Chromosome-scale haplotype-phased genome assemblies of the male and female lines of wild asparagus (*Asparagus kiusianus*), a dioecious plant species

Kenta Shirasawa 1 *,†, Saki Ueta 2 †, Kyoko Murakami 2, Mostafa Abdelrahman 3,4,5, Akira Kanno 3, and Sachiko Isobe 1

1 Kazusa DNA Research Institute, Kisarazu, Chiba 292-0818, Japan, 2 Kagawa Prefectural Agricultural Experiment Station, Ayagawa, Kagawa 761-2306, Japan, 3 Graduate School of Life Sciences, Tohoku University, Sendai, Miyagi 980-8577, Japan, 4 Biotechnology Program, Faculty of Science, Galala University, Suze, Galala 43511, Egypt, and 5 Botany Department, Faculty of Science, Aswan University, Aswan 81528, Egypt

*To whom correspondence should be addressed. Tel. +81 438 52 3935. Fax. +81 438 52 3934.

Email: shirasaw@kazusa.or.jp

† The first two authors contributed equally to the work.

Received 15 December 2021; Editorial decision 11 January 2022; Accepted 13 January 2022

Abstract

*Asparagus kiusianus* is a disease-resistant dioecious plant species and a wild relative of garden asparagus (*Asparagus officinalis*). To enhance *A. kiusianus* genomic resources, advance plant science, and facilitate asparagus breeding, we determined the genome sequences of the male and female lines of *A. kiusianus*. Genome sequence reads obtained with a linked-read technology were assembled into four haplotype-phased contig sequences (~1.6 Gb each) for the male and female lines. The contig sequences were aligned onto the chromosome sequences of garden asparagus to construct pseudomolecule sequences. Approximately 55,000 potential protein-encoding genes were predicted in each genome assembly, and ~70% of the genome sequence was annotated as repetitive. Comparative analysis of the genomes of the two species revealed structural and sequence variants between the two species as well as between the male and female lines of each species. Genes with high sequence similarity with the male-specific sex determinant gene in *A. officinalis*, *MSE1/AoMYB35/AspTDF1*, were presented in the genomes of the male line but absent from the female genome assemblies. Overall, the genome sequence assemblies, gene sequences, and structural and sequence variants determined in this study will reveal the genetic mechanisms underlying sexual differentiation in plants, and will accelerate disease-resistance breeding in garden asparagus.

Key words: dioecy, genome assembly, haplotype phasing sequence, linked reads, structural variation
1. Introduction

Asparagus kiussianus is a wild relative of garden asparagus (Asparagus officinalis). While garden asparagus is a cultivated species belonging to the Asparagaceae family and is consumed as a vegetable crop around the world, A. kiussianus is native to the coastal regions of Japan. Therefore, A. kiussianus might exhibit tolerances and/or resilience to abiotic and biotic stresses. Although A. kiussianus has been identified as a potential donor of stem-blight disease resistance in asparagus breeding programs, neither the genetic mode nor the genetic loci of resistance have been elucidated to date.

Asparagus officinalis is a dioecious species and is widely recognized as a model for sex determination in plants. Recent studies indicate that the male-specific MYB-like gene, MSEA1/AoMYB35/AspTDF1, located at the masculinization-promoting M locus of the Y-specific region in asparagus, functions in sex determination in asparagus. Since A. kiussianus is also a dioecious plant species, like garden asparagus, it is possible that both species share the same system of sex determination. Therefore, comparative genome sequence and structure analyses between the two species could provide insights into the molecular mechanisms underlying sex determination in Asparagaceae and the evolutionary processes involved therein.

Advances in sequencing technologies have enabled the whole-genome sequencing of various plant species, thus providing fundamental information required for understanding the plant biology and accelerating breeding programs. Nevertheless, while the genome sequence data of garden asparagus and transcriptome data of A. kiussianus have been made publicly available, no whole-genome sequence data have been released for A. kiussianus to date. Owing to the dioecious nature of A. kiussianus, which leads to allogamy, its genome is predicted to be highly heterozygous. Therefore, haplotype-phased genome sequence data would be useful for dissecting the allelic structure and structural variations in A. kiussianus. In this study, we employed a linked-read technology (10X Genomics, Pleasanton, CA, USA) to construct haplotype-based genome sequence assemblies of the male and female lines of A. kiussianus. The genome sequence assemblies were then used for gene prediction and sequence and structural variant discovery. Overall, the genome sequence information of A. kiussianus obtained in this study could accelerate studies on plant sex determination and facilitate asparagus breeding programs.

