Molecular identification of the wheat male fertility gene Ms1 and its prospects for hybrid breeding

Elise J. Tucker1, Ute Baumann1, Allan Koudri1, Radoslaw Suchecki1, Mathieu Baes1, Melissa Garcia1, Takashi Okada1, Chongmei Dong2, Yongzhong Wu3, Ajay Sandhu3, Manjit Singh3, Peter Langridge1, Petra Wolters3, Marc C. Albertsen3, A. Mark Cigan3 & Ryan Whitford 1

The current rate of yield gain in crops is insufficient to meet the predicted demands. Capturing the yield boost from heterosis is one of the few technologies that offers rapid gain. Hybrids are widely used for cereals, maize and rice, but it has been a challenge to develop a viable hybrid system for bread wheat due to the wheat genome complexity, which is both large and hexaploid. Wheat is our most widely grown crop providing 20% of the calories for humans. Here, we describe the identification of Ms1, a gene proposed for use in large-scale, low-cost production of male-sterile (ms) female lines necessary for hybrid wheat seed production. We show that Ms1 completely restores fertility to ms1d, and encodes a glycosylphosphatidylinositol-anchored lipid transfer protein, necessary for pollen exine development. This represents a key step towards developing a robust hybridization platform in wheat.
With the predicted growth in world population to over nine billion by 2050, the Food and Agriculture Organization of the United Nations (July 2005) set a target of 60% increased food production by that year. This is an ambitious target for two reasons: there are serious concerns about the viability of existing production systems and the sustainability of current growth rates in crop production, and the predicted environmental changes are expected to have an overall negative effect on agricultural production, with serious crop declines in some countries. Wheat is grown more widely than any other crop and delivers around 20% of our food calories and protein. To increase global production by 60% will require a lift in the rates of gain from the current 1 to 1.6% per annum. Improvements in disease resistance and stress tolerance offer opportunities for small increases in productivity but major jumps in yield are hard to achieve and are expected to come about through gains in the way we breed wheat and other crops. However, many important new genetic and genomic technologies are difficult to apply to wheat since this plant has a large complex genome, an allohexaploid, which is 50 times larger than rice.

One of the most promising options for achieving significant boosts in yield across diverse production environments is through hybrid breeding. Hybrids offer two important advantages: first, heterotic yield gains of well over 10%, and improved yield stability have been reported, and second, hybrid seed production would act as a major stimulant for investment in wheat improvement from both the public and private sectors. However, the competitiveness of wheat hybrids relative to line varieties will depend on hybrid seed production costs.

Lowering hybrid seed production costs depends on a reliable and inexpensive system that forces outcrossing. Wheat male sterility and restoration systems were first developed in the 1960s, but many of them were proved to be impractical and deemed commercially high risk. Relative to systems based on chemical hybridizing agents and cytoplasmic male sterility, the use of non-conditional nuclear-encoded recessive male steriles (ms) would offer major advantages for hybrid breeding. The value of recessive male steriles was first recognised in 1972 with the proposal of the XYZ system. This system aimed to overcome the costs associated with propagating pure stands of male steriles by cyrogenetic chromosomal manipulation. A further advantage of recessive male steriles came through the opportunity to broaden parental line choice, avoid negative alloplasmic and cytoplasmic yield penalties, as well as alleviate the problems associated with incomplete fertility restoration. A cost-effective and flexible hybridization platform that uses a recessive male sterile is Seed Production Technology (SPT) developed for maize and rice hybrid seed production (Supplementary Fig. 1). This platform overcomes many of the problems with large-scale production of male steriles for use as female parents in hybrid breeding. SPT uses a maintainer line solely for the propagation of non-GM homozygous recessive male-sterile parents; therefore, F1 hybrids provided to farmers are considered to be non-GM.

Developing an equivalent platform for hybrid wheat breeding requires the identification of a suitable non-conditional, nuclear-encoded recessive male sterile. These types of mutants are particularly rare and difficult to detect in polyploids due to genetic redundancy. Many of our major crops and food plants are polyploids, including wheat, oats, potato, sweet potato, peanut, sugarcane, cotton, kiwifruit, strawberry, and plums. For example, only ten nuclear-encoded wheat male-sterile mutants have been identified to date, in contrast to 108 mutants in diploid barley. Polyploidy not only makes it difficult to find suitable male-sterile mutations but also complicates deploying mutants since multiple mutations would be needed to deal with genetic redundancy and this increases breeding costs and population sizes needed for introgression of each additional mutation. The most cost-effective mutants would be single locus encoded. In wheat, only two of the ten mutant loci are reported to fit this criterion. These are ms1 and ms5 located on chromosomes 4BS and 3AL, respectively.

The first ms1 mutant was observed in Australia in the late 1950s. This spontaneous mutant named Pugsley's male sterile was followed by the identification of Probus and Cornerstone male sterile from an X-ray-induced wheat mutant population. Cytogenetic and linkage analysis showed these to be allelic and they were designated as ms1a, ms1b and ms1c, respectively. In 1976, additional monogenic recessive male steriles were identified from an ethyl methanesulfonate (EMS)-treated population. Three mutants were allelic to ms1, and designated as ms1d, ms1e and ms1f. While the fourth mutant was nonallelic to ms1 and designated as ms5. However, even for ms1, the variability between backgrounds and mutant alleles, and problems with male sterility penetrance were reported. In order to address these problems, it is necessary to identify the gene underlying the Ms1 locus and explain its function.

