



OPEN ACCESS

EDITED BY

Stephen Moose,
University of Illinois at Urbana-Champaign,
United States

REVIEWED BY

David Jarvis,
Brigham Young University, United States
Rie Shimizu-Inatsugi,
University of Zurich, Switzerland

*CORRESPONDENCE

Yasunari Fujita
✉ fujitay0434@jircas.go.jp

†PRESENT ADDRESS

Hideki Hirakawa,
Department of Bioscience and
Biotechnology, Faculty of Agriculture, Kyushu
University, Fukuoka, Japan
Kenichiro Fujii,
Division of Crop Genome Editing, Institute of
Agrobiological Sciences, National Agriculture
and Food Research Organization, Tsukuba,
Japan

RECEIVED 17 May 2024

ACCEPTED 22 July 2024

PUBLISHED 19 August 2024

CITATION

Kobayashi Y, Hirakawa H, Shirasawa K,
Nishimura K, Fujii K, Oros R, Almanza GR,
Nagatoshi Y, Yasui Y and Fujita Y (2024)
Chromosome-level genome assemblies
for two quinoa inbred lines from
northern and southern highlands of
Altiplano where quinoa originated.
Front. Plant Sci. 15:1434388.
doi: 10.3389/fpls.2024.1434388

COPYRIGHT

© 2024 Kobayashi, Hirakawa, Shirasawa,
Nishimura, Fujii, Oros, Almanza, Nagatoshi,
Yasui and Fujita. This is an open-access article
distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Chromosome-level genome assemblies for two quinoa inbred lines from northern and southern highlands of Altiplano where quinoa originated

Yasufumi Kobayashi¹, Hideki Hirakawa^{2†}, Kenta Shirasawa²,
Kazusa Nishimura³, Kenichiro Fujii^{1†}, Rolando Oros⁴,
Giovanna R. Almanza⁵, Yukari Nagatoshi¹,
Yasuo Yasui⁶ and Yasunari Fujita^{7,8*}

¹Biological Resources and Post-harvest Division, Japan International Research Center for Agricultural Sciences (JIRCAS), Ibaraki, Japan, ²Department of Frontier Research and Development, Kazusa DNA Research Institute, Chiba, Japan, ³Graduate School of Environmental, Life, Natural Science and Technology, Okayama University, Okayama, Japan, ⁴Fundación para la Promoción e Investigación de Productos Andinos (Fundación PROINPA), Cochabamba, Bolivia, ⁵Instituto de Investigaciones Químicas, Universidad Mayor de San Andrés, La Paz, Bolivia, ⁶Graduate School of Agriculture, Kyoto University, Kyoto, Japan, ⁷Food Program, Japan International Research Center for Agricultural Sciences (JIRCAS), Ibaraki, Japan, ⁸Graduate School of Life Environmental Science, University of Tsukuba, Ibaraki, Japan

Quinoa is emerging as a key seed crop for global food security due to its ability to grow in marginal environments and its excellent nutritional properties. Because quinoa is partially allogamous, we have developed quinoa inbred lines necessary for molecular genetic analysis. Our comprehensive genomic analysis showed that the quinoa inbred lines fall into three genetic subpopulations: northern highland, southern highland, and lowland. Lowland and highland quinoa are the same species, but have very different genotypes and phenotypes. Lowland quinoa has relatively small grains and a darker grain color, and is widely tested and grown around the world. In contrast, the white, large-grained highland quinoa is grown in the Andean highlands, including the region where quinoa originated, and is exported worldwide as high-quality quinoa. Recently, we have shown that viral vectors can be used to regulate endogenous genes in quinoa, paving the way for functional genomics to reveal the diversity of quinoa. However, although a high-quality assembly has recently been reported for a lowland quinoa line, genomic resources of the quality required for functional genomics are not available for highland quinoa lines. Here we present high-quality chromosome-level genome assemblies for two highland inbred quinoa lines, J075 representing the northern highland line and J100 representing the southern highland line, using PacBio HiFi sequencing and dpMIG-seq. In addition, we demonstrate the importance of verifying and correcting reference-based scaffold assembly with other approaches such as linkage maps. The assembled genome sizes of J075 and J100 are 1.29 and 1.32 Gb, with contigs N50 of 66.3 and 12.6 Mb, and scaffold N50 of 71.2 and 70.6 Mb, respectively, comprising 18 pseudochromosomes. The repetitive sequences of J075 and J100 represent 72.6% and 71.5% of the genome, the majority of which are long terminal repeats, representing 44.0% and 42.7% of the genome, respectively. The *de novo* assembled genomes of J075 and J100 were predicted to contain 65,303 and

64,945 protein-coding genes, respectively. The high quality genomes of these highland quinoa lines will facilitate quinoa functional genomics research on quinoa and contribute to the identification of key genes involved in environmental adaptation and quinoa domestication.

KEYWORDS

quinoa, genome, chromosome, inbred lines, altiplano, environment, origin

1 Introduction

The composition of global staple crops has remained largely unchanged for decades. However, with the rapid pace of climate change, crop diversification will be critical to securing the future of our food supply (Mayes et al., 2012; Massawe et al., 2016; Krug et al., 2023). To date, approximately 2,500 plant species in 160 families have been domesticated, but less than 300 of these species are available on the global market. The majority of the world's population relies on only a handful of plant species for their caloric intake (Dirzo and Raven, 2003; Meyer and Purugganan, 2013; Massawe et al., 2016). Among crops considered underutilized and relatively neglected, quinoa (*Chenopodium quinoa* Willd.) is in the spotlight due to its excellent nutritional properties and remarkable ability to grow under harsh environmental conditions (Massawe et al., 2016).

Quinoa is a C3 annual plant belonging to the Amaranthaceae family, which also includes crops such as sugar beet (*Beta vulgaris* L.) and spinach (*Spinacia oleracea* L.). Both its seeds and leaves are consumed as food. Notably, quinoa seeds boast an exceptional nutritional profile, offering a well-rounded balance of the five essential macronutrients: proteins, fats, carbohydrates, vitamins, and minerals (Navruz-Varli and Sanlier, 2016; Nowak et al., 2016; Vilcacundo and Hernández-Ledesma, 2017; Rodríguez et al., 2020). Specifically, quinoa is rich in high-quality proteins, encompassing all nine essential amino acids, such as lysine, which the human body cannot produce on its own (Ruales and Nair, 1992; Burrieza et al., 2019; Dakhili et al., 2019). Additionally, it is gluten-free, making it a safe option for individuals with gluten allergies (Peñas et al., 2014). Quinoa leaves, too, are a valuable source of essential amino acids, vitamins, and minerals (Pathan et al., 2019). Furthermore, quinoa contains various functional components, such as bioactive peptides, polysaccharides, saponins, polyphenols, flavonoids, and phytoestrogens, which are believed to possess antioxidant and anti-inflammatory properties, as well as potential benefits for hypertension and diabetes management (Hirose et al., 2010; Vilcacundo and Hernández-Ledesma, 2017; Ren et al., 2023). In quinoa, the dispersal unit (i.e., seed) is botanically classified as an achene—a type of fruit containing a single seed encased in a dry,

non-opening outer shell, referred to as the pericarp (Prego et al., 1998; Burrieza et al., 2019). The seed features a curved embryo positioned around the edges, enveloping a central area known as the perisperm or basal body, which provides a unique structural characteristic to quinoa seeds (Prego et al., 1998; Burrieza et al., 2019). Unlike other grains like wheat and rice, in which essential nutrients are mainly located in the endosperm and outer hull, the valuable amino acids and functional compounds in quinoa are found in the kernel (Hemalatha et al., 2016; Motta et al., 2017; Motta et al., 2019).

Quinoa has been cultivated in the Andes region of South America for over 7,000 years (Dillehay et al., 2007). Revered as the 'mother grain' in the pre-Columbian era, it was a fundamental part of the Andean indigenous peoples' diet, along with llamas and tubers (González et al., 2015; Miller et al., 2021). However, following the Spanish conquest, quinoa's cultivation and use in indigenous ceremonies were banned, contrasting with the global spread of other Andean crops like tomatoes (*Solanum lycopersicum* L.) and potatoes (*Solanum tuberosum* L.) (Gomez-Pando, 2015). Since the 1950s, quinoa has served as an indicator plant for identifying plant virus species due to its susceptibility to various plant viruses that cause local lesions (Uschdraweit, 1955; Hein, 1957; Yasui et al., 2016). Recognizing its potential, the National Academy of Sciences (NAS) and the National Research Council (NRC) highlighted quinoa as one of the underexploited crops with significant economic promise (NAS, 1975; NRC, 1989). The National Aeronautics and Space Administration (NASA) has also explored quinoa as a nutritious food source for astronauts on extended space missions (Schlick and Bubenheim, 1993; Schlick and Bubenheim, 1996). The UN's designation of 2013 as the "International Year of Quinoa" aimed to underscore its potential in enhancing food and nutrition security and supporting sustainable agriculture (Bazile et al., 2016). Recently, although Bolivia and Peru remain the leading producers, accounting for nearly 80% of the world's supply, quinoa is being researched and grown in over 120 countries (Alandia et al., 2020).

Quinoa can be grown at temperatures from below freezing to near 40°C, at low to high latitudes, and from lowlands along the coast to highlands above 4,000 m, and is highly tolerant to abiotic

stresses such as drought, high salt, low temperatures, hail, and frost (Jacobsen, 2003; Hariadi et al., 2011; Zurita-Silva et al., 2014; Yasui et al., 2016; Mizuno et al., 2020). For example, the area around Salar de Uyuni, the world's largest salt flat, located in Bolivian Altiplano (3,800 m above sea level), where annual precipitation is less than 200 mm and saline-alkaline soil prevails, has been known as a production area for high-quality quinoa, which can only be grown as a crop in this harsh environment (Bonifacio, 2019). Compared to quinoa varieties from other regions, quinoa varieties from this southern highland region germinate faster and show faster initial growth after the seeds absorb water (Mizuno et al., 2020). This diversity within quinoa species is also seen in the color of the seed coat (from dark red to yellow to white) and is related to betalain, a pigment unique to the order Caryophyllales (Stafford, 1994; Bonifacio, 2019; Imamura et al., 2019; Ogata et al., 2021). Cultivation of quinoa outside the Andean region will require exploiting diversity within quinoa species, as it will need to be more heat tolerant in warmer environments, more resistant to pests and diseases, and less sensitive to changes in day length (Gomez-Pando et al., 2019).

Quinoa is an allotetraploid species with a chromosome count of $2n = 4x = 36$ and an estimated genome size of approximately 1.5 Gb, comprising two sub-genomes, A and B (Palomino et al., 2008; Yangquanwei et al., 2013). For years, the complex genome of quinoa, a consequence of its allotetraploidy, along with genetic diversity stemming from partial outcrossing due to the presence of both hermaphrodite and female flowers on the same plant, had posed challenges to molecular analysis (Maughan et al., 2004; Christensen et al., 2007). To overcome this situation, our collaborative group (Yasui et al., 2016) and two subsequent independent groups (Jarvis et al., 2017; Zou et al., 2017) decoded the quinoa genome with the technology available at the time. Following this, we have developed more than 130 genotyped quinoa inbred lines tailored for molecular studies, elucidating genotype-phenotype relationships for salt stress responses and important growth traits (Mizuno et al., 2020). Our comprehensive genomic analyses, employing single-nucleotide polymorphisms (SNPs), revealed that these quinoa inbred lines can be categorized into three genetic sub-populations: the northern highland, southern highland, and lowland groups (Mizuno et al., 2020). Furthermore, we have established a method to analyze the function of endogenous genes in quinoa by using virus-induced gene silencing (VIGS) and virus-mediated overexpression (VOX), paving the way for functional genomics analysis (Ogata et al., 2021).