2. Materials and methods

2.1. Plant materials

Male (K1) and female (K2) lines of A. kiussianus cultivated at Kagawa Prefectural Agricultural Experiment Station (Kagawa, Japan) were used in this study. Genomic DNA was extracted from the stems of young seedlings using the modified cetyltrimethylammonium bromide method.

2.2. Genome sequencing and assembly

Genomic DNA libraries of male and female lines were prepared using the Chromium Genome Library Kit v2 (10X Genomics), and sequenced on NovaSeq 6000 (Illumina, San Diego, CA, USA) in paired-end, 150 bp mode. The sequence reads were assembled with Supernova (10X Genomics) to construct the contig sequences, scaffold the contigs, and resolve haplotype phases. DNA library preparation, sequencing, and assembly were conducted by Takara Bio (Shiga, Japan) as an outsourcing service.

The genome sizes of male and female lines were estimated based on short reads using Jellyfish. To construct pseudomolecule sequences at the chromosome level, the assembled contigs were aligned against the sequence of 10 A. officinalis chromosomes (reference) using RatGoo.

The software tools used for data analyses are listed in Supplementary Table S1.

2.3. Repetitive sequence analysis and gene prediction

Repetitive sequences in the assemblies were identified with RepeatMasker, using repeat sequences registered in Repbase and a de novo repeat library built with RepeatModeler.

RNA-Seq reads of A. kiussianus and A. officinalis were obtained from a public DNA database (GenBank Sequence Read Archive accession number: SRA1003110). The RNA-Seq reads, from which adapter sequences were trimmed with fastx_clipper in the FASTX-Toolkit, were aligned against the assembled sequences with HISAT2. Gene prediction was performed with BREAker2 using the positional information of the repeats, RNAs, and peptide sequences of the predicted genes of A. officinalis (V1.1) released in Phytozome.

2.4. Comparative genome structure analysis of Asparagus kiussianus and Asparagus officinalis

Chromosome-level genome sequence assemblies of A. kiussianus (this study) and A. officinalis (V1.1) were compared with Minimap2, and the resultant Pairwise mApping Format files were visualized with pafr.

3. Results and data description

3.1. Haplotype-phased genome assembly

Short-read sequences of the male (143.7 Gb) and female (140.0 Gb) lines of A. kiussianus were obtained in this study, and their genome sizes were estimated at 1,563.8 Mb and 1,729.4 Mb, respectively (Fig. 1).

The short-read sequences of the male line were assembled into raw contigs (total length = 3,724.6 Mb, N50 = 7.5 kb), which included gaps and all homologous sequences of the diploid genome (Supplementary Table S2). Then, the homologous sequences were flattened, and the gaps were filled by joining the sequence to its flanking sequence, thus producing megabubble sequences (total length = 1,811.6 Mb, N50 = 170.6 kb) (Supplementary Table S2). Finally, two haplotype-phased genome assemblies (each containing 111,443 sequences) were generated from the megabubble sequences (Table 1). Haplotype 1 spanned 1,618.9 Mb in total with an N50 value of 155.5 kb, while haplotype 2 spanned 1,618.5 Mb in total with an N50 length of 155.3 kb. Complete Benchmarking Single-Copy Orthologs (BUSCO) scores were 88.4% and 88.6% for haplotypes 1 and 2, respectively (Table 1). The male genome assemblies for haplotype 1 and 2 were designated as AKIK1p1 and AKIK1p2, respectively.

