Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Draft Genome Assembly of a Wild Potato (Solanum bulbocastanum) reveals disease resistance genes against Columbia Root-Knot Nematode (Meloidogyne chitwoodi)

Oluwafemi Omilani
Oregon State University
Senthilkumar Shanmugavel
Oregon State University
Kelly Vining
Oregon State University
Charles Brown
United States Department of Agriculture
Solomon Yilma
Oregon State University
Vidyasagar Sathuvalli
vidyasagar@oregonstate.edu
Oregon State University

Research Article

Keywords: diploid potato, wild species, resistance (R) genes, late blight, disease resistance breeding

Posted Date: November 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11681742/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A Draft Genome Assembly of a Wild Potato (*Solanum* bulbocastanum) reveals disease resistance genes against Columbia Root-Knot Nematode (*Meloidogyne chitwoodi*)

Oluwafemi Omilani[1#], Senthilkumar Shanmugavel[2#], Kelly Vining[3], Charles Brown[4], Solomon Yilma[1], Vidyasagar Sathuvalli[1,2*]

[1]Department of Crop and Soil Science, 109 Crop Science Building, Corvallis, OR 97331, USA.
[2]Hermiston Agricultural Research and Extension Centre, 2121 S 1st Street, Hermiston, OR 97838, USA.
[3]Department of Horticulture, 2750 SW Campus Way, Corvallis, OR 97331, USA.
[4]United States Department of Agriculture, Vegetable and Forage Crops Production Research Unit, 24106 N. Bunn Rd., Prosser, WA 99350, USA.
# authors had equal contribution to the manuscript
*E-mail for Correspondence: vidyasagar@oregonstate.edu and kelly.vining@oregonstate.edu

C. Brown is deceased

ABSTRACT

The cultivated potato is susceptible to numerous pests and diseases for which wild relatives represent an important source of genetic resistance. The wild potato, *Solanum bulbocastanum*, is one such species providing resistance to Columbia Root-Knot Nematode (CRKN). Here, we report a chromosome-scale draft genome assembly of a *S. bulbocastanum* selection 22 (SB22), generated using Illumina short reads, PacBio long reads, and scaffolded with Hi-C data into a haploid genome. The final genome assembly size was 656.37 Mb, including 12 chromosome pseudomolecules, with a genome-mode BUSCO completeness score of 95.8%. Gene prediction using BRAKER predicted 38,077 gene models and 57,200 transcripts. Repeat elements accounted for 63.8 % of the genome, with LTR retrotransposons being the most abundant. Comparative genomic analyses with SB11H21 and DM6.1 revealed an extensive conservation of gene order, alongside variations in the abundance and distribution of nucleotide binding leucine-rich repeat (NLR) loci. NLR-Annotator identified 576 candidate *R* gene loci in SB22, with 233 CNL-like and 58 TNL-like loci (Table 4). Chromosome 11 contained 104 NLR loci, with 35 concentrated in the distal 3 Mb (Table S2). Putative resistance gene loci were also identified near previously characterized CRKN resistance markers on chromosome 11. The SB22 genome assembly provides a valuable genomic resource for comparative resistance gene analyses, thereby facilitating the investigation of distinct genetic regions associated with CRKN root and tuber resistance, and the development of molecular tools for potato improvement.

KEYWORDS: diploid potato, wild species, resistance (R) genes, late blight, disease resistance breeding

INTRODUCTION

Potato *(Solanum tuberosum* L.*)* is a tuber crop which is native to the Andes of South America, and it ranks as the world's fourth most important crop after rice, wheat, and maize (Spooner et al. 2014; Dolničar 2021). Since its introduction to Europe in the late sixteenth century, it has become a major agricultural crop worldwide, with one in seven people relying on the crop for food (Wilson, 1959; Devaux et al. 2020). Global production stands at 390 million tons, and the United States, the world's third largest producer, produces 19 million tons with a crop value of $4.5 billion in 2025 (USDA, NASS, 2026). The Pacific Northwest, which includes Idaho, Oregon, and Washington, is a highly productive potato-growing region, accounting for ~60% of the US production. In 2025, the crop value in these three states was $2.21 billion (USDA, NASS, 2026). This productivity is, however, limited by many pests and diseases which reduce both the yield and marketable quality of the crop.

In the Pacific Northwest region of the United States, root-knot disease caused by a soil-borne nematode, the Columbia Root Knot Nematode (CRKN) (*Meloidogyne chitwoodi Golden* O'Bannon, Santo, and Finley), has a significant economic impact on potato production (King and Taberna 2013). CRKN infests potato roots and tubers causing small brown spots in the tuber flesh that reduce tuber quality and marketability. Standard control measures include soil fumigation, crop rotation and nematicide applications. Although soil fumigation and nematicide applications are widely used to manage CRKN, their high cost and environmental impacts highlight the need for alternative control strategies (Desaeger et al. 2020). CRKN resistant varieties represent a cost-effective and environmentally sustainable strategy for managing CRKN. While the cultivated potato lacks genetic resistance to CRKN, resistance has been identified in several closely related wild potato species including *S. bulbocastanum, S. cardiophyllum, S. brachistotrichum, S. fendleri, S. stenophyllidium* and *S. hougasii* (Janssen et al. 1995; Brown et al., 1999; Graebner et al. 2018). *S. bulbocastanum* therefore presents a promising and effective source of resistance to CRKN for disease resistance introgression and breeding cultivated potatoes (Austin *et al*., 1993; Mojtahedi *et al*., 1995).

Solanum bulbocastanum Dunal, a wild diploid potato species ($2n = 2x = 24$), is native to the region encompassing modern Mexico and Central America. Screening for CRKN resistance identified *S. bulbocastanum* (selection 22) - hereafter designated as 'SB22' - representing the 22nd selection from accession PI 275187 (HAW 1584) collected in 1961 from Michoacán de Ocampo, Mexico (USDA Agricultural Research Service (USDA-ARS), n.d.). Earlier studies showed that SB22 exhibited non-host resistance to root-knot nematodes, including race 1 and race 2 of *M. chitwoodi*, *M. persicae* and *M. euphorbiae* (Radcliffe and Lauer 1966; Brown 1989). More recent studies comparing resistant and susceptible potato genotypes indicated that SB22 nematode resistance comes from the inability of CRKN juveniles to establish feeding sites (Brown et al. 2014; Bali et al. 2019). There have been efforts to introgress SB22 CRKN resistance into commercial potato cultivars through somatic hybridization

(Austin et al. 1993) and backcrossing (Brown et al. 1995; Brown et al. 1999). PA99N82-4 - an advanced selection from the Pacific Northwest variety development program with resistance to CRKN - is the result of a successful introgression via somatic hybridization between SB22 and an improved *S. tuberosum* clone followed by multiple generations of backcrossing to *S. tuberosum* (Brown et al. 2009; Graebner et al. 2018). Although PA99N82-4 bears some phenotypic semblances with commercial russet potato varieties, its release as a commercial cultivar has been withheld due to specific agronomic deficiencies including susceptibility to growth cracks.

Potato breeding is a long-term process and typically takes over a decade from the initial hybridization process to release of improved selections as varieties. This process takes even longer when dealing with disease resistance introgression, as screening for resistance under field or greenhouse conditions typically takes multiple years. To reduce the 10-12 year development time of a potato cultivar and integrate novel resistance traits, modern potato breeding programs incorporate molecular markers and advanced genomic resources. Identification of plant disease resistance ("*R"*) genes and associated markers is valuable to potato breeding programs. Most *R* genes encode proteins that contain a putative amino-terminal signaling domain, a nucleotide binding site (NBS), and a series of carboxy-terminal leucine-rich repeats (LRR). These *R* genes are further divided into TIR-NBS-LRR (TNL), which has an amino-terminal TIR (Toll/interleukin receptor domain), and CC-NBS-LRR (CNL), which encodes an N-terminal coiled-coil domain (Wersch et al. 2020). Nematode resistance genes have been identified in Solanaceae crops: in tomato, a single CNL*, Mi-1*, confers resistance to three species of root-knot nematodes (Milligan et al. 1998; Jablonska et al. 2007); in potato, a non-TIR NBS-LRR, *Gpa2,* confers resistance to pale cyst nematode by recognizing pathogen effector variants and shares high homology with the potato virus X resistance gene *Rx1*, but an immune response is triggered only if *RanGAP2* is present (Sacco et al. 2009). Resistance to CRKN from SB22 is hypothesized to be conferred by two known *R* genes, $R_{Mctuber\ (blb)}$ and $R_{Mc\ (blb)}$ providing resistance to tuber and root, independently (Brown et al. 2009). Genetic mapping has placed R_{mc} , the locus conferring resistance to race 1 of *M. chitwoodi*, on chromosome 11 (Brown et al. 1996; Zhang et al. 2007). Broad transcriptomic profiling of susceptible and *S. bulbocastanum*-introgressed resistant clones showed upregulation of defense-related genes (Bali et al. 2019) . Despite identification of a resistance gene locus for CRKN, knowledge about molecular mechanisms and potential resistance gene candidates is limited.

The widely-used potato reference genome, DM (DM1–3 516 R44) (Pham et al. 2020; Yang et al. 2023) hereafter called DM6.1 is a doubled monoploid clone of *S. tuberosum* Group Phureja. However, this reference does not encompass the genetic diversity of cultivated potatoes, nor does it fully represent the spectrum of disease resistance genes. Therefore, it is essential to develop a comprehensive genome resource repository to enhance our understanding of potatoes. The genome of *S. bulbocastanum* (SB22) will provide valuable insights into resistance to key pathogens such as CRKN, and it can serve as a resource for plant breeding programs targeting genetic improvement for this

nematode. Two *S. bulbocastanum* assemblies, which include one chromosome level assembly of '11H21' (Hosaka et al. 2024) and one contig level assembly 'PG6241' (Tang et al. 2022). have previously been published. Given the importance of *S. bulbocastanum* as a source of pest and disease resistance, the SB22 genome sequencing project was undertaken to expand our understanding of the "resistome" of this wild potato specie, thereby providing an additional genomic resource for addressing evolving biological pressures on the cultivated potato.

MATERIALS AND METHODS

Plant material and genome sequencing

SB22 (PI 275187) was obtained from USDA-ARS, Prosser, WA which was collected and maintained by Dr Charles Brown (USDA-ARS, Prosser, WA). High molecular weight genomic DNA was isolated from 1-3g of young, unexpanded leaf tissue using the method previously described (Healey et al. 2014), with the final gDNA sample re-suspended in EB buffer (Qiagen, Hilden, Germany). Genomic DNA quality was assessed using a Nanodrop spectrophotometer with accepted 260/280 absorbance ratio values greater than or equal to 1.8 and by gel electrophoresis of both intact and HindIII-digested samples to ensure integrity and absence of contaminating RNA and polysaccharides. DNA quantity was measured by Qubit, using the Qubit® dsDNA HS Assay Kit (Molecular Probes, Eugene, OR). DNA was sequenced on a Pacific Biosciences RSII instrument at the Arizona Genomics Institute. Sequencing libraries were prepared at 20 and 35 kb and loaded on 4 SMRT cells. Illumina Libraries were prepared from gDNA in 150bp paired end configuration and sequenced using Illumina HiSeq3000 at Center for Quantitative Life Sciences, Corvallis, OR.

Genome assembly

An initial genome assembly was performed from the PacBio data using Falcon v0.5 (Chin et al. 2016). Genome size and heterozygosity were estimated using a kmer-based approach (k = 21) with Illumina reads, using Jellyfish version 2.3 (Marçais and Kingsford 2011) and GenomeScope 2.0 (Ranallo-Benavidez et al. 2020). Errors from the PacBio assembly were corrected using Illumina HiSeq3000 reads (2x150 bp) with Pilon v1.22 (Walker et al. 2014). Genome scaffolding was completed using the HiRise method by Dovetail Genomics (Scotts Valley, CA, USA) according to the methods described by Putnam et al. 2016. The final super scaffolds were then compared to DM6.1 assembly with minimap2 tool v2.1 (Li 2018), visualized using the D-GENIES tool, and merged to make final chromosome-level assemblies. The quality and contiguity of the genome assembly were assessed using the Quality Assessment Tool (QUAST v5.2.0) (Gurevich et al., 2013). BUSCO (Manni et al. 2021) was invoked to assess genome completeness using the Solanales_odb10 data set (Manni et al. 2021). As another check for completeness, the LTR assembly index (LAI) score was computed using the tools

ltr_harvest version v1.6.1 (Ellinghaus et al. 2008), LTR_FINDER_parallel v1.07 (Xu et al. 2007), and LTR_retriever (Ou and Jiang 2018).

Synteny analysis

The SB22 genome assembly was aligned to the DM6.1 potato genome (Pham et al. 2020) and the *S. bulbocastanum* (SB11H21) genome (Hosaka et al. 2024) with minimap2 v2.26 (Li 2018). The alignment was processed through Syri v1.6.3 (Goel et al. 2019) with default parameters to identify structural variations, which were then visualized using Plotsr v1.1.0 (Goel and Schneeberger 2022). To analyze the syntenic relationships and orthology shared with SB11H21 and DM6.1, we utilized GENESPACE R package (v1.4) (Lovell et al., 2022). High-quality protein coding genes and corresponding genomic coordinates (GFF3) were formatted as inputs. GENESPACE was executed using the default pipeline constraints described in the user manual for synteny block detection shared in the three genomes.

