The asparagus genome sheds light on the origin and evolution of a young Y chromosome

Alex Harkess et al.#

Sex chromosomes evolved from autosomes many times across the eukaryote phylogeny. Several models have been proposed to explain this transition, some involving male and female sterility mutations linked in a region of suppressed recombination between X and Y (or Z/W, U/V) chromosomes. Comparative and experimental analysis of a reference genome assembly for a double haploid YY male garden asparagus (Asparagus officinalis L.) individual implicates separate but linked genes as responsible for sex determination. Dioecy has evolved recently within Asparagus and sex chromosomes are cytogenetically identical with the Y, harboring a megabase segment that is missing from the X. We show that deletion of this entire region results in a male-to-female conversion, whereas loss of a single suppressor of female development drives male-to-hermaphrodite conversion. A single copy anther-specific gene with a male sterile Arabidopsis knockout phenotype is also in the Y-specific region, supporting a two-gene model for sex chromosome evolution.

DOI: 10.1038/s41467-017-01064-8

Correspondence and requests for materials should be addressed to Y.Y. (email: yinye@genomics.cn) or to J.H.L.-M. (email: jleebensmack@uga.edu) or to G.C. (email: genebksh@hotmail.com).

#A full list of authors and their affiliations appears at the end of the paper.
Over the last century, cytological and genetic evidence has implicated sex chromosomes as controlling sex determination in many dioecious species across plants, animals, and fungi but much remains unclear concerning their origin and early evolution from autosomes. Whereas the concomitant origins of separate sexes and sex chromosomes are ancient in metazoan lineages, the evolution of dioecy has occurred independently many times in the evolutionary history of flowering plants, making angiosperms ideal systems for investigating the evolution of sex chromosomes across various time scales. Roughly 6% of all angiosperms are dioecious but the genes controlling sex determination and their organization on sex chromosomes in dioecious plants are largely unknown.

On the basis of his own work in Silene latifolia (formerly Melandrium album) and a review of data from other dioecious plant systems including garden asparagus, Mogens Westergaard posited in 1958 that the evolution of an active Y chromosome from an autosome involved at least two genes: a dominant suppressor of female function and a gene essential for male function that is missing from the X chromosome. This two-gene hypothesis was later advanced by Brian and Deborah Charlesworth in an evolutionary model for the conversion of an autosomal chromosome pair to sex chromosomes in association with a transition from hermaphroditism to dioecy. Alternatively, a single-sex determination gene could dominantly repress female development and promote male function. Recently, Boualem et al. have elegantly engineered a transition from monoecy to dioecy in melon (Cucumis melo) by selecting on natural variation to synthesize a population that is fixed for a null form of the feminizing gene CmACS11 and segregating for a functional C2H2 zinc finger transcription factor gene, WIP1, that has been shown to suppress both carpel development and stamen development. As had been demonstrated earlier in maize and melon, fixation of a null mutation at one locus in a monecious population set the stage for segregation of a null mutation at a second locus, resulting in dioecy and single-gene sex determination. Indeed the experimental one-gene sex determination systems described for melon and maize involve mutations in two genes, but it is the segregation of a single functional gene that ultimately determines sex. A single non-coding RNA has also been hypothesized as a sex determination gene in persimmons (Diospyros). It is important to note that one- and two-gene models for the origin of dioecy and sex determination are not mutually exclusive explanations for the numerous independent origins of dioecy and sex chromosomes across the angiosperms.

Here we directly test the one- and two-gene hypotheses for the origin of sex determination in the genus Asparagus, using garden asparagus (Asparagus officinalis L.) as a representative of the derived dioecious clade within the otherwise hermaphroditic genus. Phylogenetic analyses suggest that dioecy evolved once.
or potentially twice within a single clade in the *Asparagus* phylogeny, coincident with increases in genome size and repetitive DNA content, as well as a range expansion from southern Africa into northern Africa and Europe. In garden asparagus, X and Y chromosomes are cytologically homomorphic, and YY "supermale" genotypes are viable and produce fertile pollen, suggesting a recent origin of dioecy and sex chromosomes in the genus.

Male and female organs are both initiated in developing garden asparagus flower buds, and differences in the timing of cessation of male pistil development and degeneration of female anthers are suggestive of a two-gene sex determination system. In XX females, anther development ceases before pollen microsporogenesis occurs and is marked by a degeneration of the tapetal layer. In males, the stylar tube and receptive stigma fail to fully form, resulting in a functionally male flower.

Rigorous testing of the one and two-gene models for sex determination requires comprehensive characterization of the sex determination locus and the function of genes that segregate perfectly with sex. To that end, we have generated an annotated genome assembly for a YY male garden asparagus genotype, and using resequencing data from additional YY male and XX female genotypes we have characterized the non-recombining portion of the Y chromosome. Further, resequencing of deletion mutants exhibiting male to female conversion patterns and a previously mapped Y-linked marker (*Asp1-T7*) identified a non-recombining region covering approximately one megabase on the 132.4 Mb (~0.75%) Y chromosome. As seen in papaya, most of the non-recombining sex determination region of the Y chromosome is hemizygous with an increased density of retrotransposons compared to the surrounding pseudoautosomal region (Fig. 1a). Twelve of thirteen gene predictions in the non-recombining sex determination region of the YY males are missing from XX females (Fig. 1b; Table 1; Supplementary Note 2; Supplementary Note 3). Five of the predicted Y-linked gene model annotations do not have clear hits against the NCBI nr or Swiss-Prot databases; these gene models could represent either lineage-specific genes or artifacts of the annotation process.

Among the Y-specific genes with identified matches in other plant species is a single copy homolog of DEFECTIVE IN TAPETUM DEVELOPMENT AND FUNCTION 1 (TDF1). Female organ development is unperturbed in *Arabidopsis TDF1* knockouts, but they do not produce pollen grains due to degeneration of the tapetum. In asparagus XX females, anthers typically abort shortly before pollen is formed, exhibiting a clear degeneration of the tapetal layer and interruption of the microsporogenesis transcriptional program. On the basis of its male-sterile knockout phenotype in *Arabidopsis*, the single copy, Y-specific *aspTDF1* is an obvious candidate for the sex-linked male promoting gene in garden asparagus. *AspTDF1* has also recently been confirmed as male-specific in all tested garden asparagus cultivars and other diecious *Asparagus* species closely related to garden asparagus. Single-gene knockouts of all Y-linked genes would be necessary to rigorously test whether *aspTDF* is the one and only Y-specific promoter of male function in garden asparagus. The influence of other Y-specific genes, including a multi-copy ethylene response gene with an APETALA 2 (AP2) domain (gene model 1.235), is unknown.

