A MYB transcription factor gene involved in sex determination in *Asparagus officinalis*

Short title: Sex determination gene in asparagus

Authors:

Kohji Murase[1]*, Shuji Shigenobu[2], Sota Fujii[1], Kazuki Ueda[1], Takanori Murata[1], Ai Sakamoto[1], Yuko Wada[1], Katsushi Yamaguchi[2], Yuriko Osakabe[3], Keishi Osakabe[3], Akira Kanno[4], Yukio Ozaki[5], Seiji Takayama[1]* [a]

Affiliations:

[1]Graduate School of Biological Sciences, Nara Institute of Science and Technology, Ikoma, Nara 630-0192, Japan

[2]National Institute for Basic Biology (NIBB) Core Research Facilities, NIBB, Okazaki, Aichi 444-8585, Japan

[3]Center for Collaboration among Agriculture, Industry and Commerce, The University of Tokushima, 3-18-15 Kuramoto-cho, Tokushima 770-8503, Japan

[4]Graduate School of Life Sciences, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan

[5]Faculty of Agriculture, Kyushu University, Kasuya, Fukuoka 811-2307, Japan

*Correspondence to: E-mail: kmurase@is.naist.jp or takayama@bs.naist.jp

[a]Present address: Department of Applied Biological Chemistry, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

Key words: Plant, Dioecy, Sex determination gene, Asparagus

Abstract:

Dioecy is a plant mating system in which individuals of a species are either male or female. Although many flowering plants evolved independently from hermaphroditism to dioecy, the molecular mechanism underlying this transition remains largely unknown. Sex determination in the dioecious plant *Asparagus officinalis* is controlled by X and Y chromosomes; the male and female karyotypes are XY and XX, respectively. Transcriptome analysis of *A. officinalis* buds revealed that a *MYB*-like gene, *Male Specific Expression 1* (*MSE1*), is specifically expressed in males. *MSE1* exhibits tight linkage with the Y chromosome, specific expression in early anther development, and loss of function on the X chromosome. Knockout of the *MSE1* orthologue in *Arabidopsis* induces male sterility. Thus, *MSE1* acts in sex determination in *A. officinalis*.

Introduction:

In order to preserve genetic variety within species, flowering plants have evolved various systems to prevent self-fertilization. In one such system, dioecy, individuals of a species are either male or female. In angiosperms, about 15,000 species (~6%) of 160 families are dioecious, and the evolution of dioecy is thought to have occurred independently more than 800 times (Charlesworth 2002; Renner 2014). According to a current theoretical model, the transition from hermaphroditism to dioecy can proceed by two evolutionary pathways: gynodioecy, in which individual plants separate into females and hermaphrodites, and androdioecy, in which plants separate into males and hermaphrodites (Charlesworth & Charlesworth 1978). The first step of the evolution of gynodioecy is a recessive male sterile mutation, followed by a dominant female sterile mutation (or gain of suppressor function) near the male mutation locus, thus creating a sex chromosome. Conversely, a female mutation is the first step in the evolution of androdioecy, but this pathway is not predominant because female mutations and androdioecious plants are very rare in nature. However, molecular mechanisms of sex determination and its evolution in flowering plants are largely unknown. The recent identification of sex-determination genes in persimmon is the only example that an autosomal homeobox transcription factor gene, *MeGI*, dominantly suppresses male organ development, whereas *OGI* on the Y chromosome encodes a small RNA that targets *MeGI* for gene silencing (Akagi *et al.* 2014).

Sex determination in the dioecious plant *Asparagus officinalis* is controlled by a single locus, the *Mating-* (*M*) locus, located on chromosome 5 (Löptien 1979; Telgmann-Rauber *et al.* 2007). The sex chromosome karyotypes of males and females are XY and XX,

respectively. Relatively large vestiges of organs corresponding to the opposite sex are observed in both male and female flowers in *A. officinalis*, suggesting that morphological sex differentiation occurs at a late stage of flower development (Fig. 1A,B). In fact, at early developmental stages, the male and female flowers look like those of hermaphrodites; the morphological differences between male and female flowers appear later when the stylar tube is formed on carpels in female flowers, and during or just before meiosis in male flowers (Caporali *et al.* 1994). To generate these morphological differences, sex-determination genes must be expressed in the appropriate tissues at the appropriate developmental stages. Genetic analysis suggested the involvement of two sex-determination genes, called "*male activator*" and "*female suppressor*", located in the *M*-locus of the Y chromosome (Marks 1973).

Results and Discussion:

To search for sex-determination genes, a transcriptome analysis was performed during early development of male and female flowers of *A. officinalis* cv. Super Welcome. *De novo* assembly of 10.5 Gb of male paired-end sequences by Trinity (Grabherr *et al.* 2011) yielded 104,937 contigs and 51,525 unigenes (Fig. S1, Table S1). Mapping of 52.8 and 56.6 million reads from males and females against the assembled contigs, respectively, revealed that 149 contigs (114 unigenes) are expressed in a male-biased manner. Because the previous transcriptome analysis of *A. officinalis* failed to identify Y-chromosome genes (Harkess *et al.* 2015), we performed further screening by mapping each of 316 million reads from the male and female genome sequencings against the 114 candidate genes. Ultimately, seven

contigs were obtained as candidates for male-specific genes (Table S2). To confirm that the candidate genes were male-specific, we amplified them by polymerase chain reaction (PCR) from bulked male and female genomes (Fig. S2). Only one gene, which we named *Male Specific Expression 1* (*MSE1*), exhibited male-specific amplification. This result was unexpected because both female RNA and genome sequence reads mapped to the *MSE1* contig, albeit at low levels (Table S2). Mapping of RNA sequence reads against the *MSE1* contig revealed that female reads only mapped to the 5' end of the transcript, whereas male reads covered the whole transcript (Fig. S3). Reverse transcription (RT)-PCR using primers that amplified the full-length *MSE1* transcript confirmed male-specific expression of *MSE1* (Fig. 1C).

To determine whether *MSE1* is on the Y chromosome, we PCR-amplified *MSE1* from the genomes of male and female individuals of *A. officinalis* cv. Super Welcome. *MSE1* specifically amplified from male individuals, but not from females (Fig. 1D). Subsequent PCR analysis of 112 independent plants confirmed male-specific amplification of *MSE1* (Fig. 4S). These results suggest that *MSE1* is on the Y chromosome gene, tightly linked to the *M*-locus.

MSE1 encodes a 276-amino acid protein containing two MYB domains at the N-terminus (Fig. 1E, Fig. S5). MSE1 belongs to the R2R3-MYB class of proteins, which includes MYB transcription factors involved in metabolism, cell fate and identity, development, and biotic and abiotic stress responses (Stracke *et al.* 2001; Dubos *et al.* 2010; Ambawat *et al.* 2013). To study the spatial and temporal pattern of *MSE1* expression, we measured the levels of *MSE1* mRNA in each tissue of male plants by quantitative RT-PCR. *MSE1* mRNA

was specifically expressed in small buds, but not in other tissues (Fig. 1F). Detailed analysis of *MSE1* expression in young buds revealed that *MSE1* was predominantly expressed in anther (Fig. 1G). These results suggest that *MSE1* acts in early stages of male organ development.

RNA and genome-sequence mapping data of *MSE1* transcripts suggested that the vestige of *MSE1* still exists on the X chromosome (Table S2). To test this hypothesis, whole-genome sequencing and assembly was performed on a male genome. BLAST searches against the assembly scaffolds revealed four scaffolds with high sequence similarity to *MSE1* cDNA. One completely matched the genomic sequence of *MSE1* in which male DNA sequence reads were specifically mapped, and was judged to represent the *MSE1* sequence on the Y chromosome (Fig. 1H, Fig. S6). The other three scaffolds shared partial similarity with the *MSE1* genome sequence on the Y chromosome (Fig. 1H). The high conservation of intergenic regions and introns of *MSE1* between these scaffolds, and the fact that they could be amplified from both male and female genomes, suggested that these three scaffolds represented X chromosome sequences. We designated *MSE1* on Y chromosome as $MSE1^{Y}$ and the putative *MSE1* sequence on the X chromosome as $MSE1^{X}$. Two scaffolds of $MSE1^{X}$ were assembled with a 19 kb PacBio sequence showing that $MSE1^{X}$ is fragmented in at least 30 kb region of X chromosome rather than $MSE1^{Y}$ is encoded within 2.5 kb (Fig. 1H). This result could explain the misamplification of *MSE1* genome fragment by PCR from female individuals because the amplicon is too large (Fig. 1D). In *A. officinalis* cv. Super Welcome, three insertions, five deletions, and 16 point mutations are present in the coding region of $MSE1^{X}$ relative to $MSE1^{Y}$ (Fig. S7). Some of these are likely to be deleterious

mutations: a one-base deletion at tyrosine 28 of MSE1^{Y} induces a frame shift, resulting in a premature stop codon; a deletion at the end of the second exon causes the loss of 26 bases of protein-coding region and a splicing signal; and a large deletion at the end of third exon also causes a 200 bp deletion of protein-coding region (Fig. 1H, Figs. S7,8). Various mutations were observed among *A. officinalis* cultivars in *MSE1*^{X}, but no SNPs were detected in *MSE1*^{Y}, suggesting that *MSE1*^{X} is no longer under selection pressure to maintain its function (Fig. S7). These results suggest that loss of function of *MSE1* has occurred on the X chromosome.

The *Asparagus* genus contains up to 300 species distributed widely around the world (Kubitzki & Rudall 1998). Phylogenetic analysis of these species revealed that the dioecious species form a single clade, suggesting that the evolutionary event leading from hermaphroditism to dioecy in *Asparagus* occurred only once (Fig. S9) (Kubota *et al.* 2012). Therefore, if *MSE1* acts as "*male activator*" in sex determination during male organ development, the system is likely to be conserved in dioecious *Asparagus* species. To test whether the *MSE1* system is conserved in dioecious *Asparagus* species, PCR amplification and sequencing of *MSE1*^{Y} and *MSE1*^{X} from genomic DNA of the male and female individuals were performed. *MSE1* genes could be amplified from the genomes of all male individuals, but not those from female individuals in three cultivars of *A. officinalis*, *A. pseudoscaber*, *A. kiusianus*, *A. schoberioides*, and *A. verticillatus* (Fig. 2A). Three conserved deleterious mutations, which are caused by the frameshift mutations, were observed in *MSE1*^{X} sequences from the female individuals of these related species (Fig. S7). Furthermore, *MSE1* could also be amplified from all tested hermaphroditic species, and the

coding protein sequences were highly conserved among dioecious and hermaphroditic species (Fig. 2B, Fig. S10). These results suggest that the arrest of male organ development in female flowers in these *Asparagus* species is caused by loss of *MSE1* function. Interestingly, *A. acutifolius*, *A. stipularis*, and *A. cochinchinensis*, which are phylogenetically most distant dioecious species from *A. officinalis*, have no deleterious mutation in *MSE1* coding regions (Fig. 2A, Fig. S11). This result suggests that the origin of male mutation have occurred in outside of *MSE1* coding region or these three species have evolved in independent pathway.

If *MSE1* mutation is responsible for the transition from hermaphroditism to dioecy, artificial mutation of *MSE1* orthologues in other plant species should convert hermaphrodites into female plants. Phylogenetic analysis of MSE1-like MYB transcription factors revealed that *MSE1* orthologues are widely conserved in monocot and dicot species, including the model plant *Arabidopsis thaliana* (Fig. S12). The second most similar MYB protein in *A. thaliana*, AtMYB103, is outside the MSE1 clade, suggesting that ancestral *MSE1* and *AtMYB103* branched before the monocot-dicot divergence (Fig. S12). The conservation of *MSE1* is assumed to reflect the functional importance of this gene in the life cycle of flowering plants. The *MSE1* orthologue of *A. thaliana* is *AtMYB35*/*TDF1* (*Tapetal Development and Function 1*), which is essential for normal anther development (Fig. S12) (Zhu *et al.* 2008). Because a T-DNA insertion line was not available, genome editing knockouts of *TDF1* were produced using the CRISPR/Cas9 system targeting three sites in the *TDF1* gene (Fig. 3A,B). Each transformant exhibited normal vegetative growth and flowering, but seedless siliques (Fig. 3C to F). No pollen grains were observed in the

transformants, suggesting that the sterility is caused by a defect in male organ development (Fig. 3G,H). These features are consistent with the previously reported phenotype of the *tdf1* mutant (Zhu *et al.* 2008). These results support the idea that *MSE1* functions in male organ development.

Charles Darwin considered male (or female) organ abortion to be the first step in the evolution of dioecy (Darwin 1877). Our results strongly suggest that *MSE1* acts as the "*male activator*" in *A. officinalis* sex determination, and that the loss-of-function mutation in $MSE1^X$ was an important step in the evolution of dioecy in *Asparagus*. Although gynodioecious (or androdioecious) *Asparagus* species have not been identified to date, our data may provide the first molecular evidence that this species evolved via the gynodioecy (or androdioecy) by the mutation of gene involved in male organ development pathway (Charlesworth & Charlesworth 1978). The *MSE1* system is clearly distinct from the *OGI–MeGI* system involved in persimmon sex determination, in which a small RNA acts as the sex determination factor (Akagi *et al.* 2014). Divergent molecular mechanisms have been described in the self-incompatibility system, which is also involved in preventing self-fertility in flowering plants (Takayama & Isogai 2005). It will be interesting to compare the molecular mechanisms and evolution of these systems. Our results will contribute to the understanding of the molecular mechanisms of sex determination, as well as the evolution of dioecy from hermaphroditism in flowering plants.

Experimental procedures:

Plant materials

A. officinalis cv. Super Welcome and Pole Tom were purchased from SAKATA SEED (Yokohama), cv. Mary Washington from Takii Seed (Kyoto); cv. Niagara from Nitto Nousan (Yokohama); and *A. pseudoscaber*, *A. verticillatus*, *A. densiflorus*, and *A. virgatus* from B & T World Seeds (Aigues-Vives). *A. plumosus* and *A. asparagoides* were purchased from public garden centers. *A. kiusianus* was collected from Keyakaigan (Fukuoka, Japan). *A. schoberioides*, *A. acutifolius*, *A. stipularis*, and *A. cochinchinensis* are described in Kubota *et al.* 2012.

Transcriptome analysis

Total RNAs were extracted from early developmental buds (0.5–0.8 mm) of male and female *A. officinalis* cv. Super Welcome using the RNeasy Plant Mini Kit (Qiagen). Library preparations and RNA sequencing by HiSeq 2000 (Illumina) were outsourced to Hokkaido System Science (Hokkaido). Adapter sequences were removed from raw sequence data using the *cutadapt* program (Martin 2011). Next, four bases of 3'-terminal sequences of the treated sequences were removed using FastX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). *De novo* assembly of RNA sequences from males was performed using Trinity with default parameters except that minimum k-mer coverage was set to 3 (Grabherr *et al.* 2011). Mapping of RNA and genome sequences from males and females were conducted using *bowtie* (Langmead *et al.* 2009). Mapping data were processed by SAMtools (Li *et al.* 2009) and visualized by the Integrative Genomics Viewer (IGV) software (Robinson *et al.* 2011). Gene functions were annotated by BLAST2GO (Conesa *et al*. 2005).

