Two independent origins of XY sex chromosomes in *Asparagus*

Authors

1. Philip C. Bentz (1,2)* https://orcid.org/0000-0003-2232-7488
2. Sarah B. Carey (2) https://orcid.org/0000-0002-6431-0660
3. Francesco Mercati (3) https://orcid.org/0000-0003-1356-2881
4. Haley Hale (2) https://orcid.org/0000-0002-3318-8383
5. Valentina Ricciardi (3) https://orcid.org/0000-0001-7081-8906
6. Francesco Sunseri (4) https://orcid.org/0000-0001-5201-5413
7. Alex Harkess (2) https://orcid.org/0000-0002-2035-0871
8. Jim Leebens-Mack (1)* https://orcid.org/0000-0003-4811-2231

Affiliations

1. Department of Plant Biology and The Plant Center, University of Georgia, Athens, Georgia, 30605, United States
2. HudsonAlpha Institute for Biotechnology, Huntsville, Alabama, 35806, United States
3. Institute of Biosciences and BioResources, Division of Palermo, National Research Council, 90129, Palermo, Italy
4. Dipartimento Agraria, Università Mediterranea degli Studi di Reggio Calabria, 89124, Reggio Calabria, Italy

* corresponding authors: P.C.B. pbentz@hudsonalpha.org; J.L.-M. jleebensmack@uga.edu

Abstract

The relatively young and repeated evolutionary origins of dioecy (separate sexes) in flowering plants enable investigation of molecular dynamics occurring at the earliest stages of sex chromosome evolution. With two independently young origins of dioecy, *Asparagus* is a model genus for studying the genetics of sex-determination and sex chromosome evolution. Dioecy first evolved in *Asparagus* ~3-4 million years ago (Ma) in the ancestor of a now widespread Eurasian clade including garden asparagus (*Asparagus officinalis*). A second origin occurred in a smaller, geographically restricted, Mediterranean Basin clade including *Asparagus horridus*. New haplotype-resolved reference genomes for garden asparagus and *A. horridus*, elucidate contrasting first steps in the origin of the sex chromosomes of the Eurasian and Mediterranean Basin clade ancestors. Analysis of the *A. horridus* genome revealed an XY system derived from different ancestral autosomes with different sex-determining genes than have been characterized for garden asparagus. We estimate that proto-XY chromosomes evolved 1-2 Ma in the Mediterranean Basin clade, following an ~2.1-megabase inversion that now distinguishes the X and Y chromosomes. Recombination suppression and LTR retrotransposon accumulation drove the expansion of the male-specific region on the Y (MSY) that reaches ~9.6-megabases in *A. horridus*. The garden asparagus genome revealed an MSY spanning ~1.9-megabases. A segmental duplication and neofunctionalization of one duplicated gene (SOFF) drove the origin of dioecy in the Eurasian clade. These findings support previous inference based on phylogeographic analysis revealing two recent origins of dioecy in *Asparagus* and establish the genus as a model for investigating sex chromosome evolution.

Key words: asparagus sex chromosomes; convergent evolution; dioecy origins; genome inversion; young plant sex chromosomes

Significance Statement

Flowering plants with separate sexes are ideal systems for investigating genome dynamics underlying the earliest stages of sex chromosome evolution across the tree of life. We use *Asparagus* as a model to better understand early sex chromosome formation more generally, by investigating how different XY sex chromosomes evolved within two young, sister clades. Genomic comparisons of garden asparagus and *Asparagus horridus* (wild related species) revealed distinct evolutionary origins of XY-chromosomes with different sex-determination mechanisms. Whereas the garden asparagus Y-chromosome originally evolved around 3-4 million years ago (Ma), following a small segmental duplication, the Y-chromosome in *Asparagus horridus* evolved more recently (~1-2 Ma) following a large structural inversion between a different chromosome pair. Interestingly, both evolutionary transitions from hermaphroditism to separate sexes occurred as ancestors of garden asparagus and *Asparagus horridus* independently dispersed northward out of southern Africa.

Introduction

Separate sexes and sex chromosomes have evolved many times across the tree of life. Sex chromosomes exhibit unique evolutionary innovations relative to autosomes, often including regions of suppressed recombination, size differences between X and Y (or Z/W) chromosomes, and the evolution of sex-specific gene content and expression patterns. Sex-determination systems vary widely across eukaryotes: while some are ancient and conserved (e.g., the XY sex chromosomes in placental mammals), others are relatively young due to frequent turnovers or independent origins. Recurrent turnovers, replacing ancestral sex-determination loci with new ones, have been observed in a variety of animal taxa including fishes and frogs (Vicoso 2019), providing an important comparative framework for investigating young sex chromosome dynamics. While several angiosperm (flowering plant) lineages exhibit translocations of conserved sex-determining loci, including strawberry (Tennessen et al. 2018) and kiwifruit (Akagi et al. 2023), no cases of full gene-level turnover have been reported in plants.

The angiosperms arose around 139-158 million years ago (Ma) and the most recent common ancestor (MRCA) was hermaphroditic with bisexual flowers, producing both male and female gametophytes (pollen and ovules, respectively) (Sauquet et al. 2017). However, separate sexes (i.e., dioecy, or unisexual flowers on different plants) have evolved in less than 10% of angiosperm species (Renner 2014). Many of the hundreds of transitions from hermaphroditism to dioecy occurred independently and relatively recently across the angiosperms (see reviews: Carey et al. 2021; Renner and Müller 2021; Marais et al. 2025), offering an opportunity to investigate the earliest stages of sex chromosome evolution (Charlesworth and Harkess 2024). It is, however, not yet clear whether the origin and evolution of dioecy occurs through a common set of genomic and molecular mechanisms, or rather, there are a myriad evolutionary paths for the transition from autosomes in hermaphroditic species to sex chromosomes in dioecious species. By studying sex chromosomes of different evolutionary ages, we may begin to understand the molecular mechanisms and ecology driving the evolution of separate sexes more broadly. Investigations of independently evolving sex chromosomes among closely related dioecious species may be especially informative for understanding the origin and evolution of sex chromosomes.

The genus *Asparagus* Tourn. ex L. (Asparagaceae) is an important model system for studying genetic sex-determination and sex chromosome dynamics in flowering plants. Investigations of the genetic basis of sex-determination in garden asparagus (*Asparagus officinalis* L.) characterized Y-specific genes responsible for the suppression of pistil (female) development and completion of pollen (male) development (Harkess et al. 2017;

Murase et al. 2017; Tsugama et al. 2017; Harkess et al. 2020). However, little is known regarding sex-determination and sex chromosomes in the other 50+ dioecious species of *Asparagus,* partly due to historical uncertainty surrounding sexual systems and species relationships across the genus. To address these limitations, we recently reviewed all sexual systems reported in the genus (Bentz, Liu, et al. 2024) and released an updated phylogeny based on 1,726 nuclear genes and robust species sampling (>150 spp.) (Bentz, Burrows, et al. 2024). Our review of sexual systems showed that all extant *Asparagus* species are either dioecious or hermaphroditic with bisexual flowers, revealing no evidence of gynodioecy or androdioecy in the genus (Bentz, Liu, et al. 2024). According to a phylogeographic analysis and the updated *Asparagus* phylogeny, two independent origins of dioecy were strongly supported within the genus: first in a widespread Eurasian clade, then again in a geographically restricted Mediterranean Basin clade (Fig. 1) (Bentz, Burrows, et al. 2024). Interestingly, both origins of dioecy in *Asparagus* occurred within ~1-2 million years of each other and were associated with separate long-distance dispersal events out of southern Africa (Norup et al. 2015; Bentz, Burrows, et al. 2024). Two independently derived dioecious systems within *Asparagus* makes the genus ideal for testing for common themes in the molecular dynamics contributing to dioecy and sex chromosome evolution.

In garden asparagus (representing the Eurasian clade), the presence of an ~1 Mb nonrecombining, male-specific region on the Y (hereafter, "MSY") was identified in the first reference genome of a double-haploid YY male (Harkess et al. 2017). The MSY in the YY garden asparagus reference included 13 annotated genes, two of which were shown to sufficiently control sex in experimental and spontaneous mutant genotypes: *SUPPRESSOR OF FEMALE FUNCTION* (*SOFF*), a *DUF247* gene that suppresses pistil development (Harkess et al. 2017); and *TAPETAL DEVELOPMENT and FUNCTION 1* (*aspTDF1*), an *R2R3*-type *MYB* transcription factor and male-promoter gene influencing tapetal and pollen development (Harkess et al. 2017; Murase et al. 2017; Tsugama et al. 2017; Harkess et al. 2020). The presence of two sex-determining genes in garden asparagus supports a two-gene hypothesis for the evolution of sex chromosomes with at least two linked mutations affecting female and male fertility, respectively (Westergaard 1958; Charlesworth and Charlesworth 1978), rather than a single master-switch sex-determining gene as described in the Salicaceae (Müller et al. 2020). The presence of a male-specific *SOFF* ortholog in *A. officinalis* and *A. cochinchinensis* suggests that a Y-linked *SOFF* is conserved across the Eurasian dioecy clade (Harkess et al. 2017) (Fig. 1). However, PCR assays revealed that whereas *aspTDF1* is male-specific in garden asparagus and its closest relatives, it is autosomal in other dioecious species in the genus including *Asparagus cochinchinensis* (Lour.) Merr., *Asparagus acutifolius* L., and

Asparagus horridus L. (Murase et al. 2017). *Asparagus cochinchinensis* falls within a subclade that split from the subclade with garden asparagus early in the evolution of the Eurasian dioecy clade (Fig. 1); suggesting that *aspTDF1* evolved Y-linkage following the origin of dioecy within the Eurasian clade. The Mediterranean Basin dioecious clade, including *A. acutifolius*, and *A. horridus*, may have independently evolved a distinct sex-determination system.

In this study, we leverage the updated *Asparagus* phylogeny and explore the origins of recombination suppression between ancestral autosomes that led to independent sex chromosome formation in the Mediterranean Basin and Eurasian dioecious clades of *Asparagus*. Specifically, we present and compare new haplotype-resolved, reference genomes for taxa from the two dioecious clades: *A. horridus* (Mediterranean Basin) (Fig. 2) and *A. officinalis* (Eurasian). Both new genomes are derived from diploid male genotypes (2n = 2x = 20). This study is the first to present a genome assembly for *A. horridus* and identify the sex chromosomes for the species and the new *A. officinalis* is more complete than the previously published genome. Our findings advance understanding of how dioecy can evolve between two closely related lineages with unique sex chromosomes, expanding the utility of *Asparagus* as a model for the study of sex chromosome evolution more broadly.

Results and Discussion

Asparagus horridus genome and sex chromosomes

Here we present the first reference genome for *A. horridus* and report an XY sex chromosome system for the species. The *A. horridus* genome assembly size was in-line with flow cytometry-based estimates (Plath et al. 2022), totaling ~1.01-1.03 Gb per haplotype (Table S1). Presence of a larger Y chromosome (~107.2 Mb) in haplotype 1, compared to the X (~99.4 Mb) in haplotype 2, largely explains the size difference between the two haplotype assemblies (Table S2). Both *A. horridus* haplotype assemblies and annotations (Figs. S1-S2) passed our quality control thresholds (see Materials and Methods) aside from ratios of multi-exon to single-exon genes, which were about two times greater than the ~0.20 ratio found in model plant genomes (Vuruputoor et al. 2023) (Table S1). Fewer genes were predicted in *A. horridus* (31,194 to 31,235 genes per haplotype), compared to the new *A. officinalis* genome annotation (Table S1), but were generally in-line with *A. officinalis* gene counts based on histone modification (ChIP-seq) data (Mendieta et al. 2021).