Deletion mutants exhibiting male to female conversion. Experimental mutants were generated through cobalt-60 gamma irradiation of XY male seeds derived from a cross of an XX seed parent with a YY pollen donor (Supplementary Note 2; 4). Combining genetic and optical map data, 93.7% of the genome assembly was placed within recombination intervals on the 10 linkage groups in the genetic map, and 29.3% of the assembly was ordered and oriented within recombination intervals. Although the genome is largely comprised of recently-inserted Long Terminal Repeat (LTR) retrotransposons (Supplementary Fig. 1), nearly 95% of the Core Eukaryotic Gene Model Annotations (CEGMA) and 88.2% of the BUSCO plant gene annotations were identified in the asparagus gene models. Sex-linked segregation patterns and a previously mapped Y-linked marker (*Asp1-T7*) identified a non-recombining region covering approximately one megabase on the 132.4 Mb (~0.75%) Y chromosome. As seen in papaya, most of the non-recombining sex determination region of the Y chromosome is hemizygous with an increased density of retrotransposons compared to the surrounding pseudoautosomal region (Fig. 1a). Twelve of thirteen gene predictions in the non-recombining sex determination region of the YY males are missing from XX females (Fig. 1b; Table 1; Supplementary Note 2; Supplementary Note 3). Five of the predicted Y-linked gene model annotations do not have clear hits against the NCBI nr or Swiss-Prot databases; these gene models could represent either lineage-specific genes or artifacts of the annotation process.

### Results

**Genome assembly and Y chromosome annotation.** To generate a high quality reference assembly and annotation for the 1.3 Gb ICh garden asparagus genome, a fully homozygous doubled haploid YY individual derived through anther culture was sequenced. We employed a suite of complementary technologies including a variety of short and long-read sequencing strategies, BioNano Genomics optical mapping, resequencing of a diversity panel of additional doubled haploid YY males and XX females, and mapping of recombination events in a population of double haploid YY and XX siblings derived from anther culture using a single XY male parent (Supplementary Notes 1–3; Supplementary Tables 1–4; Supplementary Data 1; Supplementary Figs. 1–2).

### Table 1 Annotations of the 13 hemizygous non-recombining genes on the Y chromosome

| Gene ID                 | Annotation                                                                 |
|-------------------------|---------------------------------------------------------------------------|
| AsparagusV1_01.231      | Suppress of Female Function (SOFF)                                        |
| AsparagusV1_01.232      | Unannotated                                                               |
| AsparagusV1_01.233      | Transformation/transcription domain-associated protein                     |
| AsparagusV1_01.234      | Unannotated                                                               |
| AsparagusV1_01.235      | AP2 ethylene-responsive transcription factor                               |
| AsparagusV1_01.236      | Nudix hydrolase 15, mitochondrial-like                                     |
| AsparagusV1_01.240      | DEFECTIVE IN TAPETUM DEVELOPMENT AND FUNCTION 1 (TDF1)                    |
| AsparagusV1_01.247      | Unannotated                                                               |
| AsparagusV1_01.248      | Outer envelope protein 80, chloroplastic-like isoform X1                  |
| AsparagusV1_01.272      | Photosystem I reaction center subunit XI, chloroplastic                   |
| AsparagusV1_01.273      | Photosystem I reaction center subunit XI, chloroplastic                   |
| AsparagusV1_01.274      | Unannotated                                                               |
| AsparagusV1_01.275      | Unannotated                                                               |

* Denotes gene models placed within the sex-linked recombination bin, but an small contigs that could not be anchored onto the optical map.
Supplementary Fig. 3). Seeds were sown and flowers were phenotyped 18 months later. Plants producing fruits were genotyped using microsatellite markers to verify that they were XY products of the parental XX×YY cross. Three verified XY progeny that produced only female flowers were resequenced to an average depth of ×10 (Supplementary Table 5). In line with our prediction that the non-recombining region contains all genes necessary for sex determination in males, the entire non-recombining region of the Y chromosome was found to be independently deleted in all three mutants. Consistent with either a one-gene or multi-gene model for the genetic basis of sex determination in garden asparagus, these results indicate that the deleted portions of these Y chromosomes encode one or more factors that dominantly repress female development and promote male function.

**Male to hermaphrodite conversion mutants.** Inspection of flowers from an additional berry-producing irradiation mutant revealed that all of its flowers were hermaphroditic. Resequencing of this individual indicated a localized deletion limited to a single Y-specific gene containing a DUF247 domain (Domain of Unknown Function 247, PF03140) (Fig. 2). In addition, an independently derived spontaneous XY male-to-hermaphrodite mutant was identified in a separate germplasm collection. RNA-seq analysis of this plant and targeted sequencing of the Y-linked DUF247 domain-containing gene implicated a frameshift mutation as responsible for the sexual conversion in this second XY hermaphrodite (Fig. 2; Supplementary Note 2; Supplementary Fig. 4). Altogether, these two independent mutants provide strong evidence for the Y-specific DUF247 gene as responsible for female suppression; we have named this gene **SUPPRESSOR OF FEMALE FUNCTION (SOFF)**. Interestingly, a DUF247 domain-containing gene has been identified as the male component of the self-incompatibility S-locus in perennial ryegrass. Consistent with the two-gene model for the origin of dioecy and incipient sex chromosomes, neither hermaphroditic asparagus mutants displayed disruption of anther development (Fig. 2b), strongly suggesting that at least one other Y-specific gene promotes male function in garden asparagus. As described above, we hypothesize aspTDF1 as a male promoter but functional analysis of this and all other Y-specific genes is necessary to test whether introduction of aspTDF1 into an XX genotype would be both necessary and sufficient for conversion from female to hermaphrodite flower development.

