bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Title: Positional identification of a candidate gene for *MALE STERILITY 2* (*MS2*) by linkage mapping and transcriptomic data in *Cryptomeria japonica* D. Don

Saneyoshi Ueno[1*], Yoichi Hasegawa[1], Tokuko Ujino-Ihara[1], Momi Tsuruta[1], Hiroyuki Kakui[2], Junji Iwai[3$], Satoko Hirayama[3$], Katsushi Yamaguchi[4], Shuji Shigenobu[4], Takeshi Fujino[5], Yutaka Suzuki[5], Masahiro Kasahara[5], Yoshinari Moriguchi[6*]

1 Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Forest Research and Management Organization, 1 Matsunosato, Tsukuba, Ibaraki, 305-8687, Japan

2 Institute for Sustainable Agro-ecosystem Services, Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Midoricho, Nishitokyo, Tokyo, 188-0002, Japan

3 Niigata Prefectural Forest Research Institute, 5-2249 Unotoro, Murakami, Niigata, 958-0264, Japan

4 Trans-Scale Biology Center, National Institute for Basic Biology, Okazaki, Aichi, 444-8585, Japan

5 Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561, Japan

6 Graduate School of Science and Technology, Niigata University, 8050 Ikarashi 2 no-cho, Nishi-ku, Niigata, Niigata, 950-2181, Japan

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

*Corresponding authors:

Saneyoshi Ueno

Department of Forest Molecular Genetics and Biotechnology, Forestry and Forest Products Research Institute, Forest Research and Management Organization,

1 Matsunosato, Tsukuba, Ibaraki, 305-8687, Japan

E-mail: saueno@ffpri.affrc.go.jp

Yoshinari Moriguchi

Department of Agriculture, Faculty of Agriculture, Niigata University,

8050 Ikarashi 2-no-cho, Nishi-ku, Niigata, 950-2181, Japan

E-mail: chimori@agr.niigata-u.ac.jp

$Current addresses:

Junji Iwai

Forestry Conservation Division, Department of Agriculture, Forestry and Fisheries, Niigata Prefectural Government, Niigata, Japan

Satoko Hirayama

Forestry Administration Division, Department of Agriculture, Forestry and Fisheries, Niigata Prefectural Government, Niigata, Japan

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Abstract

Background:

Japanese cedar (*Cryptomeria japonica* D. Don) is a major plantation species in Japan, but its abundant pollen production is a primary cause of seasonal allergic rhinitis (pollinosis). To mitigate this public health issue, the use of male-sterile cultivars has been promoted. Five types of recessive male-sterile mutations (*ms1–ms5*) have been identified, and the causal genes and mutations for *MS1* and *MS4* have been elucidated. However, the gene responsible for MS2-type male sterility remains unknown.

Results: We aimed to identify the candidate gene responsible for MS2-type male sterility using a map-based cloning strategy. High-resolution linkage mapping localized the *MS2* locus to a 1.56 cM interval on linkage group 5, corresponding to an 8.64 Mb region of the reference genome. Ninety-one genes in this region were subjected to functional annotation, gene expression analysis, and mutation screening. Among these, a single gene, *SUGI_0493010*, encoding a GDSL-type esterase/lipase protein (GELP), fulfilled all three criteria: it showed homology to pollen development genes in *Arabidopsis thaliana*, was specifically expressed in male strobili, and carried a deleterious amino acid substitution (S40F) within the predicted catalytic domain in *ms2* mutant. The same mutation was also detected in a heterozygous individual (*Ms2/ms2*) from a separate breeding population, whose genotype was confirmed through progeny testing. Structural annotation revealed that the affected serine residue lies within the conserved GDSL motif, suggesting a functional disruption of enzymatic activity.

Conclusions: Our results strongly suggest that SUGI_0493010 (*GELP*) is the causal gene for MS2-type male sterility in *C. japonica*. This finding enhances our understanding of male sterility mechanisms in conifers and provides a valuable genetic resource for breeding pollen-free trees. The study also demonstrates the effectiveness of combining genetic mapping with transcriptomic and mutational data in forest tree genomics.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Keywords

Cryptomeria japonica, Male sterility, GELP, GDSL esterase/lipase, map-based cloning, pollen development, Conifer genomics, pollinosis.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Introduction

Japanese cedar (*Cryptomeria japonica* D. Don), or sugi, is one of the most widely planted tree species in Japan due to its economic importance in the forest industry, occupying about 40 % of artificial forests in Japan [1]. The timber has been used not only in house construction, but also in traditional crafts and daily-use wooden products, including furniture, chopsticks, sake barrels, and shrine architecture. However, large-scale sugi plantations have contributed to a major public health issue: pollen allergy or sugi pollinosis. This seasonal allergy has become increasingly prevalent, with nationwide surveys indicating that nearly 40% of the population was affected in 2019 [2]. The primary source of the allergen is the massive production of pollen from male strobili during the spring season, varying widely among studies, ranging from approximately 100,000 to 656,000 grains per single male strobilus depending on the clone and environmental conditions [3–6]. In order to reduce pollen emissions at the source, increasing attention has been directed toward the utilization of male-sterile cultivars in afforestation.

Genetic characterization of male-sterile mutants in *C. japonica* has provided a solid foundation for this approach. Recessive mutations causing male sterility have been classified into five types (MS1–MS5) based on the specific developmental stage at which pollen formation is disrupted. These classifications were established through detailed cytological observations [7–10] and linkage analyses using controlled crosses [11–13]. Among these, the MS1-type has been the most extensively studied [8, 14, 15] and has already been utilized in public afforestation initiatives [16]. Pollen-free sugi trees homozygous for the *ms1* allele (*ms1/ms1*) have been propagated and planted to reduce airborne pollen emission from sugi forests. Recent molecular studies [17, 18] have further advanced our understanding of the genetic basis of male sterility in *C. japonica*, particularly for *MS1* and *MS4*. *MS1*, which is associated with complete pollen abortion at the tetrad stage, has been linked to a gene involved in lipid transportation. Similarly, *ms4*, which disrupts pollen formation during the microspore stage, has been associated with a gene orthologous to *Arabidopsis thaliana TKPR1*, a key enzyme in sporopollenin biosynthesis [19]. These findings have provided valuable insights into the molecular mechanisms underlying pollen development and laid the foundation for marker-assisted

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

selection of male-sterile trees in breeding programs [20, 21].

In contrast, the molecular basis of MS2-type male sterility remains unresolved. *MS2* is characterized by developmental abnormality from the tetrad stage, as previously reported in [10] (Figure 1) and was first identified in a naturally occurring mutant known as 'Shindai-1,' selected from an artificial forest in Niigata Prefecture in 1998 [9]. Previous studies have localized the *MS2* locus to linkage group 5 (LG5) [11, 22], but the causal gene has not yet been identified. Understanding the genetic mechanism underlying *MS2* is essential not only for expanding the tools for marker-assisted breeding but also for deepening our knowledge of reproductive development in conifers.

In this study, we aimed to identify the candidate gene responsible for MS2-type male sterility using a map-based cloning approach. By integrating high-resolution linkage mapping, RNA-Seq-based expression profiling, and mutation screening with functional prediction, we successfully narrowed down the candidate region and identified a GDSL-type esterase/lipase gene (*GELP*) as the most likely causal gene. Our findings provide important insights into the molecular basis of male sterility in *C. japonica* and offer a genetic resource for the practical development of pollen-free sugi trees.

