

Resource Article: Genomes Explored

Chromosome-level genome assembly of Japanese chestnut (*Castanea crenata* Sieb. et Zucc.) reveals conserved chromosomal segments in woody rosids

Kenta Shirasawa ^{1*}, Sogo Nishio², Shingo Terakami², Roberto Botta³, Daniela Torello Marinoni³, and Sachiko Isobe ¹

¹Kazusa DNA Research Institute, Chiba 292-0818, Japan, ²Institute of Fruit Tree and Tea Science, NARO, Ibaraki 305-8605, Japan, and ³Dipartimento di Scienze Agrarie, Forestali e Alimentari, Università degli Studi di Torino, Largo P. Braccini 2, 10095 Grugliasco (TO), Italy

*To whom correspondence should be addressed. Tel. +81 438 52 3935. Fax. +81 438 52 3934.
Email: shirasaw@kazusa.or.jp

Received 29 July 2021; Editorial decision 20 August 2021; Accepted 20 August 2021

Abstract

Japanese chestnut (*Castanea crenata* Sieb. et Zucc.), unlike other *Castanea* species, is resistant to most diseases and wasps. However, genomic data of Japanese chestnut that could be used to determine its biotic stress resistance mechanisms have not been reported to date. In this study, we employed long-read sequencing and genetic mapping to generate genome sequences of Japanese chestnut at the chromosome level. Long reads (47.7 Gb; 71.6× genome coverage) were assembled into 781 contigs, with a total length of 721.2 Mb and a contig N50 length of 1.6 Mb. Genome sequences were anchored to the chestnut genetic map, comprising 14,973 single nucleotide polymorphisms (SNPs) and covering 1,807.8 cM map distance, to establish a chromosome-level genome assembly (683.8 Mb), with 69,980 potential protein-encoding genes and 425.5 Mb repetitive sequences. Furthermore, comparative genome structure analysis revealed that Japanese chestnut shares conserved chromosomal segments with woody plants, but not with herbaceous plants, of rosids. Overall, the genome sequence data of Japanese chestnut generated in this study is expected to enhance not only its genetics and genomics but also the evolutionary genomics of woody rosids.

Key words: chestnut, genetic mapping, genome synteny, rosids, whole-genome sequence

1. Introduction

Chestnut is naturally distributed in temperate regions in the northern hemisphere and plays a critical role in maintaining the landscape and forest ecosystem.¹ Seeds or nuts of chestnut trees provide important nutrition for animals as well as humans, while their wood is used as a source of timber and tannins. The genus *Castanea* ($2n = 2x = 24$) includes four cultivated species: Japanese chestnut (*Castanea crenata* Sieb. et Zucc.), American chestnut [*Castanea dentata* (Marshall)

Borkh.], Chinese chestnut (*Castanea mollissima* Bl.), and European chestnut (*Castanea sativa* Mill.). Japanese chestnut is commercially cultivated in Japan and Korea, while the remaining three species are commercially grown in the USA, China, and Europe, respectively. Seguin chestnut (*Castanea seguinii* Dode.) and Henry chestnut (*Castanea henryi* Rehd. et Wils.), found in China, and Allegheny chinkapin (*Castanea pumila* Mill.), which grows in the USA, are recognized as wild species. The cultivated and wild

species are cross-compatible, and the hybrids have been used to develop genetic maps.^{2–5}

In the early 19th century, the American chestnut was devastated by chestnut blight, caused by an exotic pathogen, namely, *Cryphonectria parasitica*.⁶ European chestnut has also been damaged by chestnut blight and ink disease, caused by *C. parasitica* and *Phytophthora cinnamomi*, respectively.⁷ In addition, chestnut gall wasp (*Dryocosmus kuriphilus*) was accidentally introduced into Europe, resulting in great damage to chestnut production.⁴ Because Japanese as well as Chinese chestnut cultivars are generally resistant to diseases and wasps, they have been used as the sources for resistance in several chestnut breeding programs.⁷

Genome sequence information is useful for not only accelerating breeding programs but also understanding the phylogenetic relationships and revealing the evolutionary history of a given species through comparative genome structure analysis. To the best of our knowledge, genome sequences of three Chinese chestnut lines and one Henry chestnut line have been reported to date.^{8–11} However, the whole-genome sequence of Japanese chestnut is not yet publicly available. Advances in long-read sequencing technology have enabled the construction of a mega-base-scale genome assembly of haploid, diploid, and polyploid species.¹² Long contig sequences generated by long-read sequencing could be further concatenated at the chromosome level with high-throughput chromosome conformation analyses, optical mapping, and/or genetic mapping to establish pseudomolecule sequences.¹³ In this study, using long-read sequencing and genetic mapping, we established chromosome-level pseudomolecule sequences of the Japanese chestnut variety ‘Ginyose’, which has contributed to the spread of chestnut cultivation in Japan owing to its large nut size and excellent appearance. Despite its long history of cultivation that dates back to the 17th century, ‘Ginyose’ is still a major cultivar in Japan, accounting for ~15% of the national chestnut production. Genetic structure and parentage analysis shows that ‘Ginyose’ is the predecessor of many Japanese cultivars.¹⁴