Similarly, short-read sequences of the female line were assembled into raw contigs (total length = 3,780.7 Mb, N50 = 7.0 kb) and megabubble sequences (total length = 1,703.6 Mb, N50 = 76.2 kb) (Supplementary Table S2). The resultant haplotype-phased assemblies (each containing 125,241 sequences) spanned 1,567.6 Mb in total with an N50 value of 59.7 kb for ‘haplotype 1’, and 125,241 sequences of 1,567.3 Mb in total with an N50 value of 59.6 kb for ‘haplotype 2’ (Table 1). Complete BUSCO scores were 88.7% and 88.6% for haplotypes 1 and 2, respectively (Table 1). The female genome assemblies for haplotype 1 and 2 were designated as AKIK2p1 and AKIK2p2, respectively.
The four sets of genome sequence assemblies of *A. kiusianus* (haplotypes 1 and 2 of male and female lines) were aligned against the chromosome-scale genome assembly of *A. officinalis*. A total of 96,224 sequences (1,535.9 Mb) for haplotype 1 and 96,224 sequences for haplotype 2 (1,535.3 Mb) in the male line, and 107,875 sequences (1,491.7 Mb) for haplotype 1 and 107,864 sequences (1,489.7 Mb) for haplotype 2 in the female line, could be aligned to the 10 chromosome sequences of *A. officinalis* (Table 2). Complete BUSCO scores ranged from 91.2% (AKIK1p1) to 91.8% (AKIK2p2). The nomenclature of the pseudomolecule sequences was based on the chromosome names of *A. officinalis* (ch01–ch10), where the chromosome 1 is the sex chromosome. Sequences that were unassigned to the *A. officinalis* genome were connected and termed chromosome 0 (ch00).

### 3.2. Repetitive sequence analysis and gene prediction

Repeat sequences occupied 66.8% (AKIK1p1), 66.8% (AKIK1p2), 68.1% (AKIK2p1), and 68.1% (AKIK2p2). The most abundant repetitive sequences were long terminal repeats (LTRs) (45.4–46.3%), followed by unclassified repeats (14.9–13.7%) and DNA transposons (4.7–4.8%) (Table 3). A total of 404.2 million RNA reads for 18 samples were mapped to the genome sequences. The mapping rates of *A. kiusianus* RNA-Seq reads were 93.7–94.1%, while those of *A. officinalis* reads were 82.4–82.6%. Based on the positions of RNA-Seq reads on the genome sequences, a total of 59,208, 56,706, 57,523, and 58,694 potential protein-coding genes were predicted in AKIK1p1, AKIK1p2, AKIK2p1, and AKIK2p2, respectively (Table 2), of which 365, 380, 472, and 505 genes contained premature termination codons in their internal sequences. Complete BUSCO scores ranged from 90.1% (AKIK2p1) to 91.4% (AKIK1p2).

Next, we compared the sequences of predicted genes with *MSE1/AoMYB35/AspTDF1*, which has been reported as the male-specific sex determinant gene in *A. officinalis*. Two genes, K1p1ch01g28074 and K1p2ch01g47751, identified in the genomes of the male line exhibited high sequence similarity with the query; however, none of the genes in the female genome assemblies showed significant sequence similarity with the query.

### 3.3. Genome sequence and structural variations between *Asparagus kiusianus* and *Asparagus officinalis*

Haplotype sequences within the male and female lines were compared. A total of 386,723 single nucleotide polymorphisms (SNPs)
The transition (Ts)/transversion (Tv) ratio was 3.0, and 46,007 insertions/deletions (indels) were identified between the two haplotypes of the male line (Table 4). On the other hand, 293,196 SNPs (Ts/Tv = 3.0) and 35,732 indels were identified between the two haplotype sequences of the female line (Table 4). The haplotype sequences of male and female lines were also compared, and 321,334 SNPs and 49,921 indels on average were identified across the four haplotype combinations (Table 4).