Here, we describe the identification of the Ms1 gene sequence (TaMs1) by map-based cloning and demonstrate its function in male fertility by complementation of the ms1d mutant. TaMs1 encodes a glycosylphosphatidylinositol (GPI)-anchored lipid transfer protein, which is necessary for pollen exine development. The identification of the Ms1 gene sequence represents a key step towards developing a robust hybridization platform in wheat similar to the maize SPT.

Results
Ms1 encodes a GPI-anchored lipid transfer protein. We followed a map-based cloning approach to isolate the Ms1 gene sequence (chr. 4BS). Using syntenic regions on chromosomes 1, 3 and 4 from rice, Brachypodium and barley, respectively, we generated markers and tested their presence or absence in male-sterile mutants ms1a, ms1b and ms1c. The results revealed that ms1a and ms1c are terminal deletions while ms1b is an interstitial deletion of the chromosome 4BS covering approximately 14 centimorgans (cM) (Fig. 1). Ms1-flanking markers were identified by their presence in ms1b and their absence from ms1c. Using a population representing 7000 meioses and segregating for ms1d, we delimited the Ms1 locus to a 0.5 cM interval between markers ×27140346 and ×12360198. Probes designed within the region bounded by these markers, were used to isolate and sequence BACs from durum and hexaploid wheat. Marker development from BAC-derived sequences and analysis of 14 recombinants across the region, further delimited Ms1 between markers 007.033.1 and 007.0046.1. The mapped 251 Kb interval contains eight intact genes and one likely pseudogene (Supplementary Table 1).

RNaseq-based expression profiling identified one of these eight genes to be preferentially expressed in floral tissues (Supplementary Table 1). This gene (TaMs1) is predicted to encode a 219 amino acid polypeptide with a similarity to a large family of GPI-anchored lipid transfer proteins (LTPGs), for which it is a member of a Poaceae-specific clade (Supplementary Fig. 2). This gene was confirmed as Ms1 through in planta complementation of the ms1d mutation and identification of the causative lesions in ms1d, ms1e, ms1f and a newly identified TILLING mutant (described below).

TaMs1 is necessary for pollen exine integrity. Arabidopsis harbours over 20 LTPG genes for which only two of them have been characterised. AILTPG1 and AILTPG2 are required for cuticular wax accumulation or for export onto stem and siliques.
surfaces. Epicuticular wax has lipid precursors common to sporopollenin, the major constituent of pollen exine, which is produced in sporophytic tapetal cells and transported to developing microspores in structures called orbicules. The analysis of ms1 anthers revealed a disrupted orbicule and a pollen exine structure (Fig. 2a), which was first observed in early uninnucleate microspores and typified by ectopic exine deposition and reduced electron-dense materials at the tapetal cell surface (Supplementary Figs. 3 and 4). No differences in the surface cuticle layer were observed between ms1 and wild-type anthers (Supplementary Fig. 5). Furthermore, metabolomic profiling revealed that ms1 anthers accumulate lipid monomers of sporopollenin (C16 and C18 long-chain fatty acids) relative to the wild type (Supplementary Fig. 6). Taken together, this suggests that Ms1 is necessary for sporopollenin biosynthesis or transport. Transcriptary Fig. 6). Taken together, this suggests that Ms1 is necessary for sporopollenin biosynthesis or transport. Transcriptional β-glucuronidase fusions and homeologue-specific qRT-PCR revealed only the B-genome-derived TaMs1 is to be expressed during early microspore development (Fig. 2b, c).

**TaMs1 exhibits functional divergence from its homeologues.** Since no obvious differences in TaMs1 coding potential were detected between homeologs, the basis for TaMs1’s sub-functionalization between homeologs is likely to be due to variation in transcription. We suggest that functional homeoalleles may still exist in the cultivated germplasm pool and this could account for the reports of poor sterility penetrance dependent upon the genotype and the mutant allele22–24. In each of these cases, the loss of Ms1 is either via a large deletion (ms1a and ms1c) or chromosome arm (DT4BL) replacement; therefore, restoration of fertility is unlikely to be a consequence of the B-genome-derived TaMs1. By performing a TaMs1 homeologue-specific qRT-PCR on anthers isolated from a partially fertile homozygote for ms1c, we attempted to answer the question on whether TaMs1 homeoalleles can transcriptionally compensate for the loss of TaMs1-B. However, this does not seem to be the case since TaMs1 and homeologous transcripts from cv. *Cornerstone* were all below the detection limits (Supplementary Fig. 7). It is, therefore, possible that other genomic loci are associated with fertility restoration in lines carrying *ms1* deletions.

The variable penetrance of *ms1* large-deletion mutants led us to investigate the utility of the available *Ms1* mutant alleles derived from an EMS-treated population27. Chromosome 4BS-specific full-length coding sequences from wild-type (*Ms1*) fertiles and the EMS-derived *ms1d, ms1e* and *ms1f* steriles were isolated and sequenced. A comparison of *Ms1* to *ms1*-derived sequences identified unique single-nucleotide transitions for each mutant allele (Fig. 3a). Transition G329A is unique to ms1d and unlikely to be a natural allelic variant, considering that the wild-type G is sequenced. A comparison of *Ms1* and *ms1*-derived sequences indicated unique single-nucleotide transitions for each mutant allele (Fig. 3a). Transition G329A is unique to *ms1d* and unlikely to be a natural allelic variant, considering that the wild-type G is detected in all 192 spring wheat varieties tested (Supplementary