**Small RNAs exhibiting sex-biased expression.** Several recent studies have identified sex chromosome genes with floral function, including a sex-linked small RNA-producing gene that ultimately represses anther development in diecious persimmon. Comparisons of small RNA abundance between male and female vegetative and reproductive tissues in garden asparagus show differential sRNA patterns between sexes, but none of the genes within the non-recombining region of the Y exhibit evidence of post-transcriptional regulation involving sRNAs (Fig. 3; Supplementary Note 4; Supplementary Figs. 6–8; Supplementary Tables 8–12; Supplementary Data 2–3). Interestingly one of the 167 annotated miRNAs in the asparagus genome, a copy of miR535a, is transcribed from the hemizygous region of the Y chromosome. However, a second copy of the MIR535a locus was identified on chromosome 9. Consequently, mature miR535a transcripts were found in all tissues sampled from both males and females and there was no consistent pattern of sex-biased expression tested across three distinct garden asparagus lines (Fig. 3; Supplementary Note 4). Six miRNAs—aof-miR167b, aof-miR2118a, aof-miR2118c, aof-miR2275a, aof-miR2275b, and aof-miR8026—did exhibit consistent male-biased expression in spear tips sampled from all three garden asparagus lines (Fig. 3b).
From the hypothesized sex determination genes aspTDF1 the anther or pistil developmental pathways that act downstream sex-biased expression of miRNAs may be triggered at points in variation in the sex determination region over time expression.

Asparagus includes just over 200 species, the vast majority of which are hermaphrodites\(^1\),\(^4\),\(^18\). Previous work has shown that dioecy arose within a single clade in the Asparagus phylogeny\(^13\),\(^14\). To assess whether SOFF and aspTDF1 arose as male-specific sex determination genes with the origin of dioecy in Asparagus, we performed additional whole genome resequencing on an XY male and XX female Asparagus cochinchinensis accessions (Supplementary Note 3; Supplementary Table 6). The A. cochinchinensis and A. officinalis lineages diverged very early within the diecious Asparagus clade\(^13\),\(^15\). In addition, hermaphrodite Asparagus virgatus was resequenced, and reads from both species were aligned to the A. officinalis reference genome. A SOFF ortholog was identified in the XY male A. cochinchinensis genotype but not in the XX female genotype. Ortholog identification was based on the observation that the putatively Y-specific DUF247 genes in A. cochinchinensis and A. officinalis both have longer open reading frames in intron 1, and further phylogenetic analysis of all sampled SOFF/DUF247 homologs placed the putative A. cochinchinensis SOFF ortholog as sister to the A. officinalis SOFF gene (Fig. 4, Supplementary Note 2). These findings imply that a gene duplication event set the stage for the evolution of the female suppressing function and the origin of dioecy in Asparagus.

Interestingly, A. cochinchinensis reads from both male and female genomes mapped across the aspTDF1 exons and no
nucleotide differences were observed between reads derived from the male and female genotypes. This finding suggests that the aspTDF1 ortholog resides in a recombining portion of the A. cochinensis genome, and unlike SOFF, aspTDF1 may not have contributed to the origin of dioecy within Asparagus. The PCR-based survey of Murase et al.29 also found evidence for full length aspTDF1 genes in females of A. cochinensis and A. stipularis, two dioecious Asparagus species that are distantly related to garden asparagus within the dioecious Asparagus clade14,15. It is possible that aspTDF1 is functionally silenced in A. cochinensis and A. stipularis females. For example, epigenetic silencing of the CmWIP1 transcription factor gene in melon (Cucumis melo) has been shown to promote production of female flowers32. In any case, it seems that recombination is impeding divergence of aspTDF1 alleles sampled in the resequenced A. cochinensis male and female genotypes, suggesting that the male-specific non-recombining portion of the Y-chromosome in garden asparagus has either expanded to include aspTDF1, or aspTDF1 has been secondarily recruited from elsewhere in the genome since the origin of a proto-Y chromosome in the last common ancestor of all dioecious Asparagus species.

**Major polyploidy events predate dioecy in Asparagus.** The haploid genome size of dioecious Asparagus species has been characterized as being nearly double that of hermaphroditic species16–18. Earlier work suggested that the increased genome size may have been due to increased rates of retrotransposon accumulation in dioecious Asparagus genomes, but synteny analysis of the reference genome assembly revealed evidence for at least two ancient whole genome duplications (WGDs; Fig. 5). Phylogenomic analyses were performed in order to assess the timing of these WGD events in relationship to duplication events in the DUF247 gene family (Fig. 4) and the origin of dioecy in the Asparagus genus. All asparagus gene models were clustered into gene families with genes from 15 other sequenced flowering plant genomes (Supplementary Table 7). Transcript assemblies for hermaphroditic species Asparagus asparagoidea, Yucca aloifolia and Acorus americanus were sorted into these gene family clusters using BLAST. Phylogenetic analyses of the resulting gene family circumscriptions indicated that duplicates mapping to syntenic blocks within the garden asparagus genome were found to have diverged before the origin of Asparagus, and well before the origin of dioecy within the genus. Reconciliation of gene trees and species trees placed two WGD events within the Asparagales lineage following divergence from the Orchidaceae: Asparagales-α, prior to the divergence of A. officinalis and A. asparagoidea and another event, Asparagales-β, predating the divergence of Yucca and Asparagus (Fig. 5). Phylogenetic analyses of homeologs on the Y chromosome and a syntenic portion of chromosome 5 (Fig. 4) associated their duplication with the Asparagales-α WGD, prior to the DUF247 duplication that spawned SOFF, before the origin of dioecy within Asparagus, and before geographic expansion of the genus from southern Africa to northern Africa, Europe and Asia14,15,18.