Materials and Methods

Plant materials

The MS2-type male-sterile individual 'Shindai-1' was originally selected from an artificial forest in Niigata Prefecture, Japan [9]. This individual was used as the seed parent to produce a mapping family (S1-2), which was generated by crossing 'Shindai-1' (*ms2*/*ms2*) and 'S1NK4' (*Ms2*/*ms2*) as female and male parent, respectively [22]. The 'S1NK4' is an offspring (F1) resulting from a cross between 'Shindai-1' and 'Nakakubiki-4.' Among the offspring of this family, 129 individuals were used for refinement of the linkage map in our previous study [22]. For the fine

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

mapping of the *MS2* locus, we selected 95 of these 129 individuals (31 male-sterile and 64 male-fertile). Because the *MS2* male-sterile phenotype can be difficult to distinguish in some individuals, only those with clearly identifiable phenotypes were included in the analysis. Phenotypes were determined based on field observations of male strobili and microscopic examination of dissected male strobili. In addition, another individual from Niigata Prefecture, 'Gosenshi-1,' which had been previously identified as *Ms2/ms2* through controlled crossing experiments [23], was included to evaluate the consistency of the candidate mutation identified in this study.

RNA-Seq data collection and analysis

Male strobili were collected from fertile and sterile individuals of the S1-2 family growing at Chiyoda nursery, FFPRI (36.184269° N, 140.217564° E) at two time points (October 17th and November 6th, 2019) for RNA-Seq analysis (Additional file 1). The male strobilus samples were immediately frozen in liquid N_2 at the field, and stored at −80°C until RNA extraction. The total RNA was extracted from the male strobili using the modified cetyltrimethylammonium bromide (CTAB) method [24]. RNA-Seq was performed following Macrogen's regular workflow (Macrogen Japan Corp., Kyoto, Japan). Briefly, high-throughput sequencing libraries were constructed using the TruSeq stranded mRNA Library Kit (Illumina, San Diego, CA, USA) following the quality control of RNA. Each library was sequenced by the Illumina platform (NovaSeq6000) at 100 bp paired-end reads (on average 7.3 Gbp per sample). Raw reads were processed using fastp [25] to remove adapter sequences and low-quality regions. Cleaned reads were aligned using bwa mem (ver. 0.7.17-r1188) with default parameters. Depending on the timing of data generation, either the *Cryptomeria japonica* (SUGI_1) permissive gene set [26] or the earlier transcript reference CJ3006NRE [27] was selected as the mapping reference. BAM files aligned to the SUGI_1 permissive gene set were subsequently utilized as a common resource for downstream analyses—including SNP identification, marker development, gene expression analysis (including DEG analysis), and detection of deleterious mutations—all of which focused on the SUGI_1 standard gene set [26] consisting of 55,246 annotated genes.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Marker development using RNA-Seq data

Based on our previous linkage map [22], the *MS2* locus was initially localized to three genomic contigs—ctg668, ctg930, and ctg820—in the *C. japonica* draft genome ver. 0.1 (Fujino et al., unpublished) (Additional file 2). At that time, a chromosome-scale genome assembly and gene annotations were not yet available. Therefore, we identified genes on these contigs by performing BLASTN searches using CJ3006NRE reference transcript sequences from our previous study [27] as queries. To improve the accuracy of gene models and prioritize polymorphisms, spliced alignments were performed using Splign [28].

Polymorphisms between fertile and sterile individuals were identified using RNA-Seq data. BAM files were grouped by phenotype (fertile or sterile), and SNPs were called using VarScan (v2.4.3) [29, 30] with default parameters, generating VCF files for downstream analysis. Within each target gene, consensus sequences incorporating heterozygous sites were created using 'bcftools consensus' (v1.9) [31] with IUPAC ambiguity codes. These sequences were then post-processed to fit the input format of Fluidigm SNPType Assay software D3 (Standard Biotools).

SNP genotyping and fine mapping

Candidate SNPs were evaluated using D3 to determine whether functional primer pairs could be designed. Only SNPs that passed the D3 design criteria were selected for genotyping. Genotyping was carried out using the Fluidigm EP1 system with a 48.48 Dynamic Array according to the manufacturer's instructions. Genotypes were called with Fluidigm SNP Genotyping Analysis software (v4.5.1). The resulting genotype data were used to construct a refined linkage map. Linkage analysis was conducted using the maximum likelihood mapping algorithm in JoinMap ver. 5 (Kyazma, Wageningen, The Netherlands), assuming a backcross (BC)-type population (nn × np), as previously described [22]. Once a chromosome-scale genome assembly (SUGI_1) became available [26], the newly developed markers were mapped onto chromosome 5 (chr5).

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Candidate gene identification by functional annotation

To identify candidate genes within the *MS2* region described above, nucleotide sequences corresponding to the standard gene set [26] in this region (Additional file 3) were subjected to BLASTX searches (E-value cutoff: 1e-5) against the TAIR10 protein database of *Arabidopsis thaliana* [32]. Genes in *C. japonica* showing homology to 372 *Arabidopsis* gene loci annotated with GO:0009555 (pollen development) were selected as functional (homology-based) candidates.

Gene expression analysis for candidate gene prioritization

Gene expression analysis was performed to identify candidate genes expressed in fertile individuals within the 8.64 Mb interval on chromosome 5. Gene expression levels were quantified using transcripts per million (TPM) values calculated from the RNA-Seq data using featureCount [33], based on BAM files aligned to the SUGI_1 permissive gene set. TPM values were calculated across all annotated transcripts, and only genes corresponding to the SUGI_1 standard gene set (55,246 genes) were retained for downstream analysis. Genes with TPM values > 1 in at least one fertile individual were considered expressed and retained as MS2 candidate genes. Differential gene expression analysis was conducted using DESeq2 [34] with raw count data from the SUGI_1 permissive gene set. For interpretation, we focused on the MS2-neighboring genes. Genes with an adjusted p-value < 0.01 were considered significantly differentially expressed.

Identification of deleterious mutations and validation in additional material

To investigate whether deleterious mutations were responsible for MS2-type male sterility, polymorphisms detected within candidate genes were evaluated using PROVEAN (v1.1.3) [35]. Variant information from the previously generated VCF file was used to create sterile (ALT) haplotype sequences by using ‘bcftools consensus’ (ver. 1.9) [31], translated into protein sequences

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

by transeq in EMBOSS suite [36], and the resulting amino acid substitutions were analyzed for their functional effects using PROVEAN (v1.1.3). Variants with a PROVEAN score below –2.5 were classified as deleterious. Candidate genes with deleterious mutations were prioritized for further investigation.

To further validate the presence of the identified mutation in an independent genetic background, we also analyzed an additional heterozygous individual, 'Gosenshi-1' (*Ms2*/*ms2*), from which male strobili were collected at Niigata Prefectural Forest Research Institute on 22 October 2020 for RNA extraction. RNA extraction and mutation detection for 'Gosenshi-1' were performed using the same methods as described above.

Structural, functional, and expression profiling of candidate gene

The gene structure of a candidate gene (SUGI_0493010) was determined using the SUGI_1 genome assembly and gene models from the standard gene set [26]. To refine transcriptional boundaries, full length cDNA sequence (DDBJ accession number: FX344585) and Iso-Seq reads for male strobili from our previous studies [27, 37] were mapped using minimap2 [38] with options (-ax splice -uf -k14 --secondary=no) and visualized in CLC Genomics Workbench ver. 20 (Qiagen). Transcription start site (TSS) was inferred based on sharp increases in read coverage, and on the prediction by TSSPlant [39] with input of a 3.4-kb upstream sequence from the start codon. The 5′UTR was defined as the region between the TSS and the annotated start codon, while the 3′UTR was defined as the region between the stop codon and the major polyadenylation site inferred from Iso-Seq read ends. The predicted amino acid sequence (364 residues) was annotated using InterPro webtool [40].

In order to get insights into expressional variation among different tissues for the MS2 candidate gene (SUGI_0493010), CPM (count per million) data were downloaded from the SugiExDB database [41], which contains RNA-Seq expression profiles for various *C. japonica* tissues. Expression levels were visually compared across tissues to explore tissue-preferential

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

226 expression tendencies.

