

A haplotype-resolved nutmeg genome reveals XY sex chromosome evolution and enables molecular sexing

Received: 24 November 2025

Accepted: 18 May 2026

Cite this article as: Matvijev, K., Wang, S., Xiao, M. *et al.* A haplotype-resolved nutmeg genome reveals XY sex chromosome evolution and enables molecular sexing. *Commun Biol* (2026). <https://doi.org/10.1038/s42003-026-10355-0>

Katarina Matvijev, Sufang Wang, Mengjun Xiao, Ashini D. Mahadura, Ray Ming & Dmitry A. Filatov

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

A haplotype-resolved nutmeg genome reveals XY sex chromosome evolution and enables molecular sexing

Katarina Matvijev^{1,#}, Sufang Wang^{1,2,#}, Mengjun Xiao^{3,#}, Ashini D. Mahadura^{1,4}, Ray Ming^{3**} and Dmitry A. Filatov^{1*}

¹ Department of Biology, University of Oxford, South Parks Rd, Oxford OX1 3EL, United Kingdom

² School of Life Science and Technology, Northwestern Polytechnical University, 1 Dongxiang Road, Xi'an, Shaanxi, 710129, China

³ Centre for Genomics and Biotechnology, Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Key Laboratory of Genetics, Breeding and Multiple Utilization of Crops, Ministry of Education, Fujian Agriculture and Forestry University, Fuzhou, 350002, China

⁴ Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic.

1st co-authorship

Correspondence:

* dmitry.filatov@biology.ox.ac.uk (ORCID: 0000-0001-8077-5452)

** rayming@illinois.edu (ORCID: 0000-0002-9417-5789)

Abstract

Nutmeg (*Myristica fragrans* Houtt.) is a dioecious species that produces one of the most valuable spices and has peculiar holocentric chromosomes with diffuse centromere. The impact of holocentricity on the evolution of sex chromosomes is not well understood. Here, we present the first fully annotated haplotype-resolved chromosome-level genome of a male nutmeg plant, identify its sex chromosomes and analyse evolution of the sex-linked region (SLR). We demonstrate that nutmeg has a male heterogametic (XY) system, with a 0.23 Mb X-linked SLR and a 3.1 Mb Y-linked SLR with limited evidence for Y-degeneration. Low X:Y synonymous divergence for X- and Y-linked homologs ($K_s < 2.5\%$) indicates recent evolution of sex linkage, likely through loss(es) of homologous X-linked region(s). Leveraging the new genome and SLR sequences, we developed PCR-based molecular marker for reliable sexing of nutmeg seedlings, enabling efficient management of plantations by ensuring the optimal male to female ratio (1:10).

Keywords: Sex chromosome evolution, dioecy, nutmeg, genome sequence

Introduction

The evolution of sex chromosomes in plants is a persistently fascinating topic. Most animals have separate sexes and sex chromosomes usually evolve and re-evolve via consecutive turnovers¹. On the other hand, in flowering plants separate sexes (dioecy) are uncommon (ca. 6%,²) and sex chromosomes often evolve *de novo* during or after transition to dioecy, though turnovers in ancestrally dioecious angiosperm clades, such as willows, have also been described³. Even though significant advances in understanding of sex chromosome evolution have been made, the origins and dynamics of plants sex chromosomes are still not fully understood⁴. The majority of dioecious angiosperms with an identified reproductive system have male-heterogamety (XX/XY sex chromosomes in females and males, respectively), although female-heterogamety (ZZ/ZW) has also been reported⁵. Sex chromosomes that can be clearly distinguished under a microscope (heteromorphic) were described only in a few plants, such as *Silene*⁶ and *Rumex*⁷, with most dioecious plants having morphologically indistinguishable (homomorphic) sex chromosomes⁸. In heteromorphic sex chromosomes, the sex-linked region (SLR) usually spans most of the chromosome length whereas in homomorphic chromosomes, the SLR is usually smaller⁹. Sex chromosomes in plants were previously considered to be of a recent origin due to their tendency to be homomorphic, but accumulation of genome sequence data revealed the existence of species with old homomorphic sex chromosomes, such as ginkgo¹⁰. Although dioecy occurs in low numbers in most orders of angiosperms, certain families are exclusively or almost exclusively dioecious^{11,12}, suggesting that sex chromosomes in these families are old. However, turnovers of sex chromosomes can keep them ‘young’ even in the families where dioecy is ‘old’, such as reported for willows^{3,13}. Previously, dioecy has been considered an “evolutionary dead-end” mainly due to its rarity in angiosperms¹⁴ or because dioecious lineages have higher extinction rates¹⁵. Now it is accepted that dioecy is relatively rare most likely due to frequent reversals from dioecy to cosexuality¹¹, and that plant reproductive systems are much more evolutionary dynamic compared to animals¹⁶.

Highly evolutionary dynamic mating systems in plants provided an opportunity to study the early stages of *de novo* sex chromosome evolution when a cosexual species evolves separate sexes^{4,17}. Early genetic work in the dioecious white campion revealed the presence of two sex-determining factors on the Y chromosome – the gynoecium suppression and stamen promotion factors¹⁴, which inspired the development of the classic two factor model of sex chromosome evolution¹⁸. The analysis of sex determination in a range of species¹⁹ supported the two factor model. More recently²⁰, it was argued that a single factor model accounts for sex determination in persimmon²¹ and an artificial switch to dioecy in melon²² illustrated how a single factor sex determination could arise. Currently, it is widely accepted that the emergence of sex chromosomes in plants requires at least two mutations – male and female sterility mutations²³, which can be genetic variation at a single locus (single-factor determination,²⁰) or at two linked loci (two-factor determination,¹⁸). The power of the two factor model is in its ability to explain the cause of initial recombination suppression between the sex-determining loci – linking male-promoting and female-suppressing factors prevents the formation of sterile plants that inherited both male and female sterility mutations¹⁸. This nascent non-recombining Y (NRY = MSY, male specific Y) region around the sex-determining loci has a tendency to expand over time, forming evolutionary strata – the regions with distinct X:Y divergence, which was first described in humans²⁴, but turned out to be a general property of sex chromosomes in animals²⁵. Plant sex chromosomes typically contain small NRY⁴, with expansions reported only for a few plants with large

NRV, such as in *Silene latifolia*²⁶⁻²⁸. The causes of NRV expansion are not fully understood, but are thought to be at least partly driven by sexually antagonistic genes^{12,29,30}.

Here we report identification and analysis of homomorphic sex chromosomes in the dioecious nutmeg (*Myristica fragrans* Houtt.). Species in the nutmeg family (Myristicaceae) are generally dioecious, with 15 out of 20 genera having either exclusively or predominantly dioecious species², indicating that dioecy is ancestral in this family. However, it is unclear whether sex chromosomes, if any, are old and shared between dioecious species in this family. True nutmeg, *M. fragrans*, is a tropical tree species autochthonous to the Moluccas islands, but now present in other tropical regions due to its economic relevance in the spice industry. This species is generally described as dioecious, although there are recurrent records of monoecious individuals^{31,32}, which likely represent ‘inconstant’ males or females described in many dioecious species³³. Apart from their unisexual flowers, nutmeg males and females are indistinguishable, meaning that individuals can be sexed only once they reach reproductive maturity after 8-10 years, which is a problem for nutmeg growers who have to optimise sex ratio (1 male to ~10 females) on nutmeg plantations. The previously proposed supposedly sex-linked molecular markers³⁴ for sexing nutmeg seedlings before planting them do not reliably distinguish males and females in our analyses (see below). Thus, apart of intrinsic scientific interest of studying sex chromosomes and their evolution, identification of the nutmeg sex linked region (SLR), reported in this study, is relevant for agriculture.

Apart of economic importance, dioecy and sex chromosomes, nutmeg is interesting due to its holocentric chromosomes³⁵ and interplay between holocentricity and sex chromosome evolution. Holocentricity – diffuse centromere with the mitotic spindle attaching along the entire chromosome length, is relatively rare, but it evolved multiple times in a range of animals (e.g. Lepidoptera, nematodes) and plants (e.g. rushes and sedges)³⁶. Holocentric species have many peculiar features, such as inverted meiosis – segregation of sister chromatids before segregation of homologs^{37,38}. Another interesting feature of holocentric species is their relative robustness to chromosome fragmentation – a chromosome fragment is usually lost in monocentrics, but it usually segregates correctly in holocentrics, raising debates whether holocentricity leads to evolution of more numerous smaller chromosomes³⁹ and whether chromosome speciation is more common in holocentric than in monocentric species⁴⁰. Distribution of recombination along the chromosomes differs between mono- and holocentric species. In monocentrics, recombination is usually suppressed around the centromere⁴¹, resulting in pericentromeric region accumulating transposable elements and other repetitive DNA and losing functional genes, not dissimilar to non-recombining Y-chromosomes⁴². Accumulation of junk DNA in pericentromeric region may lead to its expansion and it was recently reported that pericentromeric expansion likely played a significant role in sex chromosome evolution in *Silene latifolia*²⁷. Pericentromeric recombination suppression was also reported to play a significant role in initial recombination suppression on nascent plant sex chromosomes⁴³. Holocentric species lack localised centromere and do not have pericentromeric recombination suppression⁴⁴, which likely has implications for sex chromosome evolution. A combination of dioecy, sex chromosomes and holocentricity is rare in plants and nutmeg provides an opportunity to study evolution of homomorphic sex chromosomes in holocentric plants.