2. Materials and methods

2.1. Plant materials and DNA extraction

A tree of the Japanese chestnut cultivar ‘Ginyose’ planted at the Institute of Fruit Tree and Tea Science, NARO (Tsukuba, Japan) was used for genome sequencing. Genomic DNA was extracted from tree leaves with Genomic-tips (Qiagen, Hilden, Germany).

2.2. Genome sequencing and analysis

Software tools used for data analyses are listed in [Supplementary Table S1](#). Genomic DNA libraries for short- and long-read sequencing were generated with the TruSeq DNA PCR-Free Kit (Illumina, San Diego, CA, USA) and the SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA), respectively. Short-read sequence data were obtained using the HiSeq2000 platform (Illumina). Genome size was estimated using Jellyfish after removing adaptor sequences (AGATCGGAAGAGC) and reads obtained from organelle genomes (GenBank accession numbers: HQ336406 and MN199236). Long-read sequence data were obtained using the Sequel system (PacBio), and primary contigs and alternate contigs, which were generated from one allele and the other, respectively, of diploid genomes, were assembled with Falcon. Then, haplotype sequences were resolved to generate haplotigs with Falcon_Unzip. Sequence errors in the contigs were corrected twice using long reads with ARROW, and potential haplotype duplications in the primary

contigs were removed with Purge_Dups. Contig sequences potentially contaminated from organelle genomes (GenBank accession numbers: HQ336406 and MN199236), which showed sequence similarity of >90% with Minimap2, were removed. Assembly completeness was evaluated with the embryophyta_odb10 data using Benchmarking Universal Single-Copy Orthologs (BUSCO).

2.3. Genetic map-based chromosome-level sequence construction

Reads obtained by double digest restriction site-associated DNA sequencing (ddRAD-Seq) of an F1 mapping population ($n = 185$), derived from a cross between ‘Bouche de Bétizac’ (*C. sativa* × *C. crenata*) and ‘Madonna’ (*C. sativa*),⁴ were mapped onto the primary contigs as a reference using Bowtie2, and sequence variants were detected with BCFtools. High-confidence single nucleotide polymorphisms (SNPs) were selected with VCFtools (parameters of $-\text{minDP } 5$ $-\text{minQ } 999$ $-\text{max-missing } 0.5$ $-\text{maf } 0.2$ $-\text{max-maf } 0.8$) and subjected to linkage analysis with Lep-Map3. The nomenclature and direction of the linkage groups were based on the map reported by Torello Marinoni *et al.*⁴ To construct pseudomolecule sequences at the chromosome level, primary contigs were assigned to the resultant genetic map with ALLMAPS.

2.4. Repeat sequence analysis

Repetitive sequences in the assembly were identified with RepeatMasker using repeat sequences registered in Repbase and a *de novo* repeat library built with RepeatModeler. Repeat elements were classified into nine types with RepeatMasker: short interspersed nuclear elements (SINES), long interspersed nuclear elements (LINEs), long terminal repeat (LTR) elements, DNA elements, small RNA, satellites, simple repeats, low complexity repeats, and unclassified.

2.5. Gene prediction and annotation

Protein-coding genes were predicted using a MAKER pipeline, based on peptide sequences predicted in the *C. mollissima* genome (National Genomics Data Center accession number: GWHANWH00000000; $n = 33,597$)¹⁰ and transcriptome sequences of three *Castanea* species,⁶ *C. crenata* (cci_ccn, $n = 13,451$), *C. dentata* (AC454_v3, $n = 45,288$), and *C. sativa* (csi_csn, $n = 12,771$), released in the Hardwood Genomics Project (<https://doi.org/10.25504/FAIRsharing.srgkaf> Last access date: 31 Aug 2021). Short gene sequences (<300 bp), genes in repeat sequences, and genes with annotation edit distance (AED) >0.5, which is proposed as a threshold for good annotations in the MAKER protocol, were removed to facilitate the selection of high-confidence genes. Functional annotation of the predicted genes was performed with Hayai-Annotation Plants.