Table 2. Statistics of the pseudomolecule sequences of A. kiusianus

|        | AKIK1p1 |                    | AKIK1p2 |                    |
|--------|---------|--------------------|---------|--------------------|
|        | Total length (bp) | Gap (%) | Number of contigs | Number of genes | Total length (bp) | Gap (%) | Number of contigs | Number of genes |
| ch01  | 179,686,458 | 11.8 | 10,537 | 6,938 | 179,663,709 | 11.8 | 10,543 | 6,706 |
| ch02  | 94,379,980 | 12.5 | 6,058 | 3,631 | 94,286,185 | 12.5 | 6,066 | 3,427 |
| ch03  | 160,690,220 | 10.4 | 10,792 | 6,135 | 161,086,385 | 10.5 | 10,791 | 6,030 |
| ch04  | 222,888,549 | 11.1 | 13,146 | 7,547 | 222,850,935 | 11.1 | 13,139 | 7,251 |
| ch05  | 175,875,079 | 12.0 | 10,573 | 6,878 | 176,035,479 | 12.0 | 10,582 | 6,668 |
| ch06  | 112,254,750 | 9.4 | 7,553 | 3,631 | 112,182,592 | 9.4 | 7,553 | 3,504 |
| ch07  | 231,128,771 | 11.0 | 13,973 | 8,285 | 230,788,849 | 11.0 | 13,928 | 7,953 |
| ch08  | 209,954,455 | 10.5 | 13,763 | 7,167 | 209,720,314 | 10.5 | 13,796 | 6,882 |
| ch09  | 73,375,337 | 10.6 | 4,435 | 2,759 | 73,241,043 | 10.7 | 4,437 | 2,638 |
| ch10  | 85,316,880 | 9.9 | 11,974 | 6,358 | 85,479,114 | 10.0 | 11,943 | 6,030 |
| Subtotal | 1,545,550,479 | 11.0 | 96,224 | 56,389 | 1,544,924,455 | 11.0 | 96,224 | 54,113 |
| ch00  | 84,496,458 | 10.4 | 15,223 | 2,819 | 84,675,281 | 10.6 | 15,223 | 2,593 |
| Total | 1,630,046,937 | 11.0 | 111,447 | 59,208 | 1,629,699,736 | 11.0 | 111,447 | 56,706 |

Table 3. Repetitive sequences in the A. kiusianus genomes

| Repeat type | AKIK1p1 |                    | AKIK1p2 |                    | AKIK2p1 |                    | AKIK2p2 |                    |
|-------------|---------|--------------------|---------|--------------------|---------|--------------------|---------|--------------------|
| Number of elements | Length occupied (bp) | % | Number of elements | Length occupied (bp) | % | Number of elements | Length occupied (bp) | % | Number of elements | Length occupied (bp) | % |
| SINEs        | 296    | 100,129 | 0.0 | 295    | 100,876 | 0.0 | 260    | 88,942 | 0.0 | 269    | 88,784 | 0.0 |
| LINEs        | 36,139 | 22,836,398 | 1.4 | 36,171 | 22,868,973 | 1.4 | 36,022 | 22,848,619 | 1.5 | 36,237 | 22,865,141 | 1.5 |
| LTR elements | 534,586 | 740,043,346 | 45.4 | 534,419 | 739,629,919 | 45.4 | 538,917 | 731,783,197 | 46.3 | 538,772 | 731,702,512 | 46.3 |
| DNA transposons | 119,624 | 77,129,754 | 4.7 | 119,369 | 77,003,389 | 4.7 | 117,911 | 76,193,533 | 4.8 | 118,264 | 76,216,494 | 4.8 |
| Unclassified | 426,795 | 242,401,476 | 14.9 | 426,435 | 242,359,956 | 14.9 | 422,640 | 239,811,052 | 15.2 | 422,494 | 239,544,071 | 15.2 |
| Small RNA    | 447    | 231,343 | 0.0 | 451    | 233,377 | 0.0 | 435    | 232,256 | 0.0 | 442    | 228,647 | 0.0 |
| Satellites   | 1,945  | 843,970 | 0.1 | 1,894  | 847,909 | 0.1 | 2,155  | 901,555 | 0.1 | 2,171  | 901,844 | 0.1 |
| Simple repeats | 197,902 | 17,199,114 | 1.1 | 198,009 | 17,248,715 | 1.1 | 195,639 | 16,992,727 | 1.1 | 195,634 | 17,024,257 | 1.1 |
| Low complexity | 32,476 | 1,822,571 | 0.1 | 32,548 | 1,822,378 | 0.1 | 32,388 | 1,816,175 | 0.1 | 32,482 | 1,817,483 | 0.1 |