227

228 Results

229 Fine mapping of the *MS2* locus

230 To refine the localization of the *MS2* gene, we constructed a high-resolution linkage map
231 using the mapping population. The locus was positioned between two flanking markers, CJt005282-
232 890 and CJt113083-201, spanning an 8.64 Mb region on chromosome 5 (Figure 2 and Additional file
233 2). RNA-Seq-derived markers served as anchor markers, bridging the genetic linkage map and the
234 physical genome map, and enabled accurate localization of genes within the MS2 locus and
235 supporting downstream candidate gene analysis. Within this region, 91 predicted genes from the
236 standard gene set (55,246 genes) were identified and used for downstream analysis (Additional file 3).

237 Identification of candidate genes related to pollen development

238 To prioritize functionally relevant candidates, we conducted a homology search against 372
239 *A. thaliana* genes associated with the GO term “pollen development” (GO:0009555), using the TAIR
240 database. A BLASTX search for the nucleotide sequences of *C. japonica* candidate genes and the
241 *Arabidopsis* gene set revealed that 11 out of the 91 MS2-neighboring genes showed significant
242 sequence similarity to pollen development-related genes (Additional file 3).

243 Expression profiling in male strobili

244 To determine whether the candidate genes were expressed in male reproductive organs, RNA-
245 Seq analysis was performed using male strobili collected in October and November from five fertile
246 and three sterile individuals of the S1-2 family (Additional file 1). An average of 7.4 Gb of RNA-Seq

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

data was obtained per sample. The reads were mapped to the SUGI_1 permissive gene set, and transcript abundance was quantified using TPM values. Among the 91 MS2-neighboring genes, 29 genes were expressed (TPM > 1) in at least one of the fertile strobilus samples (Figure 3 and Additional file 3), and all of these genes were in the SUGI_1 standard gene set, ensuring that highly expressed transcripts were not omitted from the expression analysis. Comparison of expression levels between fertile and sterile individuals revealed no statistically significant differences ($p > 0.01$) for these MS2 candidate genes. This suggests that if MS2 is caused by a mutation in one of these genes, it likely does not result from transcriptional suppression in sterile individuals.

Detection and validation of deleterious mutations

We next examined whether deleterious mutations were present in the coding regions of the 29 expressed genes. Variant calling was conducted using RNA-Seq data, and amino acid substitutions were evaluated using PROVEAN to assess their functional impact. Three deleterious amino acid substitutions (PROVEAN score ≤ –2.5) were identified in two genes (Additional file 3). Notably, one gene (CJt112073 or SUGI_0493010), encoding a GDSL-type esterase/lipase (GELP), contained a substitution from serine to phenylalanine at position 40 (p.S40F) located within the predicted enzymatic active site (Figure 4). The other two mutations were found in a gene homologous to *Arabidopsis INTS1* (SUGI_0492850), which is not known to be involved in pollen development and did not meet the homology-based candidate criteria (Figure 3, Additional file 3). Furthermore, RNA-Seq analysis showed that the *INTS1* mutation was absent in ‘Gosenshi-1’ (Additional file 4), allowing this gene to be excluded from the list of mutational candidates.

To validate the presence of the *GELP* mutation in another genetic background, we examined ‘Gosenshi-1,’ an individual previously confirmed to carry the *MS2* mutation in a heterozygous state (*Ms2*/*ms2*) through controlled crossing experiments [23]. RNA-Seq-based analysis confirmed that the same S40F substitution was present in ‘Gosenshi-1’ (Additional file 4). The consistent presence of this mutation in both *ms2*/*ms2* and *Ms2*/*ms2* individuals supports the hypothesis that *GELP* is the most likely causal gene for MS2-type male sterility in *C. japonica*.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Structural, functional and expression profiling of GELP

Structural annotation of the *GELP* gene (SUGI_0493010) revealed five exons and four introns (Figure 4, top). TSSPlant predicted multiple transcription start sites (TSSs) within the upstream region of the gene which exceeded the prediction threshold. Among these, one TSS candidate was fully supported by the full-length cDNA sequence obtained by biotinylated CAP trapper method [37]; this site was therefore considered the most likely transcription initiation site. The 5′ untranslated region (UTR) was defined as the 19-bp region between this TSS and the start codon. In contrast, Iso-Seq reads aligned one base upstream of the predicted TSS, likely due to non-templated nucleotide addition by the reverse transcriptase during cDNA synthesis, which can introduce a non-templated nucleotide at the 5′ end [42]. Taken together, these observations support the accuracy of the TSS identified by TSSPlant and the full-length cDNA.

The 3′ end of the transcript was inferred based on the termination pattern of Iso-Seq reads downstream of the stop codon. The majority of reads converged at a position where read depth reached a local minimum, approximately 28 bp downstream of a canonical polyadenylation (polyA) signal (AATAAA), and was followed by a plateau in coverage. This position likely represents the major polyadenylation cleavage site. Additionally, a minor cleavage site was suggested further downstream, approximately 26 bp beyond a secondary polyA signal.

Functional annotation of the predicted protein sequence (364 amino acids) identified a conserved GDSL esterase/lipase domain (PF00657), belonging to the SGNH hydrolase superfamily (IPR036514). The catalytic triad—Ser40, Asp329, and His332—was conserved, and the S40F missense mutation identified in *ms2/ms2* individuals lies within this catalytic region (Figure 4, middle). Notably, the affected serine (S40) is not only a part of the predicted catalytic triad (S40, D329, H332) but also corresponds to the conserved serine residue within the GDSL motif (Gly–Asp–Ser–Leu), which defines this esterase/lipase family. Substitution of this critical serine to

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

phenylalanine (S40F) is therefore expected to abolish enzymatic activity. To confirm the specificity of this mutation, haplotype sequences spanning the GELP coding region were reconstructed from RNA-Seq data for both fertile and sterile individuals in the S1-2 family. Among all identified variants, only a single nucleotide substitution (c.119C>T) corresponding to the S40F amino acid change was found, indicating that this mutation is the sole coding difference between the two haplotypes. These findings strongly support the role of SUGI_0493010 as the causal gene underlying MS2-type male sterility.

In addition to the structural and functional annotations, tissue-specific expression profiles were examined using RNA-Seq data from 13 *C. japonica* tissues available in SugiExDB [41]. CPM values were visually compared across tissues to explore tissue-preferential expression tendencies. GELP exhibited strong and specific expression in male strobili, with no detectable expression in other tissues, including needles, cambium, roots, and ovulate strobili (Figure 4, bottom). This expression pattern supports the involvement of GELP in male reproductive development.

Discussion

In this study, we identified a strong candidate gene for the MS2-type male sterility in *Cryptomeria japonica* through a comprehensive approach that combined genetic linkage mapping, gene expression analysis, and functional prediction of amino acid substitutions. Our results suggest that a GDSL-type esterase/lipase gene (*GELP*), which harbors a deleterious amino acid substitution, is the most likely causal gene for MS2-type male sterility.