---

**Fig. 1** Map-based cloning of the *Male sterility 1* locus on chromosome 4BS. Ms1 was initially mapped to the interval between ×12,21056 and ×BF292015 based on genotyping of (a) deletion mutant allele series, and then to an (b) ~0.5 cM interval between ×27140346 and ×12360198 based on 7000 F2 segregants. Fine mapping using 14 recombinants delimited *Ms1* to a (d) 251-Kb genomic region in wheat. Marker names are in italics. The numbers indicate recombinants identified for each marker interval. Coloured arrows I to XI denote the position and orientation of predicted wheat genes with a similarity to *Brachypodium* genes Bradi1g13040 (Cupin domain-containing protein), Bradi1g13040 (Cupin domain-containing protein), Bradi2g05445 (60S ribosomal protein), Bradi1g13030 (Lipid Transfer Protein-Like 94), Bradi4g44760 (F-box/LRR-repeat protein 3), Bradi1g69240 (U-box domain-containing protein), Bradi1g13000 (Lipid Transfer Protein-Like 72), Bradi1g12990 (Lipid Transfer Protein-Like 71), Bradi1g12980 (Putative Parafibromin), Bradi1g12970 (Putative GNAT family acetyltransferase) and Bradi1g12960 (DUF581 domain-containing protein), respectively. The sequence is available via NCBI GenBank accession code KX447407.
This mutation changes an aspartic acid to an asparagine (D60N) within the conserved LTP domain. Taken together, these findings indicate that both the GPI anchor and putative lipid-binding domains are necessary for Ms1 functionality.

TaMs1 functionally complements ms1d. The SPT hybridization platform incorporates a maintainer line (Supplementary Fig. 1) capable of propagating non-GM nuclear male-sterile lines for use as female parents in hybrid production. The SPT maintainer line is homozygous recessive male sterile transformed with a SPT construct containing (i) a complementary wild-type male fertility gene to restore fertility, (ii) an α-amylase gene to disrupt pollination and (iii) a seed colour marker gene. We demonstrated that the α-amylase gene and seed colour marker function in wheat (Supplementary Fig. 9). However, the remaining key component of the SPT hybridization platform requires the demonstration of complementation of male sterility to fully restore fertility9. Therefore, we tested the ability of this gene sequence to complement ms1d. A 4.4-Kb genomic fragment containing Ms1 was synthesised (TaMs1) and introduced into the wheat cultivar Gladius segregating for ms1d. Eleven independent Agrobacterium-mediated T0 transformants were generated (Supplementary Table 3). Four SNP markers closely flanking the Ms1 locus allowed the selection of T0 regenerants that were homozygous for ms1d, whilst a seed-screenable marker (DsRed) was used to confirm the presence of TaMs1. SNP and seed colour detection identified six homozygous ms1d T0 regenerants with the introduction of the TaMs1 gene. Selfed seed set analysis showed that all six T0 regenerants were fully fertile (Fig. 3c, Supplementary Table 3). Seventeen T1 progenies for two independent T0 lines (Event 1 and Event 7) were assayed for both the copy number and zygosity of the introduced TaMs1 (Supplementary Table 4). The results revealed that all progenies were homozygous for ms1d with either zero, one or two copies of the exogenous TaMs1. Those progenies with no detectable introduced TaMs1 were male sterile whilst those containing either one or two copies of the TaMs1 transgene were self-fertile. These findings demonstrate that Ms1 can fully restore fertility to the homozygous ms1d mutant.
Fig. 3 | Identity of male sterility inducing lesions within TaMs1 and in vivo complementation of ms1d/ms1d by TaMs1.

(a) Schematic representation of the TaMs1 (the three exons are shown as black boxes) gene depicting the relative positions (indicated by solid lines with arrowheads) of EMS-derived lesions (chromatogram insets) for ms1d, ms1e, ms1f and ms1h. Polymorphisms cause either a frame shift (ms1d, ms1e) or an amino acid transition in a conserved residue of TaMs1 (ms1f, ms1h). (b) Sequence chromatogram comparison between both mutant (ms1d) and wild-type genomic cDNAs. Polymorphism G329A of ms1d induces the use of a cryptic splice site (highlighted by a 7 bp insertion) within TaMs1. (c) Stable complementation of the ms1d mutant by TaMs1. Mature inflorescences of male-sterile ms1d/ms1d, and three independent transformants (Events 1–3), each homozygous for ms1d, and showing a self-seed set (arrow). Scale bars: 1 cm.
Discussion

An ambitious target of a 60% increase in wheat production by 2050 has been set to meet the predicted demand for this crop (FAO). A viable hybrid seed production system could deliver a third or more of this gain and currently, it is difficult to see any other technology that could achieve an equivalent impact. Several components are needed for an efficient hybrid system, which include optimising hybrid gains through defining heterotic pools and improving pollen vectoring by transitioning from a self-pollinating flower structure to the one that facilitates outcrossing. Based on experience in other hybrid systems, these components can be generated through targeted selection. However, forcing outcrossing through a reliable and manageable male sterility system requires the development of systems that are able to cope with the complexity of the haploid wheat genome.