Here we first present the first fully annotated haplotype-resolved chromosome-level assembly of a male nutmeg genome. We then identify the sex-linked region using multiple approaches and analyse the

features of sex-linked and adjacent regions to infer evolution of sex chromosomes in nutmegs. Our results confirm the previous reports that nutmeg is a species with 22 homomorphic chromosomes^{31,34}, but contrary to previous suggestions of female heterogamety (ZW)³⁴, we report that male nutmeg plants have XY and females have XX sex chromosomes. We report the evidence for recent expansion of nutmeg NRY and analyse evolutionary genetic processes in the old and recently added sex-linked regions. The sequence of the nutmeg Y-chromosome allowed us to develop molecular markers for reliable sexing of nutmeg seedlings to optimise sex ratio of nutmeg plantations without the need to wait ~8-10 years until flowering.

Results

Genome size and chromosome number

To estimate the genome size of *M. fragrans*, we used flow cytometry on three male individuals, with tomato (genome size $\approx$ 900 Mb) or maize B73 (genome size $\approx$ 2.3 Gb) as references. The genome size of *M. fragrans* was estimated to be $\sim$ 590 Mb using this method (Table 1). Cytological analysis confirmed that the sex chromosomes cannot be distinguished under the microscope without specific labelling (i.e. they are homomorphic), and that the haploid number of chromosomes is $n=22$ (Suppl. Figure S1).

Genome assembly, annotation and resequencing

We generated a chromosome-level haplotype-resolved assembly of the male *M. fragrans* genome using a combination of long-read PacBio HiFi sequencing and the high-resolution chromosome conformation capture (Hi-C). We generated a total of 19,542,512,782 bp (Suppl. Table S1) of raw long read data for the reference male genome assembly, as well as more than 6 Gb of RNA-seq short read data (Suppl. Table S2). The final primary male genome assembly had a length of 641 Mb across 22 fully resolved chromosomes (Figure 1, Table 2 and Suppl. Table S3), which is consistent with cytological chromosome count (Suppl. Figure S1) as well as genome size estimation based on K-mers and cytometry (Table 1). Similarly to other holocentric species⁴⁵, the genome of *M. fragrans* has relatively low GC-content: 36.7% that is evenly distributed along the chromosomes (Figure 1A).

Gene prediction using BRAKER3 v3.0.8⁴⁶ identified 19,857 and 19,904 genes on haplotype 1 (hereafter “hap1”) and haplotype 2 (hereafter “hap2”), respectively. BUSCO completeness was 97.5% and 97.2% for the two haplotypes, respectively. Transposable elements (TEs) comprised 37.42% of the hap1 and 38.60% of the hap2 sequence length (Suppl. Table S4). This relatively low TE content for a plant genome of modest size (e.g. 52% in oak⁴⁷, 68% in tomato⁴⁸ and 73% in *Silene conica*⁴⁹) is possibly due to holocentricity of nutmegs. Pericentromeric regions in monocentric chromosomes are typically rarely recombining and accumulate TEs. Such regions can be very extensive in monocentric plants, with recombination occurring mainly at the ends of the chromosomes (e.g. ^{27,50,51}). Holocentric chromosomes lack localized centromeres and therefore do not exhibit canonical pericentromeric regions. Recombination on such chromosomes is typically evenly distributed along the chromosome length⁵². This likely prevents accumulation of TEs; e.g. in holocentric butterflies, the average TE content is similar to that in nutmeg – about 39%⁵³.

Generally, genes (Figure 1B) and TEs (Figure 1C) were quite uniformly distributed across the genome. However, a ca. 2.5 Mb region on chromosome 6 of hap1 had an unusually high TE density and was fairly

gene poor (Figures 1B and 1C), which suggested this may be a non-recombining sex-linked region. The analyses presented below confirm this hypothesis.

Comparison with *Myristica yunnanensis* genome

The newly assembled nutmeg genome was compared with the recently published⁵⁴ unphased genome of congeneric *Myristica yunnanensis*. The published *M. yunnanensis* genome assembly contains 19 chromosomes and sex chromosomes were not analysed. This comparison reveals that synteny for many chromosomes is preserved, though multiple chromosome inversion, fission and fusion events are evident between *M. yunnanensis* and *M. fragrance* genomes (Figure 2A). The preservation of synteny for many chromosomes is not too surprising given the relatively low divergence between these species: mean divergence at synonymous sites across 13,577 1-to-1 orthologs across the genome is only 2.53% (Figure 2B), indicating recent divergence between these species. Non-synonymous divergence for the same set of genes is significantly lower than synonymous divergence (Figure 2B), reflecting the predominance of purifying selection eliminating deleterious mutations affecting protein sequence.

Synteny analysis for the nutmeg chromosome 6 reveals many regions with micro-synteny to different *M. yunnanensis* chromosomes, but no clearly macro-syntenic chromosomes (Figure 2C). Lack of macro-synteny of nutmeg chromosome 6 with *M. yunnanensis* chromosomes precludes identification of *M. yunnanensis* likely sex chromosome based on synteny analysis only. Small part of *M. fragrance* sex-linked region including 5 genes is micro-syntenic to a region on *M. yunnanensis* chromosome 3 (Figure 2D), but adjacent regions do not show any synteny. It is possible to hypothesize that this region on chromosome 3 corresponds the sex-determining region in *M. yunnanensis*, but this cannot be confirmed without additional work in that species. The sex of the plants used for sequencing of *M. yunnanensis* genome is not known because the plants were too young to flower (Jing Cai, personal communication), making it impossible to identify sex chromosome in this species based on published data.

Identification of the sex-linked region

To identify the location of the sex-linked region in the nutmeg genome, we used three complimentary approaches based on the analyses of single nucleotide polymorphisms (SNP) and sequence read coverage using Illumina genome resequencing data from seven males and eleven females. In total, we generated 383 Gb of raw short-read data (Suppl. Table S5). On average, 45.02 Gb (± 5.32 , SD) of genome sequence data were obtained per individual. These data were used to read mapping and SNP calling, with each male haplotype used separately as the reference.

The GWAS for sex-linkage (with Bonferroni correction, where only SNPs with a p-value lower than 1.49×10^{-8} were considered statistically significant) revealed that a ~ 3.1 Mb region on chromosome 6 of hap1 was statistically significantly associated with sex (Figure 3A). The 2681 SNPs of interest were clustered on chromosome 6 from 26.47 Mb to 29.57 Mb on hap1. There were no statistically significant SNPs observed in hap2 (Suppl. Figure S2A).

The average sequencing coverage across all individuals was 36.44 (± 7.04 , SD) for hap1 and 36.85 (± 7.1 , SD) for hap2. The coverage analysis revealed a fairly uniform distribution across the genome (Suppl. Figure S3), with variation mainly in repeat-rich regions, which are known to cause mapping artifacts. However, a ~ 3 Mb long region on chromosome 6 of hap1 showed a stark difference in coverage between sexes, where all female individuals had a coverage close to 0, while male individuals

maintained levels similar to surrounding regions, indicating this region is likely sex-linked and male-specific (Figure 3B). This pattern was absent in hap2 (Figure 3J), indicating that *M. fragrans* is male heterogametic, and that chromosome 6 of hap1 contains the Y-SLR (Figure 3F) whereas the corresponding region of hap2 is the X-SLR (Figure 3H).

The average F_{st} values between males and females at the chromosome and genome level were close to 0, but was elevated around the putative sex-linked regions on chromosome 6 on both hap1 (Figure 3C), and hap2 (Figure 3I). Net sequence divergence (D_a) between males and females was near zero all over the genome, except the Y-SLR (Figure 3C) and X-SLR (Figure 3I) on chromosome 6.

To further confirm the location of the sex-linked region and to pinpoint its boundaries, we analysed the distribution of sex-specific SNPs. For the SNPs called using hap1 (Y) as the reference, we required all females to be homozygous for the non-reference allele ("ALT" allele in the VCF file), while the males needed to be heterozygous. For hap1 such male-specific SNPs were found only between positions 26.47Mb and 29.57Mb of the chromosome 6 (Figure 3D). It is worth noting that while the average sequence coverage in females for this region was low (Figure 3B), it was not zero, allowing us to call SNPs at some positions in females in this region.