The candidate *GELP* gene resides within the 1.56 cM interval surrounding the *MS2* locus on linkage group 5, corresponding to a 8.64 Mb region of the reference genome. Among the 91 genes in this interval, *GELP* was one of 11 genes with homology to *Arabidopsis* genes involved in pollen development and one of 29 genes expressed in male strobili. Expression analysis showed that *GELP* is specifically expressed in male strobili (Figure 4), and is present in both fertile and sterile

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

individuals (Figure 2). This supports the idea that the sterility phenotype arises not from a lack of expression, but from a functional disruption caused by the amino acid substitution. Notably, the gene exhibited a serine-to-phenylalanine substitution (S40F) within a predicted active site (Figure 4), which was predicted by PROVEAN to be functionally deleterious. Although the deleterious effect of the S40F substitution was computationally predicted by PROVEAN, the affected serine residue is part of the conserved catalytic triad of GDSL esterases/lipases. This serine hydroxyl group is essential for nucleophilic attack during catalysis, and its substitution with a bulky, nonpolar phenylalanine is expected to abolish enzymatic activity. Furthermore, the same mutation was identified (Additional file 4) in a heterozygous (*Ms2*/*ms2*) individual, 'Gosenshi-1,' which was derived from an independent genetic background, indicating that this mutation is consistently associated with MS2-type male sterility in *C. japonica*.

GDSL esterases/lipases have been implicated in a variety of biological processes, including lipid metabolism, cuticle formation, and reproductive development in plants [43–46]. In *Arabidopsis thaliana*, mutations in orthologous genes have been shown to cause male sterility due to defects in pollen wall formation [47]. Given the role of sporopollenin and lipids in pollen wall structure, it is plausible that the *GELP* gene in sugi plays a similar role in pollen development. Supporting this hypothesis, several morphological characteristics observed in *ms2* mutant individuals are consistent with those reported in *A. thaliana*. Loss-of-function mutations in orthologous GDSL genes result in abnormal pollen wall formation, irregularly shaped microspores, early microspore degeneration, and partial or complete nuclear loss [46]. Notably, clumping or high-density aggregation of microspores—likely due to defects in post-meiotic microspore separation—has also been observed in MS2-type male sterility in *C. japonica*, where densely packed and irregular microspores, some lacking nuclei, are found [9, 10]. These shared phenotypes across species strongly support the idea that GELP plays a conserved and essential role in male gametophyte development, likely through contributing to proper microspore maturation.

Despite this collective evidence supporting *GELP* as the candidate gene, several questions remain to be resolved. Microscopic observations revealed that some individuals classified as sterile

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

based on genetic and phenotypic data still produced visible pollen grains (Additional file 5), confirming that the *ms2*/*ms2* genotype may not always result in complete pollen abortion [10]. One possible explanation is the presence of a neighboring gene, SUGI_0493000, which encodes another GDSL esterase/lipase protein and is located adjacent to *GELP* (SUGI_0493010). Although the expression level of SUGI_0493000 is relatively low (Figure 2 and Additional file 3), it may exert a weak compensatory effect under certain conditions, thereby allowing partial pollen development in *ms2/ms2* individuals. Alternatively, this inconsistency may be due to environmental influences, measurement variability, or the presence of modifier loci (small-effect QTLs) [48–50]. Furthermore, the segregation ratio of fertile and sterile individuals in the S1-2 family deviated from the expected 1:1 Mendelian ratio. This may indicate viability effects or segregation distortion related to the MS2 locus, possibly due to impaired pollen function or reduced viability of early *ms2*/*ms2* embryos. These observations highlight the need for further studies to clarify the penetrance and expressivity of the *MS2* mutation.

To definitively confirm the causal role of the *GELP* gene in MS2-type male sterility, functional validation is necessary. Genome editing technologies such as CRISPR/Cas9, which have recently been successfully applied to *C. japonica* [51, 52], could be employed to introduce or correct the identified mutation. This would allow direct assessment of the gene's role in pollen development and provide a powerful tool for breeding programs aimed at producing pollen-free sugi trees. Additionally, broader surveys across natural and breeding populations may reveal further allelic diversity at the MS2 locus, similar to findings from our previous study [17] that analyzed the MS1 locus using amplicon sequencing and haplotype network construction.

The identification of *GELP* as a candidate gene for MS2-type male sterility represents a significant step toward understanding the molecular basis of male sterility in *C. japonica* and developing effective strategies for reducing pollen production. This research contributes to both fundamental knowledge of conifer reproduction and practical efforts to mitigate the growing public health issue of cedar pollen allergy.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Conclusions

In this study, we identified a strong candidate gene for *MALE STERILITY 2* (*MS2*) in *Cryptomeria japonica* using a map-based approach that integrated high-resolution linkage mapping, transcriptome profiling, and mutation analysis. The *MS2* locus was localized to a 8.64 Mb region on chromosome 5, within which a GDSL-type esterase/lipase gene (*GELP*) was identified as the most likely causal gene. A deleterious amino acid substitution (S40F) was detected in sterile individuals and validated in an independent heterozygous line, supporting the association between the mutation and the *MS2* sterility phenotype. These findings provide a key genetic resource for breeding pollen-free *C. japonica* trees, which are of increasing social importance in the context of allergenic pollen mitigation. Moreover, the identification of *GELP* as a candidate gene adds to our understanding of the molecular mechanisms underlying male sterility in conifers and highlights the utility of integrative genomic approaches in forest tree research.

List of abbreviations

CDS: Coding Sequence; DEG: Differentially Expressed Gene; GELP: GDSL esterase/lipase protein; GO: Gene Ontology; INTS1: Integrator Complex Subunit 1; kb: kilobase pairs; LG: Linkage Group; Mb: megabase pairs; MS: Male Sterility; QTL: Quantitative Trait Locus; RNA-Seq: RNA Sequencing; SNP: Single Nucleotide Polymorphism; TPM: Transcripts Per Million; TSS: Transcription Start Site; UTR: Untranslated Region.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The RNA-Seq Illumina reads were deposited in the Sequence Read Archive database of the DNA Data Bank of Japan (DDBJ) under accession number DRR733425–DRR733440. All additional datasets and materials used during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work has been supported in part by JSPS KAKENHI Grant Numbers JP16H06279 (PAGS), JP20H03239, JP23K26956 and JP23H02263, the Ministry of Agriculture, Forestry and Fisheries' "Agriculture, Forestry and Fisheries and Food Industry Science and Technology Research

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

414 Promotion Project" and the research program on development of innovative technology grants
415 (JPJ007097) from the Project of the Bio-oriented Technology Research Advancement Institution
416 (BRAIN) (Project ID 28013BC), and NIBB Collaborative Research Program (15-829, 16-403, 17-
417 405, 18-408, 19-420, 20-428, 21-302, 22NIBB402, and 23NIBB405).

418 Authors' contributions

419 SU and YM conceived and designed the study. YH and MT performed the experiments. YH
420 conducted the linkage analysis. HK proposed the strategy for the phenotypic evaluation of male
421 strobili. SU performed the bioinformatic analyses. KY, SS, and YS generated early sequencing
422 datasets. TF and MK assembled the draft genome, which served as the reference for the initial linkage
423 mapping. TU contributed to discussions and provided critical comments throughout the research
424 process. JI and SH established the mapping population and provided the breeding material. SU wrote
425 the initial draft of the manuscript with input from all authors. All authors read and approved the final
426 manuscript.
427

428 Acknowledgements

429 A part of bioinformatic analysis was carried out by the supercomputer of AFFRIT, MAFF,
430 Japan. The Arboretum and Nursery Office, FFPRI was involved in the maintenance of the *C. japonica*
431 nursery. We thank Eriko Tsurisaki and Nana Matsumura of Niigata University for photographs of
432 male strobili and light microscopic images for wild-type and male-sterile *C. japonica*. Yasuyuki
433 Komatsu and Nozomi Ohmiya, FFPRI, performed DNA and RNA extraction, respectively. ChatGPT
434 (OpenAI) was used to suggest improvements in grammar and phrasing during the writing process. All
435 scientific content, data analysis, and interpretations were solely developed and validated by the
436 authors.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Figure legends

Figure 1. Comparison of male strobili and tetrad stage pollen between fertile and MS2-type male-sterile individuals of *Cryptomeria japonica*.