The use of recessive male sterility has been an attractive prospect for hybrid seed production in wheat since it was first proposed in the 1950s but translating the concept to reality has proved to be elusive. The conundrum has been to maintain the male-sterile lines for use as pollen acceptors in commercial hybrid seed production fields. For example, the cyrogenetic 4E-ms system, which utilises the novel mutant mslg allele and a fertility-restoring chromosome from Agropyron elongatum ssp. ruthenicum Beldie (4E) was initially reported to be successful in restoring chromosome from cum Beldie (4E) was initially reported to be successful in restoring chromosome from cum Beldie (4E)30. Apart from as a consequence of a transposon insertion36, 37. Further research represents a novel gene sequence expressed only in sterile anthers recent cloning and complementation of the Ms1. Ms1 represents a major cost saving for hybrid seed production as optimising ms1 as a system for the production of a hybrid seed is now possible through the adoption of advanced breeding technologies such as gene editing39. 40. Gene editing would enable the generation of novel ms1 alleles and allow simultaneous testing in different isogenic wheat backgrounds (e.g. spring and winter or hard and soft wheat). Further, once highly penetrant ms1 alleles are identified, rather than introgression through conventional backcrossing, this new variant allele could be rapidly introduced into the most elite genetics by directly editing TaMs1. Adoption of new breeding technologies is likely to be particularly valuable in targeting the results of novel alleles from other hybrid systems and can overcome the seed purity issues inherent to the 4E-ms system. Coupled with a functional α-amylase gene for wheat pollen disruption and a seed-selectable marker, the identity and ability of the TaMs1 gene sequence to completely restore viable pollen production in ms1 plants represents the last critical step towards developing SPT for wheat.

To our knowledge, this study is the first to report a G-protein anchored LTP in pollen exine development32. Given the observation that mslld anthers contain pollen deficient in exine structural integrity, and that metabolomic profiling revealed a significant accumulation in free C16 and C18 fatty acids, it is reasonable to expect this to be a consequence of disruption in sporopollenin precursor transport from sporophytic tapetal cells to developing microspores. The finding that anther epidermal wax layers were not disrupted in mslld anthers also indicates that TaMs1’s role is specific to exine development. Further studies are needed to elucidate the subcellular localisation of TaMs1 and its direct involvement in lipid binding.

TaMs1’s identification now also raises the question on whether previously reported male-sterile mutants may represent additional male-sterile alleles of this gene or its homologues. Apart from the dominant male-sterile mutants Ms233, 34 and Ms335, no other mutants have been reported on homeogroup four. Both Ms2 and Ms4 were reported to be located on chromosome 4DS. However, recent cloning and complementation of the Ms2 gene sequence WMS, reveals that it does not encode a TaMs1 homologue, but represents a novel gene sequence expressed only in sterile anthers as a consequence of a transposon insertion36, 37. Further research is necessary to determine whether Ms4 is a mutant of the TaMs1-D homologous gene sequence.

From a traditional breeding perspective, the molecular identity of the TaMs1 gene sequence now allows the development of germplasm-specific markers for fast-tracking ms1 introgression into diverse female-inbread parental lines. Moreover, complementation studies demonstrate that Ms1 is unique and contrasts with other reported wheat ms mutant alleles38 in that Ms1 behaves as a single-mutant locus in hexaploid wheat and a single copy of Ms1 restores fertility. Given the characterisation of the ms1 alleles described here and the observed variation in pene trance of sterility between these different alleles, understanding the relationship of these mutations to pollen production, as well as optimising ms1 as a system for the production of a hybrid seed is now possible through the adoption of advanced breeding technologies such as gene editing39, 40. Gene editing would enable the generation of novel ms1 alleles and allow simultaneous testing in different isogenic wheat backgrounds (e.g. spring and winter or hard and soft wheat). Further, once highly penetrant ms1 alleles are identified, rather than introgression through conventional backcrossing, this new variant allele could be rapidly introduced into the most elite genetics by directly editing TaMs1. Adoption of new breeding technologies is likely to be particularly valuable in targeting the results of novel alleles from other hybrid systems and can overcome the seed purity issues inherent to the 4E-ms system. Coupled with a functional α-amylase gene for wheat pollen disruption and a seed-selectable marker, the identity and ability of the TaMs1 gene sequence to completely restore viable pollen production in ms1 plants represents the last critical step towards developing SPT for wheat.

Methods

Plant and DNA materials. Bread wheat lines used for fertility phenotyping, molecular marker development and genetic mapping were cv. Chinese Spring, Chinese Spring-derived nullisomic tetrasomic stocks31 (N4AT4D, N4AT4B, N4BT4A, N4BT4D, N4DT4A and N4DT4B), cv. Gladis, cv. Pogseley’s male sterile (ms1)18, cv. Probus (ms1b)38, cv. Cornerstone (ms1c)33, cv. Chris and Chris-EMS-mutagenised lines FS2, FS3 (ms1e) and FS24 (ms1f). All cultivars, breeders’ lines or DNA samples for marker screening were obtained from various Australian wheat-breeding programmes, Australian Wheat and Barley Molecular Marker Program or the Australian Winter Cereals Collection (AWCC). The BAC library was derived from the Triticum turgidum ssp. durum cv. Langdon35. The EMS-mutagenised population used for TILLING was derived from the soft bread wheat cv. QAL2006.

