

## RESEARCH ARTICLE

# QTL analysis of femaleness in monoecious spinach and fine mapping of a major QTL using an updated version of chromosome-scale pseudomolecules

Kaoru Yamano<sup>1</sup>✉, Akane Haseda<sup>1</sup>✉, Keisuke Iwabuchi<sup>1</sup>✉, Takayuki Osabe<sup>2</sup>✉, Yuki Sudo<sup>1</sup>✉, Babil Pachakkil<sup>3</sup>‡, Keisuke Tanaka<sup>4,5</sup>‡, Yutaka Suzuki<sup>6</sup>‡, Atsushi Toyoda<sup>7</sup>‡, Hideki Hirakawa<sup>8</sup>\*, Yasuyuki Onodera<sup>9</sup>\*✉

**1** Graduate School of Agriculture, Hokkaido University, Sapporo, Japan, **2** School of Agriculture, Hokkaido University, Sapporo, Japan, **3** Department of International Agricultural Development, Faculty of International Agriculture and Food Studies, Tokyo University of Agriculture, Setagaya-ku, Tokyo, Japan, **4** NODAI Genome Research Center, Tokyo University of Agriculture, Setagaya-ku, Tokyo, Japan, **5** Department of Informatics, Tokyo University of Information Sciences, Chiba, Japan, **6** Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan, **7** Department of Genomics and Evolutionary Biology, National Institute of Genetics, Mishima, Japan, **8** The Department of Technology Development, Kazusa DNA Research Institute, Kisarazu, Japan, **9** The Research Faculty of Agriculture, Hokkaido University, Sapporo, Japan

✉ These authors contributed equally to this work.

‡ BP, KT, YS and AT also contributed equally to this work.

\* onodera@agr.hokudai.ac.jp (YO); hh@kazusa.or.jp (HH)

## Abstract

Although spinach is predominantly dioecious, monoecious plants with varying proportions of female and male flowers are also present. Recently, monoecious inbred lines with highly female and male conditions have been preferentially used as parents for F<sub>1</sub>-hybrids, rather than dioecious lines. Accordingly, identifying the loci for monoecism is an important issue for spinach breeding. We here used long-read sequencing and Hi-C technology to construct SOL\_r2.0\_pseudomolecule, a set of six pseudomolecules of spinach chromosomes (total length: 879.2 Mb; BUSCO complete 97.0%) that are longer and more genetically complete than our previous version of pseudomolecules (688.0 Mb; 81.5%). Three QTLs, *qFem2.1*, *qFem3.1*, and *qFem6.1*, responsible for monoecism were mapped to SOL\_r2.0\_pseudomolecule. *qFem3.1* had the highest LOD score and corresponded to the *M* locus, which was previously identified as a determinant of monoecious expression, by genetic analysis of progeny from female and monoecious plants. The other QTLs were shown to modulate the ratio of female to male flowers in monoecious plants harboring a dominant allele of the *M* gene. Our findings will enable breeders to efficiently produce highly female- and male-monoecious parental lines for F<sub>1</sub>-hybrids by pyramiding the three QTLs. Through fine-mapping, we narrowed the candidate region for the *M* locus to a 19.5 kb interval containing three protein-coding genes and one long non-coding RNA gene. Among them, only *RADIALIS-like-2a* showed a higher expression in the reproductive organs, suggesting that it might play a role in reproductive organogenesis. However, there is no evidence that it is involved in the regulation of stamen and pistil initiation, which are directly related to the floral sex



## OPEN ACCESS

**Citation:** Yamano K, Haseda A, Iwabuchi K, Osabe T, Sudo Y, Pachakkil B, et al. (2024) QTL analysis of femaleness in monoecious spinach and fine mapping of a major QTL using an updated version of chromosome-scale pseudomolecules. PLoS ONE 19(2): e0296675. <https://doi.org/10.1371/journal.pone.0296675>

**Editor:** Mojtaba Kordrostami, Nuclear Science and Technology Research Institute, ISLAMIC REPUBLIC OF IRAN

**Received:** August 10, 2023

**Accepted:** December 15, 2023

**Published:** February 23, 2024

**Peer Review History:** PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0296675>

**Copyright:** © 2024 Yamano et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All NGS data are available from DDBJ Sequence Read Archive

(accession numbers DRX241308, DRX477737, DRX241317–DRX241420, DRX094096–DRX094113, DRX477738–DRX477746). All genome and gene sequences and annotations are available at figshare (<https://doi.org/10.6084/m9.figshare.23916627.v1>).

**Funding:** This work was supported by the Japan Society for the Promotion of Science (Grants-in-Aid for Scientific Research [KAKENHI] Grant Number 21K05525 and 22H02321), Ministry of Education, Culture, Sports, Science and Technology (KAKENHI Grant Numbers 16H06279). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

differentiation system in spinach. Given that auxin is involved in reproductive organ formation in many plant species, genes related to auxin transport/response, in addition to floral organ formation, were identified as candidates for regulators of floral sex-differentiation from *qFem2.1* and *qFem6.1*.

## Introduction

Originating from Persia, spinach has since been introduced widely across East Asia, Europe, and the Americas. Spinach is now grown as a nutritious vegetable rich in vitamins, dietary fiber, minerals, and antioxidants such as carotenoids and flavonoids, and is produced in more than 50 countries around the world, with a total annual production of 32,300 kt [1, 2]. Spinach is commonly known as a dioecious species consisting of male and female individuals. However, certain varieties, lines, germplasm accessions, and crosses have been shown to produce monoecious plants, which bear both female and male flowers in various proportions in the same individual [3]. Most of the modern spinach varieties for commercial production are F<sub>1</sub>-hybrids. In the early stages of spinach hybrid breeding, the promising combinations of dioecious lines were planted in alternating rows with male plants removed before anther dehiscence from the seed-parental rows. Later, to save labor by eliminating the requirement of removing the males, monoecious lines breeding true for a highly female condition began to be used as seed parents for hybrid seed production [4, 5]. Currently, highly male monoecious lines are also frequently used as pollen parents in spinach breeding. Using monoecious lines reduces the labor required to remove male plants from the seed-parental rows. Moreover, highly male monoecious plants have a longer life-span after flowering than male plants, making it easier to synchronize pollination timing with the maternal (seed-parental) line (personal communication from Tohoku Seed Co.). Therefore, understanding the mechanisms of sex determination in dioecious lines and controlling the degree of femaleness in monoecious spinach are critical issues in hybrid breeding programs.

Most flowering plants have both male and female reproductive organs—i.e., they are hermaphroditic or monoecious—but some have separate sexes. Dioecious plants, which have separate sexes, are taxonomically widely dispersed although they are found in only a small percentage of flowering plants [6]. This suggests that having both sexes in one plant (hermaphroditism or monoecism) is the ancestral condition and that dioecy has arisen recently and independently in many evolutionary lineages. The sex of dioecious plants is often determined by genetics, and sex chromosomes have been found in a wide range of angiosperm lineages. Therefore, given the evolutionary history of dioecy, it is plausible that plant sex chromosomes also have evolved recently from autosomes in their hermaphroditic or monoecious ancestors [7, 8]. Some dioecious plants, such as white campion, sorrel, and hemp, have heteromorphic sex chromosome pairs, while others, such as spinach, asparagus, and papaya, have homomorphic sex chromosome pairs. Sex chromosome pairs are derived from autosomal pairs, strongly suggesting the evolutionary transition from homomorphic to heteromorphic sex chromosome pairs. Thus, homomorphic and heteromorphic sex chromosomes that have emerged in various angiosperm lineages may represent different evolutionary stages. In dioecious spinach lines, the sexual dimorphism is controlled by an allelic pair designated X and Y, which are located on the longest chromosome pair [9]. Specifically, the chromosome with the dominant male determinant is the Y chromosome, while its counterpart is the X chromosome [10, 11]. The spinach sex chromosomes are cytogenetically homomorphic with some exceptions and recombine across most of their length [12]. However, recombination of the Y with the X chromosome is

absent in the region surrounding the male-determining locus [13]. A recent study predicted that the non-recombining region of the Y chromosome is 17.4 Mb in length, accounting for ~8% of the total length [14].

In our previous genetic analysis using progeny families from the cross of dioecious line 03–009 with the highly male monoecious line 03–336, a single, incompletely dominant gene designated *M* was shown to be responsible for the monoecious character. The monoecious gene (*M*) is linked to the male-determining locus with a distance of ~13 cM, located at a 7.5 cM chromosomal region bracketed by variant sites at *clc-e* and *clsy3* loci. Furthermore, the genetic analysis of progeny from the cross between female and monoecious plants suggested that the male-determining gene (*Y*) is epistatic to the monoecious gene (*M*), which is epistatic to *X*: *XXmm* is female, *XYMM*, *XYMm* and *XYmm* are male, and *XXMM* and *XXMm* produce a monoecious phenotype [15–17]. Based on the phenotypic expression of plants possessing these genotypes, it can be inferred that *Y* and *M* act as male-promoting and female-suppressing factors, with the effect of *M* being weaker than that of *Y*.

Apart from the major factor that determines the development of monoecism, there are modifying genes that influence the character of the monoecism, and these are inherited quantitatively. It has been observed that selection can produce monoecious lines that have different levels of femaleness, as measured by the proportion of female to male flowers per plant. Through three generations of inbreeding, Janick and Stevenson succeeded in selecting plants with progeny mean values that ranged from 70–80% female to as low as 5% female [18]. However, the additional loci for monoecism have not yet been identified.

Having high-quality genome assemblies, such as chromosome-scale pseudomolecules, is crucial for identifying genes and loci of interest. Such pseudomolecules are particularly important as a reference because GWAS and QTL analyses require highly accurate variant data that covers nearly the whole genome. The pseudomolecules can be constructed by aligning highly accurate scaffolds in the appropriate direction and have high genome coverage rates. Furthermore, chromosome-scale pseudomolecules lay the foundation for genome-based molecular breeding strategies, such as marker-assisted and genomic selection [19].

The spinach genome consists of  $n = x = 6$  chromosomes and is estimated to be 989 Mb, based on the C-value. Recently, a genome assembly (SOL\_r1.1) for the dioecious spinach line 03–009 was generated by single-molecule sequencing technologies (e.g., PacBio, ONT, and 10x Genomics) in combination with mate-pair reads [20]. The spinach genome assembly represented >90% of the whole genome and showed high continuity ( $N_{50} = 11.3$  Mb), but it contained chimeric scaffolds caused by assembling errors. Finally, only ~70% of the genome assembly was predicted to be correctly assembled, and this portion was anchored to a set of six pseudomolecules that were collectively designated SOL\_r1.0\_pseudomolecule [20]. Using pseudomolecules with a genome coverage rate of about 70% as a reference for GWAS and QTL analyses, there is a risk that the genes of interest may not be detected and identified. Therefore, in this study, we enhanced the length and accuracy of the pseudomolecules using Hi-C, and mapped three QTLs for controlling femaleness in monoecious spinach to the updated pseudomolecules. In addition, one of the QTLs, corresponding to the *M* locus, was fine-mapped to a 19.5 kb region, and transcriptome analysis was carried out to elucidate the underlying mechanisms controlling femaleness in monoecious spinach.

## Materials and methods

### Plant materials

In this study, to identify the loci controlling femaleness in monoecious plants, we conducted QTL analyses using female plants from the dioecious line 03–009 and plants from the highly

male monoecious line 03–336 with ~5% femaleness scores as parental plants for producing the mapping population.

Line 03–009 produced strict female (100% femaleness [proportion of female flowers per plant]) and male (0% femaleness) plants, while line 03–336 showed the lowest femaleness scores among monoecious lines stored in our lab [3]. The spinach F<sub>2</sub> population (03–009 × 03–336 F<sub>2</sub>; [20]) resulted from crossing a female plant (*mm* plant) of line 03–009 with a plant (*MM* plant) from the highly male monoecious line 03–336 and was used for linkage map construction and QTL analysis. S<sub>1</sub>BC<sub>2</sub>F<sub>1</sub>–S<sub>5</sub>BC<sub>2</sub>F<sub>1</sub>, F<sub>3</sub>–F<sub>4</sub>, and BC<sub>1</sub>F<sub>4</sub> generations from the cross between 03–009 and 03–336 were used for fine mapping of the *M* locus (S1 and S2 Figs) [3]. Furthermore, the plants homozygous for *M* (*MM* plants) selected from the F<sub>2</sub> and F<sub>3</sub> generations of the 03–009 × 03–336 cross were used for the verification of minor QTLs for the femaleness phenotype (the percentage of female flowers per plant) in monoecious plants. The genomic region carrying the monoecious (*M*) gene was introgressed by five successive backcrosses from monoecious donor line 03–336 into the 03–009 genetic background. This produced the near-isogenic line NIL-*M* (S3 Fig). By using lines NIL-*M* and 03–336 we anticipated that we could clarify the influence of the different genetic backgrounds on monoecious expression by the *M* gene.

For preliminary RNA-seq, RT-qPCR, and QTL analyses, plants were cultivated in a growth chamber (LH-350S; Nippon Medical & Chemical Instruments, Osaka, Japan) at 20°C with an 8 h photoperiod for the first four weeks, at 23°C with an 8 h photoperiod for the next three weeks, and subsequently at 25°C with a 24 h photoperiod. For QTL verification and transcriptome analysis, plants were maintained at a consistent 20°C, with an 8 h photoperiod for the first four weeks, a 16 h photoperiod for the next two weeks, and finally a 24 h photoperiod. For this study, we evaluated the femaleness of plants that bore over 20 flower clusters, with each cluster comprising 5–10 flowers. For expression analysis, vegetative leaves, shoot apices (plants under short-day conditions, and plants with 3 or 7 days of long-day exposure after being grown in short-day conditions), and early-stage (~5 mm length) and mid-stage (30–40 mm length) inflorescences were harvested (S4 Fig). Total RNA was extracted using an RNeasy Plant Mini Kit combined with an RNase-Free DNase Set (QIAGEN, Venlo, Netherlands). Total cellular genomic DNA was prepared using the method of Sassa [21].