Representative images showing the morphological and cytological differences between wild-type (*Ms2*/*Ms2*) and MS2-type male-sterile (*ms2*/*ms2*) individuals. **Left**: Cross-sections of male strobili. **Right**: Light microscopic image of pollen at the tetrad stage. In fertile individuals, pollen development proceeds normally, while in sterile individuals ('Shindai-1'), tetrads exhibit developmental abnormality with clumping and dense aggregation. Scale bars for the cross-sections are estimated based on the approximate size of the male strobilus (~4 mm) and are intended for reference only.

Figure 2. Fine mapping of the MS2 locus and expression profiles of candidate genes.

Top: Partial linkage map of linkage group 5 (LG5) showing the MS2 locus positioned between two flanking markers. **Middle**: The corresponding 8.64 Mb physical interval on chromosome 5 (Chr5), with vertical lines indicating the positions of 91 positional candidate genes. **Bottom**: Heatmap showing expression levels (TPM) of the 91 genes in male strobili samples collected in October and November from both fertile (WT) and sterile (MS) individuals. Red dots indicate genes functionally annotated as homologs of *Arabidopsis* genes involved in pollen development.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Figure 3. Summary of the candidate gene (CG) screening process, illustrated as a Venn diagram.

A total of 91 positional candidate genes (Pos. CG) were initially identified based on genome sequence and linkage analysis. Among these, 11 genes were functionally annotated as homologs of *Arabidopsis* genes involved in pollen development (Homology-based candidate genes; Hom. CG), 29 genes were expressed in male strobili based on RNA-Seq data (Expressional candidate genes; Exp. CG), and 2 genes contained predicted deleterious mutations (Mutational candidate genes; Mut. CG). Only one gene, a GDSL esterase/lipase protein (*GELP*), met all three criteria and was thus identified as the most likely candidate gene responsible for MS2-type male sterility in *C. japonica*. Note: One of the two Mutational CGs (homologous to *Arabidopsis INST1*) does not fall into the Hom. CG category.

Figure 4. Structural, functional, and expression profiling of the *GELP* gene.

Top: Exon-intron structure of the *GELP* gene (SUGI_0493010). Coding sequences (CDS) are shown in blue boxes, untranslated regions (UTRs) in green, and introns as lines connecting exons. Numbers bellow the exons indicate the relative positions of the CDS nucleotides. The position of the MS2-associated SNP (c.119C>T, TCT>TTT) is indicated with an arrow, corresponding to a missense mutation in the first exon. **Middle**: Schematic representation of the predicted protein encoded by the *GELP* gene based on InterPro annotation. The signal peptide (yellow), GDSL esterase/lipase domain (light orange), and catalytic triad residues (orange bars) are indicated. The amino acid substitution (p.S40F) caused by the SNP is located within the N-terminal region of the conserved GDSL domain. Only one coding SNP (c.119C>T) leading to the S40F substitution was identified between fertile and sterile haplotypes reconstructed from RNA-Seq data in the S1-2 family. **Bottom**: Expression profile of *GELP* across 13 tissues of C. japonica based on RNA-Seq data from SugiExDB [41]. Counts per million (CPM) values are shown as boxplots. *GELP* exhibits strong and specific expression in male

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

strobili, with no detectable expression in other tissues, including needles, cambium, roots, and ovulate strobili.

References

1. Forestry Agency, Japan. State of Japan's Forests and Forest Management. Forestry Agency, Japan; 2019.

2. Matsubara A, Sakashita M, Gotoh M, Kawashima K, Matsuoka T, Kondo S, et al. Epidemiological Survey of Allergic Rhinitis in Japan 2019. Nippon Jibiinkoka Gakkai Kaiho. 2020;123:485–90.

3. Kakui H, Tsurisaki E, Shibata R, Moriguchi Y. Factors affecting the number of pollen grains per male strobilus in Japanese cedar (*Cryptomeria japonica*). Plants. 2021;10:856.

4. Ikuse M. On the number of pollen grains contained in anthers of some plants. Quat Res (Daiyonki-Kenkyu). 1965;4:144–9.

5. Kondo T, Senda M, Tabuchi K. Pollen grain number in a male flower in Sugi (*Cryptomeria japnica*). Nichirinron. 1992;103:329–30.

6. Saito H, Takeoka M. Production rates of the reproductive organ and relationship between pollen and seed production in the forests of sugi (*Cryptomeria japonica*) of Japan Sea type (in Japanese with English Abstract). Jap J Ecol. 1987;37:183–95.

7. Miyajima D, Yoshii E, Hosoo Y, Taira H. Cytological and genetic studies on male sterility in *Cryptomeria japonica* D. Don (Shindai 8) (In Japanese). J Jpn For Soc. 2010;92:106–9.

8. Saito M, Taira H, Furuta Y. Cytological and Genetical Studies on Male Sterility in *Cryptomeria japonica* D. Don. J Forest Res. 1998;3:167–73.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

503 9. Yoshii E, Taira H. Cytological and Genetical Studies on Male Sterile Sugi (*Cryptomeria japonica*
504 D. Don), Shindai 1 and Shindai 5. J Jpn For Soc. 2007;89:26–30.

505 10. Tsurisaki E, Nameta M, Shibata S, Hirayama S, Iwai J, Ohashi R, et al. Cytological analysis of
506 male-sterile MS5 Japanese cedar (*Cryptomeria japonica* D. don) and comparison with other male-
507 sterile mutants. J Plant Biol. 2024;67:11–23.

508 11. Moriguchi Y, Ueno S, Higuchi Y, Miyajima D, Itoo S, Futamura N, et al. Establishment of a
509 microsatellite panel covering the sugi (*Cryptomeria japonica*) genome, and its application for
510 localization of a male-sterile gene (ms-2). Mol Breed. 2014;33:315–25.

511 12. Moriguchi Y, Uchiyama K, Ueno S, Ujino-Ihara T, Matsumoto A, Iwai J, et al. A high-density
512 linkage map with 2560 markers and its application for the localization of the male-sterile genes *ms3*
513 and *ms4* in *Cryptomeria japonica* D. Don. Tree Genet Genomes. 2016;12:57.

514 13. Moriguchi Y, Ujino-Ihara T, Uchiyama K, Futamura N, Saito M, Ueno S, et al. The construction
515 of a high-density linkage map for identifying SNP markers that are tightly linked to a nuclear-
516 recessive major gene for male sterility in *Cryptomeria japonica* D. Don. BMC Genomics. 2012;13:95.

517 14. Futamura N, Igasaki T, Saito M, Taira H, Shinohara K. Comparison of fertile and sterile male
518 gametogenesis in *Cryptomeria japonica* D. Don. Tree Genet Genomes. 2019;15:30.

519 15. Tsubomura M, Kurita M, Watanabe A. Determination of male strobilus developmental stages by
520 cytological and gene expression analyses in Japanese cedar (*Cryptomeria japonica*). Tree Physiol.
521 2016;36:653–66.

522 16. Saito M. Comprehensive study on development of male sterile *Cryptomeria japonica* cultivars (in
523 Japanese with English abstract). Japanese Journal of Palynology. 2019;65:1–9.

524 17. Hasegawa Y, Ueno S, Wei F-J, Matsumoto A, Uchiyama K, Ujino-Ihara T, et al. Identification
525 and genetic diversity analysis of a male-sterile gene (*MS1*) in Japanese cedar (*Cryptomeria japonica*
526 D. Don). Sci Rep. 2021;11:1496.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

527 18. Kakui H, Ujino-Ihara T, Hasegawa Y, Tsurisaki E, Futamura N, Iwai J, et al. A single-nucleotide
528 substitution of *CjTKPR1* determines pollen production in the gymnosperm plant *Cryptomeria*
529 *japonica*. PNAS Nexus. 2023;2:gad236.