Plant growth and phenotyping. Plants for genetic mapping, cytological examination, expression analysis, TILLING and transformation donor material were sown at 3–6 plants per 61 (8-inches diameter) pot containing soil mix. The soil mix consisted of 75% (v v−1) CocoS, 25% (v v−1) nursery-cutting sand (sharp), 750 mg l−1 CaSO4.2H2O (gypsum) 750 mg l−1 Ca(H2PO4)2, 1.9 g l−1 FeSO4, 125 mg l−1 FeEDTA, 1.9 g l−1 Ca(NO3)2, 750 mg l−1 Scotts Micromax micronutrients and 2.5 g l−1 Osmocote Plus slow-release fertilizer (16:3:9) (Scotts Australia Pty. Ltd.). The pH was adjusted between 6.0 and 6.5 using two parts of agricultural lime to one part of hydrated lime. Potted plants were grown either in controlled environment growth rooms at 23 °C (day) and 16 °C (night) or similarly temperature moderated glasshouses in which the photoperiod was extended using 400 W high-pressure sodium lamps in combination with metal halide lamps to 12 h over winter months.

Individual plants were assessed for self-fertility by placing and sealing a glassine bag over each head before anthesis. Between three and ten heads per plant were collected for seed counting. The two basal and two apical spikelets per head were eliminated from analysis due to their incomplete development. The total seed set and numbers of florets were counted on a per-head basis. The percentage of fertility for each spike or plant was calculated as follows: 

\[
\text{Percentage of fertility} = \left(\frac{\text{Total number of 1}, 2, \text{and} 3 \text{florets per spike}}{\text{Total number of spikes per spike}}\right) \times 100
\]

A plant was deemed to be self-fertile if the total calculated percentage of fertility was greater than 60% or was equivalent to a wild-type control. Pollen viability was assessed for three isolated anthers per plant (n = 3) by either acetocarmine or Lugol (1% solution) staining. Dissected anthers were mounted on glass microscope slides and pollen grains (n > 500 per sample) counted for staining by visualising them on a Zeiss Axio Imager M2 optical microscope coupled with a CCD camera (The University of Adelaide microscopy). Stainable pollen is represented as the mean ± standard deviation for nine samples per genotype.

Histochemical staining. Anthers containing premeiotic microspores to mature pollen were isolated from wheat plants identified to contain the TaMs1-gusplus cassette. Histochemical GUS activity was detected using 5-
bromo-4-chloro-3-indolyl-beta-D-glucuronic acid (Gold Biotechnology, Inc). The samples were incubated in a 1 mM X-Gluc solution in 100 mM sodium phosphate at pH 7.0, 10 mM potassium chloride, 10 mM MgCl2 and 10 mM sodium fluoride at 37 °C. After 5 days, samples containing mesophyll cells or sterile deletion mutant lines were transferred to fresh X-Gluc solution. The stained tissue was then washed three times with sterile water and dried for 20 min. The puriﬁed DNA was extracted using a Sephadex G50 column (GE Healthcare) and denatured at 95 °C for 5 min. Denatured DNA was puriﬁed by precipitation using 3 volumes of isopropanol and centrifuged at 13,000 rpm for 10 min. The DNA pellets were washed twice with 1 ml of 70% ethanol, centrifuged for 2 min at 13,000 rpm and air dried for 20 min. DNA was precipitated by adding 60 µl of isopropanol and centrifuged at 13,000 rpm for 10 min. The DNA pellet was resuspended in 10 µl of sterile water and used for hybridisation. This represents a coverage of 5.1 genome equivalents from the Triticum turgidum ssp. durum wheat cv. Langdon.32 For prehybridisation, overnight incubation of colony ﬁlters in a hybridisation solution (2x SSPE, 0.5% SDS, 5x Denhardt’s reagent), and 40 µg ml−1 salmon sperm DNA) was done in rotary glass tubes at 65 °C. The labelled probe was mixed with 5 ml of hybridisation solution and heated at 65 °C for 10 min. Approximately 2 µg of unbound probe, the ﬁlters were washed twice in a washing solution containing 2x SSPE and 0.5% SDS and rinsed with 1x SSC. The washed ﬁlters were exposed to an X-ray ﬁlm for 1–3 days based on the signal intensity to identify positive clones. BAC clones that gave a positive signal were grown on single colonies from glycerol stocks and DNA was extracted using the PureLink® bacteria genomic DNA purification kit (Epicentre®, www.epicentre.com, Madison, Wisconsin, USA).

Restriction mapping, PCR experiments with primers corresponding to the markers previously used, determined the order of the BACs covering the region of interest. BAC libraries from the Triticum turgidum ssp. durum cv. Langdon and a Chinese Spring chromosome arm survey sequencing using Biokanga 2.76 (https://github.com/cirso-crop-informatics/biokanga), allowing 2 mismatches per 100 bp to conﬁrm that they were derived from homegroup 4. The reads were then filtered for bacterial sequence contamination, and trimmed for a vector sequence using a combination of BLASTn-based ﬁlters and custom scripts. Before assembly, the overlapping PE reads were fused using FLASH 1.2.7 (https://sourceforge.net/projects/flashpage). The fused reads along with the remaining PE reads were then assembled into contigs using MIRA 4.0.2. Contigs produced by MIRA were then scaffolded using SPSPACE36 using mate-pair anchors for each contig derived from a single-mate-pair library for all samples. Single contiguous scaffolds for each homeologous were manually ﬁnished using Gap53. Highly repetitive regions on the BACs were masked based on a powder depth of mate-pair reads (see Materials and Methods, Long-read, cv. ‘Gladius’45) exceeding 1000. The alignments to the BACs of Brachypodium genes as well as mappings of publically available RNA-seq datasets from the bread wheat cv. Chinese Spring66 facilitated gene prediction. Nucleotide sequences spanning the Ms1 region were submitted to GenBank (accession codes KX447407, KX447408 and KX447409).