### Polishing and scaffolding of the spinach *de novo* genome assembly

A phased long-read-based genome assembly, Spol\_r0.0 ([20]; S1 Table), of a male plant of dioecious line 03–009 was polished by Racon [22] with long-reads (DRX241314; [20]) and by Pilon [23] with Illumina PE reads (DRX241308). Hi-C analysis was used because it produces pairs of short reads formed by crosslinking chromatin interactions. In addition, Hi-C analysis provides genomic proximity information with a scale ranging from 1 kb to 1 Mb or longer, which allows for the accurate assignment, ordering, and orientation of genomic sequences [24]. To assemble Spol\_r0.0 into a set of more continuous genome sequences by scaffolding with Hi-C data, the Hi-C library was prepared from young leaves harvested from a single male plant of line 03–009 by using an Arima Hi-C Kit (Arima Genomics, San Diego, CA) and sequenced by a HiSeq X Instrument (Illumina, San Diego, CA) configured to generate 151-bp PE reads (DRX477737; S2 Table). Scaffolding of contigs was carried out using the Hi-C mate pairs with SALSA [25]. Illumina short reads obtained from a single male and a single female plant (DRX241308 and DRX241308; [20]) were mapped to Spol\_r0.0, and calculated the male-to-female read depth ratio (*M/F* ratio) on assembly contigs and used to identify sex-linked (*Y*-linked) regions: the *Y*-linked region tends to have a higher male (*XY*) read depth than female (*XX*) read depth, as it does not recombine with the *X* chromosome and tends to be divergent from the *X*-linked

region. In this study, contigs with  $\log_{10}(M/F \text{ ratio}) > 0.3$  were considered Y-linked. The M/F ratio scores were used to specify chimeric scaffolds composed of contigs representing X- and Y-linked regions and to filter out Y-linked regions from the chimeric scaffolds.

### Construction of linkage map and QTL analysis

The double-digest restriction site-associated DNA sequencing (ddRAD-Seq) reads (DRX241317–DRX241420) obtained from 101 F<sub>2</sub> plants and an F<sub>1</sub> plant from a 03–009 × 03–336 cross and their parents (03–009 and 03–336) were mapped to the spinach draft genome (Spol\_r1.0; [S1 Table](#)). SNPs were called as described in Hirakawa et al. [20]. The raw SNPs were filtered by using VCFtools [26] and SnpEff [27]. Construction of the linkage map was carried out by using the Lep-Map3 software package [28]. QTL analysis was performed using the CIM function of the R/qtl package [29] with the Haley–Knott regression method [30]. The LOD significance threshold for detecting QTLs was calculated by performing 1,000 iterations using the R/qtl permutation test. ANOVA was used to calculate the phenotypic variance explained (PVE) for each QTL.

### Construction of a linkage map and pseudomolecules

Scaffolds of the spinach draft genome were anchored to the SNP linkage groups and assembled into draft pseudomolecules. The Hi-C mate pairs obtained from the 03–009 were mapped to the draft pseudomolecules, and Hi-C contact maps were built by using the software Juicer 1.6 [31]. In accordance with the Hi-C contact maps, the draft pseudomolecules were manually corrected to obtain the final version of the spinach pseudomolecules. The pseudomolecules were annotated with 29,276 genes (SOL\_r1.1a; [20]) using the software GeMoMa 1.9 [32]. The quality of the assembled genome and predicted genes was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCOs). The BUSCO completeness scores of the 1416 single-copy orthologous genes defined in *embryophyta\_odb10* dataset were calculated by BUSCO v5.3.1 [33].

### Detection and annotation of nucleotide variants

Whole-genome sequencing (WGS) libraries for the spinach lines 03–009 and 03–336 were prepared as described in Hirakawa et al. [20] and sequenced using a HiSeq 2500 Instrument (Illumina) configured to generate 251-bp pair-end reads. The WGS reads (DRX401201, DRX401202; [S3 Table](#)) were mapped to the spinach pseudomolecules obtained herein using the software Bowtie2 Version 2.5.1 [34], and nucleotide variants were called. The impacts of nucleotide variants on gene function were evaluated using SnpEff 5.0e software [27] and classified as ‘HIGH’, ‘MODERATE’, ‘LOW’ or ‘MODIFIER’. A variant categorized as HIGH is expected to have a significant disruptive impact on the protein. This may lead to the truncation of the protein, loss of function, or trigger nonsense-mediated decay. A variant categorized as MODERATE is not expected to cause significant disruption but may alter protein effectiveness. A variant categorized as LOW is considered to be mostly harmless or unlikely to change the behavior of the protein. Finally, the fourth category, MODIFIER, is reserved for non-coding variants or variants affecting non-coding genes, where predictions are difficult or there is no evidence of impact.

### DNA marker analysis

Genetic markers for identifying allelic variants are listed in [S4](#) and [S5](#) Tables. Except for Sequence Characterized Amplified Region (SCAR) markers, all markers used in this study were developed by cleaved amplified polymorphic sequence (CAPS) and derived cleaved

amplified polymorphic sequence (dCAPS) marker techniques [35]. Amplification and detection of the markers were performed as described previously [17].

### Reverse transcription quantitative-PCR

The cDNA used as a template for reverse transcription quantitative-PCR (RT-qPCR) was synthesized according to the manufacturer's instructions using SuperScript<sup>®</sup> IV Reverse Transcriptase (Invitrogen). qPCR was performed with at least three biological replicates, each with two technical replicates, using PowerUp SYBR Green Master Mix (ThermoFisher). Quantification of the expression levels of the genes of interest was carried out using the comparative critical threshold ( $\Delta\Delta C_t$ ) method normalized with *Actin-7* mRNA. The primers used for the RT-qPCR are shown in [S6 Table](#).

### *In situ* hybridization

*In situ* hybridization techniques were used to determine the spatial and temporal expression pattern of *SoRL2a* during flower development. A 637-bp *SoRL2a* cDNA fragment was cloned into pBluescript II SK vector, and *in vitro* transcribed to generate sense and antisense riboprobes. The riboprobes were labeled with digoxigenin using a DIG RNA Labeling Kit (Roche Diagnostics GmbH, Mannheim, Germany). Plant tissues were infiltrated with 10% formalin neutral buffer solution (FUJIFILM Wako Pure Chemical Corporation, Osaka, Japan) at 4°C for 2 days. They were then dehydrated, embedded in Paraplast and sectioned at 5  $\mu$ m. The tissue sections were subjected to *in situ* hybridization using an ISH Reagent kit (Genostaff, Tokyo) according to the manufacturer's instructions, and were counterstained with Kernechtrot stain solution (Muto, Tokyo). DIG-labeled riboprobes were detected with anti-Digoxigenin AP Conjugate (Roche Diagnostics GmbH) and visualized using NBT/BCIP Solution (Sigma, St. Louis, MO). Tissue sections were also stained with hematoxylin-eosin to check their morphology.

### RNA-seq-based transcriptome analysis

Using the method described in Takahata et al. [16], we prepared RNA seq libraries from the RNA samples for preliminary expression analysis and sequenced using a HiSeq 2500 Instrument (Illumina) configured to generate 100-bp paired-end reads (DRX094096–DRX094113; [S7 Table](#)). The RNA-Seq reads for each RNA sample were quality filtered and adaptor trimmed, and separately mapped to the spinach genome assembly SOL\_r1.1 (<http://spinach.kazusa.or.jp>; [20]) using HISAT 2.2.1 (<https://github.com/DaehwanKimLab/hisat2>; [36]). Expression values of genes were estimated using StringTie and Ballgown software [37, 38].

For the primary expression analysis, we prepared RNA seq libraries using the NEBNext Ultra II Directional RNA Library Prep Kit (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. The library mixture was sequenced by 1  $\times$  100-bp single read sequencing using the Illumina platform NextSeq 1000 (Illumina). The read data were deposited in the DDBJ Sequence Read Archive (DRX477738–DRX477746 [S7 Table](#)). Differentially expressed gene (DEG) analysis was performed using a CLC Genomics Workbench 22 (Qiagen, Hilden, Germany). The clean reads were mapped to the reference genome. After statistical analysis based on a generalized linear model, DEGs with a change of more than |2|-fold and a false discovery rate (FDR)-adjusted p-value < 0.05 were selected.

### GO enrichment analysis

The genes predicted in SOL\_r1.1a [20] were annotated with Gene Ontology (GO) terms using eggNOG mapper [39]. To elucidate the biological and physiological meaning of the

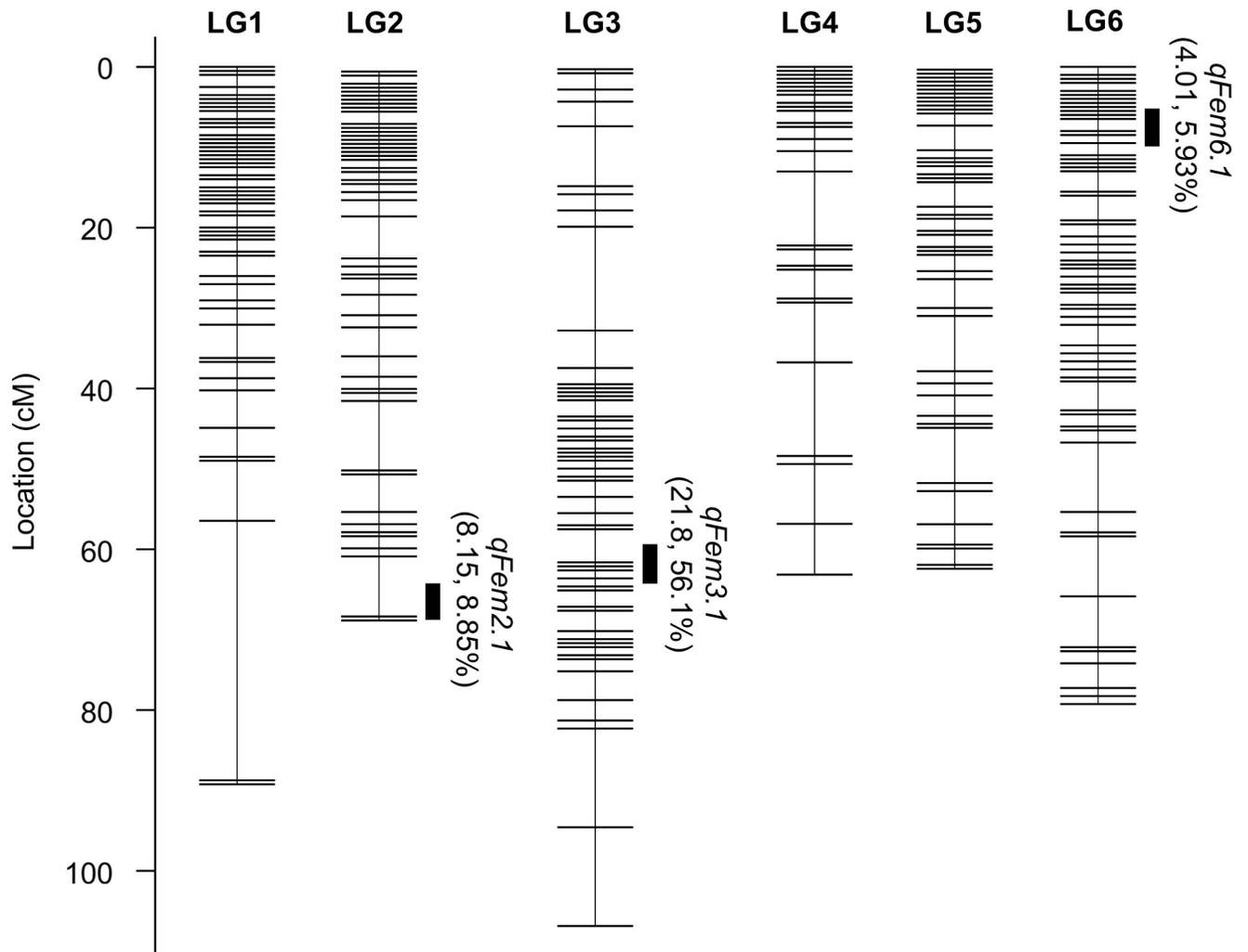
differentially expressed genes (DEGs) identified in this study, we carried out GO enrichment to identify GO terms over-represented in the DEGs. The significance of the enrichment of GO terms with DEGs was evaluated using Fisher's exact test implemented in the R package topGO [40]. In addition to the conventional Fisher's exact test (classic), the statistical significance was assessed by three other algorithms, i.e., elim, weight, and weight01, to identify significantly enriched GO terms for biological processes. In this study, a GO term was considered significantly enriched when the p-value calculated with the elim, weight, or weight01 algorithms was less than 0.01.

## Results

### Construction of a linkage map and chromosome-scale pseudomolecules for spinach line 03–009

To obtain accurate variant data covering the whole genome required for linkage and QTL analyses, we set out to create a spinach reference genome by correcting pseudomolecules (SOL\_r1.0\_pseudomolecule) for spinach line 03–009. To achieve this, the initial genome assembly Spol\_r0.0 generated by Hirakawa et al. [20] was refined using both long and short reads and then scaffolded using Hi-C mate-pair reads (S1 and S2 Tables). As shown in S1 Table, the resultant genome assembly Spol\_r1.0 consists of 323 scaffolds and has a total length of 935.4 Mb with higher continuity (N50 = 74.6 Mb) and BUSCO completeness (97.4%) in comparison to Spol\_r0.0. This finding is consistent with a recent study which showed that assemblies with relatively high N50 values consistently had high BUSCO scores [41].

