Title: Convergent evolution of sex chromosomes in palms

Authors: H. Tessarotto[1], T. Beulé[1], E. Cherif[1], J. Orjuela[1], A. Lindstrom[2], A. Lemansour[3], M. Dahme[4], S. Santoni[5], J. Käfer[6]*†, F. Aberlenc[1]*†

Affiliations:
[1] DIADE, Université de Montpellier, IRD, CIRAD; Montpellier, France.
[2] Global Biodiversity Conservancy, Thailand.
[3] United Arab Emirates University, UAE.
[4] Central Florida Palm and Cycad Society, USA.
[5] AGAP, Université Montpellier, CIRAD, INRAe, Montpellier SupAgro; Montpellier, France.
[6] ISEM, Université Montpellier, CNRS, IRD; Montpellier, France.
† These authors contributed equally to this work.
** For correspondence: frederique.aberlenc@ird.fr and jos.kafer@cnrs.fr*

Abstract: While sex chromosomes often evolve in species with separate sexes through similar processes, there are also many differences in their evolutionary dynamics. In palms, a clade with 8-10 independent evolutionary origins of dioecy, sex chromosomes have hitherto been well characterized only in the date palm genus *Phoenix*. Using genome-wide target capture sequencing, we found that the same region independently became sex-linked both in *Phoenix* and in the rather distantly related species *Kerriodoxa elegans*. Investigating the gene content of this region revealed it contains many genes involved in flower sex differentiation which might be the reason for this case of convergent evolution. This region, which could function as a "genetic toolbox" for the production of unisexual flowers or individuals, may have an ancestral functional importance in palm genomes.

Main Text:

Although anisogamy and separate sexes evolved many times independently in the eukaryotes, there are striking similarities as well as differences between the clades in which they occur. In animals, for example, hermaphroditism is rare but evolved frequently from gonochorism which most likely is the ancestral state for most phyla (*1*). Among gonochoric animals, diverse sex determination mechanisms exist among which environmental and chromosomal sex determination, all of which seem to act on a common downstream sex-determining pathway (*2*). In flowering plants, by contrast, separate sexes (dioecy) evolved many times independently and sex determination pathways might in each case be different (*3*). However, as the first sex-determining pathways are being unraveled, it appears that they also share common features, such as the involvement of plant hormones like cytokinins (*4*).

Sex determination often involves sex chromosomes, a pair of chromosomes for which recombination is suppressed in a region containing the sex-determining gene(s). Even while plant sex chromosomes share striking similarities with those of other groups, such as degeneration of the sex-limited chromosome (Y or W) due to the absence of recombination,

many differences remain unexplained. In particular, there is no obvious relation between the age of recombination suppression and the size (proportion of the genome, number of genes) of the non-recombining region, nor is it clear whether the onset of recombination suppression is triggered by one common mechanism (*5*). While only a handful of sex chromosomes were known before 2000 and their sparse distribution among the angiosperms limited the scope for comparative studies, their characterization has become more feasible in the last decades with the increasing availability of genomic data and tools.

Palms are a large mainly tropical plant family characterized by numerous transitions between sexual systems (hermaphroditism, monoecy and dioecy; (*6, 7*)). Sex chromosomes have been relatively well described in the genus *Phoenix* and in particular the date palm, *P. dactylifera* (*8–10*), and are of one common origin for the whole genus (*11*). XY sex chromosomes seem to exist in *Trachycarpus fortunei* and *Lodoicea maldivica* (*12, 13*). We here study another palm species that evolved dioecy independently, *Kerriodoxa elegans*, and identify and compare its sex chromosomes with those of *P. dactylifera*.

A common sex-linked region in *P. dactylifera* and *K. elegans* of independent origin

We developed gene capture kits based on the *P. dactylifera* genome and tested these in conjunction with our pipeline for the identification of sex-determining genes on this species before application to *K. elegans*. The capture kits "SexPhoenix" (SP) and "PalmExon" (PE) were respectively developed based on a set of known sex-linked and autosomal genes in *P. dactylifera* (*14*), and on the whole exome of this species (*15*). These kits were used for short-read sequencing (Illumina) in both species. For comparison and testing, we obtained whole-genome short-sequence data for *P. dactylifera* (bioproject PRJNA505141). All reads were mapped on the female DPV01 genome assembly (*15*) of *P. dactylifera* to facilitate the detection of XY polymorphisms (table S1,S2, S3), after which the individuals were genotyped and the coding regions were extracted for further analysis. Related or distant individuals were removed for downstream analysis in *P. dactylifera* but not in *K. elegans* which showed a rather homogeneous population structure (fig. S1, S2, table S1, S2). On the genotyped data, we ran the SDpop pipeline (*16*) to obtain posterior scores for sex linkage of each gene. Genes with high XY scores in at least one dataset were sorted into several categories based on the level of confidence: sex-linked, probably sex-linked, inconclusive, not sex-linked and missing data (Materials and Methods).

For *P. dactylifera*, our data and analyses highlighted the known region on chromosome 12 (Linkage Group 12 (*8*), NC_052403.1 (*10*)) as non-recombining (Fig. 1): between approximately 9.4 and 12.4 Mb, most genes were clearly inferred as sex-linked while none had an autosomal segregation pattern, in the data obtained using our capture kits as well as in the whole-genome resequencing data. All sex-linked genes were inferred to be XY; X-hemizygous genes were not found through a combination of the analyses of segregation patterns and female vs. male coverage (Materials and Methods; Supplementary Text 2; Data S2). Additional sex-linked genes were found outside this region, mainly on the unplaced scaffolds (Data S1), indicating that the non-recombining region might be larger than its size estimate based on the current assembly.

In *K. elegans*, analysis of the SP and PE capture kit data showed striking similarities with the results obtained in *P. dactylifera* despite the presumed independent origin of dioecy (Fig. 1).

The sex chromosomes of *K. elegans* were inferred to be of the XY type (table S4), and the sex-linked genes were concentrated in the same region as those of *P. dactylifera*. Indeed, most genes inferred as sex-linked in *P. dactylifera* (48 genes) and *K. elegans* (59 genes) were so in both species (24 genes), whereas only 2 high-confidence sex-linked genes in *P. dactylifera* and 10 in *K. elegans* were found not to be sex-linked in the other species (Figure 2, Data S1). In addition, among the genes which are sex-linked in one species and probably sex-linked in the other, 15 are probably sex-linked in *P. dactylifera* and 3 in *K. elegans*.

The currently available genome assembly of *P. dactylifera*, Barhee BC4, was obtained using a male (XY) individual (*10*), which has likely rendered the assembly of the non-recombining region more difficult than for the autosomes (*17*). As the sex-linked region in *P. dactylifera* shares synteny with a genomic region on the monoecious oil palm (*Elaeis guineensis*) chromosome 10 (*8, 9, 18, 19*), we identified the *E. guineensis* homologs of the genes found to be sex-linked. Indeed, 78% of genes inferred as sex-linked in either *P. dactylifera* or *K. elegans*, including 95% of the common sex-linked genes, were positioned in a single region between 43 and 46.1Mb of the *E. guineensis* chromosome 10 (GK000085.1, EGPMv6 assembly, Fig. 2). By contrast only 16% of the genes in the region are autosomal in *P. dactylifera* and none in *K. elegans* (Fig. 2, DataS1). Thus, a large majority of the XY genes in the two dioecious species were found to be positioned within a unique 3 Mb region on chromosome 10 of the *E. guineensis* genome (Fig. 2A), and conversely, the majority of the genes identified in this oil palm genomic region are sex-linked genes either in one or both dioecious species (Fig. 2B).