RT-PCR and real-time quantitative RT-PCR

Total RNA was extracted from each tissue as described above. Reverse transcription was performed using SuperScript III (Invitrogen). *MSE1* cDNA was PCR amplified using *Ex Taq* (Takara) with primers 646 (5'-GATCGGATCCATGGGCAGGCCTCCATGCTGCGA-3') and 647 (5'-GATCGAATTCCTACAGCAAATCATAAAAAAACTCAGG-3'), which were designed to amplify full-length *MSE1*. Real-time quantitative RT-PCR was carried out on a LightCycler 96 system (Roche) using QuantiFast SYBR Green RT-PCR Kit (Qiagen). *MSE1* and *Actin* were amplified by primers AoMSE1realtimeFw (5'-GCCCTAATTTGAAGCATGAGAG-3')/AoMSE1realtimeRv (5'-GATTTGAGAGATGGGTTGTG-3') and AoActin1F (5'-GTTCCTGCTCATAATCTAGAGCAAC-3')/AoActin1R (5'-CTTCTCACTGAGGCTCCACTCAAC-3'), respectively. *MSE1* expression was normalized against expression of *Actin*.

Linkage analysis and amplification of *MSE1* from genus *Asparagus*

For linkage analysis of *MSE1*, seeds of *A. officinalis* cv. Super Welcome were treated with *n*-propyl *N*-(3,4-dichlorophenyl) carbamate (NPC) to induce early flowering as described previously (Aneja *et al.* 1999). Genomes of each individual were extracted by Plant DNAzol Reagent (Invitrogen). *MSE1* fragments were amplified using *ExTaq* with primers 706 (5'-TGGTCGGTAATCGCACATCACCTCC-3') and 647. To sequence the coding

region of *MSE1* from other *A. officinalis* cultivars and *Asparagus* species, primers U10 (5’-AATTGGTTTCATCATCATTGTACCTCAG-3’) and U21 (5’-CTAAGATCCCAACGCACAAAC-3’) were used for PCR amplification. Sequencing was performed using BigDye Terminator v3.1 (Applied Biosystems). To amplify *MSE1* fragments from genomes of each individual of *A. officinalis* and related species, we used primer sets 646–647 or 646–648 (5’-GATCGAATTCCTAGGCTAGAGTGGTGATGGTTTCCTTG-3’). For amplification of $MSE1^{Y}$ from *A. kiusianus*, 706 and 647 primers were used. For male-specific marker, the Asp1-T7 primer set was used as described previously (Jamsari *et al.* 2004). To check the sex genotypes of *A. verticillatus* individuals, newly developed male-specific marker, designated MSM1 (male-specific marker 1), was created from male-specific scaffold in the genome assembly. For amplification of MSM1, 814 (5’-CAACTCCAGGTGACAACATTCATAG-3’) and 805 (5’-TCGTCAACGTCGACTGCAGGTAGGC-3’) primers were used. $MSE1^{X}$ fragments were amplified by primers 752 (5’-ATTGGTTTCATCATCATTGTACCTC-3’) and 754 (5’-TTGCCTGTCCATCTCACTTCTGGAT-3’) for the first and second exons, and 755 (5’-CTAACCATGATCTACACACGATCAC-3’) and 757 (5’-CCCTTCGACGTGGATTAATCGCTACC-3’) for the third exon.

Phylogenetic analysis

Protein sequences of the MYB transcription factors were multiply aligned using Clustal Omega (Sievers *et al.* 2011), with the Myb_DNA-binding domain HMM matrix (accession

No. PF00249 under Pfam database (Finn *et al.* 2014) used as the external profile HMM. Conserved selection blocks from the alignment were selected using Gblocks (Talavera & Castresana 2007) with default parameters. Phylogenetic tree was constructed based on Bayesian inference by using the MrBayes 3.2.2 program (Ronquist *et al.* 2012), using HsMYB as the outgroup sequence. Four chains of the Metropolis-coupled Markov Chain Monte Carlo processes were run for 1,000,000 generations, with trees sampled every 1,000 generations. The first 25% of trees were discarded, and the remaining trees were used to support the majority-rule consensus tree topology with posterior probabilities. For the *Asparagus* genus taxon phylogeny, five chloroplast intergenic sequences (Kubota *et al.* 2012) were aligned using MAFFT (Katoh & Standley 2013) and concatenated. The alignment was cleaned by Gblocks and subjected to MrBayes 3.2.2 analysis as described above, using the *C. stricta* sequence as outgroup, to yield the consensus tree.

Genome sequencing and assembly

For whole-genome assembly, DNA from a single male asparagus was used for genome sequencing. For screening of male-specific genes in transcriptome analysis, the female genome was also used. Genome DNA was extracted using the DNeasy Plant Mini Kit (Qiagen). The purified DNA was fragmented on a Covaris S2 sonicator (Covaris, Woburn, MA), size-selected with Pippin Prep (Sage Science, Beverly, MA), and then used to create two libraries using the TruSeq DNA PCR-Free Sample Preparation Kit (Illumina, San Diego, CA) with insert sizes of ~180 and ~800 bp. These libraries were sequenced on the Illumina HiSeq platform using a 2x 101-nt paired-end sequencing protocol. The reads were

cleaned up with *cutadapt*. Low-quality ends (<QV20) and adapter sequences were trimmed, and reads shorter than 50 bp were discarded. Total sequence of 105.3 Gb (~84× coverage of the genome, assuming a genome size of 1.26 Gb) was generated from the libraries, and then assembled using the ALLPATHS-LG assembler (Gnerre *et al.* 2011). The assembly yielded 146,894 scaffolds with an N50 length of 5.2 kb. For transcriptome analysis, 316 million reads of male and female genome sequences were used for mapping. Long read sequences were generated by PacBio RS II sequencer (Pacific Biosciences, Menlo Park, CA) with a 20 kb DNA library prepared from the female asparagus. Total sequence of 3.72 Gb in 372,292 reads was obtained from 8 SMRT cells. The N50 length was 13,054 bp. Sequences containing *MSE1* locus were searched by BLAST program (Camacho *et al.* 2009).

Genome editing of *MSE1* orthologue in *A. thaliana*

Genome editing of the *MSE1* orthologue *TDF1* was performed using the binary vector pEgP226-2A-gfbsd (Osakabe *et al.* 2016), which was designed for CRISPR/Cas9 and guide RNA–mediated genome editing. Three primer sets [735
(5’-GATTTTGGACTTGTCACAACAAGG-3’) – 736
(5’-AAACCCTTGTTGTGACAAGTCCAA-3’) for *GE1*, 737
(5’-GATTTCCATTGCACGAAAGCTTCC-3’) – 738
(5’-AAACGGAAGCTTTCGTGCAATGGA-3’) for *GE2*, and 739
(5’-GATTTAATGTTTCTGAATTCTGCA-3’) – 740
(5’-AAACTGCAGAATTCAGAAACATTA-3’) for *GE3*] were annealed to serve as guide RNA–targeting sequences. The annealed DNA fragments were subcloned into the *Bsa*I site

of pEgP226-2A-gfbsd. Transformation of *A. thaliana* (Col-0) was performed by floral dip method (Clough & Bent 1998) using *Agrobacterium* (pMP90) harboring these binary vectors. Transformants were screened in Murashige–Skoog (MS) medium containing 0.6% agar and 60 μg/mL kanamycin (Murashige & Skoog 1962), and then transferred into soil. Transformants were confirmed by genomic PCR with a primer set, GEF (5'-ACTTAAGACCTGACC-3') and GER (5'-GATGATTTGGATGGC-3').

Acknowledgments:

We thank A. Takahashi, F. Kodama, H. Asao, and A. Akita for technical assistance, and T. Nishimoto and H. Asao for a kind gift of 3-year-old *A. officinalis* cv. Super Welcome. Computational resources were provided by the Data Integration and Analysis Facility, NIBB. This work was supported by a Grant-in-Aid for Scientific Research (25252021 to S.T.) from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT), Takeda Science Foundation (to K.M.), and the Council for Science, Technology and Innovation (CSTI), Cross-ministerial Strategic Innovation Promotion Program (SIP), "Technologies for creating next-generation agriculture, forestry and fisheries" (funding agency: Bio-oriented Technology Research Advancement Institution, NARO) to K.O. Science and technology research promotion program for agriculture, forestry, fisheries and food industry (to A.K. and Y.Ozaki). S.T. supervised this project. K.M. and S.T. designed the experiments. K.M., K.U., T.M., and A.S. performed molecular experiments, with assistance from Y.W. S.F. performed phylogenetic analysis. Y. Osakabe and K.O. were responsible for genome-editing constructs. Y. Ozaki and A. K. collected and maintained the

plant materials, observed the sex phenotypes, and extracted the DNA. S.S. and K.Y. performed genome sequencing and assembly. K.M. and S.T. wrote the manuscript, helped by S.S. and S.F., and all other authors contributed to editing. The *MSE1* sequence and high-throughput sequencing data used in this study have been deposited in DNA Data Bank of Japan (DDBJ) under accession numbers LC190965, SAMD00047009, SAMD00047010, SAMD00047011, SAMD00047624, and SAMD00064393.

References:

Akagi, T., Henry, I.M., Tao, R. & Comai, L. (2014) A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646–650.

Ambawat, S., Sharma, P., Yadav, N.R. & Yadav, R.C. (2013) MYB transcription factor genes as regulators for plant responses: an overview. *Physiol. Mol. Biol. Plants* **19**, 307–321.

Aneja, M., Gianfagna, T.J., Garrison, S.A. & Durner, E.F. (1999) Rapid sex-typing of Asparagus for male hybrid seed production using *n*-propyl *N*-(3,4-dichlorophenyl)carbamate (NPC). *HortScience* **34**, 1090–1094.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. & Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421.

Caporali, E., Carboni, A., Galli, M.G., Rossi, G., Spada, A. & Marziani Longo, G.P. (1994) Development of male and female flower in *Asparagus officinalis*. Search for point of transition from hermaphroditic to unisexual developmental pathway. *Sex. Plant Reprod.* **7**, 239–249.

Charlesworth, D. (2002) Plant sex determination and sex chromosomes. *Heredity* **88**, 94–101.

Charlesworth, B. & Charlesworth, D. (1978) A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997.

Clough, S.J. & Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. & Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676.

Darwin, C. *The different forms of flowers on plants of the same species*. (John Murray, London, UK, 1877).

Dubos, C., Stracke, R., Grotewold, E., Weisshaar, B., Martin, C. & Lepiniec, L. (2010) MYB transcription factors in *Arabidopsis*. *Trends Plant Sci.* **15**, 573–581.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L., Tate, J. & Punta, M. (2014) Pfam: the protein families database. *Nucleic Acids Res.* **42**, D222–230.

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., Berlin, A.M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E.S. & Jaffe, D.B. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. USA* **108**, 1513–1518.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis,

X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N. & Regev, A. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652.

Harkess, A., Mercati, F., Shan, H.Y., Sunseri, F., Falavigna, A. & Leebens-Mack, J. (2015) Sex-biased gene expression in dioecious garden asparagus (*Asparagus officinalis*). *New Phytol.* **207**, 883–892.

Jamsari, A., Nitz, I., Reamon-Büttner, S.M. & Jung, C. (2004) BAC-derived diagnostic markers for sex determination in asparagus. *Theor. Appl. Genet.* **108**, 1140–1146.

Katoh, K. & Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780.

Kubitzki, K. & Rudall, P. J. "Asparagaceae" in *The Families and Genera of Vascular Plants* (Springer, New York, 1998), vol. 3, pp. 125–129.

Kubota, S., Konno, I. & Kanno, A. (2012) Molecular phylogeny of the genus *Asparagus* (Asparagaceae) explains interspecific crossability between the garden asparagus (*A. officinalis*) and other *Asparagus* species. *Theor. Appl. Genet.* **124**, 345–354.

Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Löptien, H. (1979) Identification of the sex chromosome pair in asparagus (*Asparagus officinalis* L.). *Z. Pflanzenz* **82**, 162–173.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12.

Marks, M. 1973. A reconsideration of the genetic mechanism for sex determination in *Asparagus officinalis*. *Proc. EUCARPIA Meeting on Asparagus* (*Asparagus officinalis* L.), pp. 123–128. Wageningen, Netherlands: EUCARPIA.

Murashige, T. & Skoog, F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol. Plant* **15**, 473–497.

Osakabe, Y., Watanabe, T., Sugano, S.S., Ueta, R., Ishihara, R., Shinozaki, K. & Osakabe, K. (2016) Optimization of CRISPR/Cas9 genome editing to modify abiotic stress responses in plants. *Sci. Rep.* **6**, 26685.

Renner, S.S. (2014) The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am. J. Bot.* **101**, 1588–1596.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G. & Mesirov, J.P. (2011) Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. & Huelsenbeck, J.P. (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D. & Higgins, D.G. (2011) Fast, scalable

generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539.

Stracke, R., Werber, M. & Weisshaar, B. (2001) The *R2R3-MYB* gene family in *Arabidopsis thaliana*. *Curr. Opin. Plant Biol.* **4**, 447–456.

Takayama, S. & Isogai, A. (2005) Self-incompatibility in plants. *Annu. Rev. Plant Biol.* **56**, 467–489.

Talavera, G. & Castresana, J. (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **56**, 564–577.

Telgmann-Rauber, A., Jamsari, A., Kinney, M.S., Pires, J.C. & Jung, C. (2007) Genetic and physical maps around the sex-determining *M*-locus of the dioecious plant asparagus. *Mol. Genet. Genomics* **278**, 221–234.

Zhu, J., Chen, H., Li, H., Gao, J.F., Jiang, H., Wang, C., Guan, Y.F. & Yang, Z.N. (2008) *Defective in Tapetal development and function 1* is essential for anther development and tapetal function for microspore maturation in Arabidopsis. *Plant J.* **55**, 266–277.

Figure legends:

Fig. 1. Characterization of the *MSE1* gene. (**A** and **B**) Photographs of male (A) and female (B) flowers of *A. officinalis* cv. Super Welcome. Arrows show the vestiges of opposite-sex organs. (**C**) RT-PCR analysis of full-length *MSE1* using mRNA extracted from early buds. (**D**) PCR amplification of *MSE1* from genomic DNA extracted from male and female individuals of *A. officinalis* cv. Super Welcome. (**E**) Domain structure of MSE1 protein. R2- and R3-type MYB domains are shown. (**F** and **G**) Quantitative RT-PCR

analysis of *MSE1* expression using mRNA extracted from each tissue of *A. officinalis* cv. Super Welcome. Expression levels were normalized against the corresponding levels of *Actin*. Means and SEs of three (F) and nine (G) replicates are shown. (**H**) Comparison of the *MSE1* locus between the X and Y chromosomes. Gray boxes show protein-coding regions. Dotted lines show the genomic regions that share DNA sequence similarities. Scaffolds were extracted from the assembly data of Illumina short read sequences from a single male DNA of *A. officinalis* cv. Super Welcome. PB means a long read sequence generated by PacBio sequencer using female genome.