To construct a linkage map, ddRAD-seq reads obtained from F<sub>2</sub> plants and an F<sub>1</sub> plant from a 03–009 × 03–336 cross and their parents (03–009 and 03–336) were mapped to Spol\_r1.0, and 171,522 SNPs were called from this assembly. Linkage analysis of 1816 filtered SNPs produced six linkage groups ranging from 62.2 to 106.6 cM with a total size of 468.6 cM (Fig 1, S8 Table). Since the ddRAD-seq reads were derived from female (03–009) and monoecious (03–336) plants having XX sex chromosome type, one of the six linkage groups—i.e., the group that was sex-linked—was considered to represent the X chromosome. Therefore, although the genome assembly Spol\_r1.0 was derived from genomic DNA of the male plant having XY sex chromosome type, we intended to construct pseudomolecules representing a haploid set containing the X chromosome by anchoring Spol\_r1.0 to the six linkage groups. The 65 scaffolds (893.1 Mb in total) accounting for 95.5% of the total length of Spol\_r1.0 were anchored to the six linkage groups. As five of the anchored scaffolds were found to contain segments (13.9 Mb in total) derived from Y-specific non-recombining regions, based on the M/F ratio score (see the Materials and Methods section for details), the scaffolds were trimmed by filtering out the segments. The trimmed scaffolds (879.2 Mb in total), accounting for 94.0% of the total length of Spol\_r1.0, were oriented and ordered with reference to the linkage map, were catenated while filling in gaps between them with Ns of 10 kb in length, and were built into a provisional set of six pseudomolecules. Hi-C contact maps of the pseudomolecules were constructed to detect errors caused by wrong ordering and orientation of the scaffolds. The final set of pseudomolecules, SOL\_r2.0\_pseudomolecule, was produced by several rounds of manual error correction guided by the Hi-C contact map (S8 Table). As shown in S5 Fig, SOL\_r2.0\_pseudomolecule was validated using its Hi-C contact map, which suggested that the scaffolds were mostly properly integrated into the final set of the pseudomolecules. Based on the BUSCO methodology [33], the pseudomolecules were assessed as representing 97.0% of genome assembly completeness. Furthermore, based on the LTR Assembly Index (LAI) [42], which evaluates assembly continuity using long terminal repeat retrotransposons (LTR-RTs), SOL\_r2.0\_pseudomolecule was shown to be of reference grade quality (10 < LAI) (S1 Table).



**Fig 1. Molecular linkage map of spinach based on the mapping population 03–009 x 03–336 F<sub>2</sub>.** Black bars on the right side of the map represent QTLs (95% CI) for femaleness (proportion of female flowers per plant). The left and right values in parentheses represent LOD values and phenotypic variance explained (PVE) values of the QTLs.

<https://doi.org/10.1371/journal.pone.0296675.g001>

It may be worth noting that the longest pseudomolecule (Chr3), which corresponded to linkage group LG3, was found to represent the sex chromosomes. This is because the protein-coding genes (*phosphoglucomutase* [*pgm*]/g22838.t1 [Chr3\_g077230.1], acyl-activating enzyme 6 [*aae6*]/g17189.t1 [Chr3\_g069500.1], aminotransferase [*tat2*]/g27450.t3 [Chr3\_g042770.1], *nitrite reductase* [*nir*]/g9296.t1 [Chr3\_g027530.1], and *chloride channel protein* [*clc-e*]/g49095.t1 [Chr3\_g017620.1]) that were previously mapped on the sex chromosomes were found on this longest pseudomolecule (DOI: [10.6084/m9.figshare.23916627](https://doi.org/10.6084/m9.figshare.23916627)) [16].

### Fine mapping of the *M* locus and expression analysis of its candidate genes

In our previous study, the locus responsible for monoecism (*M*) was identified as a major factor responsible for the expression of the monoecious character, and positioned between the *clc-e* (g49095.t1 [Chr3\_g017620.1]; Chr3 21.9 Mb) and *clsy3* (g48854.t1 [Chr3\_g019780.1]; Chr3 24.7 Mb) loci corresponding to the two DNA markers SP\_0008 and SP\_0140 on the sex chromosome. Based on the spinach reference genomes (SOL\_r2.0\_pseudomolecule; [S8 Table](#)),

the interval between *clc-e* and *clsy3* loci was estimated to be 2.8 Mb in length (S6 Fig). Here, we attempted to further narrow down the *M* candidate region in order to identify the DNA sequence corresponding to the *M* gene. As in our previous study, progeny families ( $S_3BC_2F_1$  generation) from a monoecious selection ( $S_2BC_2F_1-34-1$ ; Fig 2 in Takahata et al. [16]) in the 03-009x03-336  $S_2BC_2F_1$  population were examined to narrow down the candidate region for the *M* locus, using seven DNA markers spanning the interval of *clc-e* and *clsy3* (S6 Fig).

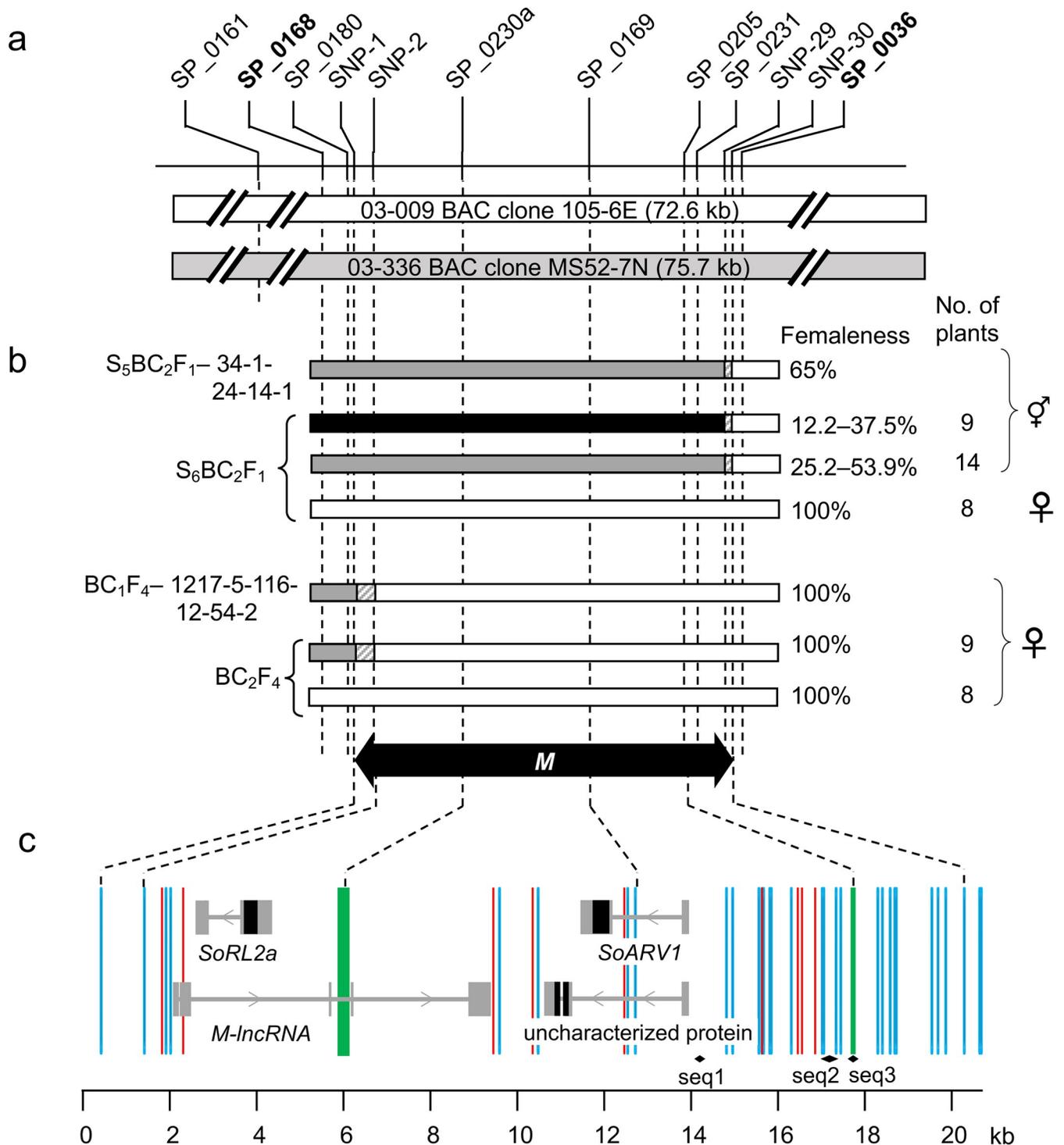
As shown in S6 Fig,  $S_2BC_2F_1-34-1$  is heterozygous at the interval of SP\_0008 and SP\_0140 except at the SP\_0008 and SP\_0124 locus, which is homozygous for the 03-009 and 03-336 genotypes, respectively. Self-pollination of  $S_2BC_2F_1-34-1$  produced monoecious and female  $S_3BC_2F_1$  plants, suggesting that the interval is heterozygous for the monoecious gene. Among the five markers bracketed by SP\_0008 and SP\_0140, only the SP\_0036 genotypes co-segregated with the monoecious phenotype in the 15  $S_3BC_2F_1$  plants examined, suggesting that the *M* locus is located around the marker locus bracketed by SP\_0124 and SP\_0054 (S6 Fig).

For fine mapping of the *M* locus, progeny families from monoecious selections ( $S_3BC_2F_1-34-1-24$  and  $F_3-1217-5-116-12$ ) in 03-009x03-336  $S_3BC_2F_1$  and in  $F_3$  populations were analyzed using nine markers spanning from SP\_0124 to SP\_0054 (S7 Fig). As shown in S7 Fig, five markers spanning from SP\_0138 to SP\_0168 were found to be co-segregated with the monoecious phenotype in the nine progeny plants from  $S_3BC_2F_1-34-1-24$ , while three markers spanning from SP\_0036 to SP\_0054/SP\_0053 were co-segregated with the phenotype in the three progeny plants from  $F_3-1217-5-116-12$ . The results from the analysis of the progeny families indicate that the locus for monoecism (*M*) is located in the interval between SP\_0168 and SP\_0036, which comprises a 20.7-kb physical distance.

To verify the location of the *M* locus, self-pollinated progeny ( $S_5BC_2F_1$  and  $S_6BC_2F_1$  generations) from the monoecious selection “34-1-24-14” in 03-009x03-336  $S_4BC_2F_1$  generation, and backcross progeny ( $BC_1F_4$  and  $BC_2F_4$ ) from the female selection “1217-5-116-12-54” in the 03-009x03-336  $F_4$  generation were examined (S7 Fig). As shown in Fig 2, phenotypes and genotypes of the progeny plants supported the notion that the *M* locus locates at the interval SP\_0168–SP\_0036. The chromosomal recombination sites in the interval of the progeny plants were determined by sequencing, which further narrowed down the *M* candidate region to a 19.5 kb interval between SNP1 and SNP30 (Fig 2).

As shown in Fig 2C and Table 1, three protein-coding genes, namely, *RADIALIS-like 2a* (*SoRL2a/g48915.t1* [Chr3\_g019240.1]), an uncharacterized protein coding gene (*g48913.t1* [Chr3\_g019250.1]) and a gene for Acyl-coA acyltransferase-related enzyme 2 required for viability (*SoARV1/g48912.t1* [Chr3\_g019260.1]), were predicted in the 19.5-kb *M* candidate region. It is worth mentioning that the *M* candidate region contains short DNA fragments having high homology with homologs for two Arabidopsis B-class floral organ identity genes, spinach *PISTILLATA* (*SpPI*) and *APETALA3* (*SpAP3*) (GQ120477 and GQ120478): the 343-bp and 97-bp sequences in the *M* candidate region were found to have high homology (89% and 73% identity) with the 5th and 3rd introns of *SpPI*, respectively (Fig 2C). The 343-bp sequence also shows high homology (85% identity) to the upstream region of *SpAP3*. Furthermore, the 75-bp sequence in the *M* candidate region was found to have high homology (90% identity) to the 6th intron of *SpAP3*.

Sequence comparison of BAC clones carrying the *M* locus of lines 03-009 and 03-336 led us to find 45 variant sites in the 19.5-kb region (30 SNPs, 9 indels, and 2 structural variants [SV]) (Fig 2C). However, these variants were not located in coding regions of the three protein-coding genes (Fig 2C), though three variants (1 indel, 2 SNPs) were found in the introns of *SoARV1* and *SoARV1-2*. By contrast, a 7-bp indel and a 327-bp indel (SV) were found in an exon and an intron of the *M-lncRNA* locus, respectively. Two SNPs and one indel were found within the 500-bp upstream flanking region of the *lncRNA* locus. Furthermore, three SNPs



**Fig 2. Physical map of the *M* locus.** **a.** Schematic diagrams representing allelic bacterial artificial chromosome clones of the *M* locus from lines 03–009 and 03–336. Positions of the DNA markers are shown on the diagrams. **b.** Graphical genotypes and femaleness of progeny families from monoecious selections in  $S_5BC_2F_1$  and  $BC_1F_4$  populations produced by crosses between 03–009 and 03–336. Black boxes represent homozygous 03–336 segments, gray boxes indicate heterozygous 03–009/03–336 regions, and white boxes represent homozygous 03–009 segments. Indices of femaleness (expressed as the percentage of female flowers per plant) are shown to the right of the boxes. **c.** Schematic diagrams of protein-coding genes and a lncRNA locus predicted in the 19.5-kb *M* candidate region. Exons and coding regions are represented as gray and black boxes, respectively. Introns are depicted as horizontal gray lines. Gene directions are indicated by arrowheads. Vertical blue, red and green lines represent SNPs, small indels ( $\leq 7$  bp) and structural variants (SV) found between lines 03–009 and 03–336. Diamonds (a–d) represent regions having significant sequence homology with the spinach *PISTILLATA* (*PI*) and *APETALA3* (*AP3*) genes. **seq1**, the

97-bp region, has homology with the third intron of *PI*; **seq2**, the 343-bp region, has homology with the 5th intron of *PI*, and the upstream region of *AP3*; **seq3**, the 75-bp region, has homology with the 6th intron of *AP3*.