We inferred the gene trees for the X and Y copies of each sex-linked gene in the dioecious species using *E. guineensis* as an outgroup. In all trees, the X and Y sequences of each of the dioecious species formed a strongly supported clade (Fig. 3a, fig. S2), indicating that recombination suppression evolved after the lineages of *Phoenix* and *Kerriodoxa* split. We furthermore investigated divergence with *E. guineensis* homologs for the genes that were sex-linked in both species (using the SP data which cover a larger part of the genes in this region, table S3) as well as for ten autosomal genes. Cumulated synonymous divergence (dS) was similar among autosomal and sex-linked genes (fig. S4), indicating that the sex-linked genes experienced substitutions at similar rates as autosomal genes. Geneconv (*20*) identified possible gene conversion in 6 of the sex-linked genes in both species (table S5). Divergence times estimated when the stretches that could have undergone gene conversion were retained were slightly higher than when discarding them but this did not change the results qualitatively (fig. S5). For further analyses, these stretches were discarded.

We then estimated the node ages in these sex-linked gene trees using previously estimated divergence times at the level of the palms (*21*) as a calibration. All estimates confirmed that recombination suppression occurred much later than the split between the *Phoenix* and *Kerriodoxa* lineages (Fig 3b). The mean age of divergence between *P. dactylifera* and *K. elegans* is 61.86 My, whereas the mean ages of splits between X and Y haplotypes for the five most divergent genes of our datasets for each species were estimated to be 22.07 Myr for *P. dactylifera* and 19.32 Myr for *K. elegans* (Figure 3b). Thus, despite these regions being similar in gene content and age, none of the genes in the sex-linked region showed any signal for recombination suppression in the last common ancestor of *P. dactylifera* and *K. elegans*, confirming they evolved independently, as one could expect based on ancestral state reconstructions of sexual systems (*6, 7*).

The sex-linked region contains many genes related to flower sexual development

For *P. dactylifera*, a hypothesis for the evolution of the sex chromosomes and dioecy has been proposed by Torres et al. (2018)(*9*) who identified the Y-specific genes GPAT3, CYP703 and LOG3. It involves two steps: first, loss of male function in some individuals leading to gynodioecy (involving recessive loss-of-function of GPAT3 and/or CYP703), followed by the loss of female function (dominant suppression by LOG3) leading to dioecy. This hypothesis is reminiscent of the classical "two locus model" proposed by Charlesworth & Charlesworth (1978)(*22*). In our analysis, LOG3 (LOC103701078) was inferred as XY (Data S1), but this is due to paralogy of the Y-specific gene and an autosomal gene, as previously found (*9*). Indeed, males had more reads than females and the reconstructed X and Y haplotypes aligned to two different loci in the date palm genome: LOC120104231 on chromosome 17 in the BC4 assembly for the X sequence, and mRNA from the Y chromosome (MH685630) for the Y sequence. Concerning the other two Y-specific genes identified previously, we confirmed by remapping previously unmapped reads (i.e. that did not align to the female DPV01 genome) to the Barhee BC4 genome that GPAT3 (LOC120108105) and CYP703 (LOC120107104) both had no female reads for *P. dactylifera* and thus are Y specific. In *K. elegans*, however, GPAT3 and CYP703 were inferred as XY gametologs, and LOG3 as autosomal, suggesting that these genes do not play the same role in sex determination in *K. elegans*.

How, then, did this region become sex-linked independently in two different lineages? Further investigation revealed that several other genes in this region have functions in flower sexual organ development, including MYB61 (LOC103695972), no pollen germination 1 (NPG1) (LOC113461716 and LOC103697114), flap endonuclease 1A (LOC103696654), CCR4–NOT (LOC103696352), and DEAD-box ATP-dependent RNA helicase (LOC103697660, LOC103698960, and LOC103699024), related to pollen, anther development, and male sterility in plants (*23–27*) (Table 1). An ABC transporter C (LOC103697353) is similar to a gene associated with gynoecious sex expression in bitter gourd (*28*). Additionally, pentatricopeptide repeat-containing protein (PPR) coding genes (LOC103720547 and LOC103705387) belongs to a protein family involved in cytoplasmic male sterility in some plants (*29*). The gene squamosa promoter-binding-like protein_6 (SPL6) (LOC103724287) in *K. elegans* is homologous to SPL8 in *Arabidopsis thaliana*, involved in anther development and male fertility (*30*). Additionally, the sex-linked transcription factor phytochrome interacting factor-like 15 (PIL) (LOC103697220) in *K. elegans* is similar to PIL family factors that interact with SPLs (*31*) and is orthologous to the oil palm LOC105059744 gene that displays sex-dependent expression (*32*). All the sex-linked genes mentioned have homologs positioned in the 43 and 46.1Mb region of the *E. guineensis* chromosome 10, except the LOG3 gene located in the chromosome 12 (Table 1).

The role in palm sex differentiation of hormones, cytokinins, auxins and jasmonate, has been demonstrated in the monoecious species *Elaeis guineensis* and *Areca catechu* through variations in their contents between female and male inflorescences (*32*, *33*). In *E. guineensis*, female inflorescence development is associated with cytokinin accumulation, increased LOG and SPL gene expression and reduced miRNA regulators, while male inflorescences show higher auxin levels and increased expression of PIL genes (*32*). An antagonistic regulatory model involving SPL, PIL, and LOG genes has been proposed for

sexual differentiation and branching activity in inflorescences (*32*). These genes, along with shaggy-related protein kinase (LOC103697930) are involved in panicle branching in rice (*34*). In monoecious maize, sex mutants reveal carpel development in the tassel suggesting shared pathways between sex determination and inflorescence meristem branching (*35*). As genes similar to SPL, PIL and LOG3 are sex-linked in *P. dactylifera* and/or *K. elegans*, the same genes could be involved in flower sex determination in monoecious species and sex determination on the plant level in dioecious species. Such a similarity in sex-determining mechanisms has for example been demonstrated experimentally in melon (*36*), and it has been argued that the modification of the expression of plant hormones such as cytokinins could be involved in several transitions from monoecy to dioecy (*4, 5, 37*).

An ancestral region linked to flower unisexualization in palm genomes

The genomic region associated with flower sex exhibits collinearity with the *Elaeis guineensis* genome and shows also synteny with two other monoecious species, *Cocos nucifera* and *Areca catechu* (*33*). It was therefore probably present in the common ancestor of the Arecoideae (to which *E. guineensis*, *C. nucifera* and *A. catechu* belong) and the Coryphoideae (with *P. dactylifera* and *K. elegans*) subfamilies, estimated to be 87 million years old (*21*), and could constitute an ancestrally syntenic region within the palm genomes. Due to its high concentration of sex-related genes, it could serve as a "genetic toolbox" for the production of unisexual flowers in palms. Clusters of genes with similar functions have been mainly found to be involved in the production of certain metabolites (*38*) but some are related to development (*39, 40*). The existence of such clusters could be due to the effect of chromatin organization on transcription that favors the co-localization of simultaneously expressed genes, but the exact expression dynamics of these genes are not yet known in palms. Furthermore, how such clusters arose and are maintained remains to be investigated (*41*). Additionally, the existence of this "toolbox" may be related to the fact that palms are among the plant families with the highest proportion of unisexual flowers (around 80% of species (*6*), compared to about 10% overall in angiosperms), and could favor transitions between monoecy, dioecy and hermaphroditism. Unraveling the role of the genes in this region will shed more light on how such transitions take place; currently, it is not known whether dioecy in *Phoenix* and *Kerriodoxa* evolved through intermediates of gynodioecy or monoecy, as their known closest relatives have hermaphroditic flowers (*6, 7*).