Repeat masking and functional annotation

De-Novo Repeat libraries for SB22 were created using RepeatModeler2 v2.0.4 (Flynn et al. 2020) package and the repeat libraries generated were then used to softmask the SB22 genome with RepeatMasker v4.1.0 (Nishimura 2000) and to estimate the total repeat content in the genome. The repeat masked genome was subjected to gene prediction with BRAKER2 v2.1.6 (Stanke et al. 2006a; Hoff et al. 2016; Brůna et al. 2021) and BRAKER3 v3.0.8 pipelines (Gabriel et al 2024). The protein models used in gene prediction were derived from Viridiplantae (https://bioinf.uni-greifswald.de/bioinf/partitioned_odb11/Viridiplantae.fa.gz), DM6.1, Atlantic v3, and RH89-039-16 (Simão et al. 2015; Pham et al. 2020; Zhou et al. 2020; Hoopes et al. 2022). The RNA-seq evidence obtained from NCBI SRA database were mapped to the SB22 genome with Hisat2 with " -dta -q" flag. To fully incorporate a homology-based approach into the gene annotation set, BRAKER2-Genome threader was run with protein models previously described. BRAKER2 was run with RNA evidence from leaf, stem, root, flower, stolon, and tuber tissues of *Solanum bulbocastanum* PG6241 (SRR15560132, SRR15560142, SRR15560143, SRR15560145, SRR15560146, SRR15560147, SRR15560148, SRR15560149, SRR15560150, SRR15560151, SRR15560152, SRR15560153) (Tang et al 2022; Achakkagari et al. 2024) and SB22 root tissue (unpublished). BRAKER3 pipeline was run with RNA-seq and protein evidence described above and BRAKER3-Isoseq utilized RNA-seq evidence from *Solanum tuberosum* tubers, stolons, stem, root, and mature and young leaves (SRR14298414, SRR14298419, SRR14298420, SRR14298435, SRR14298437, SRR1428447, and SRR1429455), which was mapped using minimap2 (-ax splice:hq -uf). Due to differences in samples derived from different sequencing platforms, we decided to run a second BRAKER3 with transcriptome evidence from *S. bulbocastanum* (leaf and tuber) (SRR23225910, SRR23225911) and *S. brevicaule* (leaf and shoot) (SRR23225938, SRR23225939) (Achakkagari et al. 2024), and the previously cited

Viridiplantae_ODB11 protein set. A final high-quality gene prediction set was obtained by concatenating all five annotation sets using TSEBRA v1.1 (Gabriel et al. 2021). The final prediction set was assessed for completeness using BUSCO in protein mode with the Solanales_odb10 set.

This final gene prediction set was subject to functional annotation using OmicsBox v3.3.0 (Götz et al. 2008) using BLASTx with "Standard" sensitivity and parameters (taxon 33090 Viridiplantae), which also merged EggNOG 5 (Huerta-Cepas et al. 2019) results with InterProScan v5.69-101.0 (Jones et al. 2014) and DiamondBlast blastx with the NR (2024-07-11) database (evalue 1e-3). Microsatellite analysis for the entire SB22 genome was carried out using MISA v2.1 (Thiel et al. 2003) using default parameters (minimum 7/di, 5/tri, 5/tetra, 5/penta and 5/hexa nucleotide repeats).

R gene prediction and analysis

Two disease resistance ("*R*") gene identification approaches were used: an *ab-initio* method, NLR-Annotator v2.1 (Steuernagel et al. 2020), which identifies NBS-LRR genes at the genome level using a motif-based search, and DRAGO3, which identifies *R* genes based on amino acid sequences of predicted gene models. The assembled genome was fed to NLR-Annotator and the predicted NBS-LRR loci were written to a bed file. The bed file was then used to extract the nucleotide sequences using seqtk v1.4 (https://github.com/lh3/seqtk) subseq function for further functional annotation using BLAST2GO v3.0 (Götz et al. 2008) within Omicsbox software. The predicted *R* gene loci were further subjected to gene predictions with MAKER2 to get final gene models with training sets from Genemark (Brůna et al 2024) and Augustus (Stanke et al 2006b) from BRAKER3 (Gabriel et al 2024) and snap training set obtained by training with final gene models along with "*R*" gene peptides obtained from 15 *S. lycopersicum* datasets (Seong et al. 2020), putative NLR genes in *Arabidopsis* (annotation version Araprot11) (Cheng et al. 2017), and experimentally validated NLR genes obtained from PRGdb 3.0 (Osuna-Cruz et al. 2018) and RefPlantNLR (Kourelis et al. 2021).

Predicted gene models from SB22 genome were subject to *R*-gene motif identification using DRAGO3 (Osuna-Cruz et al. 2018), which uses HMMER (Eddy, 2011) v3 package and 60 HMM modules along with CC and TM domains using COILS 2.2 (Lupas et al 1991) and TMHMM 2.0c (Krogh et al 2001) programs. In addition, CRKN resistance markers previously reported by (Bali et al. 2022) were mapped to the assembled SB22 genome to identify putative candidate genes.

Results

Genome Assembly

PacBio RSII sequencing yielded 39 Giga bases (Gb) of raw data for an estimated genome coverage of 36x while Illumina paired-end sequencing resulted in 120.4 Gb at 98x Coverage. Genome

size estimation by Jellyfish (Marçais and Kingsford, 2011) and GenomeScope (Vurture et al. 2017) gave an estimated haploid genome size of 741 Mb, with a heterozygosity of ~1.85% (Figure S1). Following the Falcon (Chin et al. 2016) assembly and error correction with Illumina reads using Pilon (Walker et al. 2014), an initial assembly consisting of 655.277 Mb was produced which had 1,313 contigs, ranging in size from 29.50 Mb to 805 bp (Table 1). The HiRise assembly (Dovetail Genomics, Scotts Valley, CA, USA) and further scaffolding analysis significantly increased contiguity, reducing the number of genome scaffolds to 403 (NCBI project id: PRJNA1147704), representing 12 chromosomes, and increasing the total assembly size to 656.37 Mb and N50=51.7 Mb. Alignment of whole genome scaffolds with DM6.1 and subsequent merging of the scaffolds of chromosome 1 and 10 by introducing N-mers resulted in 12 complete chromosome assemblies. A comparison of the genome assembly chromosome lengths revealed considerable similarities between the SB22 and SB11H21 assemblies, with the 12 chromosome-anchored scaffolds summing up to 637.94 Mb and 647 Mb, respectively. This represents a difference of approximately 9.06 Mb (1.40%). In contrast, the DM6.1 reference genome assembly contained 731.29 Mb of chromosome-anchored sequence. This computes to about 93.35 Mb longer than the SB22 genome. Notable differences were observed on chromosome 1, the longest chromosome which measured 71.92 Mb and 72.19 Mb in SB22 and SB11H21, respectively, when compared with 88.59 Mb in DM6.1 (Table S1). The shortest chromosome was chromosome 11, with a length of 42.10 Mb, 43.36 Mb and 46.78 Mb in SB22, SB11H21 and DM6.1, respectively. Chromosome 10 also showed a considerable variation in length measuring 47.76 Mb in SB22, 54.37 Mb in SB11H21 and 61.04 Mb in DM6.1 (Table S1). These findings are corroborated by the comparative genome and gene synteny analysis (Figure S4), which revealed broadly conserved chromosome-scale organization between SB22 and SB11H21 as well as other regions of structural variation. In contrast, there was a lower degree of synteny between the SB22 and DM6.1, suggesting a greater order of chromosome-scale organization between the two wild potato clones than with the DM6.1. The SB22 genome assembly had a high degree of genome assembly completeness using the Solanales BUSCO lineage dataset, with a reported BUSCO score of 95.8 % (Table 2) in genome mode. While slightly less than the reported genome BUSCO score of SB11H21 (98.27%), the SB22 genome assembly still captures a substantial proportion of the conserved gene landscape.

The synteny analysis showed that the SB22 genome assembly shared a high degree of synteny with the *S. tuberosum* reference DM6.1 and SB11H21 assemblies. Relative to both assemblies, the SB22 showed multiple chromosomal aberrations; specifically, it contained a higher number of chromosomal inversions than translocations relative to the DM6.1 and SB11H21 (Figure 1). A major translocation event was observed on the telomere of chromosome 5 in SB22 relative to SB11H21 and DM6.1. The synteny analysis also revealed varying degrees of duplication events across all 12 chromosomes. Chromosome 11, which has CRKN resistance markers, showed high synteny with both SB11H21 and DM6.1 across both chromosome arms, with two small inversions observed near the

centromere (Figure 1). Similar patterns of pericentromeric inversions were observed across multiple chromosomes in both DM6.1 and SB11H21.

A comparative analysis of the gene-level synteny using GENESPACE revealed an extensive conservation of chromosome-scale gene order among SB22, SB11H21 and DM6.1 (supplementary figures). Large scale syntenic blocks were observed among all 12 chromosomes, with a mostly one to one chromosomal synteny shared among the three assemblies. This conservation was more evident between the two *S. bulbocastanum* genomes (SB22 and SB11H21), although similar patterns were also observed with DM6.1. Some cross-chromosomal syntenic lines could also be observed. This suggests localized differences in genome organization or perhaps the presence of duplicated homologous regions. Overall, these findings indicate a substantial conservation of gene order among the three genomes, even with earlier noted differences in chromosome lengths as well as some of the structural rearrangements in the whole genome synteny analysis.

Repeat Analysis

De-novo repeat analysis with RepeatModeler and RepeatMasker estimated a repeat content of ~419 Mb, representing 63.84 % of the total assembled genome (Table 3). Retroelements represented 31.63 % of the genome with 185,113 repeat elements (Figure 2). LTR elements were the second most abundant, representing 29.27 % with 29,738 elements, followed by Gypsy/DIRS1 with 22.30% with 100,741 repeat elements.

Microsatellite prediction using MISA identified 62,429 microsatellites, with 18,612 in compound formation. The relative abundance of microsatellites was 95.2 per Mb in the SB22 genome. Di-nucleotide repeats were most abundant in the genome, representing 44.2% (27,649) (Figure 3), followed by tri-nucleotide and hepta-nucleotide repeats with 32.4 % (20,248) and 11.6 % (7,241), respectively. The frequency of AT repeat motifs was the most abundant in the genome, representing 87.1 % of total di-nucleotide repeats, followed by the tri-nucleotide repeat motif AAT, which represented 60.6 %.

Structural and Functional Annotation

The BRAKER gene prediction identified 38,077 gene models (putative genes/loci) and 57,200 transcripts, representing alternative splice isoforms derived from these gene models. BUSCO protein analysis indicated completeness of 96% (Solanales_odb10) with 49 % (2,913) complete single copy BUSCOs, 47% (2,799) complete duplicated BUSCOs, 1.2% (69) fragmented BUSCOs and 2.8 % (169) missing BUSCOs.

Functional annotation using OmicsBox resulted in 71% of gene models receiving GO annotations, with nearly 8.5% having no BLAST hits (Supplementary Figure S2). KEGG analysis

annotated a total of 17,227 (30.11%) transcripts (Figure 4), including Protein families: genetic information processing (3,590), genetic information processing (2884), carbohydrate metabolism (1,282) and protein families: signaling and cellular process (1,278).

Disease Resistance Genes

NLR-Annotator was used to identify NBS-LRR loci in the genome assemblies of SB22, SB11H21 and DM6.1. This analysis identified 576 NBS-LRR loci in the genome of SB22 compared with 562 in SB11H21 and 743 in DM6.1 (Table 2, Figure 5). Of these, 572, 562 and 729 were assigned to the 12 chromosomes of SB22, SB11H21 and DM6.1, respectively. Based on their conserved motif arrangements, 233 loci in SB22 were classified as CNL-like (CC-NBARC-LRR) and 58 as TNL-like (TIR-NBARC-LRR). This is compared with the 219 CNL-like and 61-TNL-like loci in SB11H21 and 312 CNL-like and 75 TNL-like loci in DM6.1. The remaining loci contained partial or alternative combinations of NLR associated motifs.

A chromosome-wise comparison of the putative *R* genes identified by NLR-Annotator showed an uneven distribution of R-genes across chromosomes in the three genome assemblies. In SB22, chromosomes 4 and 11 contained the highest numbers of predicted NLR loci with 114 and 104 loci, respectively. This represents about 38% of all NLRs on chromosomes in SB22. A similar observation was made in DM6.1, with 123 and 113 on chromosomes 4 and 11, respectively. Although, numerically comparable, SB11H21 had the smallest number of putative NLR loci on chromosomes 4 (82) and 6 (99), with chromosome 11 (115) containing the highest number. This number in SB11H21 was more than in both SB22 (67) and the DM6.1 (84). Still, chromosome 11 contained a high number of NLR loci across all three genomes, highlighting the abundance of putative resistance associated regions on the chromosomes. Thirty-five of these genes were clustered in a 3 Mb region of the distal end of chromosome 11. *R* genes were also seen to be enriched in chromosomes 1, 4, 5, 6 and 10, with more than 40 R gene loci per chromosome in SB22 (Figure S3).