**Broader phylogenomic analysis of duplication events across the sampled species refined the placement of ancient WGD events that had been inferred in previous analyses of monocot genomes (Fig. 5).** Multiple putative WGD events were identified across the taxa sampled for our phylogenomic analysis, including both previously identified and novel events. Gene family trees were queried to estimate the timing of duplication for synteny-derived paralog pairs identified in A. officinalis (Supplementary Fig. 5). Of these trees, 1149 were found to contain syntelog pairs within the garden asparagus genome. Known WGD events were identified as unique subtrees within these 1149 gene trees, including (as numbered in Fig. 5): Poaceae—(1) rho (433 gene tree clades with Bootstrap Values of 80% or greater (BSV 80)), Poales—(2) sigma (78 BSV 80), monocots—(11) tau (56 BSV 80), Orchidaceae—(10) orchid WGD (336 BSV 80), and eudicots—(19) gamma (212 BSV 80) (Fig. 5). The two events inferred within the Asparagaceae were Asparagales-β shared with Yucca (8) (135 BSV 80) and Asparagales-α just predating divergence of the two included Asparagus species (7) (262 BSV 80) (Fig. 5). More work is necessary to properly model the gene loss process following whole genome duplications in order to rigorously test whether the clear patterns seen in our synteny and phylogenomic analyses can be confidently attributed to genome duplications. Nonetheless, both synteny and gene tree analyses clearly indicate that the two rounds of duplication within the Asparagales lineage leading to Asparagus significantly predated the origin of dioecy within the genus.

**Discussion**

Taken together, the body of data presented here strongly supports that the evolution of an active Y chromosome in Asparagus was mediated by gene duplication and neofuntionalization of a female
suppressor and subsequent expansion of the non-recombining sex determination region as predicted in the two-gene model.10-11

Given the immense diversity in sexual systems across the angiosperms and numerous independent origins of dioecy,5,7 we expect different genetic factors will be regulating male and female function in unrelated dioecious lineages. The work done to date suggests that single-gene sex determination systems may evolve following fixation of null mutations in unlinked but interacting sex specification genes,8-11 or as we show here for garden asparagus, two or more linked genes may act independently during development of female or male reproductive pathways.

Methods

Genome sequencing and initial assembly. A doubled haploid YY genotype (DH00/086) was generated through anther culture33. Nearly 341 Gb of a variety of short insert paired-end libraries (170, 200, 500, 800, 2000 nt) to larger mate-pair Illumina reads were generated for DH00/086, utilizing insert sizes that ranged from 5 to 40 kb. An initial assembly was produced using SOAPdenovo24, gap-filled with GapCloser (http://sourceforge.net/projects/soapdenovo2/files/GapCloser/), and further scaffolded with SSPACE. Additional detail is provided in Supplementary Note 1.

To improve the contiguity of the Illumina-based SOAPdenovo assembly, we generated 6.07Gb of PacBio long reads for further gap-filling and scaffolding. High molecular weight DNA isolated from the reference DH00/086 individual was used as input for library preparation, size selected for >20 kb fragments using the BluePippin (Sage Science, Beverly, MA) and sequenced on a PacBio RS II. PBelly25 version PBSuite_14.7.14 was used to improve the existing scaffold assembly with parameters “-minMatch 8 -sdpTupleSize 8 -minIdentity 75 -besn 5 -nCandidates 10 -maxScore 500 -noSplitSsReads”, followed by an additional run of GapCloser. Additional detail is provided in Supplementary Note 1.

Genetic mapping. To anchor the assembly onto pseudomolecules representing the 10 haploid chromosomes in garden asparagus, we utilized a population of 74 doubled haploid individuals to generate a genetic map by low depth (>3.5 average depth of coverage) resequencing. The doubled haploid individuals, 35 XX females and 39 YY supermales, were all derived from anther culture in a single XY male that was also sequenced. Additionally, the doubled haploid male and female parents of that XY male (i.e., grandparents of the double haploid progeny array) were also resequenced, so that all possible segregating alleles could be identified.

The 77 doubled haploid offspring were resequenced to an average of >3.5 coverage using Illumina PE150 reads. All reads were aligned to the contigs using BWA version 0.7.10-r789 with default parameters. Samtools37 version 1.2 mpileup was used to combine the results for all individuals, retaining the read depth for each allele at each locus (-1 DP) and filtering for polymorphic SNPs using bcftools version 1.2. SNPs with average read depth of <1 across the population, indels, SNP with >2 alleles and SNPs with fewer than 5 non-reference sequence reads across the whole population were removed using vcf tools v0.11.12a. Genotypes for each individual were called as AA, AB, BB, or missing data from the mpileup output. Preliminary SNPs were further filtered by removing SNPs where the genotype quality score was < 900, had > 4 heterozygous individual scores called out of 77 individuals, or had a combined read depth < 120 or > 500 (average x1.66 and x6.95 coverage) across all individuals. The filtered variant set includes 3,352,321 SNPs, or an average of one SNP per 343 bp.

SNPs from the same sequence contig were combined to create a consensus genotypic score for each contig. A total of 2864 chimeric joins between contigs within assembled sequence scaffolds were identified and placed in the proper recombination bin.

The scaffolds from the sequence genome assembly were mapped onto the optical map contigs using the Comparison tool in Irysview software (BioNano Genomics; Fig. 1b, Supplementary Fig. 1). The resulting super-scaffolds were then mapped to the male optical map and genetic SNP map. Ordered and oriented segments were put together as contiguous super-scaffolds within each recombination interval. Unordered scaffolds were placed within the genetic map at the distal end of each recombination interval. Only 6.2% of the genome assembly could not be assigned to chromosomal locations in the linkage map. Additional detail is provided in Supplementary Note 1.

Assembly correction and super-scaffolding with optical maps. The genetic mapping data were used to identify chimeric contigs or scaffolds in the genome assembly resulting from mis-assembly. Contigs with five or more consecutive SNPs that mapped to two different genetic loci > 5 cm apart were cut into separate contigs. Scaffolds that contained contigs mapping to different loci were broken. A total of 788 scaffolds were placed within the genetic map data conflict with the optical map assignments. In cases where 2 or more sequence contigs mapped to the same part of an optical map, the sequence contig with the lower e-value was removed. In total 80.4% of the sequence assembly could be placed on the optical maps, and the remaining contigs were generally too small to contain enough recombination sites to assign a location. The second optical map of a sibling doubled haploid XX female asparagus plant was used to order and orient a small fraction (<1%) of the genome that was genetically mapped but not oriented by the male optical map and genetic SNP map. Ordered and oriented segments were put together as contiguous super-scaffolds within each recombination interval.