Accurate genome sequencing is crucial for advancing functional genomics in quinoa, shedding light on its unique properties, and tackling the challenges it faces in both local and global production. Thus far, genome assemblies have been produced for four quinoa lines: two from lowland lines, Kd (Yasui et al., 2016) and QQ74 (Jarvis et al., 2017), and two from southern highland lines, Real (Zou et al., 2017) and CHEN125 (Bodrug-Schepers et al., 2021). Although these assemblies are fragmented and do not fully reflect the chromosomal biology of quinoa, the chromosome-level assembly QQ74 V2 has recently been reported for the lowland quinoa line QQ74 (Rey et al., 2023). However, unlike lowland quinoa lines, no useful genomic information is available for

highland quinoa lines: there are still no reports of genome assemblies of northern highland quinoa lines or chromosome-level assemblies of southern highland quinoa lines. In particular, lowland and highland quinoa lines are the same species, but differ significantly not only in genotype but also in phenotype, for example, highland quinoa lines are difficult to grow in lowland areas, while lowland quinoa lines are difficult to grow in highland areas (Mizuno et al., 2020). Northern highland quinoa lines have traditionally been grown around Lake Titicaca in Peru and Bolivia, where quinoa is believed to originate. The quinoa lines that have been attempted to be grown outside the Andean region such as in Europe, the United States, Asia, and Africa, are basically lowland quinoa lines with relatively small grains and a darker color, while the white, large-grained quinoa lines grown in the Andean region of South America and exported to outside the Andean region such as in Europe, the United States, and Asia are highland quinoa lines. The Altiplano highlands of Peru and Bolivia, situated at elevations of 3,000 to 4,000 m, are among the world's largest quinoa-producing regions, supplying high-quality, organically cultivated quinoa globally. However, due to the harsh conditions of this environment, they are particularly vulnerable to climate change (Bonifacio, 2019). Given these circumstances, a reference genome assembly that provides accurate genomic information on highland quinoa is essential to develop effective breeding strategies to address these challenges and to understand the process of quinoa domestication, including its adaptation to harsh environments and its origins.

In this study, we present chromosome-level genome assemblies for two inbred lines of highland quinoa: J075, a representative of the northern highland, and J100, a representative of the southern highland (Mizuno et al., 2020; Ogata et al., 2021). These assemblies were achieved by integrating PacBio HiFi long-read sequencing with the dpMIG-seq method, which is a multiplexed inter simple sequence repeat genotyping by sequencing, using degenerate oligonucleotide primers (Nishimura et al., 2024). The reference genomes obtained in this study will provide the basis for advancing functional genomics in quinoa to facilitate the development of climate-adapted highland quinoa breeding materials and contribute to a better understanding of the domestication process of quinoa, including its adaptation to harsh environments and its origin. These findings have the potential to provide clues for improving various crops to make them more adaptable to climate change. In addition, we verified and corrected the reference-based scaffold assembly using a linkage map and corrected the previously reported reference genome.

2 Materials and methods

2.1 Plant materials and growth conditions

Quinoa inbred lines, derived through single-seed descent via self-crossing, were cultivated in phytotrons as described previously (Yasui et al., 2016; Mizuno et al., 2020; Ogata et al., 2021). In this study, we used J075 as a representative of the northern highland lines, J100 as a representative of the southern highland lines, and

J082 as a representative of the lowland quinoa lines (Mizuno et al., 2020). The J075 and J100 plants were grown until they produced more than a dozen fully expanded leaves, under conditions of 25 ± 5°C and a 12-hour light/12-hour dark photoperiod. To assess the color phenotype of the J075, J100, and J082, they were grown for eight weeks under the same conditions as stated above.

2.2 Processing data for the whole genome sequence of quinoa

High molecular weight DNA was extracted from fully expanded leaves of the J075 and J100 lines, using a single plant from each line and the Genomic-tip Kit (Qiagen). SMRT sequencing libraries were prepared with the SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) and sequenced on a PacBio Sequel II system, utilizing two SMRT Cell 8M units. The PacBio subreads were converted to HiFi reads using the circular consensus sequencing (CCS) program (<http://github.com/PacificBiosciences/ccs>). HiFi reads from each quinoa line were assembled using Hifiasm (Cheng et al., 2021). Chromosome-scale pseudomolecules were then created using the reference-guided scaffolding software, RagTag ver. 2.1.0 (Alonge et al., 2022).

An F₆ mapping population (n = 183) was generated from a cross between the quinoa inbred lines J100 and J027. Genomic DNA was extracted from the young leaves of each line. Using genomic DNA from the two parental lines and 183 F₆ plants, the dpMIG-seq library was constructed as described previously (Nishimura et al., 2024). The libraries were sequenced using an Illumina HiSeq X sequencer. Low quality and adapter sequences were removed using Trimmomatic ver. 0.38 (Bolger et al., 2014) with the settings 'HEADCROP:17 SLIDINGWINDOW:4:20 LEADING:20 TRAILING:20 MINLEN:25'. The resultant high quality reads were mapped onto hifiasm-assembled J100 contigs using BWA ver. 0.7.17 (Li, 2013) and sorted using samtools ver. 1.18 (Danecek et al., 2021), and then generated a vcf file by mpileup in bcftools ver. 1.16 (Danecek et al., 2021). A total of 13,526 SNPs were uniformly selected for genetic linkage map construction in software Lep-MAP3 ver. 0.5 (Rastas, 2017). In detail, the markers were identified as informative through the ParentCall2 module and anchored to LGs comparable to the contigs in the pseudomolecule assembled by RagTag using the SeparateChromosomes2 module by adjusting the LOD values (ranging from 15 to 40). Next, a total of 29,751 SNPs were extracted as anchor markers from the 151 contigs comprising scaffolds of J100 and linkage maps were reconstructed for the contigs in each scaffold using LepMap3 as described above. The high-density genetic linkage maps were visualized using ALLMAPS pipeline (Tang et al., 2015).

The genomic structures of J075 and J100 were compared to QQ74 (v2, id60716) from CoGe (<https://genomevolution.org/coge/>) using D-GENIES (Cabanettes and Klopp, 2018), with alignments performed by Minimap2 ver. 2.26 (Li, 2021). Using the same methods, the genomic structures of Real (Zou et al., 2017) and CHEN125 (Bodrug-Schepers et al., 2021) were compared to J100 genome assembly in this study.

2.3 Assessing assembly completeness

The completeness of the genome assemblies for J075 and J100 was assessed using the Benchmarking Universal Single-Copy Orthologs (BUSCO) database, ver. 5.3.2 (Simao et al., 2015), in genome mode. This assessment searched for genes conserved in embryophyte species. The data sets, embryophyta_odb10, created on 10 September 2020, include 50 genomes and 1614 BUSCOs, respectively. The completeness of the assembled genomes for J075 and J100 was further evaluated using the Long Terminal Repeat (LTR) Assembly Index (LAI) (Ou et al., 2018), which assesses genome quality based on intergenic genome information using LTR retrotransposons (LTR-RT). LTR retrotransposon candidates were identified using LTRharvest and LTR_FINDER_parallel and were further classified using LTR_retriever ver. 2.9.6 (Ou and Jiang, 2018). The continuity of repeat space in the assemblies was evaluated by the LAI, calculated using the LTR_retriever package.

2.4 Repetitive sequence detection and genome functional annotation

De novo repeat libraries for the pseudomolecule sequences J075 and J100 were constructed using RepeatModeler ver. 2.0.3 (Flynn et al., 2020). Repetitive sequences were identified and classified using these *de novo* repeat libraries along with RepeatMasker ver. 4.1.2 (Flynn et al., 2020). Putative protein-coding genes in the genome sequences were predicted using the BRAKER2 pipeline, which incorporates protein evidence via GeneMark-ES (Bruna et al., 2021). Proteins were aligned to the genome using ProfHint (Bruna et al., 2020), which integrates the Spaln (Iwata and Gotoh, 2012; Gotoh et al., 2014) and DIAMOND (Buchfink et al., 2015) protein aligners, generating specific gene model parameters. These parameters were then applied in AUGUSTUS ver. 3.4.0 (Hoff and Stanke, 2019) for gene prediction. Within the BRAKER pipeline, Viridiplantae protein sequences from the OrthoDB (Kriventseva et al., 2019) served as the evidence protein dataset.

Homology-based gene annotation was conducted using GeMoMa ver. 1.9 (Keilwagen et al., 2019), utilizing genome, gene annotation, and mapped RNA-seq data from quinoa reference lines (Yasui et al., 2016; Rey et al., 2023). This homology-based gene prediction by GeMoMa was complemented by comparison with *ab initio* gene annotation by BRAKER, with GeMoMa modules being used to obtain the final gene annotation. The quality of putative protein-coding genes was evaluated using the EnTAP functional annotation package ver. 0.10.8 (Hart et al., 2020), with curated databases from UniProtKB/Swiss-Prot (https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_sprot.fasta.gz), UniProtKB/TrEMBL (https://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/complete/uniprot_trembl.fasta.gz), and NCBI's Plant Protein (<https://ftp.ncbi.nlm.nih.gov/refseq/release/plant/>) for similarity searches (DIAMOND; E-value cut-off of 0.00001). These searches were followed by alignment to the eggNOG database ver. 4.5 using eggNOG-mapper ver. 2.1 (Huerta-Cepas et al., 2016;

Cantalapiedra et al., 2021). Gene family associations provided the basis for Gene Ontology (GO) term assignments, identification of protein domains (Pfam) (Finn et al., 2014), and associated pathways (KEGG) (Kanehisa et al., 2014). Genome circular plots were created using Circa (<https://omgenomics.com/circa/>).

2.5 Orthology prediction and phylogenomic analysis

Orthologous cluster, phylogenetic, and gene family evolution analyses of quinoa J075 and J100, along with six other flowering species (*Arabidopsis thaliana*, *Oryza sativa*, *Zea mays*, *Amaranthus hypochondriacus*, *Beta vulgaris*, and *C. quinoa* QQ74), were conducted using the OrthoVenn3 platform (Sun et al., 2023). The OrthoFinder algorithm identified orthologous gene families across these eight plants, which included quinoa lines partitioned by two subgenomes. Clustered gene families were visualized using UpSetR (Conway et al., 2017), which facilitated the analysis of overlapping families among multiple plants and highlighted unique clusters within each species. Overlapped and unique gene families were used for GO enrichment analysis in the platform. The GO terms for the biological process, molecular function, and cellular component categories and enrichments were obtained as GO Slim terms (p -value < 0.05).