Nucleic acid extraction and expression analysis. DNA extractions from all bread wheat lines were performed using either a phenol/chloroform or freeze-dried extraction protocol.37 A 15 cm leaf piece from a 2-week-old plant was frozen in liquid nitrogen, and the tissue was ground to a ﬁne powder using one large (9 mm) and three small (3 mm) ball bearings and a vortex. 700 µl of the extraction buffer (1% Sarkosyl, 100 mM Tris-HCl at pH 8.5, 100 mM NaCl, 10 mM EDTA and 2% PVP) was added to each sample. The samples were mixed for 2 min on a rotatory shaker. 700 µl of phenol/chloroform/iso-amylalcohol (25:25:1) was added and the extract was transferred to a silica matrix tube and spun at 4000 rpm for 10 min. DNA was precipitated by adding 60 µl of 3 M sodium acetate at pH 4.8 and 600 µl of isopropanol and centrifuged at 13 000 rpm for 10 min. The DNA pellet was washed with 1 ml of 70% ethanol, centrifuged for 2 min at 13 000 rpm and air dried for 20 min. The puriﬁed DNA was resuspended in 50 µl of 1x TE (10 µg ml−1 RNAase A).

TaMs1 and homeologous transcripts were detected by qRT-PCR on cDNA using total RNA extracted from wheat of cv. Chris (wild type) using an ISOLATE II RNA kit (Bioline) for reverse transcription. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed. For each sample, the respective fertility control (active fertility control) was performed.
anthers from the same floret were isolated and snap frozen in liquid nitrogen. Developmentally equivalent anthers were pooled and RNA isolated. All total RNA samples were treated with DNase I (Qiagen). First-strand cDNA was synthesised using oligo dT51 and Superscript III reverse transcriptase (Thermo Fisher). Amplification products from qRT-PCR on each tissue sample, three technical replicates and three biological replicates, were used to estimate TaMs1 transcript abundance relative to TaEFA2*, TaGAHDH 2*2 and Tctuliphenol 2*2 reference transcripts. Standard qRT-PCR assays58 were performed using primers, as listed in Supplementary Table 7.

RNA-seq analysis and expression. RNA-seq reads derived from five organs (root, leaf, stem, spike and grain) at three developmental stages each from hexaploid wheat cv. Chinese Spring have been published previously59. The reads were aligned against the repeat-masked BAC assemblies with Bowtie259 and TopHat60. The returned alignments were stringently filtered so as to remove ambiguously aligned reads and read pairs with conflicting alignments. Gene expression was computed on RNA-seq data using Cufflinks and Cuffmerge v3.2.1. RNA-seq expression data for all predicted coding regions from the BAC assemblies are presented in Supplementary Table 1.

TaMs1 sequence from mutant alleles. To identify the TaMs1 sequence variants in the ms1d, ms1c and ms1f alleles, the TaMs1 coding region was amplified from 13 individuals segregating for the sterile phenotype for each mutant allele. PCR used Phusion High-Fidelity DNA Polymerase (NEB, M0530) with Phusion GC buffer, 5% DMSO and 1 M betaine using the primer TaMs1 (coding region) listed in Supplementary Table 7. The fragments were subcloned into a PCR/8/GW/TOPO® TA Cloning Kit (Invitrogen, Fisher Scientific, Københavnsvej) and Sanger sequencing of positive clones was performed by the Australian Genome Research Facility. Sequence chromatograms were compared using Genome version 6.1.82. Sequence analysis correlated a G-to-A transition at position 329 with the sterility phenotype in the ms1d mutant. A Kompetitive Allele-Specific PCR (KASP®) (LGc Genomics) was designed to this SNP transition using Primer Picker (LGc Genomics) (007-00911. Supplementary Table 6) and was assayed using KASP Mastermix on the SNPline (LGc Genomics) on DNA from the ms1d x Gladus F1 mapping population, the ms1d mutant alleles and across a panel of 192 spring wheat germplasm.

Phylogenetic analysis. TaMs1 homologous Plocase sequences were retrieved from Phytozone (www.phytozone.net), TGAC Triticum monococcum Shotgun sequence, International Barley Genome Sequencing Consortium and Rice Genome Annotation Project (http://rice.planthbiology.msu.edu). All BLASTNs, BLASTPs, tBLASTn and BLASTx hits were retrieved using a cutoff e-value of ≤ 1 × 10−5. Default BLAST settings were used for querying with complete sequences. Two prediction tools, PredGI (gpcc.biocomp.unibo.it/predgpi) and big-PI Plant Predictor (mendel.imp.ac.at/gpi/plant_server.html), were then used to determine whether primary peptide sequences contained a putative GPI-anchored motif at the C-termini. Protein multiple-sequence alignments (MSAs) were generated using MUSCLE (default settings) implemented in Geneious analysis package (www.geneious.com). The phylogenetic alignment was performed to improve the MSAs. The phylogenetic tree was computed with MEGA7 using the maximum likelihood method under default parameters (www.megasoftware.net).