<https://doi.org/10.1371/journal.pone.0296675.g002>

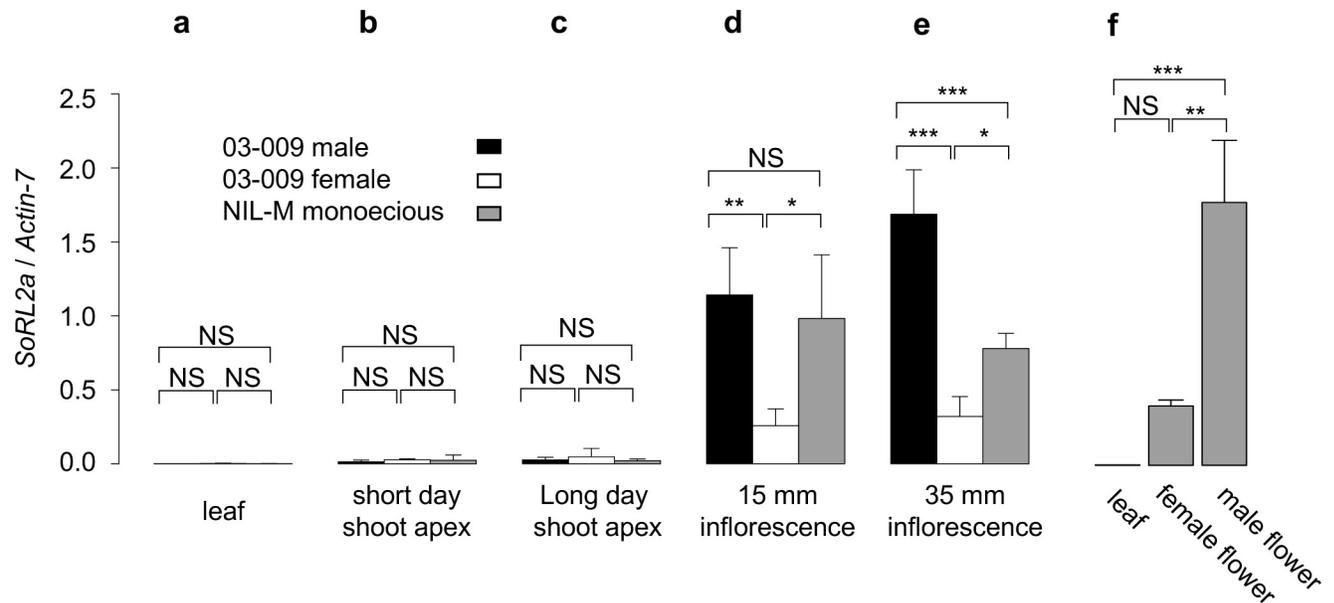
were located in the 343-bp sequence (**'seq2'** in Fig 2C) with homology to *SpPI*, and the position of the 75-bp sequence (**'seq3'** in Fig 2C) with homology to *SpAP3* was matched with an SV (110-bp indel): the 75-bp sequence was present in the *M* locus of 03–009 but absent in that of 03–336. To make it easy to determine the genotypes at the *M* candidate region in subsequent analyses, a codominant marker, SP\_0205, which targets the 110-bp indel, was developed (Fig 2 and S5 Table).

To identify the reproductive tissue-expressed gene(s) and the *M* gene from the *M*-candidate region, preliminary RNA-seq analysis of monoecious, female, and male plants was carried out without biological replicates. The female and male plants used in this analysis were from dioecious-line 03–009, and the monoecious plants were from line NIL-*M*, which is homozygous for *M* and produced by introgression of the *M* locus of line 03–336 into the genetic background of 03–009 (S3 Fig). Since 03–336 and 03–009 are late- and early-bolting/flowering lines, respectively, dioecious 03–009 and monoecious NIL-*M* lines with nearly identical bolting characteristics were used in this analysis. As shown in S8 Fig, *SoRL2a* was found to be expressed at relatively low levels in vegetative tissues (leaves) (FPKM = 0.6–4.1) and shoot apices (FPKM = 1.6–14.6) from plants of all sexes. By contrast, its expression levels in the reproductive tissues (inflorescences) of the plants were relatively high and increased with maturation. It is also worth mentioning that *SoRL2a* was estimated to be expressed more in the inflorescences of the male (FPKM = 61.4–78.9) and monoecious (FPKM = 17.1–66.2) plants than in those of the female (FPKM = 5.8–29.7) plants. In plants of all sexes, the uncharacterized protein coding gene (g48913.t1 [Chr3\_g019250.1]) had the highest TPM score (17.5–25.0) in shoot apices under short-day (SD) conditions. *SoARV1* (g48912.t1 [Chr3\_g019260.1]) was expressed at low levels in leaves (FPKM = 2.2–2.6), shoot apices (FPKM = 2.0–5.9) and reproductive tissues (inflorescences) (FPKM = 2.0–3.3) of the male, monoecious, and female spinach plants. The levels of *M-lncRNA* expression were also estimated to be constitutively low (FPKM = 0.4–1.6) in plants of all sexes. There was no obvious trend associated with reproductive stages or sex types in the expression patterns of *SoARV1*, the uncharacterized protein-coding gene, or *M-lncRNA*. *SoRL2a* is a gene that encodes a small protein, which contains a SANT/MYB domain. *SoRL2a* is a homolog of the *RADIALIS* gene in *Antirrhinum*, which controls the asymmetry of flowers. A recent study revealed that a *RADIALIS* homolog, *DkRAD*, plays a crucial role in the reversion from male to hermaphrodite flowers in hexaploid persimmon by influencing gynoecium development. Since members of the *RADIALIS* family are

**Table 1. Genes located on the 19.5-kb candidate genomic region for the *M*-locus.**

gene name	SOL_r1.1a gene id [16]	SOL_r2.0 gene id	blastp vs NR	E-value	blastp vs Araport11	E-value
<i>SoRL2a</i>	g48915.t1	Chr3_g019240.1	XP_021847620.1 RADIALIS-like 2 [ <i>S. oleracea</i> ]	4.60E-40	AT1G75250.1 RAD-like 6	1.00E-36
uncharacterized protein coding gene	g48913.t1	Chr3_g019250.1	XP_021847648.1 uncharacterized protein isoform X2 [ <i>S. oleracea</i> ]	3.00E-46	AT4G39235.2 hypothetical protein	4.00E-33
<i>SoARV1</i>	g48912.t1	Chr3_g019260.1	XP_021847636.1 protein arv1 homolog isoform X1 [ <i>S. oleracea</i> ]	2.90E-118	AT1G01020.1 ARV1 family protein	5.00E-61
			blastn vs NR	E-value		
<i>M-lncRNA</i>	g48914.t2	SOL_r2.0_LOC110787320	XR_002533067.1 uncharacterized transcript variant X2, ncRNA [ <i>S. oleracea</i> ]	0		

<https://doi.org/10.1371/journal.pone.0296675.t001>



**Fig 3. RT-qPCR analysis of *SoRL2a*.** a-e, RT-qPCR analysis of *SoRL2a* in male and female plants from the dioecious line 03-009, and monoecious plants from line NIL-M; f, RT-qPCR analysis of *SoRL2a* in leaves, and female and male flowers of monoecious plants. The amount of *SoRL2a* mRNA was normalized to that of *Actin-7*.

<https://doi.org/10.1371/journal.pone.0296675.g003>

involved in the formation of floral organs and sexual differentiation, and *SoRL2a* is preferentially expressed in reproductive organs, we considered that *SoRL2a* would be a promising candidate gene for male organogenesis and monoecious expression. We performed further verification to confirm this. It is worth noting that, among *RAD*-like genes in plants, *SoRL2a* was closely related to *RL2* (XM\_010692578) in *Beta vulgaris* in our phylogenetic analysis (S9 Fig).

*SoRL2a* mRNA levels relative to *Actin-7* mRNA were quantified by RT-qPCR, to further characterize the *SoRL2a* expression profile in monoecious (NIL-M), female (03-009), and male (03-009) plants. As shown in Fig 3A–3E, the RT-qPCR assay reconfirmed that *SoRL2a* was predominantly expressed in inflorescences. Our RT-qPCR assay also showed that *SoRL2a* expression was 3.7–4.4-fold greater in the early-stage inflorescences of the male and monoecious plants than in those of the female plants. Furthermore, among inflorescences at the subsequent developmental stage, the male plant had the highest level of *SoRL2a* expression.

As shown in Fig 3F, RT-qPCR assay of tissues from inflorescences of monoecious plants showed that *SoRL2a* mRNA expression was approximately four-fold higher in male flowers than in female flowers. Its expression in leaves was shown to be almost absent, but not significantly different from that in female flowers. The lack of significance was likely due to the limited number of biological replicates ( $n = 3$ ). Together with the results above, these findings indicate that *SoRL2a* mRNA is expressed in flowers of monoecious, female, and male plants, and the higher expression of *SoRL2a* mRNA in male and monoecious inflorescences may be due to its higher expression in male flowers than in female flowers.

To elucidate the detailed temporal and spatial expression profiles of *SoRL2a* mRNA within male and female flowers, *in situ* hybridization analysis was carried out on tissue sections from a male and a female inflorescence (S10 Fig). As shown in the photos of male flower sections (sections a and a' in S10 Fig), *SoRL2a* mRNA was expressed in the endothecia and middle layers of anthers before the start of meiosis (male flower 1 in sections a and a') but was absent in

the anthers at the earlier stage (male flower 2 in sections **a** and **a'**). Furthermore, *SoRL2a* mRNA was found to persist for a short period after meiosis (sections **b** and **b'** in S10 Fig). *In situ* hybridization analysis of female flowers showed that *SoRL2a* mRNA was expressed at the base of the outer and inner integuments. However, *SoRL2a* mRNA was absent in the female flowers at the earlier stage when gynoceium primordia are initiated (data not shown). The higher expression of *SoRL2a* mRNA in male flowers than in female flowers was likely attributable to the regions where it was expressed in the male flowers being larger than those in the female flowers.

### Mapping of QTLs controlling femaleness in monoecious spinach

To investigate whether loci other than the *M* locus might be involved in the monoecious expression, female plants (*mm*) from line 03–009 ( $n = 7$ ), monoecious plants (*MM*) from line 03–336 ( $n = 8$ ), and  $F_1$  ( $n = 8$ ) and  $F_2$  ( $n = 92$ ) generations from the cross 03–009 x 03–336 were evaluated for femaleness. Monoecious plants from NIL-*M* ( $n = 8$ ) were also subjected to the analysis.

As shown in S11 Fig, female plants of genotype *mm* from line 03–009 and monoecious plants of genotype *MM* from line 03–336, which were grown as control populations, displayed 100% and 4.8% femaleness scores on average, respectively. The  $F_1$  plants (*Mm*) showed femaleness scores (64.7–77.4%, average = 72.7%) intermediate between those of the parental lines, which is consistent with our previous finding that *M* acts as an incomplete dominant gene [15]. Femaleness scores of  $F_2$  plants from 03–009 x 03–336 ranged from 0% to 100% (average = 65.3%). When the  $F_2$  plants were genotyped with marker SP\_0205 at the *M* locus (S5 Table), 22 of the 25 *mm* plants (homozygous for the 03-009-derived allele of SP\_0205) were found to show the female character (100% femaleness); the three exceptions were plants that were not strictly female but exhibited highly female monoecious conditions (92–97% femaleness). By using the 19 markers shown in S4 Table and S6 and S7 Figs, two (92% and 97% femaleness) of the three highly female monoecious plants were confirmed to be homozygous for the 03–009 genotype in the interval between SP\_0008 and SP\_0069 (~135.3 Mb in size), which includes the *M*-candidate region. The other one (93% femaleness) was also shown to be homozygous for the 03–009 genotype in the interval between SP\_0125 and SP\_0069 (~133.7 Mb in size), which includes the *M*-candidate region. Twenty-two *MM* plants (homozygous for the 03-336-derived allele of SP\_0205) exhibited the monoecious character with relatively lower female scores (0–64%, average = 30.0%). Furthermore, 48 of 49 *Mm* plants (homozygous and heterozygous for the 03-336-derived allele of SP\_0205) exhibited monoecious conditions with a wide range of femaleness scores from 0% to 99% (average = 65.2%), with the exception being a single plant with a strict female condition with a femaleness score of 100%. This result was likely attributable to the incomplete dominant nature of the *M* gene: the *Mm* genotype often produces plants with higher femaleness scores than the *MM*. These results supported our conclusion that *M* is the major factor regulating the monoecious expression [15, 17].

The respective *MM* and *Mm* plants in the  $F_2$  population had a wider range of femaleness scores compared to monoecious plants from 03–336 (*MM*) and the  $F_1$  generation (*Mm*) (S11 Fig). Furthermore, monoecious plants from NIL-*M* ( $n = 8$ ) that were homozygous for *M* and were produced by introgression of the *M* locus of 03–336 into the genetic background of 03–009 showed femaleness scores (38–88%, average = 63.5%) significantly higher than those exhibited by the 03–336 plants (Exact Wilcoxon rank sum test,  $W = 64$ ,  $p < 0.001$ ). These findings may suggest that monoecious expression (degree of femaleness) is controlled by not only the *M* locus but also other loci.

In order to identify the additional loci for monoecism, other than *M*, QTL analysis for femaleness was performed using the 03–336 x 03–009 F<sub>2</sub> population. We detected three QTLs with LOD scores greater than 4.0; these were designated *qFem2.1* (LOD = 8.15), *qFem3.1* (LOD = 21.8), and *qFem6.1* (LOD = 4.0) on LGs 2, 3, and 6, respectively (Fig 1). However, after performing the permutation test, only two (*qFem2.1* and *qFem3.1*) QTLs remained significant. Referring to our pseudomolecules, the chromosomal region for *qFem3.1* (99% CI) was found to harbor the *M* candidate region.