Also, considering the importance of several palm species as vital crops, where only the female sex produces fruits and seeds, these genes could be used to develop genetic sex markers that allow early selection of female plants with high added value, thereby avoiding to wait before the sexual phenotype becomes apparent, a process that can take up to 15 years (*42*). More broadly, the existence of a region determining the unisexuality of inflorescences and the knowledge of the genes it contains offer perspectives for controlling the sex ratio of cultivated palm species, whether they are dioecious or monoecious (*19*).

Despite their independent origins, the sex-linked regions in *P. dactylifera* and *K. elegans* share striking similarities including the size of the sex-linked region, rather weak divergence given the age of recombination suppression, and weak degeneration (no loss of Y genes). We hypothesize that this region evolved recombination suppression twice due to the high concentration of genes involved in flower sexual development. An inversion is presumed to have created the proto-Y-chromosome of *P. dactylifera* (*9*) but how this non-recombining

region extended is not known. Several genes in this region have relatively high divergence between their X and Y copies (Figure 3, Data S1) in *P. dactylifera* and some of them are involved in the development of flower sexual organs, so they could have contributed to extending the non-recombining region through sex-antagonistic selection. In *K. elegans*, in the absence of a genome assembly, how recombination was suppressed remains an open question.

The evolution of sex chromosomes has been studied in a few groups or genera with multiple closely related dioecious species. In some genera (e.g. *Silene* (*43*)) sex chromosomes have evolved after two independent transitions to dioecy but with striking differences; in others, where dioecy is ancestral for a whole clade, either frequent turnovers have been observed (*Salix* (*44*)) or on the contrary a long-term stability (*Cannabis* and *Humulus* (*45*)). The case we report here, of convergent evolution of a sex-linked region in two species of palms that diverged 62 million years ago from a non-dioecious ancestor, adds yet another dimension to the wide range of evolutionary scenarios observed in sex chromosome evolution in plants (*5, 46*). Thus, it provides further opportunities to test how sex chromosomes evolve, a question that has recently regained interest with the publication of several modeling studies (*47–49*), as well as to shed light on the evolution of dioecy (*22, 50, 51*).

Reference:

1. D. A. Sasson, J. F. Ryan, A reconstruction of sexual modes throughout animal evolution. *BMC Evol Biol* **17**, 1–10 (2017).

2. D. Bachtrog, J. E. Mank, C. L. Peichel, M. Kirkpatrick, S. P. Otto, T.-L. Ashman, M. W. Hahn, J. Kitano, I. Mayrose, R. Ming, N. Perrin, L. Ross, N. Valenzuela, J. C. Vamosi, The Tree of Sex Consortium, Sex Determination: Why So Many Ways of Doing It? *PLoS Biol* **12**, e1001899 (2014).

3. I. M. Henry, T. Akagi, R. Tao, L. Comai, One Hundred Ways to Invent the Sexes: Theoretical and Observed Paths to Dioecy in Plants. *Annu. Rev. Plant Biol.* **69**, 553–575 (2018).

4. A. P. Leite Montalvão, B. Kersten, M. Fladung, N. A. Müller, The Diversity and Dynamics of Sex Determination in Dioecious Plants. *Frontiers in Plant Science* **11**, 2280 (2021).

5. S. S. Renner, N. A. Müller, Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. *Nat. Plants* **7**, 392–402 (2021).

6. S. Nadot, E. Alapetite, W. J. Baker, J. W. Tregear, A. S. Barfod, The palm family (Arecaceae): a microcosm of sexual system evolution. *Bot. J. Linn. Soc.* **182**, 376–388 (2016).

7. C. Cássia-Silva, C. G. Freitas, L. Jardim, C. D. Bacon, R. G. Collevatti, In situ radiation explains the frequency of dioecious palms on islands. *Ann Bot* **128**, 205–215 (2021).

8. L. S. Mathew, M. Spannagl, A. Al-Malki, B. George, M. F. Torres, E. K. Al-Dous, E. K. Al-Azwani, E. Hussein, S. Mathew, K. F. Mayer, Y. A. Mohamoud, K. Suhre, J. A. Malek, A first genetic map of date palm (Phoenix dactylifera) reveals long-range genome structure conservation in the palms. *BMC Genomics* **15**, 285 (2014).

9. M. F. Torres, L. S. Mathew, I. Ahmed, I. K. Al-Azwani, R. Krueger, D. Rivera-Nuñez, Y. A. Mohamoud, A. G. Clark, K. Suhre, J. A. Malek, Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix. *Nat Commun* **9**, 3969 (2018).

10. K. M. Hazzouri, M. Gros-Balthazard, J. M. Flowers, D. Copetti, A. Lemansour, M. Lebrun, K. Masmoudi, S. Ferrand, M. I. Dhar, Z. A. Fresquez, U. Rosas, J. Zhang, J. Talag, S. Lee, D. Kudrna, R. F. Powell, I. J. Leitch, R. R. Krueger, R. A. Wing, K. M. A. Amiri, M. D. Purugganan, Genome-wide association mapping of date palm fruit traits. *Nat Commun* **10**, 4680 (2019).

11. E. Cherif, S. Zehdi-Azouzi, A. Crabos, K. Castillo, N. Chabrillange, J.-C. Pintaud, A. Salhi-Hannachi, S. Glémin, F. Aberlenc-Bertossi, Evolution of sex chromosomes prior to speciation in the dioecious *Phoenix* species. *J. Evol. Biol.* **29**, 1513–1522 (2016).

12. Jousson A, Naciri Y, Christe C, Marazzi B, Stauffer F, Not just females and males: Unravelling the complex sex determinism of the hemp palm, Trachycarpus fortunei. *American journal of botany* **110** (2023).

13. E. J. Morgan, C. N. Kaiser-Bunbury, P. J. Edwards, M. Scharmann, A. Widmer, F. Fleischer-Dogley, C. J. Kettle, Identification of sex-linked markers in the sexually cryptic coco de mer: are males and females produced in equal proportions? *AoB Plants* **12** (2020).

14. E. K. Al-Dous, B. George, M. E. Al-Mahmoud, M. Y. Al-Jaber, H. Wang, Y. M. Salameh, E. K. Al-Azwani, S. Chaluvadi, A. C. Pontaroli, J. DeBarry, V. Arondel, J. Ohlrogge, I. J. Saie, K. M. Suliman-Elmeer, J. L. Bennetzen, R. R. Kruegger, J. A. Malek, De novo genome sequencing and comparative genomics of date palm (Phoenix dactylifera). *Nat Biotechnol* **29**, 521–527 (2011).

15. I. S. Al-Mssallem, S. Hu, X. Zhang, Q. Lin, W. Liu, J. Tan, X. Yu, J. Liu, L. Pan, T. Zhang, Y. Yin, C. Xin, H. Wu, G. Zhang, M. M. Ba Abdullah, D. Huang, Y. Fang, Y. O. Alnakhli, S. Jia, A. Yin, E. M. Alhuzimi, B. A. Alsaihati, S. A. Al-Owayyed, D. Zhao, S. Zhang, N. A. Al-Otaibi, G. Sun, M. A. Majrashi, F. Li, null Tala, J. Wang, Q. Yun, N. A. Alnassar, L. Wang, M. Yang, R. F. Al-Jelaify, K. Liu, S. Gao, K. Chen, S. R. Alkhaldi, G. Liu, M. Zhang, H. Guo, J. Yu, Genome sequence of the date palm Phoenix dactylifera L. *Nat Commun* **4**, 2274 (2013).