We identified the sex chromosomes and delimited the putative nonrecombining (X- or Y-specific) region from the surrounding pseudo-autosomal regions (PARs), which are

thought to recombine relatively freely (Otto et al. 2011), based on 1) structural variation between haplotypes (Figs. 3a, S3), 2) the highest density of male-specific *k*-mer (hereafter "Y-mer") alignments (Figs. 3a, S4), and 3) a lower ratio of female:male read mapping depth/coverage (Fig. S4). Y-mers were identified from whole genome sequencing reads for 8 total males and 7 females of *A. horridus* (Table S3), wherein putative Y-mers were required to be present in all males and absent in all females. Approximately 96% of Y-mers mapped to an ~9.6 Mb region on chromosome 3 of haplotype 1, corresponding to the putative MSY and Y chromosome in *A. horridus* (Table S2). Intriguingly, the entire length of the *A. horridus* MSY is encompassed by a single inversion between the X and Y; the endpoints of which mark the boundaries between the PARs and the nonrecombining MSY (yellow block in Fig. 3a). A small region (~0.65 Mb), containing three genes, within the ~9.6 Mb MSY was found to be oriented in the same direction as the X (dark grey region within the large red inversion in Fig. 3a). However, we cannot infer whether the large, MSY-spanning, or small, internal inversion occurred first, since there is complete overlap in the per-site synonymous substitution estimates for nonrecombining X-Y homologs shared between the sex chromosomes (i.e., gametologs) within and outside the smaller internal inversion (shaded region in Fig. 3d). Read and Y-mer mapping depths (Fig. S4) and an X-Y dot plot (Fig. S5) provide additional support for the inferred boundaries of the nonrecombining region and nested inversions, respectively.

The MSY in *A. horridus* is relatively gene poor (122 non-TE gene predictions, see Table S4 for functional annotations) relative to adjacent PAR segments and the homologous region of the X chromosome and is highly enriched in TE content, especially of *Ty3*- and *Ty1*-type long-terminal repeat (LTR) retrotransposons (Fig. 3a; Table S2). Retrotransposons, and repeats more generally, are thought to lead to artificially inflated gene model estimates, especially of single-exon genes (Vuruputoor et al. 2023). In the *A. horridus* MSY, 62 of the predicted genes were mono-exonic (Tables S2) and ~71% of those single-exon genes lacked orthologs in other species and therefore may include erroneous gene models. The *A. horridus* MSY corresponds to a collinear, but inverted, region on the X that spans ~2.1 Mb (Fig. S3) with 77 gene models (~23% mono-exonic) and considerably lower repeat content than the MSY (Table S2). Finally, and perhaps most importantly, no shared homologs were found between the *A. horridus* and *A. officinalis* MSYs, supporting independent origins of genetic sex-determination and XY chromosomes in the two dioecious clades in the genus *Asparagus*.

Origin of XY sex chromosomes in the Mediterranean Basin clade

Sex chromosomes evolved only once in the Mediterranean Basin clade, as indicated by the conservation of male-specific sequences (Y-mers) shared between *A. horridus* and *A. acutifolius* (Fig. S4). The ~9.6 MSY in *A. horridus* is syntenic with ~2.1 Mb on chromosome 3

in *A. officinalis*, *A. setaceus*, and the *A. horridus* X. This suggests that proto-XYs in the Mediterranean Basin clade evolved following an ~2.1 Mb inversion between the ancestral chromosome 3 pair, inhibiting recombination throughout the region and allowing for X-Y divergence. Expansion of the nonrecombining MSY—as a consequence of TE accumulation—is thought to be a dominant process contributing to disparate sizes between the sex chromosomes of some lineages (Steinemann and Steinemann 2005). Compared to the X-limited region, the MSY in *A. horridus* exhibits a roughly four times greater ratio of repeats (Table S2) supporting TE enrichment as a major driver of its expansion over time (see significantly increased TE density throughout the MSY in Fig. 3a).

Little or no degeneration in the nonrecombining MSY, as indicated by stretches of genic synteny (syntologs) shared with autosomes from other species, may suggest a young evolutionary age. To estimate the timing of dioecy evolution in the Mediterranean Basin clade, relative to total divergence from Eurasian clade lineages, we measured the per-site synonymous substitution rate (d_S) between X-Y gametologs (median = 0.034; teal curve in Fig. 3b), then compared that d_S distribution with genome-wide d_S values for *A. officinalis*–*A. horridus* orthologs (median = 0.084; yellow curve in Fig. 3b), and d_S between the *A. horridus* MSY genes and their orthologs in *A. officinalis* (median = 0.071; purple curve in Fig. 3b). We found that median d_S measurements between the *A. horridus* X-Y gametologs were approximately 47.9% (*p*-value = 1.43e-10) and 40.5% (*p*-value = 4.54e-09) less than the genome-wide and *A. horridus* MSY comparisons with *A. officinalis* orthologs, respectively. Synonymous substitution differences were not significantly different between the two comparisons with *A. officinalis* orthologs (*p*-value = 0.10), which were used as an experimental control. Based on d_S comparisons and the estimated age of the MRCA of *A. horridus* and *A. officinalis* (i.e., ~2.78-3.78 Ma), the *A. horridus* X-Y nonrecombining regions began diverging around 1.33-1.81 Ma or 1.13-1.53 Ma (Fig. 3c), which is less than 1 million years younger than the origin of the Mediterranean Basin crown group (i.e., ~1.9-2.9 Ma) (Bentz, Liu, et al. 2024).

The young *A. horridus* MSY is clearly delimited by a large inversion that spans the length of the nonrecombining region, therefore evolutionary strata (i.e., large stepwise expansions of the nonrecombining region) were not expected nor implicated in a Spearman's correlation test which showed no significant changes in d_S across the MSY (*p*-value = 0.62, r_S = 0.08). We posit that an ancestral, ~2.1 Mb inversion coincided with the origin of dioecy and recombination suppression between proto-XY chromosomes ~1.13-1.81 Ma in the Mediterranean Basin clade, and accumulation of LTR retrotransposons drove the expansion of the MSY to ~9.6 Mb over time. It is, however, possible that a novel sex-determination gene evolved before the inversion took place, which could have been selected for sex-specific benefits, effectively locking sexually antagonistic mutations to

the sex-determining gene. Either way, our results highlight the role of large-scale rearrangements (e.g., inversions) in the cessation of recombination and subsequent accelerated divergence between gametologs (Fig. 3d).

Asparagus officinalis genome and sex chromosomes

Publication of a YY double-haploid genome, for *A. officinalis* in 2017 (Harkess et al. 2017), and an XX double haploid genome in 2020 (Harkess et al. 2020), along with analyses of experimental mutants in both studies, documented a Y-linked two-gene sex-determination system for the species. Recent improvements in genome sequencing and assembly technologies have enabled generation of more accurate and contiguous, diploid-phased reference genomes. The new *A. officinalis* haplotype assemblies and annotations (Figs. S6-S7) were above quality control thresholds, except for the ratio of mono:multi-exon genes, which were similar to those observed in *A. horridus* (~0.40 on average) (Table S1). The *A. officinalis* genome presented here yielded more complete pseudo-chromosome assemblies compared to the previous YY double-haploid (Harkess et al. 2017) (Fig. S8; Table S5). Total gene content in the new *A. officinalis* reference (34,316 to 34,681 genes per haplotype) was higher than predicted for the earlier *A. officinalis* genomes (Harkess et al. 2017; Harkess et al. 2020) and the hermaphroditic species *Asparagus setaceus* (Li et al. 2020); but are closer to updated gene counts based on ChIP-seq data that revealed as many as 4,640 additional protein coding genes missing from the earlier genome annotation (Mendieta et al. 2021).

We inferred the X- and Y-specific regions in the new *A. officinalis* diploid-phased assembly based on structural differences between the X-Y haplotypes (Fig. S9) and the highest density of Y-mer mapping (Fig. 4a). As performed for *A. horridus,* Y-mers were identified for *A. officinalis* by comparing *k*-mers from 3 males and 3 females (Table S3). Using gene trees with X-Y gametologs, we precisely defined PAR boundaries and compared those estimates with previous results (Harkess et al. 2017; Harkess et al. 2020). Comparison of the two *A. officinalis* haplotypes revealed a fully hemizygous region spanning ~1.87 Mb on chromosome 1 of haplotype 2 corresponding to the MSY, and an ~0.13 Mb X-specific region in haplotype 1 (Table S2; Fig. 4b). The Y chromosome assembly was ~145.2 Mb total, whereas the X chromosome was ~144.1 Mb (Table S2). Technological advancements in genome sequencing and assembly explain the different size estimates of the *A. officinalis* sex-limited regions from this study compared to previous work (Harkess et al. 2017; Harkess et al. 2020), as no scaffolding was required to assemble those regions in this study. In the YY double-haploid *A. officinalis* assembly, 6 of the 13 MSY genes were identified on sex-linked contigs that were not anchored to the physical genome map, raising the possibility of misplacement (Harkess et al. 2017). Two of those 6, originally unanchored, gene models were collapsed into a single gene model, and 3 others were

reassigned to the PAR in our haplotype-resolved assemblies (Table S6). Ten total genes (non-TE-associated) were predicted in the updated *A. officinalis* MSY, including male-specific copies of *SOFF* (Fig. 4c) and *aspTDF1* (Fig. 4d). As expected, given the hemizygosity of the MSY, no gametologs were found between the Y- and X-specific regions.

Only one gene (*aspWIP2*) was found in the X-specific region (Fig. 4b). In *Arabidopsis thaliana* (hereafter, "Arabidopsis"), *WIP2* is a zinc finger transcriptional regulator (*C2H2*-type) required for normal pollen tube growth and transport to ovules for fertilization (Crawford et al. 2007). The function of *aspWIP2* in *A. officinalis* has not been tested, but its specificity on the X leaves open the possibility of a dosage advantage in females (two copies) relative to males (one copy) and potential sexually antagonistic function (Harkess et al. 2020). A second gene model, an *outer envelope protein 80* homolog, showed evidence of linkage with both the X- and Y- specific regions in earlier work (Harkess et al. 2017; Harkess et al. 2020), but results from our analyses were inconclusive because none of its exons contained a Y-mer, although separate clades of X- vs. Y-linked orthologs were moderately or poorly supported, respectively (Fig. S10). We placed this *outer envelope protein 80* homolog in PAR2 of both haplotypes (Table S6); however, the germplasm used here differs from the previous studies, so it is possible that the boundary of the nonrecombining MSY varies within the species. In sum, the new genome for *A. officinalis* provides improved assembly of the X-Y nonrecombining regions and sex-limited gene annotations, due its increased contiguity enabled by PacBio HiFi+Omni-C sequencing. Additionally, by applying Y-mer mapping and phylogenetic methods, we found increased resolution of the PAR boundaries in *A. officinalis* (Table S2).