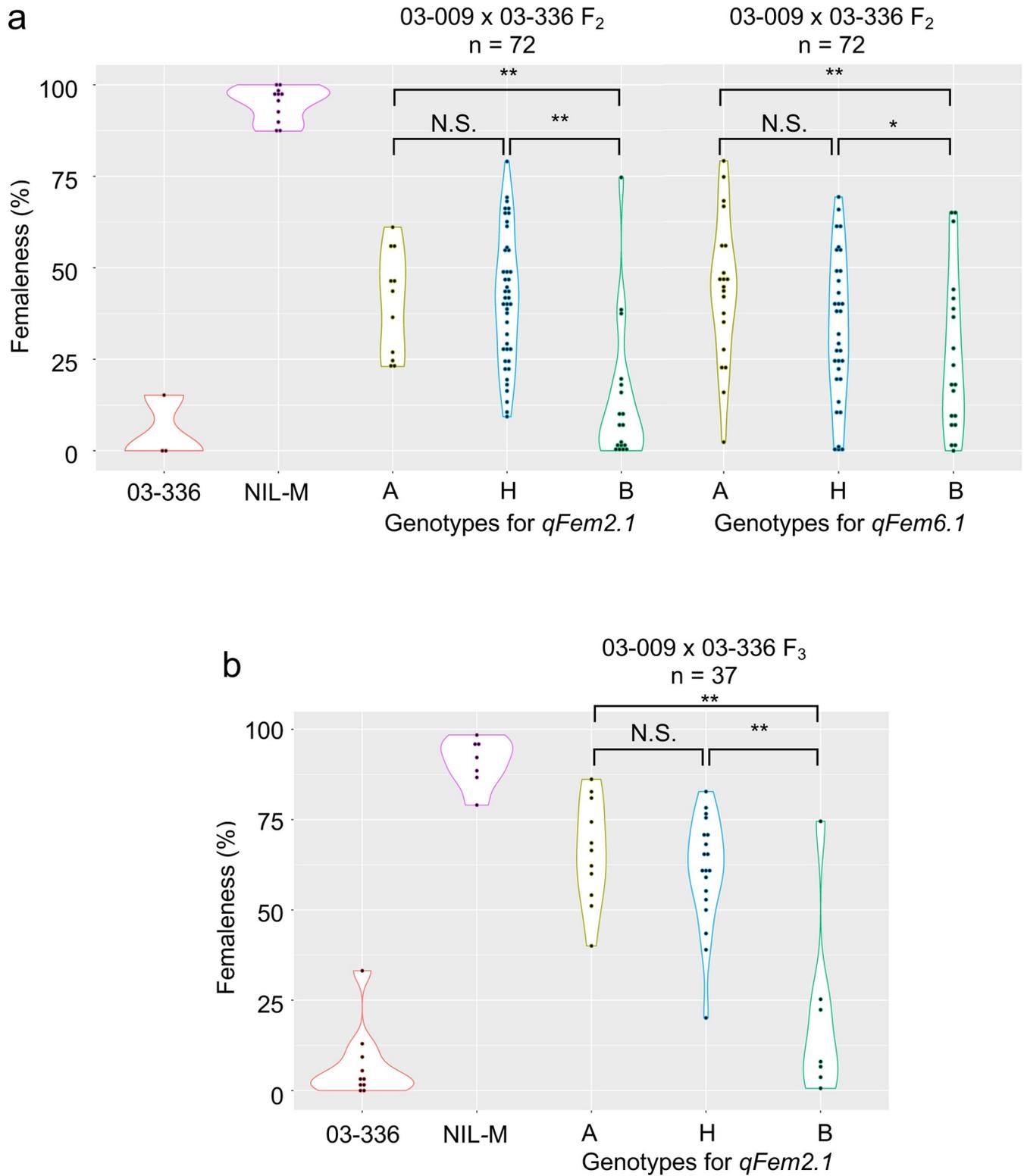
### ***qFem2.1* and *qFem6.1* act as modifiers of the *M* gene function**

Next, we plotted the means of femaleness scores in the F<sub>2</sub> population against genotypes at the QTLs, as shown in S12 Fig. The plots suggested that *qFem2.1* and *qFem6.1* may affect femaleness in the monoecious plants carrying the *M* allele—i.e., *MM* and *Mm* plants in the F<sub>2</sub> population tended to exhibit lower femaleness as the dose of 03-336-derived alleles for *qFem2.1* and *qFem6.1* increased. Still, this trend was not observed in *mm* plants. The “fitqtl” function of the software package R/QTL detected the interaction effect between *qFem3.1/M* and *qFem2.1* (ANOVA *F*-test *p*-value < 0.05; S9 Table) in the F<sub>2</sub> population but failed to detect an interaction effect between *qFem3.1/M* and *qFem6.1* (data not shown). The plants used in the analyses described above were grown at a high temperature (25°C) during a reproductive developmental stage (the periods when plants were grown under long-day photoperiods), since the high-temperature conditions bring about a shift towards maleness in monoecious plants [3, 43]. Therefore, compared to low-temperature conditions, high-temperature conditions can be expected to make *Mm* plants exhibit obvious monoecious conditions with relatively low femaleness scores. To verify the effect of *qFem2.1* and *qFem6.1* on femaleness, the F<sub>2</sub> progeny plants from the cross between 03–336 x 03–009 were grown again, and then 72 *MM* plants were selected from the F<sub>2</sub> progeny. The *MM* F<sub>2</sub> plants and control plants (females [*mm*] from 03–009, monoecious [*MM*] from NIL-M and 03–336) were grown under a low temperature (20°C), since these *MM* plants were expected to exhibit a wide range of femaleness scores from low to high depending on their genetic backgrounds. Subsequently, the *MM* plants were genotyped for *qFem2.1* and *qFem6.1* using dCAPS markers (SP\_0319 and SP\_0325; S5 Table).

As shown in Fig 4A, female plants from 03–009 displayed the strictly female condition (100% femaleness score), and femaleness scores exhibited by the plants from NIL-M (*n* = 11) and 03–336 (*n* = 3) were 94.9% and 5.1% on average, respectively: the difference in femaleness scores of the two monoecious lines grown under the low-temperature condition was more remarkable than that under the high-temperature condition (S11 Fig). In the 72 *MM* plants selected from the F<sub>2</sub> generation, the mean of femaleness scores exhibited by plants homozygous for 03-336-derived alleles for *qFem2.1* and *qFem6.1* was significantly lower than that of plants heterozygous and homozygous for 03-009-derived alleles (Fig 4A). Furthermore, the effect of *qFem2.1* on femaleness was confirmed again in the F<sub>3</sub> progeny (*n* = 37) from an F<sub>2</sub> plant with the *MM* genotype, which was heterozygous at the *qFem2.1* locus, but homozygous for 03-009-derived alleles at the *qFem6.1* locus (Fig 4B). The examination for *qFem6.1* using the F<sub>3</sub> generation was not carried out due to the failure to produce F<sub>3</sub> progeny from F<sub>2</sub> plants heterozygous at the *qFem6.1* locus.

### **Transcriptome analysis of 03–009, 03–336 and NIL-M**

To identify candidates for genes involved in the regulation of femaleness in monoecious plants, transcriptome profiles of early-stage inflorescences from highly male monoecious (03–336), highly female monoecious (NIL-M), and female (03–009) plants grown at a low temperature (20°C) were analyzed by RNA-seq. As shown in S13 Fig, 1,834 up-regulated and 2,153



**Fig 4. Violin plots showing the distribution of femaleness exhibited by MM plants selected from F<sub>2</sub> and F<sub>3</sub> progeny of the cross between 03-009 and 03-336. Statistical significance was determined by Tukey-Kramer test (N.S.  $p \geq 0.05$ ; \*  $p < 0.05$ ; \*\*  $p < 0.01$ ). A, homozygous for the 03-009 allele; H, heterozygous for the 03-009 and 03-336 alleles; B, homozygous for the 03-336 allele.**

<https://doi.org/10.1371/journal.pone.0296675.g004>

down-regulated differentially expressed genes (DEGs) were identified by comparing 03–336 vs. 03–009, NIL-M vs. 03–009, and 03–336 vs. NIL-M plants. However, the protein-coding genes in the *M*-candidate region were not included in the DEGs.

One hundred thirty-six genes were found to be commonly up-regulated in highly male monoecious (03–336) and highly female monoecious (NIL-M) plants compared with their expression in female plants (03–009). These genes were further characterized by GO enrichment analysis (S13 Fig). As shown in S10 Table, the shared up-regulated DEGs were significantly enriched with GO terms including ‘stamen formation’ (GO: 0048455) and ‘specification of floral organ identity’ (GO: 0010093). Both DEGs (g3323.t1 [Chr5\_g058930.1] and g3324.t1 [Chr5\_g058920.1]) annotated with the GO terms were homologs (*SpPI*) for Arabidopsis *PIS-TILLATA* (*PI*). However, the homologs were located on Chr5 (LG5), which does not harbor an *M* locus or other QTLs. Karrikin (GO:0080167), RNA 5'-end processing (GO:0000966), behavior (GO:0007610), and vernalization response (GO:0010048) were also highly ranked among the enriched, shared up-regulated DEGs. 175 down-regulated DEGs were found to be shared between the comparison pairs 03–336 vs. 03–009 and NIL-M vs. 03–009 (S13 Fig). However, the shared DEGs were significantly enriched with biological process GO terms that were not directly related to flower organ formation or reproductive development processes, such as mismatch repair (GO:0006298), wax biosynthetic process (GO:0010025), and mitochondrial mRNA processing (GO:0090615) (S11 Table).

Eight-hundred thirty-six genes were found to be up-regulated in highly male monoecious (03–336) plants compared to both highly female monoecious (NIL-M) and female (03–009) plants and were characterized based on GO enrichment analysis (S13 Fig). The up-regulated DEGs also included *PI* homologs and were shown to be significantly enriched with four GO terms related to reproductive organogenesis (GO:0009554, GO:0048658, GO:0009556, and GO:0048455) (S12 Table). However, the DEGs were most significantly enriched with the GO term ‘defense response to fungus’ (GO:0050832). ‘Auxin export across the plasma membrane’ (GO:0010315) was another GO term significantly enriched among the DEGs.

One thousand and five genes were found to be down-regulated in highly male monoecious (03–336) plants compared to both highly female monoecious (NIL-M) and female (03–009) plants (S13 Fig). The down-regulated DEGs were shown to be significantly enriched with GO terms related to the reproductive process, such as ‘vernalization response’ (GO:0010048), ‘regulation of flower development’ (GO:0009909) and ‘positive regulation of reproductive process’ (GO:2000243) (S13 Table). The down-regulated DEGs were also significantly enriched with genes related to defense response (GO:0006952 and GO:0098542), as in the up-regulated DEGs.

### Candidate genes for the QTLs *qFem2.1* and *qFem6.1*

Among the following conditions, the genes satisfying conditions (1) & (2) or conditions (1) & (3) were selected from the *qFem2.1* region (3.2 Mb, 95%CI) as the candidate genes. (1) The genes were expressed above 1 TPM, in early-stage inflorescences of at least one line. (2) The genes were predicted to be associated with reproductive, floral-organogenesis, or hormone metabolism/transport/response, and had nucleotide variants classified as exerting ‘HIGH’ or ‘MODERATE’ impact that were assumed to disrupt functions or change the effectiveness of encoded proteins. (3) The genes were predicted to be associated with the biological processes mentioned above, to be differentially expressed among the different lines (03–009, NIL-M and 03–336), and to have nucleotide variants classified as exerting ‘LOW’ or ‘MODIFIER’ impact that were assumed to be unlikely to change the functions of encoded proteins.

Four candidate genes were identified after searching the *qFem2.1* region (S14 Table). The candidate genes g23600.t1 (Chr2\_g000510.1) and g23706.t3 (Chr2\_g001540.1), which had

HIGH- and MODERATE-impact variants, are homologs of the Arabidopsis genes AT5G50850 (*MABI*) and AT5G13010 (*EMB3011 / PINP1*) associated with auxin transport and auxin-mediated organ development, respectively [44, 45]. The gene g23617.t1 (Chr2\_g000670.1) with a MODIFIER-impact variant was predicted to encode an auxin-responsive protein and was expressed more highly in 03–336 than in NIL-M or 03–009. The gene g23612.t1 (Chr2\_g000620.1) had a MODIFIER-impact variant and is a homolog of the Arabidopsis gene (AT5G14070.1.1/*ROXY2*) encoding a thioredoxin superfamily protein that was shown to be involved in anther development [46] and to be expressed more highly in 03–336 than in NIL-M or 03–009.

The two genes satisfying only conditions (1) & (2) above were selected from the *qFem6.1* region (57.1 Mb, 95%CI) as the candidate genes, since this region was too wide to select candidates based on the same conditions as in the case of *qFem2.1* (S14 Table). G49564.t1 (Chr6\_g038890.1) and g13001.t2 (Chr6\_g032140.1), which had MODERATE-impact variants, were predicted to encode LRR receptor-like kinase (LRR-RLK). The former gene is a homolog of the Arabidopsis gene, *CLAVATA3 INSENSITIVE RECEPTOR KINASEs (CIKs / AT3G02130)*, which was shown to regulate early anther development [47]. The latter one is a homolog of the Arabidopsis gene and *CLAVATA 1 (CLV1 / AT1G75820)*, respectively, which are required for maintaining floral meristem identity [48].

## Discussion

In this study, we successfully constructed a set of chromosome-scale pseudomolecules, SOL\_r2.0\_pseudomolecule, representing the genome of a dioecious line 03–009, by assembling PacBio long reads with the help of Hi-C-based chromatin proximity information and a RAD-Seq-derived SNP linkage map.

Our previous version of the genome assembly, SOL\_r1.1 [20], covered over 90% of the whole genome and had a high degree of continuity (N50 = 11.3 Mb, LAI = 15.91). However, we discovered that 29 out of the 112 scaffolds in SOL\_r1.1, which were anchored with the linkage map, were chimeric due to scaffolding errors. This may have been due to the size of the spinach genome, which is relatively large (~1000 Mb) and contains many repetitive elements. Additionally, the mate-pair libraries used to construct SOL\_r1.1 were limited to 3, 6, 10, and 15 kb inserts, which provided proximity information for only a small area. Hi-C analysis generates genome-wide chromatin interaction data, providing genomic proximity information with a scale ranging from 1 kb to 1 Mb or longer, to assign, order, and orient genomic sequences, achieving high accuracy in genome assembly for humans and goats, both of which have large-size genomes [25, 49]. Therefore, combining long-read sequencing technologies with Hi-C proximity data may be appropriate to construct chromosome-scale assembly for this genome, rather than combining those with mate-pair reads. Finally, SOL\_r2.0\_pseudomolecule had a total length of 879.2 Mb and scored 97.0 for BUSCO (gene completeness) and 19.94 for LAI (sequence continuity). It outperformed the SOL\_r1.0\_pseudomolecule constructed from the genome assembly SOL\_r1.1 in terms of total length, BUSCO score (81.5), and LAI score (17.08) (S1 Table). In recent years, by combining long-read and Hi-C data, pseudomolecules of the genomes for the spinach lines Monoe-Viroflay, Viroflay, and Cornell-No. 9 have been constructed [14, 50] (S1 Table). The pseudomolecules for Monoe-Viroflay and Virelay have good gene completeness scores (BUSCO completeness  $\geq$  97%) and high assembly continuity scores (LAI > 19.0). Similarly, the SOL\_r2.0\_pseudomolecule is of high quality and may be useful as a reference genome for future genetic studies toward identifying loci that control agronomic traits. Monoe-Viroflay and Viroflay likely derive from a European heirloom cultivar, since ‘Viroflay’ is generally known to be an old French spinach variety.