For the SNPs called using hap2 (X) as the reference, we required all females to be homozygous for the reference allele, while all males needed to be heterozygous. Such SNPs were found only between positions 25.4Mb and 26.7Mb on the hap2 of the chromosome 6 (Figure 3H). They form a major peak around position 26.5Mb and a minor peak around positions 25.5Mb. The minor peak between positions 25,474,965 and 25,506,534 includes 64 male-specific SNPs, while the major peak between positions 26,488,672 and 26,705,386 includes 278 male-specific SNPs. Two peaks with similar positions are also visible in D_a and F_{st} plots (Figure 3I), suggesting that either the entire region between 25.465Mb and 26

The results of the analyses reported above indicate that the region between 26.47 Mb and 29.57 Mb of the chromosome 6 hap1 corresponds to the Y-SLR, which is 3.1 Mb long, while the X-SLR is located between positions 26.49 and 26.72 and is only 0.28 Mb long. The difference in length between the X- and Y-linked SLRs is due to presence of male-specific Y-linked (MSY) regions in the haplotype 1 (Figure 3E). The rest of the chromosome 6 is likely pseudoautosomal, as evidenced by complete collinearity (Figure 3F), lack of sequence divergence and sex-specific SNPs in the regions to the left (PAR1) and right (PAR2) of the sex-linked region (Figures 3E and 3G).

Composition of the sex-linked region

Eight and 32 intact genes were annotated within the SLR on the X and Y haplotypes, respectively. The identified sex-linked genes had various functions, including transcription, cellular processes, respiratory processes, biosynthetic processes, and developmental processes (Suppl. Table S6). Several genes were identified as having protein domains related to plant development, seed and flower growth, and hormonal regulation (KH, Zf-CCHC and DnaJ protein domains), of which only the DnaJ protein domain gene was found both in the X- and Y-SLR. However, a more detailed functional analysis of these genes would be necessary to further explore their function and study their possible role in sex determination.

The comparison of gene order along the two haplotypes of the chromosome 6 revealed nearly complete collinearity, except the Y-SLR, where 24 Y-linked genes lacked gametologs in hap2 (Figure 3F). The male-specific genes in Y-SLR are arranged in two blocks that we called MSY1 and MSY2 (Figure 3E). MSY1 and

MSY2 are separated by a 101-kb long X-homologous region 1 (Xh1) that contains 6 genes. Five of these genes have gametologs in the X-SLR (Figure 3F) and homologs in the outgroup species *M. yunnanensis* (Figure 2D). Phylogeny for these genes, rooted by the outgroup sequences shows that X- and Y-linked genes of *M. fragrans* cluster together, while the outgroup sequence is more divergent and equidistant from the X- and Y-linked genes (Figure 3M). The clustering of the *M. fragrans* X- and Y-linked genes implies evolution of sex linkage in this region after divergence from *M. yunnanensis*, but other parts of the *M. fragrans* SLR may be considerably older.

The second region of homology with the X-SLR (Xh2) is present at the end of Y-SLR (Figure 3E). This region contains three genes all of which have intact X- and Y-linked gametologs. Nucleotide divergence between X- and Y-linked gametologs in both Xh1 and Xh2 regions is low, $K_s < 2.5\%$ for most X:Y pairs (Figure 3K), indicating recent cessation of X:Y recombination, or a translocation from one to the other sex chromosome. Given the collinearity of hap1 and hap2 is preserved in this region, the translocation is a less likely explanation than the recent cessation of recombination due to expansion of the non-recombining sex-linked region.

The analysis of TEs revealed that MSY2 is significantly enriched with LTR retrotransposons, while LINEs and TIRs are significantly depleted in this region compared to the adjacent regions (Figure 4) and the rest of the genome. TEs often accumulate in non-recombining regions, but TE density in MSY1 is similar to that in the PAR1 and significantly differs from MSY2 (Figure 4), suggesting that MSY1 is a recent addition to the SLR, while the MSY2 has lacked recombination for a longer period of time. This is consistent with low nucleotide divergence (K_s) between X- and Y-linked gametologs of genes in the region between the MSY1 and MSY2 (Figure 3K). We hypothesize that the MSY2 represents the older stratum, while MSY1 and XhR evolved sex linkage in the last ~5 million years ($= g K_s / 2m$), based on average X:Y $K_s = 0.7\%$ for sex-linked genes with homologs present on the X and the Y chromosomes (Figure 3K) and assuming $g = 10$ years per generation (as nutmeg first flowers at the age ~8 years) and $m = 7 \times 10^{-9}$ per nucleotide per generation mutation rate (see Table 1 in Krasovec et al ⁵⁵).

The distribution of sequence divergence between copies of TE of the same family across the genome reveals that most TE families were active a long time ago, with the major peak of divergence at ~28% (Figure 4K). There is also a distinct peak including multiple types of TEs at ~10% sequence divergence, indicating a short burst of TE activity around 71 million years ago (MYA), assuming 10 years generation time and 7×10^{-9} per nucleotide per generation mutation rate, based on the mutation rates reported for various angiosperms (see Table 1 in Krasovec et al ⁵⁵). However, it is worth noting that generation time has likely changed over evolutionary time, and the mutation rate for TE may differ from nuclear genes used by Krasovec et al ⁵⁵, creating considerable uncertainty in dating ancient bursts or TE activity. LTR retrotransposons, dominated by copia, show a peak of recent activity (left peak at Figure 4K), indicating that currently, or very recently, this TE is the most actively transposing type in the nutmeg genome. The divergence between two LTRs of the same copy provides a proxy for time since insertion of individual retrotransposon copies. The distribution of LTR divergence peaks at 2% (Figure 4L), indicating that these TEs were most active around 14 MYA and since then its activity declined. Interestingly, the MSY2 peak at 2% LTR divergence is higher than the rest of the genome, indicating preferential accumulation of LTR retrotransposons in the non-recombining region. Assuming that accumulation of LTR-TE in this region was caused by recombination cessation, it is possible to infer that MSY2 is at least 14 MY old. The other types of TEs are underrepresented in the MSY2 (Figure 4B) possibly because they were inactive in the

time since this region evolved recombination cessation and accumulation of LTR retrotransposons 'diluted' their abundance. This interplay of TE activity/inactivity and the evolution of recombination cessation in the Y-SLR can also explain no excess of LTR retrotransposons in the MSY1 and XhR if these regions evolved sex linkage more recently than the time when LTRs were actively transposing.

Y-chromosome degeneration

Y-SLR contains 32 annotated protein coding genes, 8 of which have annotated homologs in X-SLR. Synonymous site divergence (K_s) for the pairs of homologous X- and Y-linked genes is low (Figure 3K) suggesting recent cessation of recombination between X- and Y-linked gametologs due to inclusion of these genes in the sex-linked region. Y-degeneration can lead to complete or partial gene loss from the Y-chromosome. However, all 8 X-SLR genes have intact Y-linked homologs. Gene loss and accumulation of repetitive DNA typically lead to lower gene density on the Y-chromosome compared to the X and non-sex-linked regions. Consistent with this, gene density in X-SLR is not significantly different from genome-wide average of ~31 genes/Mb, while gene density of the Y-SLR is much lower (Figure 3L), likely due to gene loss and accumulation of TEs.

Sequence analysis identified 91 pseudogenes in the Y-SLR, totalling 34,746 nucleotides in length. Half (46 out of 91) of the pseudogenes had homology to Y-SLR genes with a mean sequence identity of 91.6%. These pseudogenes represent partial duplicates of Y-SLR genes. The other half were homologs of annotated genes elsewhere in the genome, with a mean sequence identity of 77.7%. No Y-SLR pseudogenes homologous to X-SLR genes were detected. The pseudogene density in the Y-SLR (30.3 pseudogenes per Mb, with total length 12kb) was three times higher than observed outside Y-SLR on the same chromosome (10.5 pseudogenes per Mb, with total length 3.5kb). Most pseudogenes (78) were located in MSY2 (31.2 pseudogenes per Mb), while only 2 pseudogenes were found in the MSY1 (7.6 pseudogenes per Mb). This pattern is consistent with a longer period of recombination suppression and accumulation of pseudogenes in MSY2, suggesting that MSY2 represents an older sex-linked region.