2.6. Comparative genome structure analysis

The genome structure of Japanese chestnut cultivar ‘Ginyose’ was aligned to the genome sequence of Chinese chestnut (*C. mollissima*), as a reference, with Unimap. In addition, publicly available chromosome-level sequences of 114 plants,¹³ which, in accordance with Angiosperm Phylogeny Group (APG) classification IV,¹⁵ consisted of 20 monocots, 93 eudicots (including 66 rosids and 23 asterids), and one member of Nymphaea, were also employed as references. Dot-plot charts based on the alignments were drawn with D-Genies.

3. Results and data description

3.1. Genome sequence and assembly

In accordance with *k*-mer frequency analysis of 63.3 Gb short reads, the genome size of *C. crenata* was estimated as 665.9 Mb (Fig. 1). A total of 2.7 million long reads (47.7 Gb, 71.6× genome coverage, *N*₅₀ = 29.8 kb) obtained from four single-molecule real-time (SMRT) cells were assembled into two haplotype-resolved sequences: primary contigs and haplotigs. The resultant assemblies of primary contigs (CCR_r1.0) and haplotigs (CCR_r1.0.haplotigs) consisted of 781 and 3,355 sequences, respectively (Table 1). Total lengths of primary contigs (CCR_r1.0) and haplotigs (CCR_r1.0.haplotigs) were 721 Mb (*N*₅₀ = 1.6 Mb) and 629.2 Mb (*N*₅₀ = 275.8 kb), respectively. The completeness of primary contigs indicated that 96.6% of sequences were complete BUSCOs (Table 1).

3.2. Pseudomolecule sequence construction

To construct a genetic map of chestnut, we used ddRAD-Seq data of an interspecific F1 mapping population reported previously.⁴ A total of 903 million high-quality ddRAD-Seq reads were obtained from the mapping population and parental lines were mapped onto the primary contigs. The mapping rate of the F1 progeny (mean 86.5%),

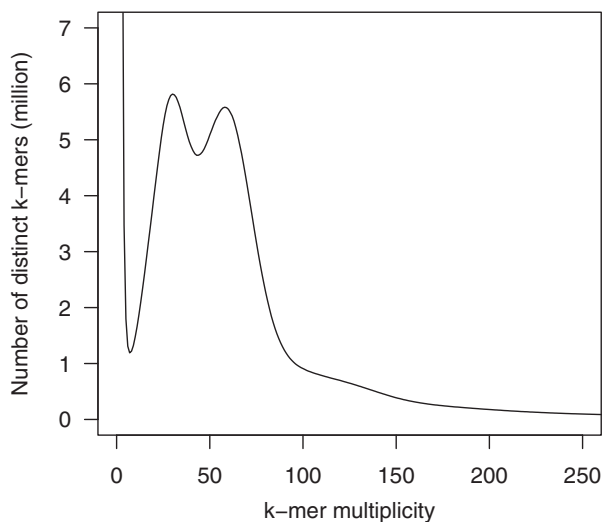


Figure 1. Estimation of the genome size of Japanese chestnut (*Castanea crenata*), based on *k*-mer analysis (*k* = 17) with the given multiplicity values.

Table 1. Statistics of the contig sequences of Japanese chestnut

| | CCR_r1.0 | CCR_r1.0.haplotig |
|--|-------------|-------------------|
| Total contig size (bases) | 721,168,657 | 629,166,897 |
| Number of contigs | 781 | 3,355 |
| Contig <i>N</i> ₅₀ length (bases) | 1,595,543 | 275,768 |
| Longest contig size (bases) | 13,981,633 | 1,688,135 |
| Gap (bases) | 0 | 0 |
| Complete BUSCOs | 96.6 | n.a. |
| Single-copy BUSCOs | 91.9 | n.a. |
| Duplicated BUSCOs | 4.7 | n.a. |
| Fragmented BUSCOs | 0.9 | n.a. |
| Missing BUSCOs | 2.5 | n.a. |
| #Genes | 69,980 | n.a. |

n.a., not analyzed.

‘Bouche de Bétizac’ (*C. sativa* × *C. crenata*) (86.9%), and ‘Madonna’ (*C. sativa*) (86.9%) were even, suggesting high sequence similarity between the genomes of *C. sativa* and *C. crenata*. SNPs were detected and filtered. Consequently, 15,364 high-quality SNPs were identified, of which 14,973 were grouped into 12 linkage groups, which corresponds to the number of chromosomes in *C. crenata*. Then, SNP order and the map distance between adjacent SNPs were calculated (Table 2, Supplementary Table S2). The nomenclature of linkage groups was based on our previous study.⁴ Since sequence and structure similarities were presumed between the genomes of *C. sativa* and *C. crenata*,^{2–5} the interspecific genetic map was used to establish pseudomolecule sequences of *C. crenata*. A total of 575 primary contigs (683.8 Mb) were assigned to this genetic map and connected with 100 Ns, while the remaining 206 contigs (37.4 Mb) were not assigned to any linkage groups (Table 2).