Sex chromosome evolution in the Eurasian clade

Investigation into the evolutionary origin of the *A. officinalis* MSY has been hindered by the hemizygous nature of the X- and Y-limited regions (Harkess et al. 2017; Harkess et al. 2020), leaving inference of the genomic mechanism(s) responsible for the origin of proto-XY recombination suppression unresolved for the Eurasian clade. We leveraged recently published, chromosome-scale, reference genomes representing two additional Asparagaceae subfamilies (Agavoideae and Nolinoideae) (B.-Z. Chen et al. 2024; DOE-JGI; DOE-JGI) to investigate the MSY origin in the *Asparagus* Eurasian clade. Inference of syntologs vs. lineage-specific structural rearrangements (summarized in Fig. 5a) revealed no structural variation associated with the PAR boundaries in *A. officinalis*. However, PAR-linked regions, immediately adjacent to the *A. officinalis* MSY on chromosome 1, exhibited large blocks of syntologs on one autosome (chromosome 5) in *Asparagus* (Asparagoideae), two in *Dracaena* (Nolinoideae) (Fig. 5b), and three in *Yucca* (Agavoideae) (Table S7). One *SOFF* homolog was found on chromosome 5 in *A. officinalis*, but not in a syntenic block. To that end, no syntologs were identified for any of the ten MSY-linked genes from *A.*

officinalis, altogether suggesting that these genes entered the MSY in a stepwise manner following the establishment of a nonrecombining *SOFF* locus on an ancestral proto-Y. Interestingly, we found syntologs of the X-specific *aspWIP2* on chromosome 5 in all analyzed *Asparagus* species (Table S7), thus we hypothesize that the ancestral Y-linked allele was lost sometime following the origin of dioecy in the Eurasian clade. We then tested whether the observed relationship between the *Asparagus* chromosomes 1 and 5 could be traced back to a whole genome duplication (WGD) or a smaller, segmental duplication, and if either were associated with the origin of dioecy in the Eurasian clade. Analysis of d_S indicates that the syntologs observed between chromosomes 1 and 5 in *Asparagus* arose from an ancient WGD shared with other Asparagaceae subfamilies >41 Ma (Asparagoideae-Nolinoideae d_S tests shown in Figs. 5c, S11), well before either origin of dioecy in *Asparagus*. This inference agrees with previous analysis of copy number variation (paralogs vs. orthologs) in de novo transcriptome comparisons of Agavoideae and Asparagoideae taxa (Harkess et al. 2017).

Analysis of the *DUF247* gene family across multiple Asparagaceae taxa revealed no closely related *SOFF* orthologs outside of *Asparagus* (Fig. 5d; Supplemental File10), nor were any identified in a separate analysis with wider sampling (Zhu et al. 2025). Phylogenetic analysis of *SOFF/DUF247* homologs from the hermaphroditic species *Asparagus setaceus* and three Eurasian dioecious species (*A. officinalis, A. kiusianus* and *A. cochinchinensis*) supports the hypothesis that a male-specific *SOFF* arose following a more recent single or tandem gene duplication in the MRCA of the Eurasian dioecy clade and that the *SOFF/DUF247* homolog on chromosome 5 likely represents an older paralog (Fig. 4c). The less well supported placement of *A. cochinchinensis SOFF/DUF247* homologs in Fig. 4c implies an independent set of duplications in the *A. cochinchinensis* lineage, but understanding the timing and nature of those duplications will require genome assemblies for *A. cochinchinensis* and close relatives (see Fig. 1). As seen in earlier work (Zhu et al. 2025), phylogenetic analysis of *DUF247* genes shows many instances of gene family expansions by tandem duplications and variation in copy number across Asparagaceae lineages (Fig. 5d). Rampant copy number variation across *DUF247* homolog clades in Asparagaceae may also explain the absence of a closely related *SOFF* ortholog in *A. horridus*, suggesting that *SOFF* arose from a duplication event in the MRCA of the Eurasian clade. Regardless, phylogenetic analysis indicates that *SOFF* originally evolved following an *Asparagus*-specific *DUF247* gene family expansion (Fig. 5d).

Evolution of two XY sex-determination systems in Asparagus

In this study, we use genomic and evolutionary analysis to test for support for two independent origins of dioecy and sex chromosomes in *Asparagus* (Bentz, Burrows, et al. 2024). We show that each origin of dioecy in the genus involved different ancestral

autosomes: chromosome 1 in the Eurasian clade and chromosome 3 in the Mediterranean Basin clade (Y chromosomes are bolded in Fig. 5a). In *A. horridus*, the nearly 10 Mb MSY is considerably larger than the almost 2 Mb MSY in *A. officinalis*, despite being ~1.7-2 million years younger; which supports the hypothesis that expansion of recombination suppression, as a measure of age and total MSY size, is not correlated in plants (Renner and Müller 2021) although both regions do appear to have expanded over time. Aside from the presence of non-orthologous *AP2*-like genes and a higher ratio of repetitive sequences—compared to the X and autosomes—the only common patterns observed between the *A. officinalis* and *A. horridus* MSYs was the secondary recruitment of LTR retrotransposons and other sequence content (shown as duplications in Figs. S3 and S9) driving their stepwise expansions. Intriguingly, however, both dioecy origins occurred within ~1-2 million years of each other (Fig. 3c), within the same major clade in the genus (Fig. 1 shows the "Asparagus clade" in the genus *Asparagus*), and in association with long-distance dispersals out of southern-central Africa to Eurasia (Norup et al. 2015; Bentz, Burrows, et al. 2024). Considering ancestral biogeography (southern-central Africa) and timeliness (~1.1-3.8 Ma) of dioecy evolution, it is plausible that founder events associated with historical climate oscillations across central-northern Africa helped set the stage for both independent transitions from hermaphroditism to dioecy in the genus (Bentz, Liu, et al. 2024).

The origin of dioecy in the Eurasian clade is marked by the evolution of a male-specific *SOFF* co-opted for sex-determination, which was followed by the stepwise recruitment of additional genes including *aspTDF1* in some lineages (Fig. 1). The ancestral Y-linked *SOFF* may have evolved from a tandem duplication of an autosomal *DUF247* gene, which have since been lost in *A. officinalis,* but may still be present in *A. cochinchinensis* (Fig. 4c). Therefore, a single-gene hypothesis may explain the origin of dioecy in the Eurasian clade, since early diverging lineages (i.e., the *A. cochinchinensis* subclade) exhibit a male-specific copy of *SOFF* (Fig. 4c) but not *aspTDF1* (Fig. 4d) (Harkess et al. 2017; Murase et al. 2017). *SOFF* knockouts in *A. officinalis* males result in functioning bisexual flowers (Harkess et al. 2017), whereas *aspTDF1* knockouts are male-sterile (Harkess et al. 2020), indicating that *SOFF* expression does not impact pollen development in the species. Thus, experimental investigation of *SOFF* function in the *A. cochinchinensis* subclade is necessary to elucidate the ancestral sex-determination mechanisms in the Eurasian group. A single-gene model for the origin of dioecy in the Eurasian clade would require that the ancestral *SOFF* had some function in pollen development which was lost following the co-option of *aspTDF1* into the MSY, in an ancestor of *A. officinalis*. Sexually antagonistic genes are predicted to accumulate in nonrecombining sex-limited regions over time and are thought to lead to sexual

dimorphism (Rice 1984). The stepwise recruitment of *aspTDF1* into the MSY may have been a consequence of sexually antagonistic selection (i.e., removal of *aspTDF1* from the autosomes ensures that females can no longer produce pollen) in the MRCA of *A. officinalis* and *A. filicinus* or *A. verticillatus* (see Fig. 1).

The genome data presented here, together with phylogeographic analysis of the origin of dioecy in *Asparagus* (Bentz, Burrows, et al. 2024) and functional work on other dioecious plant species (Marais et al. 2025), indicate that there are many potential molecular mechanisms for the shift from hermaphroditism to dioecy in flowering plants. Continued work on dioecious lineages of *Asparagus* offers opportunities for improved understanding of the ecological drivers of the origin and persistence of dioecy. For instance, two of the eight independent range expansions out of southern Africa were associated with dioecy origins in *Asparagus* (Bentz, Burrows, et al. 2024) suggesting that inbreeding avoidance associated with dispersal and founder events may have promoted transitions to dioecy, but adaptive specialization on male or female function may have contributed to its maintenance across time. In any case, integrated phylogenetic, genomic, and functional investigations of dioecy in model taxa such as *Asparagus* will continue to yield deeper understanding of the origins and evolution of separate sexes across the tree of life.

Candidate genes for sex-determination and sexual antagonism in A. horridus

Testing for candidate sex-determination genes requires functional validation (e.g., genetic knockouts) and is outside the scope of this study. However, comparisons of X-Y gene content, expression, and molecular evolutionary analysis can yield focused lists of gene candidates for sex-determination or other sex-specific phenotypes. Here, we investigate MSY-linked genes from *A. horridus* that may be promoting male fertility and/or repressing female fertility in XY males, since diverging from their X-linked gametologs ~1-2 Ma. First, we assess sex-biased floral transcription profiles for 5 males and 5 females of *A. horridus* genotypes from a wild population in Capo Rama, Terrasini, Palermo, Sicily, Italy (38°08'23.6"N, 13°03'34.6"E; https://wwfcaporama.it/en/). Secondly, we test for shifts in the ratio of nonsynonymous to synonymous substitution rates (d_N/d_S) in MSY-linked genes relative to X-linked gametologs and autosomal orthologs from other *Asparagus* species. Of the MSY-linked genes in *A. horridus,* we found that 20 were significantly up-regulated and 7 were down-regulated in male flowers compared to female flowers across 3 combined developmental stages (red data points in Fig. 3e; see Table S8 for tissue sampling details and Table S9 for expression analysis results). Six of those 27 genes also showed shifts to positive selection ($d_N/d_S > 1.0$) in the MSY-linked ortholog in gene trees with X-gametologs and orthologs from other *Asparagus* species (bolded in Table S10; see Fig. S12 for phylograms).