The comparison of synonymous / non-synonymous rate ratios (K_a/K_s) did not reveal significant difference between the X- and Y-linked genes for which orthologs from outgroup *M. yunnanensis* were available. The analysis of gene expression revealed that median expression of Y-linked genes is somewhat lower compared to X-linked and pseudoautosomal genes (Figure 5A), but this difference is not significant. No expression was detected in 8 out of 32 annotated Y-linked genes, while all X-linked genes were expressed. The proportion of unexpressed genes in MSY is only marginally ($P=0.076$, 2×2 $\chi^2=3.1$) significantly higher compared to the proportion of unexpressed pseudoautosomal genes (183 out of 1129). The Y-linked genes lacking expression are located in MSY1 and MSY2 (3 and 5 genes, respectively), while for the 8 genes with X- and Y-linked gametologs expression was almost identical (Figure 5B), revealing no evidence for Y-degeneration in gene expression. Thus, the only evidence for Y-degeneration we found in nutmeg is the accumulation of TE and pseudogenes, which likely contributed to increase of Y-SLR size.

Development of a sex-specific PCR marker for sexing nutmeg seedlings

To provide a reliable marker to sex nutmeg seedlings in agriculture, we developed a sex-specific PCR marker. The availability of molecular markers to reliably distinguish male from female seedlings is important in nutmeg production, as it can reduce costs related to growing unproductive individuals (i.e. males). To develop a sex-linked marker, we tested 26 primer pairs (Suppl. Figure S5, Tables S7 for primer

sequences and S8 for PCR protocol) and found one pair that reliably distinguished males from females (Suppl. Figure S5). Primer pair Mf_7216M2 amplifies ~0.6kb long fragment of Y-SLR gene g7232 (located at position 29.489Mb in hap1) in males and not in females. It is worth mentioning that this marker also amplifies in monoecious individuals (marked “H” in Suppl. Figure S5) that occasionally appear in nutmegs. This observation suggests that monoecious nutmeg plants may carry Y-linked sequences and could represent either unstable males or individuals with a modified Y chromosome that has reduced ability to suppress the development of female reproductive organs, similar to the Yh chromosome in papaya⁵⁶.

Discussion

In this study, we present the first assembly and analysis of *Myristica fragrans* 0.6 Gb long genome, which is on the lower side of the range of genome sizes in Spermatophytes and is within the genome size range of other magnoliid species⁵⁷. The GC-content, gene and TE density in the nutmeg genome is more evenly distributed compared to most other sequenced plant genomes that typically have low gene density in fairly large central (usually pericentromeric) part of the chromosomes. The even distribution in nutmeg likely reflects the lack of a single localised centromere and pericentromeric region in the holocentric chromosomes. Holocentric chromosomes, characterised by a diffuse centromere, are uncommon in eukaryotes³⁶, and the impact of holocentricity on the evolution of sex chromosomes is not well understood.

The newly sequenced nutmeg genome was used to locate and characterise the SLR in *M. fragrans*. Our finding that this species has a male-heterogametic (XY) reproductive system contradicts the conclusion of the previous RAPD-based study that reported the presence of female heterogamety in *M. fragrans*³⁴. However, RAPD markers can be unreliable, while our conclusion is based on several complimentary approaches (F_{st} , D_a , presence of male-specific SNPs, GWAS and sequence coverage) that all consistently support male heterogamety and the location of SLR close to one end of chromosome 6. The Y-SLR is ~3Mb long, located between positions 26.47 Mb to 29.57 Mb on haplotype 1 of the chromosome 6. The X-SLR is much shorter – only 0.23 Mb and located between positions 26.49 Mb and 26.72 Mb on haplotype 2 of the chromosome 6. This difference in size between the Y- and X-linked SLR is due to presence of a Y-specific region. It seems plausible that deletion of this region from the nascent X chromosome may have played a key role in evolution of sex chromosomes in nutmeg, though we cannot rule out an alternative scenario that the absence of this region is ancestral and it evolved on the Y-chromosome. As dioecy is ancestral in this plant family, there are no sufficiently close outgroup species to provide information about the likely ancestral state and distinguish these scenarios of MSY evolution.

The genome analysis of the closely related dioecious *M. yunnanensis* revealed considerable synteny preservation across many chromosomes, but no macrosynteny was found for the nutmeg sex chromosome (chr6) that has small homologous regions scattered across many *M. yunnanensis* chromosomes (Figure 2C). The nutmeg Y-specific region was not found in *M. yunnanensis*, suggesting that the sequenced genome belongs to a female (sex of the sequenced individual was not reported in⁵⁴). Interestingly, the homologs of nutmeg sex-linked genes in *M. yunnanensis* are equidistant from X- and Y-linked gametologs (Figure 3M) suggesting that at least the genes in Xh1 region, and possibly the entire nutmeg SLR evolved sex-linkage after the divergence between *M. fragrans* and *M. yunnanensis*. If

the nutmeg SLR evolved since the divergence from *M. yunnanensis*, then there must have been a recent sex chromosome turnover and sex-linked region in *M. yunnanensis* would be expected to be located in a different region of the genome. Lack of synteny for the nutmeg chromosome 6 with *M. yunnanensis* chromosomes is consistent recent sex chromosome turnover. Such turnovers are common in dioecious plants^{3,13}.

However, the high density of repeat elements, pseudogenes and the lower density of genes than genomic average on the Y chromosome (Figure 4) suggest that the SLR on chromosome 6 is at least a few million years old. X:Y divergence for the SLR genes with shared X- and Y-linked homologs is quite low ($K_s < 0.025$, Figure 3K), suggesting that nutmeg sex chromosomes are of recent origin. However, these genes may be located in the region recently added to the sex-linked region (i.e. it is a recent stratum), while the MSY2 may be older, as suggested by the prevalence of dioecy in the entire family Myristicaceae. The absence of X-linked homologs for the Y-linked genes in the MSY region does not allow us to estimate the age of nutmeg sex chromosomes more accurately. Identification of SLR in more Myristicaceae species will reveal how long ago the sex chromosomes have evolved and whether they have undergone any turnovers in this family.

Although other groups of land plants, such as bryophytes, have ancient sex chromosomes⁵⁸, most dioecious angiosperm species have SLRs of a fairly recent origin (e.g. *Amborella trichopoda*⁵⁹), wherein the SLR is relatively small with little to no accumulation of repetitive sequences. However, there are a few exceptions, such as *Silene latifolia*, where the X and Y sex chromosomes are heteromorphic and the oldest sex-linked region is estimated to be 11 Ma^{26,55}. We could not definitively identify the sex-determining genes in *M. fragrans*. However, the two F-box genes in SLR may have a potential role in floral development, a feature that is crucial in dioecy. To our knowledge, this is the first time that this group of proteins has been found in the SLR, although four MADS-box genes have been identified in *Ginkgo biloba*⁶⁰. MADS-box proteins are another large domain of proteins that has a role in floral development. Most of the other genes annotated in the SLR have an unknown function (Suppl. Table S7).

We identified 10 and 13 male-specific genes in *M. fragrans* MSY1 and MSY2 regions, respectively. While being male-specific, most of these genes have divergent non-sex-linked paralogs elsewhere in the genome, which created difficulties in developing male-specific primers. The Y-linked gene g7232 that was used for primer development is located at the border between the MSY2 and Xh2 and it has incomplete broken copy in the X-SLR that is sufficiently divergent from g7232 to allow us design male-specific primers, which will likely be useful for nutmeg growers to sex the seedlings in order to optimise the sex ratio in nutmeg plantations. It is worth mentioning that the previously published supposedly sex-specific markers from Shibu et al.³⁴ were not sex specific in our samples (Suppl. Figure S6).

To our knowledge, this is the first species within the whole Myristicaceae family for which the sex chromosomes have been identified and characterised. Considering that the predominant reproductive system within the family is dioecy, it could be presumed that this is the ancestral state of the family, with monoecy being the derived state. However, taxonomic descriptions of several species are inconsistent with field observations, calling for a revaluation of the reproductive systems within the family. For example, *M. insipida*, *M. fragrans*, and *Iryanthera hostmannii* are considered dioecious, but have reports of monoecious individuals⁶¹⁻⁶³. In *Myristica fragrans*, occasional reports of individuals bearing both male and female flowers are more plausibly explained as instances of *leaky dioecy* rather

than evidence for stable monoecy. Similar leakiness has been increasingly documented in other dioecious angiosperms with sex chromosomes⁶⁴, including the well-studied case of *Mercurialis annua*, where environmental factors such as pollen limitation can induce opposite-sex flower production⁶⁵. In nutmeg, such plasticity may function as a reproductive assurance mechanism under conditions of mate or pollen scarcity, while the species retains an underlying dioecious sexual system.

Concluding remarks

We produced the first chromosome-level haplotype resolved assembly of a male nutmeg (*M. fragrans*) and identified chromosome 6 as the sex chromosome, with a 3 Mb long sex-linked region. While the age of nutmeg SLR cannot be established, it is clear that at least part of this region, including X-homologous region, evolved sex-linkage recently (since nutmeg divergence from closely related *M. yunnanensis*). The rest of SLR is probably older and may have evolved as a result of deletion(s) on the nascent X-chromosome, making most of SLR male-specific, which should prevent recombination on the Y-chromosome. It is tempting to speculate that deletion(s) of this region on the X-chromosome may have been the cause of initial recombination suppression between the proto-X and proto-Y chromosomes. Identification and characterisation of SLR in other nutmeg species will help to shed light on the origin and evolutionary dynamics of sex chromosomes in this primarily dioecious understudied plant family.

