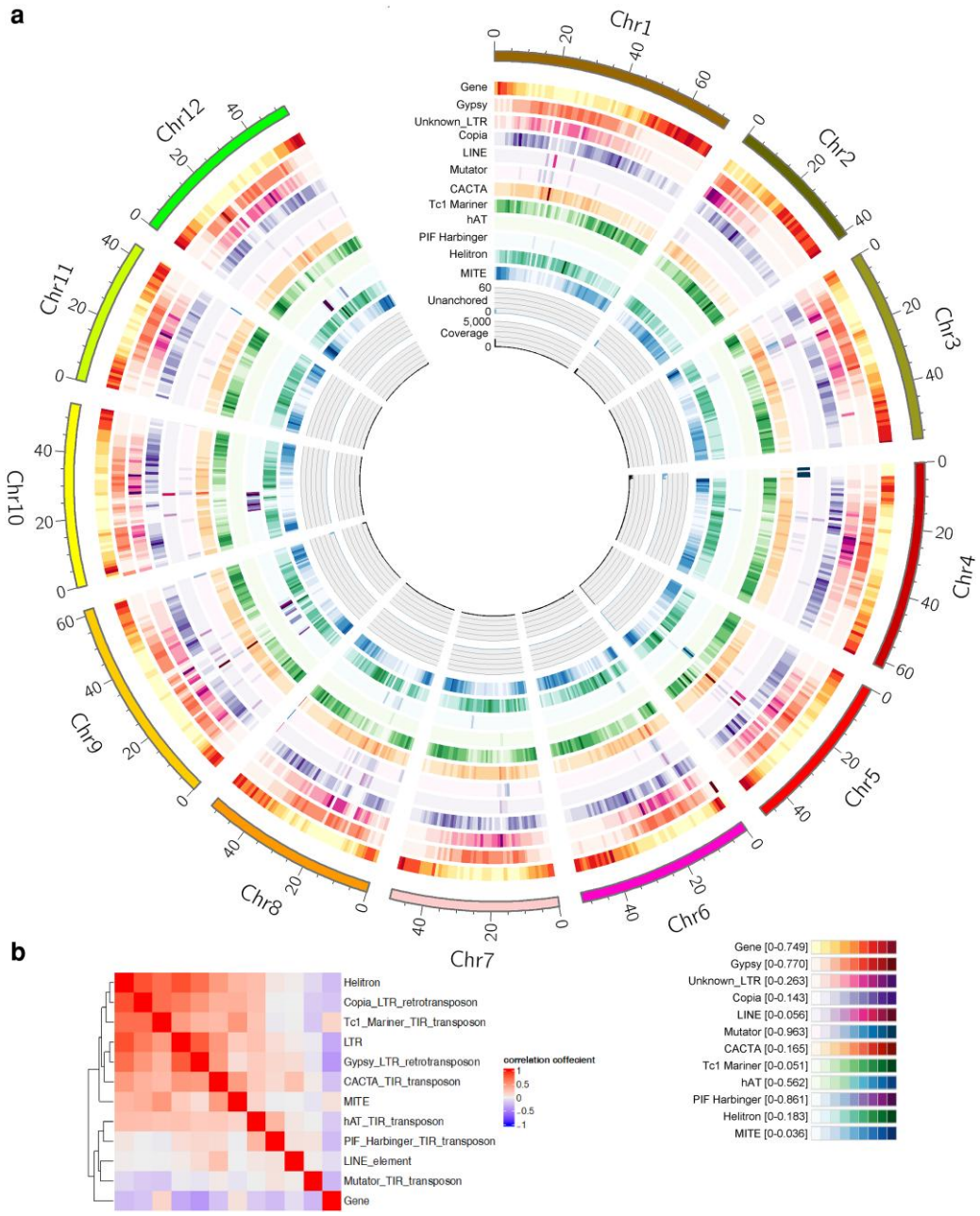

Fig. 1. Chromosomal distribution of HiFi reads, unanchored contigs, genes, and transposons (a) and the correlations of their locations (b).

this group was clustered with *S. bulbocastanum*, forming a tuber-bearing species group (=section *Petota* or Petota clade). The tuber-bearing species group was then clustered with a group of *S. etuberosum* (Etuberosum clade) and tomato (Tomato clade).