To further characterize the predicted NLR regions in SB22, gene prediction using the MAKER2 pipeline was used. This analysis returned 744 putative gene models. Unlike the NLR-Annotator which only identifies genomic regions containing the characteristic NLR-like motifs, MAKER2 predicts protein-coding gene structures. The resulting gene models in the analysis, therefore, represent a complementary annotation of the resistance associated genomic regions, and not just a direct measure of the NLR locus abundance.

In addition to NLR-annotator, the finalized gene models predicted using BRAKER were analyzed with DRAGO3 to identify putative resistance related genes based on conserved protein domains.

This analysis identified 2,901 unique *R* genes with the resistance domain signatures distributed throughout the SB22 genome. This list included genes which code for NLR-like proteins (CNL and TNL), receptor-like proteins (RLP), receptor-like kinases (RLK), and other protein coding domains associated with resistance. Of these, 1,899 (65.5%) contained kinase domains, 952 (32.5%) contained LRR domains, and 361 (12.4%) contained NBS domains. Additionally, 314 candidates contained Coiled-Coil (CC) domains, and 64 had Toll/Interleukin Receptor-like (TIR) domains. In all, 140 candidates contained both NBS and LRR domains, including 36 classified as CNL, and 28 classified as TNL proteins. These results represent a complementary characterization of the putative resistance gene landscape in SB22

Mapping the positions of R genes relative to previously identified CRKN resistance markers on chromosome 11 of the SB22 genome assembly (Figure 6) placed the markers near 11 R gene loci (Table S2). Four of these loci, which encode TMV resistance-like proteins, lie 3 to 26.7kb from the CRKN resistance marker SB_MC1Chr11-SSR20. Additionally, a locus encoding an RPM1-like disease resistance protein was identified approximately 4.1kb from the marker SB_MC1Chr11-PotVar0066518. This proximity suggests that these R genes may contribute to CRKN resistance.

DISCUSSION

Here, we present the chromosome-level assembly of *S. bulbocastanum* (selection 22). We compared the *S. bulbocastanum* SB22 assembly with a previously published chromosome-level assembly (SB11H21; Hosaka et al. 2024) and with reported genome assembly metrics for a scaffold-level *S. bulbocastanum* assembly (Tang et al. 2022). BUSCO completeness in genome mode for the present assembly was 95.8 % with 90.4 % complete single copy and 5.4 % duplicated BUSCOs. Although this score is slightly lower than the 98.27% genome BUSCO score of SB11H21, the high BUSCO completeness of the SB22 assembly indicates that it is largely complete and suitable for our downstream comparative genomic analyses. The LAI index of 8.73 of SB22 is lower compared to other more complete genomes like DM6.1 which has an LAI score of 13.56 (Pham et al. 2020) but it is comparable to available assemblies of *S. bulbocastanum* genomes with LAI scores of 9.63 (SB11H21) and 8.97 (PG6241). The highly accurate genomes of rice and maize have LAI scores of 21.71 and 20.7 respectively (Ou et al. 2018; Wu et al. 2023). Gene prediction with BRAKER2-genomethreader, BRAKER2 and BRAKER3 and subsequent merging of gene prediction with TSEBRA resulted in 57,200 transcripts and 38,077 gene models which is close to 56,032 transcripts reported for PG6241 but higher than 35,523 gene models in 11H21. The BUSCO completeness of the SB22 genome annotation (96%) was higher than that of SB11H21 (91.76%) and comparable to the PG6241 (95.04%), likely reflecting differences in annotation pipelines, including the availability and use of RNA-seq and protein evidence. The SB22 genome annotation, however, exhibited a notably higher proportion of duplicated BUSCO genes (47%) compared to single-copy genes (49%), even though the assembly-level BUSCO

assessment showed predominantly single-copy genes (90.4% single-copy, 5.4% duplicated). This observed difference between the genome assembly and genome annotation BUSCO metrics could be the result of the retention of alternative transcript isoforms or the annotation of allelic variants as distinct gene models in this heterozygous wild species. The higher number of gene models in SB22 relative to SB11H21 may corroborate this position. Alternatively, this observation may point to technical aspects of the annotation pipeline applied to this heterozygous, wild diploid genome, providing an opportunity for future bioinformatic optimization. This pattern is, however, not anomalous in potato genome annotations. For example, the phased tetraploid genomes of 'Castle Russet' and 'Atlantic' reported duplicated BUSCO scores of 89.1 and 91.4% respectively for their high confidence gene models, although this in both tetraploid assemblies may point to gene duplications which are more common in polyploid genomes (Hoopes et al. 2022).

A comparison of the SB22 assembly with DM6.1 revealed several chromosomal aberrations in the genome, with genome inversions being more widespread. Previous studies have shown *that S. bulbocastanum* exhibits visible variations during cell division along with structural variations in the genome (Hermsen and Ramanna 1976; Jackson and Hanneman 1999). These structural variations agree with an earlier linkage map study indicating the presence of inversions in chromosomes of the *S. bulbocastanum* genome compared to *S. tuberosum* and tomato genomes (Andolfo et al. 2014).

The repeat scape in SB22 is similar to that of the DM6.1 genome, in which 66.8% (495.7Mb) consisted of repeat elements (Pham et al. 2020). The earlier assemblies published by (Tang et al. 2022) and (Hosaka et al. 2024) reported a repeat percentage of 59 % and 56.2 % in their assemblies which is slightly lower than 63.84 % reported in SB22. Retroelements and unclassified repeats represented similar proportions of repeat elements in DM6.1 (36.83 and 25.81 %) and SB22 (31.63% and 27.64 %) genomes. The slightly higher proportion of unclassified repeats in SB22 relative to DM6.1 may be due to the quality of repeat libraries used. Microsatellites are widely distributed in plant genomes, and they remain useful in plant breeding programs, fingerprinting and mapping studies, species discrimination, and marker-assisted selection programs (Subramanya et al. 2025). They can also be used for a variety of plant genetics applications due to their reproducibility, multiallelic nature, co-dominant inheritance, relative abundance, and good genome coverage (Remya et al. 2010). In the present study, we identified 62,429 microsatellites in the SB22 genome. The findings of our study are in line with the abundance of microsatellites in four other potato genomes (DM, RH, M6, and *S. commersonii*), reported earlier by Jian et al. 2021, where di-nucleotide repeats were reported to be the most abundant followed by tri-, tetra- and hepta-nucleotide repeats, with penta-nucleotide repeats being the least abundant. Our study reports ~1 microsatellite per ~10kb, which is less than the earlier study, perhaps due to the parameter set used for prediction and assembly quality.

Plant defense involves the recognition of pathogen associated molecules and effectors by plant immune receptors, including intracellular NLR proteins which are encoded by many R genes. The comparative analysis of the NLR landscape in SB22, SB11H21, and DM v6.1 revealed similarities and differences in the abundance and distribution of putative loci associated with resistance. NLR-Annotator identified 576 loci in SB22, compared with 562 in SB11H21 and 743 in DM6.1. In all three assemblies, CNL-like loci were more abundant than TNL-like loci. This is consistent with previous reports of NLR diversity in potato (Jupe et al. 2012; Tang et al. 2022). Tang et al. (2022) identified between 478 NLR genes in *S. morelliforme* and 1,976 in *S. chacoense*, suggesting a considerable diversification of the NLR landscape. Earlier studies also reported variations in NLR gene numbers in potato, with Jupe et al. (2012) identifying 438 NLR genes in the DM reference genome, which was subsequently expanded to 755 loci through resistance gene enrichment sequencing (RenSeq) (Jupe et al. 2013). These differences highlight the influence of genome assembly quality, gene annotation approaches, and detection methods on estimates of NLR gene content. A larger genome size does not always translate into a higher number of NLRs, and so, although DM6.1 contained more predicted NLR loci than either *S. bulbocastanum* assembly, this difference may reflect a combination of biological variation and methodological factors rather than differences in resistance gene content alone. The chromosomal distribution of NLR loci also revealed notable similarities and differences among the three assemblies. Chromosomes 4 and 11 were particularly enriched in predicted NLR loci, together accounting for approximately 32–38% of chromosome-assigned loci across the three genomes, despite representing only approximately 16% of their chromosome-anchored sequences. Chromosome 11 consistently contained a high number of NLR loci, with 104, 115, and 113 identified in SB22, SB11H21, and DM6.1, respectively. In contrast, more pronounced differences were observed on chromosomes 4 and 6, suggesting variation in the organization and abundance of NLR-associated regions between the two *S. bulbocastanum* accessions. These observations are consistent with previous reports that resistance genes are frequently organized in genomic clusters, which may undergo duplication, recombination, and sequence diversification (Jupe et al. 2012; Barragan and Weigel, 2021; Tang et al. 2022). However, further comparative analyses would be necessary to determine whether the differences observed in the present study reflect genuine variation in NLR gene content or differences in genome assembly and annotation.

The abundance of NLR loci on these chromosomes is particularly interesting given the previously characterized resistance genes and loci associated with several potato diseases and pests. For example, chromosome 4 contains the *R2* and *Rpi-blb3* resistance loci associated with late blight resistance, while chromosome 6 harbors *Rpi-blb2*, a late blight resistance gene originally identified in *S. bulbocastanum*. Similarly, chromosome 11 contains the characterized late blight resistance genes *R3a* and *R3b*, in addition to the previously mapped R_{Mc1}(blb) locus associated with resistance to CRKN (Torres Ascurra et al. 2026). Other important resistance loci include *Gro1-4* on chromosome 7, associated with resistance to *Globodera rostochiensis*, and *Gpa2* and *Rx* on chromosome 12, which confer

resistance to *G. pallida* and potato virus X, respectively (Gebhardt, 2023). The identification of NLR-rich regions in SB22 therefore provides an opportunity to investigate the conservation and diversification of resistance-associated genes through comparative genomics. Combining these NLR annotations with the syntenic relationships shared among SB22, SB11H21, and DM6.1 could facilitate the identification of conserved resistance gene families, accession-specific candidates, and genomic regions potentially associated with resistance traits. This approach is especially relevant to chromosome 11, where recent genetic studies suggest that CRKN root and tuber resistance factors occupy distinct genomic regions (Hu, 2025). The SB22 genome could provide a valuable framework for investigating additional candidate resistance genes within these regions, particularly those associated with tuber resistance that may not be adequately represented by markers targeting the previously characterized root resistance locus.

Some differences were also observed between the NLR annotations generated by NLR-Annotator and DRAGO3. While NLR-Annotator identified 576 putative NLR loci directly from the SB22 genomic sequence, DRAGO3 identified 140 candidate genes containing both NBS and LRR domains among the predicted protein sequences. It is worth mentioning that these values are not directly comparable because NLR-Annotator detects conserved motif arrangements independently of gene prediction, whereas DRAGO3 identifies resistance-associated protein domains within previously predicted gene models. Consequently, fragmented gene models, incomplete predictions, or NLR-associated regions lacking annotated protein-coding genes may contribute to the lower recovery of NBS-LRR candidates by DRAGO3. The clustered organization and sequence diversity of NLR genes can further complicate their annotation using conventional gene prediction approaches (Barragan and Weigel, 2021; Tang et al. 2022). Notwithstanding, the two approaches complement each other, with NLR-Annotator facilitating the identification of potential NLR-associated genomic regions and DRAGO3 providing broader characterization of resistance-associated domains in predicted proteins. Together, these analyses establish a foundation for further refinement of the SB22 resistance gene landscape and the identification of candidate genes for functional investigation.

The genetic basis of resistance to CRKN in *S. bulbocastanum* has been investigated extensively following the identification and introgression of resistance from SB22 into cultivated potato. Brown et al. (1996) initially mapped resistance to CRKN Race 1 reproduction in roots to a dominant locus, R_{Mc1}(blb), on chromosome 11. Their later studies showed that resistance to nematode reproduction in roots and resistance to tuber infection $R_{Mctuber}$(blb) could be expressed independently. Brown et al. (2009) later identified recombinant progeny exhibiting contrasting root and tuber resistance phenotypes and therefore proposed the existence of two distinct but linked resistance factors on chromosome 11, with an estimated genetic distance of 38 ± 5 cM. More recently, Bali et al. (2022) developed molecular markers associated with CRKN resistance, providing additional tools for marker-assisted selection. We

mapped those markers to chromosome 11 of the assembled SB22 genome and identified several putative CRKN resistance gene loci near the markers, thereby providing potential candidates for further exploration. Recent mapping studies using progeny derived from PA99N82-4 × AOR13260-3adg further support the genetic separation of root and tuber resistance, suggesting that both factors occupy distinct regions of chromosome 11 (Hu, 2025). These findings suggest that the genetic control underlying root and tuber resistance may be physically separated, highlighting the importance of exploring both resistance-associated regions independently seeing as genetic resistance to nematode reproduction in roots may not necessarily prevent tuber infection.

The chromosome-level SB22 genome assembly provides a valuable resource for further investigations into the genomic organization of these resistance regions. While previously developed molecular markers have facilitated the identification of candidate genes associated with root resistance, the evidence-supported physical separation of the tuber resistance region suggests that there may be other sources of resistance on chromosome 11. The combination of available genetic mapping data with the SB22 reference genome can support the discovery of genomic intervals associated with both root and tuber resistance, the identification of additional candidate resistance genes, and the characterization of their sequence and structural variation. Further improvements in assembly resolution and refined resistance gene prediction approaches can strengthen the characterization of these loci. Ultimately, these efforts can further the development of molecular markers which target both resistance factors, thereby facilitating the selection of potato breeding lines which combine both root and tuber resistance to CRKN.