Two Y-linked pectin methylesterase inhibitor (*PMEI*) genes exhibited interesting patterns linked to sex: the first (*PMEI* a) showed evidence of elevated d_S compared to its X-linked counterpart (d_S = 0.11; Fig. 3d), positive selection (d_N/d_S = 68.3; Table S10), and significantly lower expression in male flowers compared to female flowers (log2 FC -3.3; Table S10); while the second gene (*PMEI* b) was highly up-regulated (log2 FC +5.6; Table S9; Fig. 3e) with no signatures of positive selection (d_N/d_S = 1; Table S11). Pectin methylesterase (PME) enzymes are post-transcriptionally regulated by PMEI proteins (Xu et al. 2022) and PME/PMEI activity plays important roles in many growth processes, including pollen tube development (Tian et al. 2006) and fruit ripening regulation (Louvet et al. 2006) in Arabidopsis and the dioecious kiwifruit (*Actinidia deliciosa*) (Irifune et al. 2004). *PMEIs* were also shown to regulate pollen tube growth in pear (*Pyrus bretschneideri*) (Zhu et al. 2021). Interestingly, *PMEIs* have also been found in the MSY of a dioecious night shade (*Solanum appendiculatum*) (Wu et al. 2021), thus potentially representing a gene family that commonly neofunctionalizes in male heterogametic (XY) systems, in association with loss of selection pressure to maintain function in fruit development and increased pressure for optimization of male-specific functions. We also found evidence of positive selection and male-biased expression (log2 FC +8.1) of a chalcone synthase (*CHS*) gene (Fig. 3e; Table S10). *CHS* genes are involved in pollen development and show sex-biased expression in several other dioecious systems (Liao et al. 2020; Huang et al. 2022; Keefover-Ring et al. 2022). For example, chemically induced male sterility experiments in wheat (*Triticum aestivum*) revealed that *CHS* expression decreased in male-sterile plants (Ba et al. 2017), suggesting a functional role in pollen development and/or viability. The remaining MSY genes with evidence of both positive selection and sex-biased expression include a fructose-bisphosphate aldolase (transcriptional activator of glycolytic enzymes) encoding gene (d_S = 0.27 compared to X-gametolog; Fig. 3d), remorin C-terminal domain (plasma membrane-associated protein), metallophos domain (*DUF4073*), and an unannotated (i.e., no blast hits to genes with known functions) gene model (Table S10).

Five other MSY genes exhibited signatures of positive selection, but lacked sex-biased expression patterns, including a *DUF1295* gene encoding for 3-oxo-5-alpha-steroid 4-dehydrogenase (d_S = 0.23 compared to X-gametolog; Fig. 3d); an APETALA2/ethylene-responsive element-binding factor domain (*AP2*/*ERF*); a *SLOW WALKER2* (*SWA2*) homolog, and another gene model without a functional annotation (Table S10). The Y-linked *AP2*/*ERF* and *SWA2*-like genes are evolving under strong positive selection in *A. horridus* (d_N/d_S = 121.2 and 110.7; Table S10) and are especially interesting due to their possible functions in other flowering plants. In Arabidopsis, the *AP2*/*ERF* gene family includes *AP2,* an A-class

homeotic gene in the ABC flower development model (Kunst et al. 1989; Bowman and Meyerowitz 1991; Jofuku et al. 1994). Down-regulation of an *AP2* ortholog in rice (*Oryza sativa*) leads to reduced stamens, fused anthers, additional pistils, lower seed efficacy, and decreased pollen viability and germination altogether suggesting a major role in male fertility (Zhao et al. 2006). The *A. horridus* Y-linked *AP2*/*ERF* homolog is distantly related to the Arabidopsis *AP2* gene model At4g36920 (~37% identity in amino acid alignment). Another distantly related *AP2*/*ERF* gene was also found in the *A. officinalis* MSY, but it was not specifically implicated in sex-determination for the species (Harkess et al. 2017) and is a paralog of the *A. horridus* MSY *AP2*/*ERF* gene. *SWA2* is required for coordinated cell cycle progression during female gametophyte and pollen development in Arabidopsis (Pagnussat et al. 2005; N. Li et al. 2009). Mutant *swa2* genotypes in Arabidopsis exhibit arrested female gametophyte development, with asynchronous embryo sac development, in which ovules within the same pistil abort at different stages (Pagnussat et al. 2005; N. Li et al. 2009). Pollen cell cycles were also disrupted in Arabidopsis *swa2* mutants—though to a lesser extent compared to the impaired development of female gametophytes—leading to defective pollen development in a small percentage of mutants (N. Li et al. 2009). Moreover, differential expression analysis among sterile vs. fertile ovules in Chinese pine (*Pinus tabuliformis*) revealed significantly lower expression of a *SWA2* homolog in sterile ovules compared to the latter, suggesting a conserved functional role in ovule development (Guo et al. 2014). *AspSWA2* was not differentially expressed between male and female flowers sampled from a wild population of *A. horridus* but we did not assay expression in young flower buds. Considering the broadly conserved role of *SWA2* in female function, the apparent late abortion of pistil development in *A. horridus* male flowers (Fig. 2 shows the typical *A. horridus* staminate flower phenotype with vestigial pistils), and a significantly elevated d_N/d_S ratio, the Y-linked *aspSWA2* gametolog may be evolving functions related to sex-determination in *A. horridus*.

Materials and Methods

Biological materials

To analyze both the X and Y sex chromosomes, male plants from both species were selected for genome assembly. Fresh cladodes were collected and flash frozen with liquid nitrogen for all DNA-seq experiments. We sampled several tissue types, in triplicates, at different developmental stages (Table S8), for transcriptome sequencing from the two genome-lines *A. officinalis* (accession pb81m) and *A. horridus* (accession pb32m), which we used for structural annotation predictions. We also sampled male and female flowers across different developmental stages (five replicates each), from a wild population of *A.*

horridus identified at Capo Rama reserve WWF - Terrasini (PA) Italy for differential expression analysis between the sexes (see Table S8 for sampling stages). Tissue for transcriptome sequencing was flash frozen with liquid nitrogen immediately after sampling and all tissue was sampled at the same time, if used together in an analysis. Details about each biological from this study can be found in Table S3.

DNA and RNA sequence data generation

PacBio HiFi, Omni-C, and Illumina (PE150) libraries were prepared and sequenced at HudsonAlpha Genome Sequencing Center (Huntsville, Alabama, USA) using SMRTbell® Prep Kit v2.0 (Pacific Biosciences, Menlo Park, California, USA), Dovetail Genomics Omni-C® Kit (Cantata Bio, Scotts Valley, California, USA), and NEBNext Ultra II DNA PCR-free Library Prep Kit (New England Biolabs Inc., Ipswich, Massachusetts, USA), respectively. PacBio HiFi long-reads were sequenced on the SEQUEL II platform, while Omni-C and all other short-read sequencing experiments were performed on the Illumina (San Diego, California, USA) NovaSeq 6000. High molecular weight DNA extraction was performed using the Takara NucleoBond® HMW DNA kit (Takara Bio USA, Inc., San Jose, California, USA) prior to PacBio HiFi library preparation. The DNeasy Plant Mini kit (Qiagen, Hilden, Germany) was used for DNA isolation prior to Illumina library preparation. For the *A. horridus* and *A. officinalis* genome-lines, total RNA was extracted from the various tissue types using RNeasy Plant Mini Kit (Qiagen) and libraries were prepared using Illumina TruSeq Stranded mRNA Library Prep Kit. Sampled RNA replicates, for each genome-line, were also pooled for sequencing full-length cDNA (PacBio Iso-Seq) on the SEQUEL II. Additional RNA-seq datasets were generated for male and female plants of *A. horridus* from the wild Italian population as follows: 1) total RNA extraction with RNeasy Plant Mini Kit; 2) shipped from Italy to the U.S. on GenTegraRNA columns (GenTegra, Pleasanton, California, USA) to ensure RNA stability; 3) mRNA libraries prepared by Novogene Corporation Inc. (Sacramento, California, USA) using in-house protocols; 4) sequencing on the Illumina NovaSeq X-Plus (10B, PE150) platform.

Genome assembly

We used HIFIAsm+HiC v0.16.1 (Cheng et al. 2021) to build initial contigs and YaHS v1.1 (Zhou et al. 2023) to scaffold contigs into chromosome-scale, diploid-phased assemblies. Prior to scaffolding, we used BWA-MEM v0.7.17 with the flag *-5SP* (Li 2013), SAMBLASTER v0.1.24 (Faust and Hall 2014), and SAMtools v1.16.1 (H. Li et al. 2009) to map Omni-C reads to contigs, mark duplicate alignments, and remove duplicates, respectively. HIFIAsm contigs <50,000 nt were removed prior to scaffolding. The final haplotype assemblies were curated using the JUICER v1.6 (Durand, Shamim, et al. 2016) and Juicebox v1.11.08 (Durand, Robinson, et al. 2016) pipelines and ordered and oriented

chromosomes to match the Aspof.V1 YY haploid assembly (Harkess et al. 2017). Final assembly completeness was assessed based on presence of >90% of conserved Viridiplantae BUSCO orthologs, using BUSCO v6.0.0 (viridiplantae_odb12) (Manni et al. 2021). Separate and combined haplotype assembly completeness was assessed with Merqury v1.3 and Meryl v1.4.1 (Rhie et al. 2020) which were used to count all unique 21-mers found in the HiFi reads, and then calculate the percentage of the *k*-mers found in the reads that were also observed in the final assemblies. *K*-mer completeness should be lower when analyzed separately for each haplotype, based on heterozygosity levels, but near 100% of the *k*-mers should be present in concatenated haplotypes (i.e., analyzed together).

Genome structural annotation

We annotated repetitive elements and protein coding genes using repeat-soft-masked haplotype assemblies after generating repeat libraries de novo for both species using RepeatModeler v2.0.2 (Flynn et al. 2020) and soft-masking with RepeatMasker v4.1.2 and the options *-cutoff 250* and *-nolow*. Curated repeats from the Repbase database for monocots (Bao et al. 2015) were combined with our de novo library prior for the RepeatMasker analysis. We used the long-read protocol (see long_read_protocol.md, github.com/Gaius-Augustus/BRAKER/) for BRAKER v3.0.3 (Hoff et al. 2016; Brůna et al. 2021; Gabriel et al. 2023) and its many dependencies (Lomsadze et al. 2005; Stanke et al. 2006; Gotoh 2008; Stanke et al. 2008; Barnett et al. 2011; Iwata and Gotoh 2012; Buchfink et al. 2015; Tang et al. 2015; Hoff et al. 2019; Brůna et al. 2020) for gene prediction based on extrinsic evidence from short- and long-read transcriptome sequencing of various tissue samples and published protein sequences from *Asparagus officinalis* (Harkess et al. 2017), *Asparagus setaceus* (Li et al. 2020), and Viridiplantae (OrthoDB v11) (Kuznetsov et al. 2023). Illumina transcriptome reads were aligned to soft-masked haplotypes with STAR v2.7.10 (Dobin et al. 2013). Full-length (non-concatemer) consensus Iso-Seq reads were mapped to soft-masked haplotypes using pbmm2 v1.3.0 (Li 2018), then isoforms were collapsed in the mapped transcripts. Gene predictions were parsed and filtered using AGAT v1.1.0 (Dainat) and GffRead v0.12.7 (Pertea and Pertea 2020), discarding genes with 1) in-frame stop codons (or adjusting the CDS phase when possible), 2) single-exon transcripts when absent on the opposite strand, 3) missing start codons, or 4) total CDS <300-nt. EnTAP v1.0.0 (Hart et al. 2020) was used to further assess gene prediction accuracy via reciprocal functional annotation based on the UniProt/Swiss-Prot (Bairoch et al. 2005) and EggNOG v5.0 (Huerta-Cepas et al. 2019) databases, aiming for >70% of genes with functional annotations. Mono:multi-exonic gene ratios were also used for gene prediction quality control assessment, for which a mono:multi ratio of ~0.20 has been suggested as ideal based on model plant genomes (Vuruputoor et al. 2023). TE predictions

were performed with EDTA v2.2.2 (flags: *--sensitive 1 --anno 1 --evaluate 1*) (Ou et al. 2019) using the RepeatModeler library of classified elements, BRAKER gene predictions, and the 'out' file from RepeatMasker. MSY gene predictions were further curated by removing models that were assigned TE-associated annotations or >90% soft-masked. Completeness of gene predictions were assessed using BUSCO v6.0.0 (viridiplantae_odb12 proteins) aiming for >90% complete BUSCOs per haplotype.