Phylogenetic inference

OrthoFinder assigned 1,289,183 genes identified in 36 *Solanum* species and clones into 36,033 orthogroups (Supplementary Table 6). Among the 36,033 orthogroups, 5,905 (16.4%) were present in all the species and clones analyzed. These were used to infer the species' phylogeny using OrthoFinder with the STAG algorithm (Supplementary Fig. 5a). Of the 5,905 common orthogroups, 1,624 were single-copy orthogroups, which were used to infer the species' phylogeny using the MSA method with FastTree 2 and IQ-TREE 2 (Supplementary Fig. 5b and Fig. 3b, respectively). As expected, three phylogenetic trees indicated the present *S. bulbocastanum* located at the same node of *S. bulbocastanum* PG6241, showing that the two genomes are sister genomes and, together with other Mexican diploid species, formed a distinct cluster from the A-genome species, most of which are distributed in South America and created a large group including cultivated species and its wild ancestor (Fig. 3b). Noteworthy, *S. verrucosum*, only A-genome species distributed in Mexico, was located in the cluster of A-genome species. The presence of *S. verrucosum* in Mexico supports the suggestion that an A-genome species dispersed to Mexico from South America (Hawkes and Jackson 1992). The phylogenetic tree in Fig. 3b was similar to that based on sequence similarity in 3,971 single-copy genes, both using the same software IQ-TREE 2 (Tang *et al.* 2022). The two phylogenetic trees suggested that E-genome species share a most recent common

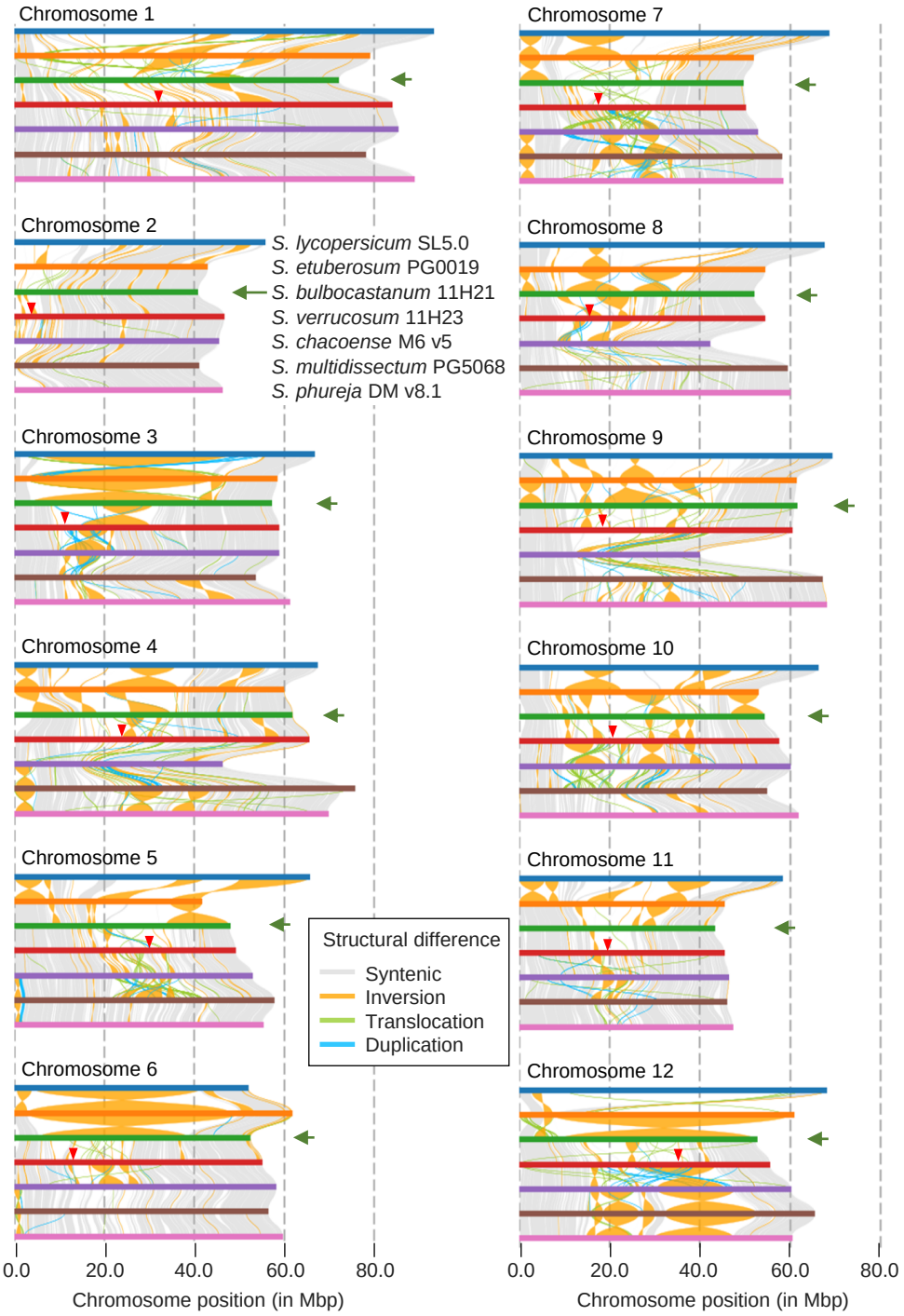

Fig. 2. Length and structural variation among seven chromosome-scale assemblies. Arrowed chromosomes are those of *S. bulbocastanum*. Arrowheads indicate the locations of putative centromeres in the *S. verrucosum* genome (Hosaka *et al.* 2022).