The phylogenetic and gene family contraction and expansion analysis were automatically run on the platform. A total of 2,137 single-copy orthologs from these plants were aligned using MUSCLE, with conserved sequences subsequently trimmed using TrimAl. Phylogenetic trees were then constructed using the maximum likelihood method, based on the JTT+CAT evolutionary model implemented in FastTree2. In addition to these analyses, CAFE5 was employed to detect gene family expansion and contraction by extrapolating species gene family sizes and evolutionary timelines. Divergence times among lineages containing each quinoa genome were estimated using the TimeTree5 database (<http://www.timetree.org/>), with divergence events for *A. thaliana* vs *O. sativa*, *A. hypochondriacus* vs *O. sativa*, *B. vulgaris* vs *O. sativa*, *A. thaliana* vs *A. hypochondriacus*, *A. thaliana* vs *B. vulgaris*, and *A. hypochondriacus* vs *B. vulgaris* occurring at 160, 160, 160, 118, 118, and 49 mya, respectively. GO enrichment analysis was also automatically performed on the platform to explore the functional attributes of the expanded and contracted gene families as mentioned above.

2.6 Genome-wide replication analysis

The synonymous substitution rate (K_s) distributions between the two sub-genomes of quinoa, *C. pallidicaule* (A-progenitor genome, v4.0, id53872) (Mangelson et al., 2019) and *C. suecicum* (B-progenitor genomes, v2, id52047) (Rey et al., 2023) from CoGe (<https://genomeevolution.org/coge/>) were calculated by wgd ver. 2.0.26 (Chen and Zwaenepoel, 2023). For the calculation of K_s distributions for one-to-one orthologs, the dmd module was utilized to extract orthologs via an all-versus-all BLASTP search

using the DIAMOND algorithm. Subsequently, the ksd module was employed to construct the one-to-one ortholog K_s distributions among the genomes.

2.7 Genome synteny analysis

Syntenic gene pairs among quinoa lines were identified using the MCScan pipeline implemented in python [[https://github.com/tanghaibao/jcvi/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version))] (Tang et al., 2008). Within the pipeline, the “jcvi.compara.catalog.ortholog” function was used to search for syntenic regions. The “jcvi.compara.synteny.screen” function was employed to identify macro-syntenic blocks, which are defined as having more than 30 collinear genes within a syntenic block. The “jcvi.compara.synteny.mcscan” function was utilized for micro-synteny analysis at the gene level. Finally, the “jcvi.graphics.karyotype” function was used to visually display each synteny as a karyotype.

For identification of structural variants (SVs) calling using SyRI (ver.1.6) (Goel et al., 2019), minimap2 (ver. 2.24) (Li, 2021) was used to generate pairwise genome alignments with parameters ‘-ax asm5’. The alignment result was subsequently passed to SyRI, and SVs consisting of insertions, deletions, inversions, and translocations were kept using default parameters.

3 Results

3.1 Chromosome-level genome assemblies for southern and northern highland inbred quinoa lines

Based on the results of previous population structure and phenotypic analyses (Mizuno et al., 2020; Ogata et al., 2021), we selected J075 as a representative line of northern highland quinoa and J100 as a representative line of southern highland quinoa. The whole genome sequences of northern highland J075 and southern highland J100 were obtained using PacBio Sequel II sequencing platform, yielding 59.9 Gb and 38.9 Gb of data, respectively (Supplementary Table 1). The PacBio data obtained for J075 and J100 were assembled into 155 and 963 primary contigs, respectively (Supplementary Table 2). For J075, the HiFi-based contig sequences totaled 1.29 Gbp with an N50 of 66.3 Mbp, and for J100, they totaled 1.32 Gbp with a contig N50 of 12.6 Mbp (Supplementary Table 2). Next, we performed reference-guided scaffolding of the HiFi-based contigs, using RagTag in conjunction with the quinoa reference genome (QQ74, v2, id60716) (Rey et al., 2023), which includes all 18 chromosomes of quinoa (Chromosomes 1A to 9A and 1B to 9B). Subsequently, 30 contigs of J075 and 151 contigs of J100 were each mapped to the 18 chromosomes of quinoa (Cq1A-Cq9A and Cq1B-Cq9B; Supplementary Table 3). RagTag analysis resulted in total scaffold lengths of 1.28 Gbp for J075 and 1.27 Gbp for J100, with average scaffold lengths of 71.0 Mbp and 70.7 Mbp, respectively (Supplementary Table 3).

The 30 contigs of J075 could be mapped directly to the reference genome sequence, but the mapping of the 151 contigs of J100 was

unclear in some places, so scaffolding was performed using a linkage map created from the J100 x J027 (Mizuno et al., 2020) mapping population. Initially, SNPs were extracted from all contigs assembled using HiFi reads to serve as anchor markers for linkage analysis. In the linkage groups, we confirmed that the majority of contigs mapped to the scaffold assembled by RagTag. Consequently, 29,751 SNPs were extracted as anchor markers from the 151 contigs comprising these scaffolds for linkage analysis. The contigs were then mapped to linkage groups, covering 3,205 cM, using these SNPs (Figure 1 and Supplementary Table 4). Dot plot analysis conducted with D-GENIES showed a significant improvement in assembly continuity for the J100 assembly compared to the previously published genomic scaffolds for the Real line (Zou et al., 2017) and the CHEN125 line (Bodrug-Schepers et al., 2021) (Supplementary Figure 1). These findings indicate that the J100 assembly represents the most contiguous, highest-quality assembly of southern highland quinoa lines to date.

The pseudomolecules were searched against chloroplast and mitochondrial genome sequences obtained from NCBI RefSeq (all genome sequences available in the mitochondrion and plastid directories as of June 20, 2024) using blastn v2.12.0 with an *E*-value cut-off of 1e-20. No reference organelle sequences with more than 90% alignment length were identified within the genomic regions of the pseudomolecules. Comparing the pseudomolecule lengths of both J075 and J100 lines and reference genome line QQ74, the N50 lengths of J075 and J100 were both 1.06 times longer than those of QQ74 (Table 1). Dot plot analysis using D-GENIES, comparing the assemblies of J075 and J100 with the reference genome line QQ74, revealed structural variations in Cq7A and Cq3B (Figure 2 and Supplementary Figure 2). For the order of physical positions of contigs in the scaffolds obtained with Hifiasm and RagTag, contigs that were clearly inconsistent with their positions on the linkage map were reassembled according to the

order of the linkage map (Supplementary Table 4). These results show that the J075 and J100 assemblies have superior genome completeness with long repetitive regions compared to the QQ74 assembly of the reference genome.

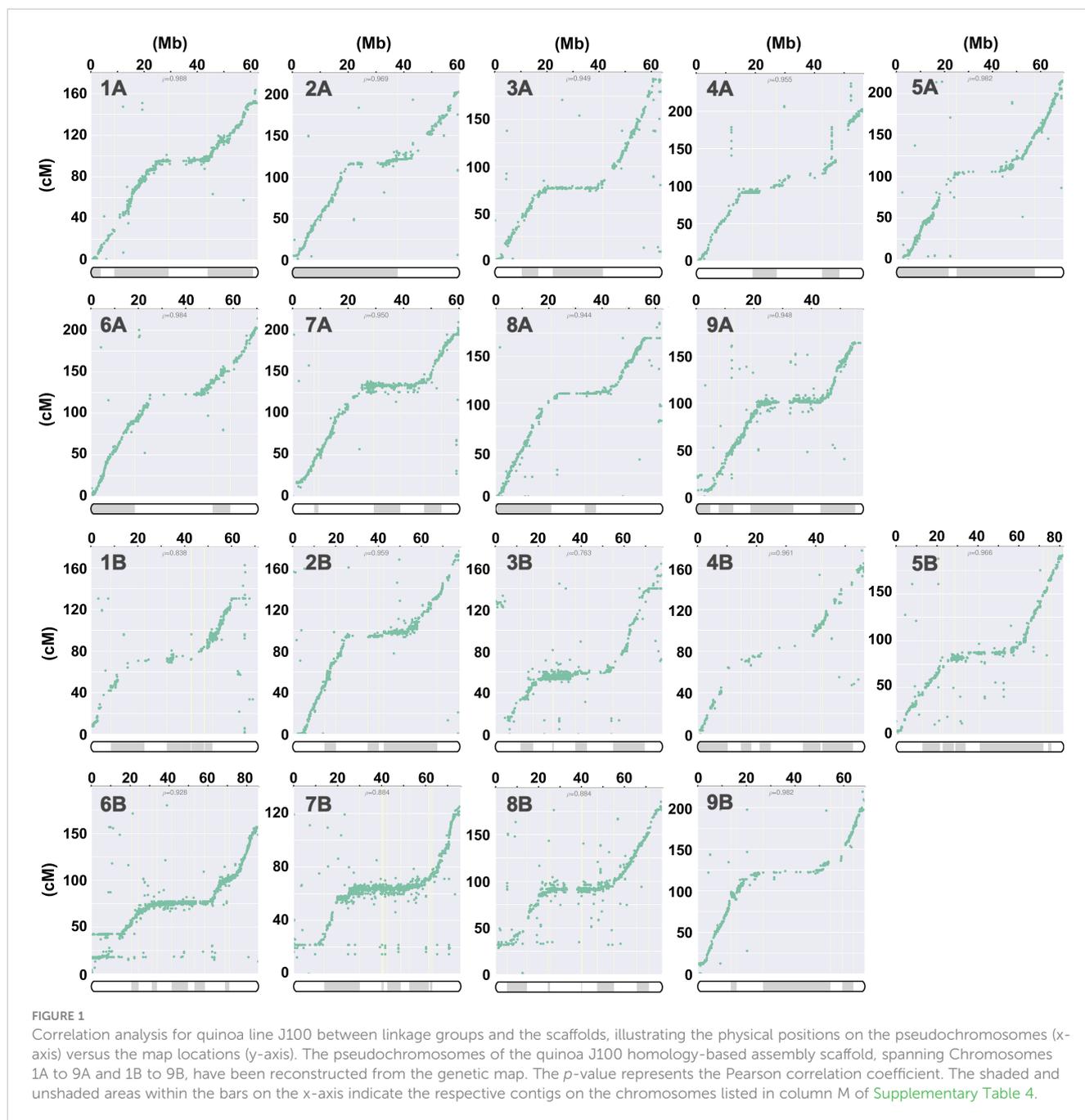
3.2 Genome annotation and gene prediction

The completeness of the genome assembly for all scaffolds and non-localized contigs was quantified using the BUSCO database. We used the embryophyta_odb10 dataset, which contains 1,614 core genes. The scaffolds and non-localized contigs of J075 and J100 contained benchmark genes with high quality values of 99.2% and 99.1%, respectively, and almost all of them (89.7% and 89.3%, respectively) were duplicated as isoform genes, according to our BUSCO analyses (Supplementary Table 5). The benchmarking gene preservation of J075 and J100 was almost equivalent to that of reference genome line QQ74. The completeness of these genome assemblies was further validated using the LAI that evaluates assembly continuity using LTR-RTs. A higher LAI score indicates better quality of repetitive and intergenic sequence spaces due to the identification of more intact LTR-RTs. J075 and J100 presented relatively higher LAI scores, at 17.40 and 17.75, respectively (Supplementary Table 6). Based on these quality assessment results, the J075 and J100 genomes were classified as being of reference quality based on the assembly of repetitive and intergenic sequence spaces (draft quality, LAI < 10; reference quality, LAI between 10 and 20; gold quality, LAI > 20). We then analyzed repetitive and transposable sequences in J075 and J100. The repetitive sequences comprised 938.9 Mb (72.6%) and 944.0 Mb (71.5%) of the pseudomolecule sequences of J075 and J100, respectively (Supplementary Table 6). The most common LTR-RTs superfamily, *Gypsy* and *Copia*, comprised 44.0% and 42.7% of the genomes of J075 and J100, respectively (Supplementary Table 6). These results indicated that both J075 and J100 contain a similar proportion of repetitive and transposable sequences. The GC contents ratio of the two genomes averaged 37% across the genome length, but this percentage was lower in the central region of most chromosomes, where a region containing an unknown LTR retrotransposon was detected (Figure 3 and Supplementary Figure 3).