Constructs. A 4.3-Kb genomic fragment containing approximately 1.5 Kb upstream of TaMs1 promoter and 1 Kb downstream of the stop codon was synthesised and introduced upstream of the visible marker MoPAT-DsRED (a translational fusion of the bialaphos resistance gene, phosphorothionin-N-acetyltransferase, and the red fluorescent protein DsRED) transcribed by the maize Ubiquitin promoter66 and this 8.1 kb DNA fragment replaced the 2.1 kb coding region was amplified from a T0 plant homozygous for TaMs1 mutation with primers M2″ (Event-1 and Event-7) (Supplementary Table 4). One set of T1 progenies was derived from two separate T0 plants, each with independent T-DNA insertions (Event-1 and Event-7) (Supplementary Table 4). One set of T1 progenies was derived from a T0 plant homozygous for ms1d mutation (ms1d/ms1d) and containing the TaMs1-DsRed cassette (Event-1). The second set of T1 progenies was derived from a T0 plant heterozygous for ms1d mutation (Ms1/ms1d) and containing the TaMs1-DsRed cassette (Event-7). Genotyping for ms1d zygosity and the presence of the T-DNA insertion for plants derived from both sets were determined using flanking markers, as described above, and the expression of the DsRed colour marker.

Data availability. MiSeq BAC-sequencing data have been deposited in the NCBI SRA database under BioProject ID PRJNA396428. The assembled genomic DNA of BACs derived from the Ms1 locus (KX447407) as well as its A (KX447408) and D (KX447409) genome-derived homeologues have been deposited in NCBI GenBank. Further data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 18 December 2016 Accepted: 8 August 2017 Published online: 11 October 2017