ancestor with the clade comprising tomato and tuber-bearing species. However, the phylogenetic trees generated by STAG and FastTree 2 suggest that the Tomato clade is a sister to a clade comprising E-genome and tuber-bearing species' shared ancestor (Supplementary Fig. 5a, b). Both branching patterns (the Etuberosum clade sister to the Tomato-Petota clade or the Tomato clade sister to the Etuberosum-Petota clade) have previously been reported and argued (Hosaka *et al.* 1984; Spooner *et al.* 1993; Huang *et al.* 2019; Gagnon *et al.* 2022; Tang *et al.* 2022). Structural differences among these genomes exhibited a slightly different branching pattern as described earlier (Fig. 3a). Tang *et al.* (2022) mentioned that a homologous gene to *Identity of Tuber 1*, an essential gene for initiating potato tubers, was identified but unfunctional in E-genome species. The TE profiles support the idea that *S. etuberosum* is more similar to the potato than tomato species (Gaiero *et al.* 2019). Further research is needed to solve this discordance using multiple species and accessions of the E-genome species.

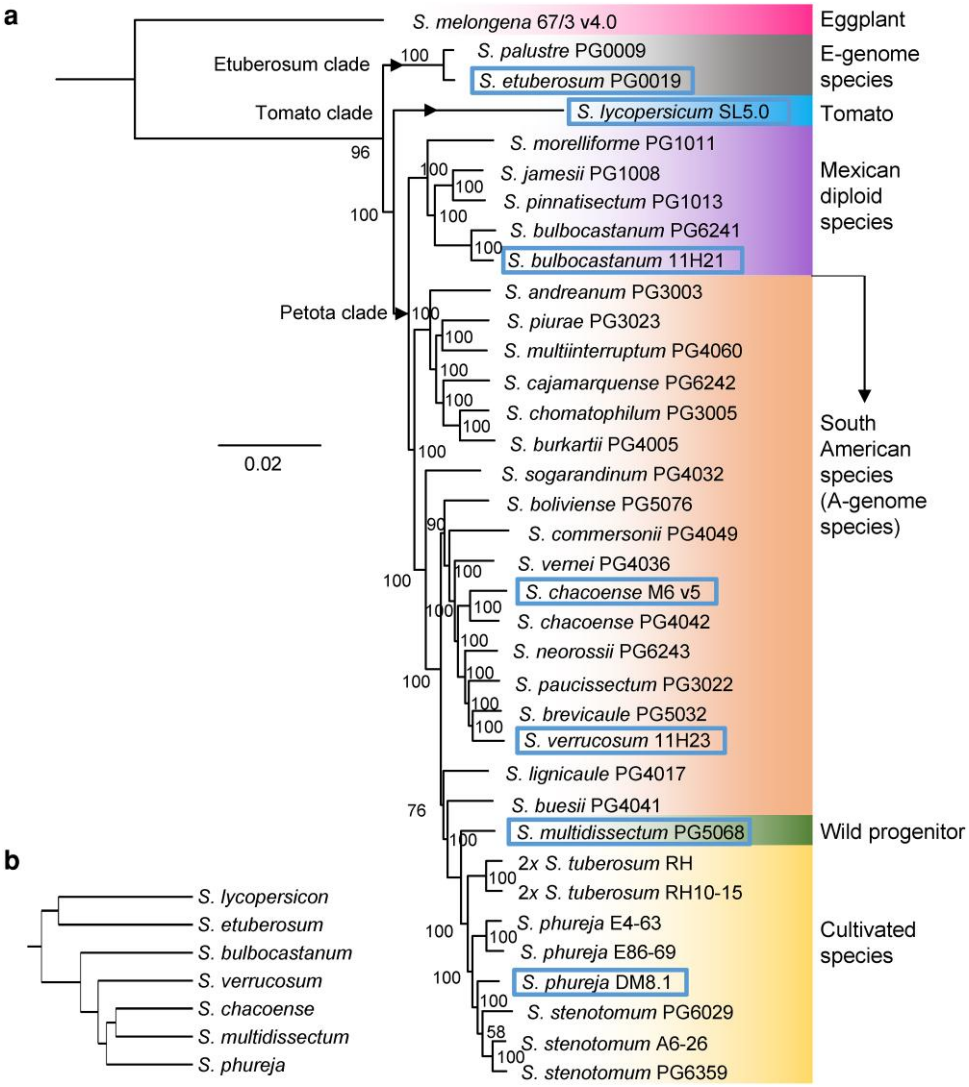

Fig. 3. Species relationships based on genome structures (a) and gene sequence similarities (b). a) A dendrogram, constructed using an UPGMA, based on pairwise ratios of syntenic regions among seven chromosome-scale assemblies, showing genome structure similarity. b) A phylogenetic tree generated using IQ-TREE 2, constructed using single-copy gene sequence similarities in 1,624 orthogroups commonly present in all 36 species. A bootstrap value is given in each node. Species with chromosome-scale assemblies are shown in rectangles. The distinctiveness of *S. bulbocastanum* and other Mexican diploid species from the A-genome species is disclosed. Note that the species names used are those described by Hawkes (1990). According to Spooner's taxonomy, *S. multidissectum* is included in a superspecies *S. candolleanum*, while *S. stenotomum* and *S. phureja* are included in *S. tuberosum* (Spooner *et al.* 2014).