Delimitation of sex chromosomes

To identify X/Y haplotypes, we mapped male-specific *k*-mers (Y-mers), identified in Illumina datasets by comparing males vs. females of *A. officinalis* (3 males + 3 females) and *A. horridus* (8 males + 7 females, see Table S3 for additional sampling details), to each haplotype for both species. Unique 21-bp *k*-mers were counted in the Illumina reads using JELLYFISH v2.3.0 (Marcais and Kingsford 2012) and were filtered by removing *k*-mers present at low (<10) or high (>250) frequencies. Y-mers were subset by selecting for *k*-mers conserved across all males and absent in all females for each species (Carey et al. 2024). We used BWA-MEM (flags: *-k 21 -T 21 -c 10 -a*) to map the Y-mers, then delimited each Y chromosome and MSY according to scaffolds and regions with the highest density of Y-mer alignment coverage, respectively (Carey et al. 2024). In a separate analysis, *k*-mers from *A. acutifolius* (1 male + 1 female) were processed with those from *A. horridus* to test for shared Y-mers and a conserved MSY across the two species. Normalized Illumina reads from all *A. horridus* samples (Table S3) were mapped to both haplotypes for the species, to perform coverage comparisons between the sexes for MSY delimitation (Palmer et al. 2019), using BWA-MEM, requiring a 35 bp minimum seed length, <10 multi-mapped alignments, and >30 mapping quality. BBMap v38.93 *reformat.sh* (Bushnell 2018) was used to normalize read depth by random down-subsampling to ~30x. Using rough MSY coordinates based on Y-mer mapping density, we assessed gene tree topologies for the relative placement of X-Y gametologs/orthologs for genes near putative PAR boundaries. MSYs and adjacent PARs (PAR1 and PAR2) were defined according to the first and last Y-specific gene/allele, as indicated by strongly supported clades of either Y- or X-linked orthologs (e.g., genes in a clade of only Y-linked orthologs were assigned to the MSY). Orthologs/gametologs were identified using OrthoFinder v2.5.5 (Emms and Kelly 2019) and GENESPACE v1.3.1 (Lovell et al. 2022) with both new haplotypes from *Asparagus officinalis* and *Asparagus horridus,* as well as other monocot relatives: *Asparagus setaceus* (Li et al. 2020)*, Asparagus kiusianus* (Shirasawa et al. 2022)*, Dracaena cambodiana* (B.-Z. Chen et al. 2024), *Yucca aloifolia* (DOE-JGI), and *Ananas comosus* (pineapple) (Ming et al. 2015). Chromosome labels from the *A. setaceus* genome were renamed here to match those from this and previous studies (Harkess et al. 2017; Harkess et al. 2020). GENESPACE results were also used to identify structural variants among

syntenic, orthologous gene blocks. Haplotype-haplotype alignments were generated to test for structural variation between XY chromosomes within each species, using minimap2 v2.26 (Li 2018) and SyRI v1.6.3 (Goel et al. 2019), respectively. Structural variants were plotted with plotsr v1.1.0 (Goel and Schneeberger 2022).

Sex chromosome evolution in the Mediterranean Basin clade

We performed pairwise d_S comparisons between the MSY+flanking Y-PAR genes vs. X-gametologs in *A. horridus*. First, we tested for a linear correlation between X chromosome position and d_S to determine whether stepwise MSY expansion events caused by large-scale structural changes (evolutionary strata) could be detected via significant changes in d_S across the nonrecombining region. Gene positions on the X chromosome are typically more preserved over time compared to the MSY, due to no meiotic crossing over in the latter; therefore, a Spearman's rank correlation test was performed comparing X-gametolog positions vs. d_S, measured between MSY genes and X-gametologs, using coefficient (r_S) and a p-value <0.05 significance threshold, calculated with base R (v4.2.2) (RC-Team 2020).

We then used d_S measurements between genome-wide orthologs from *A. officinalis* and *A. horridus*, to estimate total species divergence and performed three separate Wilcoxon signed-rank tests because the data were not normally distributed. Significant differences (p-value <0.5) in d_S were explored for the following comparisons among single copy homologs: 1) *A. horridus* MSY genes vs. X-gametologs (N=47); 2) *A. horridus* MSY genes vs. *A. officinalis* orthologs (N=41); and 3) *A. horridus* vs. *A. officinalis* genome-wide orthologs (N=12,646). All d_S estimates were calculated with KaKs_Calculator 3.0 (Zhang 2022) from protein alignments converted to nucleotide codon alignments with MAFFT v7.505 (Katoh and Standley 2013) and pal2nal v14 (Suyama et al. 2006), respectively.

To estimate the absolute timing of sex chromosome origins in the Mediterranean Basin clade, we used previous divergence time estimates for the MRCA of *A. horridus* and *A. officinalis* and multiplied the 95% confidence intervals by the percentage of test 1 to tests 2 and 3 from above. We then tested for signs of positive selection among MSY genes in *A. horridus* by calculating d_N/d_S ratios with PAML v4.8 (Yang 2007) CODEML branch-sites model (M2a) and two nulls: the M1a sites model for nearly neutral selection; and the M2a branch-sites null which fixes omega (d_N/d_S) to 1 (neutral) for the foreground branch and is more stringent than the former (Álvarez-Carretero et al. 2023). Unrooted ML gene trees were estimated for CODEML consisting of *Asparagus* orthologs identified by OrthoFinder. Gene trees were inferred using IQ-TREE v1.6.12 (Nguyen et al. 2015) with 1000 ultrafast bootstraps (UFBoot) and the best fit substitution model (Kalyaanamoorthy et al. 2017). Trees with only three tips were reconstructed based on species relationships (Bentz,

Burrows, et al. 2024). M1a sites model results were assessed using a likelihood ratio test (LRT) and chi-squared distribution to compute *p*-values (*P*) and a sequential Bonferroni type procedure to control the false discovery rate by computing the expected rate of false rejection (*Q*), requiring *P*<*Q* (Benjamini and Hochberg 1995). To compare the nested M2a models, LRT and chi-squared critical value thresholds were applied based on 1 degree of freedom (i.e., 3.84 LRT = *P* of 0.05; 6.63 LRT = *P* of 0.01) to compute relative *P* (<0.05 cut-off).

Male vs. female floral expression of X-Y gametologs in A. horridus

Sex-specific gene expression patterns were compared between flowers from male and female genotypes sampled at different developmental stages from a single Italian population of *A. horridus*. RNA extractions with sufficient yield were sequenced (Table S8), then transcriptomes were assembled *de novo* with StringTie v2.2.1 (Kovaka et al. 2019), using *A. horridus* haplotype 1 transcript alignments generated with STAR. Differential expression was measured among single copy gametologs, shared between the MSY and X-specific region (N=47), using DESeq2 (Love et al. 2014) with the StringTie gene count matrix. Expression profile similarities across flower sampling treatments were assessed based on a PCA (Fig. S13) and clustered heatmap (Fig. S14). We tested whether sex (male vs. female) explained significant differential expression patterns among the floral sampling treatments, rather than comparing individual developmental stages, because expression profiles from each treatment were too overlapping among the successful RNA-seq libraries. Significantly different expression profiles were assessed based on adjusted *p*-value (>0.05) and log2 fold change (log2 FC) indicating significant up-regulation (log2 FC >1) or down-regulation (log2 FC <-1).

Sex chromosome evolution in the Eurasian clade

Along with the genomes presented here, we leveraged published genomes for the Asparagaceae taxa *Yucca aloifolia* (Agavoideae) (DOE-JGI), *Dracaena cambodiana* (Nolinoideae) (B.-Z. Chen et al. 2024), and *Asparagus setaceus* (Asparagoideae) (Li et al. 2020) to investigate the origin of the recombination suppression and the MSY in the Eurasian clade. First, we used GENESPACE and OrthoFinder results to infer homologous, syntenic gene blocks among *A. officinalis, A. horridus, A. setaceus, Dracaena,* and *Yucca*—with a focus on the MSY+flanking Y-PARs near the left end of chromosome 1 in *A. officinalis*. Then, we measured d_S among syntologs using wgd v2 (H. Chen et al. 2024) and compared results from two different treatments: 1) between paralogs from the *A. officinalis* chromosomes 1 and 5 to estimate the relative timing for the duplication event in question; 2) between genome-wide orthologs from *Asparagus* and *Dracaena* to estimate the relative timing of species divergence. Homologs/paralogs from the *A. officinalis*

chromosomes 1 and 5 included only those identified from the first 400 gene models on chromosome 1 of haplotype 1 (i.e., selecting genes that span the PAR1-MSY-PAR2 borders) that exhibited syntologs on chromosome 5 (86/400 genes fit these criteria).

SOFF and *aspTDF1* multiple sequence alignments were inferred using MAFFT v7.490 (flags: *--maxiterate 1000 --localpair*), then trimmed to remove poorly aligned regions with trimAl and the flag *-automated1*. Exons and introns were included in gene alignments. Gene trees were inferred with IQ-TREE v1.6.12, using 1000 ultrafast bootstraps, and the best fit substitution model as before. Homologs were identified in published assemblies for *Asparagus setaceus* (Li et al. 2020), a male and female of *Asparagus kiusianus* (Shirasawa et al. 2022), a female of *Asparagus officinalis* (Harkess et al. 2020), and genes from the outgroup species *Agave tequilana* (DOE-JGI) were used for rooting each tree. The *A. kiusianus* assemblies lacked *SOFF* annotations, so we used a local BLAST search to identify homolog sequences. In the *A. setaceus* annotations, a *SOFF* homolog on chromosome 1 was split into two gene models that we concatenated according to BLAST alignments. We used BLAST to test for and extract *SOFF* and *aspTDF1* gene sequences from de novo short-read assemblies, generated by SPAdes v3.15.5 (Bankevich et al. 2012) for a male and female of *A. cochinchinensis* (Harkess et al. 2017) and a female of *A. horridus* (F1, see Table S3). The *SOFF* and *aspTDF1* trees were rooted with the *Agave tequilana* gene models *AgateH1.23G025400.1.v2.1* and *Agave_AgateH1.26G044700.1.v2.1*, respectively. The phylogenetic placement of the ancestral *DUF247* duplication, that preceded the origin of *SOFF,* is unknown, thus we chose the *SOFF* gene tree root (marked by a black square in Fig. 5d) based on a larger phylogenetic analysis of the *DUF247* gene family (see Supplemental File10) across *Asparagus, Agave, Dracaena*, and rice clade references from Zhu et al. (2025).

For the wider *DUF247* analysis, all *DUF247* gene predictions were selected for phylogenetic analysis, based on inference of gene functions by EnTAP (performed as described above), from all four *Asparagus* haplotypes from this study and the cited *Asparagus setaceus, Agave tequilana,* and *Dracaena cambodiana* assemblies. Prior to running EnTAP, genome annotations from other studies were re-filtered using AGAT and GffRead, as executed previously in this study. The *DUF247* gene family tree was inferred as before but using amino acid alignments instead of nucleotides due to increased sequence diversity across the *DUF247* family. *Asparagus kiusianus* genes were not included in the broader *DUF247* analysis because the published genomes were missing protein predictions and exon-intron boundaries for those genes.