Bold indicates subtotal (ch01 to ch10) and total values (ch01 to ch10 and ch00).
Table 4. Completeness of gene predictions, based on BUSCO assessment

|                  | Within a line | Between male and female lines |
|------------------|---------------|-------------------------------|
|                  | AKIK1p1 vs AKIK1p2 | AKIK2p1 vs AKIK2p2 | AKIK1p1 vs AKIK2p1 | AKIK1p1 vs AKIK2p2 | AKIK1p2 vs AKIK2p1 | AKIK1p2 vs AKIK2p2 |
| Base substitutions | 386,723       | 293,196                       | 324,361           | 327,460           | 314,704           | 318,810           |
| Ts/Tv ratio      | 3.0           | 3.0                           | 2.9               | 2.9               | 2.9               | 2.9               |
| 1-bp indels      | 23,833        | 17,283                        | 20,344            | 20,379            | 19,692            | 19,970            |
| 2-bp indels      | 5,799         | 4,477                         | 5,722             | 5,763             | 5,645             | 5,691             |
| 3- to 49-bp indels | 11,100        | 8,641                         | 13,931            | 13,897            | 13,647            | 13,777            |
| 50- to 1000-bp indels | 1,871       | 1,472                         | 7,141             | 7,127             | 7,096             | 7,096             |
| ⩾1000-bp indels  | 3,404         | 3,859                         | 3,199             | 3,229             | 3,160             | 3,179             |

Figure 2. Comparative analysis of the genome sequence and structure of A. kiusianus and A. officinalis. Dots represent similarities between the genome sequence and structure of the two species. Genomes of the male and female lines of A. kiusianus are indicated as AKIK1 and AKIK2, respectively. Haplotype phases 1 and 2 are indicated as p1 and p2, respectively. AsparagusV1 indicates A. officinalis genome.
While the chromosome structures were conserved within *A. kiusianus* lines and between *A. kiusianus* and *A. officinalis* (Fig. 2), genomic rearrangements were observed at the local level. For instance, at the sex-related region including the male-specific gene *MSE1/AoMYB35/AspTDF1* of *A. officinalis*, sequence collinearity was disrupted by inversions and translocations between the male and female lines (Fig. 3). Although sequence similarity was low between the male haplotypes of *A. kiusianus* and *A. officinalis*, sequence collinearity was moderately conserved (Fig. 3).

### 4. Conclusion and future perspectives

We present the chromosome-level haplotype-phased genome assemblies of the male and female lines of *A. kiusianus*, a wild relative of garden asparagus. The genome size of *A. kiusianus* was estimated to be ∼1.6 Gb (Fig. 1), which was 300 Mb larger than that of garden asparagus (ca. 1.3 Gb). This estimation was reflected in the difference between the assembly sizes of *A. kiusianus* (1.6 Gb) (Table 1) and garden asparagus (1.2 Gb). Of the 1.6 Gb assembly, 1.5 Gb could be aligned to the pseudomolecule sequence of asparagus, without any structural rearrangements (Fig. 2 and Table 2). Since we determined haplotype-phased genome sequences for the male and female lines of *A. kiusianus*, it was possible to compare the sequence and structure of the M locus between the Y-specific region of the male line and the corresponding region of the female line (Fig. 3). The result suggested dynamic genome rearrangements between the male and female lines, similar to that reported in jojoba, which might lead to presence/absence variation of the male-specific gene *MSE1/AoMYB35/AspTDF1* between the male and female lines.

The genome of *A. kiusianus* harbours valuable genes that could be used for the breeding of elite garden asparagus cultivars. Because of cross-compatibility between the two species, important genetic loci, such as those imparting resistance to stem blight, which causes considerable production losses, could be transferred from *A. kiusianus* into garden asparagus. However, while DNA markers linked to the genes would facilitate the selection of disease-resistant lines in breeding programs, the genetic loci responsible for disease resistance have not been reported so far. The chromosome-level genome sequence of *A. kiusianus* presented in this study could serve as a reference for genetic mapping and the identification of resistance genes, as well as for transcriptome analysis and the determination of gene functions and mechanisms underlying the resistance and susceptible phenotypes.