Characterization of the unanchored contigs

Since 1,083 contigs remained unanchored, we further characterized these unanchored contigs. A pairwise comparison of these contigs by Minimap2 (Supplementary Fig. 6a) indicated their highly repetitive nature and shared similarities. Most of these contigs were aligned to the telomeric regions of chromosomes 1–5, and 9, and the subtelomeric region of chromosome 6 (Supplementary Fig. 6b). Interestingly, in these regions, the coverages of HiFi reads were abnormally high, suggesting that their highly repetitive nature made it difficult to assemble even with long reads and caused underestimation of repeat numbers and many unanchored contigs. Based on the coverages of the unanchored contigs and HiFi reads, 14 regions (1a–9) were tentatively identified as unanchored-contig clustered regions (UCCRs) with a size range from 39.7 kb (1b) to 1.14 Mb (4e) (Fig. 4, Supplementary Table 7). The UCCRs were composed of various repeat sequences and arrays of specific TEs: Gypsy (1a, 1b, 4c, 5, 6, and 9), helitron (3b), CACTA (2 and 4c), or Mutator (2, 3a, 4a, 4b, 4c, 4d, 4e, and 4f) (Supplementary Table 7). The UCCRs were aligned using Minimap2 with >80% homology to the contigs constructed from 23 wild species and 6 clones of three cultivated species (Tang *et al.* 2022). The UCCRs 1b and 6 were found in all species, while 1a and 5 were found in 19 and 11 among 29 species and clones, respectively (Supplementary Table 8). The other UCCRs were rare. In *S. bulbocastanum* PG6241, only five (1b, 3a, 4b, 4e, and 6) were found, but the most significant number of UCCRs was found. The UCCR sequences were aligned with >80% homology to seven chromosome-scale assemblies (Supplementary Table 9). Except for the present *S. bulbocastanum* genome, only a few had the same UCCRs: the UCCR 2 in *S. multidissectum* and 1a, 1b, 6, and 9 in *S. lycopersicum*. Therefore, not all but most of these highly repetitive sequences, or UCCRs, are very specific to the present

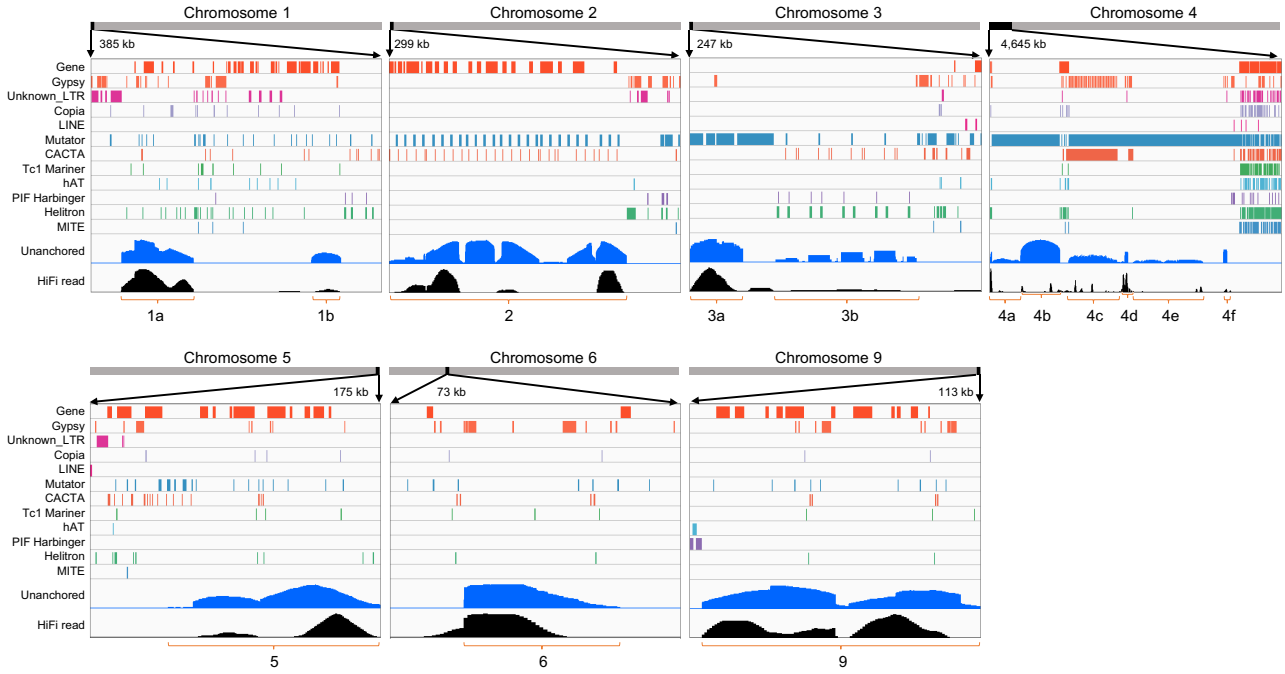

Fig. 4. Distribution of genes, transposons, and dosages of HiFi reads and unanchored contigs in the unanchored-contig clustered regions (UCCRs 1a to 9).