Methods

Genome size estimation

Genome size was estimated with flow cytometry using nutmeg young leaves, with tomato (*Solanum lycopersicum*, genome size $\approx$ 900 Mb) and maize, inbred line B73 (*Zea mays*, genome size $\approx$ 2.3 Gb) as internal standards. Fresh tissue was chopped in ice-cold Mg^{2+} buffer, filtered through a 40- μ m nylon mesh, and nuclei were stained with propidium iodide (50 μ g/mL) in the presence of RNase A. Stained nuclei were analyzed on a BD FACScalibur flow cytometer with 488-nm excitation, collecting at least 10,000 nuclei per sample, and data were processed with ModFit LT v3.0. Genome size was calculated from the ratio of mean fluorescence intensities of the sample and the internal standard, multiplied by the known genome size of the standard.

DNA extraction and sequencing for reference genome assembly

The male individual used to produce the reference genome in this study was grown at the Xinglong Tropical Botanical Garden (Wanning, Hainan Province, PR of China). High molecular weight DNA was isolated from fresh young leaves of a single male plant using the CTAB method: the sample was ground in liquid nitrogen, incubated with CTAB buffer, and extracted with chloroform/isoamyl alcohol (24:1). The DNA was then precipitated with isopropanol and sodium acetate, washed with 75% ethanol, air-dried, dissolved in TE buffer, and stored at -20°C . Genomic DNA quality was assessed using 1% agarose electrophoresis. Genomic DNA was sheared using g-Tubes (Covaris), and concentrated with AMPure PB magnetic beads. Then, the DNA was quantified using the Qubit TM3 Fluorometer (Invitrogen). The SMRT bell library was constructed using the SMRTbell® Express Template Prep Kit 2.0, followed by primer annealing and binding of SMRT bell templates to polymerases with the DNA/Polymerase Binding Kit. Sequencing was performed on the Pacific Bioscience Sequel II platform (AnnoRoad Gene Technology) to generate HiFi reads. For Hi-C library construction, young leaves were

cross-linked with 2% formaldehyde at ambient temperature in a hybrid rotary hearth furnace for 90 min. Cells were lysed and digested using the Hind III restriction enzyme. DNA ends were filled and labelled with biotin. Hi-C libraries were created using the Mate-pair Kit. Complexes containing the biotin-labelled ligation products were purified and sheared to 300–700 bp, followed by pulling down. Paired-end reads (2×150 bp) were sequenced on an Illumina HiSeq X Ten platform, yielding a total of 66,928,937,824 bp (Suppl. Table S1). The resulting HiFi and Hi-C data are available from European Nucleotide Archive under project number PRJEB112374.

Genome assembly and annotation

We produced a haplotype-resolved male genome assembly using Hifiasm⁶⁶ (version 0.16.1, with parameter `-s=0.2`) in combination with Hi-C reads, while the Hi-C reads were assembled with HiC-Pro⁶⁷ (version 3.1.0, with default parameters). Duplicate reads were removed with Khaper⁶⁸ (the program developer did not provide version information), and reads were polished with NextPolish2⁶⁹ (version 0.2.1, with default parameters). Finally, EndHiC⁷⁰ (version 1.0.0, with parameter `--rounds=3`, `--binnumstep=5`) was used to anchor the chromosomes.

Transposable elements were identified and quantified using the default parameters of EDTA v2⁷¹ with the parameter `--cds`, that allows the pipeline to exclude coding regions from the TE library. Structural gene prediction and annotation was performed using BRAKER3 v3.0.8⁴⁶ by combining the RNA sequences, the soft-masked genome produced by EDTA, and the Viridiplantae protein dataset. Functional annotation was performed using emapper v2⁷² and OrthoDB⁷³. Pseudogenes were identified using PseudogenePipeline⁷⁴.

Genome resequencing

Short read genome sequencing was done for 18 individuals of known sex: seven males and eleven females. Samples for 16 individuals were kindly provided by the Singapore Botanic Garden (Republic of Singapore) and two individuals were obtained from the Eden Project (Cornwall, United Kingdom) (Suppl. Table S5). Total DNA was extracted from silica-dried leaves following the MATAB protocol as described in⁷⁵. DNA was quantified using Qubit 2.0 (Invitrogen). Genomic libraries were constructed at Azenta Life Sciences (Germany), and short-read (150 bp) sequencing was performed on an Illumina NovaSeq platform at the same facility with at least 30x coverage. The sequence data for 11 females and 7 males are available from NCBI SRA under project number PRJNA1301071.

RNA extraction and sequencing

Total RNA was isolated from fresh leaves from the same male individual that was used for sequencing the reference genome, using the TRIzol reagent (Thermo Fisher Scientific) following the manufacturer's protocol. To account for any variability in environmental conditions, we sampled and sequenced three replicates per tissue, with sampling done at different days. The concentration of extracted nucleic acids was detected with the Nanodrop2000 Spectrophotometer (Thermo Fisher Scientific). RNA concentration and integrity were quantified with the Agilent 2100 RNA nano 6000 assay kit (Agilent Technologies). RNA-seq library construction was performed from extracted samples and libraries were sequenced on the Illumina NovaSeq platform with the NovaSeq 6000 S4 Reagent Kit. The analysis of gene expression was conducted with RSEM⁷⁶, using gene sequences from the annotated genome as the reference for RNA-seq read mapping.

Bioinformatic processing

Raw reads were trimmed and filtered with fastp v0.23.4⁷⁷, and then mapped to haplotypes 1 (hereafter hap1) and 2 (hereafter hap2) of the male reference using bowtie2 v2.5.4⁷⁸. Mapped reads were then sorted, indexed, and piled up using the *sort*, *index*, and *mpileup* commands of samtools v1.7⁷⁹. Variant calling was performed with vcftools v0.1.16⁸⁰ and initial filtering was done with bcftools v1.21⁸¹. After the quality of the data was visually inspected, vcftools was used to filter out indels, multiallelic sites, sites with more than 80% missing data, sites with coverage less than 5, and a minor allele frequency of at least 0.1.

Identification of candidate sex-linked region

As there is no prior knowledge on the size and age of the SDR, we used several approaches for SDR identification. First, we explored the datasets by performing a coverage analysis on the mapped reads using mosdepth v0.3.8⁸² with 250 kb sliding windows. All polymorphism analyses, including calculation of D_a and F_{st} between males and females were done in proSeq4⁸³. In addition, we performed a genome-wide association study (GWAS) with plink v1.9⁸⁴, where we defined sex as the phenotypic trait and used the *-assoc* parameter for a basic association study. Finally, we identified sex-specific SNPs (i.e. with one of the alleles present in one but not the other sex) using filtering of SNPs in Microsoft Excel.

The comparison with congeneric genome

Synteny between *M. fragrans* and *M. yunnanensis* genomes was analysed with MCScanX⁸⁵ and two way nucleotide blast⁸⁶ for annotated coding regions. The sequences for coding regions from 1-to-1 orthologs of the two species were aligned in ProSeq4⁸³. The same program was used to check the alignments and calculate pairwise synonymous (K_s) and nonsynonymous (K_a) divergence between the species.

Phylogenetic analyses

Phylogenetic analysis for concatenated coding regions of genes with available X-, Y-linked and outgroup *M. yunnanensis* orthologs was conducted in MEGA⁸⁷. Synonymous site divergence was estimated using Nei-Gojobori method with Jukes-Cantor distances for synonymous sites. The phylogeny-based comparison of substitution rates in the X- and Y-linked genes with outgroup sequence available were conducted in codeml program from PAML package⁸⁸. Omega (K_a/K_s) ratio was estimated in branch model allowing for separate rates for X- and Y-linked genes. For testing significance of K_a/K_s ratios between the branches we used likelihood ratio test comparing log-likelihoods for two branch models, one that allows different rates for X- and Y-linked genes and the model allowing these branches to have the same rate.

Development of sex-specific PCR markers

We developed sex-specific primers to validate the identified SLR. More specifically, we checked which genes within the SLR were present in one sex, but not the other. Then, we chose one gene of interest and used blast⁸⁶ to query the gene sequences against the genome to verify that the sequence was specific to only that region. If the query had a high similarity to another region in the genome, we developed primers so that at least one of the last five bases on the 3' end is different in the other region. Theoretically, this should ensure a high specificity of the primers. Additionally, we tested previously developed sex-specific primers for *Myristica fragrans*³⁴. We first tested each developed

primer on two males and two females. The primers that were highly sex-specific were then used on all available samples. The PCR conditions are listed in Suppl. Table S8. Each 50 μ L PCR reaction contained approximately 14–15 ng of genomic DNA and 29–40 ng of each primer (equivalent to 2.5 pmol per primer, depending on primer length).