530 19. Grienenberger E, Kim SS, Lallemand B, Geoffroy P, Heintz D, Souza C de A, et al. Analysis of
531 TETRAKETIDE α-PYRONE REDUCTASE function in *Arabidopsis thaliana* reveals a previously
532 unknown, but conserved, biochemical pathway in sporopollenin monomer biosynthesis. Plant Cell.
533 2010;22:4067–83.

534 20. Watanabe M, Ueno S, Hasegawa Y, Hirayama S, Iwai J, Kakui H, et al. Development and
535 application of a KASP marker for marker-assisted selection against the male-sterile gene *MALE*
536 *STERILITY 4* (*MS4*) in Japanese cedar (*Cryptomeria japonica* D. Don). New Forests. 2024;55:1349–
537 62.

538 21. Moriguchi Y, Ueno S, Hasegawa Y, Tadama T, Watanabe M, Saito R, et al. Marker-Assisted
539 Selection of Trees with *MALE STERILITY 1* in *Cryptomeria japonica* D. Don. Forests. 2020;11:734.

540 22. Hasegawa Y, Ueno S, Matsumoto A, Ujino-Ihara T, Uchiyama K, Totsuka S, et al. Fine mapping
541 of the male-sterile genes (*MS1*, *MS2*, *MS3*, and *MS4*) and development of SNP markers for marker-
542 assisted selection in Japanese cedar (*Cryptomeria japonica* D. Don). PLoS One. 2018;13:e0206695.

543 23. Hirayama S, Iwai J, Higuchi Y, Kaneko T, Moriguchi Y. Selection of trees with male sterile genes
544 except for *MALE STERILITY 1* in *Cryptomeria japonica* D. don (In Japanese). J Jpn For Soc.
545 2021;103:161–7.

546 24. Ueno S, Uchiyama K, Moriguchi Y, Ujino-Ihara T, Matsumoto A, Wei F-J, et al. Scanning RNA-
547 Seq and RAD-Seq approach to develop SNP markers closely linked to *MALE STERILITY 1* (*MS1*) in
548 *Cryptomeria japonica* D. Don. Breed Sci. 2019;69:19–29.

549 25. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
550 Bioinformatics. 2018;34:i884–90.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

551 26. Fujino T, Yamaguchi K, Yokoyama TT, Hamanaka T, Harazono Y, Kamada H, et al. A
552 chromosome-level genome assembly of a model conifer plant, the Japanese cedar, *Cryptomeria*
553 *japonica* D. Don. BMC Genomics. 2024;25:1039.

554 27. Wei F-J, Ueno S, Ujino-Ihara T, Saito M, Tsumura Y, Higuchi Y, et al. Construction of a
555 reference transcriptome for the analysis of male sterility in sugi (*Cryptomeria japonica* D. Don)
556 focusing on *MALE STERILITY 1 (MS1)*. PLoS One. 2021;16:e0247180.

557 28. Kapustin Y, Souvorov A, Tatusova T, Lipman D. Splign: algorithms for computing spliced
558 alignments with identification of paralogs. Biol Direct. 2008;3:20.

559 29. Koboldt DC, Larson DE, Wilson RK. Using VarScan 2 for germline variant calling and somatic
560 mutation detection. Curr Protoc Bioinformatics. 2013;44:15.4.1–17.

561 30. Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, et al. VarScan: variant
562 detection in massively parallel sequencing of individual and pooled samples. Bioinformatics.
563 2009;25:2283–5.

564 31. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of
565 SAMtools and BCFtools. Gigascience. 2021;10:giab008.

566 32. Berardini TZ, Reiser L, Li D, Mezheritsky Y, Muller R, Strait E, et al. The Arabidopsis
567 information resource: Making and mining the "gold standard" annotated reference plant genome.
568 Genesis. 2015;53:474–85.

569 33. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning
570 sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

571 34. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq
572 data with DESeq2. Genome Biol. 2014;15:550.

573 35. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid
574 substitutions and indels. Bioinformatics. 2015;31:2745–7.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

575 36. Rice P, Longden I, Bleasby A. EMBOSS: The European molecular biology open software suite.
576 Trends Genet. 2000;16:276–7.

577 37. Futamura N, Totoki Y, Toyoda A, Igasaki T, Nanjo T, Seki M, et al. Characterization of expressed
578 sequence tags from a full-length enriched cDNA library of *Cryptomeria japonica* male strobili. BMC
579 Genomics. 2008;9:383.

580 38. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–
581 100.

582 39. Shahmuradov IA, Umarov RK, Solovyev VV. TSSPlant: a new tool for prediction of plant Pol II
583 promoters. Nucleic Acids Res. 2017;45:e65.

584 40. Blum M, Andreeva A, Florentino LC, Chuguransky SR, Grego T, Hobbs E, et al. InterPro: the
585 protein sequence classification resource in 2025. Nucleic Acids Res. 2025;53:D444–56.

586 41. Takata N, Sato R, Nagano S, Ueno S. SugiExDB: data repository for *Cryptomeria japonica* RNA-
587 sequencing read counts in TogoDB. J Wood Sci. 2025;71:1–6.

588 42. Wulf MG, Maguire S, Humbert P, Dai N, Bei Y, Nichols NM, et al. Non-templated addition and
589 template switching by Moloney murine leukemia virus (MMLV)-based reverse transcriptases co-
590 occur and compete with each other. J Biol Chem. 2019;294:18220–31.

591 43. Chepyshko H, Lai C-P, Huang L-M, Liu J-H, Shaw J-F. Multifunctionality and diversity of GDSL
592 esterase/lipase gene family in rice (*Oryza sativa* L. japonica) genome: new insights from
593 bioinformatics analysis. BMC Genomics. 2012;13:309.

594 44. Girard A-L, Mounet F, Lemaire-Chamley M, Gaillard C, Elmorjani K, Vivancos J, et al. Tomato
595 GDSL1 is required for cutin deposition in the fruit cuticle. Plant Cell. 2012;24:3119–34.

596 45. Yuan G, Zou T, Zhang X, Liu M, Luo T, He Z, et al. A rice GDSL esterase/lipase protein (GELP)
597 is required for anther and pollen development. Mol Breed. 2020;40:1–15.

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

598 46. Tsugama D, Fujino K, Liu S, Takano T. A GDSL-type esterase/lipase gene, GELP77, is necessary
599 for pollen dissociation and fertility in *Arabidopsis*. Biochem Biophys Res Commun. 2020;526:1036–
600 41.

601 47. Zhu J, Lou Y, Shi Q-S, Zhang S, Zhou W-T, Yang J, et al. Slowing development restores the
602 fertility of thermo-sensitive male-sterile plant lines. Nat Plants. 2020;6:360–7.

603 48. Peng G, Liu Z, Zhuang C, Zhou H. Environment-sensitive genic male sterility in rice and other
604 plants. Plant Cell Environ. 2023;46:1120–42.

605 49. Shahinnia F, Geyer M, Block A, Mohler V, Hartl L. Identification of Rf9, a gene contributing to
606 the genetic complexity of fertility restoration in hybrid wheat. Front Plant Sci. 2020;11:577475.

607 50. Kammenga JE. The background puzzle: how identical mutations in the same gene lead to different
608 disease symptoms. FEBS J. 2017;284:3362–73.

609 51. Nanasato Y, Mikami M, Futamura N, Endo M, Nishiguchi M, Ohmiya Y, et al. CRISPR/Cas9-
610 mediated targeted mutagenesis in Japanese cedar (*Cryptomeria japonica* D. Don). Sci Rep.
611 2021;11:16186.