S. bulbocastanum genome. Although a further survey is needed, these UCCRs likely originated by in vitro propagation that continued over half a century, as Bozan *et al.* (2023) suggested.

Conclusions

A de novo chromosome-scale assembly of the *S. bulbocastanum* genome was constructed, which was slightly shorter and had less content of TEs compared with the A-genome. It was distinct from the A-genome in a syntenic relationship and sequence homology of orthologous genes. As noted in our and other studies, it is most closely related to the Mexican diploid species, excluding *S. verrucosum* (Hosaka *et al.* 1984; Spooner *et al.* 1991, 2008; Pendinen *et al.* 2008; Rodríguez and Spooner 2009; Li *et al.* 2018; Huang *et al.* 2019; Tang *et al.* 2022; Bozan *et al.* 2023). These findings provide important insights into understanding the genome evolution of wild potato species and the genetic diversity of Mexican diploid species. Due to reproductive barriers, the use of Mexican diploid species in potato breeding has been mainly limited to that for late blight resistance (Song *et al.* 2003; Lokossou *et al.* 2010; Sanetomo *et al.* 2019). Whole-genome sequences of *S. bulbocastanum* could help to understand and break a genetic mechanism for reproductive barriers, which facilitates the incorporation of valuable traits such as resistance to various viruses, aphids, Colorado potato beetle, and nematodes, and heat tolerance (Hawkes 1958, 1990; Hanneman 1989) into breeder's gene pools.

Data availability

The raw DNA sequencing reads, genome assembly, and annotation have been deposited into the National Center for Biotechnology Information under BioProject Number PRJNA1009588. Supplemental material is available at G3 online.

Supplemental material available at G3 online.

Acknowledgments

We thank Dr Sarah Mathews for constructive comments to improve the manuscript and Kazumi Tanaka for assistance in in vitro plant maintenance. KH and RS designed the research. KH prepared the DNA sample for sequencing. AH performed genome sequencing, assembly, annotation, and comparative genome analysis. AH and KH wrote the manuscript with input from all authors. The Human Genome Center at the University of Tokyo provided supercomputing resources. All the authors have read and approved the final manuscript.

Funding

This research was supported by potato industries and growers: Calbee, Inc., Hokkaido Potato Growers Association, KENKO Mayonnaise Co., Ltd., KOIKE-YA Inc., Yamayoshi Seika Co., Ltd., Yamazaki-Biscuit Co., Ltd., KIKUSUI-Do Co., Ltd., and Fukagawa Oil & Fat Co., Ltd.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Literature cited

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 23(1):258. doi:10.1186/s13059-022-02823-7.

Austin S, Pohlman JD, Brown CR, Mojtahedi H, Santo GS, Douches DS, Helgeson JP. 1993. Interspecific somatic hybridization between *Solanum tuberosum* L. and *S. bulbocastanum* Dun. as a means of transferring nematode resistance. Am Potato J. 70(6):485–495. doi:10.1007/BF02849067.

Aversano R, Contaldi F, Ercolano MR, Grosso V, Iorizzo M, Tatino F, Xumerle L, Molin AD, Avanzato C, Ferrarini A, *et al.* 2015. The *Solanum commersonii* genome sequence provides insights into adaptation to stress conditions and genome evolution of wild potato relatives. Plant Cell. 27(4):954–968. doi:10.1105/tpc.114.135954.

Bali S, Brown C, Majtahedi H, Yilma S, Ingham RE, Cimrhakl L, Quick R, Sathuvalli V. 2022. Genomic markers linked to *Meloidogyne chitwoodi* resistance introgressed from *Solanum bulbocastanum* to cultivated potato and their utility in marker–assisted selection. Mol Breed. 42(3):12. doi:10.1007/s11032-022-01285-w.

Barchi L, Rabanus-Wallace MT, Prohens J, Toppino L, Padmarasu S, Portis E, Rotino GL, Stein N, Lanteri S, Giuliano G. 2021. Improved genome assembly and pan-genome provide key insights into eggplant domestication and breeding. Plant J. 107(2): 579–596. doi:10.1111/tpj.15313.

Bozan I, Achakkagari SR, Anglin NL, Ellis D, Tai HH, Strömvik MV. 2023. Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. Proc Natl Acad Sci USA. 120(31):e2211117120. doi:10.1073/pnas.2211117120.

Brown CR, Mojtahedi H, Santo GS. 1995. Introgression of resistance to Columbia and Northern root-knot nematodes from *Solanum bulbocastanum* into cultivated potato. Euphytica. 83(1):71–78. doi: 10.1007/BF01677863.

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 6:e4958. doi:10.7717/peerj.4958.

Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 18(1):188–196. doi:10.1101/gr.6743907.

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34(17):i884–i890. doi:10.1093/bioinformatics/bty560.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 18(2):170–175. doi:10.1038/s41592-020-01056-5.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, *et al.* 2017. De novo assembly of the *Aedes aegypti* genome using Hi–C yields chromosome-length scaffolds. Science. 356(6333):92–95. doi:10.1126/science.aal3327.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi–C experiments. Cell Syst. 3(1):95–98. doi:10.1016/j.cels.2016.07.002.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16(1):157. doi:10.1186/s13059-015-0721-2.

Emms DM, Kelly S. 2018. STAG: species tree inference from all genes. bioRxiv 267914. https://doi.org/10.1101/267914, preprint: not peer reviewed.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238. doi:10.1186/s13059-019-1832-y.

Gagnon E, Hilgenhof R, Orejuela A, McDonnell A, Sablok G, Aubriot X, Giacomin L, Gouvêa Y, Bragionis T, Stehmann JR, *et al.* 2022. Phylogenomic discordance suggests polytomies along the backbone of the large genus *Solanum*. Am J Bot. 109(4):580–601. doi: 10.1002/ajb2.1827.

Gaiero P, Vaio M, Peters SA, Schranz ME, de Jong H, Speranza PR. 2019. Comparative analysis of repetitive sequences among species from the potato and the tomato clades. Ann Bot. 123(3): 521–532. doi:10.1093/aob/mcy186.

Goel M, Schneeberger K. 2022. Plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics. 38(10):2922–2926. doi:10.1093/bioinformatics/btac196.

Goel M, Sun H, Jiao W-B, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biol. 20(1):277. doi:10.1186/s13059-019-1911-0.

Graham KM, Niederhauser JS, Servin L. 1959. Studies on fertility and late blight resistance in *Solanum bulbocastanum* Dun. in Mexico. Can J Bot. 37(1):41–49. doi:10.1139/b59-003.

Hanneman RE Jr. 1989. The potato germplasm resource. Am Potato J. 66(10):655–667. doi:10.1007/BF02853985.

Hawkes JG. 1958. Significance of wild species and primitive forms for potato breeding. Euphytica. 7(3):257–270. doi:10.1007/BF00025267.

Hawkes JG. 1990. The Potato: Evolution, Biodiversity and Genetic Resources. London: Belhaven Press.

Hawkes JG, Jackson MT. 1992. Taxonomic and evolutionary implications of the endosperm balance number hypothesis in potatoes. Theor Appl Genet. 84(1–2):180–185. doi:10.1007/BF00223998.

Hermsen JGTh, Ramanna MS. 1976. Barriers to hybridization of *Solanum bulbocastanum* Dun. and *S. verrucosum* Schlechtd. and structural hybridity in their F_1 plants. Euphytica. 25(1):1–10. doi: 10.1007/BF00041523.

Hoopes G, Meng X, Hamilton HP, Achakkagari SR, de Alves Freitas Guesdes F, Bolger ME, Coombs JJ, Esselink D, Kaiser NR, Kodde L, *et al.* 2022. Phased, chromosome-scale genome assemblies of tetraploid potato reveals a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. Mol Plant. 15(3):520–536. doi:10.1016/j.molp.2022.01.003.

Hosaka AJ, Sanetomo R, Hosaka K. 2022. A *de novo* genome assembly of *Solanum verrucosum* schlechtendal, a Mexican diploid species geographically isolated from other diploid A-genome species of potato relatives. G3 (Bethesda). 12(8):jkac166. doi:10.1093/g3journal/jkac166.

Hosaka K, Ogihara Y, Matsubayashi M, Tsunewaki K. 1984. Phylogenetic relationship between the tuberous *Solanum* species as revealed by restriction endonuclease analysis of chloroplast DNA. Jpn J Genet. 59(4):349–369. doi:10.1266/jjg.59.349.

Hosmani PS, Flores-Gonzalez M, van de Geest H, Maumus F, Bakker LV, Schijlen E, van Haarst J, Cordewener J, Sanchez-Perez G, Peters S, *et al.* 2019. An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv 767764. https://doi.org/10.1101/767764, preprint: not peer reviewed.

Huang B, Ruess H, Liang Q, Colleoni C, Spooner DM. 2019. Analyses of 202 plastid genomes elucidate the phylogeny of *Solanum* section *Petota*. Sci Rep. 9(1):4454. doi:10.1038/s41598-019-40790-5.

Irikura Y, Sakaguchi S. 1972. Induction of 12-chromosome plants from anther culture in a tuberous *Solanum*. Potato Res. 15(2): 170–173. doi:10.1007/BF02355964.

Jansky S, Hamernik A. 2009. The introgression of 2*x* 1EBN *Solanum* species into the cultivated potato using *Solanum verrucosum* as a bridge. Genet Resour Crop Evol. 56(8):1107–1115. doi:10.1007/s10722-009-9433-3.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780. doi:10.1093/molbev/mst010.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotech. 37(8):907–915. doi:10.1038/s41587-019-0201-4.