3.3. Repeat sequence analysis and gene prediction

Repetitive sequences (425.5 Mb) accounted for 59.0% of the Japanese chestnut genome assembly (Table 3). LTR retroelements were the most abundant repeat sequences (172.7 Mb, 24.0%), followed by unclassified repeat sequences (154.0 Mb, 21.3%), i.e. those unavailable in public databases.

A total of 195,950 potential protein-coding sequences were identified in the Japanese chestnut genome assembly, based on *ab initio* prediction and amino acid sequence homology among four *Castanea* species: *C. crenata*, *C. dentata*, *C. mollissima*, and *C. sativa*. Subsequently, protein-coding sequences were filtered to remove 89,194 genes with AED > 0.5, 13,389 protein-coding sequences in repeat sequences, 23,386 short sequences (< 300 bp), and one redundant sequence. Finally, 69,980 sequences were identified as high-confidence genes, including 83.6% complete BUSCOs. Functional annotation analysis showed that of the 69,980 predicted genes, 4,370, 7,803, and 5,219 sequences were assigned to Gene Ontology slim terms in the biological process, cellular component, and molecular function categories, respectively, and 1,267 genes had enzyme commission numbers (Supplementary Table S3).

Next, peptide sequences of *C. mollissima* and transcriptome sequences of *C. crenata*, *C. dentata*, and *C. sativa* were aligned to the genome sequence of Japanese chestnut obtained in this study. A total of 24,561 genes predicted in the Japanese chestnut genome sequence were covered by transcriptome sequences from at least one of the three *Castanea* species (Fig. 2), while the remaining 28,773 genes did not correspond to transcriptome sequences from any *Castanea* species.

3.4. Comparative genome structure analysis

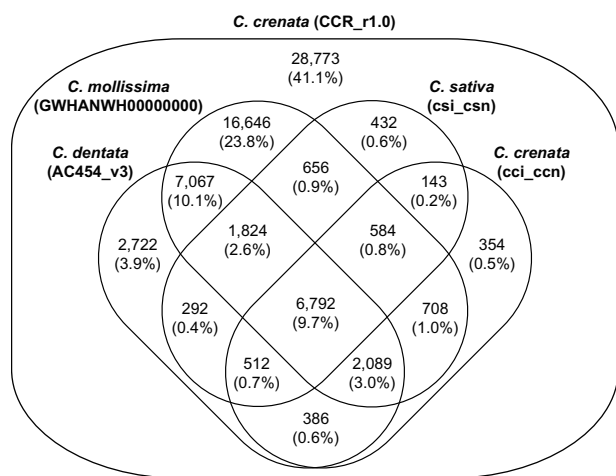
The genome structure of the Japanese chestnut was well-conserved with that of Chinese chestnut (Fig. 3), as expected. Interestingly, the *C. crenata* genome locally showed sequence and structure similarity, or synteny, with the genomes of at least 14 plant species (*Acer yangbiense*, *Citrus sinensis*, *Fragaria vesca*, *Malus × domestica*, *Manihot esculenta*, *Populus trichocarpa*, *Prunus avium*, *Prunus dulcis*, *Prunus mume*, *Prunus persica*, *Pyrus betulifolia*, *Theobroma cacao*, *Vitis vinifera*, and *Ziziphus jujuba*) (Fig. 3), all of which represent woody tree species belonging to Malpighiales, Malvales, Rosales, Sapindales, and Vitales in rosids.¹⁵ On the other hand, the *C. crenata* genome exhibited much less sequence and structure similarity to genomes of Brassicales, Cucurbitales, and Fabales belonging to herbaceous rosids than the woody members (Supplementary Fig. S1).

Table 2. Statistics of Japanese chestnut pseudomolecule sequences (CCR_r1.0.pmol)