Next, gene models for the J075 and J100 genomes were developed using a comprehensive approach that included *ab initio* and homology-based predictions. Initially, BRAKER2 and GeMoMa predicted a total of 100,169 and 97,309 genes, respectively, from the softmasked scaffolds of J075 and J100 using protein evidence. These genes underwent further analysis, including more sensitive DIAMOND searches against the UniProtKB and NCBI NR databases and domain searches against the Pfam database. Genes containing transposon-related keywords were excluded from the similarity search results. Furthermore, genes filtered by the 'Viridiplantae' and 'Eukaryotes' categories in the EggNOG Tax Scope were classified as 'best' genes. As a result, 65,303 and 64,945 genes from J075 and J100, respectively, were classified as 'best' genes (Supplementary Table 7). The completeness

TABLE 1 Scaffolding statistics of quinoa J075 and J100 after assembly using RagTag or in combination with Lep-MAP3.

Statistics	J075	J100
N50 (bp)	71,226,514	70,609,007
L50	9	9
N90 (bp)	59,638,252	59,948,908
L90	17	17
Number of scaffolds	143	813
Gaps	12	149
N count (bp)	1,200	14,900
Largest scaffold (bp)	84,258,211	86,085,297
Smallest scaffold (bp)	28,991	18,494
Average scaffold length (bp)	9,041,347	1,623,460
Total length (bp)	1,292,912,647	1,319,872,862
Percentage in 18 largest scaffolds	98.9	96.4



of the annotation based on the BUSCO was 96.3% for both J075 and J100 lines. These predicted genes were found to be densely distributed at the terminal ends of chromosomes ([Figure 3](#) and [Supplementary Figure 3](#)).

3.3 Gene family analysis and genome evolution in quinoa

To explore the evolutionary history of the quinoa gene family, orthologous gene families were identified using the protein sequences of J075 and J100 predicted in this study, along with those from six other flowering species (*A. thaliana*, *O. sativa*, *Z.*

mays, *A. hypochondriacus*, *B. vulgaris*, and *C. quinoa* QQ74). Given that quinoa is allotetraploid, their protein sequences were differentiated into the A and B genomes. In total, 26,989 and 27,663 genes from the A and B genomes of J075 (representing 99.4% and 99.5% of the total, respectively) clustered into 18,793 and 18,723 gene families. Similarly, 26,873 and 27,197 genes from the A and B genomes of J100 (accounting for 99.5% of the total each) clustered into 18,777 and 18,696 gene families ([Supplementary Table 8](#)). Among these, 8,986 gene families were shared across all genomes from the eight plants ([Figure 4](#)). We also discovered that the A genomes of J075 and J100 have 28 and 27 unique gene families (with 104 and 101 unique genes, respectively), respectively, while the B genomes have 23 and 24 unique gene families (with 124

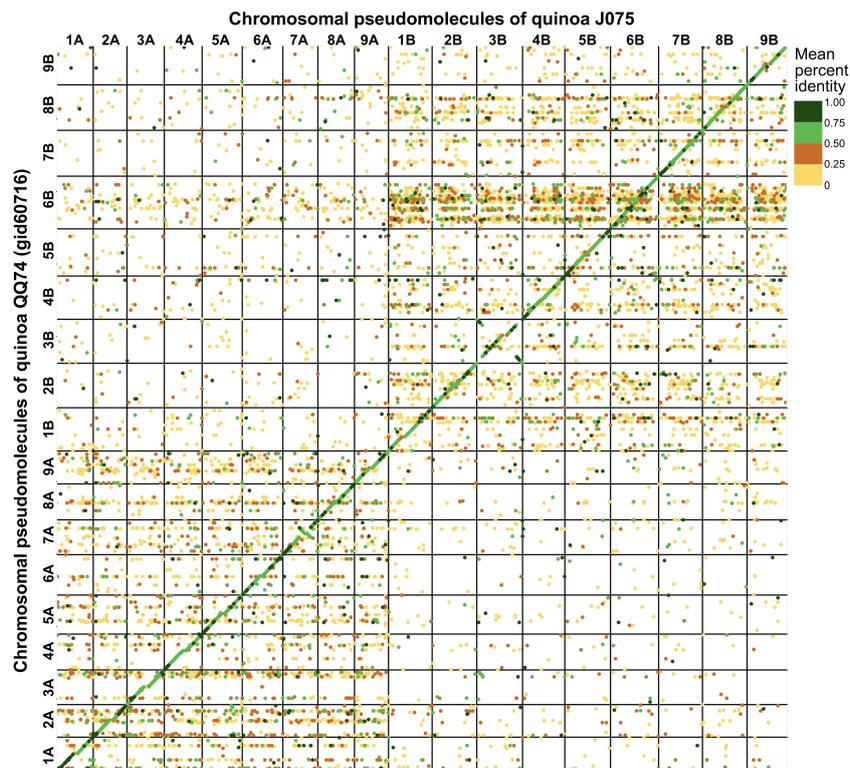


FIGURE 2

Dotplot image comparing the quinoa J075 pseudomolecules with the reference chromosomes. Dotplot analysis was conducted using minimap2 aligner, and the result was subsequently visualized on the D-GENIES platform (<https://dgenies.toulouse.inra.fr>). The colors in the dotplot represent identity levels.

and 108 unique genes, respectively) (Supplementary Figure 4 and Supplementary Table 8). Classification of these unique orthologous gene families revealed that nucleic acid metabolic processes and defense response-related GO terms are enriched across multiple sub-genomes in quinoa (Supplementary Table 9).

We constructed a high-confidence phylogenetic tree and estimated the divergence times of these plant species based on the amino acid sequences from 2,137 single-copy gene families (Figure 5). As expected, the phylogenetic trees generated using concatenated methods confirmed that the A and B genomes of quinoa were divided into two separate clades. Subsequently, the expansion and contraction of orthologous gene families were analyzed using CAFE5. In the lineage leading to the highland quinoa population, 85 and 62 gene families were expanded, while 91 and 63 gene families were contracted in the A and B genomes, respectively. In the lineage leading to the southern highland J100, 94 and 81 gene families were extended, which were fewer than the 142 and 161 extended gene families found in the reference quinoa's A and B genomes (Figure 5 and Supplementary Tables 10, 11). To investigate the effects of gene gain and loss on quinoa population development, we conducted GO enrichment analysis for the significantly changed gene families in each quinoa line (Supplementary Tables 10, 11). GO terms common to two or more lines included nucleic acid and protein metabolic processes, as well as defense response-related terms. These results mirror the results of comparative analysis of orthologous gene families among

plant species, with an increase in gene families that are functionally similar but differ in sequence homology. The uniquely expanded gene families in J075A and J075B that were selected for the top 5 were rRNA base methylation, gibberellic acid-mediated signaling pathway, regulation of intracellular pH, tRNA methylation, and multicellular organism development, and (-)-secologanin biosynthesis process, ATP synthesis-coupled electron transport, protein-chromophore linkage, aerobic electron transport chain, and lipid transport, respectively. On the other hand, the unique expanded gene families in J100A and J100B were copper ion stress response, cell surface receptor signaling pathway, DNA replication, microsporogenesis and unidimensional cell growth, and cellular aromatic compound metabolic process, positive regulation of hydrogen peroxide metabolic process, maintenance of chromatin silencing, anatomical structure morphogenesis, and auxin-activated signaling pathway. These expanded and contracted gene families are suggested to reflect trait differences in the J075 and J100 lines. GO terms involved in electron transport and intracellular pH homeostasis may partially explain the unique energy-generating function of J075. In J100, the specific GO terms involved in cell growth and auxin response may also differentiate it from other lines to plant growth characteristics.

Ks distributions in the allotetraploid quinoa genome displayed a peak not observed in the ancestral diploid species, suggesting that whole genome duplication may have been caused by hybridization of the ancestral species (Jarvis et al., 2017). Therefore, we calculated

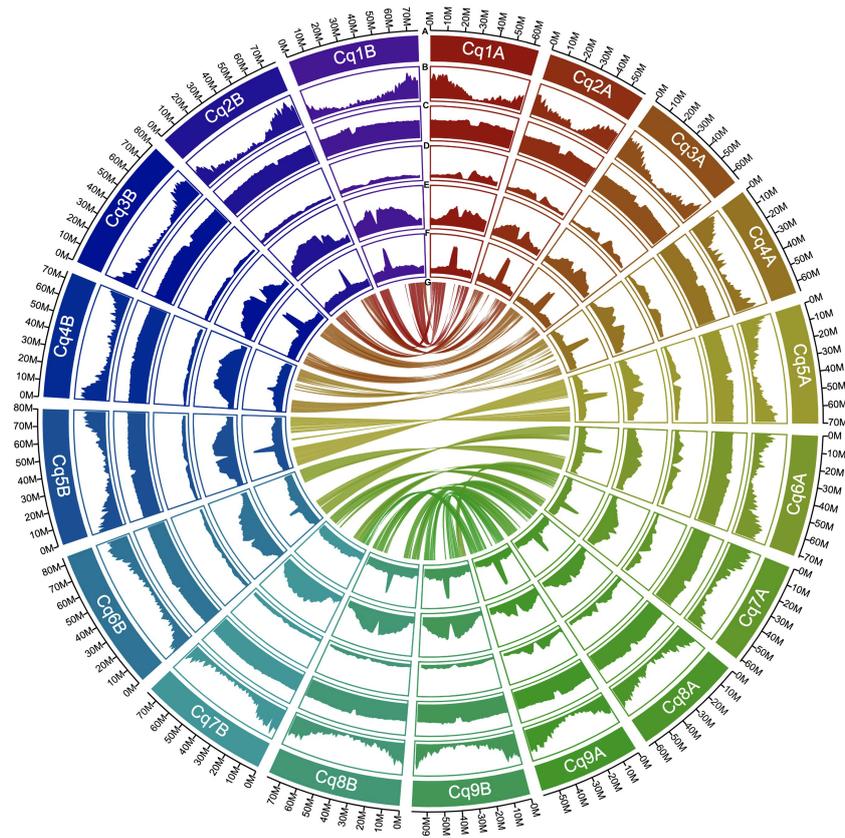


FIGURE 3 Distribution of genomic features in quinoa J075. From the outermost to the innermost, the tracks represent the following: assembled 18 pseudochromosomes (A), gene density (B), GC content (C), *copia*-type retrotransposon fragments (D), *gypsy*-type retrotransposon fragments (E), unknown retrotransposon fragments (F), and intergenomic collinearity (G).