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. 20(1):278. doi:10.1186/s13059-019-1910-1.

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2019. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Res. 47(D1):D807–D811. doi:10.1093/nar/gky1053.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19(9):1639–1645. doi:10.1101/gr.092759.109.

Leisner CP, Hamilton JP, Crisovan E, Manrique-Carpintero NC, Marand AP, Newton L, Pham GM, Jiang J, Douches DS, Jansky SH, *et al.* 2018. Genome sequence of M6, a diploid inbred clone of the high-glycoalkaloid-producing tuber-bearing potato species *Solanum chacoense*, reveals residual heterozygosity. Plant J. 94(3):562–570. doi:10.1111/tpj.13857.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094–3100. doi:10.1093/bioinformatics/bty191.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics. 25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

Li Y, Colleoni C, Zhang J, Liang Q, Hu Y, Ruess H, Simon R, Liu Y, Liu H, Yu G, *et al.* 2018. Genomic analyses yield markers for identifying agronomically important genes in potato. Mol Plant. 11(3):473–484. doi:10.1016/j.molp.2018.01.009.

Lokossou AA, Rietman H, Wang M, Krenek P, van der Schoot H, Henken B, Hoekstra R, Vleeshouwers VGAA, van der Vossen EAG, Visser RGF, *et al.* 2010. Diversity, distribution, and evolution of *Solanum bulbocastanum* late blight resistance genes. Mol Plant Microbe Interact. 23(9):1206–1216. doi:10.1094/MPMI-23-9-1206.

Magoon ML, Cooper DC, Hougas RW. 1958. Cytogenetic studies of some diploid solanums section *Tuberarium*. Am J Bot. 45(3):207–221. doi:10.1002/j.1537-2197.1958.tb12211.x.

Marks GE. 1968. Structural hybridity in a tuberous *Solanum* hybrid. Can J Genet Cytol. 10(1):18–23. doi:10.1139/g68-003.

Matsubayashi M. 1991. Phylogenetic relationships in the potato and its related species. In: Tsuchiya T, Gupta PK, editors. Chromosome Engineering in Plants: Genetics, Breeding, Evolution, Part B. Amsterdam: Elsevier. p. 93–118.

Matsubayashi M, Misoo S. 1977. Species differentiation in *Solanum*, sect. Tuberarium. IX. Genomic relationships between three Mexican diploid species. Jap J Breed. 27(3):241–250. doi:10.1270/jsbbs1951.27.241.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 37(5):1530–1534. doi:10.1093/molbev/msaa015.

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al.* 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20(1):275. doi:10.1186/s13059-019-1905-y.

Pendinen G, Gavrilenko T, Jiang J, Spooner DM. 2008. Allopolyploid speciation of the Mexican tetraploid potato species *Solanum stoloniferum* and *S. hjertingii* revealed by genomic in situ hybridization. Genome. 51(9):714–720. doi:10.1139/G08-052.

Pham GM, Hamilton JP, Wood JC, Burke JT, Zhao H, Vaillancourt B, Ou S, Jiang J, Buell CR. 2020. Construction of a chromosome-scale long-read reference genome assembly for potato. GigaSci. 9(9):giaa100. doi:10.1093/gigascience/giaa100.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS One. 5(3):e9490. doi:10.1371/journal.pone.0009490.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841–842. doi:10.1093/bioinformatics/btq033.

Ramanna MS, Hermsen JGTh. 1981. Structural hybridity in the series *Etuberosa* of the genus *Solanum* and its bearing on crossability. Euphytica. 30(1):15–31. doi:10.1007/BF00033655.

Rodríguez F, Spooner DM. 2009. Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota*) genomes with emphasis on the origins of the polyploid species. Syst Bot. 34(1):207–219. doi:10.1600/036364409787602195.

Sanetomo R, Habe I, Hosaka K. 2019. Sexual introgression of the late blight resistance gene *Rpi-blb3* from a Mexican wild diploid species *Solanum pinnatisectum* Dunal into potato varieties. Mol Breed. 39(1):13. doi:10.1007/s11032-018-0924-9.

Sanetomo R, Hosaka K. 2021. Re-evaluation of monohaploid *Solanum verrucosum* and *S. bulbocastanum* ($2n = x$=12) and dihaploid *S. stoloniferum* and *S. acaule* ($2n$=$2x$=24), all derived from anther culture. Am J Potato Res. 98(4):333–343. doi:10.1007/s12230-021-09847-y.

Sharma P, Masouleh AK, Topp B, Furtado A, Henry RJ. 2022. De novo chromosome level assembly of a plant genome from long read sequence data. Plant J. 109(3):727–736. doi:10.1111/tpj.15583.

Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 11(10):e0163962. doi:10.1371/journal.pone.0163962.

Shumate A, Salzberg SL. 2021. Liftoff: accurate mapping of gene annotations. Bioinformatics. 37(12):1639–1643. doi:10.1093/bioinformatics/btaa1016.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210–3212. doi:10.1093/bioinformatics/btv351.