Data Availability Statement

Genome assemblies, comprehensive structural annotations and sequence files, and other analysis result files are available for download at https://doi.org/10.5281/zenodo.10802232. Curated Y-mer and repeat libraries can be found on GitHub along with *SOFF* and *aspTDF1* matrices: https://github.com/bentzpc/Asparagus_genomes. Sequencing reads and assemblies are available under the NCBI BioProject PRJNA1320942 (https://www.ncbi.nlm.nih.gov/bioproject/1320942).

Acknowledgements

This work was supported by United States National Science Foundation (NSF) Division of Environmental Biology (DEB) no. 2110875 (J.L.-M.); NSF IOS-PGRP CAREER no. 2239530 (A.H.); NSF IOS-EDGE no. 2335775 (A.H.); University of Georgia, Plant Center Doctoral Dissertation Improvement Grant (P.C.B.); Botanical Society of America, Bill Dahl Graduate Student Research Award (P.C.B.); Society for the Study of Evolution, R.C. Lewontin Early Award (P.C.B.). We thank Laura Genco and Davide Bonaviri (WWF Italy, Capo Rama reserve), Tony Avent (Juniper Level Botanical Garden, Raleigh, NC, USA), and Mason McNair (Michigan State University, East Lansing, MI, USA) for their support with *Asparagus horridus* sampling; as well as Zach Stansell (USDA-ARS, Geneva, NY, USA) for providing seed collections for *Asparagus officinalis* from US Department of Agriculture's Germplasm Resources Information Network. We thank the University of Georgia and HudsonAlpha for computational, laboratory, and greenhouse support. We also thank two anonymous reviewers for providing insightful feedback on the manuscript. Lastly, we are grateful to the late plantsman Alan Galloway for collecting wild *A. horridus* (pb32m) germplasm from Majorca, Spain in 2018, which led to the first reference genome for the species as presented here.

Author Contributions

P.C.B. wrote the manuscript and performed computational analysis. P.C.B., S.B.C., F.M., F.S., A.H., and J.L.-M. conceived the study and performed early analysis. P.C.B., F.M., V.R., and H.H. conducted field and/or laboratory experiments. P.C.B., F.M., and F.S. collected and curated biological samples. All authors reviewed manuscript drafts.

References

Akagi T, Varkonyi-Gasic E, Shirasawa K, Catanach A, Henry IM, Mertten D, Datson P, Masuda K, Fujita N, Kuwada E, et al. 2023. Recurrent neo-sex chromosome evolution in kiwifruit. *Nat. Plants* 9:393–402.

Álvarez-Carretero S, Kapli P, Yang Z. 2023. Beginner's guide on the use of PAML to detect positive selection. *Mol. Biol. Evol.* 40:msad041.

Ba Q, Zhang Gaisheng, Li G, Lijuan Zhou, Zhang Gensheng, Fu Z, Chen C, Song Y. 2017. Analysis of chalcone synthase gene in chemically induced male sterility (CIMS) in wheat (*Triticum aestivum* L.). *Indian J. Genet. Plant Breed.* 77:508–512.

Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, et al. 2005. The universal protein resource (UniProt). *Nucleic Acids Res.* 33:D154–D159.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* 19:455–477.

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:1–6.

Barnett DW, Garrison EK, Quinlan AR, Strömberg MP, Marth GT. 2011. BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* 27:1691–1692.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57:289–300.

Bentz PC, Burrows JE, Burrows SM, Mizrachi E, Liu Z, Yang J, Mao Z, Popecki M, Seberg O, Petersen G, et al. 2024. Bursts of rapid diversification, dispersals out of southern Africa, and two origins of dioecy punctuate the evolution of *Asparagus*. *Genome Biol. Evol.* 16:evae200.

Bentz PC, Liu Z, Yang J-B, Zhang L, Burrows S, Burrows J, Kanno A, Mao Z, Leebens-Mack J. 2024. Young evolutionary origins of dioecy in the genus *Asparagus*. *Am. J. Bot.* 111:e16276.

Bowman JL, Meyerowitz EM. 1991. Genetic control of pattern formation during flower development in Arabidopsis. In: Symposia of the Society for Experimental Biology. Vol. 45. p. 89–115.

Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics Bioinforma.* 3:lqaa108.

Brůna T, Lomsadze A, Borodovsky M. 2020. GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. *NAR Genomics Bioinforma*. 2:lqaa026.

Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12:59–60.

Bushnell B. 2018. BBTools: a suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. *BBMap* [Internet]. Available from: https://sourceforge.net/projects/bbmap/

Carey S, Yu Q, Harkess A. 2021. The diversity of plant sex chromosomes highlighted through advances in genome sequencing. *Genes* 12:381.

Carey SB, Aközbek L, Lovell JT, Jenkins J, Healey AL, Shu S, Grabowski P, Yocca A, Stewart A, Jones T, et al. 2024. ZW sex chromosome structure in *Amborella trichopoda*. *Nat. Plants*:1–11.

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

Charlesworth D, Harkess A. 2024. Why should we study plant sex chromosomes? *Plant Cell*:koad278.

Chen B-Z, Li D-W, Wang W-J, Xin Y-X, Wang W-B, Li X-Z, Hao T-T, Dong Y, Yu W-B. 2024. Chromosome-level and haplotype-resolved genome assembly of *Dracaena cambodiana* (Asparagaceae). *Sci. Data* 11:873.

Chen H, Zwaenepoel A, Van de Peer Y. 2024. wgd v2: a suite of tools to uncover and date ancient polyploidy and whole-genome duplication. *Bioinformatics* 40:btae272.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18:170–175.

Crawford BC, Ditta G, Yanofsky MF. 2007. The NTT gene is required for transmitting-tract development in carpels of *Arabidopsis thaliana*. *Curr. Biol.* 17:1101–1108.

Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format (Version 1.1.0). *Zenodo* [Internet]. Available from: https://www.doi.org/10.5281/zenodo.3552717

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21.

DOE-JGI. *Yucca aloifolia* Ya24Inoko v2.1. Available from:

https://phytozome.jgi.doe.gov/info/YaloifoliaYa24Inoko_v2_1

DOE-JGI. *Agave tequilana* var. Weber's Blue v2.1. Available from: https://phytozome-next.jgi.doe.gov/info/Atequilanavar_WebersBlueHAP1_v2_1

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3:99–101.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3:95–98.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20:1–14.

Faust GG, Hall IM. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* 30:2503–2505.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* 117:9451–9457.

Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. 2023. BRAKER3: Fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. *bioRxiv*.

Goel M, Schneeberger K. 2022. plotsr: visualizing structural similarities and rearrangements between multiple genomes. *Bioinformatics* 38:2922–2926.

Goel M, Sun H, Jiao W-B, Schneeberger K. 2019. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 20:1–13.

Gotoh O. 2008. A space-efficient and accurate method for mapping and aligning cDNA sequences onto genomic sequence. *Nucleic Acids Res.* 36:2630–2638.

Guo A, Zheng CX, Yang YY. 2014. Differential expression of *SLOW WALKER2* homologue in ovules of female sterile mutant and fertile clone of *Pinus tabulaeformis*. *Russ. J. Dev. Biol.* 45:78–84.

Harkess A, Huang K, van der Hulst R, Tissen B, Caplan J, Koppula A, Batish M, Meyers BC, Leebens-Mack J. 2020. Sex determination by two Y-linked genes in garden asparagus. *Plant Cell* 32:1790–1796.

Harkess A, Zhou J, Xu C, Bowers JE, van der Hulst R, Ayyampalayam S, Mercati F, Riccardi P, McKain MR, Kakrana A, et al. 2017. The asparagus genome sheds light on the origin

and evolution of a young Y chromosome. *Nat. Commun*. 8:1279.

Hart AJ, Ginzburg S, Xu M, Fisher CR, Rahmatpour N, Mitton JB, Paul R, Wegrzyn JL. 2020. EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Mol. Ecol. Resour.* 20:591–604.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* 32:767–769.

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. *Gene Predict. Methods Protoc.*:65–95.

Huang H, Wang H, Hu X, Zhang Z-Q. 2022. Identification of candidate genes associated with sex differentiation and determination of gender diphasic plant *Lilium apertum* (Liliaceae). *Sci. Hortic.* 306:111431.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* 47:D309–D314.

Irifune K, Nishida T, Egawa H, Nagatani A. 2004. Pectin methylesterase inhibitor cDNA from kiwi fruit. *Plant Cell Rep.* 23:333–338.

Iwata H, Gotoh O. 2012. Benchmarking spliced alignment programs including Spaln2, an extended version of Spaln that incorporates additional species-specific features. *Nucleic Acids Res.* 40:e161–e161.

Jofuku KD, Den Boer B, Van Montagu M, Okamuro JK. 1994. Control of Arabidopsis flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6:1211–1225.

Kalyaanamoorthy S, Minh BQ, Wong TK, von Haeseler A, Jermiin LS. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14:587–589.

Katoh K, Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30:772–780.

Keefover-Ring K, Carlson CH, Hyden B, Azeem M, Smart LB. 2022. Genetic mapping of sexually dimorphic volatile and non-volatile floral secondary chemistry of a dioecious willow. *J. Exp. Bot.* 73:6352–6366.

Kovaka S, Zimin AV, Pertea GM, Razaghi R, Salzberg SL, Pertea M. 2019. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20:1–13.

Kunst L, Klenz JE, Martinez-Zapater J, Haughn GW. 1989. *AP2* gene determines the identity of perianth organs in flowers of *Arabidopsis thaliana*. *Plant Cell* 1:1195–1208.

Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. 2023. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. *Nucleic Acids Res.* 51:D445–D451.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Prepr. ArXiv13033997*.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079.

Li N, Yuan L, Liu N, Shi D, Li X, Tang Z, Liu J, Sundaresan V, Yang W-C. 2009. *SLOW WALKER2*, a *NOC1/MAK21* homologue, is essential for coordinated cell cycle progression during female gametophyte development in Arabidopsis. *Plant Physiol.* 151:1486–1497.

Li S-F, Wang J, Dong R, Zhu H-W, Lan L-N, Zhang Y-L, Li N, Deng C-L, Gao W-J. 2020. Chromosome-level genome assembly, annotation and evolutionary analysis of the ornamental plant *Asparagus setaceus*. *Hortic. Res.* 7:1–11.

Liao Q, Du R, Gou J, Guo L, Shen H, Liu H, Nguyen JK, Ming R, Yin T, Huang S, et al. 2020. The genomic architecture of the sex-determining region and sex-related metabolic variation in Ginkgo biloba. *Plant J.* 104:1399–1409.

Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 33:6494–6506.

Louvet R, Cavel E, Gutierrez L, Guénin S, Roger D, Gillet F, Guerineau F, Pelloux J. 2006. Comprehensive expression profiling of the pectin methylesterase gene family during silique development in *Arabidopsis thaliana*. *Planta* 224:782–791.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15:1–21.

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *Elife* 11:e78526.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and

streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* 38:4647–4654.