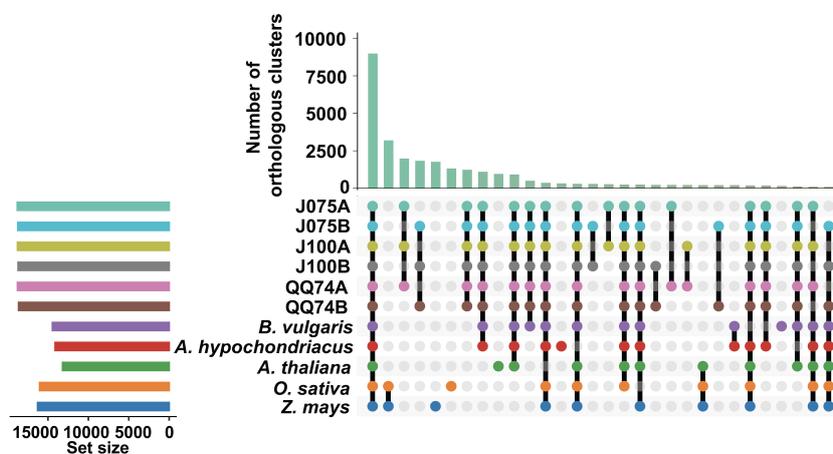


FIGURE 4 Orthologous cluster analysis of several plant species, including the A and B genomes of quinoa. An UpSet table displays both unique and shared orthologous clusters among the species. The horizontal bar chart on the left shows the number of orthologous clusters per species, while the vertical bar chart on the right indicates the number of clusters shared among the species. Lines represent intersecting sets. The results of the intersecting combinations of J075, J100, QQ74, *B. vulgaris* and *A. hypochondriacus* from this dataset are illustrated as Venn diagrams in [Supplementary Figure 4](#).

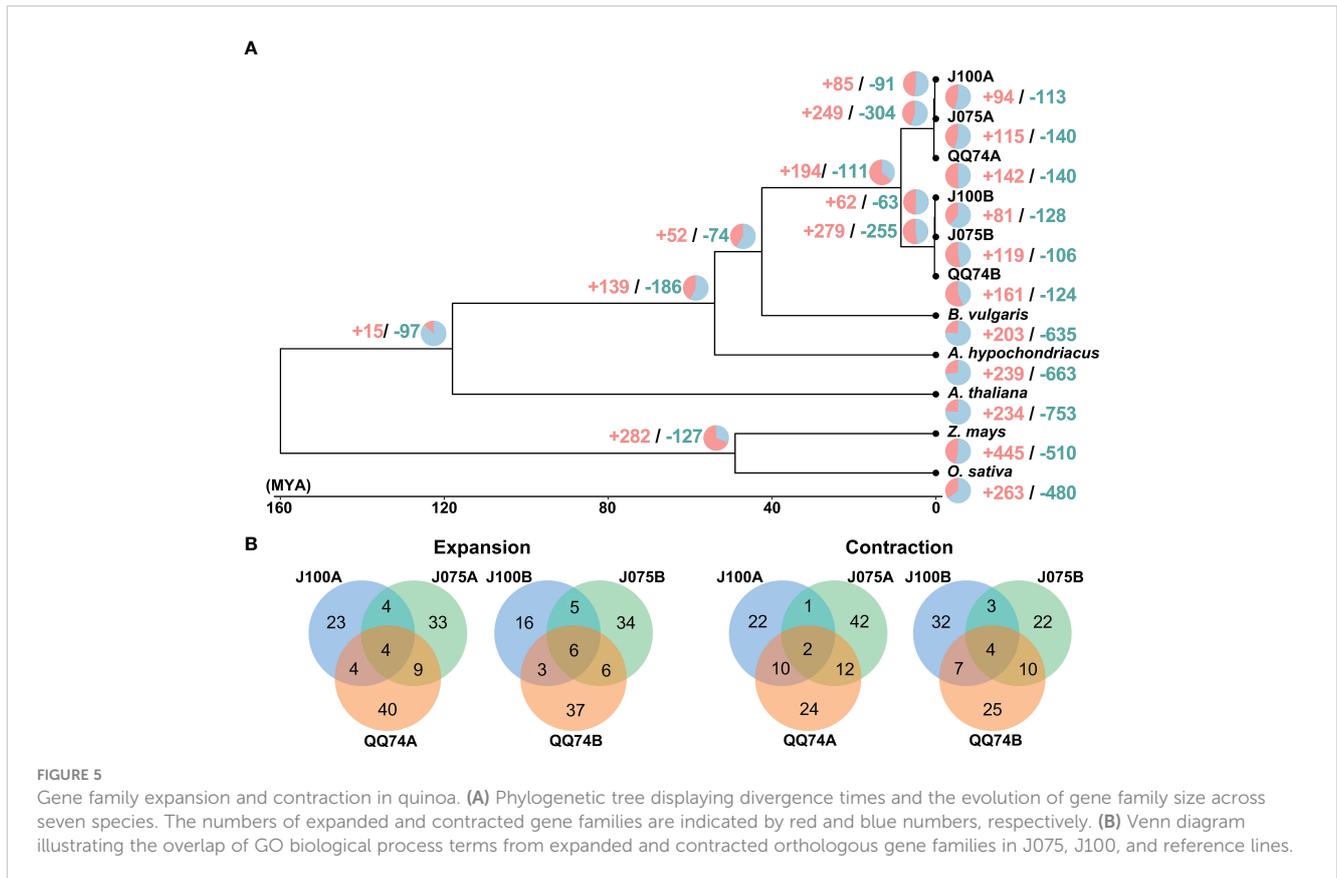


FIGURE 5

Gene family expansion and contraction in quinoa. (A) Phylogenetic tree displaying divergence times and the evolution of gene family size across seven species. The numbers of expanded and contracted gene families are indicated by red and blue numbers, respectively. (B) Venn diagram illustrating the overlap of GO biological process terms from expanded and contracted orthologous gene families in J075, J100, and reference lines.

the K_s distribution for one-to-one orthologous genes between both genomes of quinoa and their progenitor genomes. By comparing the divergence between the A and B genomes (J075A and J075B, J100A and J100B, and A- and B-progenitors), we found that the K_s values in the highland quinoa lines J075 and J100 were consistent with those between the A and B progenitor genomes, as well as with

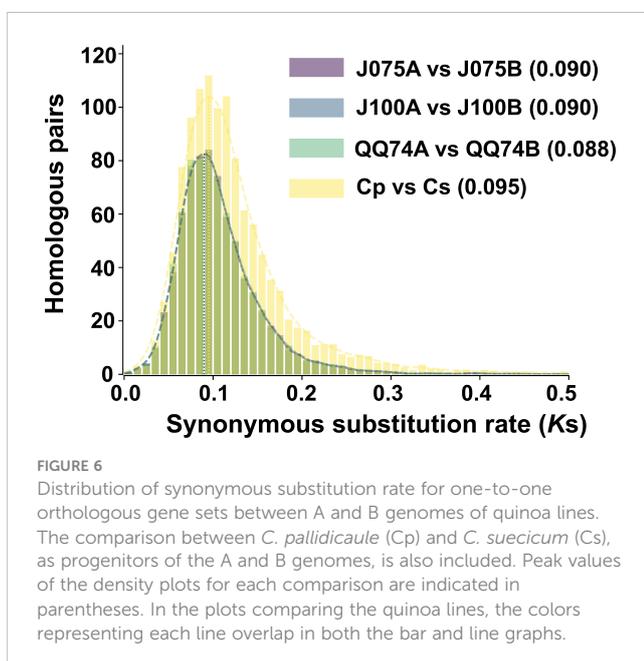
those reported for the previously studied lowland quinoa line QQ74 (Figure 6). The result suggests that there was hybridization between the diploid ancestral species in the lines representing each population.

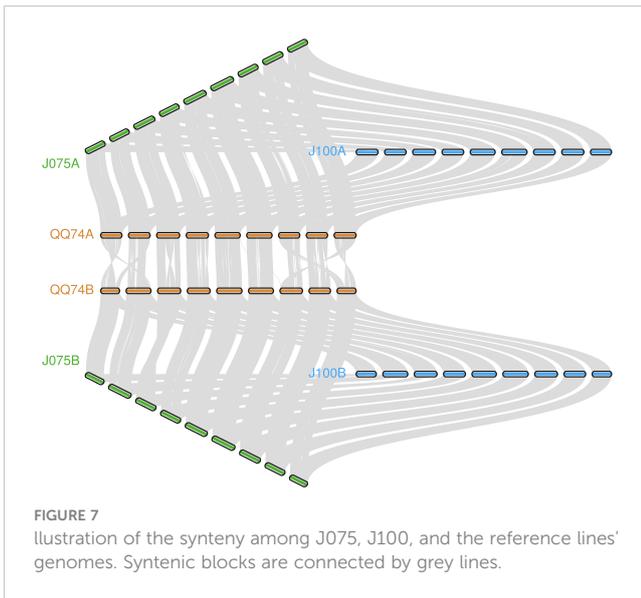
3.4 Synteny analysis and structural-variant detection in quinoa genomes

The collinearity of chromosomes was assessed by comparing the sequences of J075, J100, and reference lines. Based on JCVI-filtered information, we provided the distribution of collinear genes across the chromosomes. The number of collinear blocks in the J075-QQ74, J100-QQ74, and J075-J100 pairs was 71, 70, and 51, respectively, encompassing 44,966, 45,143, and 53,429 genes. These results demonstrated a clear one-to-one syntenic relationship, with overall gene synteny being largely conserved (Figure 7). Despite the high level of synteny across all 18 chromosomes, SyRI identified structural rearrangements when comparing the J075 genome to the J100 genome. Specifically, we identified 100 inversions ranging from 217 bp on chromosome 9B to 9.2 Mb on chromosome 5A (Figure 8 and Supplementary Table 12).

3.5 Variation of betalain biosynthesis gene family in highland quinoa lines

We have been working on phenotypic and genotypic associations, such as responses to salt stress and key growth traits,



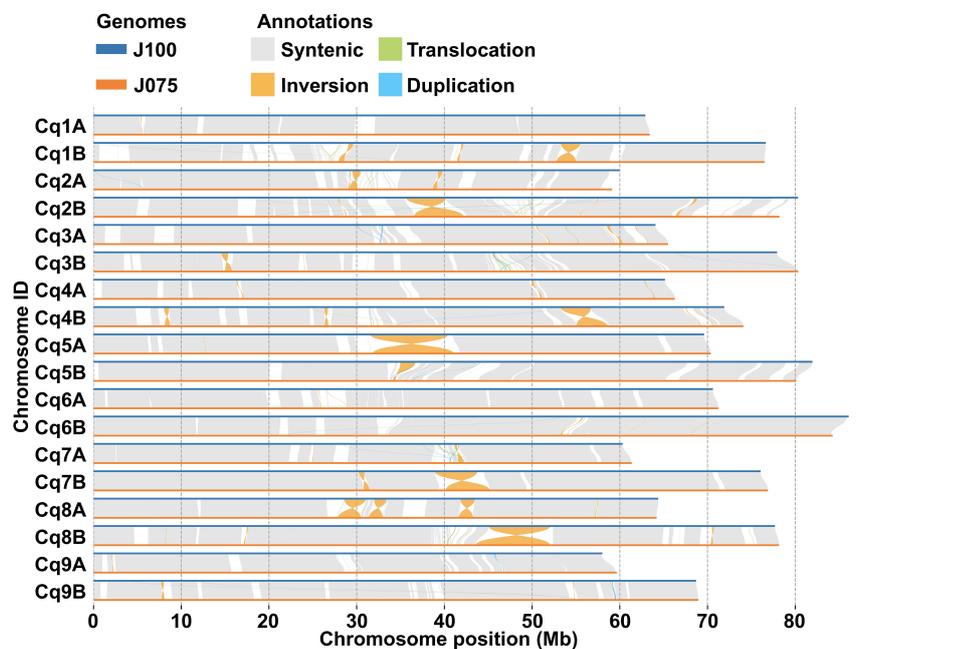


among genetically classified subpopulations of the quinoa inbred lines produced by our research group (Mizuno et al., 2020). Although synteny is maintained in most genomic regions among quinoa lines, synteny differences in localized regions may harbor genetic factors responsible for phenotypic differences. A biosynthetic pathway in quinoa that uses L-tyrosine as a precursor to synthesize betanin or amaranthin has been previously reported (Imamura et al., 2019; Ogata et al., 2021) (Supplementary Table 13). We selected genes homologous to the reported betalain biosynthetic gene sequences from each line and assessed genomic structural variations (Supplementary Table 14).