Repetitive element annotation. LTRHarvest (GenomeTools v1.5.1) was used to annotate full-length LTR retrotransposons using default parameters except for ‘–minIdentity 0.5 –maxdist 40000’ to identify older insertions and longer fragments. A total of 57,045 full-length LTR retrotransposons were annotated with LTRHarvest and classified as either Gypsy or Copia using RepeatClassifier. The 57,045 full-length LTR retrotransposons were additionally annotated using the RepeatClassifier program (v1.0.8), part of the RepeatModeler package.

We also utilized RepeatMasker with a custom transposon database with options ‘RepeatMasker -e wublast -pa 10 -lib Asparagus.repeatexplorer.annotatedRepeats.fa’ to identify older insertions and longer fragments. A total of 57,045 full-length LTR retrotransposons were annotated with LTRHarvest and classified as either Gypsy or Copia using RepeatClassifier. The 57,045 full-length LTR retrotransposons were additionally annotated using the RepeatClassifier program (v1.0.8), part of the RepeatModeler package.

We also utilized RepeatMasker with a custom transposon database with options ‘RepeatMasker -e wublast -pa 10 -lib Asparagus.repeatexplorer.annotatedRepeats.fa -no_iso -html -gff’. The custom transposon database was derived from clustering and annotation of low coverage whole genome shotgun 454 reads48 using the RepeatExplorer pipeline, freely available in Dryad (http://dx.doi.org/10.5061/dryad.ik4s3). In total, 74.08% of the genome was masked using the custom repeat database.
**Gene annotation.** We leveraged a suite of ab initio and evidence-based programs for gene annotation. RNA-Seq reads from Harkness et al.27 of vegetative shoot and reproductive speckle tissue, in addition to root tissue (BioSample SAMN05366843) were individually aligned to the genome using TopHat2 v2.0.14 with the fragment size set to 250 bp and std. deviation to 100 bp. The min and max intron size were set to 30 and 25,000 bp, respectively. Each of the mapped RNA-Seq libraries were then processed using the Cufflinks transcript assembler (http://cole-trapnell-lab.github.io/cufflinks/ version 2.2.1). Multiple Cufflinks assemblies were then merged using Cuffmerge application to produce a combined cufflinks transcriptome assembly. A total of 81,549 transcripts were identified in the combined assembly. A comprehensive transcriptome database was built using de novo and genome-guided RNA-Seq assembly. The PASA pipelines (http://pasqpipeline.github.io) hybrid approach to transcript reconstruction protocol was used to construct the transcriptome database. The de novo and genome-guided RNA-Seq assembly was performed using Trinity RNA-Seq assembler (Ver. 2.0.2). The de novo assembly was done with in-silico normalization set to ×50 coverage. The combined Cufflinks assembly was also used. The final transcript assembly was database with 88,465 genes. The Augustus ab initio gene finder (http://bioinf.uni-greifswald.de/augustus/ version 2.5.5) was used to identify gene models. The gene finder was trained using 500 full length gene models identified from the PASA built transcriptome database. A total of 27,334 genes were identified by Augustus. The SNAP (Semi-HMM-based Nucleic Acid Parser, version 2006-07-28) ab initio gene finder was also used to identify gene models. The gene finder was trained using the same 500 full length gene models from the PASA built transcriptome database. A total of 17,096 gene models were identified by SNAP.

To leverage annotations from additional genomes to provide additional evidence for asparagus gene models, we aligned protein models from five species using Exonerate version 2.2.0: Arabidopsis thaliana, Musa acuminata, Oryza sativa, Phalaenopsis equestris, and Vitis vinifera.40–43. We combined evidence from RNA Seq, ab initio and homology-based approaches utilizing Evidence Modeler version 1.1.044 to predict 27,656 genes which were used for all downstream analysis. These gene models were compared against the Genome Gene Model Annotation (CEGMA)45 v2.5 and the BUSCO46 Plant Gene databases to assess completeness of the annotation.

**Female suppressing gene identification.** To identify a female suppressing gene on the Y chromosome we identified and sequenced several independent hermaphroditic lines that were each derived from a different XY male. Nearly 40,000 XY sequences were treated with cobalt-60 gamma irradiation to identify a single field-grown hermaphroditic mutant (G033) for whole genome resequencing (Supplementary Fig. 5). A wild type XY untreated control plant (K232) was also resequenced. Additional gamma irradiation produced 3 more male-to-female mutants (Lim_mut1, Lim_mut3, and Lim_mut4), and a control XY male plant (Lim_mut5) was also resequenced. Further, a spontaneous XY male-to-hermaphroditic mutant (Lim3) was discovered in a glasshouse. Floral bud RNA-seq of this individual, followed by genomic PCR-based confirmation, revealed a single base pair deletion in a coding exon of the FTF female suppressing gene that results in a frameshift mutation and a premature stop codon. Additional detail regarding the refinement of the non-recombining region assembly as well as mutant identification is provided in Supplementary Note 2.

**Small RNA annotation and analysis.** We generated and analyzed small RNA (sRNA) libraries from different organs and developmental stages. Asparagus officinalis samples were collected from a commercial field in the T.S. Smith and Son's Farm (http://www.tsstsmithandsons.com/), Bridgeville, Delaware, as well as from female, male, and supernumerate whole spear tip tissue from a previous study27. Samples were collected and anthers were dissected using a 2 mm stage micrometer (Wards Science, cat. #949910) in a stereo microscope, and immediately frozen in liquid nitrogen until total RNA isolation was performed. Total RNA was isolated using the PureLink Plant RNA Reagent (ThermoFisher Scientific, cat. #12322012) following the manufacturer’s instructions. Small RNAs (20–30 nt) were size selected in a 15% polyacrylamide/urea gel and used for small RNA library preparation using the Illumina TruSeq Small RNA Prep Kit. Detailed description of analysis can be found in Supplementary Note 4.