| Linkage group | No. of SNPs | Genetic distance (cM) | No. of bins | No. of contigs | % | Total length (bp) | % |
|---------------|-------------|-----------------------|-------------|----------------|------|-------------------|------|
| CCR1.0A | 2,483 | 173.7 | 284 | 59 | 7.6 | 89,228,932 | 12.4 |
| CCR1.0B | 1,514 | 172.2 | 202 | 53 | 6.8 | 53,118,407 | 7.4 |
| CCR1.0C | 1,232 | 152.6 | 204 | 44 | 5.6 | 56,655,719 | 7.9 |
| CCR1.0D | 1,083 | 137.5 | 161 | 57 | 7.3 | 52,875,950 | 7.3 |
| CCR1.0E | 1,342 | 165.2 | 228 | 61 | 7.8 | 71,160,265 | 9.9 |
| CCR1.0F | 1,339 | 132.3 | 154 | 42 | 5.4 | 61,824,531 | 8.6 |
| CCR1.0G | 924 | 137.2 | 153 | 44 | 5.6 | 44,683,286 | 6.2 |
| CCR1.0H | 816 | 138.9 | 170 | 34 | 4.4 | 53,252,922 | 7.4 |
| CCR1.0I | 1,199 | 157.9 | 168 | 42 | 5.4 | 45,426,588 | 6.3 |
| CCR1.0J | 1,036 | 148.0 | 157 | 32 | 4.1 | 40,655,448 | 5.6 |
| CCR1.0K | 1,026 | 144.9 | 196 | 62 | 7.9 | 67,434,223 | 9.4 |
| CCR1.0L | 979 | 147.4 | 140 | 45 | 5.8 | 47,439,869 | 6.6 |
| Total | 14,973 | 1,808 | 2,217 | 575 | 73.6 | 683,756,140 | 94.8 |

Table 3. Repetitive sequences in the Japanese chestnut genome

| Repeat type | No. of elements | Physical length (bp) | % |
|-----------------|-----------------|----------------------|------|
| SINEs | 3,222 | 349,743 | 0.0 |
| LINEs | 58,824 | 28,244,400 | 3.9 |
| LTR elements | 173,629 | 172,741,932 | 24.0 |
| DNA transposons | 206,227 | 47,812,722 | 6.6 |
| Unclassified | 672,305 | 153,965,018 | 21.3 |
| Small RNA | 1,548 | 851,212 | 0.1 |
| Satellites | 1,888 | 1,522,847 | 0.2 |
| Simple repeats | 308,508 | 12,661,150 | 1.8 |
| Low complexity | 43,949 | 2,095,795 | 0.3 |

**Figure 2.** Number of genes in the genomes of Japanese chestnut supported by members of the *Castanea*. Predicted genes in the genomes of Japanese chestnut (CCR_r1.0) and Chinese chestnut (GWHANWH00000000), and the transcriptome sequences of American chestnut (AC454_v3), European chestnut (csi_csn), and Japanese chestnut (cci_ccn).

4. Conclusion and future perspectives

Here, we report the chromosome-level genome sequence of Japanese chestnut for the first time. The genome information on Japanese chestnut could be used as a standard reference for genetics,

genomics, and breeding of Japanese chestnut, their relatives, and woody rosids.

Comparative genome structure analysis showed that the genome structure of Japanese chestnut was quite similar to that of Chinese chestnut (Fig. 3). This result suggests that agronomically important genetic loci controlling nut weight and harvesting time^{3,16,17} could be transferred from Chinese chestnut to Japanese chestnut and vice versa. Chromosomal segments in the Japanese chestnut genome shared synteny with the peach genome² (Fig. 3); this relationship has also been reported in Chinese chestnut. More interestingly, regions in the Japanese chestnut genome showed sequence and structure similarity with the genomes of Malpighiales, Malvales, Rosales, Sapindales, and Vitales belonging to woody rosids (Fig. 3), but not with Brassicales, Cucurbitales, and Fabales belonging to herbaceous rosids (Supplementary Fig. S1). Genome structure conservation among rosids, e.g. Rosales (sweet cherry, apple, and Japanese fig) and Malvales (cacao), have been suggested in our previous investigation of 114 plant species.¹³ This result suggests that woody members of rosids share a closely related genome sequence and structure with each other, although this relationship contradicts that suggested by the APG classification IV.¹⁵ Indeed, none of the plant species in Cucurbitales, which is suggested to be the closest order to Fagales according to APG IV,¹⁵ showed genome structure similarity to Japanese chestnut. Thus, comparative genome structure analysis provides new insights into the evolutionary history of angiosperm.

The genome sequence data generated in this study would enhance the genetics and genomics of Japanese chestnut, in which classical molecular markers, such as simple sequence repeats, previously played a key role in (i) the identification of genetic loci affecting agronomically important traits^{3,4,18}; (ii) chestnut breeding programs with marker-assisted selection¹⁹; and (iii) the assessment of the genetic structure of cultivated and wild chestnut populations.²⁰ Furthermore, the highly conserved genome sequence and structure of woody members of rosids might help us better understand their evolutionary history and facilitate the identification of common genetic mechanisms affecting agronomic traits across various orders, families, species, and genera.