Song J, Bradeen JM, Naess SK, Raasch JA, Wielgus SM, Haberlach GT, Liu J, Kuang H, Austin-Phillips S, Buell CR, *et al.* 2003. Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proc Natl Acad Sci USA. 100(16):9128–9133. doi:10.1073/pnas.1533501100.

Spooner DM, Anderson GJ, Jansen RK. 1993. Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). Am J Bot. 80(6):676–688. doi:10.1002/j.1537-2197.1993.tb15238.x.

Spooner DM, Ghislain M, Simon R, Jansky SH, Gavrilenko T. 2014. Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. Bot Rev. 80(4):283–383. doi:10.1007/s12229-014-9146-y.

Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ. 2005. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proc Natl Acad Sci USA. 102(41):14694–14699. doi:10.1073/pnas.0507400102.

Spooner DM, Rodríguez F, Polgár Z, Ballard HE Jr, Jansky SH. 2008. Genomic origins of potato polyploids: GBSSI gene sequencing data. Crop Sci. 48(S1):S27–S36. doi:10.2135/cropsci2007.09.0504tpg.

Spooner DM, Sytsma KJ, Conti E. 1991. Chloroplast DNA evidence for genome differentiation in wild potatoes (*Solanum sect.* Petota: *Solanaceae*). Am J Bot. 78(10):1354–1366. doi:10.1002/j.1537-2197.1991.tb12602.x.

Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res. 32(Web Server):W309–W312. doi:10.1093/nar/gkh379.

Sukhotu T, Hosaka K. 2006. Origin and evolution of Andigena potatoes revealed by chloroplast and nuclear DNA markers. Genome. 49(6):636–647. doi:10.1139/g06-014.

Sun H, Jiao WB, Krause K, Campoy JA, Goel M, Folz-Donahue K, Kukat C, Huettel B, Schneeberger K. 2022. Chromosome-scale and haplotype-resolved genome assembly of a tetraploid potato cultivar. Nat Genet. 54(3):342–348. doi:10.1038/s41588-022-01015-0.

Tang D, Jia Y, Zhang J, Li H, Cheng L, Wang P, Bao Z, Liu Z, Feng S, Zhu X, *et al.* 2022. Genome evolution and diversity of wild and cultivated potatoes. Nature. 606(7914):535–541. doi:10.1038/s41586-022-04822-x.

Tiwari JK, Devi S, Ali N, Luthra SK, Kumar V, Bhardwaj V, Singh RK, Rawat S, Chakrabarti SK. 2018. Progress in somatic hybridization research in potato during the past 40 years. Plant Cell Tiss Organ Cult. 132(2):225–238. doi:10.1007/s11240-017-1327-z.

Tiwari JK, Rawat S, Luthra SK, Zinta R, Sahu S, Varshney S, Kumar V, Dalamu D, Mandadi N, Kumar M, *et al.* 2021. Genome sequence analysis provides insights on genomic variation and late blight resistance genes in potato somatic hybrid (parents and progeny). Mol Biol Rep. 48(1):623–635. doi:10.1007/s11033-020-06106-x.

Toxopeus HJ. 1964. Treasure-digging for blight resistance in potatoes. Euphytica. 13(3):206–222. doi:10.1007/BF00023099.

Yan L, Zhang Y, Cai G, Qing Y, Song J, Wang H, Tan X, Liu C, Yang M, Fang Z, *et al.* 2021. Genome assembly of primitive cultivated potato *Solanum stenotomum* provides insights into potato evolution. G3 (Bethesda). 11(10):jkab262 doi:10.1093/g3journal/jkab262.

Yang X, Zhang L, Guo X, Xu J, Zhang K, Yang Y, Yang Y, Jian Y, Dong D, Huang S, *et al.* 2023. The gap-free potato genome assembly reveals large tandem gene clusters of agronomical importance in highly repeated genomic regions. Mol Plant. 16(2):314–317. doi:10.1016/j.molp.2022.12.010.

Zavallo D, Crescente JM, Gantuz M, Leone M, Vanzetti LS, Masuelli RW, Asurmendi S. 2020. Genomic re-assessment of the transposable element land scape of the potato genome. Plant Cell Rep. 39(9):1161–1174. doi:10.1007/s00299-020-02554-8.

Zhou Q, Tang D, Huang W, Yang Z, Zhang Y, Hamilton JP, Visser RGF, Bachem CWB, Robin Buell C, Zhang Z, *et al.* 2020. Haplotype-resolved genome analyses of a heterozygous diploid potato. Nat Genet. 52(10):1018–1023. doi:10.1038/s41588-020-0699-x.

Zhou Y, Zhang Z, Bao Z, Li H, Lyu Y, Zan Y, Wu Y, Cheng L, Fang Y, Wu K, *et al.* 2022. Graph pangenome captures missing heritability and empowers tomato breeding. Nature. 606(7914):527–534. doi:10.1038/s41586-022-04808-9.

Editor: S. Mathews