The possibility that root and tuber resistance loci occupy physically separated regions of chromosome 11 has important implications for understanding CRKN resistance in potato. Although the resistance factors were previously shown to be genetically linked, their distinct chromosomal locations would suggest that resistance expressed in roots and tubers may involve separate genetic determinants rather than a single locus with tissue-dependent effects. This interpretation is also consistent with the previously observed differences in resistance responses to CRKN races and resistance-breaking pathotypes.

CONCLUSION

The draft chromosome scale genome assembly of SB22 (*S. bulbocastanum*) provides a valuable genomic resource for potato breeding programs aimed at improving disease resistance. This wild potato is an important source of resistance to various pests and diseases, including CRKN. The comparative genomics analyses with SB22 and DM6.1 revealed extensive conservation of gene order, alongside multiple variations in the distribution of loci associated with putative resistance genes. The NLR annotation also identified some R-gene enriched chromosomes, especially chromosomes 4 and 11. Three

CNL-like candidate genes were identified near previously characterized CRKN resistance markers on chromosome 11. These provide additional targets for further investigation. The availability of this genetic resource also provides opportunities to explore the distinct genetic regions associated with root and tuber resistance factors, as well as other additional candidate resistance genes. Further characterization and downstream functional validation of these candidates can facilitate the development of diagnostic molecular markers which will support the introgression and selection of resistance traits in the cultivated potato using both conventional and molecular breeding approaches.

DATA AVAILABILITY STATEMENT

The raw reads of SB22 can be found at NCBI BioProject ID PRJNA1147704.

The authors confirm that the data supporting the findings of this study are available within the article [and/or] its supplementary materials. Additional data can be requested from the corresponding author.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

DECLARATIONS

Ethics, Consent to Participate, and Consent to Publish declarations: not applicable

ACKNOWLEDGMENTS

The authors wish to thank Samuel Talbot for his contributions to some of the comparative genomics analyses in the manuscript

FUNDING

We thank The Northwest Potato Research Consortium and Oregon State University startup funds for Dr. Vidyasagar Sathuvalli for funding and supporting this project.

AUTHOR CONTRIBUTIONS

OO: Data curation, Formal analysis, Methodology, Writing - review & editing; SS: Data curation, Formal analysis, Methodology, Writing - original draft, Writing - review & editing; SY: Data curation, Methodology, Writing - review & editing; KV: Conceptualization, Data curation, Methodology, Project administration, Writing - review & editing; CB: Resources; VS: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing - original draft, Writing - review & editing

REFERENCES

Achakkagari, S. R., Bozan, I., Camargo-Tavares, J. C., McCoy, H. J., Portal, L., Soto, J., Bizimungu, B., Anglin, N. L., Manrique-Carpintero, N., Lindqvist-Kreuze, H., Tai, H. H., & Strömvik, M. V. (2024). The phased Solanum okadae genome and Petota pangenome analysis of 23 other potato wild relatives and hybrids. *Scientific data*, *11*(1), 454. https://doi.org/10.1038/s41597-024-03300-5

Andolfo G, Jupe F, Witek K, Etherington GJ, Ercolano MR, Jones JDG. 2014. Defining the full tomato NB-LRR resistance gene repertoire using genomic and cDNA RenSeq. BMC Plant Biol. 14(1):120. doi:10.1186/1471-2229-14-120.

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. American Potato Journal. 70(6):485–495. doi:10.1007/BF02849067.

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. Molecular Breeding : New Strategies in Plant Improvement. 42(3). doi:10.1007/S11032-022-01285-W. /pmc/articles/PMC10248663/ /pmc/articles/PMC10248663/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10248663/.

Bali S, Vining K, Gleason C, Motjahedi H, Brown CR, Sathuvalli V. 2019. Transcriptome profiling of resistance response to Meloidogyne chitwoodi introgressed from wild species Solanum bulbocastanum into cultivated potato. BMC Genomics. 20(1):1–18. doi:10.1186/S12864-019-6257-1/TABLES/3.

Barragan AC, Weigel D. 2021. Plant NLR diversity: the known unknowns of pan-NLRomes. The Plant Cell. 33(4):814–831. doi:10.1093/plcell/koaa002.

Brown CR. 1989. Comparison of Reproductive Efficiency of Meloidogyne chitwoodi on Solanum bulbocastanum in Soil and In Vitro Tests. Plant Disease. 73(12):957. doi:10.1094/PD-73-0957.

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/METRICS.

Brown CR, Mojtahedi H, Santo GS. 1999. Genetic Analysis of Resistance to Meloidogyne chitwoodi Introgressed from Solanum hougasii into Cultivated Potato. Journal of Nematology. 31(3):264.

Brown CR, Mojtahedi H, Zhang L-H, Riga E. 2009. Genetics and Resistance Independent Resistant Reactions Expressed in Root and Tuber of Potato Breeding Lines with Introgressed Resistance to Meloidogyne chitwoodi. 99(9):1085. doi:10.1094/PHYTO-99-9-1085.

Brown CR, Yang CP, Mojtahedi H, Santo GS, Masuelli R. 1996. RFLP analysis of resistance to Columbia root-knot nematode derived from Solanum bulbocastanum in a BC2 population. Theoretical and Applied Genetics. 92(5):572–576. doi:10.1007/BF00224560/METRICS.

Brown CR, Zhang L, Mojtahedi H. 2014. Tracking the R_{Mc1} Gene for Resistance to Race 1 of Columbia Root-Knot Nematode (Meloidogyne chitwoodi) in Three Mexican Wild Potato Species with Different Ploidies. American Journal of Potato Research. 91(2):180–185. doi:10.1007/S12230-013-9336-X.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. NAR Genomics and Bioinformatics. 3(1):1–11. doi:10.1093/NARGAB/LQAA108.

Brůna, T., Lomsadze, A., & Borodovsky, M. (2024). GeneMark-ETP significantly improves the accuracy of automatic annotation of large eukaryotic genomes. Genome research, 34(5), 757–768. https://doi.org/10.1101/gr.278373.123

Cheng C-Y, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. 2017. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. The Plant Journal. 89(4):789–804. doi:10.1111/tpj.13415.

Chin, C. S., Peluso, P., Sedlazeck, F. J., Nattestad, M., Concepcion, G. T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., Cramer, G. R., Delledonne, M., Luo, C., Ecker, J. R., Cantu, D., Rank, D. R., & Schatz, M. C. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nature methods, 13(12), 1050–1054. https://doi.org/10.1038/nmeth.4035.

Desaeger, J., Wram, C., & Zasada, I. (2020). New reduced-risk agricultural nematicides - rationale and review. Journal of nematology, 52, e2020-91. https://doi.org/10.21307/jofnem-2020-091

Devaux, A., Goffart, J. P., Petsakos, A., Kromann, P., Gatto, M., Okello, J., ... & Hareau, G. (2020). Global Food Security, Contributions from Sustainable Potato Agri-Food Systems. In: Campos, H., Ortiz, O. (eds) *The potato crop: Its agricultural, nutritional and social contribution to humankind* (pp. 3-35). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-28683-5_1

Dolničar P. 2021. Importance of Potato as a Crops and Practical Approaches to Potato Breeding. In: Dobnik D, Gruden K, Ramšak Ž, Coll A, editors. Solanum tuberosum: Methods and Protocols. New York, NY: Springer US. p. 3–20. [accessed 2024 Sep 28]. https://doi.org/10.1007/978-1-0716-1609-3_1.

Eddy S. R. (2011). Accelerated Profile HMM Searches. PLoS computational biology, 7(10), e1002195. https://doi.org/10.1371/journal.pcbi.1002195

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 9(1):1–14. doi:10.1186/1471-2105-9-18/TABLES/5.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences of the United States of America. 117(17):9451–9457. doi:10.1073/PNAS.1921046117/SUPPL_FILE/PNAS.1921046117.SAPP.PDF.

Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. 2021. TSEBRA: transcript selector for BRAKER. BMC Bioinformatics. 22(1):1–12. doi:10.1186/S12859-021-04482-0/FIGURES/3.

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. bioRxiv : the preprint server for biology, 2023.06.10.544449. https://doi.org/10.1101/2023.06.10.544449

Gebhardt C. (2023). A physical map of traits of agronomic importance based on potato and tomato genome sequences. *Frontiers in genetics*, *14*, 1197206. https://doi.org/10.3389/fgene.2023.1197206

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 WB, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biology. 20(1):1–13. doi:10.1186/S13059-019-1911-0/FIGURES/5.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research. 36(10):3420–3435. doi:10.1093/NAR/GKN176.

Graebner RC, Brown CR, Ingham RE, Hagerty CH, Mojtahedi H, Quick RA, Hamlin LL, Wade N, Bamberg JB, Sathuvalli V. 2018. Resistance to Meloidogyne chitwoodi Identified in Wild Potato Species. American Journal of Potato Research. 95(6):679–686. doi:10.1007/S12230-018-9674-9.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics (Oxford, England)*, *29*(8), 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Healey A, Furtado A, Cooper T, Henry RJ. 2014. Protocol: A simple method for extracting next-generation sequencing quality genomic DNA from recalcitrant plant species. Plant Methods. 10(1):1–8. doi:10.1186/1746-4811-10-21/FIGURES/4.

Hermsen JGT, Ramanna MS. 1976. Barriers to hybridization of Solanum bulbocastanum Dun. and S. Verrucosum Schlechtd. and structural hybridity in their F1 plants. Euphytica. 25(1):1–10. doi:10.1007/BF00041523.

Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M. (2016). BRAKER1: Unsupervised RNA-Seq-Based Genome Annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics (Oxford, England)*, *32*(5), 767–769. https://doi.org/10.1093/bioinformatics/btv661

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

Hosaka AJ, Sanetomo R, Hosaka K. 2024. A de novo genome assembly of Solanum bulbocastanum Dun., a Mexican diploid species reproductively isolated from the A-genome species, including cultivated potatoes. G3 Genes|Genomes|Genetics. 14(6). doi:10.1093/G3JOURNAL/JKAE080. https://dx.doi.org/10.1093/g3journal/jkae080.

Hu, S. (2025). Breeding, genetics and genomic analysis of Columbia Root-Knot nematode resistance in potatoes.: Oregon State University. Accessed on June 2026. Available at: https://ir.library.oregonstate.edu/concern/graduate_thesis_or_dissertations/1g05fm689

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S. K., Cook, H., Mende, D. R., Letunic, I., Rattei, T., Jensen, L. J., von Mering, C., & Bork, P. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic acids research, 47(D1), D309–D314. https://doi.org/10.1093/nar/gky1085

Jablonska B, Ammiraju JSS, Bhattarai KK, Mantelin S, de Ilarduya OM, Roberts PA, Kaloshian I. 2007. The Mi-9 Gene from Solanum arcanum Conferring Heat-Stable Resistance to Root-Knot Nematodes Is a Homolog of Mi-1. Plant Physiology. 143(2):1044–1054. doi:10.1104/pp.106.089615.

Jackson SA, Hanneman RE. 1999. Crossability between cultivated and wild tuber-and non-tuber-bearing Solanums. Euphytica. 109(1):51–67. doi:10.1023/A:1003710817938/METRICS.

Janssen GJW, Norel A van, Verkerk-Bakker B, Janssen R. 1995. Resistance to Meloidogyne chitwoodi, M. fallax and M. hapla in wild tuber-bearing Solanum spp. Euphytica: International Journal of Plant Breeding. 92(3):287–294. doi:10.1007/BF00037110/METRICS.

Jian Y, Yan W, Xu J, Duan S, Li G, Jin L. 2021. Genome-wide simple sequence repeat markers in potato: abundance, distribution, composition, and polymorphism. DNA Research. 28(6). doi:10.1093/DNARES/DSAB020. https://dx.doi.org/10.1093/dnares/dsab020.

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*(9), 1236–1240. https://doi.org/10.1093/bioinformatics/btu031

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

Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJA, Wright F, Sharma SK, Bolser D, Bryan GJ, Jones JDG, et al. 2012. Identification and localisation of the NB-LRR gene family within the potato genome. BMC Genomics. 13(1):1–14. doi:10.1186/1471-2164-13-75/FIGURES/5.

Jupe F, Witek K, Verweij W, Sliwka J, Pritchard L, Etherington GJ, Maclean D, Cock PJ, Leggett RM, Bryan GJ, et al. 2013. Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. Plant J. 76(3):530–544. doi:10.1111/tpj.12307.

King BA, Taberna JP. 2013. Site-Specific Management of Meloidogyne chitwoodi in Idaho Potatoes Using 1,3-Dichloropropene; Approach, Experiences, and Economics. Journal of Nematology. 45(3):202.

Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of molecular biology, 305(3), 567–580. https://doi.org/10.1006/jmbi.2000.4315

Kourelis J, Sakai T, Adachi H, Kamoun S. 2021. RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. PLOS Biology. 19(10):e3001124. doi:10.1371/journal.pbio.3001124.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 34(18):3094–3100. doi:10.1093/BIOINFORMATICS/BTY191.