Multiple copies of the genes *CqDODA1* and *CqCYP76AD1*, which are involved in the conversion of L-tyrosine to betanine, are located at 74.53-74.71 Mb, 74.60-74.82 Mb, and 69.74-69.91 Mb on chromosome 1B, and at 58.22-58.33 Mb, 57.36-57.44 Mb, and 52.04-52.10 Mb on chromosome 2A in J075, J100, and QQ74, respectively. Additionally, chromosomes 4A and 4B had clusters of *CqDODA1s* in J075, J100, and QQ74 at 6.59-6.66 Mb, 6.70-6.76 Mb, and 6.74-6.80 Mb and 6.66-6.71 Mb, 6.54-6.59 Mb, and 6.79-6.84 Mb, respectively (Figure 9A). In particular, regions on chromosomes 1B and 2A containing clusters of betalain synthesis genes were less conserved in synteny compared to adjacent regions in each line. Several genes, such as *CqCYP76AD1* on 1B, *CqDODA1* on 4A, and *CqDOPA5GT* on 4B, were more duplicated in the J075 and J100 lines than in the lowland QQ74 line. The leaf color at the shoot apex of the highland J075 and J100 lines was redder compared to that of J082, which is the inbred line derived from PI614886, the same accession as QQ74 (Mizuno et al., 2020; Ogata et al., 2021). Comparing the two highland lines, J075 displayed a darker red color at the shoot apex leaves and hypocotyl (Figure 9B). These color differences may be related to variations in genomic sequences in regions harboring betalain biosynthesis genes, including their intergenic and promoter regions.

4 Discussion

Here, we present chromosome-scale high-quality *de novo* genome assemblies for two quinoa inbred lines from the northern and southern highlands of the Andean Altiplano of South America, where quinoa originated, using a combination of PacBio long-read sequencing with the dpMIG-seq method (Nishimura et al., 2024).



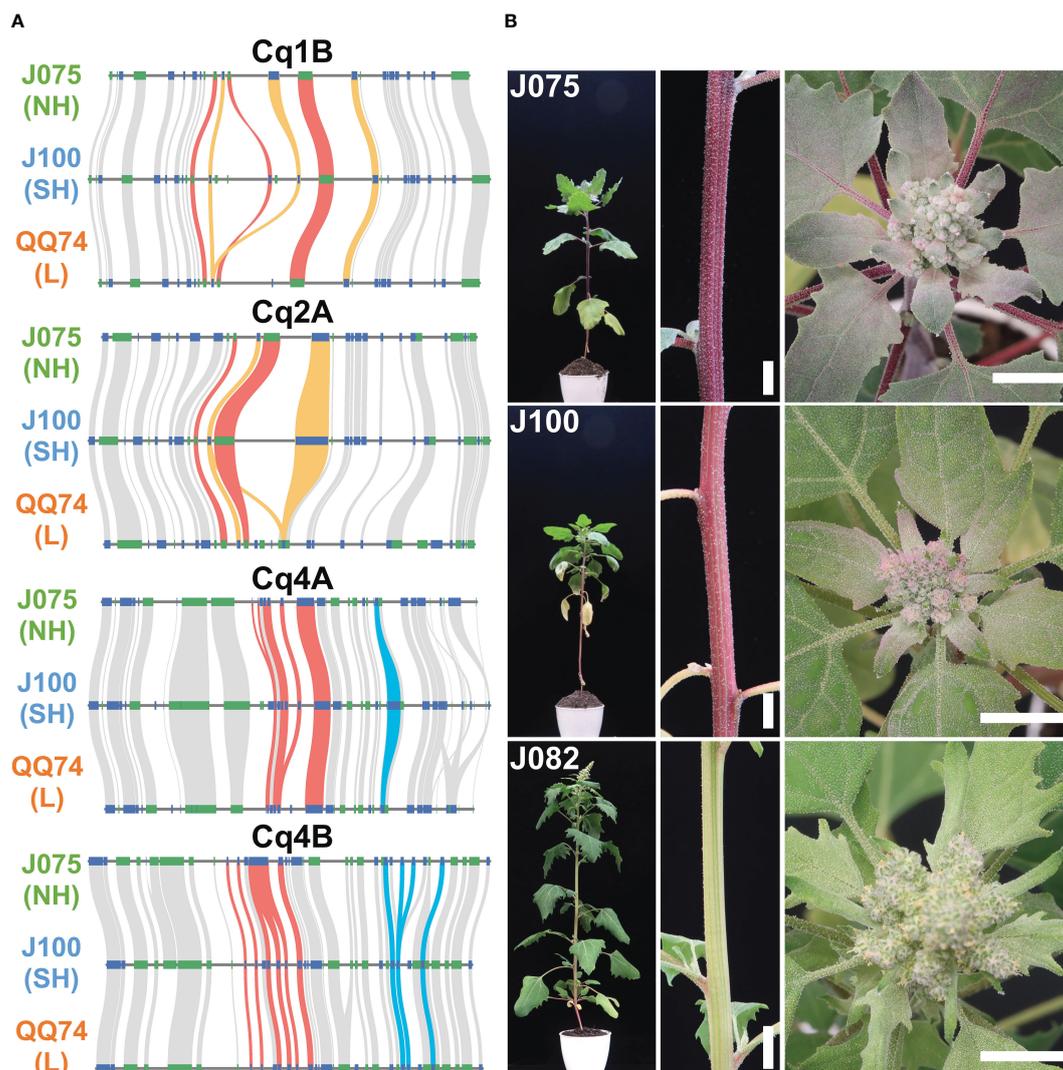


FIGURE 9

Syntenic betalain biosynthesis gene clusters and plant color phenotypes in representative quinoa lines. **(A)** Synteny analysis of betalain biosynthesis genes on chromosomes 1B, 2A, 4A, and 4B among J075, J100, and J082. The red, orange, and blue lines highlight the syntenic gene pairs *CqDODA1*, *CqCYP76AD1*, and *CqDOPA5GT*, respectively. NH, SH, and L indicate northern highland, southern highland and lowland, respectively. **(B)** Phenotypes of stem and shoot apex colors in J075, J100, and J082 after 56 days of growth. J082 is the inbred line derived from PI614886, the same accession as QQ74. Bars indicate 1 cm.

Because quinoa can partially outcross due to the presence of both hermaphrodite and female flowers on the same plant (Maughan et al., 2004; Christensen et al., 2007), high-quality quinoa inbred lines need to be used for detailed molecular analysis, and high-quality genomes of highland quinoa lines grown close to their place of origin are essential for understanding the origin of quinoa and the process of domestication. These quinoa genomes of northern highland line J075 and southern highland line J100 are 1.29 Gb and 1.32 Gb, respectively, of which approximately 98.9% (1.28 Gb, Table 1) and 96.4% (1.27 Gb, Table 1) can be anchored to 18 chromosomes. The quality of these two genome assemblies of highland quinoa lines was higher than that of several other published quinoa genomes, including those of Kd (Yasui et al.,

2016), QQ74 V1 (Jarvis et al., 2017), Real (Zou et al., 2017), CHEN125 (Bodrug-Schepers et al., 2021), and QQ74 V2 (Rey et al., 2023). It is worth noting that 72.6% and 71.5% of the repetitive elements and 99.2% and 99.1% of the plant single-copy orthologs were detected in the assembly genomes of J075 and J100, which is a higher percentage than detected in Kd, QQ74 V1, Real, CHEN125, or QQ74 V2. Taken together, the assemblies of J075 and J100 are relatively accurate and complete. The quality and length of the HiFi reads obtained in this study suggest that a reliable, high-quality chromosome-level genome can be rapidly constructed by exploiting the homology between the spatial chromosome conformation and the reference genome constructed based on the linkage map, as long as the number of contigs that make up the

chromosome-level scaffold is less than a few. These are the most recent reference genomes for quinoa and will lay the foundation for understanding the expansion of quinoa cultivation areas and adaptation during intraspecies diversification in harsh environments and will provide important resources for the further investigation of genetic diversity in quinoa and related species.

The expansion and contraction of orthogroup gene families have contributed to plant diversification during evolution (Demuth and Hahn, 2009). Among these gene families, those related to nucleic acid metabolism and defense responses were particularly abundant in the genome of quinoa lines (Figure 5 and Supplementary Tables 10, 11). Transposons, related to nucleic acid metabolism, have been identified as important in adaptive evolution (Li et al., 2018), suggesting that the expansion of these gene families was necessary for the adaptive evolution of quinoa populations, which are distributed in widely differing environments. Defense-related genes have also been reported to expand through tandem duplication (Hanada et al., 2008), with the occurrence of multiple genes with similar functions possibly leading to diversification among quinoa lines. Beyond these common gene families, we identified features from gene families specific to each line, presumed to reflect the unique characteristics of the quinoa populations (Figure 5 and Supplementary Tables 10, 11). The expansion of the auxin-related gene family in J100, and the post-embryonic development gene family in both J100 and QQ74, may be linked to enhanced post-germination growth in lines from the southern highland population compared to other populations (Mizuno et al., 2020). Furthermore, the expansion of gene families related to photosynthesis in both J075 and QQ74 suggests that variations in photosynthesis-related genes have evolved to adapt to the significant elevational and latitudinal differences in the regions where they are distributed. Regarding the comparison of orthologous gene families in each subgenome, it is possible that translocations may have occurred between subgenomes. Families with significant changes in gene number due to these translocations may be one of the factors involved in phenotypic differences among quinoa lines.