'Monoe-Viroflay' was reported to be provided by a European company (Enza Zaden, <https://www.enzazaden.com/>). In contrast, the breeding line 03–009 originates from a Japanese cultivar. Although the details of its pedigree information are not publicly available, line 03–009 is known to exhibit typical phenotypes observed in East Asian cultivars: early bolting, hastate and flat leaves, and production of prickly fruits [20]. Given that the use of only a single reference could bias the analysis toward a comparison within the chromosomal regions highly conserved with the reference due to difficulties in adequately aligning short-reads to highly divergent regions, the pseudomolecules for the 03–009 genome, SOL\_r2.0\_pseudomolecule, might be helpful as a reference for East Asian spinach genomes [51]. In particular, SOL\_r2.0\_pseudomolecule can be expected to serve as a reference to obtain highly accurate variant information in spinach germplasm collections derived from native varieties in East Asia, including Japan. Furthermore, the variant information obtained here will further understanding of the genetic structure of these genetic resource collections and is expected to be useful when using these genetic resource collections as breeding materials.

This study successfully identified three QTLs regulating femaleness in monoecious spinach plants by using variant data of F<sub>2</sub> progeny plants from the cross between 03–009 (female) and 03–336 (highly male monoecious) detected from SOL\_r2.0\_pseudomolecule. These QTLs were mapped to the SOL\_r2.0\_pseudomolecule, and the QTL with the highest LOD score, *qFem3.1*, corresponds to the major locus (*M*) for monoecism we previously identified [15–17]. PCR-based DNA markers targeting the QTLs were also successfully designed by referring to variants detected in the SOL\_r2.0\_pseudomolecule. These results indicate that the SOL\_r2.0\_pseudomolecule is suitable for NGS-based variant detection and genetic mapping studies and is a platform with enormous potential for developing PCR-based DNA markers that can aid in agronomic trait selection. Furthermore, our research provides valuable insights into the complex genetic regulatory mechanisms that govern the variability of femaleness in monoecious spinach plants. We have elucidated interaction patterns of the three QTLs: the *qFem3.1* (*M* locus) QTL is a crucial factor that is necessary for monoecious expression, while the *qFem2.1* and *qFem6.1* QTLs serve as modifiers for the *M* locus that affect femaleness. These findings will be valuable for spinach hybrid breeding programs using parental lines with highly male and highly female monoecious characters. The monoecious line 03–336 used in our study is a useful source of highly male monoecious characters that are desirable for paternal parents for hybrids. By pyramiding *qFem2.1*, *qFem6.1*, and *M* loci derived from monoecious line 03–336, using the variants and markers at the loci, it will be possible to efficiently produce paternal breeding lines. Similarly, the monoecious line NIL-*M* can be used as a source of highly female monoecious characters by utilizing the variants and markers present in the loci.

Of the three protein-coding genes in the *M* candidate region (the 19.5-kb interval between SNP1 and SNP30; Fig 2), only *SoRL2a* was found to be expressed in reproductive organs. *SoRL2a* is a homolog of *RADIALIS* that has been well characterized as a gene encoding a SANT/MYB domain transcription factor controlling floral asymmetry in *Antirrhinum majus* [52]. In tomato, a *RADIALIS*-like gene, *FSM1*, controls cell expansion and fruit development [53]. In Arabidopsis, the *RADIALIS-like 2* (*RSM1*) gene was reported to be involved in early morphogenesis [54, 55]. A recent study reported that a *RADIALIS* homolog, *DkRAD*, is a key factor for reversions from male to hermaphrodite flowers via gynoceium development in the monoecious hexaploid persimmon [56]. *DkRAD* was shown to be induced by abscisic acid and the cytokinin signaling responsible for the reversions to hermaphrodite flowers. However, in dioecious diploid persimmon, abscisic acid and cytokinin treatments cannot induce the reversions to hermaphrodite flowers, and thus cannot induce *DkRAD* expression. By contrast, *SoRL2a* was shown to be expressed not only in monoecious plants but also male and female

plants, suggesting that the role of this gene may be different from that of *DkRAD* in monoecious persimmon.

Sex differentiation in spinach flowers is not involved in the developmental arrest or abortion of one sex organ in bisexual floral primordia, since rudimentary opposite-sex organs were not found in unisexual flowers of spinach [57]. The floral sex differentiation system in spinach is, therefore, assumed to regulate the initiation of the stamen and pistil primordia. Spinach B class genes (*SoPI* and *SoAP3*) were shown to be involved in sex differentiation, since the suppression of the genes in male plants resulted in homeotic transformation of stamens into carpels and the presence of a fourth whorl [58]. In our study, *SoPI* was observed to be up-regulated in monoecious plants compared to female plants. However, the expression of *SoRL2a* mRNA was not observed in floral organ primordia, though it was observed in developing anthers and pistils (S10 Fig). Therefore, we cannot reasonably conclude that *SoRL2a* directly regulates the initiation of the sex organs in cooperation with the floral organ identity genes. According to Takahata et al. [16], the region of the spinach chromosome that contains the *M* locus shares a high degree of gene synteny with the corresponding region of the hermaphroditic species *Beta vulgaris* (sugar beet). Sugar beet belongs to the *Amaranthaceae* family, which also includes spinach. The chromosomal region in sugar beet that corresponds to the spinach *M* locus contains *BvRL2* (XM\_010692578), the ortholog of *SoRL2a* (S9 Fig). Since the function of *BvRL2* might be similar to that of *SoRL2a* in the hermaphroditic species, it is reasonable to assume that *SoRL2a* does not play a role in regulating floral sex determination, but could be involved in the development of anthers and ovules based on its expression profile.

Papaya is a species that has three types of individuals: male, female, and hermaphrodite. Profiling analyses of small non-coding RNAs (sRNA) from the flowers of papaya showed that in male flowers, most of the expressed miRNAs were related to the auxin signaling pathways. On the other hand, the miRNAs that are more highly expressed in female flowers have the potential to regulate the apical meristem identity genes. These findings suggest that these miRNAs could play critical roles in the sex expression in papaya [59]. In diploid persimmon, siRNAs generated from a Y-chromosomal gene, designated *OGI*, play a role in male determination by suppressing the expression of its autosomal paralog, *MeGI*, which is a male function suppressor [60]. Furthermore, epigenetic regulation, triggered by *OGI*-derived siRNA, of *MeGI* was shown to be responsible for monoecious expression in hexaploid persimmon [61]. A recent study found that two Y-chromosomal genes are involved in poplar sex determination: one produces siRNAs that repress the expression of its autosomal paralog, which acts as a female-promoting factor, while the other generates long non-coding RNAs promoting maleness [62]. In this context, we should consider the possibility that unidentified non-coding RNA might be involved in the monoecious expression in spinach. Given that the *SoPI* and *SoAP3*-homologous sequences in the *M* candidate region are transcribed and generate siRNA, the *M* locus could control the monoecious expression via modulation of the B class genes. Nevertheless, more empirical evidence is necessary to verify this speculation.

The shared up-regulated DEGs were shown to be enriched with the GO term 'stamen formation' (GO: 0048455), as expected (S10 and S12 Tables). The up-regulated DEGs with the GO term include two copies of *PI* homologs, suggesting that the B class genes play a crucial role in the monoecious expression. Chailakhyan and Khryanin [63] reported that the exogenous application of gibberellic acid (GA3) on dioecious spinach plants resulted in an increase in the proportion of male plants. The level of expression of the spinach homolog of Arabidopsis *GAI* (AT1G14920), a member of the GA-repressor DELLA family, was reported to be significantly lower in males than in females [64]. Subsequent knockdown analysis of the spinach *GAI* suggested that it regulates the B class genes and is involved in sex determination in

spinach [64]. However, our transcriptome analysis did not detect significant differences in the expression levels of spinach homologs (g15409.t1 [Chr5\_g029780.1] and g20972.t1 [Chr1\_g022500.1]) of Arabidopsis DELLA family members (*GAI* and *RGA1* [AT2G01570]) between the female and monoecious plants (data not shown), and our GO enrichment analysis failed to reveal an association of GA with the monoecious expression. Furthermore, the spinach *GAI* and *RGA1* homologs are located on Chr.1 and Chr.5, which are unrelated to the QTLs detected herein, suggesting that the regulatory mechanisms of the B class genes might be different between male and monoecious plants.

The up-regulated and down-regulated DEGs were found to be enriched with GO terms including 'bacterium response' (GO:0009617) and 'defense response' (GO:0006952, GO:0050832, GO:0098542) (S11 and S12 Tables). This might suggest that pathways for sex-differentiation and stress/defense response have some commonalities. Indeed, stress-response-related genes (e.g., defense response, drought/oxidative stress response) were shown to be involved in the sex-differentiation of persimmon [56]. We note that the RNA helicase gene (g23706.t3 [Chr2\_g001540.1]), which was located in the *qFem2.1* region, was annotated with GO:0006952 (defense response) and was included among down-regulated DEGs shared between 03–336 vs. 03–009 and 03–336 vs. NIL-M (S13 Table). Its Arabidopsis homolog (AT5G13010/ EMB3011 /*PINP1*) was previously shown to be associated not only with auxin-mediated organ development but also with plant immunity [44, 65].

The genes annotated with the GO term 'cellular response to toxic substance' (GO:0097237) in the up-regulated DEGs (S10 Table) encode a homolog of Arabidopsis UDP-glucosyltransferase (AT1G05680) induced by H<sub>2</sub>O<sub>2</sub>, and peroxidase superfamily proteins. Interestingly, UDP-glucosyltransferase was shown to be involved in auxin homeostasis [66]. Furthermore, the genes annotated with the GO term 'auxin export across the plasma membrane' (GO:0010315) were enriched among up-regulated DEGs shared between the comparison pairs 03–336 vs. 03–009 and 03–336 vs. NIL-M (S12 Table). Among genes located in the *qfem2.1* and *qfem6.1* regions that carry variants potentially affecting their function, auxin-related genes were found in addition to floral-organogenesis genes.

Auxins are involved in various plant developmental processes, including floral meristem initiation, floral organ boundary establishment, floral organ outgrowth and floral meristem determinacy [67]. *ARF6* and *ARF8*, members of the auxin response factor gene family, were reported to control stamen elongation and flower maturation in Arabidopsis [68]. A double mutant with defects of both genes has short petals, short stamen filaments, late dehiscent anthers, and immature gynoecia. *AUXIN RESPONSE FACTOR 3* is involved in auxin-regulated gynoecium patterning [69]. The Arabidopsis *SUPERMAN* (*SUP*) gene encodes a C2H2 zinc finger transcription factor that is expressed at the boundary of the stamen and carpel whorls [70]. This gene negatively regulates the expression of the auxin biosynthesis genes *YUCCA1/4* [71]. In *sup* mutants, the auxin levels increase at the boundary, leading to an increase in the number and the prolonged maintenance of floral stem cells [70, 71]. This results in supernumerary stamens and reduced or absent carpel tissue, as the fate of cells in whorl 4 changes from female to male [67]. Auxin plays a crucial role in reproductive organogenesis, which is why it may be involved in the sexual differentiation of flowers in certain dioecious and monoecious plants. In fact, auxins were reported to be potentially involved in sex differentiation in several dioecious plants, including spinach. Moreover, auxin application was shown to raise the proportion of female plants in a dioecious spinach cultivar with an even ratio of female to male plants under control conditions [67]. Comparative transcriptomic analysis of male and female floral buds in papaya suggested that auxin transporters might be involved in sex differentiation [72]. Heslop-Harrison [73] provided evidence that treating male dioecious hemp with auxins resulted in the development of female flowers. A SAUR (small auxin-up

RNA) gene was reported to be one of the putative candidate genes for controlling sexual traits in grapevine [74]. All these facts suggest that the auxin-related genes in the *qfem2.1* region could be promising candidate genes for mediating the sexual differentiation of flowers and regulating femaleness in monoecious spinach.

*CLAVATA3 (CLV3)* encodes a small secreted peptide that negatively regulates stem cell activity in Arabidopsis [75]. A recent study suggested that *CLV3* plays a role in regulating the size of floral meristems in coordination with *AGAMOUS (AG)* and *SUPERMAN (SUP)* [76]. Proper regulation of floral meristem activity is crucial for the normal development of reproductive organs, as suggested by *sup* mutants. Therefore, it is possible that homologs of *CLV3* could affect the sexual differentiation of flowers. A *CLV3*-like gene was recently identified as a potential sex-determining factor (gynoecium suppression factor) in *Silene latifolia* [77]. In the context of monoecious spinach, variants in the homologs of Arabidopsis *CLAVATA1* and *RPK2* [78, 79], which are receptors of *CLV3*, located in the *qfem6.1* region, may also be considered to affect femaleness. However, these ideas are preliminary and require further investigation.

## Supporting information

**S1 Fig. Crossing scheme to produce self-pollinated progeny families from a monoecious selection in a 03-009×03-336 BC<sub>2</sub>F<sub>1</sub> population.**

(PDF)

**S2 Fig. Crossing scheme to produce 03-009×03-336 BC<sub>1</sub>F<sub>4</sub>.**

(PDF)

**S3 Fig. Crossing scheme to obtain the line NIL-M bearing the XXMM genotype in the background of the dioecious line 03-009.**

(PDF)

**S4 Fig. Representative images of plant tissues used for RNA extraction.**

(PDF)

**S5 Fig. Hi-C contact map of the chromosomal pseudomolecules SOL\_r2.0\_pseudomolecule.**

(PDF)

**S6 Fig. Schematic diagrams showing the physical position of a chromosomal region harboring the *qFem3.1* and *M* loci. a.** A schematic diagram showing the physical positions of *qFem3.1* and DNA markers on a pseudomolecule, Chr3. **b.** Graphical genotypes and femaleness of self-pollinated progeny families from a monoecious selection in an S<sub>2</sub>BC<sub>2</sub>F<sub>1</sub> population produced by crosses between 03-009 and 03-336. Black boxes represent homozygous 03-336 segments, gray boxes indicate heterozygous 03-009/03-336 regions, and white boxes represent homozygous 03-009 segments. Marker positions and names are indicated above the boxes. Indices of femaleness (expressed as the percentage of female flowers per plant) are shown to the right of the boxes.