Lupas, A., Van Dyke, M., & Stock, J. (1991). Predicting coiled coils from protein sequences. Science (New York, N.Y.), 252(5009), 1162–1164. https://doi.org/10.1126/science.252.5009.1162

Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. Current Protocols. 1(12):e323. doi:10.1002/CPZ1.323.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27(6):764–770. doi:10.1093/BIOINFORMATICS/BTR011.

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM. 1998. The root knot nematode resistance gene Mi from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell. 10(8):1307–1319. doi:10.1105/tpc.10.8.1307.

Mojtahedi, H., Brown, C. R., & Santo, G. S. (1995). Characterization of Resistance in a Somatic Hybrid of Solanum bulbocastanum and S. tuberosum, to Meloidogyne chitwoodi. *Journal of nematology*, *27*(1), 86–93.

Nishimura D. 2000. RepeatMasker. Biotech Software & Internet Report. 1(1–2):36–39. doi:10.1089/152791600319259.

Osuna-Cruz CM, Paytuvi-Gallart A, Donato AD, Sundesha V, Andolfo G, Cigliano RA, Sanseverino W, Ercolano MR. 2018. PRGdb 3.0: a comprehensive platform for prediction and analysis of plant

disease resistance genes. Nucleic Acids Research. 46(D1):D1197–D1201. doi:10.1093/NAR/GKX1119.

Ou, S., Chen, J., & Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic acids research*, *46*(21), e126. https://doi.org/10.1093/nar/gky730

Ou, S., & Jiang, N. (2018). LTR_retriever: A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. *Plant physiology*, *176*(2), 1410–1422. https://doi.org/10.1104/pp.17.01310

Pham, G. M., Hamilton, J. P., Wood, J. C., Burke, J. T., Zhao, H., Vaillancourt, B., Ou, S., Jiang, J., & Buell, C. R. (2020). Construction of a chromosome-scale long-read reference genome assembly for potato. *GigaScience*, *9*(9), giaa100. https://doi.org/10.1093/gigascience/giaa100

Putnam, N. H., O'Connell, B. L., Stites, J. C., Rice, B. J., Blanchette, M., Calef, R., ... & Green, R. E. (2016). Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome research, 26(3), 342-350.

Radcliffe, E. B., & Lauer, F. I. (1966). A Survey of aphid resistance in the tuber-bearing *Solanum* (Tourn.) *L. species*. https://bonnus.ulb.uni-bonn.de/permalink/49HBZ_ULB/hr53dr/alma990073180880206441

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nature Communications 2020 11:1. 11(1):1–10. doi:10.1038/s41467-020-14998-3.

Remya KS, Joseph S, Lakshmi PK, Akhila S. 2010. Microsatellites in varied arenas of research. Journal of Pharmacy and Bioallied Sciences. 2(2):141. doi:10.4103/0975-7406.67004.

Sacco MA, Koropacka K, Grenier E, Jaubert MJ, Blanchard A, Goverse A, Smant G, Moffett P. 2009. The Cyst Nematode SPRYSEC Protein RBP-1 Elicits Gpa2- and RanGAP2-Dependent Plant Cell Death. PLOS Pathogens. 5(8):e1000564. doi:10.1371/journal.ppat.1000564.

Seong K, Seo E, Witek K, Li M, Staskawicz B. 2020. Evolution of NLR resistance genes with noncanonical N-terminal domains in wild tomato species. New Phytologist. 227(5):1530–1543. doi:10.1111/nph.16628.

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.

Spooner, D.M., Ghislain, M., Simon, R. Jansky SH, Gavrilenko T. 2014. Systematics, Diversity, Genetics, and Evolution of Wild and Cultivated Potatoes. *Bot. Rev.* **80**, 283–383 (2014). https://doi.org/10.1007/s12229-014-9146-y

Stanke M, Schöffmann O, Morgenstern B, Waack S. (2006a). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics. 7(1):1–11. doi:10.1186/1471-2105-7-62/TABLES/2.

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., & Morgenstern, B. (2006b). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic acids research*, *34*(Web Server issue), W435–W439. https://doi.org/10.1093/nar/gkl200

Steuernagel, B., Witek, K., Krattinger, S. G., Ramirez-Gonzalez, R. H., Schoonbeek, H. J., Yu, G., Baggs, E., Witek, A. I., Yadav, I., Krasileva, K. V., Jones, J. D. G., Uauy, C., Keller, B., Ridout, C. J., & Wulff, B. B. H. (2020). The NLR-Annotator Tool Enables Annotation of the Intracellular Immune Receptor Repertoire. *Plant physiology*, *183*(2), 468–482. https://doi.org/10.1104/pp.19.01273

Sunil Subramanya, A.E., Antre, S.H., Ravikumar, R.L. *et al.* Comparative genomics analysis gives insights into differential microsatellite distribution in selected cereals and legumes. *Discov. Plants* **2**, 313 (2025). https://doi.org/10.1007/s44372-025-00389-9

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 2022 606:7914. 606(7914):535–541. doi:10.1038/s41586-022-04822-x.

Thiel T, Michalek W, Varshney RK, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). TAG Theoretical and applied genetics Theoretische und angewandte Genetik. 106(3):411–422. doi:10.1007/S00122-002-1031-0.

Torres Ascurra, Y. C., Schrammel, B., Monino-Lopez, D., Merino Martin, G., Wiedmer, J., & Vleeshouwers, V. G. A. A. (2026). Diversification and deployment of PRR and NLR immune receptors in potato. *The Plant journal : for cell and molecular biology*, *127*(1), e71002. https://doi.org/10.1111/tpj.71002

U.S. Department of Agriculture, National Agricultural Statistics Service. (2026, September). *Potatoes 2025 summary*. U.S. Department of Agriculture. https://esmis.nal.usda.gov/sites/default/release-files/796078/pots0926.pdf. Accessed October 2026.

U.S. Department of Agriculture, Agricultural Research Service. (n.d.). PI 275187, Solanum bulbocastanum Dunal, HAW 1584 [Germplasm accession record]. GRIN-Global, National Plant Germplasm System. Retrieved September 2025, from https://npgsweb.ars-grin.gov/gringlobal/accessiondetail?id=1206286

Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics (Oxford, England), 33(14), 2202–2204. https://doi.org/10.1093/bioinformatics/btx153

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PloS one. 9(11). doi:10.1371/JOURNAL.PONE.0112963. https://pubmed.ncbi.nlm.nih.gov/25409509/.

Wersch S van, Tian L, Hoy R, Li X. 2020. Plant NLRs: The Whistleblowers of Plant Immunity. Plant Communications. 1(1). doi:10.1016/J.XPLC.2019.100016. /pmc/articles/PMC7747998/ /pmc/articles/PMC7747998/?report=abstract https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7747998/.

Wilson MT. 1959. Americans Learn to Grow the Irish Potato. The New England Quarterly. 32(3):333. doi:10.2307/362827.

Wu D, Xie L, Sun Y, Huang Y, Jia L, Dong C, Shen E, Ye C-Y, Qian Q, Fan L. 2023. A syntelog-based pan-genome provides insights into rice domestication and de-domestication. Genome Biology 2023 24:1. 24(1):1–22. doi:10.1186/S13059-023-03017-5.

Xu, Z., & Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic acids research*, *35*(Web Server issue), W265–W268. https://doi.org/10.1093/nar/gkm286

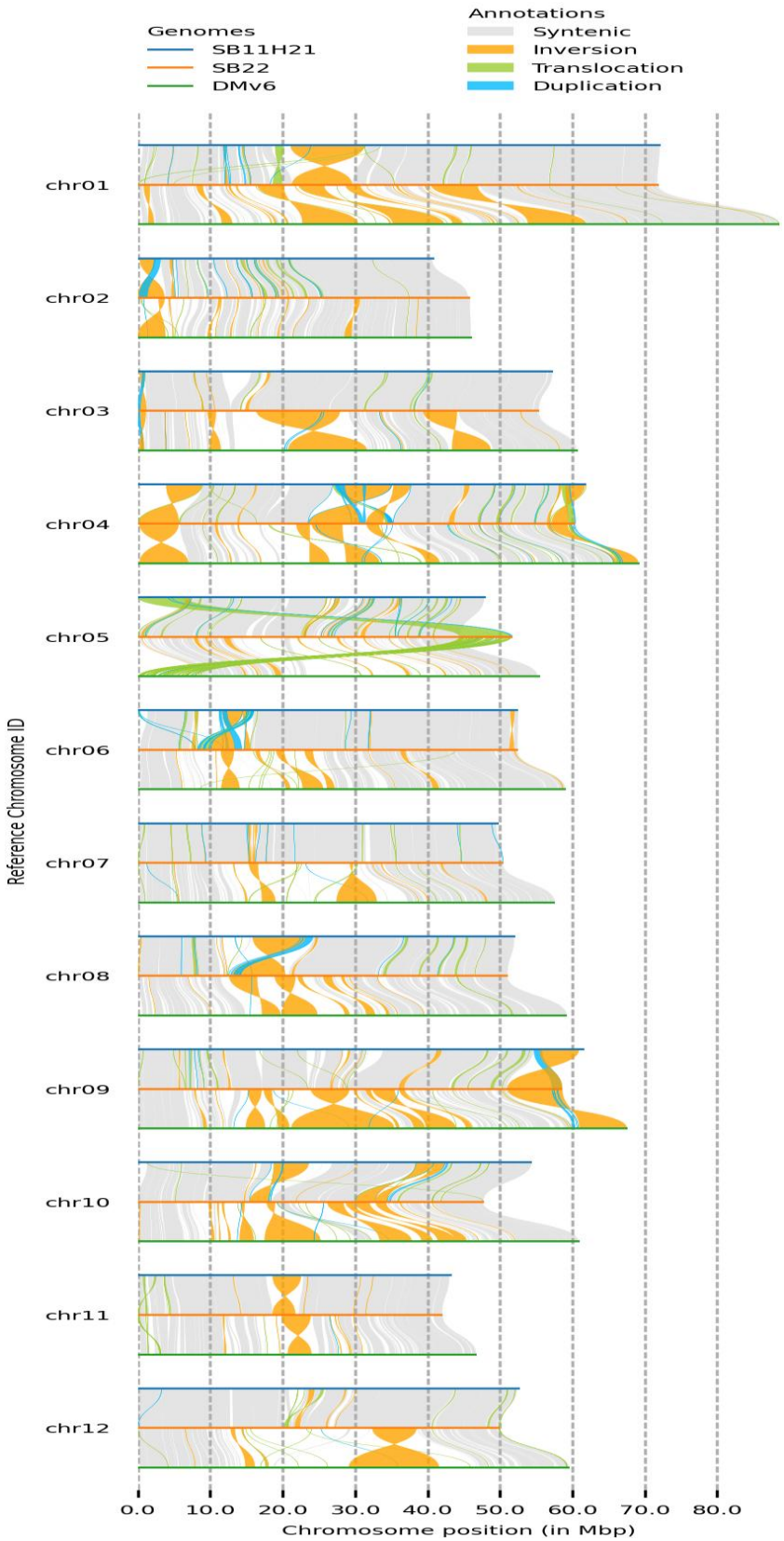
Yang X, Zhang L, Guo X, Xu J, Zhang K, Yang Yinqing, Yang Yu, Jian Y, Dong D, Huang S, et al. 2023. The gap-free potato genome assembly reveals large tandem gene clusters of agronomical importance in highly repeated genomic regions. Mol Plant. 16(2):314–317. doi:10.1016/j.molp.2022.12.010.

768 Zhang LH, Mojtahedi H, Kuang H, Baker B, Brown CR. 2007. Marker-assisted selection of Columbia
769 root-knot nematode resistance introgressed from Solanum bulbocastanum. Crop Science. 47(5):2021–
770 2026. doi:10.2135/CROPSCI2007.01.0003.