16. J. Käfer, N. Lartillot, G. A. B. Marais, F. Picard, Detecting sex-linked genes using genotyped individuals sampled in natural populations. *Genetics* **218**, iyab053 (2021).

17. S. B. Carey, J. T. Lovell, J. Jenkins, J. Leebens-Mack, J. Schmutz, M. A. Wilson, A. Harkess, Representing sex chromosomes in genome assemblies. *Cell Genomics* **2** (2022).

18. M. F. Torres, Y. A. Mohamoud, S. Younuskunju, K. Suhre, J. A. Malek, Evidence of Recombination Suppression Blocks on the Y Chromosome of Date Palm (Phoenix dactylifera). *Front Plant Sci* **12**, 634901 (2021).

19. A.-L. Ong, C.-K. Teh, S. Mayes, F. Massawe, D. R. Appleton, H. Kulaveerasingam, An Improved Oil Palm Genome Assembly as a Valuable Resource for Crop Improvement and Comparative Genomics in the Arecoideae Subfamily. *Plants (Basel)* **9**, 1476 (2020).

20. W. J. Baker, T. L. P. Couvreur, Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. *Journal of Biogeography* **40**, 274–285 (2013).

21. Sawyer S, Statistical tests for detecting gene conversion. *Molecular biology and evolution* **6** (1989).

22. B. Charlesworth, D. Charlesworth, A Model for the Evolution of Dioecy and Gynodioecy. *The American Naturalist* **112**, 975–997 (1978).

23. S. Moritoh, D. Miki, M. Akiyama, M. Kawahara, T. Izawa, H. Maki, K. Shimamoto, RNAi-mediated Silencing of OsGEN-L (OsGEN-like), a New Member of the RAD2/XPG Nuclease Family, Causes Male Sterility by Defect of Microspore Development in Rice. *Plant Cell Physiol* **46**, 699–715 (2005).

24. Li X, Gao X, Wei Y, Deng L, Ouyang Y, Chen G, Li X, Zhang Q, Wu C, Rice APOPTOSIS INHIBITOR5 coupled with two DEAD-box adenosine 5'-triphosphate-dependent RNA helicases regulates tapetum degeneration. *The Plant cell* **23** (2011).

25. S.-B. Shin, M. Golovkin, A. S. N. Reddy, A pollen-specific calmodulin-binding protein, NPG1, interacts with putative pectate lyases. *Sci Rep* **4**, 1–8 (2014).

26. L. Sun, G. Sun, C. Shi, D. Sun, Transcriptome analysis reveals new microRNAs-mediated pathway involved in anther development in male sterile wheat. *BMC Genomics* **19**, 1–17 (2018).

27. Wang Y, Zhou H, He Y, Shen X, Lin S, Huang L, MYB transcription factors and their roles in the male reproductive development of flowering plants. *Plant science : an international journal of experimental plant biology* **335** (2023).

28. Vinay N. D., Hideo Matsumura, Anilabha Das Munshi, Ranjith Kumar Ellur, Viswanathan Chinnusamy, Ankita Singh, Mir Asif Iquebal, Sarika Jaiswal, Gograj Singh Jat, Ipsita Panigrahi, Ambika Baladev Gaikwad, A. R. Rao, Shyam Sundar Dey, Tusar Kanti Behera, Molecular mapping of genomic regions and identification of possible candidate genes associated with gynoecious sex expression in bitter gourd. *Front. Plant Sci.* **14**, 1071648 (2023).

29. K. Kitazaki, K. Oda, A. Akazawa, R. Iwahori, Molecular genetics of cytoplasmic male sterility and restorer-of-fertility for the fine tuning of pollen production in crops. *Theor Appl Genet* **136**, 1–16 (2023).

30. Xing S, Salinas M, Höhmann S, Berndtgen R, Huijser P, miR156-targeted and nontargeted SBP-box transcription factors act in concert to secure male fertility in Arabidopsis. *The Plant cell* **22** (2010).

31. L. Zhang, G. He, Y. Li, Z. Yang, T. Liu, X. Xie, X. Kong, J. Sun, PIL transcription factors directly interact with SPLs and repress tillering/branching in plants. *New Phytologist* **233**, 1414–1425 (2022).

32. Tregear Jw, Richaud F, Collin M, Esbelin J, P. H, Cochard B, Nodichao L, Morcillo F, Adam H, Jouannic S, Micro-RNA-Regulated SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Expression and Cytokinin Accumulation Distinguish Early-Developing Male and Female Inflorescences in Oil Palm (Elaeis guineensis). *Plants (Basel, Switzerland)* **11** (2022).

33. G. Zhou, H. Yin, F. Chen, Y. Wang, Q. Gao, F. Yang, C. He, L. Zhang, Y. Wan, The genome of Areca catechu provides insights into sex determination of monoecious plants. *New Phytologist* **236**, 2327–2343 (2022).

34. X. Zhang, W. Meng, D. Liu, D. Pan, Y. Yang, Z. Chen, X. Ma, W. Yin, M. Niu, N. Dong, J. Liu, W. Shen, Y. Liu, Z. Lu, C. Chu, Q. Qian, M. Zhao, H. Tong, Enhancing rice panicle branching and grain yield through tissue-specific brassinosteroid inhibition. *Science*, doi: 10.1126/science.adk8838 (2024).

35. Q. Li, B. Liu, Genetic regulation of maize flower development and sex determination. *Planta* **245**, 1–14 (2017).

36. Boualem A, Troadec C, Camps C, Lemhemdi A, Morin H, Sari Ma, Fraenkel-Zagouri R, Kovalski I, Dogimont C, Perl-Treves R, Bendahmane A, A cucurbit androecy gene reveals how unisexual flowers develop and dioecy emerges. *Science (New York, N.Y.)* **350** (2015).

37. E. M. Golenberg, N. W. West, Hormonal interactions and gene regulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. *American J of Botany* **100**, 1022–1037 (2013).

38. Balounova V, Gogela R, Cegan R, Cangren P, Zluvova J, Safar J, Kovacova V, Bergero R, Hobza R, Vyskot B, Oxelman B, Charlesworth D, Janousek B, Evolution of sex determination and heterogamety changes in section Otites of the genus Silene. *Scientific reports* **9** (2019).

39. Hu N, Sanderson Bj, Guo M, Feng G, Gambhir D, Hale H, Wang D, Hyden B, Liu J, Smart Lb, DiFazio Sp, Ma T, Olson Ms, Evolution of a ZW sex chromosome system in willows. *Nature communications* **14** (2023).

40. D. Prentout, N. Stajner, A. Cerenak, T. Tricou, C. Brochier-Armanet, J. Jakse, J. Käfer, G. A. B. Marais, Plant genera *Cannabis* and *Humulus* share the same pair of well-differentiated sex chromosomes. *New Phytol* **231**, 1599–1611 (2021).

41. C. Zhan, S. Shen, C. Yang, Z. Liu, A. R. Fernie, I. A. Graham, J. Luo, Plant metabolic gene clusters in the multi-omics era. *Trends in Plant Science* **27**, 981–1001 (2022).

42. E. Khavkin, E. H. Coe, Mapped genomic locations for developmental functions and QTLs reflect concerted groups in maize (Zea mays L.). *Theor Appl Genet* **95**, 343–352 (1997).