**Assembly of SOFF across Asparagus species.** To identify variation in the Y chromosome across additional species in the genus, we resequenced the hermaphroditic A. virgata and a male and female accession of dioecious A. cochlearis each to 2×20X coverage with single end 150 nt reads on an Illumina NextSeq500. To in silico assemble SOFF homologs across these individuals, SE150 reads were aligned to the genome with bwa mem v0.7.5a. Using bedtools v2.25.0, we extracted every read that aligned to the SOFF and aspTDF1 gene model exons with at least 1nt overlap. Extracted reads were then assembled for each exon in cap3 (v. 0.77) resulting in a 59% overlap identity over at least 40nt. Next they were then aligned using PASTA (https://github.com/smirarab/pasta) and a RAXML tree was generated using the CAT model within PASTA.

**Synteny and phylogenomic analysis.** We performed pairwise genome alignments between a total of 12 selected genomes, including Amborella (Amborella trichopoda),22 Musa acuminata (Musa Musa acuminata)23, date palm (Phoenix dactylifera)24, duckweed (Spirodela polyrhiza)25, grape (Vitis vinifera)26, oil palm (Elaeis guineensis)27, orchid (Phalaenopsis equestris)28, pineapple (Ananas comosus)29, rice (Oryza sativa)30, sorghum (Sorghum bicolor)31, Zosteria (Zosteria marina)32, and garden asparagus (Asparagus officinalis). For each pairwise genome alignments, the coding sequences of predicted gene models were compared to each other using LAST (last.align)28. Our synteny search pipeline defines syntetic blocks by chaining the LAST hits with a distance cutoff of 20 genes apart, also requiring at least 4 gene pairs per syntetic block (Supplementary Fig. 5).

**Gene family circumscription and gene tree estimation.** Gene models from 14 genomes (Supplementary Table 7) were clustered into gene families using OrthoFinder ver. 0.4.4.33 Gene models from the Zosteria marina and Dendrobiun catenatum33 genomes and transcriptome assemblies for Acorus curosus, Yuca aloifolia and Asparagus asparagusoides were sorted into the gene family clusters based on best BLASTX (NCBI blast+)2.2.29 matches to the gene models in the OrthoFinder circumscribed gene families.

**Interspecie peptide sequences** from the family cluster with more than five genes were using MAFFT v6.3.217 and coding nucleotide sequences were codon aligned on top of the amino acid alignments using pal2nal37. Gene trees were estimated from the resulting codon alignments using the GTR_KAMMA model in RAxML ver. 8.2.4.35.

**Phylogenetic placement of whole genome duplication events.** The timing of whole genome duplication events relative to branching events in the species tree was estimated using the PUG algorithm and software (https://github.com/mrmckain/PUG). A species tree estimated based on coalescence analysis of 1,584 gene families that were found in to exist in one and only one copy in at least 13 of the 15 genomes clustered or sorted into gene families. Multi-gene bootstrap analyses were performed in ASTRAL ver. 4.10.148 to estimate relationships among all of the species listed in Supplementary Table 7. The resulting species tree was used to infer the timing of ancient whole genome duplications (WGD) evident in syntetic blocks (see Supplementary Fig. 5) by assessing the timing of duplication events in multi-copy gene trees.

To infer the timing of WGDs, the coalesence node for each paralogs pair in each rooted multi-copy gene tree was identified and species composition of the paralog subtree above each duplication node was compared against the species tree to identify the lower bound on timing of the gene duplication event in the species tree. The node of paralogs coalescence in the gene tree represents the timing of gene duplication and in whole genome duplication, the point of genome duplication and divergence. The lineage sister to the paralog subtree was also queried for species composition to provide additional support for the lower bound of the duplication event verifying adequate sampling of the species tree in the gene tree. If the timing of a gene duplication was mapped to a single internal branch in the species tree, bootstrap values for the coalesence node in the gene tree was recorded. The timing of WGDs were inferred based on the number of gene duplications mapping to each internal node on the species tree. Possible gene duplications on terminal branches were not considered since paralogs are difficult to distinguish from single locus isofoms. Figures representing the placement of gene duplications were made using the PUG_Figure_Maker.R script available in the PUG software package (https://github.com/mrmckain/PUG). When displaying duplicated genes on species trees, only nodes with at least 10% of the maximum number of duplicated genes found at a node are shown in Fig. 5.

**Data availability.** All sequence data have been deposited in NCBI’s reference genome database (BioProject Accessions PRJNA317340 (genome shotgun reads), PRJNA259909 (RNA-Seq), PRJNA326431 (Asparagus asparagusoides transcriptome). Small RNA and PARE data are available at https://mps.danforthcenter.org/dbs/index.php?SITE=asparagus_sRNA and https://mpss.danforthcenter.org/dbs/index.php?SITE=asparagus_pare. Gene alignments and trees are available at https://github.com/alexharkess/asparagus_genome. The genome is also available through CoGe (genomestrology.org/coge/OrganismView.pl?dsgid=33908) and bulk downloads of all data are available through https://asparagus.uga.edu/ along with a genome browser and BLAST portal. The authors declare that all other data supporting the findings of this study are available from the corresponding author upon request.