Data availability

Raw sequence reads were deposited into the Sequence Read Archive (SRA) database of the DNA Data Bank of Japan (DDBJ) under

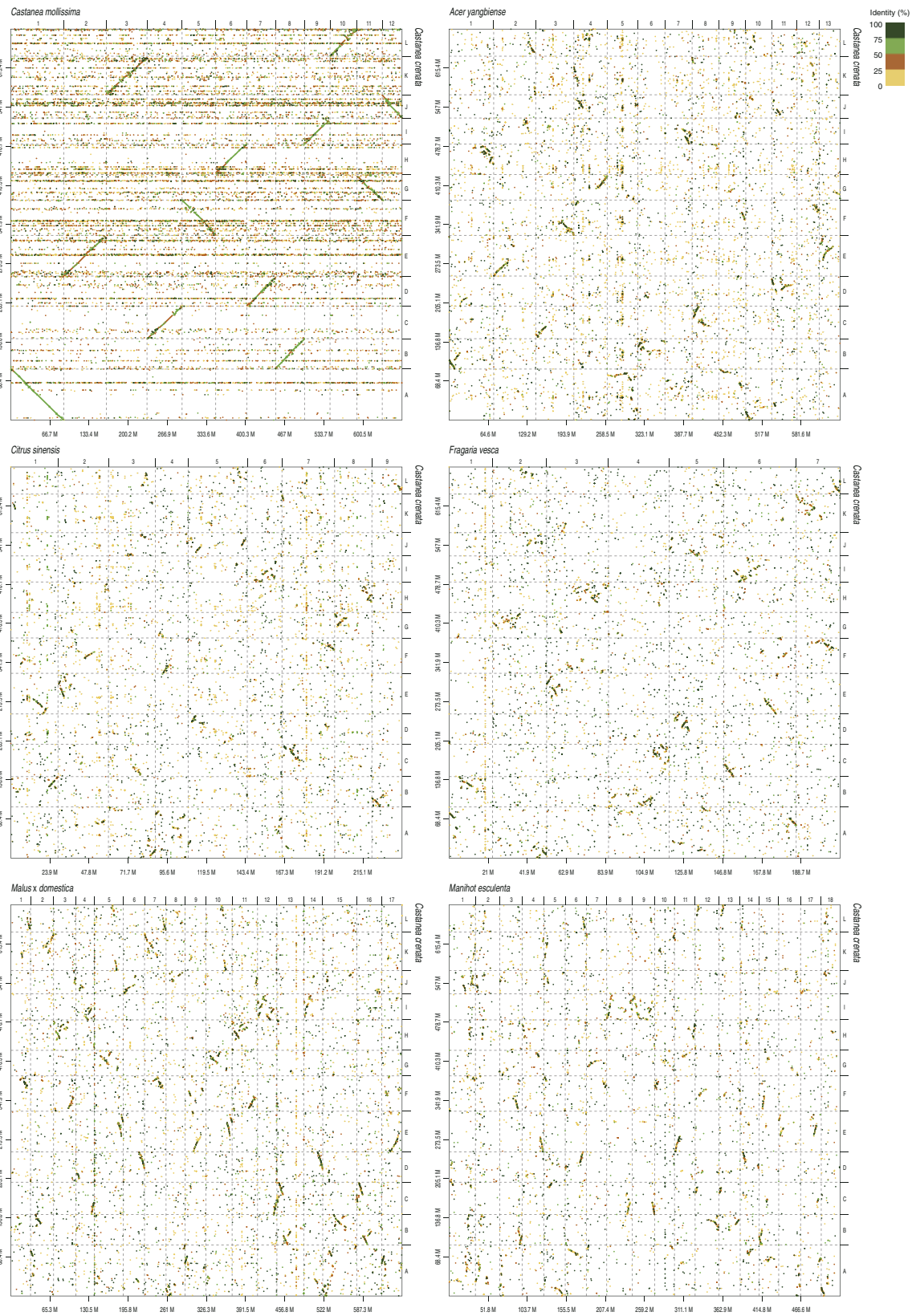


Figure 3. Comparative analysis of the genome sequence and structure of Japanese chestnut and other woody members of rosids. Dots and colors indicate genome structure and sequence similarity, respectively, among *Castanea mollissima*, *Acer yangbiense*, *Citrus sinensis*, *Fragaria vesca*, *Malus × domestica*, *Manihot esculenta*, *Populus trichocarpa*, *Prunus persica*, *Pyrus betulifolia*, *Theobroma cacao*, *Vitis vinifera*, and *Ziziphus jujuba*. Chromosome numbers are indicted on upper side of x-axis and right side of y-axis and genome sizes (Mb) are shown on lower side of x-axis and left side of y-axis.