Although the plant genomics era started with the whole-genome sequencing of *Arabidopsis thaliana*, an undomesticated species, the advanced approaches of plant genomics have been applied more frequently to agronomically important crops rather than to wild plant species. Wild plants have the potential to accelerate the pace of breeding programs and to further the field of plant science. The genome sequence information of *A. kiusianus* generated in this study will help to reveal the genetic mechanisms underlying sexual differentiation in plants and will accelerate disease-resistance breeding in asparagus.

### Data availability

The genome sequence information generated in this study is available at Plant GARDEN (https://plantgarden.jp).

### Acknowledgements

We thank T. Ikeshi and T. Nakamura (Kagawa Prefectural Agricultural Experiment Station) for providing the plant material, and A. Watanabe (Kazusa DNA Research Institute) for technical assistance.

### Accession numbers

Sequence reads generated in this study are available from the Sequence Read Archive (DRA) of the DNA Data Bank of Japan (DDBJ) under accession number: DRA012987. The DDBJ accession numbers of the assembled genome sequences are BQKN01000001-BQKN01111443 (AKIK1p1), BQKO01000001-BQKO01111443 (AKIK2p1), BQKP01000001-BQKP01125241 (AKIK2p2), and BQKQ01000001-BQKQ01125241 (AKIK2p2).

### Funding

This work was partially supported by the Kazusa DNA Research Institute Foundation.

### Conflict of interest

None declared.

### Supplementary data

Supplementary data are available at DNARES online.

### References

1. Ohwi, J. 1965. *Asparagus L.*, In: Meyer, F.G., Walker, E.H. (eds) *Flora of Japan*, Smithsonian Institution, Washington, D.C., pp. 300.
2. Iwato, M., Kosaza, M., Takeuchi, Y., et al. 2014. Stem blight resistance of *Asparagus kiusianus* and its hybrid with *A. officinalis*, *Adv. Hort. Sci.*, 28, 202–7.
3. Harless, A., Zhou, J., Xu, C., et al. 2017. The asparagus genome sheds light on the origin and evolution of a young Y chromosome, *Nat. Commun.*, 8, 1279.
4. Murase, K., Shigenobu, S., Fujii, S., et al. 2017, MYB transcription factor gene involved in sex determination in *Asparagus officinalis*, *Genes Cells*, 22, 115–23.

5. Tsugama, D., Matsuyama, K., Ide, M., Hayashi, M., Fujino, K. and Masuda, K. 2017, A putative MYB35 ortholog is a candidate for the sex-determining genes in *Asparagus officinalis*, *Sci. Rep.*, 7, 41497.

6. Abdelrahman, M., Nakabayashi, R., Mori, T., et al. 2020, Comparative metabolome and transcriptome analyses of susceptible *Asparagus officinalis* and resistant wild *A. kiusianus* reveal insights into stem blight disease resistance, *Plant Cell Physiol.*, 61, 1464–76.

7. Abdelrahman, M., Suzumura, N., Mitoma, M., et al. 2017, Comparative de novo transcriptome profiles in *Asparagus officinalis* and *A. kiusianus* during the early stage of *Phomopsis asparagi* infection, *Sci. Rep.*, 7, 2608.

8. Porebski, S., Bailey, L.G. and Baum, B.R. 1997, Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components, *Plant Mol. Biol. Rep.*, 15, 8–15.

9. Al-Dossary, O., Alsulmaie, R., Kharabian-Masouleh, A., Al-Mssallem, I., Furtado, A. and Henry, R.J. 2021, The jojoba genome reveals wide divergence of the sex chromosomes in a dioecious plant, *Plant J.*, 108, 1283–94.

10. Ito, T., Konno, I., Kubota, S., et al. 2011, Production and characterization of interspecific hybrids between *Asparagus kiusianus* Makino and *A. officinalis* L, *Euphytica*, 182, 285–294.

11. The Arabidopsis Genome Initiative. 2000, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*, *Nature*, 408, 796–815.

12. 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.

13. Bobra, A., Kilhan, B., Sivasankar, S., et al. 2021, Reap the crop wild relatives for breeding future crops. *Trends Biotechnol*, in press.

14. Migicovsky, Z. and Myles, S. 2017, Exploiting wild relatives for genomics-assisted breeding of perennial crops, *Front. Plant Sci.*, 8, 460.