Fig. 2. Conservation of *MSE1* in genus *Asparagus*. PCR amplification of *MSE1* from male (M) and female (F) individuals of *A. officinalis* three cultivars and dioecious (**A**) and eight individuals of hermaphroditic species (**B**) in genus *Asparagus*. Their sexes were determined by male-specific PCR markers or flower phenotypes.

Fig. 3. Genome editing of *MSE1* orthologue, *TDF1*, in the model plant *A. thaliana*. (**A**) Genomic region of the *TDF1* locus in *A. thaliana*. Gray boxes show protein-coding regions. Arrows show target sites (*GE1*, position 32–10 in *MSE1* cDNA; *GE2*, 463–485; *GE3*, 610–588) of the guide RNA–CRISPR/Cas9 complex. (**B**) Transgenes were amplified from genomic DNA of each transformant with primers for amplifying the guide RNA region. (**C** to **F**) Phenotypes of transgenic plants obtained by *TDF1* genome editing. The transformants were designated as *tdf1-GE1* to *3*. Photographs show 2–3-week-old shoots of wild type (Col-0) (C), *tdf1-GE1* (D), *tdf1-GE2* (E), and *tdf1-GE3* (F). Arrows show siliques. Bars, 1

cm. (**G** and **H**) Male sterile phenotype of *tdf1-GE3*. Pollen grains were observed on anther and stigma in wild type (G), but not in *tdf1-GE3* (H). Bars, 1 mm.

Supporting Information:

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 Summary of screening for male-specific genes in *A. officinalis*.

Figure S2 PCR amplification of male-specific gene candidates in male and female genomes.

Figure S3 Female RNA sequence fragments map to the 5' terminus of *MSE1* cDNA.

Figure S4 *MSE1* is tightly linked to the Y chromosome in *A. officinalis*.

Figure S5 Alignment of predicted amino-acid sequences of MSE1 and other homologues.

Figure S6 Male DNA sequence reads are specifically mapped to the Scaffold_47312.

Figure S7 Alignment of DNA sequences of $MSE1^Y$ and $MSE1^X$ in *A. officinalis*.

Figure S8 One-base deletion creates a stop codon in $MSE1^X$ in *A. officinalis*.

Figure S9 Phylogeny of genus *Asparagus*.

Figure S10 Alignment of predicted amino-acid sequences of MSE1 orthologues in asparagus and hermaphroditic species.

Figure S11 Alignment of predicted amino-acid sequences of MSE1 orthologues from male and female individuals in three dioecious species that have no deleterious mutation in *MSE1* coding region.

Figure S12 Phylogeny of MSE1-related MYB domain transcription factors.

Table S1 Summary of transcriptome analysis of developing *A. officinalis* flowers

Table S2 List of contigs enriched in male genome sequencings

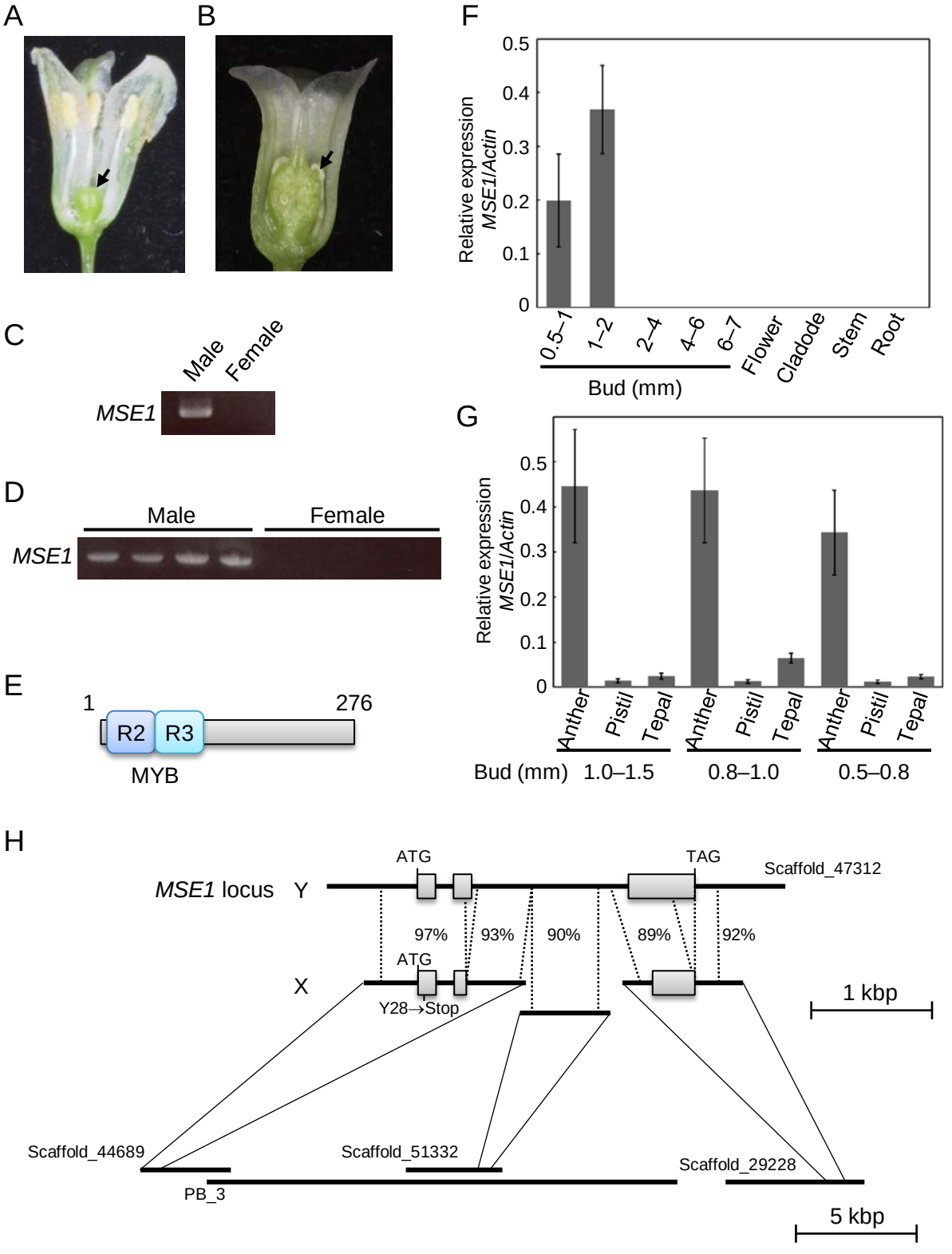
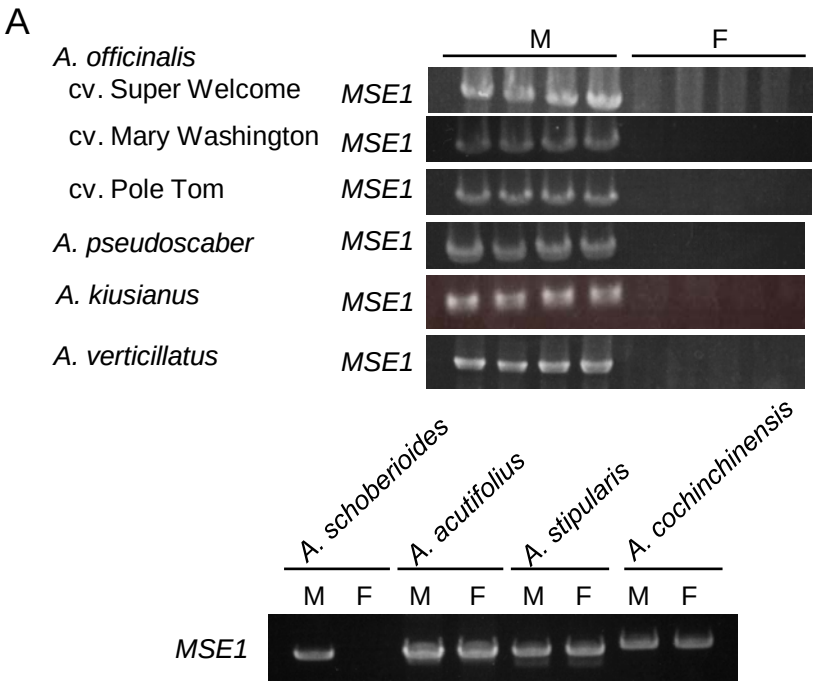

A
B
F
Relative expression
MSE1/Actin
0.5
0.4
0.3
0.2
0.1
0
0.5–1
1–2
2–4
4–6
6–7
Flower
Cladode
Stem
Root
Bud (mm)
C
Male
Female
MSE1
D
Male
Female
MSE1
G
Relative expression
MSE1/Actin
Anther
Pistil
Tepal
Bud (mm) 1.0–1.5
0.8–1.0
0.5–0.8
E
1
276
R2
R3
MYB
H
MSE1 locus Y
ATG
TAG
Scaffold_47312
97%
93%
90%
89%
92%
X
ATG
Y28→Stop
1 kbp
Scaffold_44689
Scaffold_51332
Scaffold_29228
PB_3
5 kbp

Fig. 1


A
A. officinalis
cv. Super Welcome
cv. Mary Washington
cv. Pole Tom
A. pseudoscaber
A. kiusianus
A. verticillatus
MSE1
M
F
A. schoberioides
A. acutifolius
A. stipularis
A. cochinchinensis
M F M F M F M F
MSE1