612 52. Nishiguchi M, Futamura N, Endo M, Mikami M, Toki S, Katahata S-I, et al. CRISPR/Cas9-
613 mediated disruption of CjACOS5 confers no-pollen formation on sugi trees (*Cryptomeria japonica* D.
614 Don). Sci Rep. 2023;13:11779.

615

616 s

617 Additional files

618

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

Additional file 1

File name: Additional file 1

File format: .xlsx

Title of data: **Supplementary Table 1** RNA-Seq sample information

Description of data: List of sample ID, collection date, library ID, sequencing depth, and fertility status.

Additional file 2

File name: Additional file 2

File format: .xlsx

Title of data: **Supplementary Table 2** Details of markers and the fine mapping

Description of data: Marker IDs and their primer sequences; Draft genome contig IDs of marker origin; SUGI_1 (chromosome, SNP position and gene IDs) of marker origin; linkage map positions (cM); and genotyping results for 129 individuals in S1-2 family. Marker IDs beginning with 'AX-' are used in our previous linkage map [22], and their genotypes were compiled from <https://doi.org/10.1371/journal.pone.0206695.s009> .

Additional file 3

File name: Additional file 3

File format: .xlsx

Title of data: **Supplementary Table 3.** List of genes located on MS2 locus, and their characteristics

Description of data: Genes located around MS2 locus around 1.56 cM, functional similarity to *Arabidopsis thaliana* genes, TPM values, and non-synonymous amino acid mutations with its PROVEAN score. In addition to the homology search against *Arabidopsis* pollen-related genes (TAIR GO:0009555) described in the Methods section, previously obtained functional annotations based on the full ARAPORT11 protein from our earlier study [26] are included for reference. These

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

supplementary annotations were not used for candidate gene selection but are provided here to assist interpretation. The functional annotation dataset is available at https://forestgen.ffpri.go.jp/en/info_sugi1.html.

Additional file 4

File name: Additional file 4

File format: .pdf

Title of data: **Supplementary Figure 1.** Amino-acid sequence alignment of the *MS2* mutational candidate genes in *Cryptomeria japonica*: (A) *INTS1* and (B) *GELP*.

Description of data: Alignment of the amino acid sequences of SUGI *standard gene set* (SUGI_0492850 and SUGI_0493010 for *INTS1* and *GELP*, respectively), Gosenshi-1_a1 (wild-type allele), Gosenshi-1_a2 (mutant allele), and the *ms2* mutant from the S1-2 family, generated in CLC Genomics Workbench ver. 20.0.4 (Qiagen). Residues are colored according to side-chain polarity. Red arrows indicate positions of deleterious amino acid substitutions predicted by PROVEAN (score ≤ –2.5). For INTS1, two deleterious amino acid substitutions (at positions 613 and 1047) were identified in the S1-2 family, but these substitutions were absent in 'Gosenshi-1.' For GELP, two substitutions are observed relative to SUGI_0493010 (S40F and C352Y), but only S40F is deleterious and uniquely distinguishes the *ms2* mutant from the wild-type allele. In addition, the 352nd residue in 'Gosenshi-1' (*Ms2*/*ms2*) is homozygous for tyrosine (Y/Y), as indicated by the blue arrow, further indicating that C352Y is unlikely to be the causal mutation.

Additional file 5

File name: Additional file 5

File format: .pdf

Title of data: **Supplementary Figure 2** Microscopic cross-section of a male strobilus from a sterile individual (Y677) in S1-2 family

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

668 Description of data: A partial formation of pollen was observed in the male strobilus of an individual
669 with the *ms2/ms2* genotype. The sample was collected from the Chiyoda nursery on 20 December
670 2023. The scale bar is based on an approximate size of the male strobilus (~4 mm) and is provided as
671 a reference only.