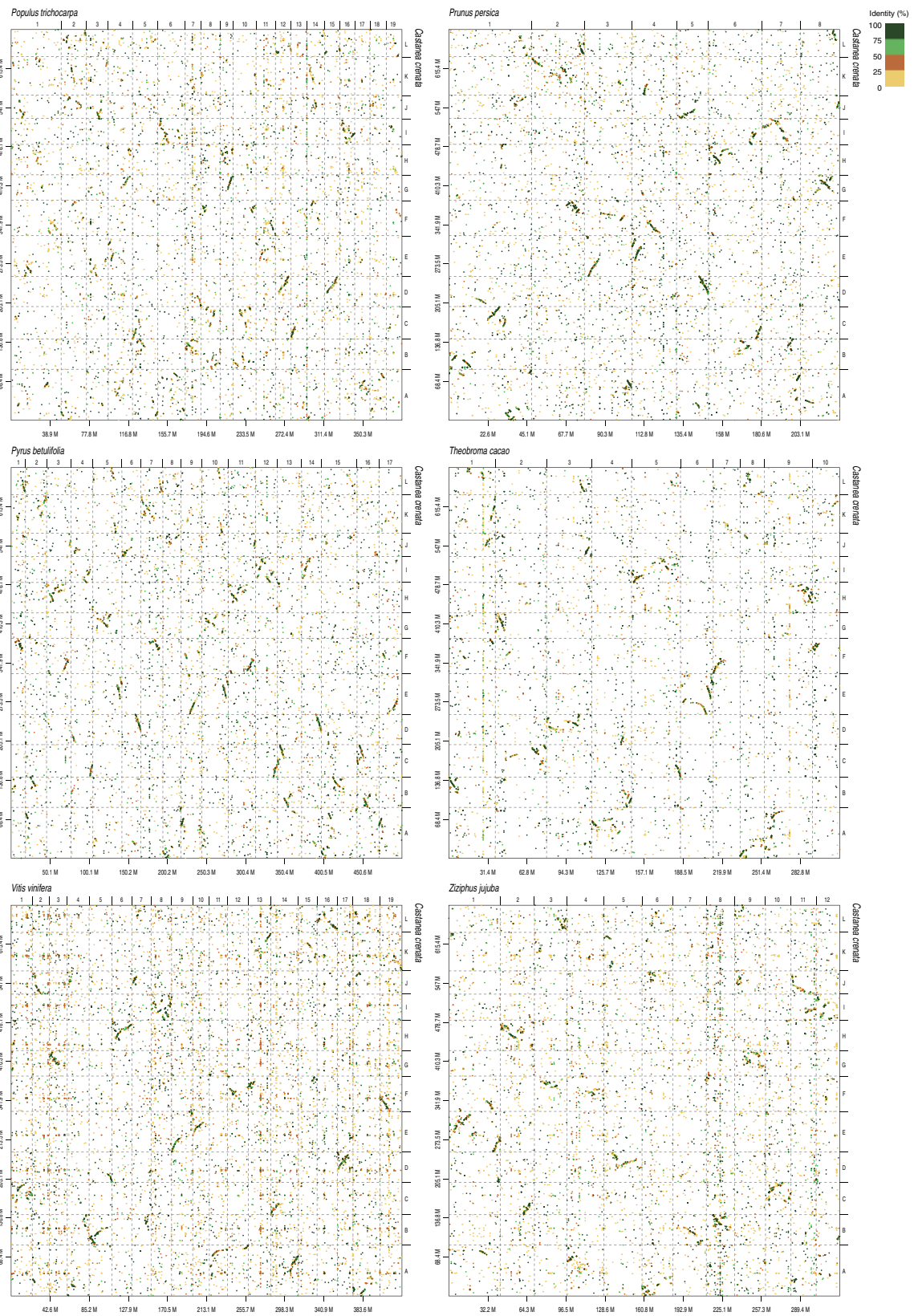


Figure 3. Continued.

accession number DRA012289. Assembled sequences are available at DDBJ under accession numbers BPMU01000001–BPMU01000781. Genome information generated in this study is available at Plant GARDEN (<https://plantgarden.jp>).

Acknowledgements

We thank Y. Kishida, C. Minami, H. Tsuruoka, and A. Watanabe (Kazusa DNA Research Institute) for providing technical assistance.

Funding

This work was supported in part by the JSPS KAKENHI (20K15524) and the Kazusa DNA Research Institute Foundation.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at DNARES online.