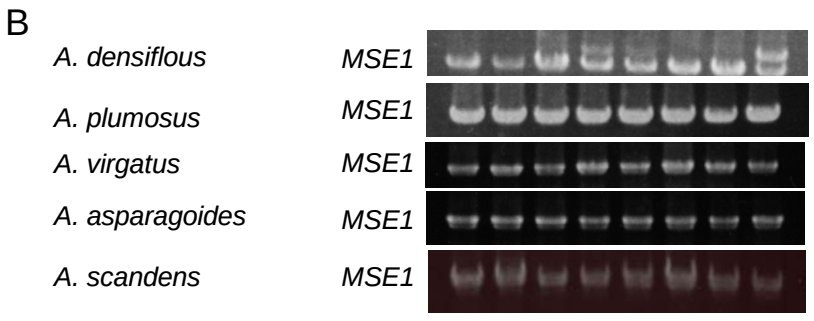

B
A. densiflous
A. plumosus
A. virgatus
A. asparagoides
A. scandens
MSE1

Fig. 2

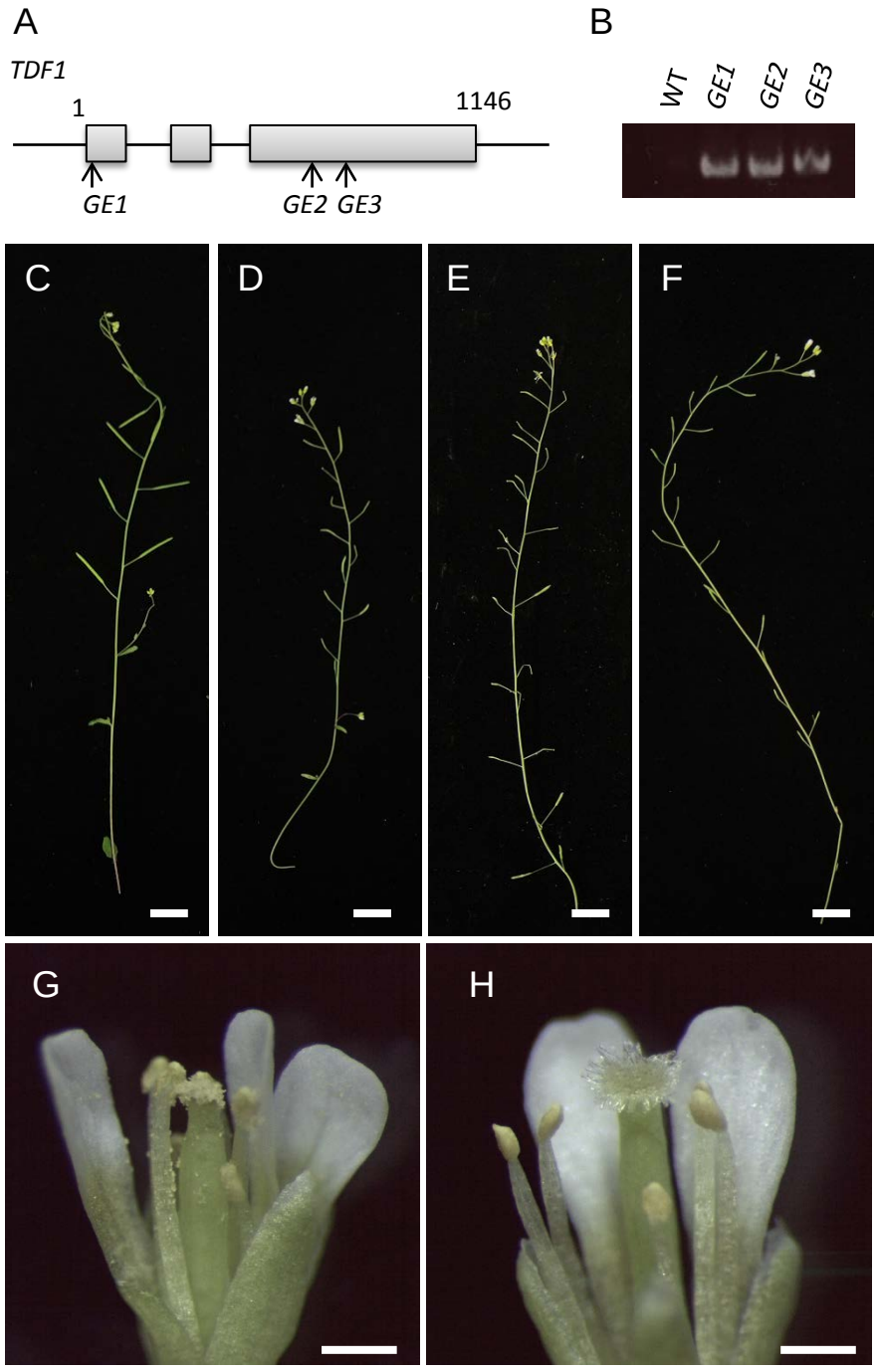


Fig. 3

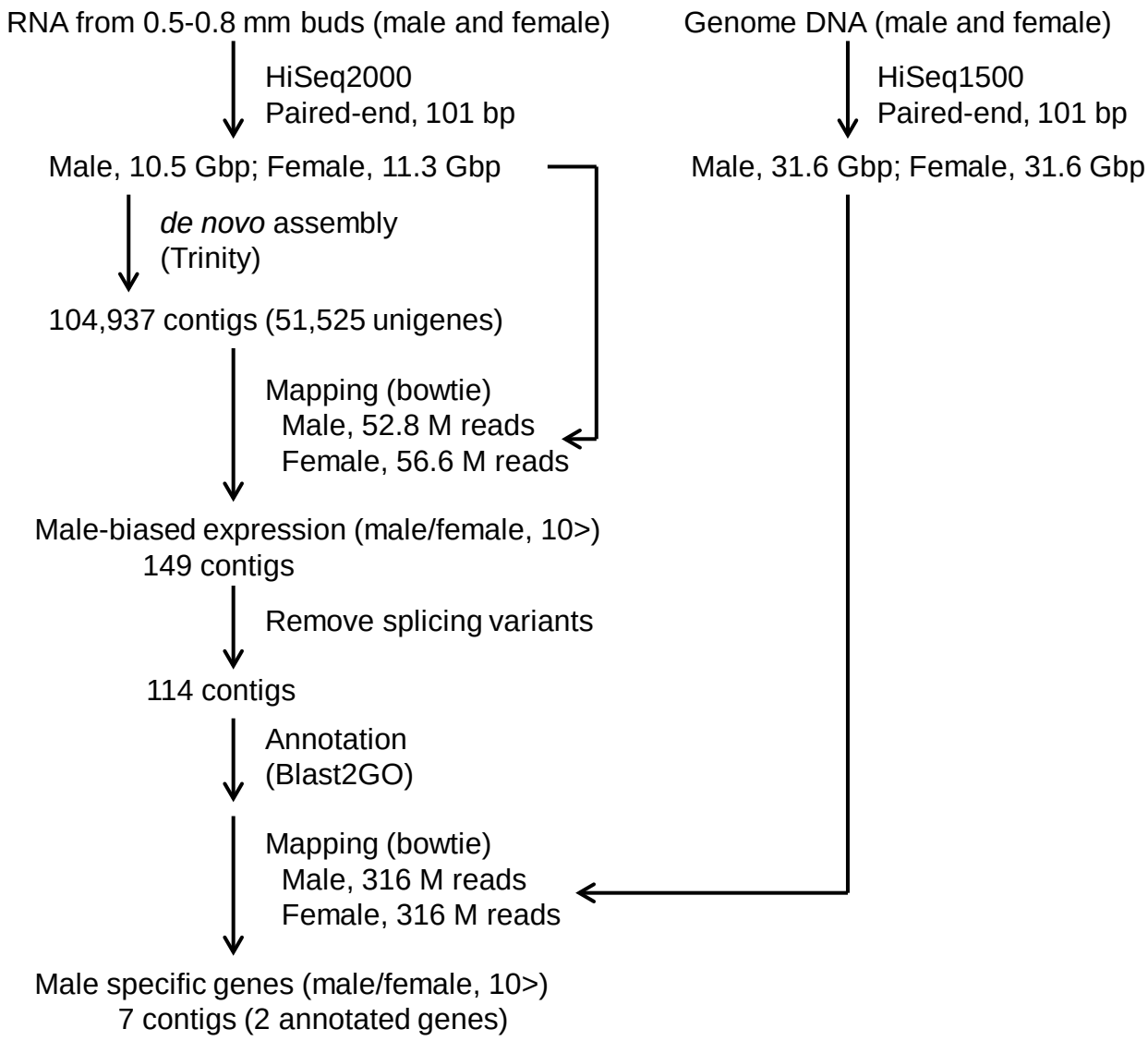


Fig. S1. Summary of screening for male-specific genes in *A. officinalis*.
RNA extracted from early flowers of male and female *A. officinalis* cv. Super Welcome plants was sequenced on an Illumina HiSeq instrument. Assembly of RNA reads was performed with the Trinity program using male reads, yielding 104,937 contigs. Mapping of male and female RNA sequence reads against the assembled contigs revealed 149 contigs with male-biased expression (i.e., ≥10-fold more mapped reads in males than in females, and more than 50 mapped reads in males). After removal of splicing variants, 114 male-specific candidates were annotated by BLAST2GO, and further mapping of genome reads was performed. Ultimately, seven contigs were screened as male-specific gene.

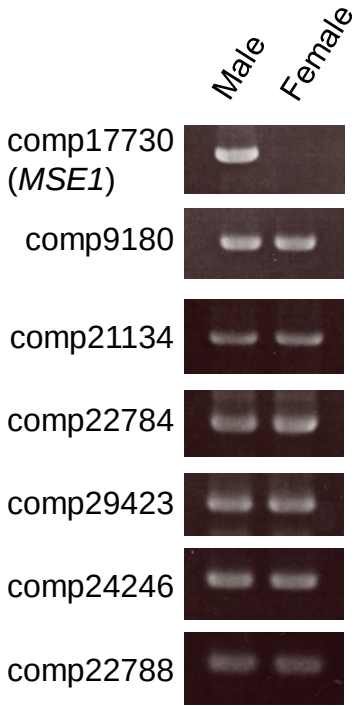


Fig. S2. PCR amplification of male-specific gene candidates in male and female genomes. PCR was performed each male-specific candidate with specific primers using male and female genomic DNA from *A. officinalis* cv. Super Welcome (pools of four individuals for each sex). Only comp17730 (*MSE1*) was specifically amplified in this experiment.

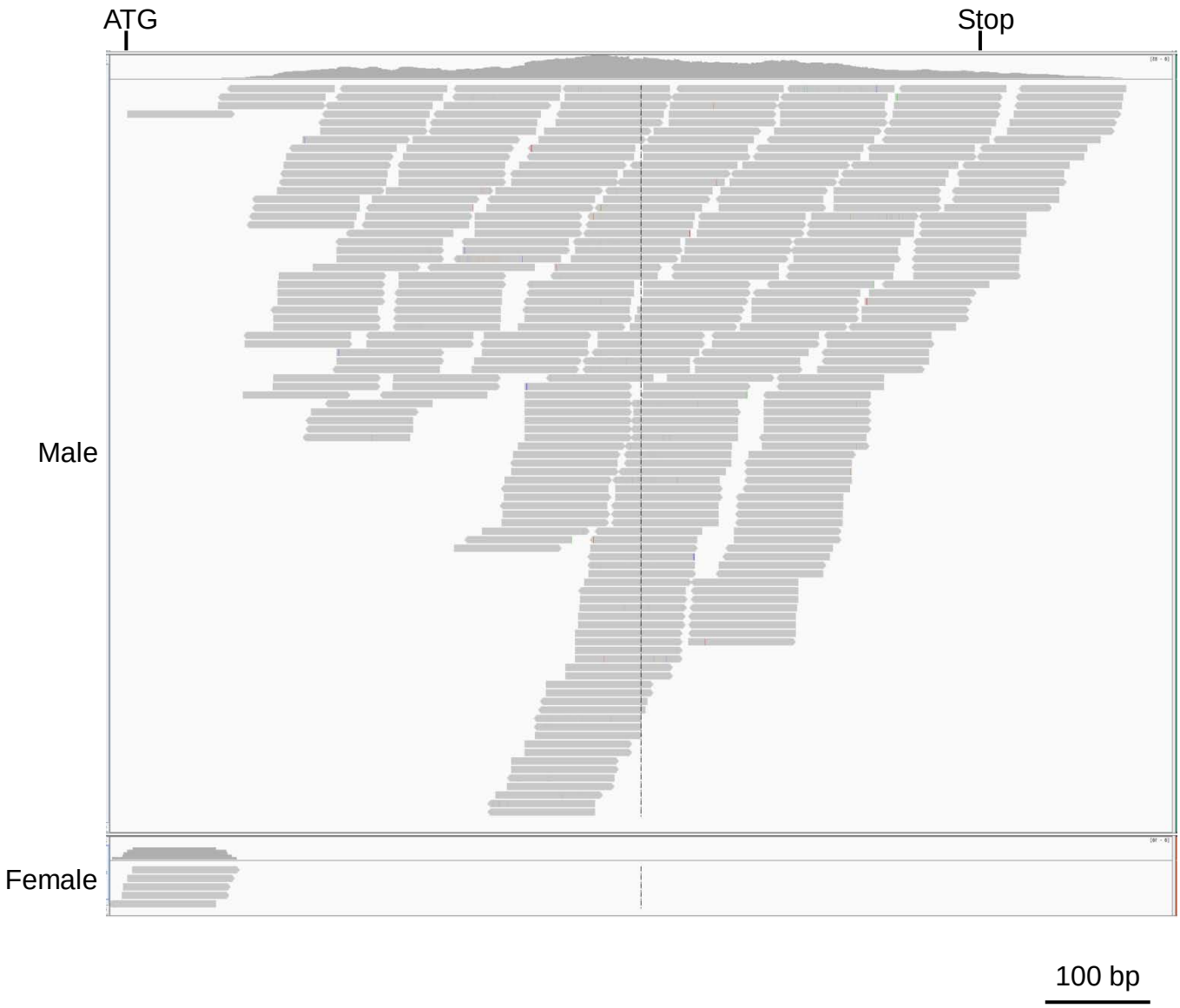


Fig. S3. Female RNA sequence fragments map to the 5' terminus of *MSE1* cDNA.
Mapping data of male and female RNA sequence reads against *MSE1* cDNA was visualized using the IGV software. Gray bars show single RNA sequence reads mapped to *MSE1*. Colored vertical lines represent mismatches with the assembled sequence, caused by SNPs or sequencing errors.

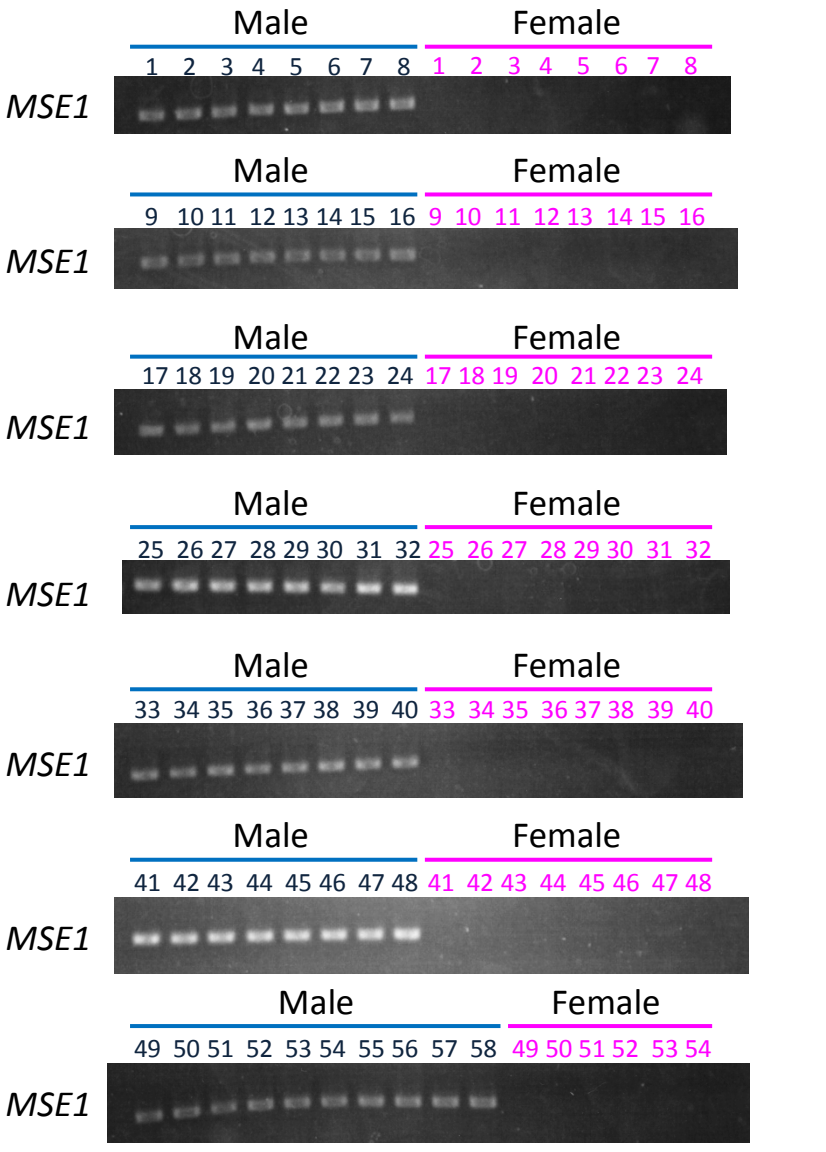


Fig. S4. *MSE1* is tightly linked to the Y chromosome in *A. officinalis.*
Seeds of *A. officinalis* cv. Super Welcome were treated with APC. After a month, about 60% of the seedlings had flowered. Individuals with abnormal or bisexual flowers were removed from this experiment. Genomic DNA was extracted from each individual and used as template for PCR amplification with *MSE1*-specific primers.

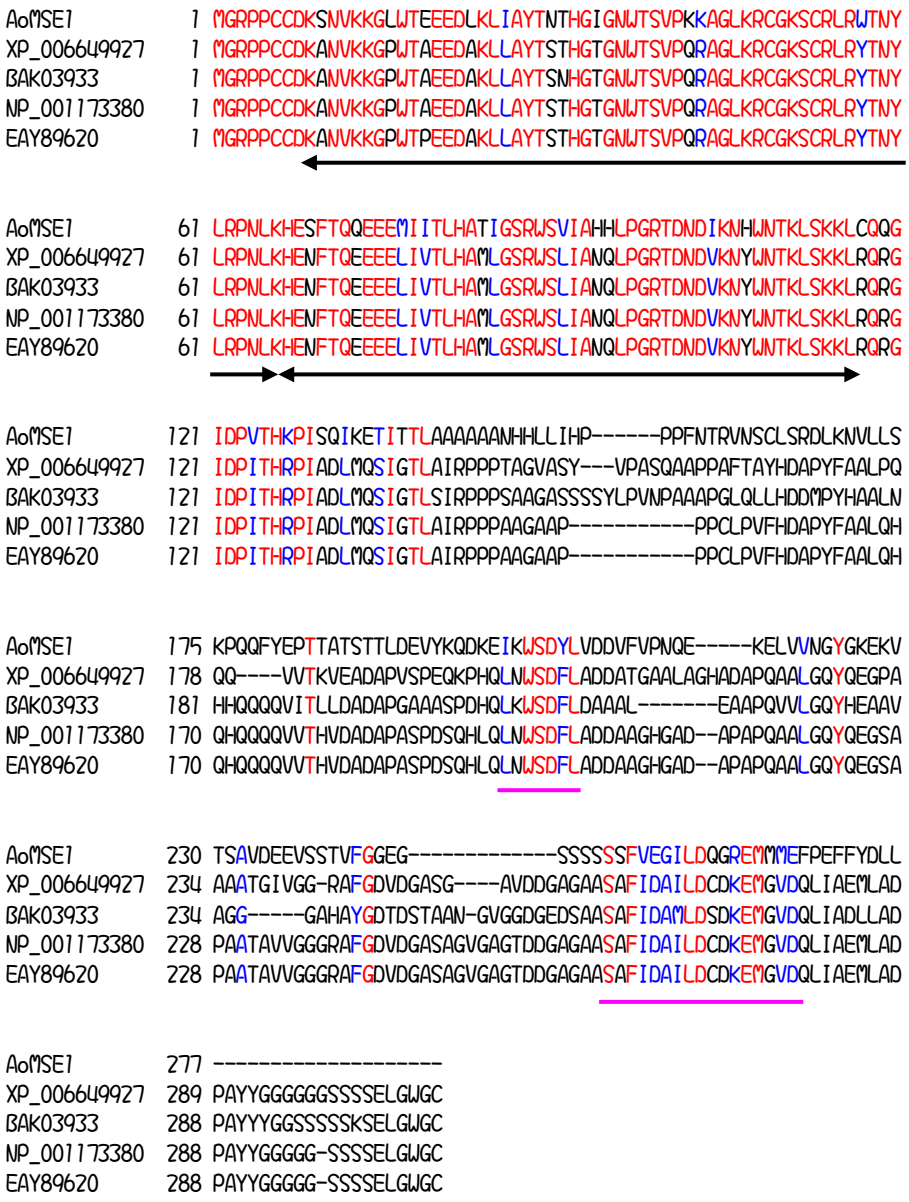


Fig. S5. Alignment of predicted amino-acid sequences of MSE1 and other homologues. Double-headed black bars show MYB domains. Magenta bars represent unknown conserved motifs. Aligned protein sequences are AoMSE1 (*A. officinalis*); XP_006649927, *Oryza brachyantha* (wild rice; accession no. XP_006649927); BAK03933, *Hordeum vulgare* (barley; BAK03933); NP_001173380, *Oryza sativa* (Japonica Group; NP_001173380); and EAY89620, *Oryza sativa* (Indica Group; EAY89620). Colored letters indicate conserved (red) or similar (blue) residues in this alignment.