Marais GA, Branco C, Rocheta M, Dufay M, Tonnabel J. 2025. Plant sex-determining genes and the genetics of the evolution towards dioecy. *J. Exp. Bot.*:eraf224.

Marcais G, Kingsford C. 2012. Jellyfish: A fast k-mer counter. *Tutorialis E Manuais* 1:1–8.

Mendieta JP, Marand AP, Ricci WA, Zhang X, Schmitz RJ. 2021. Leveraging histone modifications to improve genome annotations. *G3* 11:jkab263.

Ming R, VanBuren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang M-L, Chen J, Biggers E, et al. 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* 47:1435–1442.

Müller NA, Kersten B, Leite Montalvão AP, Mähler N, Bernhardsson C, Bräutigam K, Carracedo Lorenzo Z, Hoenicka H, Kumar V, Mader M, et al. 2020. A single gene underlies the dynamic evolution of poplar sex determination. *Nat. Plants* 6:630–637.

Murase K, Shigenobu S, Fujii S, Ueda K, Murata T, Sakamoto A, Wada Y, Yamaguchi K, Osakabe Y, Osakabe K, et al. 2017. *MYB* transcription factor gene involved in sex determination in *Asparagus officinalis*. *Genes Cells* 22:115–123.

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32:268–274.

Norup MF, Petersen G, Burrows S, Bouchenak-Khelladi Y, Leebens-Mack J, Pires JC, Linder HP, Seberg O. 2015. Evolution of *Asparagus* L. (Asparagaceae): Out-of-South-Africa and multiple origins of sexual dimorphism. *Mol. Phylogenet. Evol.* 92:25–44.

Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. 2011. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet.* 27:358–367.

Ou S, Su W, Liao Y, Chougule K, Agda JR, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, et al. 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* 20:1–18.

Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, Capron A, Xie L-F, Ye D, Sundaresan V. 2005. Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. *Development* 132:603–614.

Palmer DH, Rogers TF, Dean R, Wright AE. 2019. How to identify sex chromosomes and their turnover. *Mol. Ecol.* 28:4709–4724.

Pertea G, Pertea M. 2020. GFF utilities: GffRead and GffCompare. *F1000Research* 9:304.

Plath S, Klocke E, Nothnagel T. 2022. Karyological and nuclear DNA content variation of the genus *Asparagus*. *Plos One* 17:e0265405.

RC-Team. 2020. R: A language and environment for statistical computing.

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

Renner SS, Müller NA. 2021. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nat. Plants* 7:392–402.

Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 21:1–27.

Rice WR. 1984. Sex chromosomes and the evolution of sexual dimorphism. *Evolution*:735–742.

Sauquet H, von Balthazar M, Magallón S, Doyle JA, Endress PK, Bailes EJ, de Morais EB, Bull-Hereñu K, Carrive L, Chartier M. 2017. The ancestral flower of angiosperms and its early diversification. *Nat. Commun.* 8:1–10.

Shirasawa K, Ueta S, Murakami K, Abdelrahman M, Kanno A, Isobe S. 2022. Chromosome-scale haplotype-phased genome assemblies of the male and female lines of wild asparagus (*Asparagus kiusianus*), a dioecious plant species. *DNA Res.* 29:dsac002.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* 24:637–644.

Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:1–11.

Steinemann S, Steinemann M. 2005. Retroelements: tools for sex chromosome evolution. *Cytogenet. Genome Res.* 110:134–143.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.

Tang S, Lomsadze A, Borodovsky M. 2015. Identification of protein coding regions in RNA

transcripts. *Nucleic Acids Res.* 43:e78–e78.

Tennessen JA, Wei N, Straub SC, Govindarajulu R, Liston A, Ashman T-L. 2018. Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biol.* 16:e2006062.

Tian G-W, Chen M-H, Zaltsman A, Citovsky V. 2006. Pollen-specific pectin methylesterase involved in pollen tube growth. *Dev. Biol.* 294:83–91.

Tsugama D, Matsuyama K, Ide M, Hayashi M, Fujino K, Masuda K. 2017. A putative *MYB35* ortholog is a candidate for the sex-determining genes in *Asparagus officinalis*. *Sci. Rep.* 7:41497.

Vicoso B. 2019. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nat. Ecol. Evol.* 3:1632–1641.

Vuruputoor VS, Monyak D, Fetter KC, Webster C, Bhattarai A, Shrestha B, Zaman S, Bennett J, McEvoy SL, Caballero M, et al. 2023. Welcome to the big leaves: Best practices for improving genome annotation in non-model plant genomes. *Appl. Plant Sci.* 11:e11533.

Westergaard M. 1958. The mechanism of sex determination in dioecious flowering plants. *Adv. Genet.* 9:217–281.

Wu M, Haak DC, Anderson GJ, Hahn MW, Moyle LC, Guerrero RF. 2021. Inferring the genetic basis of sex determination from the genome of a dioecious nightshade. *Mol. Biol. Evol.* 38:2946–2957.

Xu P, Chen H, Hu J, Pang X, Jin J, Cai W. 2022. Pectin methylesterase gene *AtPMEPCRA* contributes to physiological adaptation to simulated and spaceflight microgravity in Arabidopsis. *Iscience* 25.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Zhang Z. 2022. KaKs_calculator 3.0: Calculating selective pressure on coding and non-coding sequences. *Genomics Proteomics Bioinformatics* 20:536–540.

Zhao L, Xu S, Chai T, Wang T. 2006. *OsAP2-1*, an *AP2*-like gene from *Oryza sativa*, is required for flower development and male fertility. *Sex. Plant Reprod.* 19:197–206.

Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* 39:btac808.

Zhu X, Tang C, Li Q, Qiao X, Li X, Cai Y, Wang P, Sun Y, Zhang H, Zhang S, et al. 2021. Characterization of the pectin methylesterase inhibitor gene family in Rosaceae and

role of *PbrPMEI23/39/41* in methylesterified pectin distribution in pear pollen tube. *Planta* 253:118.

Zhu Y, Ahmad Z, Lv Y, Zhang Y, Chen G. 2025. Insight into the Characterization of Two Female Suppressor Gene Families: *SOFF* and *SyGI* in Plants. *Genes* 16:280.

Figures

Figure 1. Transitions from an ancestral bisexual state to derived lineages with separate sexes—dioecy evolved twice in the genus *Asparagus*. The dioecious Eurasian clade is widespread across Eurasia, including garden asparagus (*A. officinalis*) and >50 additional species. The Mediterranean Basin dioecy clade, with *A. horridus*, is less speciose (~3-4 spp.) and geographically restricted to regions around the Mediterranean Sea. Two Y-linked genes control sex-determination in garden asparagus: a female suppressor (*SOFF*) and male promoter (*aspTDF1*) (Harkess et al. 2017). *SOFF* is male-specific, or Y-linked, across all analyzed Eurasian clade taxa and is associated with the origin of dioecy in the Eurasian clade. A Y-linked *aspTDF1* gene arose later and is absent in Mediterranean Basin clade taxa and in early diverging Eurasian clade lineages (Murase et al. 2017). Species tree adapted from Bentz, Burrows, et al. (2024). Two dioecy origins are marked by separate sex symbols. Branch lengths shown in coalescent units. Local posterior probability branch support was 1.0 for all branches. *Sex-linkage of *aspTDF1* has not been tested in *A. filicinus*.

Figure 2. Photograph of the *Asparagus horridus* male genome plant (accession pb32m) and a species-typical staminate flower produced by XY genotypes—exhibiting vestigial (non-functional) female reproductive organs (pistils). Seed for pb32m was wild collected from Mallorca (Majorca), Spain during April 2018.

Figure 3. XY sex chromosomes in *A. horridus* evolved from the ancestral chromosome 3 pair in the Mediterranean Basin dioecy clade. **a)** A large inversion marks the boundaries of the nonrecombining, male-specific region on the Y (MSY) (~9.6 Mb; yellow block) and corresponding X-specific region (~2.1 Mb). An X-Y haplotype alignment and structural annotation densities (X=top; Y=bottom, shown in 1 Mb windows) show increased Y-mer (male-specific *k*-mers identified between 8 males and 7 females) and LTR retrotransposon densities throughout the MSY, corresponding to the ~9.6 Mb inversion break points. A small region (~0.65 Mb), nested within the MSY boundaries, was found to be oriented in the same direction as the X and contains 3 genes (see shaded region in panel d). **b)** Synonymous substitution (d_S) estimates from three different comparisons suggest that the *A. horridus* MSY (teal curve) is younger than the total divergence between *A. horridus* and *A. officinalis* (purple and yellow curves). Yellow: *A. horridus* vs. *A. officinalis* genome-wide orthologs. Purple: *A. horridus* MSY genes vs. *A. officinalis* orthologs. Teal: *A. horridus* MSY genes vs. X-linked gametologs. **c)** Dioecy evolved ~1.13-1.81 million years ago in the Mediterranean Basin clade, ~1-2 million years later than in the Eurasian clade (Bentz, Liu, et al. 2024).

Dotted branches represent hermaphroditic lineages and the ancestral state. **d)** Compared to the pseudo-autosomal region (PAR), d_S values were consistently elevated between MSY genes and X-gametologs, aside from three outlier genes (filled circles labeled with gene annotations). **e)** Male vs. female (flower) transcription levels, of MSY genes with X-gametologs, revealed significant (adjusted *p*-value<0.05, log2 fold change >1 or <-1: red circles), sex-specific differences; compared to females, 20 genes were up-regulated, and 7 genes were down-regulated in males. Notable genes are labeled with annotations. *PMEI* a and *PMEI* b are different pectin methylesterase inhibitor encoding genes. *CYP* is a cytochrome P450 enzyme encoding gene.

Figure 4. The garden asparagus (*A. officinalis*) XY sex chromosomes correspond to chromosome 1**. a)** X-Y haplotype alignment and structural annotations (X=top; Y=bottom) show support for a hemizygous, male-specific region on the Y, or MSY (yellow block), where Y-mer (male-specific *k*-mers identified between 3 males and 3 females) and LTR retrotransposon densities peak, compared to the surrounding pseudo-autosomal regions. **b)** Hemizygous regions between the XY pair mark the Y-specific MSY and X-specific region. The MSY contains 10 genes, including two sex-determination genes: *SOFF* and *aspTDF1* (diamonds on Y) (Harkess et al. 2017), whereas the X-specific region contains a single gene (*aspWIP2*) which is missing a Y-linked gametolog. **c)** A *SOFF* (*DUF247* gene family) phylogeny revealed strong support for separate clades with either male-specific (Y-linked) or autosomal-linked homologs (on chromosome 5) from *A. officinalis* and *A. kiusianus*; and uncertain placement of a male-specific homolog from *A. cochinchinensis*. The *SOFF* tree supports a recent duplication of an ancestral *DUF247* gene preceded neofunctionalization of sex-determining function in the most recent common ancestor (MRCA) of the Eurasian clade, which may have been lost in a common ancestor of *A. officinalis* and *A. kiusianus* but maintained in *A. cochinchinensis*. **d)** The *aspTDF1* tree supports the stepwise recruitment of *aspTDF1* into the MSY of a common ancestor of *A. officinalis* and *A. kiusianus*, following divergence from the MRCA shared with *A. cochinchinensis*. *Asparagus setaceus* is a hermaphroditic species, representing the ancestral condition for the genus. Orthologs from *Agave* were used to root both gene trees. Bootstraps shown for branches with <100% support. Asterisks (*) mark genes from the new *A. officinalis* genome.