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

774

775 **FIGURES**



776

777 **Figure 1.** SyRI plot showing the genomic synteny among SB22, SB11H21 and DM6.1 genomes for the
778 12 main chromosomes

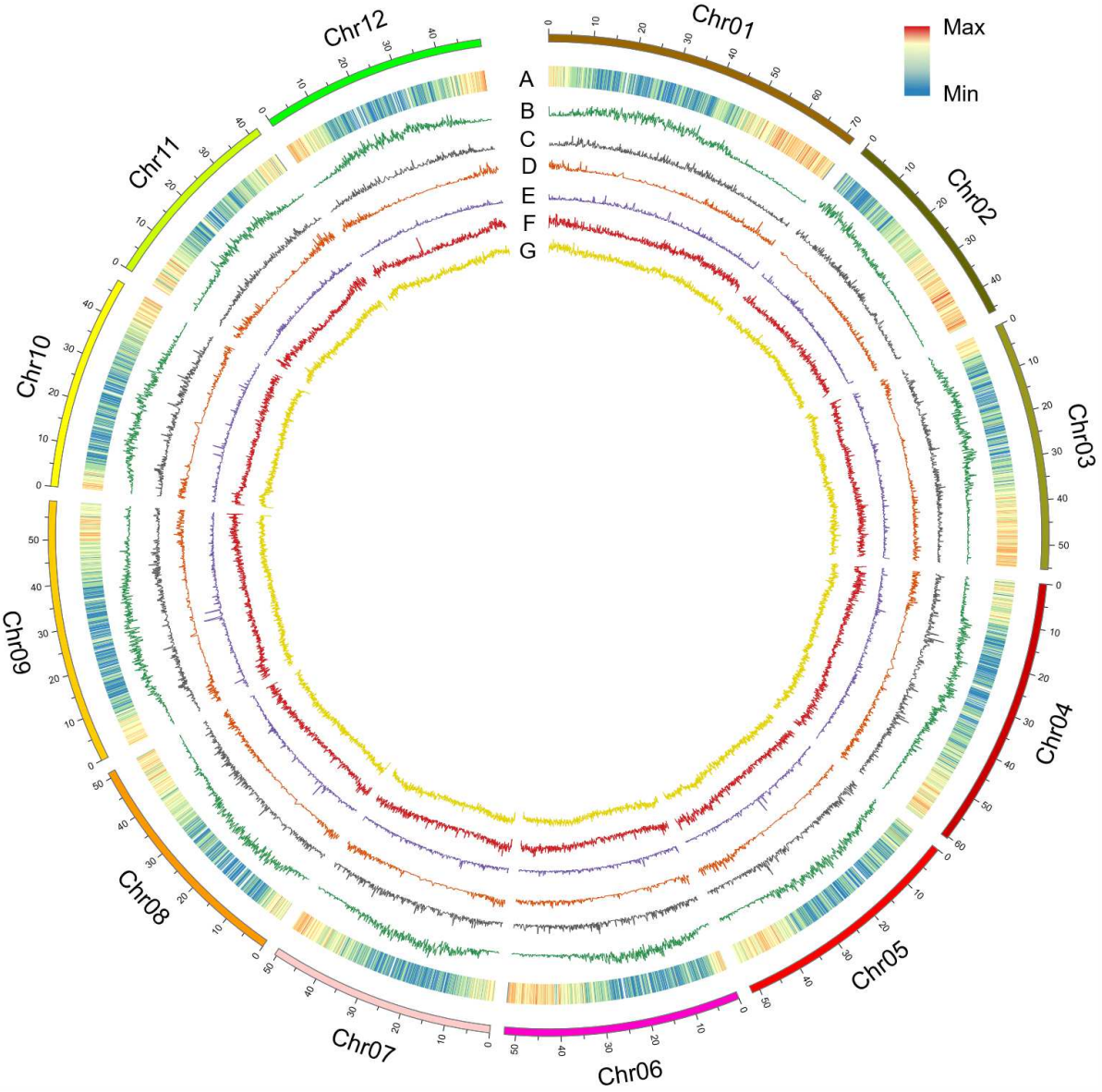
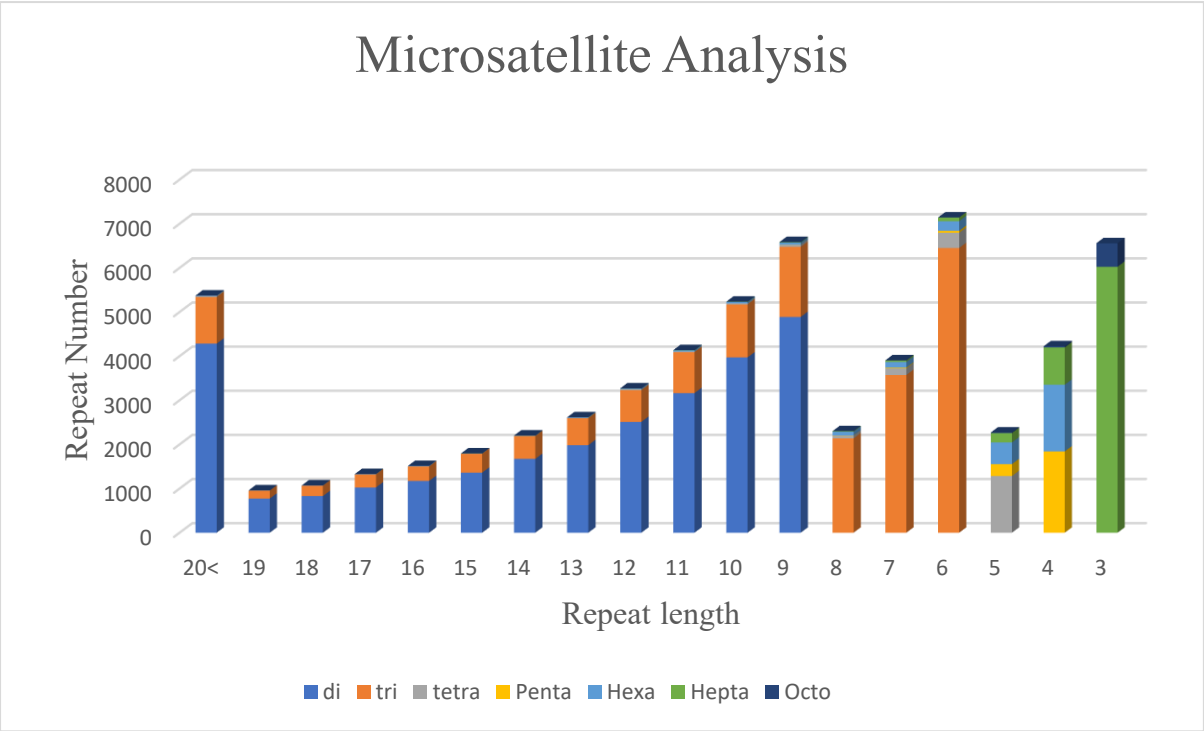


Figure 2. Genomic Features of SB22 Genome. From the outer circle gene density, repeat positions and positions of CNL and TNL genes are identified in the chromosomes. A - Gene Models, B - LTR-Gypsy, C - LTR-Copia, D - LINE-RTE-BovB, E - LINE-L1, F - Simple_repeats, G – unknown repeats



783

784 **Figure 3.** Repeat motif, length, and total number of microsatellites identified in the SB22 genome using

785 MISA program.

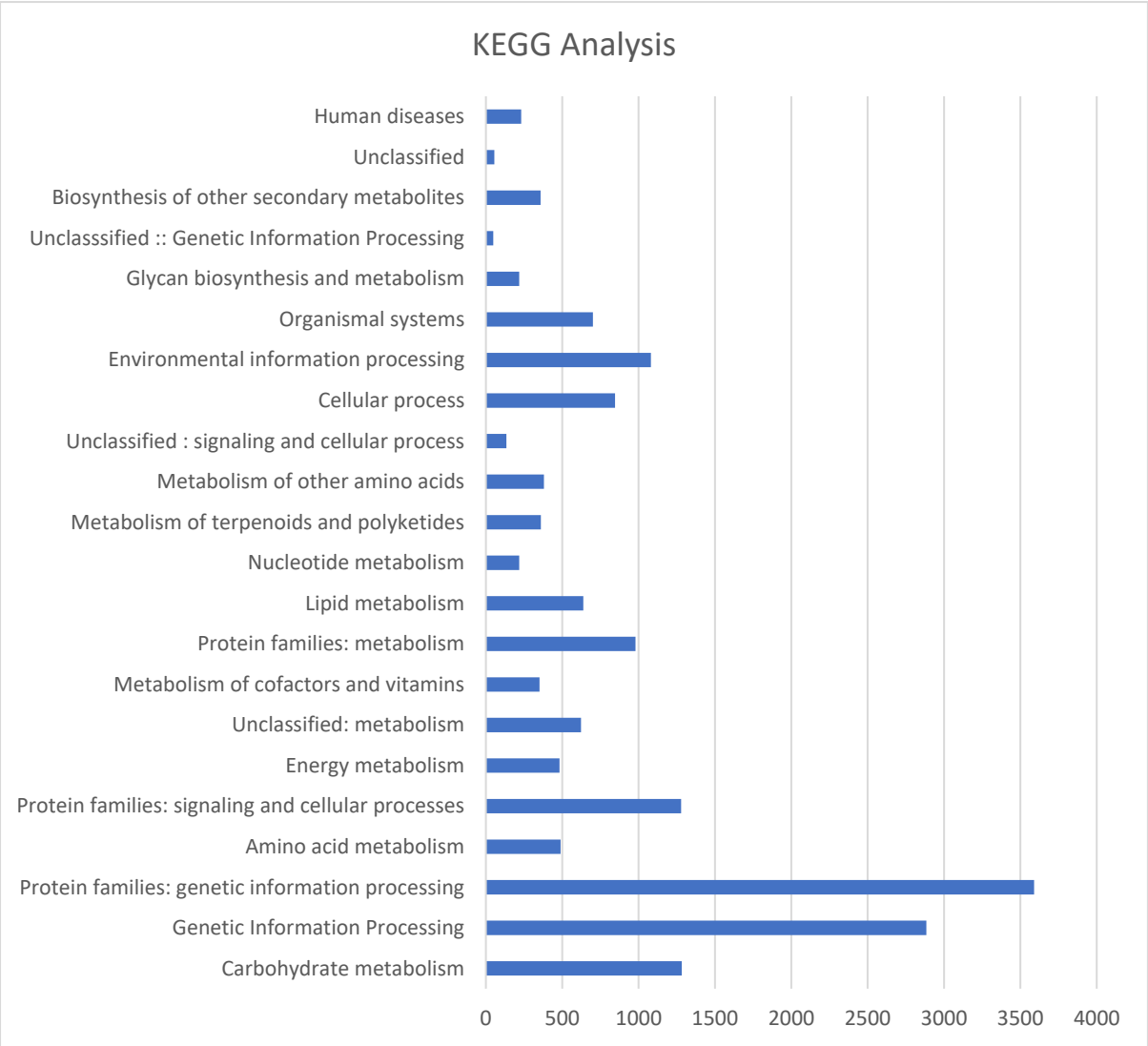


Figure 4. Distribution of SB22 transcripts across various gene families annotated using KEGG database

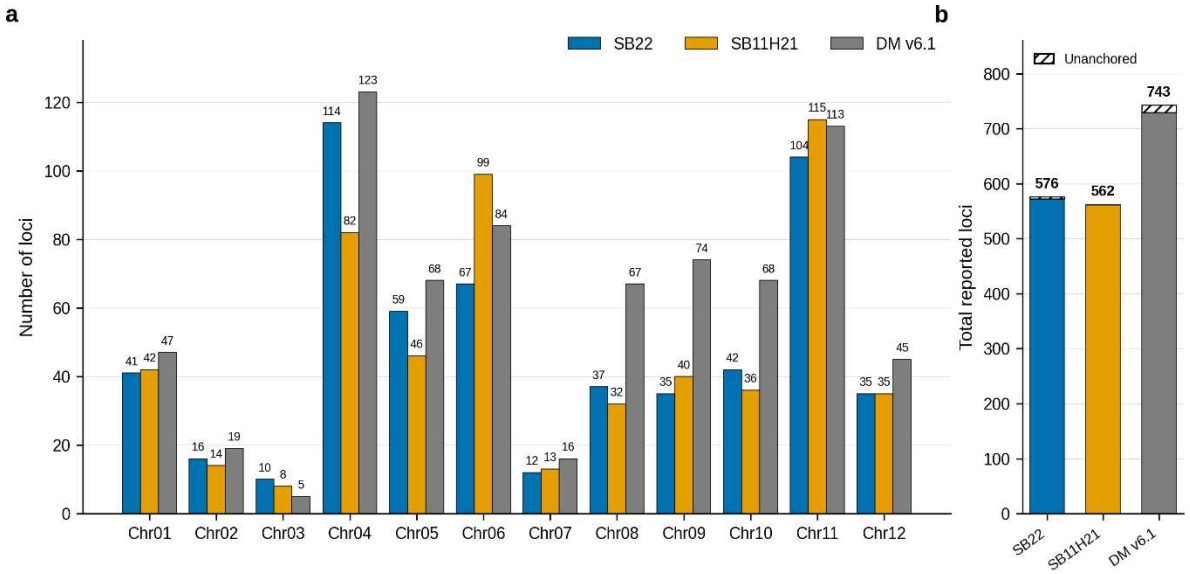
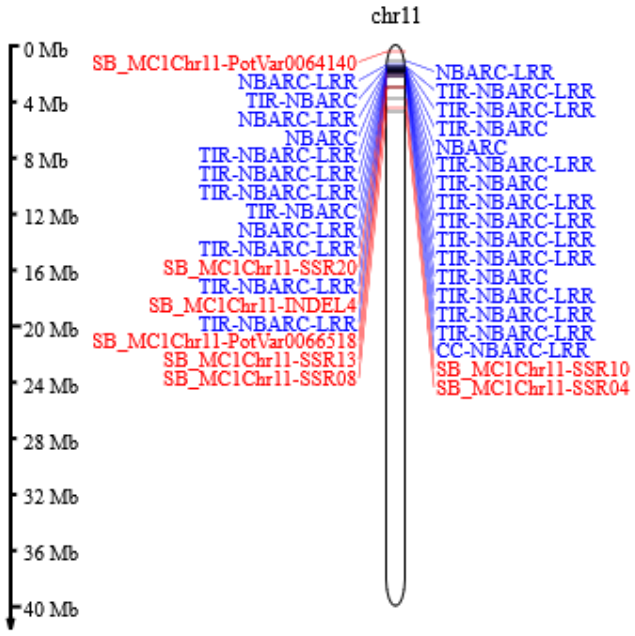


Figure 5. Total number of *R*-genes predicted per chromosome for *S. bulbocastanum* assemblies (SB22, SB11H21) and *S. tuberosum* reference genome DM6.1

Panel a shows the chromosome distribution across the 12 chromosomes, and panel b represents the totals (including those identified in non-chromosome anchored scaffolds).