(PDF)

**S7 Fig. Graphical genotypes and femaleness of progeny families from monoecious selections in a S<sub>3</sub>BC<sub>2</sub>F<sub>1</sub> and a F<sub>2</sub> population produced by crosses between 03-009 and 03-336.**

Black boxes represent homozygous 03-336 segments, gray boxes indicate heterozygous 03-009/03-336 regions, and white boxes represent homozygous 03-009 segments. Markers are indicated above the boxes. Indices of femaleness (expressed as the percentage of female flowers

per plant) are shown to the right of the boxes.  
(PDF)

**S8 Fig. Expression levels of the genes located in the 19.5-kb chromosomal region candidate for the monoecious locus measured by RNA-seq.**  
(PDF)

**S9 Fig. Maximum likelihood tree of RADIALS-like genes.** *S. oleracea* (So), *Beta vulgaris* (Bv), *Arabidopsis thaliana* (At), *Diospyros kaki* (Dk), *Antirrhinum majus* (Am) and *Oryza sativa* (Os). Bootstrap values above 50% from 1,000 iterations are indicated below the nodes.  
(PDF)

**S10 Fig. Spatial expression analysis of *SoRL2a*.** a-c, Hematoxylin-eosin-stained sections of flower(s) from a male (a and b) and a female (c) plant. Two male flowers at different developmental stages are shown in panel a and numbered 1 and 2; a'-c', *In situ* hybridization of flower sections from a male (a' and b') and a female (c') plant with anti-sense *SoRL2a* riboprobes; a''-c'', Flower sections from a male (a'') and a female (b'' and c'') plant probed with sense *SoRL2a*. E, epidermis; En, endothecium; ML, middle layer; MMC, microspore mother cells; T, tapetum; Tds, tetrads; G, gynoecium; II, inner integument; OI, outer integument; Nu, nucellus; Ch, chalazal.  
(PDF)

**S11 Fig. Frequency distribution of the femaleness in the 96 F<sub>2</sub> progeny plants from the cross between 03-009 and 03-336.** Arrowheads and horizontal bars indicate average numbers and ranges, respectively, of femaleness of the parental lines, 03-009 and 03-336, the F<sub>1</sub> progeny, and NIL-M (S<sub>2</sub>BC<sub>5</sub>F<sub>1</sub>).  
(PDF)

**S12 Fig. Means of femaleness scores in the 03-009 x 03-336 F<sub>2</sub> population against genotypes at the QTLs.** A, homozygous for the 03-009-derived allele; H, heterozygous for the 03-009 and 03-336-derived alleles; B, homozygous for the 03-336-derived allele.  
(PDF)

**S13 Fig. Venn diagrams showing the genes differentially expressed between early-stage inflorescences from the spinach lines 03-336, NIL-M and 03-009.** 03-336, highly male monoecious; NIL-M, highly female monoecious; 03-009, female.  
(PDF)

**S1 Table. Status of genome assemblies for spinach line 03-009 in the five assembly steps.**  
(PDF)

**S2 Table. Hi-C mate-pair reads used in this study.**  
(PDF)

**S3 Table. Summary of whole genome sequencing reads used for variant discovery.**  
(PDF)

**S4 Table. Primer sequences and annealing temperatures of the spinach genetic markers used for fine mapping of the M-locus.**  
(PDF)

**S5 Table. Primer sequences and annealing temperatures of the spinach genetic markers flanking the QTLs controlling monoecious expression.**  
(PDF)

**S6 Table. Primers used for the RT-qPCR assay.**

(PDF)

**S7 Table. RNA-Seq reads used for transcriptome analysis.**

(PDF)

**S8 Table. A linkage map and pseudomolecules of spinach.**

(PDF)

**S9 Table. ANOVA table showing the genetic interaction between loci *qFem3.1/M* and *qFem2.1*.**

(PDF)

**S10 Table. BP GO terms enriched in 136 up-regulated DEGs shared between the comparison pairs 03–336 vs. 03–009 and NIL-M vs. 03–009.**

(PDF)

**S11 Table. BP GO terms enriched in 175 down-regulated-DEGs shared between the comparison pairs 03–336 vs. 03–009 and NIL-M vs. 03–009.**

(PDF)

**S12 Table. BP GO terms enriched in 836 up-regulated DEGs shared between the comparison pairs 03–336 vs. 03–009 and 03–336 vs. NIL-M.**

(PDF)

**S13 Table. BP GO terms enriched in 1005 down-regulated DEGs shared between the comparison pairs 03–336 vs. 03–009 and 03–336 vs. NIL-M.**

(PDF)

**S14 Table. Candidate genes underlying QTLs for the monoecious condition.**

(PDF)

## Acknowledgments

Computations were partially performed on the NIG supercomputer at the ROIS National Institute of Genetics. The authors thank the Tohoku Seed Co. (Utsunomiya, Tochigi, Japan) for providing the spinach breeding lines used in this study. We also appreciate the technical assistance provided by Mrs. H. Yokomoto.

## Author Contributions

**Conceptualization:** Hideki Hirakawa, Yasuyuki Onodera.

**Data curation:** Babil Pachakkil, Keisuke Tanaka, Yasuyuki Onodera.

**Formal analysis:** Kaoru Yamano, Akane Haseda, Keisuke Iwabuchi, Takayuki Osabe, Yuki Sudo, Babil Pachakkil, Keisuke Tanaka, Hideki Hirakawa, Yasuyuki Onodera.

**Funding acquisition:** Hideki Hirakawa, Yasuyuki Onodera.

**Investigation:** Kaoru Yamano, Akane Haseda, Keisuke Iwabuchi, Takayuki Osabe, Yuki Sudo, Yutaka Suzuki, Atsushi Toyoda, Hideki Hirakawa, Yasuyuki Onodera.

**Methodology:** Yasuyuki Onodera.

**Project administration:** Yasuyuki Onodera.

**Supervision:** Yasuyuki Onodera.

**Validation:** Hideki Hirakawa, Yasuyuki Onodera.

**Visualization:** Yasuyuki Onodera.

**Writing – original draft:** Keisuke Tanaka, Hideki Hirakawa, Yasuyuki Onodera.

**Writing – review & editing:** Yasuyuki Onodera.

## References

1. Morelock TE, Correll JC. Spinach. In: Tomas JP, Nuez F, editors. Vegetables I (Handbook of Plant Breeding). New York: Springer; 2008. p. 189–218.
2. FAOSTAT Statistics Database [Internet]. Food and Agriculture Organization of the United Nations. 2023. Available from: <https://www.fao.org/faostat/en/#home>.
3. Onodera Y, Yonaha I, Niikura S, Yamazaki S, Mikami T. Monoecy and gynodioecy in *Spinacia oleracea* L.: Morphological and genetic analyses. *Scientia Horticulturae*. 2008; 118(3):266–9. <https://doi.org/10.1016/j.scienta.2008.06.008> ISI:000259656800014.
4. Janick J. Hybrids in horticultural crops. In: Lamkey K, R., Staub J, E., editors. Concepts and breeding of heterosis in crop plants. Crop Science Society of America Special Publication: Crop Science Society of America; 1998. p. 45–56.
5. van der Vossen H, A., M. *Spinacia oleracea* L. In: Grubben G, J., H., Denton O, A., editors. PROTA 2: Vegetables/Légumes. Wageningen, Netherlands: PROTA; 2004.
6. Renner S, Ricklefs R. Dioecy and its correlates in the flowering plants. *American Journal of Botany*. 1995; 82(5):596–606. <https://doi.org/10.2307/2445418> WOS:A1995QY91500005.
7. Westergaard M. The mechanism of sex determination in dioecious flowering plants. *Advances in Genetics Incorporating Molecular Genetic Medicine*. 1958; 9:217–81. [https://doi.org/10.1016/s0065-2660\(08\)60163-7](https://doi.org/10.1016/s0065-2660(08)60163-7) WOS:A1958WY36900006. PMID: 13520443
8. Charlesworth D. Plant sex chromosome evolution. *Journal of Experimental Botany*. 2013; 64(2):405–20. <https://doi.org/10.1093/jxb/ers322> WOS:000313618900003. PMID: 23125359
9. Janick J, Stevenson E. A genetic study of the heterogametic nature of the staminate plant in spinach. *Proceedings of the American Society for Horticultural Science*. 1954; 63:444–6.
10. Janick J, Mahoney D, Pfahler P. The trisomics of *Spinacia oleracea*. *J Hered*. 1959; 50:46–50.
11. Ellis JR, Janick J. The Chromosomes of *Spinacia oleracea*. *American Journal of Botany*. 1960; 47(3):210–4.
12. Bose S, Janick J. Karyo-Races in *Spinacia oleracea*. *American Journal of Botany*. 1961; 48(3):238–41.
13. Kudoh T, Takahashi M, Osabe T, Toyoda A, Hirakawa H, Suzuki Y, et al. Molecular insights into the non-recombining nature of the spinach male-determining region. *Mol Genet Genomics*. 2018; 293(2):557–68. Epub 2017/12/08. <https://doi.org/10.1007/s00438-017-1405-2> PMID: 29222702.
14. Ma X, Yu L, Fatima M, Wadlington WH, Hulse-Kemp AM, Zhang X, et al. The spinach YY genome reveals sex chromosome evolution, domestication, and introgression history of the species. *Genome Biol*. 2022; 23(1):75. Epub 20220307. <https://doi.org/10.1186/s13059-022-02633-x> PMID: 35255946; PubMed Central PMCID: PMC8902716.
15. Onodera Y, Yonaha I, Masumo H, Tanaka A, Niikura S, Yamazaki S, et al. Mapping of the genes for dioecism and monoecism in *Spinacia oleracea* L.: evidence that both genes are closely linked. *Plant Cell Reports*. 2011; 30(6):965–71. <https://doi.org/10.1007/s00299-010-0998-2> WOS:000290542800002. PMID: 21301852
16. Takahata S, Yago T, Iwabuchi K, Hirakawa H, Suzuki Y, Onodera Y. Comparison of Spinach Sex Chromosomes with Sugar Beet Autosomes Reveals Extensive Synteny and Low Recombination at the Male-Determining Locus. *J Hered*. 2016; 107(7):679–85. Epub 2016/08/25. <https://doi.org/10.1093/jhered/esw055> PMID: 27563071.
17. Yamamoto K, Oda Y, Haseda A, Fujito S, Mikami T, Onodera Y. Molecular evidence that the genes for dioecism and monoecism in *Spinacia oleracea* L. are located at different loci in a chromosomal region. *Heredity (Edinb)*. 2014; 112(3):317–24. <https://doi.org/10.1038/hdy.2013.112> PMID: 24169648; PubMed Central PMCID: PMC3931174.
18. Janick J, Stevenson E. Genetics of the Monoecious Character in Spinach. *Genetics*. 1955; 40(4):429–37. <https://doi.org/10.1093/genetics/40.4.429> PMID: 17247563.
19. Simko I, Jia M, Venkatesh J, Kang B, Weng Y, Barcaccia G, et al. Genomics and Marker-Assisted Improvement of Vegetable Crops. *Critical Reviews in Plant Sciences*. 2021; 40(4):303–65. <https://doi.org/10.1080/07352689.2021.1941605>

20. Hirakawa H, Toyoda A, Itoh T, Suzuki Y, Nagano AJ, Sugiyama S, et al. A spinach genome assembly with remarkable completeness, and its use for rapid identification of candidate genes for agronomic traits. *DNA Res.* 2021; 28(3). <https://doi.org/10.1093/dnares/dsab004> PMID: 34142133; PubMed Central PMCID: PMC8231376.
21. Sassa H. A technique to isolate DNA from woody and herbaceous plants by using a silica-based plasmid extraction column. *Anal Biochem.* 2007; 363(1):166–7. <https://doi.org/10.1016/j.ab.2007.01.004> PMID: 17288984.
22. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 2017; 27(5):737–46. Epub 20170118. <https://doi.org/10.1101/gr.214270.116> PMID: 28100585; PubMed Central PMCID: PMC5411768.
23. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One.* 2014; 9(11):e112963. Epub 2014/11/19. <https://doi.org/10.1371/journal.pone.0112963> PMID: 25409509; PubMed Central PMCID: PMC4237348.
24. Sedlazeck FJ, Lee H, Darby CA, Schatz MC. Piercing the dark matter: bioinformatics of long-range sequencing and mapping. *Nat Rev Genet.* 2018; 19(6):329–46. <https://doi.org/10.1038/s41576-018-0003-4> PMID: 29599501.
25. Ghurye J, Pop M, Koren S, Bickhart D, Chin CS. Scaffolding of long read assemblies using long range contact information. *BMC Genomics.* 2017; 18(1):527. Epub 2017/07/12. <https://doi.org/10.1186/s12864-017-3879-z> PMID: 28701198; PubMed Central PMCID: PMC5508778.
26. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics.* 2011; 27(15):2156–8. Epub 2011/06/07. <https://doi.org/10.1093/bioinformatics/btr330> PMID: 21653522; PubMed Central PMCID: PMC3137218.
27. Cingolani P, Platts A, Wang IL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012; 6(2):80–92. <https://doi.org/10.4161/fly.19695> PMID: 22728672; PubMed Central PMCID: PMC3679285.
28. Rastas P. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics.* 2017; 33(23):3726–32. <https://doi.org/10.1093/bioinformatics/btx494> PMID: 29036272.
29. Broman KW, Wu H, Sen S, Churchill GA. R/qtl: QTL mapping in experimental crosses. *Bioinformatics.* 2003; 19(7):889–90. <https://doi.org/10.1093/bioinformatics/btg112> PMID: 12724300.
30. Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity (Edinb).* 1992; 69(4):315–24. <https://doi.org/10.1038/hdy.1992.131> PMID: 16718932.
31. Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* 2016; 3(1):95–8. <https://doi.org/10.1016/j.cels.2016.07.002> PMID: 27467249; PubMed Central PMCID: PMC5846465.
32. Keilwagen J, Hartung F, Grau J. GeMoMa: Homology-Based Gene Prediction Utilizing Intron Position Conservation and RNA-seq Data. *Methods Mol Biol.* 2019; 1962:161–77. [https://doi.org/10.1007/978-1-4939-9173-0\\_9](https://doi.org/10.1007/978-1-4939-9173-0_9) PMID: 31020559.
33. Manni M, Berkeley MR, Seppely M, Zdobnov EM. BUSCO: Assessing Genomic Data Quality and Beyond. *Curr Protoc.* 2021; 1(12):e323. <https://doi.org/10.1002/cpz1.323> PMID: 34936221.
34. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 2012; 9(4):357–9. Epub 2012/03/04. <https://doi.org/10.1038/nmeth.1923> PMID: 22388286; PubMed Central PMCID: PMC3322381.
35. Neff MM, Turk E, Kalishman M. Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet.* 2002; 18(12):613–5. [https://doi.org/10.1016/s0168-9525\(02\)02820-2](https://doi.org/10.1016/s0168-9525(02)02820-2) PMID: 12446140.
36. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015; 12(4):357–60. Epub 20150309. <https://doi.org/10.1038/nmeth.3317> PMID: 25751142; PubMed Central PMCID: PMC4655817.
37. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015; 33(3):290–5. Epub 20150218. <https://doi.org/10.1038/nbt.3122> PMID: 25690850; PubMed Central PMCID: PMC4643835.
38. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc.* 2016; 11(9):1650–67. Epub 20160811. <https://doi.org/10.1038/nprot.2016.095> PMID: 27560171; PubMed Central PMCID: PMC5032908.
39. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. *Mol Biol*