43. J. Reimegård, S. Kundu, A. Pendle, V. F. Irish, P. Shaw, N. Nakayama, J. F. Sundström, O. Emanuelsson, Genome-wide identification of physically clustered genes suggests chromatin-level co-regulation in male reproductive development in Arabidopsis thaliana. *Nucleic Acids Res* **45**, 3253–3265 (2017).

44. Hurst Ld, Pál C, Lercher Mj, The evolutionary dynamics of eukaryotic gene order. *Nature reviews. Genetics* **5** (2004).

45. E. Cherif, S. Zehdi, K. Castillo, N. Chabrillange, S. Abdoulkader, J. Pintaud, S. Santoni, A. Salhi-Hannachi, S. Glémin, F. Aberlenc-Bertossi, Male-specific DNA markers provide genetic evidence of an XY

chromosome system, a recombination arrest and allow the tracing of paternal lineages in date palm. *New Phytol* **197**, 409–415 (2013).

46. Charlesworth D, Harkess A, Why should we study plant sex chromosomes? *The Plant cell*, doi: 10.1093/plcell/koad278 (2024).

47. Jeffries Dl, Gerchen Jf, Scharmann M, Pannell Jr, A neutral model for the loss of recombination on sex chromosomes. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **376** (2021).

48. T. Lenormand, D. Roze, Y recombination arrest and degeneration in the absence of sexual dimorphism. *Science* **375**, 663–666 (2022).

49. P. Jay, E. Tezenas, A. Véber, T. Giraud, Sheltering of deleterious mutations explains the stepwise extension of recombination suppression on sex chromosomes and other supergenes. *PLOS Biology* **20**, e3001698 (2022).

50. J. Käfer, G. A. B. Marais, J. R. Pannell, On the rarity of dioecy in flowering plants. *Mol Ecol* **26**, 1225–1241 (2017).

51. T. Lesaffre, J. Pannell, C. Mullon, A model for the gradual evolution of dioecy and heterogametic sex determination. *bioRxiv* (2024).

52. J. R. Comer, W. B. Zomlefer, C. F. Barrett, J. I. Davis, D. W. Stevenson, K. Heyduk, J. H. Leebens-Mack, Resolving relationships within the palm subfamily Arecoideae (Arecaceae) using plastid sequences derived from next-generation sequencing. *American Journal of Botany* **102**, 888–899 (2015).

53. N. Rohland, D. Reich, Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Res.* **22**, 939–946 (2012).

54. Mascher M, Richmond Ta, Gerhardt Dj, Himmelbach A, Clissold L, Sampath D, Ayling S, Steuernagel B, Pfeifer M, D'Ascenzo M, Akhunov Ed, Hedley Pe, Gonzales Am, Morrell Pl, Kilian B, Blattner Fr, Scholz U, Mayer Kf, Flavell Aj, Muehlbauer Gj, Waugh R, Jeddeloh Ja, Stein N, Barley whole exome capture: a tool for genomic research in the genus Hordeum and beyond. *The Plant journal : for cell and molecular biology* **76** (2013).

55. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

56. A. Shumate, S. L. Salzberg, Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).

57. Quinlan Ar, Hall Im, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26** (2010).

58. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

59. M. Lefouili, K. Nam, The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Sci Rep* **12**, 11331 (2022).

60. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics* **81**, 559 (2007).

61. A. Manichaikul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W.-M. Chen, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).

62. Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* **24**, 1586–1591 (2007).

63. Edgar Rc, MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32** (2004).

64. Löytynoja A, Phylogeny-aware alignment with PRANK. *Methods in molecular biology (Clifton, N.J.)* **1079** (2014).

65. R. Bouckaert, T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina, J. Heled, G. Jones, D. Kühnert, N. D. Maio, M. Matschiner, F. K. Mendes, N. F. Müller, H. A. Ogilvie, L. du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C.-H. Wu, D. Xie, C. Zhang, T. Stadler, A. J. Drummond, BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology* **15**, e1006650 (2019).

66. C. Lanave, G. Preparata, C. Sacone, G. Serio, A new method for calculating evolutionary substitution rates. *J Mol Evol* **20**, 86–93 (1984).

67. Z. Yang, Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol Biol Evol* **10**, 1396–1401 (1993).

68. Z. Yang, Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. *J Mol Evol* **39**, 306–314 (1994).

69. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, Relaxed Phylogenetics and Dating with Confidence. *PLOS Biology* **4**, e88 (2006).

70. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* **67**, 901–904 (2018).

71. Yu G, Using ggtree to Visualize Data on Tree-Like Structures. *Current protocols in bioinformatics* **69** (2020).

Data and materials availability:

Raw data are available under the European Bioinformatics Institute (EBI) – European Nucleotide Archive (ENA) Project PRJEB76653

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors

Acknowledgments:
We thank the ISO 9001 certified IRD i-Trop HPC (South Green Platform) at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this paper, as well as Audrey Weber from the Genotyping plateform, INRAe Arcad AGAP, for sequencing. We also thank Thomas Lenormand, James Tregear and Gabriel Marais for suggestions that helped to improve the manuscript.
Funding:
LabEx AGRO 2011-LABX-002 (under I-site Muse framework) coordinated by Agropolis Fondation : Project (ID 2101-014 and ID 2201-007)

Author contributions:

Conceptualization: FA

Methodology: HT, TB, EC, JO, SS, JK, FA

Investigation: HT, TB, JK, FA

Visualization: HT, JK

Funding acquisition: FA, JK

Project administration: FA, JK

Resources : ALi,ALe,MD

Supervision: FA, JK

Writing – original draft: HT, JK, FA

Writing – review & editing: HT, TB, EC, JO, SS, JK, FA

Competing interests: Authors declare that they have no competing interests.