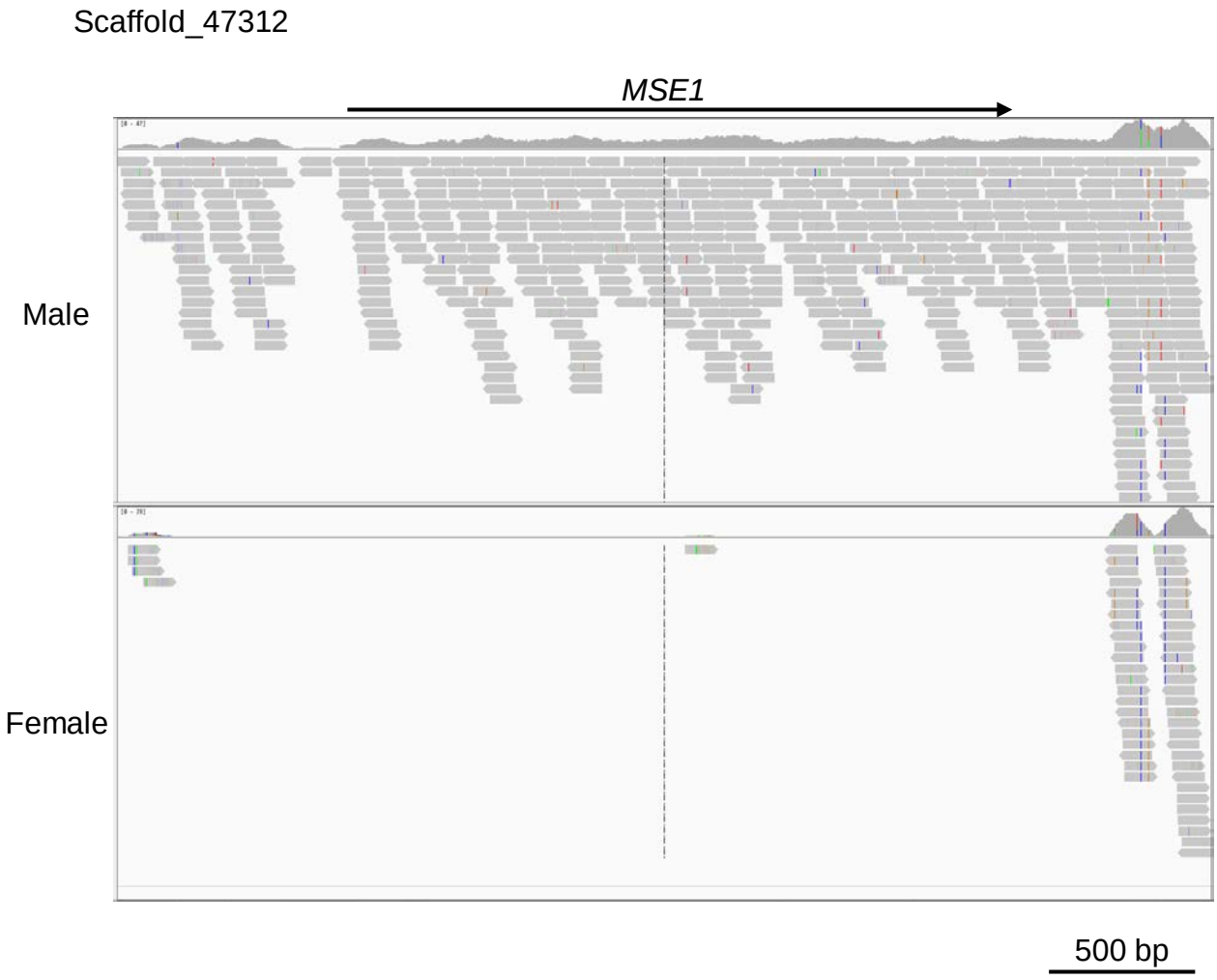


Fig. S6. Male DNA sequence reads are specifically mapped to the Scaffold_47312. Male and female DNA sequence reads were mapped to the Scaffold_47312 by *bowtie*. Mapping data was visualized using the IGV software. Arrow shows *MSE1* coding region including introns. Gray bars show single DNA sequence reads mapped to *MSE1*. Colored vertical lines represent mismatches with the assembled sequence, caused by SNPs or sequencing errors.

A

```
SW_Y1     1 TTTTGTAGAATTAGGAGATGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGG
MW_Y1     1 TTTTGTAGAATTAGGAGATGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGG
NA_Y1     1 TTTTGTAGAATTAGGAGATGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGG
SW_X1     1 TTTTGTAGAATTAGGAGATGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGG
MW_X1     1 TTTTGTAGAATTAGGAGAGGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGG
NA_X1     1 TTTTGTAGAATTAGGAGATGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGG
Aps_X1    1 TTTTGTAGAATTAGGAGAGGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGG
Aki_X1    1 TTTTGTAGAATTAGGAGATGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGG
Asb_X1    1 TTTTGTAGAATTAGGAGATGGGCAGGCCTCCATGCTGCGGTAAATCCAACGTGAAGAAGG
Ave_X1    1 TTTTGTAGAATTAGGAGACGGGCAGGCCTCCATGCTGCGGTAAATCCAACGTGAAGAAGG

SW_Y1    61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAGCTAATAGCTTATACCAACACTCATGG
MW_Y1    61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAGCTAATAGCTTATACCAACACTCATGG
NA_Y1    61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAGCTAATAGCTTATACCAACACTCATGG
SW_X1    61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAGCTAATAGCTTA-ACCAACACTCATGG
MW_X1    61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAGCTAATAGCTTA-ACCAACACTCATGG
NA_X1    61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAGCTAATAGCTTA-ACCAACACTCATGG
Aps_X1   61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAGCTAATAGCTTA-ACCAACACTCATGG
Aki_X1   61 GACTTTAATTTGGACTGAGCAAGAAGATTTGAAGCTAATAGCTTA-ACCAACACTCATGG
Asb_X1   61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAACTAACAGCTTA-ACCAACACTCATGG
Ave_X1   61 GAC-----TTTGGACTGAGGAAGAAGATTTGAAACTAATAGCTTA-ACCAACACTCATGG

SW_Y1   116 AATAGGAAATTGGACTTCTGTTCCAAAGAAAGCAGGTTCTTTTTACGTAGCTAATTGGTT
MW_Y1   116 AATAGGAAATTGGACTTCTGTTCCAAAGAAAGCAGGTTCTTTTTACGTAGCTAATTGGTT
NA_Y1   116 AATAGGAAATTGGACTTCTGTTCCAAAGAAAGCAGGTTCTTTTTACGTAGCTAATTGGTT
SW_X1   115 AATAGGAAATTGGACATCTGTTCCAAAGAAAGCAGGTTCTTTTTATGTAGCTAATTGGTT
MW_X1   115 AATAGGAAATTGGACATCTGTTCCAAAGAAAGCAGGTTCTTTTTATGTAGCTAATTGGTT
NA_X1   115 AATAGGAAATTGGACATCTGTTCCAAAGAAAGCAGGTTCTTTTTATGTAGCTAATTGGTT
Aps_X1  115 AATAGGAAATTGGACATCTGTTCCAAAGAAAGCAGGTTCTTTTTATGTAGCTAATTGGTT
Aki_X1  120 AATAGGAAATTTGACTTCTGTTCCAAAGAAAGCAGGTTCTTTTAATGTAGCTAATTGGTT
Asb_X1  115 AATAGGATATTGGACTTCTGTTCCAAAGAAAGCAGGTTCTTTTTATGTAGCTAATTGGTT
Ave_X1  115 AATAAGATATTGGACTTCTGTTCCAAAGAAAGCAGGTTCTTTTTATGTAGCTAATTGATT

SW_Y1   176 GATTTTCTTCAAAATAATATTTACGTTATTGATTTTTTATTTATTTATTGGTAATGTGAG
MW_Y1   176 GATTTTCTTCAAAATAATATTTACGTTATTGATTTTTTATTTATTTATTGGTAATGTGAG
NA_Y1   176 GATTTTCTTCAAAATAATATTTACGTTATTGATTTTTTATTTATTTATTGGTAATGTGAG
SW_X1   175 GATTTTCTTCAAAATAATATTTACCTTATTGATTTTTTTCTTATTTATTGGTAATGTCAG
MW_X1   175 GATTTTCTTCAAAATAATATTTACCTTATTGATTTTTTTCTTATTTATTGGTAATGTGAG
NA_X1   175 GATTTTCTTCAAAATAATATTTACCTTATTGATTTTTTTCTTATTTATTGGTAATGTCAG
Aps_X1  175 GATTTTCTTCAAAATAATATTTACCTTATTGATTTTTTTCTTATTTATTGGTAATGTGAG
Aki_X1  180 GATTTTCTTCAAAATAATATTTACCTTATTGATTTTTTTCTTATTTATTGGTAATGTGAG
Asb_X1  175 GATTTTCTTCAAAAAAATATTTACCTTATTGATTTTTTTCTTATTTATTGGTAACGTGAG
Ave_X1  175 GATTTTCTTCAAAAAAATATTTACCTTATTGATTTTTTTCTTATTTATTGGTAATGTGAG

SW_Y1   236 TATGCAATTTTAGGGTTGAAGAGATGTGGGAAGAGCTGTAG----GCTAAGATGGACTAA
MW_Y1   236 TATGCAATTTTAGGGTTGAAGAGATGTGGGAAGAGCTGTAG----GCTAAGATGGACTAA
NA_Y1   236 TATGCAATTTTAGGGTTGAAGAGATGTGGGAAGAGCTGTAG----GCTAAGATGGACTAA
SW_X1   235 TATGCAATTTTAGGGTTTAAGAGATGTGGGAAGAGCTGTAG----GCTGAGATGGACTAA
MW_X1   235 TATGCAATTTTATGGTTTAAGAGATGTGGGAAGAGCTGTAG----GCTGAGATGGACTAA
NA_X1   235 TATGCAATTTTAGGGTTTAAGAGATGTGGGAAGAGCTGTAG----GCTGAGATGGACTAA
Aps_X1  235 TATGCAATTTTATGGTTTAAGAGATGTGGGAAGAGCTGTAG----GCTGAGATGGACTAA
Aki_X1  240 TAAGCAATATTAGGGTTTAAGAGATGTGGGAAGAGCTGTAGCTAGGCTGAGATGGACTAA
Asb_X1  235 TATGCAATTTTAGGGTTGAAGAGATGTGGGAAGAGCTGTAG----GCTGAGATGGACTAA
Ave_X1  235 TATGCAATTTTAGGGTTGAAGAGATATGGGAAGAGCTGTAG----GCTGAGATGGACTAA
```

Fig. S7-1

```
SW_Y1   292 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
MW_Y1   292 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
NA_Y1   292 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
SW_X1   291 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
MW_X1   291 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
NA_X1   291 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
Aps_X1  291 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
Aki_X1  300 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
Asb_X1  291 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT
Ave_X1  291 TTATCTGCGCCCTAATTTGAAGCATGAGAGCTTCACTCAACAAGAAGAGGAGATGATTAT

SW_Y1   352 AACACTTCATGCTACAATCGGAAGCCGGTATTGTTCTCTCTATCAATTTATTTGATT-GA
MW_Y1   352 AACACTTCATGCTACAATCGGAAGCCGGTATTGTTCTCTCTATCAATTTATTTGATT-GA
NA_Y1   352 AACACTTCATGCTACAATCGGAAGCCGGTATTGTTCTCTCTATCAATTTATTTGATT-GA
SW_X1   351 ------------------------------------------------------------
MW_X1   351 AACGCTTCAAGCTACAATCGGAAGCCGGTATTGTTCTCTCTCTCAATTTATTTGATT-GA
NA_X1   351 ------------------------------------------------------------
Aps_X1  351 AACGCTTCAAGCTACAATCGGAAGCCGGTATTGTTCTCTCTCTCAATTTATTTGAT--GA
Aki_X1  360 AACACTTCATGCTACAATCGGAAGCCGGTATTGTTCTCTCTCTCAATTTATTTGATT-GA
Asb_X1  351 ------TCATGCTACAATCGGAAGCCGGTATTGTTCTCTCTCTCAATTTATTTGATTTGA
Ave_X1  351 AACACTTGATGCTATGATCGGCAGCCGGTATTGTTCTCTCTTTCAATTTATTTGATT-GA

SW_Y1   411 GGATTTTATATCATGATGTTAA---CTACTGTAGTTTTCATTTTAGTCAATTAGGAACAT
MW_Y1   411 GGATTTTATATCATGATGTTAA---CTACTGTAGTTTTCATTTTAGTCAATTAGGAACAT
NA_Y1   411 GGATTTTATATCATGATGTTAA---CTACTGTAGTTTTCATTTTAGTCAATTAGGAACAT
SW_X1   351 ---------ATCATGATGTTAA---CTACTGTAGTCCTCATTTTAGTCAATTAGGAACAA
MW_X1   410 GGATTTTATATCATGATGTTAA---CTACTGTAGTTCTCATTTTAGTCAATTAGGAACAA
NA_X1   351 ---------ATCATGATGTTAA---CTACTGTAGTCCTCATTTTAGTCAATTAGGAACAA
Aps_X1  409 GGATTTTATATCATGATGTTAA---CTACTGTAGTTCTCATTTTAGTCAATTAGGAACAA
Aki_X1  419 GGATTTTATATCATGATGTTAA---CTACTGTAGTTCTCATTTTAGTCAATTAGGAACAA
Asb_X1  405 GGATTTTATATCATGATGTTAA---CTACTGTAGTTCTCATTTTAGTCAATTAGGAACAA
Ave_X1  410 GGATTTTATATCATGATGTTAAGTACTACTGTAGTTCTCATTTTAGTCAATTAGGAACAA

SW_Y1   468 TATTTTGTTTGAATAGTTTTAACCTAGAAAGAATTTCTAGTGTAGAGCTGCTAACACAAA
MW_Y1   468 TATTTTGTTTGAATAGTTTTAACCTAGAAAGAATTTCTAGTGTAGAGCTGCTAACACAAA
NA_Y1   468 TATTTTGTTTGAATAGTTTTAACCTAGAAAGAATTTCTAGTGTAGAGCTGCTAACACAAA
SW_X1   399 TATTTTGTTTGAATAGTTTTAACCTAGAATGAATTTCTAGTGTAGTGCTGCTAACACAAA
MW_X1   467 TATTTTGTTTGAATAATTTTAACCTAGAATGAATTTCTAGTGTAGTGCTGCTAACACAAA
NA_X1   399 TATTTTGTTTGAATAGTTTTAACCTAGAATGAATTTCTAGTGTAGTGCTGCTAACACAAA
Aps_X1  466 TATTTTGTTTGAATAATTTTAACCTAGAATGAATTTCTAGTGTAGTGCTGCTAACACAAA
Aki_X1  476 TATTTTGTTTGAATAGTTTT----------------------------------------
Asb_X1  462 TATTTTGTTTGAATAGTTTTAACCTAGAATGAATTTCTAGTGTAGTGCTGCTAACACAAA
Ave_X1  470 TATTTTGTTTGAATAGTTTTAACC-----------------------------------
```