The phylogenetic tree shows a closer relationship between J075 and J100 compared to QQ74 in both subgenomes, suggesting that after the divergence of the lowland and highland populations, the highland population split into the northern and southern highland populations (Figure 5). Additionally, genomic structural variations were found between the southern and northern highland lines (Figure 8 and Supplementary Table 12). We have reported that agronomic traits and environmental stress responses are associated with quinoa genetic populations (Mizuno et al., 2020). In quinoa, plant color, derived from betalain, is one of the major phenotypic traits. We observed color differences among the highland lines J075 and J100, and the lowland line J082 (Figure 9B), which is the inbred line derived from PI614886, the same accession as QQ74 (Mizuno et al., 2020; Ogata et al., 2021). A comparison of gene-level synteny suggested that the highland lines J075 and J100 retained more

copies of betalain biosynthesis genes in clustered regions of chromosomes than the lowland reference line (Figure 9). The color differences among lines should be considered in light of the low conservation of these regions and the influence of other genes in the betalain biosynthetic pathway. Although several sites of structural variation were identified in the sequence comparisons between lines (Figure 8), further validation is needed to determine which gene families associated with structural variation lead to the various phenotypic changes observed among quinoa populations. The comparative genomic analysis platform for lines J075 and J100, representing the northern and southern highland populations, has the potential to expand our understanding of genetic diversity among quinoa populations and will accelerate the identification of genes associated with phenotypes.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <https://www.ddbj.nig.ac.jp/PRJDB18071> and <https://www.ncbi.nlm.nih.gov/PRJDB18071>. The genome sequence information generated in this study is also available at Plant GARDEN (<https://plantgarden.jp>).

Author contributions

YK: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. HH: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing, Funding acquisition. KS: Data curation, Formal analysis, Investigation, Methodology, Writing – review & editing, Funding acquisition. KN: Formal analysis, Investigation, Methodology, Writing – review & editing, Funding acquisition. KF: Investigation, Writing – review & editing. RO: Writing – review & editing, Project administration, Conceptualization. GA: Conceptualization, Writing – review & editing, Project administration. YN: Funding acquisition, Project administration, Resources, Writing – review & editing, Investigation, Visualization. YY: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Writing – review & editing. YF: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work

was supported by Grants-in-Aid for Scientific Research (KAKENHI) from the Japan Society for the Promotion of Science (JSPS) (Grant Nos. JP22K05374 to YK; JP22H05172 and JP22H05181 to KS; JP23KK0113 to KN, YY and YF; JP21H02158 and JP23K18036 to YN and YF), the Cabinet Office, Government of Japan, Moonshot Research and Development Program for Agriculture, Forestry and Fisheries (Funding agency: Bio-oriented Technology Research Advancement Institution; Grant No. JPJ009237), the Japan International Cooperation Agency (JICA) for the Science and Technology Research Partnership for Sustainable Development (SATREPS; Grant No. JPMJSA1907), the Kazusa DNA Research Institute Foundation, and the Japan International Research Center for Agricultural Sciences (JIRCAS) under the Ministry of Agriculture, Forestry and Fisheries (MAFF) of Japan.

Acknowledgments

We thank the staff of JIRCAS, M. Toyoshima, N. Hisatomi, Y. Masamura, K. Ozawa, Y. Shirai, N. Saito, Y. Nakamura, Y. Nonoue, Y. Takiguchi, A. Aoyama, Y. Ishino, N. Seko, T. Nada, and M. Nozawa for their excellent technical assistance, T. Ogata, Y. Murata, and T. Kashiwa for their useful discussion and support, the staff of the Kazusa DNA Research Institute, Y. Kishida, C. Minami, K. Ozawa, H. Tsuruoka, and A. Watanabe for skillful technical support, Y. Tokura, K. Katsura, I. Molaes, A. Bonifacio, W. Rojas, M. Patricia, J. Quezada, Y. Flores, J. Tanaka, S. Kushino for their cooperation in quinoa research. Computations were

performed in part on the NIG supercomputer at ROIS National Institute of Genetics.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2024.1434388/full#supplementary-material>

References

- Alandia, G., Rodriguez, J. P., Jacobsen, S. E., Bazile, D., and Condori, B. (2020). Global expansion of quinoa and challenges for the Andean region. *Glob. Food Secur.* 26, 100429. doi: 10.1016/j.gfs.2020.100429
- Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., et al. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biol.* 23, 258. doi: 10.1186/s13059-022-02823-7
- Bazile, D., Pulvento, C., Verniau, A., Al-Nusairi, M. S., Ba, D., Breidy, J., et al. (2016). Worldwide evaluations of quinoa: preliminary results from post international year of quinoa FAO projects in nine countries. *Front. Plant Sci.* 7. doi: 10.3389/fpls.2016.00850
- Bodrug-Schepers, A., Stralis-Pavese, N., Buerstmayr, H., Dohm, J. C., and Himmelbauer, H. (2021). Quinoa genome assembly employing genomic variation for guided scaffolding. *Theor. Appl. Genet.* 134, 3577–3594. doi: 10.1007/s00122-021-03915-x
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bonifacio, A. (2019). Improvement of Quinoa (*Chenopodium quinoa* Willd.) and Qañawa (*Chenopodium pallidicaule* Aellen) in the context of climate change in the high Andes. *Cien. Inv. Agr.* 46, 113–124. doi: 10.7764/RCA
- Bruna, T., Hoff, K. J., Lomsadze, A., Stanke, M., and Borodovsky, M. (2021). BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom. Bioinform.* 3, lqaa108. doi: 10.1093/nargab/lqaa108
- Bruna, T., Lomsadze, A., and Borodovsky, M. (2020). GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genom. Bioinform.* 2, lqaa026. doi: 10.1093/nargab/lqaa026
- Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Burrieza, H. P., Rizzo, A. J., Vale, E. M., Silveira, V., and Maldonado, S. (2019). Shotgun proteomic analysis of quinoa seeds reveals novel lysine-rich seed storage globulins. *Food Chem.* 293, 299–306. doi: 10.1016/j.foodchem.2019.04.098
- Cabanettes, F., and Klopp, C. (2018). D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6, e4958. doi: 10.7717/peerj.4958
- Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P., and Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol. Biol. Evol.* 38, 5825–5829. doi: 10.1093/molbev/msab293
- Chen, H., and Zwaenepoel, A. (2023). Inference of ancient polyploidy from genomic data. In: Van de Peer Y (ed.), *Polyploidy: Methods and protocols*. (New York: Springer US), 2545, 3–18.
- Cheng, H. Y., Concepcion, G. T., Feng, X. W., Zhang, H. W., and Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18, 170–175. doi: 10.1038/s41592-020-01056-5
- Christensen, S. A., Pratt, D. B., Pratt, C., Nelson, P. T., Stevens, M. R., Jellen, E. N., et al. (2007). Assessment of genetic diversity in the USDA and CIP-FAO international nursery collections of quinoa (*Chenopodium quinoa* Willd.) using microsatellite markers. *Plant Genet. Res.* 5, 82–95. doi: 10.1017/S1479262107672293
- Conway, J. R., Lex, A., and Gehlenborg, N. (2017). UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* 33, 2938–2940. doi: 10.1093/bioinformatics/btx364
- Dakhili, S., Abdolalizadeh, L., Hosseini, S. M., Shojaee-Aliabadi, S., and Mirmoghataie, L. (2019). Quinoa protein: Composition, structure and functional properties. *Food Chem.* 299, 125161. doi: 10.1016/j.foodchem.2019.125161
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFtools. *Gigascience* 10, giab008. doi: 10.1093/gigascience/giab008