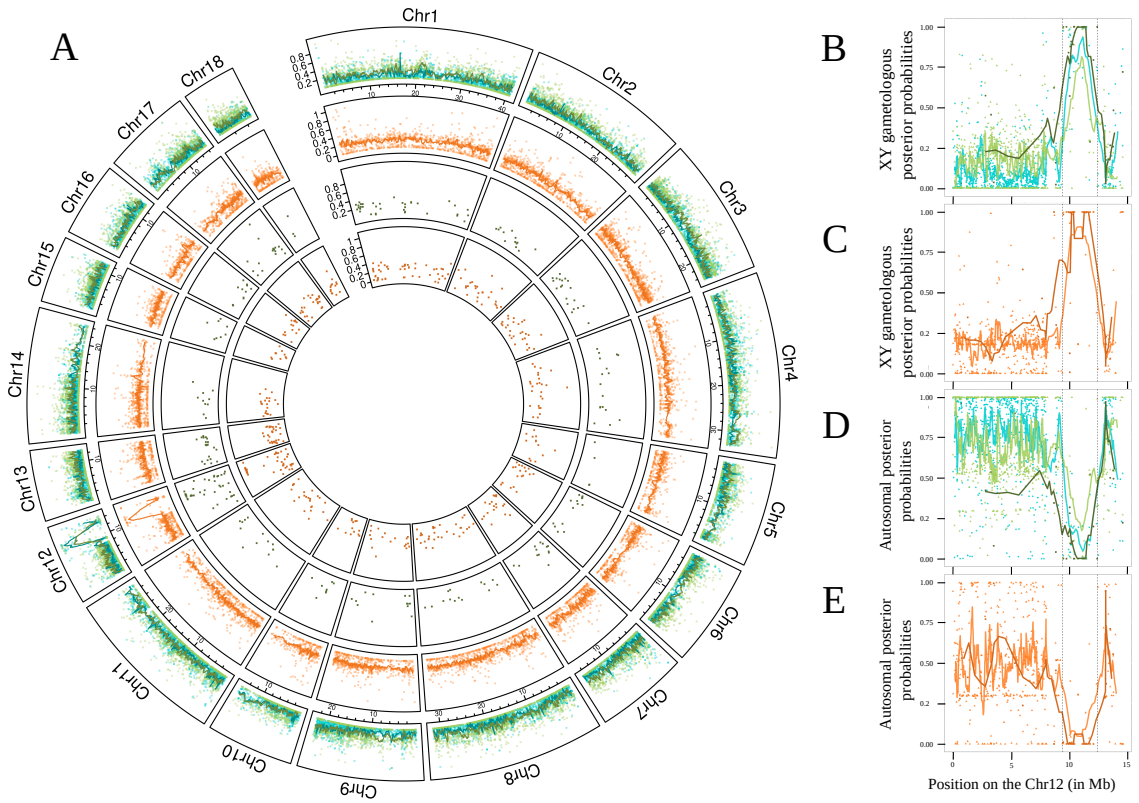
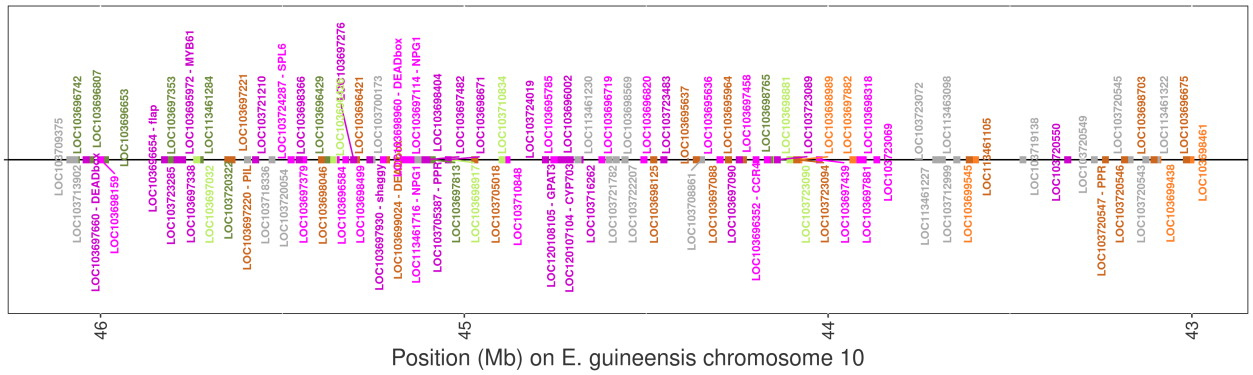


Fig. 1. Probability of sex linkage (XY) of each gene in *Phoenix dactylifera* and *Kerriodoxa elegans*. The position of the genes is taken from the Barhee BC4 assembly of *P. dactylifera* (*10*). *P. dactylifera* is indicated in blue (whole genome resequencing data) and green (capture kit data), *K. elegans* in orange. (**A**) Genome-wide distribution of XY posterior probabilities obtained with SDpop. Outer two rings show whole-genome resequencing and capture (PalmExons) data, the inner two the SexPhoenix capture kit. (B,C) XY posterior probabilities per gene on LG12. (D,E) Autosomal posterior probabilities per gene. Lines show the rolling mean with a window size of fifteen genes for the PalmExons kit (light green or orange) and whole genome resequencing data (blue) and five genes for the SexPhoenix kit (dark green or orange).

A



B

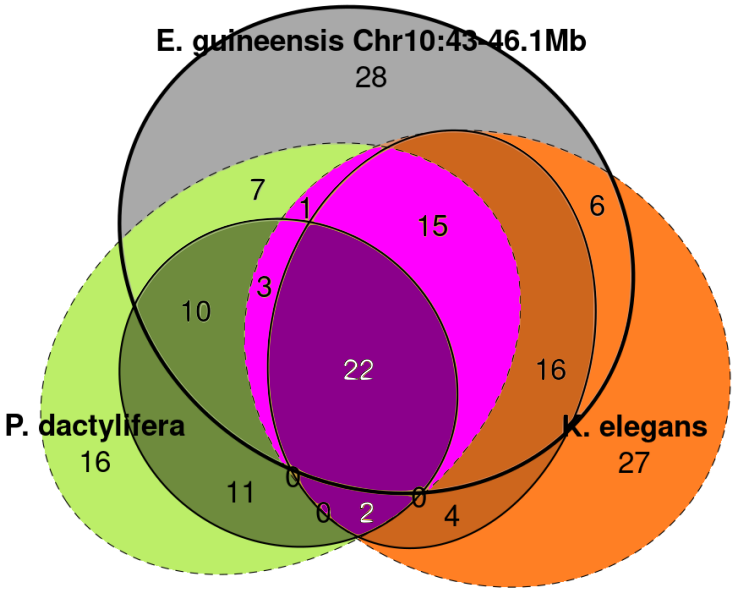


Fig. 2. Position of *Kerriodoxa elegans* and *Phoenix dactylifera* sex-linked genes on *Elaeis guineensis* chromosome 10. Genes with posterior probability to be XY above 0.8 in at least one dataset or present between 43 and 46.1Mb on chromosome 10 of the *E. guineensis* EGPMv6 genome were catergorized as follows: sex-linked with high confidence, probably sex-linked, inconclusive, not sex-linked or missing data. **(A)** Positions of genes between 43 and 46.1Mb on chromosome 10 of *E. guineensis*. Sex-linked genes were colored in green for *P. dactylifera*, orange for *K. elegans*, and magenta for both, while the shade of the color (dark or light) indicates the level of confidence (high confidence and probable, respectively). Genes from other categories (i.e. either inconclusive or absent in at least one species or autosomal in at most one) were colored in grey. (B) Euler diagram of the number of genes confidently (darker) or probably (lighter) sex-linked in *P. dactylifera* (green) or *K. elegans* (orange) or positioned between 43 and 46.1Mb on the chromosome 10 of *E. guineensis* (grey).