References
1. FAOSTAT. http://faostat3.fao.org (2013–2015).
2. HW-HY. WHEAT. Scientific and technological cooperation in plant genome research as basis of the ‘Knowledge-Based Bio-Economy’ (Plant - KBBE) 2010 Project HW-HY. WHEAT. https://www.agence-nationale-recherche.fr/110205069-KR-BB0004 (2012).
3. Longin, C. F. H. et al. Hybrid wheat: quantitative genetic parameters and consequences for the design of breeding programs. Theor. Appl. Genet. 126, 2791–2801 (2013).
4. Muhleisen, J., Piepho, H. P., Maurer, H. P., Longin, C. F. & Reif, J. C. Yield stability of hybrids versus lines in wheat, barley, and triticale. Theor. Appl. Genet. 127, 309–316 (2014).
5. Longin, C. F. H., Reif, J. C. & Würschum, T. Long-term perspective of hybrid versus line breeding in wheat based on quantitative genetic theory. Theor. Appl. Genet. 127, 1635–1641 (2014).
6. Whifft, R. et al. Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. J. Exp. Bot. 64, 5411–5428 (2013).
7. Driscoll, C. J. XXY system of producing hybrid wheat. Crop Sci. 12, 516–517 (1972).
8. Driscoll, C. J. Modified XXY system of producing hybrid wheat. Crop Sci. 25, 1115–1116 (1985).
11. Franckowiak, J. Revised linkage maps for morphological markers in barley, 
12. Cigan, A. M. et al. Targeted Mutagenesis of a conserved anther-expressed P450 
gene confers male sterility in monotoc. Plant Biotechnol. J. 15, 379–389 (2016). 
13. Klíndworth, D. L., Williams, N. D. & Maan, S. S. Chromosomal location of 
gene confers male sterility in four mutants of hexaploid wheat. Crop Sci. 42, 
14. Pugsley, A. T. & Oram, R. N. Genetic male sterility in wheat. 
15. Fossati, A. & Ingold, M. A male sterile mutant in Triticum aestivum. 
16. Islam, A. & Driscoll, C. Latent male fertility in 
23. Joshi, G. P., Li, J., Nasuda, S. & Endo, T. R. Development of a self-fertile 
24. Kim, H. et al. Characterization of glycosylphosphatidylinositol-anchored lipid 
25. Borner, G. H. H., Lilley, K. S., Stevens, T. J. & Dupree, P. Identi 
26. José-Estanyol, M., Gomis-Rüth, F. X. & Puigdomènech, P. The eight-cysteine 
35. Maan, S. & Kianian, S. Third dominant male sterility gene in common wheat. 
36. Ni, F. et al. Unsafe Male Sterility in Monocot. 
37. Voytas, D. F. & Gao, C. Precision genome engineering and agriculture: 
opportunities and regulatory challenges. Plant Biol. 12, e010877 (2014). 
38. Singh, M., Kumar, M., Thiğes, K., Cho, M.-J. & Cigan, A. M. MS26/CYP704B 
39. Vovlas, S. E. & Gao, C. Targeted mutagenesis, precision gene editing, and site-specific 
gene insertion in maize using Cas9 and guide RNA. Plant Physiol. 169, 931–945 
2015). 
40. Searls, E. R. in Chromosome Manupulations and Plant Genetics, 29–45 (Springer, 
41. Cigan, A. et al. Construction and characterization of a half millinone clone BAC 
library of durum wheat (Triticum turgidum ssp. durum). Theor. Appl. Genet. 107, 
42. Reynolds, E. S. The use of lead citrate at high pH as an electron opaque stain 
in electron microscopy. J. Cell. Biol. 17, 208 (1963). 
43. Mayer, K. F. et al. A chromosome-based draft sequence of the hexaploid 
barley genome by chromosomal and comparative genomics. Plant Cell 23, 1249–1263 (2011). 
44. Mayer, K. F. et al. A chromosome-based draft sequence of the hexaploid 
barley genome (Triticum aestivum) genome. Science 345, 1251788 (2014). 
45. Australian, Grain Technologies Pty. Ltd. Gliadins. Plant Varieties J. 20, 41 
2007). 
46. He, C., Holme, J. & Anthony, J. in Methods in Molecular Biology, Vol. 1145 
ed (eds D. Fleury & R. Whitford) Ch. 7, 75–86 (Springer, 2014). 
47. Cavanagh, C. R. et al. Genome-wide comparative diversity uncovers multiple 
targets of selection for improvement in hexaploid wheat landraces and 
multivars. Proc. Natl Acad. Sci. 110, 8057–8062 (2013). 
48. Wang, S. et al. Characterization of polyploid wheat genomic diversity using 
a high-density 90,000 single nucleotide polymorphism array. Plant Biotechnol. J. 12, 
49. Hannover, Germany, 1999) 
50. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding 
pre-assembled contigs using SPSPACE. Bioinformatics 27, 578–579 (2011). 
51. Benfield, J. K. & Whitham, A. Gap5—editing the billion fragment sequence 
assembly. Bioinformatics 28, 1703–1703 (2012). 
52. Edwards, D. et al. Bread matters: a national initiative to profile the genetic 
diversity of Australian wheat. Plant Biotechnol. J. 10, 703–708 (2012). 
53. Choulet, F. et al. Structural and functional partitioning of bread wheat 
chromosome 3B. Science 345, 1249721 (2014). 
54. Kovalchuk, N. in Methods in Molecular Biology, Vol. 1145 (eds D. Fleury 
& R. Whitford) 239–252 (Springer, 2014). 
55. Burton, R. A., Shirley, N. J., King, B. J., Harvey, A. J. & Fincher, G. B. The CesA 
gene family of barley. Quantitative analysis of transcripts reveals two groups of 
co-expressed genes. Plant Physiol 134, 224–236 (2004). 
56. Langmead, B. & Saltzman, S. I. Fast gapped-read alignment with Bowtie 2. Nat. 
Methods 9, 357–359 (2012). 
57. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of 
insertions, deletions and gene fusions. Genome Biol. 14, R36 (2013). 
58. Trapnell, C. et al. Differential gene and transcript expression analysis of RNA-seq 
experiments with TopHat and Cufflinks. Nat. Protoc. 7, 562–578 (2012). 
59. Kearse, M. et al. Geneious basic: an integrated and extendable desktop software 
platform for the organization and analysis of sequence data. Bioinformatics 28, 
1647–1649 (2012). 
60. Ananiev, E. V. et al. Artificial chromosome formation in maize (Zea mays L.). 
Chromosoma 118, 157–177 (2009). 
61. Rudrappa, T. & Girhepuje, P. V. Method of sunflower regeneration and 
transformation using radicle free embryonic axis. US patent US9013377 (2014). 
62. Komats, Y., Hiei, Y., Saito, Y., Murai, N. & Kumashiro, T. Vectors carrying two 
separate T-DNAs for co-transformation of higher plants mediated by 
Agrobacterium tumefaciens and segregation of transformants free from 
selection markers. Plant J. 10, 165–174 (1996). 
63. Gao, H. et al. Heritable targeted mutagenesis in maize using a designed 
edendronuclease. Plant J. 61, 176–187 (2010). 
64. Vickers, C. E., Schenk, P. M., Li, D., Mullineaux, P. M. & Gresshoff, P. M. 
pGFPUSPlus, a new binary vector for gene expression studies and optimising 
transformation systems in plants. Biotechnol. Lett. 29, 1793–1796 (2007). 
65. Iwasa, Y., Tsunashima, M., Hiei, Y. & Komari, T. Wheat (Triticum aestivum 
L.) transformation using immature embryos. Methods Mol. Biol. 1, 189–198 
(2015). 
66. Cho, M. J., Ellis, S. R., Gordon-Kamm, W. J. & Zhao, Z. Y. Methods and 
disproportions of producing and selecting transgenic plants Worldwide patent 
WO2014093445 (2014). 
67. Fletcher, S. J. qPCR for quantification of transgene expression and 
determination of transgene copy number. Crop Breed. Methods Protoc. 1145, 
213–237 (2014). 

Acknowledgements 
We thank J. Hayes for critical discussions and reading of the manuscript, P. Warner, 
M. Ovittigala, S. Manning, A. Okada, Y. Li, and M. Kumar for technical assistance. 
We also thank C. Kastner and M.-J. Cho for transgenic line generation, D. Dias and U. Roessner 
for metabolomic profiling and G. Mayo for microscopy assistance. This research was
funded by DuPont Pioneer Hi-Bred International Inc. We are grateful for the support provided by The University of Adelaide, Australian Research Council, Grains Research and Development Corporation and the South Australian State Government.

Author contributions
E.J.T., U.B., A.K., M.B., T.O., M.G., C.D., Y.W., A.S., P.L., P.W., M.C.A., A.M.C. and R.W.: Designed the experiments. E.J.T., A.K., M.B., T.O., C.D., R.S., A.S. and M.S.: Performed the experiments. E.J.T., U.B., A.K., C.D., R.S., A.M.C., and R.W.: Analysed the data. E.J.T., P.L., A.M.C., M.C.A., and R.W.: Wrote the manuscript.

Additional information
Supplementary Information accompanies this paper at doi:10.1038/s41467-017-00945-2.

Competing interests: M.A., U.B., M.C., M.S., E.T. and R.W. have filed a patent on this work. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2017