795

796 **Figure 6.** Positions of fine-mapped markers linked to CRKN root resistance and R genes identified in
797 Chromosome 11 of SB22. Blue text indicates NLR genes and red text indicates CRKN resistance
798 markers.

799

800 **Table 1.** Assembly metrics of *S. bulbocastanum.*

Main genome scaffold total:	403
Main genome contig total:	1313
Main genome scaffold sequence total:	656.37 Mb
Main genome contig sequence total:	656.77 Mb (0.015% gap)
Main genome scaffold N50:	51.74 Mb
Main genome contig N50:	2.21 Mb
Max scaffold length:	71.91 Mb
Length of unplaced scaffolds:	18 Mb
Ns per 100 kbp	14.71

801

802 **Table 2.** BUSCO analysis of SB22 genome against Solanales-odb10 database.

	No of Busco's	Completeness
Genome assembly	5703 (95.8%)	Complete BUSCOs (C)
	5379 (90.4%)	Complete and single-copy BUSCOs (S)
	324 (5.4%)	Complete and duplicated BUSCOs (D)
	58 (1.0%)	Fragmented BUSCOs (F)
	189 (3.2%)	Missing BUSCOs (M)
	5,950	Total BUSCO groups searched
Genome annotation	No of Busco's	Completeness
	5712 (96.0%)	Complete BUSCOs (C)
	2913 (49.0%)	Complete and single-copy BUSCOs (S)
	2799 (47.0%)	Complete and duplicated BUSCOs (D)
	69 (1.2%)	Fragmented BUSCOs (F)
	169 (2.8%)	Missing BUSCOs (M)
	5,950	Total BUSCO groups searched

803

804 **Table 3**. Abundance of repeats in categories in the SB22 genome predicted by Repeat mod-
805 eler and Repeat masker.

	No. of repeat elements*	length occupied	Sequence percentage
Retroelements	**185,113**	**207,642,457 bp**	**31.63 %**
SINEs:	1,599	414,562 bp	0.06 %
Penelope	815	147,112 bp	0.02 %
LINEs:	38,422	15,110,171 bp	2.3 %
RTE/Bov-B	18,265	4,044,098 bp	0.62 %
L1/CIN4	19,342	10,918,961 bp	1.66 %
LTR elements:	**1,45,092**	**192,117,724 bp**	**29.27 %**
BEL/Pao	0	0 bp	0 %
Ty1/Copia	29,738	24,797,547 bp	3.78 %
Gypsy/DIRS1	100,741	146,399,589 bp	22.3 %
Retroviral	613	131,712 bp	0.02 %
DNA transposons	**36,231**	**19,457,070 bp**	**2.96 %**
hobo-Activator	6,646	4,251,814 bp	0.65 %
Tc1-IS630-Pogo	1,394	667,839 bp	0.1 %
Tourist/Harbinger	5,110	1,619,510 bp	0.25 %
Rolling-circles	**6,513**	**2,552,644 bp**	**0.39 %**
Unclassified:	**651,554**	**181,390,620 bp**	**27.64 %**
Small RNA:	**2,696**	**592,784 bp**	**0.09 %**
Satellites:	**0**	**0 bp**	**0 %**
Simple repeats:	**108,396**	**5,356,117 bp**	**0.82 %**
Low complexity:	**20,636**	**1,990,654 bp**	**0.3 %**
		Total repeats	**63.84%**

806

807 **Table 4.** Comparison of numbers of R genes identified using NLR-Annotator in SB22,
808 SB11H21 and DM6.1
809
810 **NLR motif classification**

Classification	SB22	SB11H21	DM6.1
CC-NBARC-LRR (CNL-like)	233	219	312
TIR-NBARC-LRR (TNL-like)	58	61	75
Others (and partial NLR motif configurations)	285	282	356
Total	576	562	743

811

Table S1: Genome assembly chromosome-level comparison of SB22, SB11H21 and DM6.1

Chromosome	SB22	SB11H21	DM v6.1	SB22 - SB11H21 (chromosome differential)	SB22 − DM v6.1 (chromosome differential)
Chr 01	71.92	72.19	88.59	- 0.27	- 16.68
Chr 02	45.85	40.94	46.10	+ 4.91	- 0.25
Chr 03	55.38	57.31	60.71	- 1.94	- 5.33
Chr 04	60.44	61.93	69.24	- 1.49	- 8.79
Chr 05	51.74	48.12	55.60	+ 3.63	- 3.86
Chr 06	52.51	52.53	59.09	- 0.02	- 6.58
Chr 07	50.45	49.78	57.64	+ 0.67	- 7.19
Chr 08	51.17	52.11	59.23	- 0.94	- 8.06
Chr 09	58.62	61.63	67.60	- 3.01	- 8.98
Chr 10	47.76	54.37	61.04	- 6.61	- 13.28
Chr 11	42.10	43.36	46.78	- 1.26	- 4.67
Chr 12	49.99	52.73	59.67	- 2.74	- 9.68
Total	637.94	647.00	731.29	- 9.06	- 93.35

+ sign represents a relatively larger chromosome size for SB22. - sign represents a relatively smaller chromosome size for SB22.

Table S2. Functional annotation of *R* gene transcripts identified adjacent to the CRKN resistance markers in chromosome 11 of the SB22 genome.

ID	NLR type	Start	Stop
chr11_nlr104	NBARC-LRR	1172067	1174021
chr11_nlr1	NBARC-LRR	1453952	1455890
chr11_nlr2	TIR-NBARC-LRR	1573811	1578549
chr11_nlr3	TIR-NBARC	1587199	1588805
chr11_nlr4	TIR-NBARC-LRR	1612378	1616883
chr11_nlr5	NBARC-LRR	1623307	1624498
chr11_nlr6	TIR-NBARC	1649638	1651011
chr11_nlr7	NBARC	1712800	1713118
chr11_nlr8	NBARC	1737487	1738355
chr11_nlr9	TIR-NBARC-LRR	1788724	1793327
chr11_nlr10	TIR-NBARC-LRR	1802347	1807758
chr11_nlr11	TIR-NBARC-LRR	1811291	1816370
chr11_nlr12	TIR-NBARC	1834657	1836284
chr11_nlr13	TIR-NBARC-LRR	1845712	1850658
chr11_nlr14	TIR-NBARC-LRR	1856545	1859323
chr11_nlr15	TIR-NBARC	1872971	1874637
chr11_nlr16	TIR-NBARC-LRR	1901730	1904488
chr11_nlr17	NBARC-LRR	1912070	1913872
chr11_nlr18	TIR-NBARC-LRR	1923368	1928056
chr11_nlr19	TIR-NBARC-LRR	1931242	1935146
chr11_nlr20	TIR-NBARC-LRR	1995882	2002224
chr11_nlr21	TIR-NBARC	2008785	2010403
chr11_nlr22	TIR-NBARC-LRR	2017962	2020464
chr11_nlr23	TIR-NBARC-LRR	2032017	2036879
chr11_nlr24	TIR-NBARC-LRR	2256366	2261008
chr11_nlr25	TIR-NBARC-LRR	2261608	2267904
chr11_nlr26	TIR-NBARC-LRR	2273263	2276843
chr11_nlr103	CC-NBARC-LRR	3065472	3067844
chr11_nlr102	NBARC-LRR	5263377	5265814
chr11_nlr27	CC-NBARC-LRR	6024535	6026308
chr11_nlr101	CC-NBARC-LRR	6611542	6613762
chr11_nlr100	CC-NBARC-LRR	6621943	6624164
chr11_nlr99	NBARC-LRR	6638328	6640003
chr11_nlr98	NBARC-LRR	6667587	6669457
chr11_nlr97	NBARC	6672474	6672915

chr11_nlr96	CC-NBARC-LRR	6682563	6684762
chr11_nlr95	CC-NBARC-LRR	6698035	6700258
chr11_nlr94	NBARC	6712645	6713186
chr11_nlr93	CC-NBARC-LRR	6716791	6718993
chr11_nlr92	CC-NBARC	6772354	6773615
chr11_nlr91	NBARC	6785978	6786521
chr11_nlr90	CC-NBARC-LRR	6790145	6792348
chr11_nlr89	NBARC-LRR	6804020	6805865
chr11_nlr28	CC-NBARC	6914705	6915959
chr11_nlr29	CC-NBARC-LRR	6926771	6929223
chr11_nlr30	NBARC	6943900	6944335
chr11_nlr31	CC-NBARC-LRR	6959177	6961389
chr11_nlr32	NBARC	6965572	6967584
chr11_nlr33	NBARC-LRR	6982250	6984168
chr11_nlr34	CC-NBARC-LRR	6987857	6989646
chr11_nlr35	NBARC	7001918	7002674
chr11_nlr36	CC-NBARC-LRR	7006884	7009121
chr11_nlr37	CC-NBARC-LRR	7011583	7013761
chr11_nlr88	TIR-NBARC-LRR	7893460	7896388
chr11_nlr38	TIR-NBARC-LRR	10354109	10358483
chr11_nlr39	CC-NBARC-LRR	12378386	12381391
chr11_nlr87	CC-NBARC-LRR	15489818	15492227
chr11_nlr86	NBARC-LRR	16481264	16483161
chr11_nlr85	TIR-NBARC	17883396	17886066
chr11_nlr40	TIR-NBARC-LRR	18456913	18460589
chr11_nlr84	CC-NBARC-LRR	32429033	32431855
chr11_nlr83	NBARC-LRR	32971308	32972247
chr11_nlr82	CC-NBARC-LRR	34077011	34080593
chr11_nlr81	NBARC-LRR	34113834	34114431
chr11_nlr80	TIR	35129050	35129413
chr11_nlr41	NBARC-LRR	35802477	35805475
chr11_nlr42	CC-NBARC-LRR	38376844	38379151
chr11_nlr43	TIR-CC-NBARC	38836273	38837089
chr11_nlr44	CC-NBARC-LRR	38839154	38842658
chr11_nlr45	CC-NBARC-LRR	39415238	39418820
chr11_nlr46	CC-NBARC-LRR	39439567	39443074
chr11_nlr47	CC-NBARC-LRR	39456053	39459569
chr11_nlr48	CC-NBARC-LRR	39485053	39487244
chr11_nlr49	CC-NBARC-LRR	39493993	39497515

chr11_nlr79	CC-NBARC-LRR	39722486	39725837
chr11_nlr50	CC-NBARC-LRR	39881889	39890913
chr11_nlr51	CC-NBARC-LRR	39894969	39898437
chr11_nlr52	CC-NBARC-LRR	39919834	39923641
chr11_nlr53	CC-NBARC-LRR	39931484	39934898
chr11_nlr54	CC-NBARC-LRR	39950631	39957880
chr11_nlr78	CC-NBARC-LRR	40436861	40440272
chr11_nlr77	NBARC-LRR	40471798	40474840
chr11_nlr76	NBARC-LRR	40477033	40477579
chr11_nlr75	CC-NBARC-LRR	40495178	40498520
chr11_nlr74	CC-NBARC-LRR	40503370	40506154
chr11_nlr73	CC-NBARC-LRR	40540417	40543837
chr11_nlr72	CC-NBARC-LRR	40570862	40573445
chr11_nlr71	CC-NBARC-LRR	40583779	40587173
chr11_nlr70	NBARC	40614727	40614889
chr11_nlr69	CC-NBARC-LRR	40653443	40656908
chr11_nlr68	CC-NBARC-LRR	40666685	40669751
chr11_nlr67	NBARC-LRR	40695456	40697247
chr11_nlr66	NBARC-LRR	40705337	40708376
chr11_nlr65	NBARC	40712331	40712708
chr11_nlr64	CC-NBARC-LRR	40720465	40723786
chr11_nlr63	CC-NBARC-LRR	40742528	40745108
chr11_nlr62	CC-NBARC-LRR	40777695	40781112
chr11_nlr61	NBARC	40784706	40785081
chr11_nlr60	CC-NBARC-LRR	40792864	40796415
chr11_nlr55	CC-NBARC-LRR	41118588	41122131
chr11_nlr56	CC-NBARC-LRR	41127442	41130970
chr11_nlr59	NBARC-LRR	41158609	41161626
chr11_nlr58	CC-NBARC-LRR	41270745	41274291
chr11_nlr57	CC-NBARC-LRR	41280402	41284014

834

835

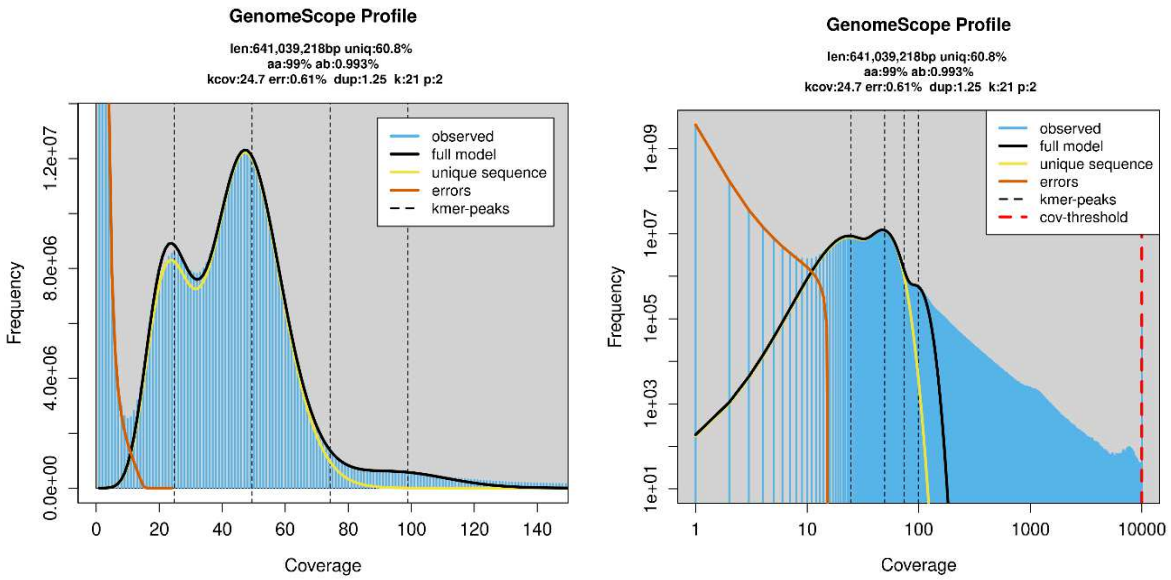


Figure S1. linear plot and log plot created using K-mer (21) based genome size and heterozygosity estimation using *Solanum bulbocastanum* raw Illumina Hiseq reads with Genoscope2 web-program.