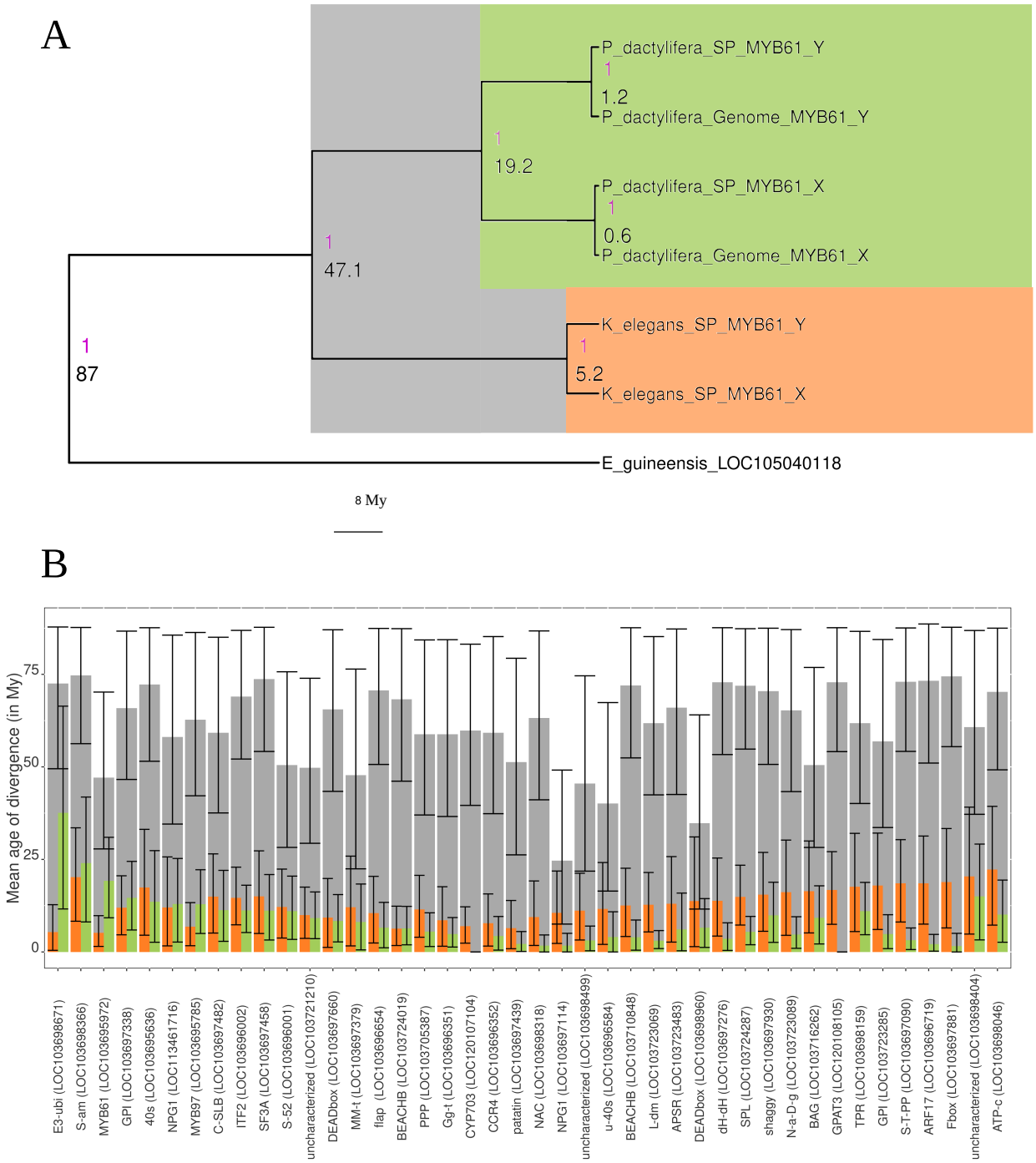


Fig. 3. Ages of species and gametolog divergence for common sex-linked genes in *Phoenix dactylifera* and *Kerriodoxa elegans*. Age estimates are colored in grey for the divergence of the *Phoenix* and *Kerriodoxa* lineages, in green for *P. dactylifera* X-Y divergence and in orange for *K. elegans* X-Y divergence. **(A)** Phylogenetic tree for the MYB61 (LOC103695972). Node values correspond to the mean age for the node's most recent common ancestor (MRCA; black) and the support of the node (magenta). (B) Mean age of divergence of X and Y sequences for common sex-linked genes compared to the age of divergence between the MRCA of *P. dactylifera* and *K. elegans*. Error bars correspond to the 95% probability density of the node age.

1

***P. dactylifera* gene**	**Putative gene function in sex**	***E. guineensis* location (Mb)**	***P. dactylifera* gene category**	***K. elegans* gene category**
LOG3 (LOC103701078) cytokinin riboside 5'-monophosphate phosphoribohydrolase	Female flower development (*52*, *53*)	Chr 12 10.9	Y specific	Autosomal
pentatricopeptide repeat protein At3g14580 (LOC103720547)	Cytoplasmic male sterility (*29*)	Chr10 43.2	Autosomal	XY
CCR4-not (LOC103696352) CCR4-NOT transcription complex subunit 9-like	Anther development (*26*)	Chr10 44.2	Probably XY	XY
CYP 703 (LOC120107104) cytochrome P450 703A2	Pollen and anther development (*54*, *55*)	Chr10 44.7	Y specific	XY
GPAT 3 (LOC120108105) glycerol-3-phosphate acyltransferase RAM2-like	Pollen and anther development (*56*)	Chr10 44.7	Y specific	XY
NPG1 (LOC103697114) protein NPG1-like	Pollen germination (*25*)	Chr10 45.1	Probably XY	XY
pentatricopeptide repeat protein At1g02420 (LOC103705387)	Cytoplasmic male sterility (*29*)	Chr10 45.1	XY	XY
NPG1 (LOC113461716) protein NPG1-like	Pollen germination (*25*)	Chr10 45.2	XY	Probably XY
DEAD-box ATP-dependent RNA helicase protein DB10 (LOC103698960)	Male sterility linked to tapetum degeneration (*24*)	Chr10 45.2	Probably XY	XY
DEAD-box ATP-dependent RNA helicase 40 (LOC103699024)	Male sterility linked to tapetum degeneration (*24*)	Chr10 45.2	inconclusive	XY
shaggy (LOC103697930) shaggy-related protein kinase alpha-like	rice panicle branching (*34*, *35*)	Chr10 45.3	XY	XY
SPL (LOC103724287) squamosa promoter-binding-like protein 6	Anther development and male fertility (*30*)	Chr10 45.5	Probably XY	XY
PIL (LOC103697220) transcription factor phytochrome interacting factor-like 15-like	Interacts with SPL and involved in rice panicle branching. (*31*)	Chr10 45.6	Autosomal	XY
MYB61 (LOC103695972) transcription factor MYB61	Male reproductive development (*27*)	Chr10 45.8	XY	XY
flap endonuclease 1A (LOC103696654)	Early microspore development (*23*)	Chr10 45.8	XY	XY
DEAD-box ATP-dependent RNA helicase 3B (LOC103697660)	Male sterility linked to tapetum degeneration (*24*)	Chr10 46.0	XY	XY
ABC transporter C 1 (LOC103697353)	Gynoecious sex expression (*28*)	Chr10 45.8	XY	inconclusive

2
3

Table 1. Sex-linked genes with functions related to flower development. Sex-linked genes in either *Phoenix dactylifera* or *Kerriodoxa elegans* with identified functions related to sexual development are listed, as well as their position on the *Elaeis guineensis* EGPMv6 genome. The level of confidence according to our analysis (Material and Methods) is indicated, as well as the reference for the function.

Supplementary Materials for

Convergent evolution of sex chromosomes in palms

Authors: H. Tessarotto[1], T. Beulé[1], E. Cherif[1], J. Orjuela[1], A. Lindstrom[2], A. Lemansour[3], M. Dahme[4], S. Santoni[5], J. Käfer[6]*†, F. Aberlenc[1]*†

Corresponding author: *frederique.aberlenc@ird.fr* and *jos.kafer@cnrs.fr*

The PDF file includes:

Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S5

Other Supplementary Materials for this manuscript include the following:

Data S1 to S2

Materials and Methods

Genomic data and treatment

Leaf samples of 10 females and 10 males of *P. dactylifera* were collected in Al Ain, United Arab Emirates and 14 females and 11 males of *K. elegans* in Puerto Rico (Central Cordillera) and in Thailand (Nong Nooch Tropical Botanical Garden). DNA was extracted from 15 mg of dried leaves with the Chemagic DNA Plant Kit (Perkin Elmer Chemagen, Baesweller, DE, Part # CMG-194), according to the manufacturer's instructions. The protocol is adapted to the use of the KingFisher Flex™ (Thermo Fisher Scientific, Waltham, MA, USA) automated DNA purification workstation. To target gene sequences, we developed two sequence capture kits. First, SexPhoenix is a capture kit composed of 20,000 baits (120 mers) targeting sex-linked regions of *P. dactylifera* (*14*) as well as a selection of nuclear and chloroplastic genes (*57*). Second, PalmExons consists of 60,000 baits (90 mers) targeting the exomic regions of *P. dactylifera*. For the exact protocol, see Supplementary Text 1. In short, genomic library preparation for multiplexed individuals and enrichment step by capture follow published protocols (*58, 59*) with some modifications. The sequences were obtained by sequence capture using the two kits of baits (myBaits, Arbor Biosciences). The final pooled library was sequenced using the Illumina paired-end protocol on a single lane of a HiSeq3000 sequencer, for 2 x 150 cycles. For comparison with the capture kits, we downloaded data of 20 females accessions and 12 males accessions of *P. dactylifera* from NCBI among the bioproject PRJNA505141 (*10*).