References

- Bounous, G. and Marinoni, D.T. 2004, Chestnut: botany, horticulture, and utilization, *Hortic. Rev.*, **31**, 291–347.
- Kubisiak, T.L., Nelson, C.D., Staton, M.E., et al. 2013, A transcriptome-based genetic map of Chinese chestnut (*Castanea mollissima*) and identification of regions of segmental homology with peach (*Prunus persica*), *Tree Genet. Genom.*, **9**, 557–71.
- Nishio, S., Terakami, S., Matsumoto, T., et al. 2018, Identification of QTLs for agronomic traits in the Japanese chestnut (*Castanea crenata* Sieb. et Zucc.) breeding. *Hortic. J.*, **87**, 43–54.
- Torello Marinoni, D., Nishio, S., Valentini, N., et al. 2020, Development of high-density genetic linkage maps and identification of loci for chestnut gall wasp resistance in *Castanea* spp., *Plants (Basel)*, **9**, 1048.
- Santos, C., Nelson, C.D., Zhebentyayeva, T., Machado, H., Gomes-Laranjo, J. and Costa, R.L. 2017, First interspecific genetic linkage map for *Castanea sativa* × *Castanea crenata* revealed QTLs for resistance to *Phytophthora cinnamomi*, *PLoS One*, **12**, e0184381.
- Barakat, A., DiLoreto, D.S., Zhang, Y., et al. 2009, Comparison of the transcriptomes of American chestnut (*Castanea dentata*) and Chinese chestnut (*Castanea mollissima*) in response to the chestnut blight infection, *BMC Plant Biol.*, **9**, 51.
- Pereira-Lorenzo, S., Ballester, A., Corredoira, E., et al. 2012, Chestnut. In: Badenes, M.L. and Byrne, D.H. (eds), *Fruit breeding*, Springer US, Boston, MA, p. 729–69.
- Staton, M., Addo-Quaye, C., Cannon, N., et al. 2020, A reference genome assembly and adaptive trait analysis of *Castanea mollissima* ‘Vanuxem,’ a source of resistance to chestnut blight in restoration breeding, *Tree Genet. Genom.*, **16**, 57.
- Sun, Y., Lu, Z., Zhu, X. and Ma, H. 2020, Genomic basis of homoploid hybrid speciation within chestnut trees, *Nat. Commun.*, **11**, 3375.
- Wang, J., Tian, S., Sun, X., et al. 2020, Construction of pseudomolecules for the Chinese chestnut (*Castanea mollissima*) genome, *G3 (Bethesda)*, **10**, 3565–74.
- Xing, Y., Liu, Y., Zhang, Q., et al. 2019, Hybrid de novo genome assembly of Chinese chestnut (*Castanea mollissima*), *Gigascience*, **8**, giz112.
- Sato, K., Abe, F., Mascher, M., et al. 2021, Chromosome-scale genome assembly of the transformation-amenable common wheat cultivar ‘Fielder’, *DNA Res.*, **28**, dsab008.
- Shirasawa, K., Harada, D., Hirakawa, H., Isobe, S. and Kole, C. 2021, Chromosome-level de novo genome assemblies of over 100 plant species, *Breed. Sci.*, **71**, 117–24.
- Nishio, S., Iketani, H., Fujii, H., et al. 2014, Use of population structure and parentage analyses to elucidate the spread of native cultivars of Japanese chestnut, *Tree Genet. Genom.*, **10**, 1171–80.
- The Angiosperm Phylogeny Group. 2016, An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV, *Bot. J. Linn. Soc.*, **181**, 1–20.
- Ji, F., Wei, W., Liu, Y., et al. 2018, Construction of a SNP-based high-density genetic map using genotyping by sequencing (GBS) and QTL analysis of nut traits in Chinese chestnut (*Castanea mollissima* Blume), *Front. Plant Sci.*, **9**, 816.
- Nishio, S., Hayashi, T., Yamamoto, T., et al. 2018, Bayesian genome-wide association study of nut traits in Japanese chestnut, *Mol. Breed.*, **38**, 99.
- Nishio, S., Takada, N., Yamamoto, T., et al. 2013, Mapping and pedigree analysis of the gene that controls the easy peel pellicle trait in Japanese chestnut (*Castanea crenata* Sieb. et Zucc), *Tree Genet. Genom.*, **9**, 723–30.
- Bernatzky, R. and Mulcahy, D.L. 1992, Marker-aided selection in a back-cross breeding program for resistance to chestnut blight in the American chestnut, *Can. J. For. Res.*, **22**, 1031–5.
- Nishio, S., Takada, N., Terakami, S., et al. 2021, Genetic structure analysis of cultivated and wild chestnut populations reveals gene flow from cultivars to natural stands, *Sci. Rep.*, **11**, 240.