Figure 5. Two XY sex chromosome systems evolved from different ancestral autosomes in *Asparagus*, a genus in the Asparagaceae subfamily Asparagoideae. The chromosome 1 pair represents the XYs in *A. officinalis* (Eurasian clade) whereas the chromosome 3 pair represents the XYs in *A. horridus* (Mediterranean Basin clade). X chromosome haplotypes not shown here. **a)** Syntolog relationships across three Asparagaceae subfamilies

(*Dracaena*=Nolinoideae; *Yucca*=Agavoideae) illustrate rampant genome rearrangements that have occurred across >41 million years of lineage divergence. **b)** The male-specific region on the Y (MSY) in *A. officinalis* is nested between syntologs conserved on the *Asparagus* chromosome 5 and the *Dracaena* chromosomes 2 and 7 (note: these are not homologous with chromosomes 2 and 7 in *Asparagus*). **c)** Synonymous substitution (d_S) distributions, measured between *A. officinalis* chromosome 1 (regions adjacent to the MSY) and chromosome 5 syntologs (purple curve), are congruent with d_S between all *Asparagus–Dracaena* syntologs (yellow curve); suggesting that an ancient genome duplication is responsible for the observed homology between chromosomes 1 and 5 in *Asparagus*. Importantly, the shared duplication predates both origins of dioecy in *Asparagus* by at least 36-38 million years. **d**) Lineage-specific *DUF247* gene family expansions are common across Asparagaceae taxa and usually occur via tandem duplications (blue clades). The Y-specific *DUF247* (*SOFF*) in *A. officinalis* arose in an *Asparagus*-specific clade with autosomal homologs from chromosomes 1 and 5. Single gene duplications=yellow clades. Clades were collapsed and labeled with the number of Asparagaceae subfamilies (out of the 3 subfamilies analyzed here) that share the indicated duplication pattern. Black box tip marks the *SOFF* gene tree root used for Fig. 4c. Grey branches are *Oryza sativa* (rice) homologs used as a control for major clades (Zhu et al. 2025). Bootstraps shown for branches with <100% support.

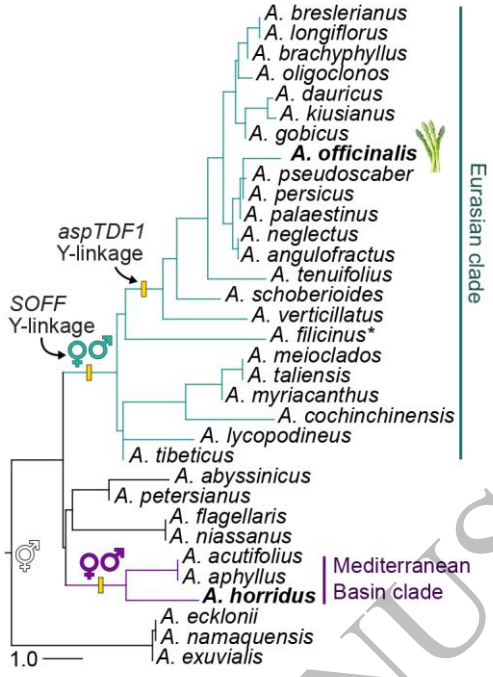

Figure 1
64x88 mm (x DPI)

Figure 2
63x34 mm (x DPI)

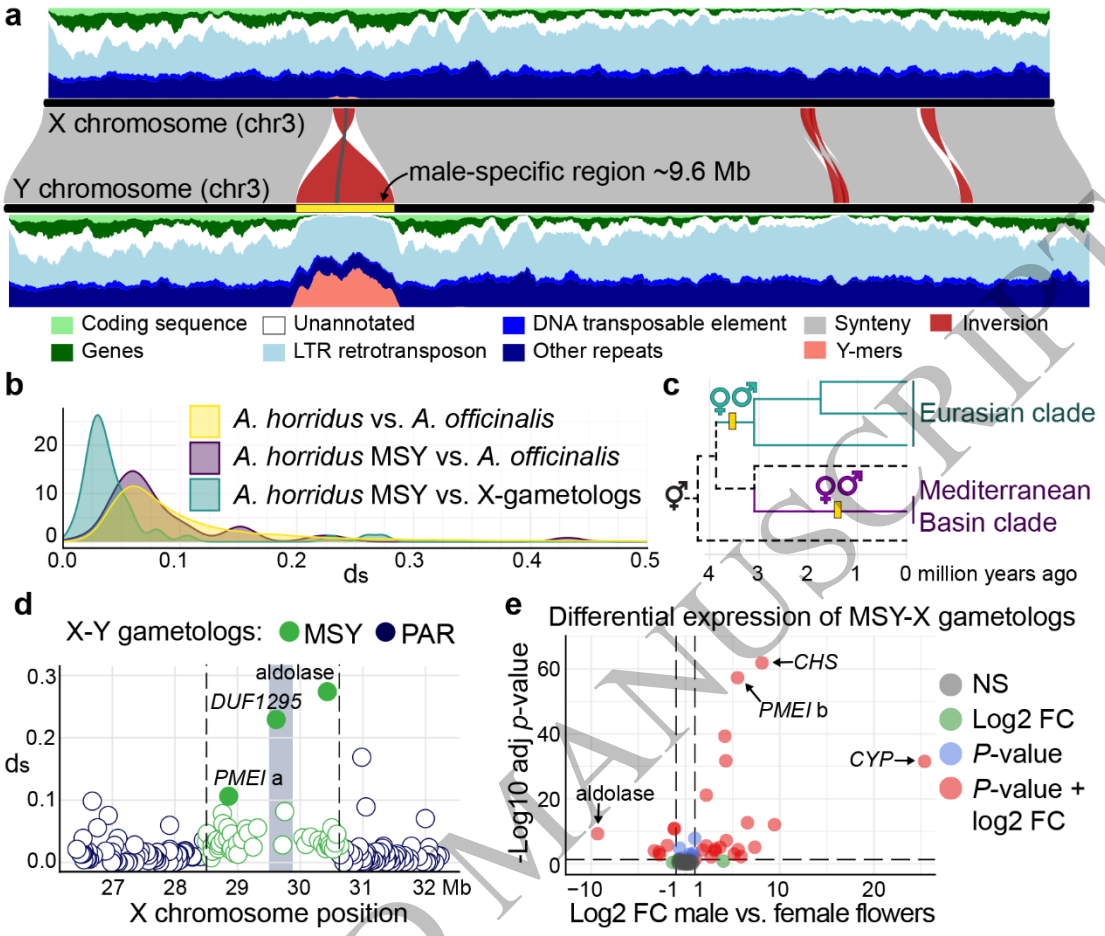

Figure 3
145x122 mm (x DPI)

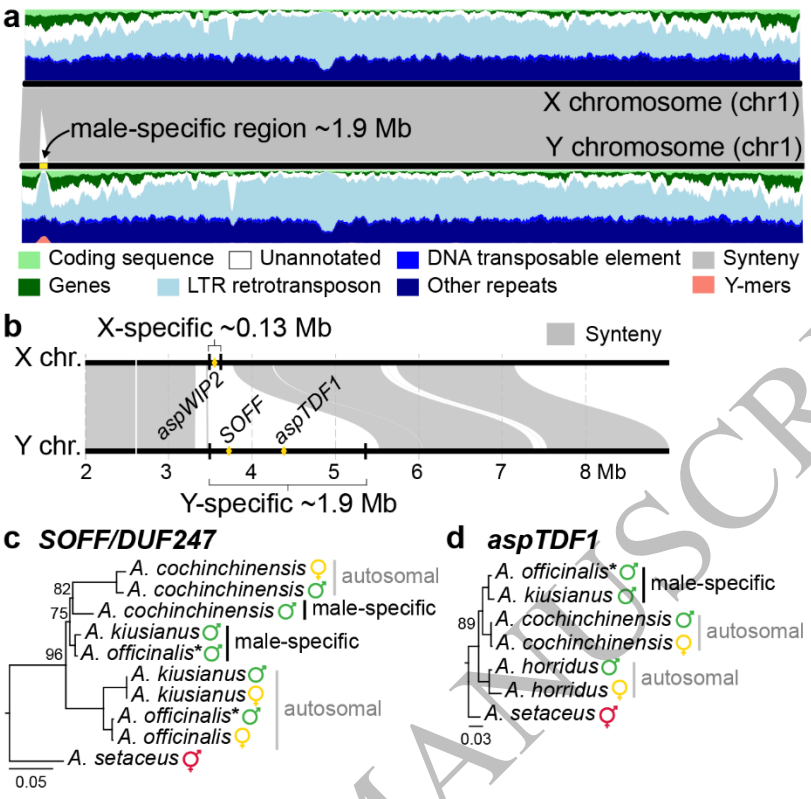

Figure 4
107x104 mm (x DPI)

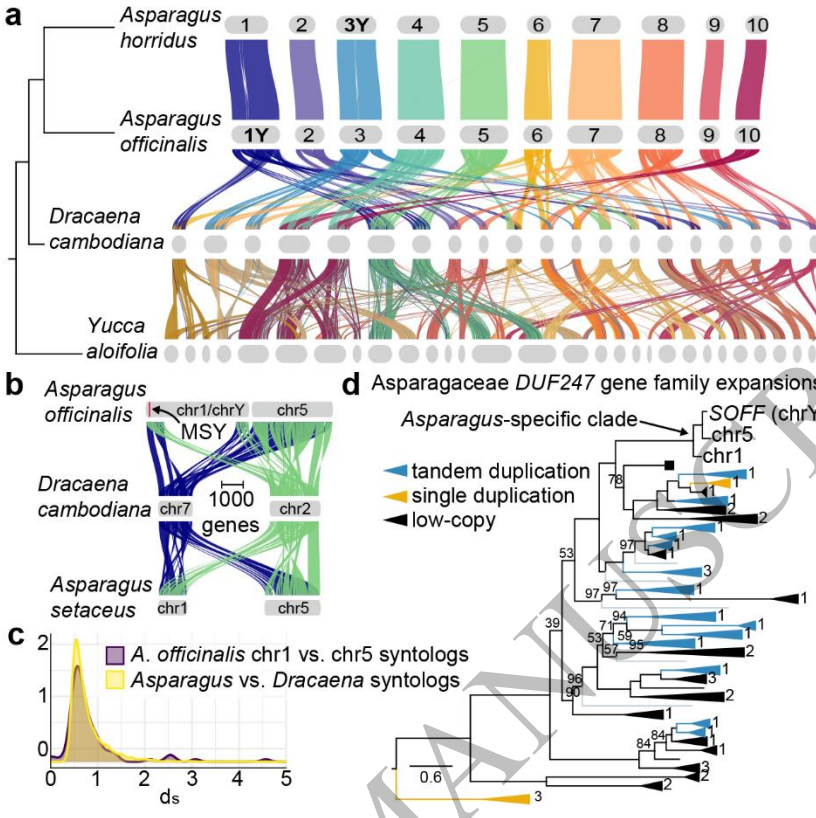

Figure 5
110x106 mm (x DPI)