Acknowledgements

We thank the staff from the institutions which kindly provided us the herbarium and fresh plant material, namely Junhao Chen at the Singapore Botanic Garden (SING), Jianhe Wei at the Peking Union Medical College (Beijing, China) and Rosie Chinn at the Eden Project (Cornwall, UK). We also thank Helen Prescott for help with the lab work.

Funding

KM was supported by the *Fondation Wiener-Anspach* (FWA, Belgium), SW was supported by the China Scholarship Council. DAF acknowledges funding from the BBSRC grant BB/Y009169/1. RM acknowledges funding from Fujian Agriculture and Forestry University.

Author contributions

KM and DAF conceived the study, planned the work, conducted the analyses and wrote the paper. SW and MX assembled and annotated the genome and contributed to the analysis. MX, RM and KM collected the samples and generated sequence data. KM and ADM contributed to development of sex-specific molecular markers. DAF and RM oversaw the work in the project.

Data availability

Genome and transcriptome sequence data generated in this project are available from NCBI (<https://www.ncbi.nlm.nih.gov/>) under project number PRJNA1301071. The genome assembly data are available from European Nucleotide Archive (<https://www.ebi.ac.uk/ena/browser/home>) under project number PRJEB112374. The genome assemblies are also available in NGDC Genome Warehouse repository (<https://ngdc.cncb.ac.cn/gwh/>) under project number PRJCA063189.

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1. Circos plot for male genome assembly of *Myristica fragrans*. Rings indicate (A) percentage of GC content, (B) gene density and (C) TE density, all calculated at 1 Mb intervals. Hi-C heatmap of male genome assembly is shown in the centre of the circos plot.

Figure 2. Synteny of nutmeg chromosomes with congeneric *M. yunnanensis* genome. A) Synteny between the two species is preserved for many chromosomes, but not for the nutmeg chromosome 6. B) Genome-wide distribution of synonymous (K_s) and non-synonymous (K_a) divergence between annotated nutmeg genes and their 1-to-1 homologs in *M. yunnanensis* genome. C) Nutmeg chromosome 6 is fragmented in *M. yunnanensis* genome. D) The synteny and positions for five nutmeg sex-linked genes (from Xh1 region, see Figure 3E) is shown with red lines.

Figure 3. Localisation of the sex-linked region on chromosome 6 of the nutmeg genome. Panels A to E use hap1 (Y) as the reference. Panels G to J use hap2 (X) as the reference. A) Genome-wide association study using sex as the phenotypic trait. B) and L) Short read sequence coverage in 7 males and 11 females using 250kb long bins. C) and K) F_{st} between males and females with 100kb long bins. D) and J) D_a between males and females with 100kb long bins. E) and I) The number of male-specific SNPs per 10kb bin. F) and H) diagrams of SLRy and SLRx locations. MSY1 and MSY2 represent male-specific Y-linked regions that are absent from the X-chromosome, while Xh1 and Xh2 represent Y-linked regions of homology with the X-chromosome. F) Green lines connect homologous genes on the two haplotypes in the proximity of the SLR. Red lines show duplications from SLRy to PAR1 that likely creates spurious signal of sex-linkage at position 25.5 of the hap2. Small black circles on hap1 and hap2 lines show locations of annotated genes. J) The number of annotated genes per 100kb and K) Synonymous and non-synonymous sequence divergence between X- and Y-linked gametologs in Xh1 and Xh2 regions; L) the percent of sequence length that corresponds to annotated genes in PAR1, MSY1, XhR, MSY2, PAR2, SLR-X and SLR-Y (combined MSY1, XhR and MSY2). M) NJ phylogeny of nutmeg Y-SLR and X-SLR and outgroup *M. yunnanensis* based on concatenation of CDSs of SLR-Xh1 genes (Figure 2D) for which Y-, X-linked and outgroup sequences were available.

Figure 4. Distribution of TEs in and around SLR on nutmeg chromosome 6. Panels A to C and G to J represent hap1 (Y), while panels E and F represent hap2 (X). Panel A shows the entire length of hap1, while panel B shows only the region distal to position 22Mb of chr6 with higher resolution. In panels A, B and F the left axis shows the number of LTR retrotransposons or TIR transposons per 100 kb, while the right axis shows the number of LINEs and SINEs per 500kb. The dotted, dashed and solid lines on panels A, B and F represent 5 data point moving average, except the LINEs and SINEs in panels A and B where the lines connect the adjacent data points. C) and E) show diagrams of SLRy and SLRx locations. MSY1 and MSY2 represent male-specific Y-linked regions that are absent from the X-chromosome, while Xh1 and Xh2 represent Y-linked regions of homology with the X-chromosome. D) Green lines connect homologous genes on the two haplotypes in the proximity of the SLR. Panels G to J compare the density of different types of TEs per 100kb between the PAR1, MSY1, MSY2 and PAR2. MSY2 is significantly (χ^2 , $P < 0.01$) different from other regions on all four plots. The error bars above and below the boxes show the range of values after outlier exclusion, while the box shows the first and third quartiles as well as the median, as standard in boxplots in R. K) Percent of the genome represented with different types of TEs plotted against total sequence divergence between the copies of each element type. Note the predominance of LTR retrotransposons with low sequence divergence at the left of the plot. L) The

distribution of sequence divergence between LTRs of the same copy of the element (proxy for age of insertion) in the MSY2, PAR1 and genome wide. Note that MSY1, XhR and PAR2 are not shown due to too few TE copies with both LTRs present.

Figure 5. Expression of sex-linked and pseudoautosomal genes. A) The comparison of gene expression ($\log_2(\text{FPKM})$) in the pseudoautosomal regions and the X- and Y-linked genes. The difference between these groups of genes is non-significant. The boxplot shows the actual data points as circles, the error bars above and below the boxes show the range of values after outlier exclusion, while the box shows the interquartile range and median. B) The comparison of gene expression between Y-linked genes and their X-linked gametologs. Gene expression is almost identical in the Y-linked genes and their X-linked gametologs.

Tables.**Table 1.** Flow cytometry-based estimation of the genome size of *M. fragrans*

Nutmeg sample	Internal reference	Fluorescence intensity of internal reference	Fluorescence intensity of the nutmeg sample	Ratio	Estimate for nutmeg genome (Gb)
Male-1	Tomato	25.43	16.55	0.65	0.59
Male-1	maize	70.27	17.48	0.25	0.57
Male-2	Tomato	24.96	16.31	0.65	0.59
Male-3	Tomato	24.79	16.32	0.66	0.59

Table 2. Summary statistics of the assembled male genome.

	Primary	Haplotype 1	Haplotype 2
Scaffold length (bp)	640,827,773	633,402,135	627,476,598
Number of scaffolds	109	157	90
GC content (%)	37%	37%	37%
N50 (scaffolds)	27,097,807	27,088,625	26,872,488
Pseudochromosomes	22	22	22
Anchor rate (%)	97.61%	95.26%	95.48%
Number of gaps	5	28	22
BUSCO score	97.3%	97.3%	97.3%