Fig. S7-2

B

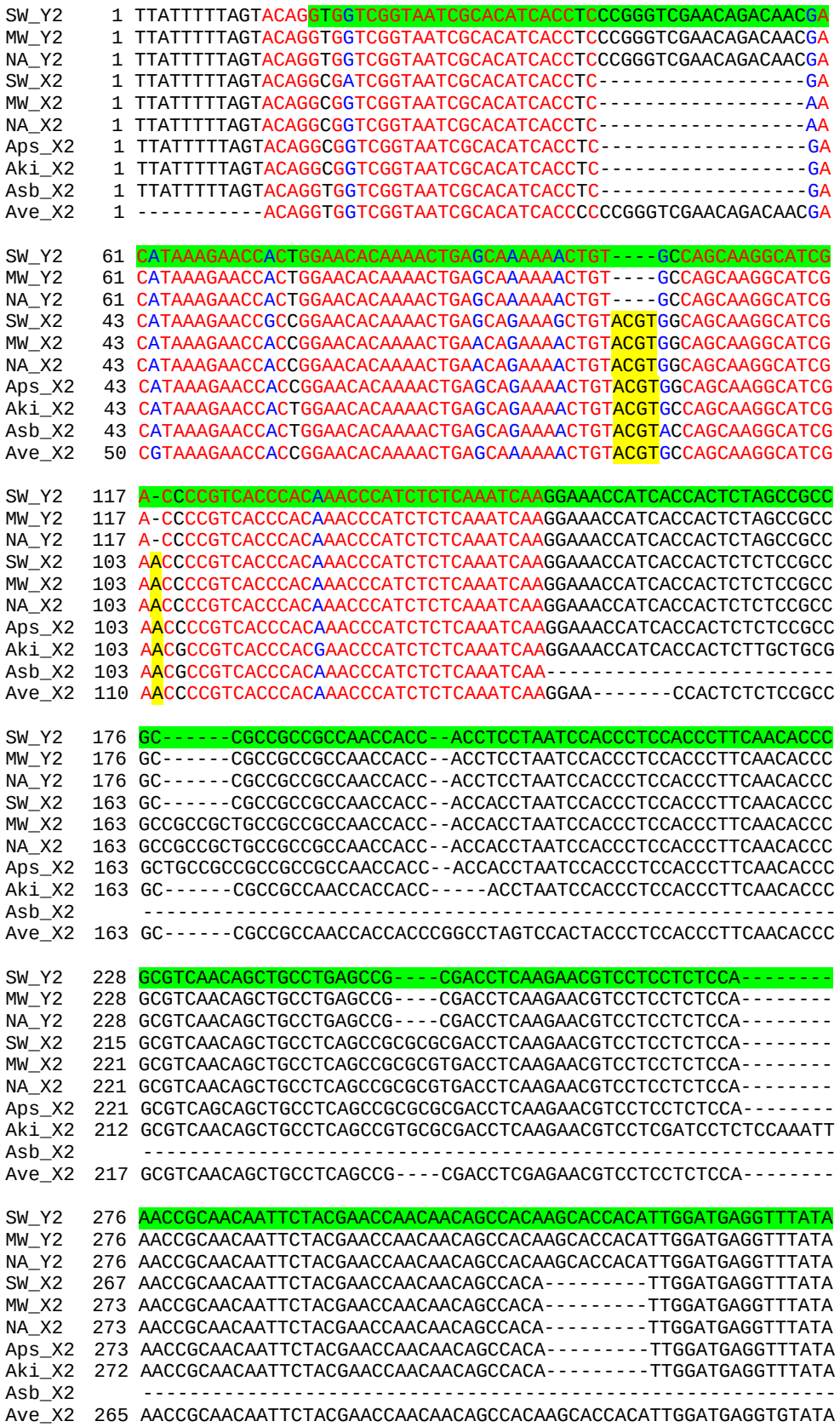

SW_Y2 1 TTATTTTTAGTACAGGTGGTCGGTAATCGCACATCACCTCCCGGGTCGAACAGACAACGA
MW_Y2 1 TTATTTTTAGTACAGGTGGTCGGTAATCGCACATCACCTCCCGGGTCGAACAGACAACGA
NA_Y2 1 TTATTTTTAGTACAGGTGGTCGGTAATCGCACATCACCTCCCGGGTCGAACAGACAACGA
SW_X2 1 TTATTTTTAGTACAGGCGATCGGTAATCGCACATCACCTC----------GA
MW_X2 1 TTATTTTTAGTACAGGCGGTCGGTAATCGCACATCACCTC----------AA
NA_X2 1 TTATTTTTAGTACAGGCGGTCGGTAATCGCACATCACCTC----------AA
Aps_X2 1 TTATTTTTAGTACAGGCGGTCGGTAATCGCACATCACCTC----------GA
Aki_X2 1 TTATTTTTAGTACAGGCGGTCGGTAATCGCACATCACCTC----------GA
Asb_X2 1 TTATTTTTAGTACAGGTGGTCGGTAATCGCACATCACCTC----------GA
Ave_X2 1 ----------ACAGGTGGTCGGTAATCGCACATCACCCCCGGGTCGAACAGACAACGA
SW_Y2 61 CATAAAGAACCACTGGAACACAAAACTGAGCAAAAAACTGT----GCCAGCAAGGCATCG
MW_Y2 61 CATAAAGAACCACTGGAACACAAAACTGAGCAAAAAACTGT----GCCAGCAAGGCATCG
NA_Y2 61 CATAAAGAACCACTGGAACACAAAACTGAGCAAAAAACTGT----GCCAGCAAGGCATCG
SW_X2 43 CATAAAGAACCGCCGGAACACAAAACTGAGCAGAAAGCTGTACGTGGCAGCAAGGCATCG
MW_X2 43 CATAAAGAACCACCGGAACACAAAACTGAACAGAAAACTGTACGTGGCAGCAAGGCATCG
NA_X2 43 CATAAAGAACCACCGGAACACAAAACTGAACAGAAAACTGTACGTGGCAGCAAGGCATCG
Aps_X2 43 CATAAAGAACCACCGGAACACAAAACTGAGCAGAAAACTGTACGTGGCAGCAAGGCATCG
Aki_X2 43 CATAAAGAACCACTGGAACACAAAACTGAGCAGAAAACTGTACGTGCCAGCAAGGCATCG
Asb_X2 43 CATAAAGAACCACTGGAACACAAAACTGAGCAGAAAACTGTACGTACCAGCAAGGCATCG
Ave_X2 50 CGTAAAGAACCACCGGAACACAAAACTGAGCAAAAAACTGTACGTGCCAGCAAGGCATCG
SW_Y2 117 A-CCCCGTCACCCACAAACCCATCTCTCAAATCAAGGAAACCATCACCACTCTAGCCGCC
MW_Y2 117 A-CCCCGTCACCCACAAACCCATCTCTCAAATCAAGGAAACCATCACCACTCTAGCCGCC
NA_Y2 117 A-CCCCGTCACCCACAAACCCATCTCTCAAATCAAGGAAACCATCACCACTCTAGCCGCC
SW_X2 103 AACCCCGTCACCCACAAACCCATCTCTCAAATCAAGGAAACCATCACCACTCTCTCCGCC
MW_X2 103 AACCCCGTCACCCACAAACCCATCTCTCAAATCAAGGAAACCATCACCACTCTCTCCGCC
NA_X2 103 AACCCCGTCACCCACAAACCCATCTCTCAAATCAAGGAAACCATCACCACTCTCTCCGCC
Aps_X2 103 AACCCCGTCACCCACAAACCCATCTCTCAAATCAAGGAAACCATCACCACTCTCTCCGCC
Aki_X2 103 AACGCCGTCACCCACGAACCCATCTCTCAAATCAAGGAAACCATCACCACTCTTGCTGCG
Asb_X2 103 AACGCCGTCACCCACAAACCCATCTCTCAAATCAA----------
Ave_X2 110 AACCCCGTCACCCACAAACCCATCTCTCAAATCAAGGAA-------CCACTCTCTCCGCC
SW_Y2 176 GC------CGCCGCCGCCAACCACC--ACCTCCTAATCCACCCTCCACCCTTCAACACCC
MW_Y2 176 GC------CGCCGCCGCCAACCACC--ACCTCCTAATCCACCCTCCACCCTTCAACACCC
NA_Y2 176 GC------CGCCGCCGCCAACCACC--ACCTCCTAATCCACCCTCCACCCTTCAACACCC
SW_X2 163 GC------CGCCGCCGCCAACCACC--ACCACCTAATCCACCCTCCACCCTTCAACACCC
MW_X2 163 GCCGCCGCTGCCGCCGCCAACCACC--ACCACCTAATCCACCCTCCACCCTTCAACACCC
NA_X2 163 GCCGCCGCTGCCGCCGCCAACCACC--ACCACCTAATCCACCCTCCACCCTTCAACACCC
Aps_X2 163 GCTGCCGCCGCCGCCGCCAACCACC--ACCACCTAATCCACCCTCCACCCTTCAACACCC
Aki_X2 163 GC------CGCCGCCAACCACCACC-----ACCTAATCCACCCTCCACCCTTCAACACCC
Asb_X2 ----------
Ave_X2 163 GC------CGCCGCCAACCACCACCCGGCCTAGTCCACTACCCTCCACCCTTCAACACCC
SW_Y2 228 GCGTCAACAGCTGCCTGAGCCG----CGACCTCAAGAACGTCCTCCTCTCCA--------
MW_Y2 228 GCGTCAACAGCTGCCTGAGCCG----CGACCTCAAGAACGTCCTCCTCTCCA--------
NA_Y2 228 GCGTCAACAGCTGCCTGAGCCG----CGACCTCAAGAACGTCCTCCTCTCCA--------
SW_X2 215 GCGTCAACAGCTGCCTCAGCCGCGCGCGACCTCAAGAACGTCCTCCTCTCCA--------
MW_X2 221 GCGTCAACAGCTGCCTCAGCCGCGCGTGACCTCAAGAACGTCCTCCTCTCCA--------
NA_X2 221 GCGTCAACAGCTGCCTCAGCCGCGCGTGACCTCAAGAACGTCCTCCTCTCCA--------
Aps_X2 221 GCGTCAGCAGCTGCCTCAGCCGCGCGCGACCTCAAGAACGTCCTCCTCTCCA--------
Aki_X2 212 GCGTCAACAGCTGCCTCAGCCGTGCGCGACCTCAAGAACGTCCTCGATCCTCTCCAAATT
Asb_X2 ----------
Ave_X2 217 GCGTCAACAGCTGCCTCAGCCG----CGACCTCGAGAACGTCCTCCTCTCCA--------
SW_Y2 276 AACCGCAACAATTCTACGAACCAACAACAGCCACAAGCACCACATTGGATGAGGTTTATA
MW_Y2 276 AACCGCAACAATTCTACGAACCAACAACAGCCACAAGCACCACATTGGATGAGGTTTATA
NA_Y2 276 AACCGCAACAATTCTACGAACCAACAACAGCCACAAGCACCACATTGGATGAGGTTTATA
SW_X2 267 AACCGCAACAATTCTACGAACCAACAACAGCCACA---------TTGGATGAGGTTTATA
MW_X2 273 AACCGCAACAATTCTACGAACCAACAACAGCCACA---------TTGGATGAGGTTTATA
NA_X2 273 AACCGCAACAATTCTACGAACCAACAACAGCCACA---------TTGGATGAGGTTTATA
Aps_X2 273 AACCGCAACAATTCTACGAACCAACAACAGCCACA---------TTGGATGAGGTTTATA
Aki_X2 272 AACCGCAACAATTCTACGAACCAACAACAGCCACA---------TTGGATGAGGTTTATA
Asb_X2 ----------
Ave_X2 265 AACCGCAACAATTCTACGAACCAACAACAGCCACAAGCACCACATTGGATGAGGTGTATA