- Demuth, J. P., and Hahn, M. W. (2009). The life and death of gene families. *Bioessays* 31, 29–39. doi: 10.1002/bies.080085
- Dillehay, T. D., Rossen, J., Andres, T. C., and Williams, D. E. (2007). Pre-ceramic adoption of peanut, squash, and cotton in northern Peru. *Science* 316, 1890–1893. doi: 10.1126/science.1141395
- Dirzo, R., and Raven, P. H. (2003). Global state of biodiversity and loss. *Annu. Rev. Environ. Resour.* 28, 137–167. doi: 10.1146/annurev.energy.28.050302.105532
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 117, 9451–9457. doi: 10.1073/pnas.1921046117
- Goel, M., Sun, H., Jiao, W. B., and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 20, 277. doi: 10.1186/s13059-019-1911-0
- Gomez-Pando, L. (2015). “Quinoa breeding.” *Quinoa: Improvement and Sustainable Production*. Eds. K. Murphy and J. Matanguihan (Hoboken, New Jersey: John Wiley & Sons, Inc), 87–107.
- Gomez-Pando, L. R., Aguilar-Castellanos, E., and Ibañez-Tremolada, M. (2019). “Quinoa (*Chenopodium quinoa* Willd.) Breeding,” in *the Advances in Plant Breeding Strategies: Cereals Volume 5*. Eds. J. M. Al-Khayri, S. M. Jain and D. V. Johnson, (Cham: Springer), 259–316.
- González, J. A., Eisa, S. S. S., Hussin, S. A. E. S., and Prado, F. E. (2015). “Quinoa: An Incan Crop to Face Global Changes in Agriculture,” in *the Quinoa: Improvement and Sustainable Production*. Eds. K. Murphy and J. Matanguihan (Hoboken, New Jersey: John Wiley & Sons, Inc), 1–18.
- Gotoh, O., Morita, M., and Nelson, D. R. (2014). Assessment and refinement of eukaryotic gene structure prediction with gene-structure-aware multiple protein sequence alignment. *BMC Bioinf.* 15, 189. doi: 10.1186/1471-2105-15-189
- Hanada, K., Zou, C., Lehti-Shiu, M. D., Shinozaki, K., and Shiu, S. H. (2008). Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. *Plant Physiol.* 148, 993–1003. doi: 10.1104/pp.108.122457
- Hariadi, Y., Marandon, K., Tian, Y., Jacobsen, S. E., and Shabala, S. (2011). Ionic and osmotic relations in quinoa (*Chenopodium quinoa* Willd.) plants grown at various salinity levels. *J. Exp. Bot.* 62, 185–193. doi: 10.1093/jxb/erq257
- Hart, A. J., Ginzburg, S., Xu, M. S., Fisher, C. R., Rahmatpour, N., Mitton, J. B., et al. (2020). EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Mol. Ecol. Resour.* 20, 591–604. doi: 10.1111/1755-0998.13106
- Hein, A. (1957). “Contributions to the knowledge of the virus diseases of weeds. II. The Lucerne mosaic and the Lamium yellow mosaic viruses. *Phytopathologische Zeitschrift* 29, 79–116.
- Hemalatha, P., Bomzan, D. P., Rao, B. V. S., and Sreerama, Y. N. (2016). Distribution of phenolic antioxidants in whole and milled fractions of quinoa and their inhibitory effects on α -amylase and α -glucosidase activities. *Food Chem.* 199, 330–338. doi: 10.1016/j.foodchem.2015.12.025
- Hirose, Y., Fujita, T., Ishii, T., and Ueno, N. (2010). Antioxidative properties and flavonoid composition of *Chenopodium quinoa* seeds cultivated in Japan. *Food Chem.* 119, 1300–1306. doi: 10.1016/j.foodchem.2009.09.008
- Hoff, K. J., and Stanke, M. (2019). Predicting genes in single genomes with AUGUSTUS. *Curr. Protoc. Bioinf.* 65, e57. doi: 10.1002/cpbi.57
- Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M. C., et al. (2016). eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248
- Imamura, T., Isozumi, N., Higashimura, Y., Miyazato, A., Mizukoshi, H., Ohki, S., et al. (2019). Isolation of *amaranthin synthetase* from *Chenopodium quinoa* and construction of an amaranthin production system using suspension-cultured tobacco BY-2 cells. *Plant Biotechnol. J.* 17, 969–981. doi: 10.1111/pbi.13032
- Iwata, H., and Gotoh, O. (2012). Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* 40, e161. doi: 10.1093/nar/gks708
- Jacobsen, S.-E. (2003). The worldwide potential for quinoa (*Chenopodium quinoa* Willd.). *Food Rev. Int.* 19, 167–177. doi: 10.1081/FRI-120018883
- Jarvis, D. E., Ho, Y. S., Lightfoot, D. J., Schmockel, S. M., Li, B., Borm, T. J., et al. (2017). The genome of *Chenopodium quinoa*. *Nature* 542, 307–312. doi: 10.1038/nature21370
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42, D199–D205. doi: 10.1093/nar/gkt1076
- Keilwagen, J., Hartung, F., and Grau, J. (2019). “GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data,” in *the Gene prediction: Methods and protocols*. Ed. M. Kollmar (New York: Humana), 161–177.
- Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A., et al. (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, D807–D811. doi: 10.1093/nar/gky1053
- Krug, A. S., Drummond, E., B., Van Tassel, D. L., and Warschefsky, E. J. (2023). The next era of crop domestication starts now. *Proc. Natl. Acad. Sci. U.S.A.* 120, e2205769120. doi: 10.1073/pnas.2205769120
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv* 1303, 3997. doi: 10.48550/arXiv.1303.3997
- Li, H. (2021). New strategies to improve minimap2 alignment accuracy. *Bioinformatics* 37, 4572–4574. doi: 10.1093/bioinformatics/btab705
- Li, Z. W., Hou, X. H., Chen, J. F., Xu, Y. C., Wu, Q., Gonzalez, J., et al. (2018). Transposable elements contribute to the adaptation of arabidopsis thaliana. *Genome Biol. Evol.* 10, 2140–2150. doi: 10.1093/gbe/evy171
- Mangelson, H., Jarvis, D. E., Mollinedo, P., Rollano-Penalosa, O. M., Palma-Encinas, V. D., Gomez-Pando, L. R., et al. (2019). The genome of *Chenopodium pallidicaule*: An emerging Andean super grain. *Appl. Plant Sci.* 7, e11300. doi: 10.1002/aps.3.11300
- Massawe, F., Mayes, S., and Cheng, A. (2016). Crop diversity: an unexploited treasure trove for food security. *Trends Plant Sci.* 21, 365–368. doi: 10.1016/j.tplants.2016.02.006
- Maughan, P. J., Bonifacio, A., Jellen, E. N., Stevens, M. R., Coleman, C. E., Ricks, M., et al. (2004). A genetic linkage map of quinoa (*Chenopodium quinoa*) based on AFLP, RAPD, and SSR markers. *Theor. Appl. Genet.* 109, 1188–1195. doi: 10.1007/s00122-004-1730-9
- Mayes, S., Massawe, F. J., Alderson, P. G., Roberts, J. A., Azam-Ali, S. N., and Hermann, M. (2012). The potential for underutilized crops to improve security of food production. *J. Exp. Bot.* 63, 1075–1079. doi: 10.1093/jxb/err396
- Meyer, R. S., and Purugganan, M. D. (2013). Evolution of crop species: genetics of domestication and diversification. *Nat. Rev. Genet.* 14, 840–852. doi: 10.1038/nrg3605
- Miller, M. J., Kendall, L., Capriles, J. M., Bruno, M. C., Evershed, R. P., and Hastorf, C. A. (2021). Quinoa, potatoes, and llamas fueled emergent social complexity in the Lake Titicaca Basin of the Andes. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2113395118. doi: 10.1073/pnas.2113395118
- Mizuno, N., Toyoshima, M., Fujita, M., Fukuda, S., Kobayashi, Y., Ueno, M., et al. (2020). The genotype-dependent phenotypic landscape of quinoa in salt tolerance and key growth traits. *DNA Res.* 27, dsaa022. doi: 10.1093/dnares/dsaa022
- Motta, C., Castanheira, I., Gonzales, G. B., Delgado, I., Torres, D., Santos, M., et al. (2019). Impact of cooking methods and malting on amino acids content in amaranth, buckwheat and quinoa. *J. Food Compos. Anal.* 76, 58–65. doi: 10.1016/j.jfca.2018.10.001
- Motta, C., Delgado, I., Matos, A. S., Gonzales, G. B., Torres, D., Santos, M., et al. (2017). Foliates in quinoa (*Chenopodium quinoa*), amaranth (*Amaranthus* sp.) and buckwheat (*Fagopyrum esculentum*): Influence of cooking and malting. *J. Food Compos. Anal.* 64, 181–187. doi: 10.1016/j.jfca.2017.09.003
- NAS (1975). *Underexploited tropical plants with promising economic value* (Washington, DC: National Academy of Sciences).
- NRC (1989). *Lost Crops of the Incas: Little known plants of the Andes Promise for Worldwide Cultivation*. (Washington, DC: National Academy Press).
- Navruz-Varli, S., and Sanlier, N. (2016). Nutritional and health benefits of quinoa (*Chenopodium quinoa* Willd.). *J. Cereal Sci.* 69, 371–376. doi: 10.1016/j.jcs.2016.05.004
- Nishimura, K., Kokaji, H., Motoki, K., Yamazaki, A., Nagasaka, K., Mori, T., et al. (2024). Degenerate oligonucleotide primer MIG-seq: an effective PCR-based method for high-throughput genotyping. *Plant J* 118, 2296–2317. doi: 10.1111/tjpi.16708
- Nowak, V., Du, J., and Charrodiere, U. R. (2016). Assessment of the nutritional composition of quinoa (*Chenopodium quinoa* Willd.). *Food Chem.* 193, 47–54. doi: 10.1016/j.foodchem.2015.02.111
- Ogata, T., Toyoshima, M., Yamamizo-Oda, C., Kobayashi, Y., Fujii, K., Tanaka, K., et al. (2021). Virus-mediated transient expression techniques enable functional genomics studies and modulations of betalain biosynthesis and plant height in quinoa. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.643499
- Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* 46, e126. doi: 10.1093/nar/gky730
- Ou, S., and Jiang, N. (2018). LTR retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176, 1410–1422. doi: 10.1104/pp.17.01310
- Palomino, G., Hernández, L. T., and Torres, E. D. (2008). Nuclear genome size and chromosome analysis in *Chenopodium quinoa* and *C. berlandieri* subsp. *nutalliae*. *Euphytica* 164, 221–230. doi: 10.1007/s10681-008-9711-8
- Pathan, S., Eivazi, F., Valliyodan, B., Paul, K., Ndunguru, G., and Clark, K. (2019). Nutritional composition of the green leaves of quinoa (*Chenopodium quinoa* willd.). *J. Food Res.* 8, 55–65. doi: 10.5539/jfr.v8n6p55
- Peñas, E., Uberti, F., di Lorenzo, C., Ballabio, C., Brandolini, A., and Restani, P. (2014). Biochemical and immunochemical evidences supporting the inclusion of quinoa (*Chenopodium quinoa* willd.) as a gluten-free ingredient. *Plant Foods Hum. Nutr.* 69, 297–303. doi: 10.1007/s11130-014-0449-2
- Prego, I., Maldonado, S., and Otegui, M. (1998). Seed structure and localization of reserves in *Chenopodium quinoa*. *Ann. Bot.* 82, 481–488. doi: 10.1006/anbo.1998.0704
- Rastas, P. (2017). Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics* 33, 3726–3732. doi: 10.1093/bioinformatics/btx494

- Ren, G., Teng, C., Fan, X., Guo, S., Zhao, G., Zhang, L., et al. (2023). Nutrient composition, functional activity and industrial applications of quinoa (*Chenopodium quinoa* Willd.). *Food Chem.* 410, 135290. doi: 10.1016/j.foodchem.2022.135290
- Rey, E., Maughan, P. J., Maumus, F., Lewis, D., Wilson, L., Fuller, J., et al. (2023). A chromosome-scale assembly of the quinoa genome provides insights into the structure and dynamics of its subgenomes. *Commun. Biol.* 6, 1263. doi: 10.1038/s42003-023-05613-4
- Rodriguez, J. P., Rahman, H., Thushar, S., and Singh, R. K. (2020). Healthy and resilient cereals and pseudo-cereals for marginal agriculture: molecular advances for improving nutrient bioavailability. *Front. Genet.* 11. doi: 10.3389/fgene.2020.00049
- Ruales, J., and Nair, B. M. (1992). Nutritional quality of the protein in quinoa (*Chenopodium-quinoa*, willd) seeds. *Plant Foods Hum. Nutr.* 42, 1–11. doi: 10.1007/BF02196067
- Schlick, G., and Bubenheim, D. L. (1993). "Quinoa: An emerging new crop with potential for CELSS," in *NASA technical paper* (Moffett Field, CA), 3422.
- Schlick, G., and Bubenheim, D. L. (1996). *Quinoa: candidate crop for NASA's controlled ecological life support systems* (Alexandria: ASHS), 632-640.
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351
- Stafford, H. A. (1994). Anthocyanins and betalains: evolution of the mutually exclusive pathways. *Plant Sci.* 101, 91–98. doi: 10.1016/0168-9452(94)90244-5
- Sun, J., Lu, F., Luo, Y., Bie, L., Xu, L., and Wang, Y. (2023). OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res.* 51, W397–W403. doi: 10.1093/nar/gkad313
- Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., and Paterson, A. H. (2008). Synteny and collinearity in plant genomes. *Science* 320, 486–488. doi: 10.1126/science.1153917
- Tang, H., Zhang, X., Miao, C., Zhang, J., Ming, R., Schnable, J. C., et al. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. *Genome Biol.* 16, 3. doi: 10.1186/s13059-014-0573-1
- Uschdraweit, H. (1955). "The green mottle mosaic of Cucumber. *Chenopodium quinoa* as test plant for Cucumber mosaic. *NachrBl. dtsh. PflSchDienst.* 7, 150–151.
- Vilcacundo, R., and Hernández-Ledesma, B. (2017). Nutritional and biological value of quinoa (*Chenopodium quinoa* Willd.). *Curr. Opin. Food Sci.* 14, 1–6. doi: 10.1016/j.cofs.2016.11.007
- Yangquanwei, Z., Neethirajan, S., and Karunakaran, C. (2013). Cytogenetic analysis of quinoa chromosomes using nanoscale imaging and spectroscopy techniques. *Nanoscale Res. Lett.* 8, 463. doi: 10.1186/1556-276X-8-463
- Yasui, Y., Hirakawa, H., Oikawa, T., Toyoshima, M., Matsuzaki, C., Ueno, M., et al. (2016). Draft genome sequence of an inbred line of *Chenopodium quinoa*, an allotetraploid crop with great environmental adaptability and outstanding nutritional properties. *DNA Res.* 23, 535–546. doi: 10.1093/dnares/dsw037
- Zou, C., Chen, A., Xiao, L., Muller, H. M., Ache, P., Haberer, G., et al. (2017). A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Res.* 27, 1327–1340. doi: 10.1038/cr.2017.124
- Zurita-Silva, A., Fuentes, F., Zamora, P., Jacobsen, S.-E., and Schwember, A. R. (2014). Breeding quinoa (*Chenopodium quinoa* Willd.): potential and perspectives. *Mol. Breed.* 34, 13–30. doi: 10.1007/s11032-014-0023-5