Received: 11 August 2016 Accepted: 16 August 2017

References

1. Morgan, T. H. Heredity and sex. Z. Indukt. Abstamm. Vererbungsl. 12, 159–161 (1914).
2. Morgan, T. H. Recent results relating to chromosomes and genetics. Q. Rev. Biol. 1, 186 (1926).
3. Ming, R., Bendahmane, A. & Renner, S. S. Sex chromosomes in land plants. Annu. Rev. Plant Biol. 62, 485–514 (2011).
4. Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and gynodioecy. Am. Nat. 112, 975–997 (1978).
5. Charlesworth, D. & Charlesworth, B. Population genetics of partial male-sterility and the evolution of monoecy and dioecy. Heredity 41, 137–153 (1978).
6. Liu, Z. Y. et al. A primitive flowering plant Arabidopsis thaliana. Nature 408, 796–815 (2000).
7. International Rice Genome Sequencing Project. The map-based sequence of the rice genome. Nature 436, 793–800 (2005).
8. Tsugama, D. et al. A putative MYB35 ortholog is a candidate for the sex-determining gene in Arabidopsis. Plant Mol. Biol. Rep. 27, 2078–2079 (2009).
9. Bowers, J. E. et al. Development of an ultra-dense genetic map of the sunflower genome based on single-feature polymorphisms. PLoS ONE 7, 721–729 (2012).
10. Jones, D. F. Unisexual maize plants and their bearing on sex differentiation in other plants and in animals. Genetics 19, 552–567 (1934).
11. Dellaporta, S. L. & Calderon-Urrea, A. The sex determination process in maize. Science 266, 1501–1505 (1994).
12. Akagi, T., Henry, I. M., Tao, R. & Comai, L. A Y-chromosome and genome evolution across the Poales. Genome Biol. Evol. 6, 103–110 (2014).
13. Renner, S. S. Pathways for making unisexual flowers and unisexual plants: moving beyond the "Two mutations linked on one chromosome" model. Am. J. Bot. 103, 587–589 (2016).
14. Kubo, S., Konno, I. & Kanno, A. A Molecular phylogeny of the genus Asparagus (Asparagaceae) explains interspecific crossability between the garden asparagus (A. officinalis) and other Asparagus species. Theor. Appl. Genet. 124, 345–354 (2012).
15. Norup, M. F. et al. Evolution of Asparagus L. (Asparagaceae): Out-of-South-Africa and multiple origins of sexual dimorphism. Mol. Phylogenet. Evol. 92, 25–44 (2015).
16. Harkess, A. et al. Retrotransposon proliferation coincident with the evolution of dioecy in Asparagus. G3: Genes Genomes Genet. 6, 2679–2685 (2016).
17. Boetzer, M., Henkel, C. V., Jansen, H. J., Butler, D. & Pirovano, W. Scaffolding pre-assembled contigs using SSPACE. Brief Bioinform. 15, W609–W619 (2014).
18. Wang, W. et al. The Spireola polyrhiza genome reveals insights into its neotenic reduction fast growth and aquatic lifestyle. Nat. Commun. 5, 3311 (2014).
19. Singh, R. et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature 500, 335–339 (2013).
20. Ming, R. et al. The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. 47, 1435–1442 (2015).
21. Paterson, A. H. et al. The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551–556 (2009).
22. Olsen, L. et al. The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. Nature 530, 331–335 (2016).
23. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157 (2015).
24. Zhang, G.-Q. et al. The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. Sci. Rep. 6, 19010 (2016).
25. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple alignment program. Brief Bioinform. 9, 286–298 (2008).
26. Suyama, M., Torrones, D. & Bork, P. PAL2NAL: robust conversion of protein sequences to genome annotations. Nucleic Acids Res. 34, W609–W612 (2006).
27. Vezzi, L. et al. Defective in Tapetal Development and Function 1 is essential for anther development and tapetal function for microspore maturation in Arabidopsis. Plant J. 55, 266–277 (2008).
28. Renner, S. S. Pathways for making unisexual flowers and unisexual plants: moving beyond the "Two mutations linked on one chromosome" model. Am. J. Bot. 103, 587–589 (2016).
29. Harkess, A. et al. Retrotransposon proliferation coincident with the evolution of dioecy in Asparagus. G3: Genes Genomes Genet. 6, 2679–2685 (2016).
30. Wang, W. et al. The Spireola polyrhiza genome reveals insights into its neotenic reduction fast growth and aquatic lifestyle. Nat. Commun. 5, 3311 (2014).
31. Singh, R. et al. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature 500, 335–339 (2013).
32. Ming, R. et al. The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. 47, 1435–1442 (2015).
33. Paterson, A. H. et al. The Sorghum bicolor genome and the diversification of grasses. Nature 457, 551–556 (2009).
34. Olsen, L. et al. The genome of the seagrass Zostera marina reveals angiosperm adaptation to the sea. Nature 530, 331–335 (2016).
35. Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157 (2015).
36. Zhang, G.-Q. et al. The Dendrobium catenatum Lindl. genome sequence provides insights into polysaccharide synthase, floral development and adaptive evolution. Sci. Rep. 6, 19010 (2016).
37. Katoh, K. & Toh, H. Recent developments in the MAFFT multiple alignment program. Brief Bioinform. 9, 286–298 (2008).
38. Suyama, M., Torrones, D. & Bork, P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34, W609–W612 (2006).
39. Vezzi, L. et al. Defective in Tapetal Development and Function 1 is essential for anther development and tapetal function for microspore maturation in Arabidopsis. Plant J. 55, 266–277 (2008).
40. Harkess, A. et al. Sex-biased gene expression in dioecious garden asparagus (Asparagus officinalis). New Phytol. 207, 883–892 (2015).
41. Murase, K. et al. MB transcription factor gene involved in sex determination in Asparagus officinalis. Genes Cells 22, 115–123 (2016).
42. Tsgama, D. et al. A putative MYB35 ortholog is a candidate for the sex-determining genes in Asparagus officinalis. Sci. Rep. 7, 41497 (2017).
43. Manzanarez, C. et al. A gene encoding a DUFC247 domain protein co-segregates with the self-incompatibility locus in perennial ryegrass. Mol. Biol. Evol. 33, 1–40 (2015).
44. Song, M., Ma, K., Ci, D., Zhang, Z. & Zhang, D. Sexual dimorphism of floral microRNA profiling and target gene expression in andromonoecious poplar (Populus tomentosa). PLoS ONE 8, e62681 (2013).
45. Martin, A. et al. A transposon-induced epigenetic change leads to sex determination in melon. Nature 461, 1135–1138 (2009).
46. Qiao, Y. M. & Falavigna, A. An improved in vitro anther culture method for obtaining doubled-haploid clones of asparagus. Acta Hortic. 271, 145–150 (1990).
Acknowledgements

We acknowledge funding from the National Science Foundation (U.S.A.; Grants DEB-0841988, IOS-0922742, IOS-1444567, DEB-1501589) and the National Natural Science Foundation of China (Grants Nos. 30771376, 31260352, 31360426 and 31560557). We also appreciate support from Neil Stone and Mikel Roose (UC Riverside). We acknowledge the help of Alex Hastie and BioNano genomics for generating the BioNano Irisy data. Finally, we thank reviewers for their helpful comments and suggestions.