Fig. S7-3

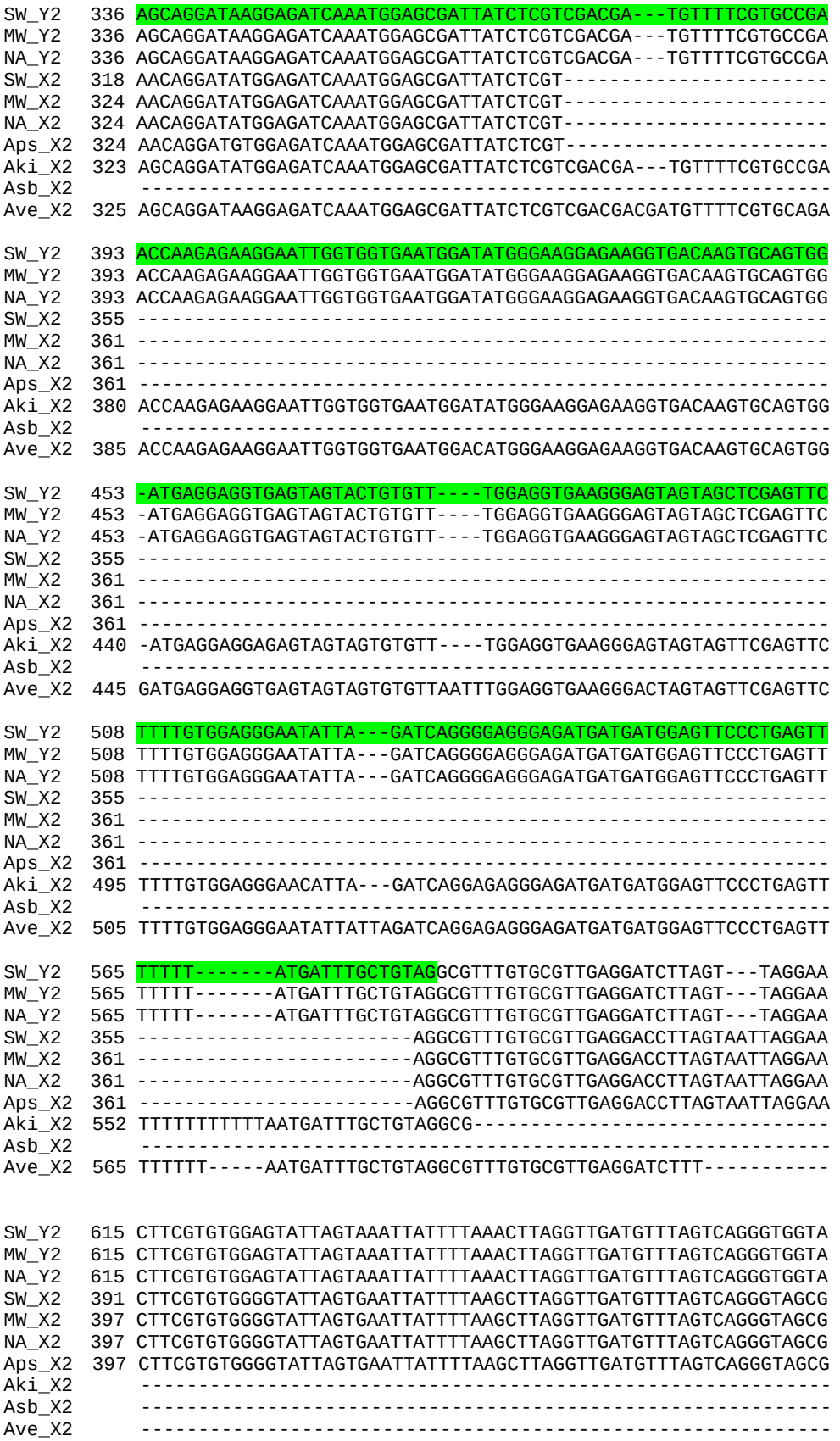

SW_Y2 336 AGCAGGATAAGGAGATCAAATGGAGCGATTATCTCGTCGACGA---TGTTTTCGTGCCGA
MW_Y2 336 AGCAGGATAAGGAGATCAAATGGAGCGATTATCTCGTCGACGA---TGTTTTCGTGCCGA
NA_Y2 336 AGCAGGATAAGGAGATCAAATGGAGCGATTATCTCGTCGACGA---TGTTTTCGTGCCGA
SW_X2 318 AACAGGATATGGAGATCAAATGGAGCGATTATCTCGT----------
MW_X2 324 AACAGGATATGGAGATCAAATGGAGCGATTATCTCGT----------
NA_X2 324 AACAGGATATGGAGATCAAATGGAGCGATTATCTCGT----------
Aps_X2 324 AACAGGATGTGGAGATCAAATGGAGCGATTATCTCGT----------
Aki_X2 323 AGCAGGATATGGAGATCAAATGGAGCGATTATCTCGTCGACGA---TGTTTTCGTGCCGA
Asb_X2 ----------
Ave_X2 325 AGCAGGATAAGGAGATCAAATGGAGCGATTATCTCGTCGACGACGATGTTTTCGTGCAGA
SW_Y2 393 ACCAAGAGAAGGAATTGGTGGTGAATGGATATGGGAAGGAGAAGGTGACAAGTGCAGTGG
MW_Y2 393 ACCAAGAGAAGGAATTGGTGGTGAATGGATATGGGAAGGAGAAGGTGACAAGTGCAGTGG
NA_Y2 393 ACCAAGAGAAGGAATTGGTGGTGAATGGATATGGGAAGGAGAAGGTGACAAGTGCAGTGG
SW_X2 355 ----------
MW_X2 361 ----------
NA_X2 361 ----------
Aps_X2 361 ----------
Aki_X2 380 ACCAAGAGAAGGAATTGGTGGTGAATGGATATGGGAAGGAGAAGGTGACAAGTGCAGTGG
Asb_X2 ----------
Ave_X2 385 ACCAAGAGAAGGAATTGGTGGTGAATGGACATGGGAAGGAGAAGGTGACAAGTGCAGTGG
SW_Y2 453 -ATGAGGAGGTGAGTAGTACTGTGTT----TGGAGGTGAAGGGAGTAGTAGCTCGAGTTC
MW_Y2 453 -ATGAGGAGGTGAGTAGTACTGTGTT----TGGAGGTGAAGGGAGTAGTAGCTCGAGTTC
NA_Y2 453 -ATGAGGAGGTGAGTAGTACTGTGTT----TGGAGGTGAAGGGAGTAGTAGCTCGAGTTC
SW_X2 355 ----------
MW_X2 361 ----------
NA_X2 361 ----------
Aps_X2 361 ----------
Aki_X2 440 -ATGAGGAGGAGAGTAGTAGTGTGTT----TGGAGGTGAAGGGAGTAGTAGTTCGAGTTC
Asb_X2 ----------
Ave_X2 445 GATGAGGAGGTGAGTAGTAGTGTGTTAATTTGGAGGTGAAGGGACTAGTAGTTCGAGTTC
SW_Y2 508 TTTTGTGGAGGGAATATTA---GATCAGGGGAGGGAGATGATGATGGAGTTCCCTGAGTT
MW_Y2 508 TTTTGTGGAGGGAATATTA---GATCAGGGGAGGGAGATGATGATGGAGTTCCCTGAGTT
NA_Y2 508 TTTTGTGGAGGGAATATTA---GATCAGGGGAGGGAGATGATGATGGAGTTCCCTGAGTT
SW_X2 355 ----------
MW_X2 361 ----------
NA_X2 361 ----------
Aps_X2 361 ----------
Aki_X2 495 TTTTGTGGAGGGAACATTA---GATCAGGAGAGGGAGATGATGATGGAGTTCCCTGAGTT
Asb_X2 ----------
Ave_X2 505 TTTTGTGGAGGGAATATTATTAGATCAGGAGAGGGAGATGATGATGGAGTTCCCTGAGTT
SW_Y2 565 TTTTT-------ATGATTTGCTGTAGGCGTTTGTGCGTTGAGGATCTTAGT---TAGGAA
MW_Y2 565 TTTTT-------ATGATTTGCTGTAGGCGTTTGTGCGTTGAGGATCTTAGT---TAGGAA
NA_Y2 565 TTTTT-------ATGATTTGCTGTAGGCGTTTGTGCGTTGAGGATCTTAGT---TAGGAA
SW_X2 355 ----------AGGCGTTTGTGCGTTGAGGACCTTAGTAATTAGGAA
MW_X2 361 ----------AGGCGTTTGTGCGTTGAGGACCTTAGTAATTAGGAA
NA_X2 361 ----------AGGCGTTTGTGCGTTGAGGACCTTAGTAATTAGGAA
Aps_X2 361 ----------AGGCGTTTGTGCGTTGAGGACCTTAGTAATTAGGAA
Aki_X2 552 TTTTTTTTTTAATGATTTGCTGTAGGCG----------
Asb_X2 ----------
Ave_X2 565 TTTTTT-----AATGATTTGCTGTAGGCGTTTGTGCGTTGAGGATCTTT----------
SW_Y2 615 CTTCGTGTGGAGTATTAGTAAATTATTTTAAACTTAGGTTGATGTTTAGTCAGGGTGGTA
MW_Y2 615 CTTCGTGTGGAGTATTAGTAAATTATTTTAAACTTAGGTTGATGTTTAGTCAGGGTGGTA
NA_Y2 615 CTTCGTGTGGAGTATTAGTAAATTATTTTAAACTTAGGTTGATGTTTAGTCAGGGTGGTA
SW_X2 391 CTTCGTGTGGGGTATTAGTGAATTATTTTAAGCTTAGGTTGATGTTTAGTCAGGGTAGCG
MW_X2 397 CTTCGTGTGGGGTATTAGTGAATTATTTTAAGCTTAGGTTGATGTTTAGTCAGGGTAGCG
NA_X2 397 CTTCGTGTGGGGTATTAGTGAATTATTTTAAGCTTAGGTTGATGTTTAGTCAGGGTAGCG
Aps_X2 397 CTTCGTGTGGGGTATTAGTGAATTATTTTAAGCTTAGGTTGATGTTTAGTCAGGGTAGCG
Aki_X2 ----------
Asb_X2 ----------
Ave_X2 ----------

Fig. S7-4

Fig. S7. Alignment of DNA sequences of *MSE1*^{Y} and *MSE1*^{X} in *A. officinalis* and related species.
MSE1^{Y} and *MSE1*^{X} were PCR-amplified from three *A. officinalis* cultivars [SW (Super Welcome), MW (Mary Washington), and NA (Niagara)] and related species [Aps (*A. pseudoscaber*), Aki (*A. kiusianus*), Asb (*A. schoberioides*), Ave (*A. verticillatus*)] and sequenced. Y1 and X1 show *MSE1*^{Y} and *MSE1*^{X} containing the first and second exons (A), and Y2 and X2 show the third exon (B). Green highlight indicates the protein-coding region of *MSE1*^{Y}. Yellow highlight indicates conserved deleterious mutations in *MSE1*^{X} among these species. Although X2 sequence of *A. schoberioides* is partial, there is no conserved deleterious mutation among other dioecious species in the non-sequenced *MSE1*^{X} region of *A. schoberioides*.

MSE1^{Y}

```
1       10        20        30        40        50        60
ATGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGGGACTTTGGACTGAGGAA
M  G  R  P  P  C  C  D  K  S  N  V  K  K  G  L  W  T  E  E

        70        80        90       100       110       120
GAAGATTTGAAGCTAATAGCTTATACCAACACTCATGGAATAGGAAATTGGACTTCTGTT
E  D  L  K  L  I  A  Y  T  N  T  H  G  I  G  N  W  T  S  V
```

MSE1^{X}

```
1       10        20        30        40        50        60
ATGGGCAGGCCTCCATGCTGCGATAAATCCAACGTGAAGAAGGGACTTTGGACTGAGGAA
M  G  R  P  P  C  C  D  K  S  N  V  K  K  G  L  W  T  E  E

        70        80        90       100       110       120
GAAGATTTGAAGCTAATAGCTTA-ACCAACACTCATGGAATAGGAAATTGGACATCTGTT
E  D  L  K  L  I  A  *
```

Fig. S8. One-base deletion creates a stop codon in *MSE1*^{X} in *A. officinalis.*
DNA and translated amino acids of the first 120 bp from the start codon in *MSE1*^{Y} and *MSE1*^{X}. One-base deletion in *MSE1*^{X} results in a conversion from TAT (Y) to TAA (stop codon).

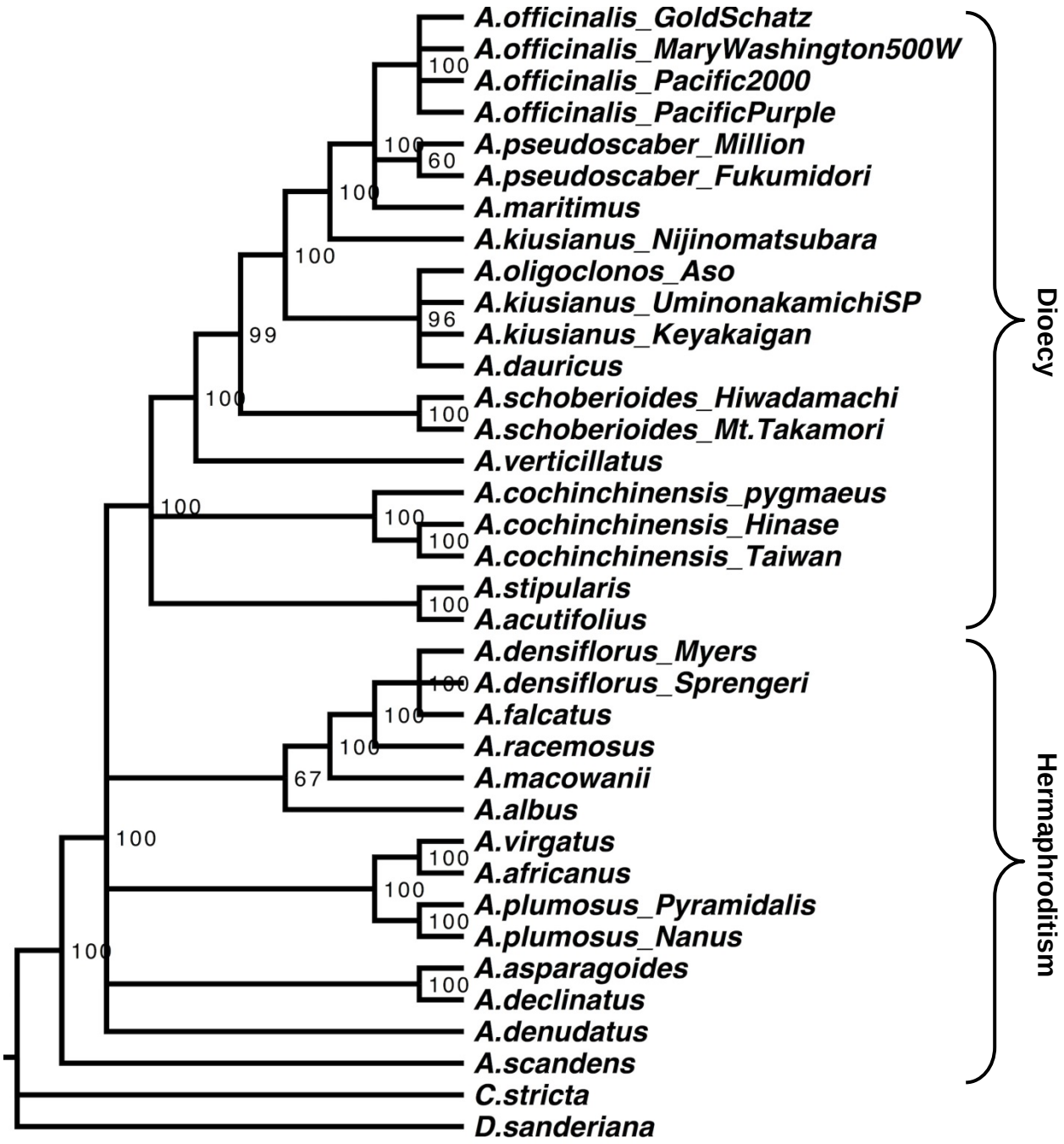


Fig. S9. Phylogeny of genus *Asparagus*.
The majority-rule consensus tree based on Bayesian inference using the five chloroplast intergenic sequences reported by Kubota *et al.* 2012. The *Cordyline stricta* sequence was used as the outgroup.