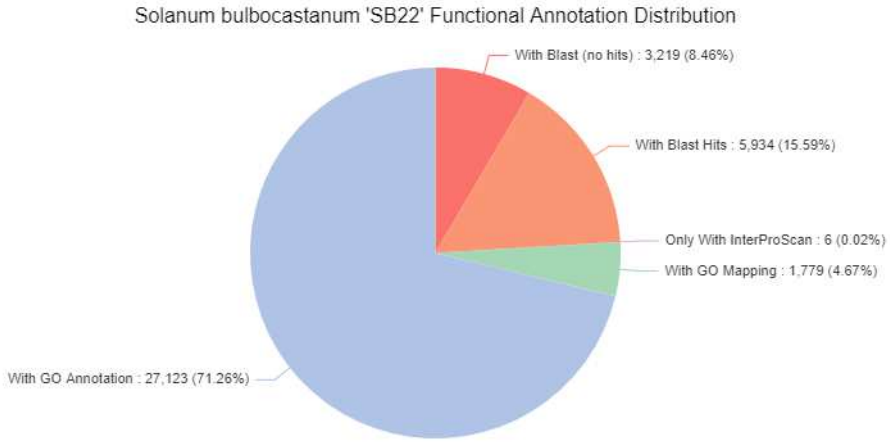


Figure S2. 'SB22' Functional Annotation Results by OmicsBox.

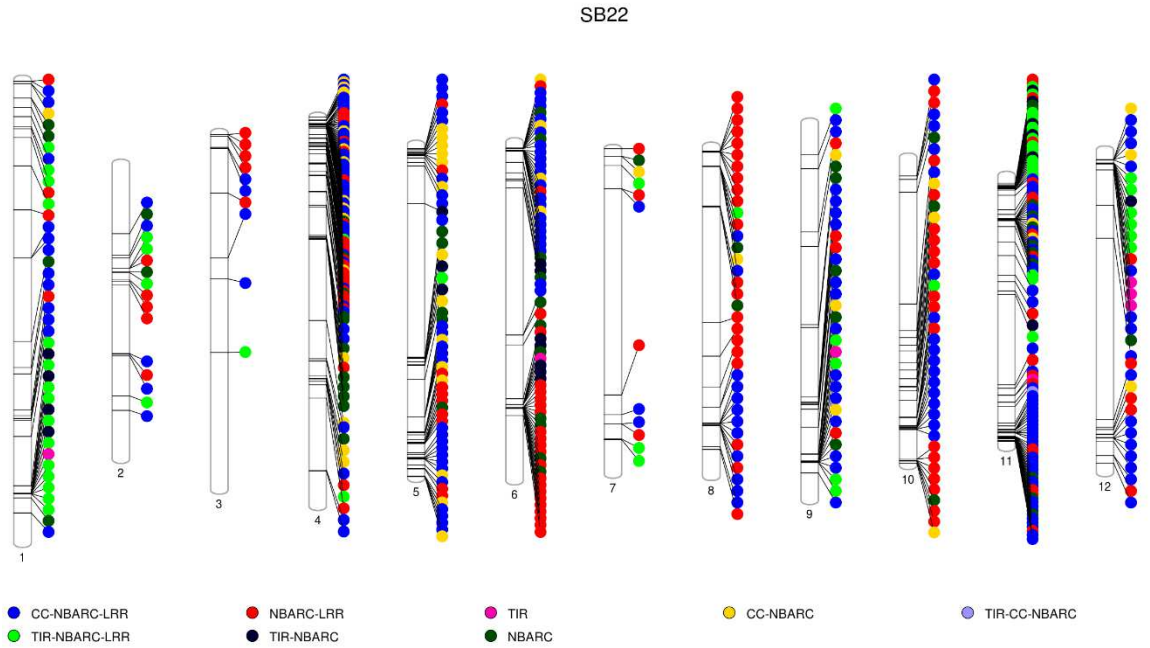


Figure S3. Positions of R gene loci predicted using NLR-Annotator prediction and SB22 genome across the 12 chromosomes.

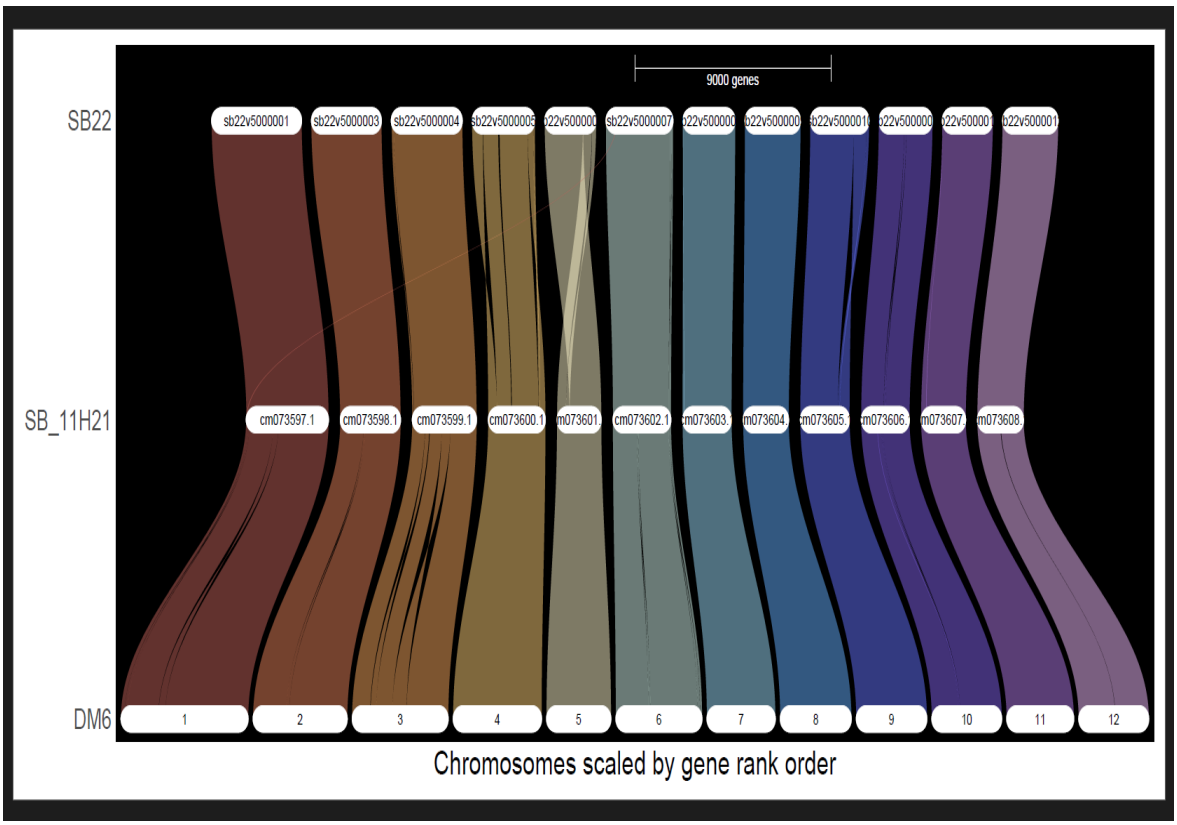


Figure S4. Genespace riparian plot of SB22, SB11H21 and DMv6 based on the DMv6 gene order

Comparative gene-level synteny among the *Solanum bulbocastanum* SB22 and SB11H21 genome assemblies and the *S. tuberosum* DM v6.1 reference genome. The GENESPACE riparian plot illustrates conserved syntenic gene blocks across the 12 chromosomes of the three assemblies. Colored ribbons represent corresponding syntenic regions, while chromosome positions are scaled according to gene rank order rather than physical sequence length. Broad conservation of chromosome-scale gene organization is evident across the three assemblies, with relatively few cross-chromosomal syntenic connections

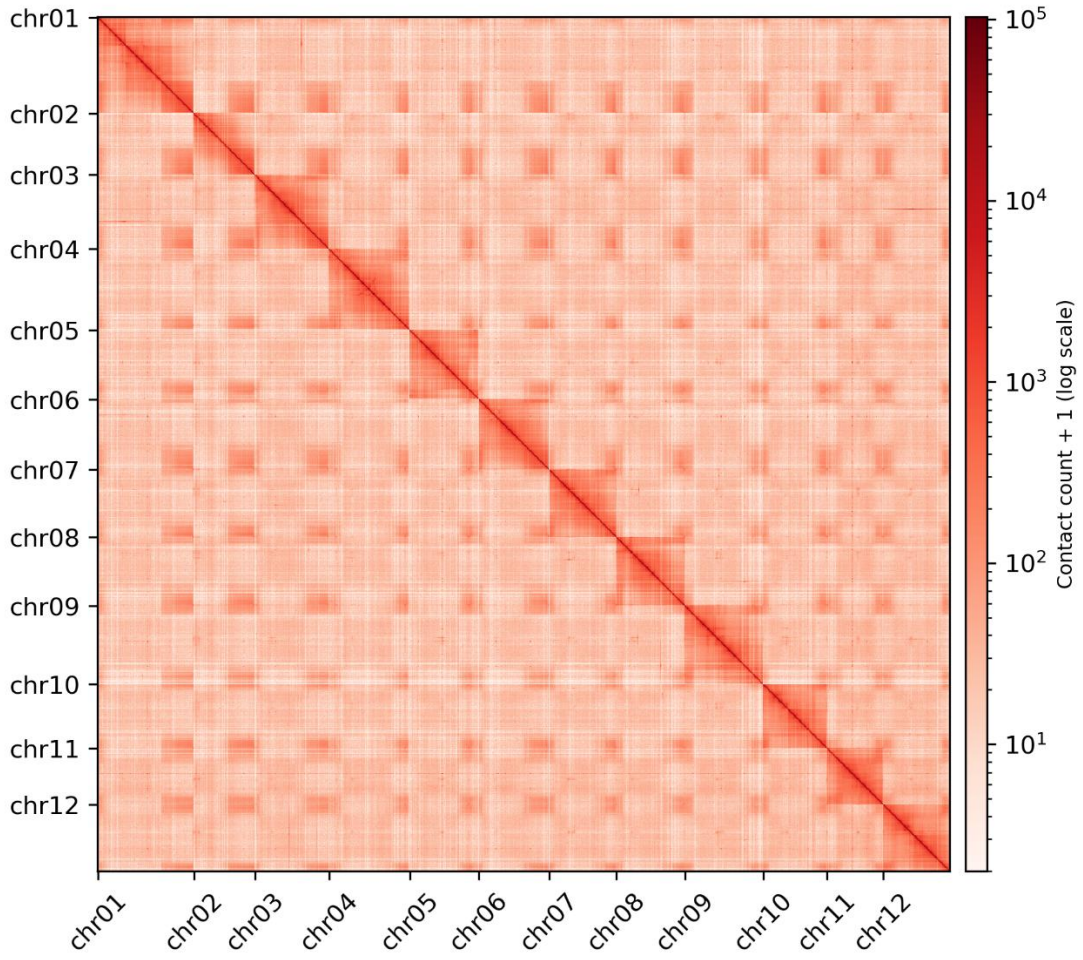


Figure S5. Link density histogram of SB22. X and Y axes show the mapping positions of the first and second read pair respectively, grouped into bins for the 12 chromosomes. Each square gives the number of read pairs within that bin. Scaffolds less than 1 Mb are excluded