Author contributions

A.H., J.L.-M., R.V.d.H., K.Y., and G.C. framed the research; J.E.B., A.H., C.X., J.G., J.R., A.T.-R., Z.Y., H.C., W.L., Y.C., XX., J.C.P., M.L., D.K., R.A.W., and Y.Y. contributed to sequencing, mapping and genome assembly; S.A., A.H., C.X., F.M., Y.Y., and J.L.-M. contributed to genome annotation; A.K., T.A., S.M., M.N., and B.C.M. performed small RNA analyses; A.H., J.H., C.X., R.V.d.H., J.G., M.R.M., H.T., H.S., A.T.-R., A.K., F.M., P.R., J.C.P., H.K., F.S., A.F., C.X. and Y.Y. contributed to all other evolutionary analyses including phylogenomic analysis of genome duplications; J.Z., Y.Z., S.L. and G.C. conducted asparagus research at JAASX; H.C., J.G., and K.Y. conducted asparagus research at CATAS; Z.M. conducted asparagus research at YAU; R.V.d.H. and P.L. provided material for reference genome, generated irradiation mutants and conducted spontaneous male-to-hermaphrodite mutant at CREA.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01064-8.

Competing interests: The 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.

\[\text{Alex Harkess}^{1,25}, \text{Jinsong Zhou}^{2}, \text{Chunyan Xu}^{3}, \text{John E. Bowers}^{1}, \text{Ron Van der Hulst}^{4}, \text{Saravanaraj Ayvampalayam}^{1}, \text{Francesco Mercati}^{1,6}, \text{Paolo Riccardi}^{7,26}, \text{Michael R. McKain}^{8,9}, \text{Atul Kakrana}^{10}, \text{Haibao Tang}^{11}, \text{Jeremy Ray}^{1}, \text{John Groenendijk}^{12}, \text{Siwaret Arik}^{10,27}, \text{Sandra M. Mathioni}^{8,10}, \text{Mayumi Nakano}^{8,10}, \text{Hongyang Shan}^{13}, \text{Alexa Telgmann-Rauber}^{1,14}, \text{Akira Kanno}^{15}, \text{Zhen Yue}^{3}, \text{Haixin Chen}^{3}, \text{Wenqi Li}^{3}, \text{Yanling Chen}^{3}, \text{Xiangyang Xu}^{3}, \text{Yueping Zhang}^{2}, \text{Shaochun Luo}^{2}, \text{Helong Chen}^{16}, \text{Jianming Gao}^{17}, \text{Zichao Mao}^{18}, \text{J. Chris Pires}^{19}, \text{Meizhong Luo}^{20}, \text{Dave Kudrava}^{21}, \text{Rod A. Wing}^{21}, \text{Blake C. Meyers}^{8,10}, \text{Kexian Yi}^{16,17}, \text{Hongzhi Kong}^{13}, \text{Pierre Lavrijsen}^{4}, \text{Francesco Sunseri}^{5}, \text{Agostino Falavigna}^{7,22}, \text{Yin Ye}^{3,23,24}, \text{James H. Leebens-Mack}^{1} \& \text{Guangyu Chen}^{2}\]

1Department of Plant Biology, University of Georgia, Athens, GA 30602, USA. 2Institute of Vegetables and Flowers, Jiangxi Academy of Agricultural Sciences, 330200 Nanchang, China. 3BGI Genomics, BGI-Shenzhen, Shenzhen 518083, China. 4Limgroup B.V., 5961 NV Horst, The Netherlands. 5Dipartimento AGRARIA, Università Mediterranea degli Studi di Reggio Calabria, 89124 Reggio Calabria, Italy. 6Institute of Biosciences and BioResources, Division of Palermo, National Research Council, 90129 Palermo, Italy. 7Consiglio per la Ricerca in Agricoltura e l’Analisi dell’economia Agraria (CREA), Research Unit for Vegetable Crops, 26836 Montanaso Lombardo, Lodi, Italy. 8Institute of Biosciences and BioResources, Division of Palermo, National Research Council, 90129 Palermo, Italy. 9Department of Plant Biology, University of Georgia, Athens, GA 30602, USA. 10College of Agriculture and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Haixia Institute of Science and Technology (HIST), Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China. 11Green Acres Life Science, De Dreijenlaan 3, 6703 HA Wageningen, The Netherlands. 12Center for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Haixia Institute of Science and Technology (HIST), Fujian Agriculture and Forestry University, Fuzhou, Fujian 350002, China. 13Department of Plant Biology, University of Delaware, Newark, DE 19711, USA. 14Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA. 15College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China. 16Arizona Genomics Institute, School of Plant Sciences and Department of Ecology and Evolutionary Biology, Tucson, AZ 85750, USA. 17College of Agriculture and Biotechnology, Yunnan Agricultural University, Kunming 650201, China. 18State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing, China. 19KWS SAAT AG, 37574 Einbeck, Germany. 20Graduate School of Life Sciences, Tohoku University, Katahira 2-1-1, Aoba-ku, Sendai 980-8577, Japan. 21Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Hainan 571101, China. 22Environment and Plant Protection Research Institute, Chinese Academy of Tropical Agricultural Sciences, Hainan 571101, China. 23College of Agriculture and Biotechnology, Yunnan Agricultural University, Kunming 650201, China. 24Division of Biological Sciences, University of Missouri, Columbia, MO 65211, USA. 25College of Life Sciences and Technology, Huazhong Agricultural University, Wuhan, Hubei 430070, China. 26Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences, Hainan 571101, China. 27College of Agriculture and Biotechnology, Yunnan Agricultural University, Kunming 650201, China. 28Present address: Danish Danforth Plant Science Center, De Dreijenlaan 3, 6703 HA Wageningen, The Netherlands. 29Present address: Bayer, Via Ghiarone 2, 40019 Sant’Agata Bolognese, Italy. 30Present address: Department of Agronomy, Faculty of Agriculture at Kamphaeng Saen and Rice Science Center, Kasetsart University, Kamphaeng Saen, Nakhon Pathom 73140, Thailand. Alex Harkess, Jinsong Zhou and Chunyan Xu contributed equally to this work.