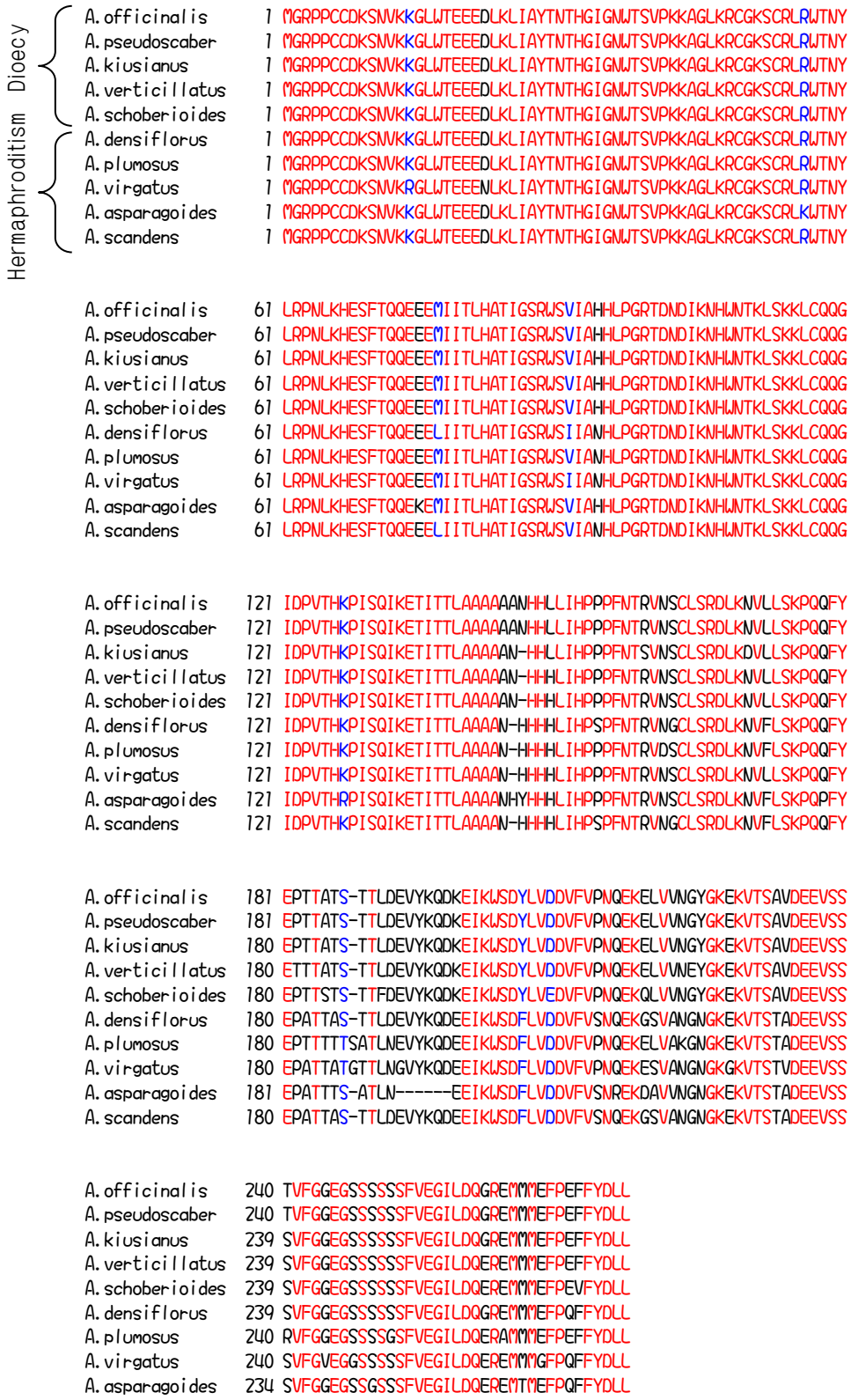


Fig. S10. Alignment of predicted amino-acid sequences of MSE1 orthologues in asparagus and hermaphroditic species.

MSE1 orthologues of hermaphroditic species in genus *Asparagus* were PCR-amplified and sequenced. Colored letters indicate conserved (red) or similar (blue) residues in this alignment.

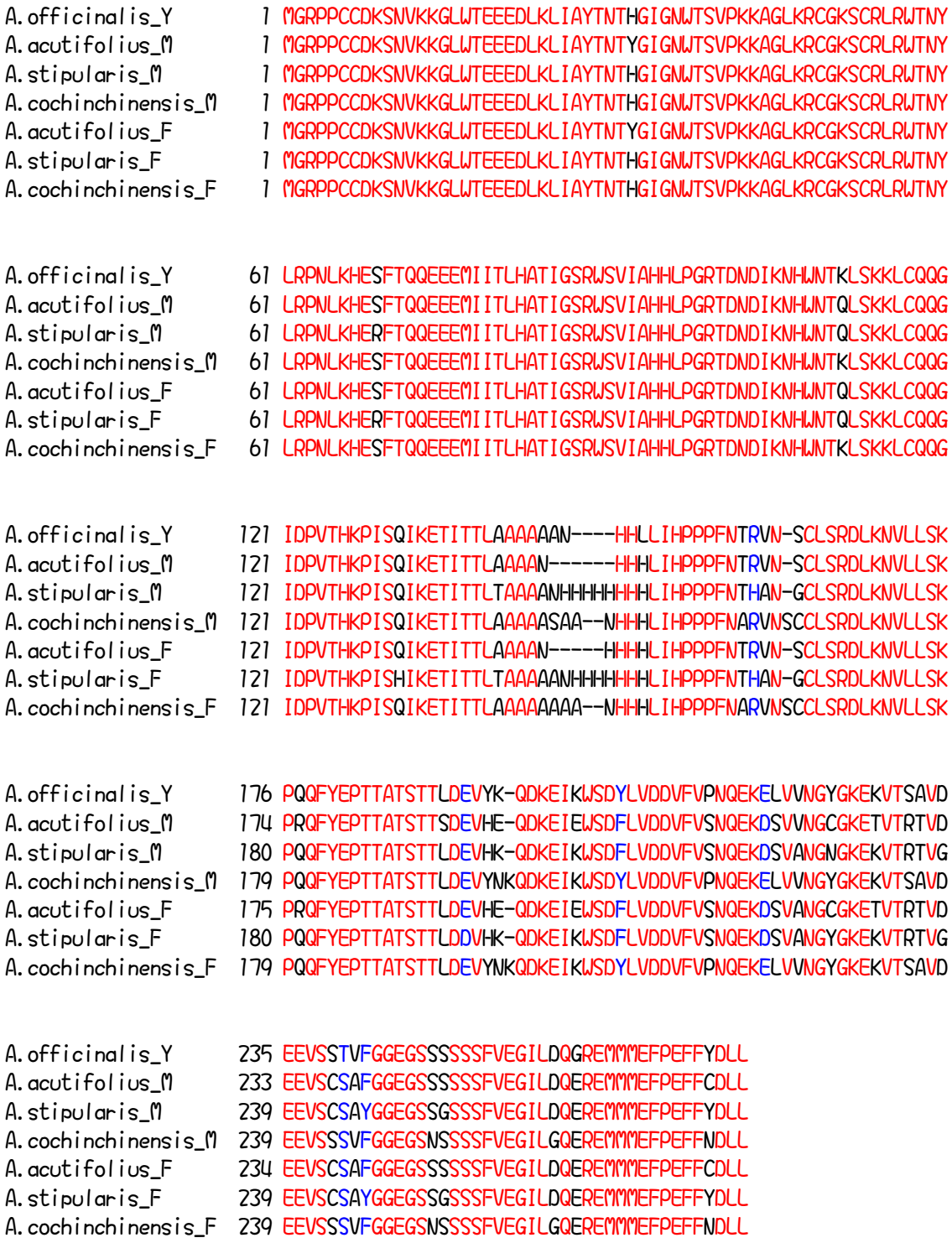


Fig. S11. Alignment of predicted amino-acid sequences of MSE1 orthologues from male and female individuals in three dioecious species that have no deleterious mutation in *MSE1* coding region.

MSE1 genome fragments were PCR-amplified and sequenced from male and female individuals of *A. acutifolius*, *A. stipularis*, and *A. cochinchinensis*. Colored letters indicate conserved (red) or similar (blue) residues in this alignment.

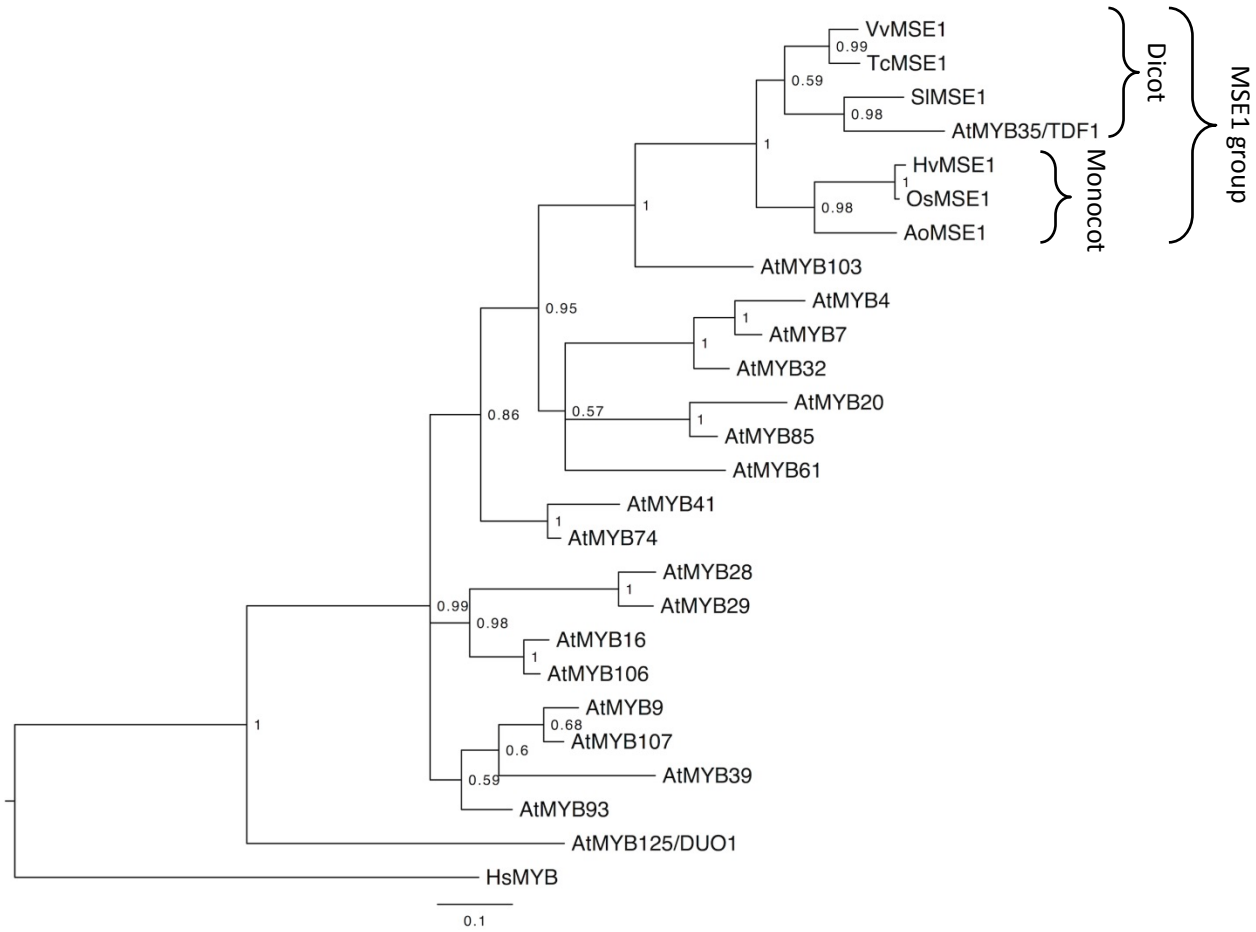


Fig. S12. Phylogeny of MSE1-related MYB domain transcription factors.
The majority-rule consensus tree based on Bayesian inference. Node support values indicate Bayesian posterior probabilities. *Homo sapiens* MYB (HsMYB; accession no. NP_001155129) protein was used as the outgroup. Amino-acid sequences used in this analysis are AoMSE1 (*A. officinalis*), VvMSE1 (grape; accession no. CAN75378), TcMSE1 (cacao; XP_007099739), SlMSE1 (tomato; XP_004234868), HvMSE1 (barley; BAK03933), and OsMSE1 (rice; NP_001173380), and MYB proteins from *A. thaliana* (AtMYB35/TDF1, NP_189488; AtMYB103, NP_200422; AtMYB28, NP_200950; AtMYB41, NP_194540; AtMYB16, NP_197035; AtMYB9, NP_197179; AtMYB107, NP_186944; AtMYB74, NP_192419; AtMYB4, NP_195574; AtMYB32, NP_195225; AtMYB106, NP_186763; AtMYB29, NP_196386; AtMYB93, NP_174726; AtMYB20, NP_176797; AtMYB7, NP_179263; AtMYB39, NP_567540; AtMYB85, NP_567664; AtMYB61, NP_172425; AtMYB125/DUO1, NP_191605).

Table S1. Summary of transcriptome analysis of developing *A. officinalis* flowers

	Male	Female
No. of reads (RNA)	105,588,640	113,110,016
Used reads	98,278,009	105,828,189
No. of contigs	104,937	
No. of unigenes	51,525	
N50 (bp)	1,920	
Male-biased expression*	114	
No. of reads (Genome)	316,018,228	316,258,290
Male-specific genes**	7	
Annotated genes	2	

*No. of tags in male >50, male/female >10

**No. of tags in male >10, male/female >10

Table S2. List of contigs enriched in male genome sequencings

ID	Length (bp)	RNA-seq			Genome reads			Blast2GO annotation
		Male (tag)	Female (tag)	M/F*	Male (tag)	Female (tag)	M/F*	
comp17730_c0	1,030	338	5	72.3	91	5	18.2	transcription factor myb86
comp9180_c0	934	56	0	-	117	0	-	uncharacterized protein LOC102611758
comp21134_c0	968	188	0	-	250	18	13.9	NA**
comp22784_c0	1,041	110	0	-	113	4	33.3	NA
comp29423_c0	747	63	0	-	106	1	106.0	NA
comp24246_c0	469	98	1	104.9	43	0	-	NA
comp22788_c1	232	143	3	51.0	41	1	41.0	NA

*M, male; F, female; M/F values were normalized against the number of mapping reads.
**NA, Not available.