References

- 1 Vicoso, B. Molecular and evolutionary dynamics of animal sex-chromosome turnover. *Nat Ecol Evol* **3**, 1632-1641 (2019). <https://doi.org/10.1038/s41559-019-1050-8>
- 2 Renner, S. S. The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. *Am J Bot* **101**, 1588-1596 (2014). <https://doi.org/10.3732/ajb.1400196>
- 3 Wang, Y. *et al.* Evolution of sex-linked genes and the role of pericentromeric regions in sex chromosomes: insights from diploid willows. *Mol Biol Evol* **41** (2024). <https://doi.org/10.1093/molbev/msae235>
- 4 Charlesworth, D. & Harkess, A. Why should we study plant sex chromosomes? *Plant Cell* (2024). <https://doi.org/10.1093/plcell/koad278>
- 5 Garcia, S., Janousek, B., Pascual-Diaz, J. P. & Renner, S. S. Sex-chrom v. 2.0: a database of green plant species with sex chromosomes. *Chromosoma* **132**, 55-58 (2023). <https://doi.org/10.1007/s00412-023-00786-7>
- 6 Blackburn, K. B. Sex chromosomes in plants. *Nature* **112**, 687-688 (1923). <https://doi.org/doi:10.1038/112687c0>
- 7 Navajas-Perez, R. *et al.* The evolution of reproductive systems and sex-determining mechanisms within *Rumex* (Polygonaceae) inferred from nuclear and chloroplastidial sequence data. *Mol Biol Evol* **29**, 1929-1939 (2005). <https://doi.org/10.1093/molbev/msi186>
- 8 Filatov, D. A. Homomorphic plant sex chromosomes are coming of age. *Mol Ecol* **24**, 3217-3219 (2015). <https://doi.org/10.1111/mec.13268>
- 9 Muyle, A., Shearn, R. & Marais, G. A. The evolution of sex chromosomes and dosage compensation in plants. *Genome Biol Evol* **9**, 627-645 (2017). <https://doi.org/10.1093/gbe/evw282>
- 10 Gong, W. & Filatov, D. A. Evolution of the sex-determining region in *Ginkgo biloba*. *Philos Trans R Soc Lond B Biol Sci* **377**, 20210229 (2022). <https://doi.org/10.1098/rstb.2021.0229>
- 11 Kafer, J., Marais, G. A. & Pannell, J. R. On the rarity of dioecy in flowering plants. *Mol Ecol* **26**, 1225-1241 (2017). <https://doi.org/10.1111/mec.14020>
- 12 Renner, S. S. & Müller, N. A. Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nature Plants* (2021). <https://doi.org/10.1038/s41477-021-00884-3>
- 13 Wang, Y. *et al.* Gap-free X and Y chromosome assemblies of *Salix arbutifolia* reveal an evolutionary change from male to female heterogamety in willows, without a change in the position of the sex-determining locus. *New Phytol* **242**, 2872-2887 (2024). <https://doi.org/10.1111/nph.19744>
- 14 Westergaard, M. The mechanism of sex determination in dioecious flowering plants. *Adv. Genet.* **9**, 217-281 (1958).
- 15 Heilbut, T. C. Lower species richness in dioecious clades. *The American Naturalist* **156**, 221-241 (2000).
- 16 Janousek, B. & Mrackova, M. Sex chromosomes and sex determination pathway dynamics in plant and animal models. *Biol J Linn Soc* **100**, 737-752 (2010).
- 17 Charlesworth, D. Young sex chromosomes in plants and animals. *New Phytol* **224**, 1095-1107 (2019). <https://doi.org/10.1111/nph.16002>
- 18 Charlesworth, B. & Charlesworth, D. Model for evolution of dioecy and gynodioecy. *American Naturalist* **112**, 975-997 (1978). <https://doi.org/Doi 10.1086/283342>
- 19 Harkess, A. *et al.* Sex determination by two Y-linked genes in garden asparagus. *Plant Cell* (2020). <https://doi.org/10.1105/tpc.19.00859>

- 20 Renner, S. S. Pathways for making unisexual flowers and unisexual plants: Moving beyond the "two mutations linked on one chromosome" model. *American Journal of Botany* **103**, 587-589 (2016). <https://doi.org/10.3732/ajb.1600029>
- 21 Akagi, T., Henry, I. M., Tao, R. & Comai, L. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646-650 (2014). <https://doi.org/10.1126/science.1257225>
- 22 Boualem, A. *et al.* A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science* **350**, 688-691 (2015). <https://doi.org/10.1126/science.aac8370>
- 23 Charlesworth, D. Plant sex chromosomes. *Annu Rev Plant Biol* **67**, 397-420 (2016). <https://doi.org/10.1146/annurev-arplant-043015-111911>
- 24 Lahn, B. T. & Page, D. C. Four evolutionary strata on the human X chromosome. *Science* **286**, 964-967 (1999). <https://doi.org/10.1126/science.286.5441.964>
- 25 Charlesworth, D. Evolution of recombination rates between sex chromosomes. *Philos Trans R Soc Lond B Biol Sci* **372** (2017). <https://doi.org/10.1098/rstb.2016.0456>
- 26 Yue, J. *et al.* The origin and evolution of sex chromosomes, revealed by sequencing of the *Silene latifolia* female genome. *Current Biology* **33**, 2504–2514 (2023).
- 27 Filatov, D. A. Evolution of a plant sex chromosome driven by expanding pericentromeric recombination suppression. *Sci Rep-Uk* **14**, 1373 (2024). <https://doi.org/10.1038/s41598-024-51153-0>
- 28 Akagi, T. *et al.* Rapid and dynamic evolution of a giant Y chromosome in *Silene latifolia*. *Science* **387**, 637-643 (2025). <https://doi.org/10.1126/science.adk9074>
- 29 Filatov, D. A. Recent expansion of the non-recombining sex-linked region on *Silene latifolia* sex chromosomes. *J Evol Biol* **35**, 1696-1708 (2022). <https://doi.org/10.1111/jeb.14063>
- 30 Charlesworth, D. Why and how do Y chromosome stop recombining? *J Evol Biol* (2023). <https://doi.org/10.1111/jeb.14137>
- 31 Nair, R. R. Chromosome number analysis in different sex types and open-pollinated seedlings of nutmeg (*Myristica fragrans* Houtt.). *Journal of Plantation Crops* **47**, 197-201 (2019). <https://doi.org/10.25081/jpc.2019.v47.i3.6056>
- 32 Saji, K. V., Anandaraj, M., Aarthi, S., Rema, J. & Sasikumar, B. Diversity in floral characters of monoecious nutmeg (*Myristica fragrans* Houtt.). *Journal of Spices and Aromatic Crops* **27**, 66-70 (2018). <https://doi.org/10.25081/josac.2018.v27.i1.1018>
- 33 Ehlers, B. K. & Bataillon, T. 'Inconstant males' and the maintenance of labile sex expression in subdioecious plants. *New Phytol* **174**, 194-211 (2007). <https://doi.org/10.1111/j.1469-8137.2007.01975.x>
- 34 Shibu, M. P. *Identification of sex-specific DNA markers in the dioecious tree nutmeg (Myristica Fragrans Houtt.)*. (University of Agricultural Sciences, 1998).
- 35 Kuo, Y. T. *et al.* The holocentricity in the dioecious nutmeg (*Myristica fragrans*) is not based on major satellite repeats. *Chromosome Res* **32**, 8 (2024). <https://doi.org/10.1007/s10577-024-09751-1>
- 36 Senaratne, A. P., Cortes-Silva, N. & Drinnenberg, I. A. Evolution of holocentric chromosomes: drivers, diversity, and deterrents. *Semin Cell Dev Biol* **127**, 90-99 (2022). <https://doi.org/10.1016/j.semcdb.2022.01.003>
- 37 Cabral, G., Marques, A., Schubert, V., Pedrosa-Harand, A. & Schlogelhofer, P. Chiasmatic and achiasmatic inverted meiosis of plants with holocentric chromosomes. *Nat Commun* **5**, 5070 (2014). <https://doi.org/10.1038/ncomms6070>
- 38 Heckmann, S. *et al.* Alternative meiotic chromatid segregation in the holocentric plant *Luzula elegans*. *Nat Commun* **5**, 4979 (2014). <https://doi.org/10.1038/ncomms5979>