- Evol. 2021; 38(12):5825–9. <https://doi.org/10.1093/molbev/msab293> PMID: 34597405; PubMed Central PMCID: PMC8662613.
40. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*. 2006; 22(13):1600–7. Epub 20060410. <https://doi.org/10.1093/bioinformatics/btl140> PMID: 16606683.
  41. Jauhal AA, Newcomb RD. Assessing genome assembly quality prior to downstream analysis: N50 versus BUSCO. *Mol Ecol Resour*. 2021; 21(5):1416–21. Epub 20210309. <https://doi.org/10.1111/1755-0998.13364> PMID: 33629477.
  42. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res*. 2018; 46(21):e126. <https://doi.org/10.1093/nar/gky730> PMID: 30107434; PubMed Central PMCID: PMC6265445.
  43. Janick J, Stevenson E. Environmental influences on sex expression in monoecious lines of spinach. *Proceedings of the American Society for Horticultural Science*. 1955; 65:416–22.
  44. Tsugeki R, Tanaka-Sato N, Maruyama N, Terada S, Kojima M, Sakakibara H, et al. CLUMSY VEIN, the Arabidopsis DEAH-box Prp16 ortholog, is required for auxin-mediated development. *Plant J*. 2015; 81(2):183–97. Epub 20141216. <https://doi.org/10.1111/tpj.12721> PMID: 25384462.
  45. Tsugeki R, Terada S. The Arabidopsis ortholog of the DEAH-box ATPase Prp16 influences auxin-mediated development. *Plant Signal Behav*. 2015; 10(10):e1074369. <https://doi.org/10.1080/15592324.2015.1074369> PMID: 26237376; PubMed Central PMCID: PMC4883861.
  46. Murmu J, Bush MJ, DeLong C, Li S, Xu M, Khan M, et al. Arabidopsis basic leucine-zipper transcription factors TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are redundantly required for anther development. *Plant Physiol*. 2010; 154(3):1492–504. Epub 20100830. <https://doi.org/10.1104/pp.110.159111> PMID: 20805327; PubMed Central PMCID: PMC2971623.
  47. Cui Y, Hu C, Zhu Y, Cheng K, Li X, Wei Z, et al. ClK Receptor Kinases Determine Cell Fate Specification during Early Anther Development in Arabidopsis. *Plant Cell*. 2018; 30(10):2383–401. Epub 20180910. <https://doi.org/10.1105/tpc.17.00586> PMID: 30201822; PubMed Central PMCID: PMC6241272.
  48. Clark SE, Williams RW, Meyerowitz EM. The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in Arabidopsis. *Cell*. 1997; 89(4):575–85. [https://doi.org/10.1016/s0092-8674\(00\)80239-1](https://doi.org/10.1016/s0092-8674(00)80239-1) PMID: 9160749.
  49. Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet*. 2017; 49(4):643–50. Epub 2017/03/06. <https://doi.org/10.1038/ng.3802> PMID: 28263316; PubMed Central PMCID: PMC5909822.
  50. Cai X, Sun X, Xu C, Sun H, Wang X, Ge C, et al. Genomic analyses provide insights into spinach domestication and the genetic basis of agronomic traits. *Nat Commun*. 2021; 12(1):7246. Epub 20211213. <https://doi.org/10.1038/s41467-021-27432-z> PMID: 34903739; PubMed Central PMCID: PMC8668906.
  51. Schneeberger K, Hagmann J, Ossowski S, Warthmann N, Gesing S, Kohlbacher O, et al. Simultaneous alignment of short reads against multiple genomes. *Genome Biol*. 2009; 10(9):R98. Epub 20090917. <https://doi.org/10.1186/gb-2009-10-9-r98> PMID: 19761611; PubMed Central PMCID: PMC2768987.
  52. Corley SB, Carpenter R, Copsey L, Coen E. Floral asymmetry involves an interplay between TCP and MYB transcription factors in *Antirrhinum*. *Proc Natl Acad Sci U S A*. 2005; 102(14):5068–73. Epub 20050324. <https://doi.org/10.1073/pnas.0501340102> PMID: 15790677; PubMed Central PMCID: PMC555980.
  53. Machemer K, Shaiman O, Salts Y, Shabtai S, Sobolev I, Belausov E, et al. Interplay of MYB factors in differential cell expansion, and consequences for tomato fruit development. *Plant J*. 2011; 68(2):337–50. Epub 20110804. <https://doi.org/10.1111/j.1365-313X.2011.04690.x> PMID: 21707804.
  54. Hamaguchi A, Yamashino T, Koizumi N, Kiba T, Kojima M, Sakakibara H, et al. A small subfamily of Arabidopsis *RADIALIS-LIKE SANT/MYB* genes: a link to HOOKLESS1-mediated signal transduction during early morphogenesis. *Bioscience, Biotechnology, and Biochemistry*. 2008; 72(10):2687–2696.
  55. Yang B, Song Z, Li C, Jiang J, Zhou Y, Wang R, et al. RSM1, an Arabidopsis MYB protein, interacts with HY5/HYH to modulate seed germination and seedling development in response to abscisic acid and salinity. *PLoS Genet*. 2018; 14(12):e1007839. Epub 20181219. <https://doi.org/10.1371/journal.pgen.1007839> PMID: 30566447; PubMed Central PMCID: PMC6317822.
  56. Masuda K, Ikeda Y, Matsuura T, Kawakatsu T, Tao R, Kubo Y, et al. Reinvention of hermaphroditism via activation of a *RADIALIS*-like gene in hexaploid persimmon. *Nat Plants*. 2022; 8(3):217–24. Epub 20220317. <https://doi.org/10.1038/s41477-022-01107-z> PMID: 35301445.
  57. Sherry RA, Eckard KJ, Lord EM. Flower development in dioecious *Spinacia oleracea* (Chenopodiaceae). *American Journal of Botany*. 1993:283–91. ISI:A1993KT07300006.

58. Sather DN, Jovanovic M, Golenberg EM. Functional analysis of B and C class floral organ genes in spinach demonstrates their role in sexual dimorphism. *Bmc Plant Biology*. 2010; 10. <https://doi.org/10.1186/1471-2229-10-46> ISI:000278022200001. PMID: 20226063
59. Aryal R, Jagadeeswaran G, Zheng Y, Yu Q, Sunkar R, Ming R. Sex specific expression and distribution of small RNAs in papaya. *BMC Genomics*. 2014; 15(1):20. Epub 20140113. <https://doi.org/10.1186/1471-2164-15-20> PMID: 24410969; PubMed Central PMCID: PMC3916515.
60. Akagi T, Henry IM, Tao R, Comai L. Plant genetics. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science*. 2014; 346(6209):646–50. Epub 2014/10/30. <https://doi.org/10.1126/science.1257225> PMID: 25359977.
61. Akagi T, Henry IM, Kawai T, Comai L, Tao R. Epigenetic Regulation of the Sex Determination Gene *MeG1* in Polyploid Persimmon. *Plant Cell*. 2016; 28(12):2905–15. Epub 2016/12/12. <https://doi.org/10.1105/tpc.16.00532> PMID: 27956470; PubMed Central PMCID: PMC5240738.
62. Xue L, Wu H, Chen Y, Li X, Hou J, Lu J, et al. Evidences for a role of two Y-specific genes in sex determination in *Populus deltoides*. *Nat Commun*. 2020; 11(1):5893. Epub 20201118. <https://doi.org/10.1038/s41467-020-19559-2> PMID: 33208755; PubMed Central PMCID: PMC7674411.
63. Chailakhyan MK, Khryanin VN. Effect of growth regulators and role of roots in sex expression in spinach. *Planta*. 1978; 142(2):207–10. <https://doi.org/10.1007/BF00388214> PMID: 24408104.
64. West NW, Golenberg EM. Gender-specific expression of *GIBBERELLIC ACID INSENSITIVE* is critical for unisexual organ initiation in dioecious Spinacia oleracea. *New Phytol*. 2018; 217(3):1322–34. Epub 20171211. <https://doi.org/10.1111/nph.14919> PMID: 29226967.
65. Qiao Y, Shi J, Zhai Y, Hou Y, Ma W. Phytophthora effector targets a novel component of small RNA pathway in plants to promote infection. *Proc Natl Acad Sci U S A*. 2015; 112(18):5850–5. Epub 20150420. <https://doi.org/10.1073/pnas.1421475112> PMID: 25902521; PubMed Central PMCID: PMC4426467.
66. Tognetti VB, Van Aken O, Morreel K, Vandenbroucke K, van de Cotte B, De Clercq I, et al. Perturbation of indole-3-butyric acid homeostasis by the UDP-glucosyltransferase *UGT74E2* modulates *Arabidopsis* architecture and water stress tolerance. *Plant Cell*. 2010; 22(8):2660–79. Epub 20100826. <https://doi.org/10.1105/tpc.109.071316> PMID: 20798329; PubMed Central PMCID: PMC2947170.
67. Cucinotta M, Cavalleri A, Chandler JW, Colombo L. Auxin and Flower Development: A Blossoming Field. *Cold Spring Harb Perspect Biol*. 2021; 13(2). Epub 20210201. <https://doi.org/10.1101/cshperspect.a039974> PMID: 33355218; PubMed Central PMCID: PMC7849340.
68. Nagpal P, Ellis CM, Weber H, Ploense SE, Barkawi LS, Guilfoyle TJ, et al. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development*. 2005; 132(18):4107–18. Epub 20050817. <https://doi.org/10.1242/dev.01955> PMID: 16107481.
69. Sessions RA, Zambryski PC. *Arabidopsis* gynoceium structure in the wild and in ettin mutants. *Development*. 1995; 121(5):1519–32. <https://doi.org/10.1242/dev.121.5.1519> PMID: 7789281.
70. Prunet N, Yang W, Das P, Meyerowitz EM, Jack TP. *SUPERMAN* prevents class B gene expression and promotes stem cell termination in the fourth whorl of *Arabidopsis thaliana* flowers. *Proc Natl Acad Sci U S A*. 2017; 114(27):7166–71. Epub 20170620. <https://doi.org/10.1073/pnas.1705977114> PMID: 28634297; PubMed Central PMCID: PMC5502645.
71. Xu Y, Prunet N, Gan ES, Wang Y, Stewart D, Wellmer F, et al. *SUPERMAN* regulates floral whorl boundaries through control of auxin biosynthesis. *EMBO J*. 2018; 37(11). Epub 20180515. <https://doi.org/10.15252/embj.201797499> PMID: 29764982; PubMed Central PMCID: PMC5983216.
72. Liu J, Chen LY, Zhou P, Liao Z, Lin H, Yu Q, et al. Sex biased expression of hormone related genes at early stage of sex differentiation in papaya flowers. *Hortic Res*. 2021; 8(1):147. Epub 20210701. <https://doi.org/10.1038/s41438-021-00581-4> PMID: 34193826; PubMed Central PMCID: PMC8245580.
73. Heslop-Harrison J. Auxin and Sexuality in *Cannabis sativa*. *Physiologia Plantarum*. 1956; 9(4):Physiologia Plantarum.
74. Carrasco D, Bellido A, Vargas A, Picq S, Bacilieri R, This P, et al. Comparative analysis of the expression of sex candidate genes in flower of dioecious and hermaphrodite grapevine (*Vitis vinifera* L. ssp.). *Scientia Horticulturae*. 2020; 274.
75. Ito Y, Nakanomyo I, Motose H, Iwamoto K, Sawa S, Dohmae N, et al. Dodeca-CLE peptides as suppressors of plant stem cell differentiation. *Science*. 2006; 313(5788):842–5. <https://doi.org/10.1126/science.1128436> PMID: 16902140.
76. Uemura A, Yamaguchi N, Xu Y, Wee W, Ichihashi Y, Suzuki T, et al. Regulation of floral meristem activity through the interaction of *AGAMOUS*, *SUPERMAN*, and *CLAVATA3* in *Arabidopsis*. *Plant Reprod*. 2018; 31(1):89–105. Epub 20171207. <https://doi.org/10.1007/s00497-017-0315-0> PMID: 29218596.

77. Kazama Y, Kito M, Kobayashi T, Ishii K, Krasovec M, Yasui Y, et al. A *CLAVATA3*-like Gene Acts as a Gynoecium Suppression Function in White Campion. *Mol Biol Evol.* 2022; 39(10). <https://doi.org/10.1093/molbev/msac195> PMID: 36166820; PubMed Central PMCID: PMC9550985.
78. Willmann MR. CLV1 and CLV3: negative regulators of SAM stem cell accumulation. *Trends Plant Sci.* 2000; 5(10):416. [https://doi.org/10.1016/s1360-1385\(00\)01763-5](https://doi.org/10.1016/s1360-1385(00)01763-5) PMID: 11044716.
79. Kinoshita A, Betsuyaku S, Osakabe Y, Mizuno S, Nagawa S, Stahl Y, et al. RPK2 is an essential receptor-like kinase that transmits the CLV3 signal in *Arabidopsis*. *Development.* 2010; 137(22):3911–20. <https://doi.org/10.1242/dev.048199> PMID: 20978082.