672

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

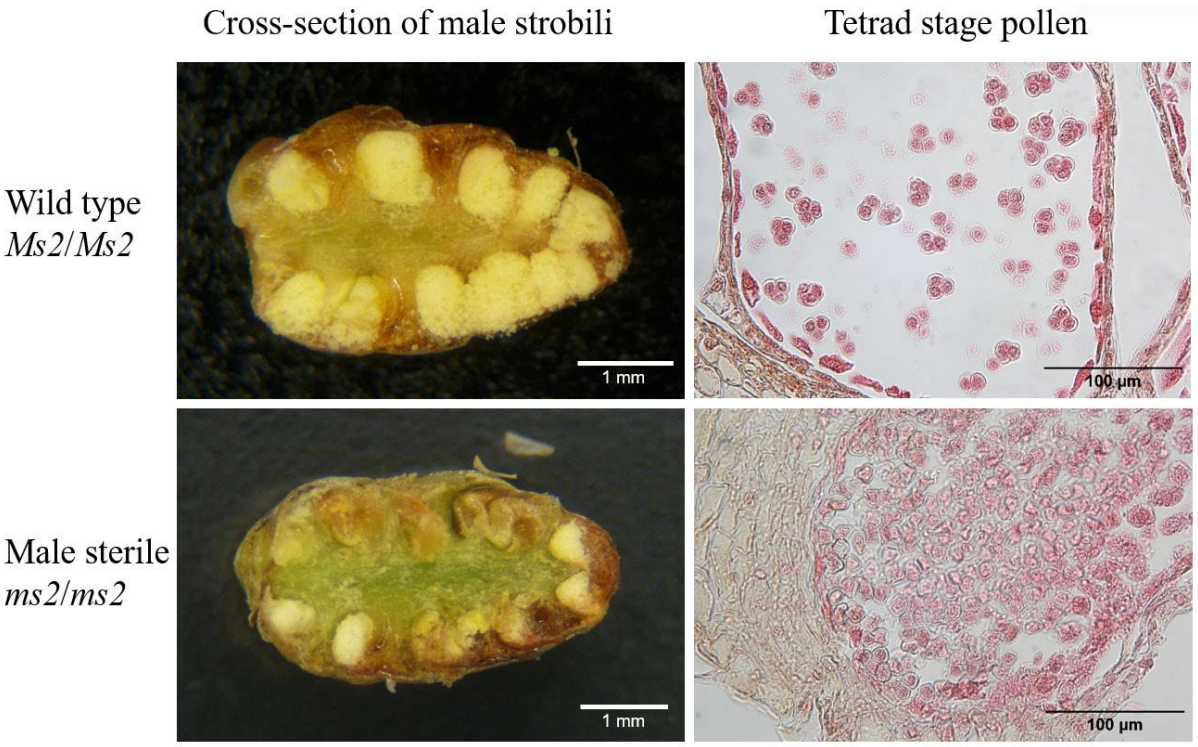
673 # Tables

674

675 # Figures

676

677 **Figure 1**



678

679

680

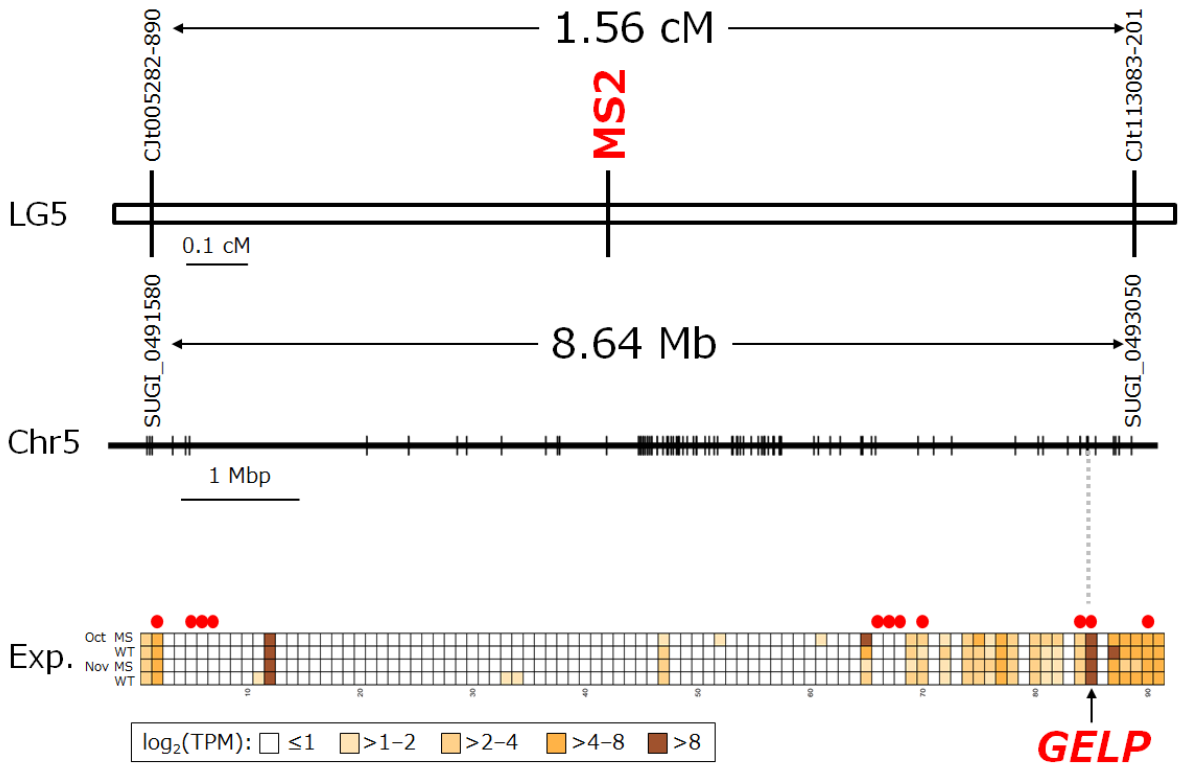
681

682

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

683 **Figure 2**

684



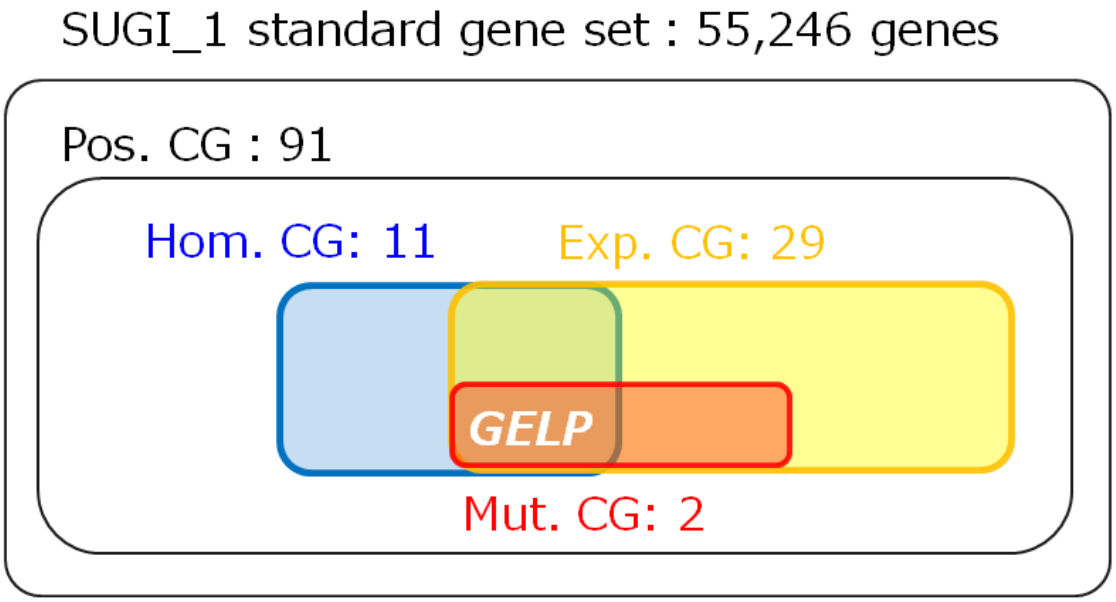
685

686

687

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

688 **Figure 3**

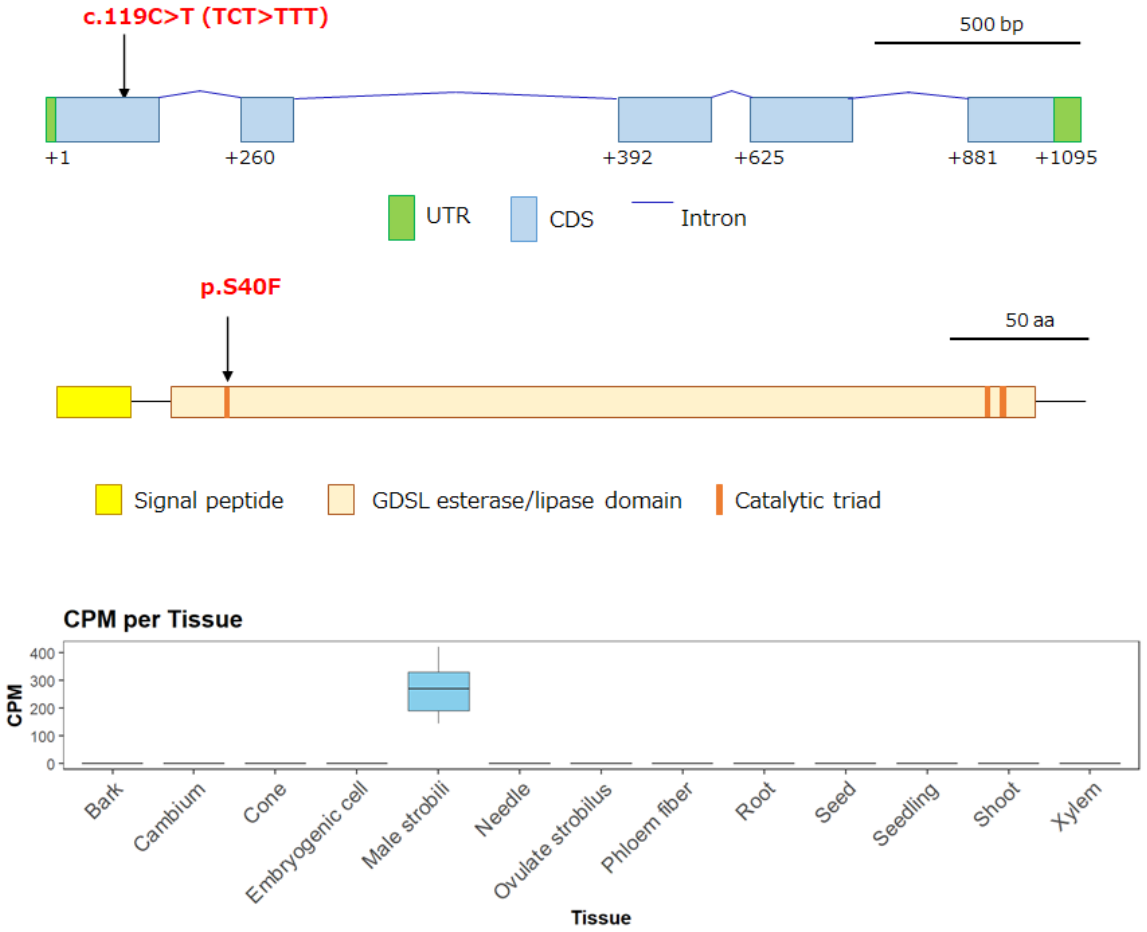


689

690

bioRxiv preprint doi: https://doi.org/10.1101/2025.09.11.674818; this version posted September 12, 2025. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

691 **Figure 4**



692