- 39 Ruckman, S. N., Jonika, M. M., Casola, C. & Blackmon, H. Chromosome number evolves at equal rates in holocentric and monocentric clades. *PLoS Genet* **16**, e1009076 (2020). <https://doi.org/10.1371/journal.pgen.1009076>
- 40 Lucek, K., Augustijnen, H. & Escudero, M. A holocentric twist to chromosomal speciation? *Trends Ecol Evol* **37**, 655-662 (2022). <https://doi.org/10.1016/j.tree.2022.04.002>
- 41 Brazier, T. & Glemin, S. Diversity and determinants of recombination landscapes in flowering plants. *PLoS Genet* **18**, e1010141 (2022). <https://doi.org/10.1371/journal.pgen.1010141>
- 42 Charlesworth, D. The timing of genetic degeneration of sex chromosomes. *Philos Trans R Soc Lond B Biol Sci* **376**, 20200093 (2021). <https://doi.org/10.1098/rstb.2020.0093>
- 43 Rifkin, J. L. *et al.* Recombination landscape dimorphism and sex chromosome evolution in the dioecious plant *Rumex hastatulus*. *Philos Trans R Soc Lond B Biol Sci* **377**, 20210226 (2022). <https://doi.org/10.1098/rstb.2021.0226>
- 44 Hofstatter, P. G., Thangavel, G., Castellani, M. & Marques, A. Meiosis progression and recombination in holocentric plants: what is known? *Front Plant Sci* **12**, 658296 (2021). <https://doi.org/10.3389/fpls.2021.658296>
- 45 Smarda, P. *et al.* Ecological and evolutionary significance of genomic GC content diversity in monocots. *Proc Natl Acad Sci U S A* **111**, E4096-4102 (2014). <https://doi.org/10.1073/pnas.1321152111>
- 46 Gabriel, L. *et al.* BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. *Genome Res* **34**, 769-777 (2024). <https://doi.org/10.1101/gr.278090.123>
- 47 Plomion, C. *et al.* Oak genome reveals facets of long lifespan. *Nat Plants* **4**, 440-452 (2018). <https://doi.org/10.1038/s41477-018-0172-3>
- 48 Jouffroy, O., Saha, S., Mueller, L., Quesneville, H. & Maumus, F. Comprehensive repeatome annotation reveals strong potential impact of repetitive elements on tomato ripening. *BMC Genomics* **17**, 624 (2016). <https://doi.org/10.1186/s12864-016-2980-z>
- 49 Fields, P. D., Weber, M. M., Waneka, G., Broz, A. K. & Sloan, D. B. Chromosome-level genome assembly for the angiosperm *Silene conica*. *Genome Biol Evol* **15** (2023). <https://doi.org/10.1093/gbe/evad192>
- 50 Gaut, B. S., Wright, S. I., Rizzon, C., Dvorak, J. & Anderson, L. K. Recombination: an underappreciated factor in the evolution of plant genomes. *Nat Rev Genet* **8**, 77-84 (2007). <https://doi.org/10.1038/nrg1970>
- 51 Rifkin, J. L. *et al.* Widespread recombination suppression facilitates plant sex chromosome evolution. *Mol Biol Evol* **38**, 1018-1030 (2021). <https://doi.org/10.1093/molbev/msaa271>
- 52 Mandrioli, M. & Manicardi, G. C. Holocentric chromosomes. *PLoS Genet* **16**, e1008918 (2020). <https://doi.org/10.1371/journal.pgen.1008918>
- 53 An, H. & Nam, K. Lepidopteran genomes have denser transposable elements in smaller chromosomes, likely driven by non-allelic homologous recombination. *Genome Biol Evol* **17** (2025). <https://doi.org/10.1093/gbe/evaf137>
- 54 Zhang, Y. *et al.* Chromosome-level genome assembly of *Myristica yunnanensis* sheds light on genomic limitations underlying its critically endangered status in China. *Plant Diversity* (2025). <https://doi.org/10.1016/j.pld.2025.08.004>
- 55 Krasovec, M., Chester, M., Ridout, K. & Filatov, D. A. The mutation rate and the age of the sex chromosomes in *Silene latifolia*. *Current Biology* **28**, 1832-1838 (2018). <https://doi.org/10.1016/j.cub.2018.04.069>
- 56 VanBuren, R. *et al.* Origin and domestication of papaya Yh chromosome. *Genome Res* **25**, 524-533 (2015). <https://doi.org/10.1101/gr.183905.114>

- 57 Pellicer, J., Hidalgo, O., Dodsworth, S. & Leitch, I. J. Genome size diversity and its impact on the
evolution of land plants. *Genes (Basel)* **9** (2018). <https://doi.org/10.3390/genes9020088>
- 58 Renner, S. S., Heinrichs, J. & Sousa, A. The sex chromosomes of bryophytes: Recent insights,
open questions, and reinvestigations of *Frullania dilatata* and *Plagiochila asplenioides*. *J Syst
Evol* **55**, 333-339 (2017). <https://doi.org/10.1111/jse.12266>
- 59 Carey, S. B. *et al.* ZW sex chromosome structure in *Amborella trichopoda*. *Nat Plants* **10**, 1944-
1954 (2024). <https://doi.org/10.1038/s41477-024-01858-x>
- 60 Liao, Q. *et al.* The genomic architecture of the sex-determining region and sex-related metabolic
variation in *Ginkgo biloba*. *Plant J* **104**, 1399-1409 (2020). <https://doi.org/10.1111/tpj.15009>
- 61 Armstrong, J. E. & Irvine, A. K. Flowering, sex ratios, pollen-ovule ratios, fruit set, and
reproductive effort of a dioecious tree, *Myristica insipida* (Myristicaceae), in two different rain
forest communities. *American Journal of Botany* **76**, 74-85 (1989).
- 62 Queenborough, S. A., Burslem, D. F., Garwood, N. C. & Valencia, R. Determinants of biased sex
ratios and inter-sex costs of reproduction in dioecious tropical forest trees. *Am J Bot* **94**, 67-78
(2007). <https://doi.org/10.3732/ajb.94.1.67>
- 63 Sharma, M. V. & Armstrong, J. E. Pollination of *Myristica* and other nutmegs in natural
populations. *Tropical Conservation Science* **6**, 595-607 (2013).
- 64 Kral, J. *et al.* Insights into the karyotype and genome evolution of haplogyne spiders indicate a
polyploid origin of lineage with holokinetic chromosomes. *Sci Rep* **9**, 3001 (2019).
<https://doi.org/10.1038/s41598-019-39034-3>
- 65 Cossard, G. G. & Pannell, J. R. A functional decomposition of sex inconstancy in the dioecious,
colonizing plant *Mercurialis annua*. *Am J Bot* **106**, 722-732 (2019).
<https://doi.org/10.1002/ajb2.1277>
- 66 Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly
using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170-175 (2021).
<https://doi.org/10.1038/s41592-020-01056-5>
- 67 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome
Biol* **16**, 259 (2015). <https://doi.org/10.1186/s13059-015-0831-x>
- 68 Zhang, X. *et al.* Haplotype-resolved genome assembly provides insights into evolutionary history
of the tea plant *Camellia sinensis*. *Nat Genet* **53**, 1250-1259 (2021).
<https://doi.org/10.1038/s41588-021-00895-y>
- 69 Hu, J. *et al.* NextPolish2: A Repeat-aware Polishing Tool for Genomes Assembled Using HiFi Long
Reads. *Genomics Proteomics Bioinformatics* **22** (2024). <https://doi.org/10.1093/gpbjnl/qzad009>
- 70 Wang, S. *et al.* EndHiC: assemble large contigs into chromosome-level scaffolds using the Hi-C
links from contig ends. *BMC Bioinformatics* **23**, 528 (2022). <https://doi.org/10.1186/s12859-022-05087-x>
- 71 Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a
streamlined, comprehensive pipeline. *Genome Biol* **20**, 275 (2019).
<https://doi.org/10.1186/s13059-019-1905-y>
- 72 Cantalapiedra, C. P., Hernandez-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-
mapper v2: Functional annotation, orthology assignments, and domain prediction at the
metagenomic scale. *Mol Biol Evol* **38**, 5825-5829 (2021).
<https://doi.org/10.1093/molbev/msab293>
- 73 Kuznetsov, D. *et al.* OrthoDB v11: annotation of orthologs in the widest sampling of organismal
diversity. *Nucleic Acids Res* **51**, D445-D451 (2023). <https://doi.org/10.1093/nar/gkac998>
- 74 Zou, C. *et al.* Evolutionary and expression signatures of pseudogenes in Arabidopsis and rice.
Plant Physiol **151**, 3-15 (2009). <https://doi.org/10.1104/pp.109.140632>

- 75 Kusuma, J. *et al.* Microsatellite markers development for Indonesian nutmeg (*Myristica fragrans* Houtt.) and transferability to other Myristicaceae spp. *Mol Biol Rep* **47**, 4835-4840 (2020). <https://doi.org/10.1007/s11033-020-05535-y>
- 76 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). <https://doi.org/10.1186/1471-2105-12-323>
- 77 Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *Imeta* **2**, e107 (2023). <https://doi.org/10.1002/imt2.107>
- 78 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012). <https://doi.org/10.1038/nmeth.1923>
- 79 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009). <https://doi.org/10.1093/bioinformatics/btp352>
- 80 Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011). <https://doi.org/10.1093/bioinformatics/btr330>
- 81 Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* **10** (2021). <https://doi.org/10.1093/gigascience/giab008>
- 82 Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867-868 (2018). <https://doi.org/10.1093/bioinformatics/btx699>
- 83 Filatov, D. A. ProSeq4: A user-friendly multiplatform program for preparation and analysis of large-scale DNA polymorphism datasets. *Mol Ecol Resour* **24**, e13962 (2024). <https://doi.org/10.1111/1755-0998.13962>
- 84 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-575 (2007). <https://doi.org/10.1086/519795>
- 85 Wang, Y. *et al.* MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012). <https://doi.org/10.1093/nar/gkr1293>
- 86 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009). <https://doi.org/10.1186/1471-2105-10-421>
- 87 Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol* **38**, 3022-3027 (2021). <https://doi.org/10.1093/molbev/msab120>
- 88 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591 (2007). <https://doi.org/10.1093/molbev/msm088>

Editorial Summary:

The analysis of genomic differences between male and female nutmeg plants identified the sex-determining region, revealed sex chromosome evolution in nutmegs and yielded a molecular marker to identify males and females in nutmeg seedlings.

Peer Review Information:

Communications Biology thanks Takashi Akagi, Zdenek Kubat, and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Jorge Duitama and Rupali Sathe.