Sequencing data were mapped on the DPV01 genome of *P. dactylifera* (*15*), the only female and annotated genome available at the time of analysis. We used the BWA mem2 algorithm (version 2.2.1; (*60*)) to map the reads, lowering the penalty for mismatches from 4 to 2 (option -B 2), to allow more divergent reads (Y chromosome, *K. elegans*) to be correctly mapped. For comparison with the current chromosome-level reference assembly (Barhee BC4, male genome, (*10*)) we identified the correspondence of the DPV01 annotation with liftoff (*61*). For all plots, the position of the genes on the Barhee BC4 genome was used. Liftoff was also used to position DPV01 assembly genes on the *Elaeis guineensis* EGPMv6 assembly (*19*). Y-specific genes not present on the DPV01 XX assembly were studied by recovering unmapped reads and aligning them on the Barhee BC4 XY assembly.

Coverage depth of each base was calculated using bedtools 2.30 genomecov (*62*). We calculated the mean depth for 100 base windows for all samples, as well as the average for each sex. To calculate the mean depth per gene, we used the windows with at least 50% overlap with the CDS of the gene. To calculate the proportion of female reads for each 100 base windows, we accounted for possible read depth differences between samples by normalizing the depth of each individual. To do so, for each window, the mean depth was divided by the median of the mean depth of all windows of each individual.

SNP calling was done with bcftools mpileup (version 1.9; (*63, 64*)). Genotypes obtained with less than 7 reads for the capture data and less than 10 for the downloaded whole-genome resequencing data were discarded. The resulting vcf file was then filtered to keep only the sites within coding sequences according to the DPV01 annotation. We conducted a Principal Component Analysis (PCA) of the genetic variation between individuals using plink 1.90 (*65*) to

check the population structure, and estimated relatedness between the individuals using the option "relatedness2" in vcftools (*66*).

Sex-linked genes

SDpop (*16*) was used to infer the presence and type (XY or ZW) of sex chromosomes, as well as to identify the sex-linked genes. In short, SDpop is a probabilistic method that uses population genetics equilibria to model sex linkage (i.e. gametology, XY or ZW, or hemizygosity, when Y or W copies are absent), as well as some frequently encountered sources of error. We furthermore used SDpop to reconstruct the X and Y haplotypes of the gametologous genes. CodeML (PAML version 1.9; (*67*)) was used to infer pairwise dS between X and Y haplotypes. Genes were sorted into different categories. They were considered as sex-linked candidates if they had posterior scores for sex-linkage higher than 0.8 in the SDpop output of at least one dataset. We then compared between datasets and considered the dataset with either the most polymorphisms detected for one gene or, if equal, with more individuals genotyped on average. To discard possible false positives, we then only considered as sex-linked, genes with a synonymous divergence (dS) higher than 0.001. Genes with lower dS or posterior probabilities between 0.5 and 0.8 (who still had posterior scores above 0.8 in other datasets) were considered as probably sex-linked. Genes with autosomal posterior scores above 0.8 were considered as not sex-linked. Finally if genes had on average less than 10 individuals genotyped per site, we could not assert sex-linkage or its absence and considered them inconclusive. All genes not fitting any of the previous criteria were classified in the inconclusive category as well.

Gene divergence and phylogeny

We identified genes in the oil palm *Elaeis guineensis* genome that are orthologous to the inferred sex-linked genes by selecting the best hit in a blast search against nr. Sequences were aligned using muscle (*68*) or PRANK (*69*) for more difficult alignments containing indels shifting the reading frame. Geneconv (*20*) was used to detect gene conversion in these alignments. Geneconv calculates two p-values, simulated p-values are considered more accurate yet less conservative than Karlin-Altschul (KA) p-values. In order to avoid underestimating the age of recombination arrest of genes, we considered pairwise or global inner-fragments to have undergone gene conversion when at least one of the two p-values was lower than 0.05 between X and Y sequences from the same dataset. For *P. dactylifera*, the same fragment of the gene was required to fit these criteria for both the Whole-Genome Sequencing (WGS) and targeted-capture datasets.
The phylogeny was then inferred using BEAST2 v2.7.5 (*70*) with a GTR + Ɣ model with 4 gamma categories (*71–73*). The length of the chain was 250,000,000 steps and a sample was taken every 100 steps after a pre burn in of 250 samples. A relaxed clock model was chosen (*74*). The prior for the most recent common ancestor (MRCA) age for the root was set to normal distribution with a mean of 87.025 and a standard deviation of 0.625 fitting the 95% confidence interval of the age of this node in previous dated phylogenies (*21*). Other parameters and priors were not modified. If convergence of the model wasn't reached, the length of the chain was increased and the program stopped after all parameters converged and reached an effective sample size (ESS) of 400 on Tracer v1.7.2 (*75*). A maximum clade credibility tree was then

generated with 10% burn in with node heights corresponding to MRCA age (*70*). The figures of the inferred trees were then generated using ggtree v3.11.1 (*76*).

Supplementary Text 1: Capture protocol

Target capture began with target preparation and construction of barcoded genomic libraries. For each individual, 100 ng of total DNA (in 13 µL of water) are sheared using the NEBNext Ultra II FS DNA Modul (New England Biolabs, Ipswich, MA, USA, Part # E7810) by mixing 3.5 µL of NEBNext Ultra II FS Reaction Buffer and 1 µL of NEBNext Ultra II FS Enzyme Mix and an incubation step of 10 min at 37°C. A clean-up step is performed with 1 x volume of Agencourt AMPure XP magnetic beads (Beckman Coulter, Indianapolis, IN, USA, Part # A63881). The elution volume is 40 µL. Fragmented DNA is ligated with 4 pmol of PE-P5 and MPE-P7 adapters. Each PE-P5 and each MPE-P7 adapter carries the same specific hexamer barcode (Rohland & Reich, 2012). Reactions are conducted in 50 µl final volume with 2 unit of T4 DNA ligase (Invitrogen, Thermo Fischer Scientific, Part # 15224025) for 4 hours at 22 °C followed by a heat inactivation step at 65°C for 10 minutes. A clean-up step is performed with 1 x volume of Agencourt AMPure XP magnetic beads. The elution volume is 40 µL. For each individual sample, a pre-hybridization PCR is performed using the KAPA® HiFi HotStart ReadyMix PCR Kit (Roche Diagnostics Indianapolis, IN, USA, Part # KK2602) with 200 nM PreHyb-PE_F (CTTTCCCTACACGACGCTCTTC) and 200 nM PreHyb-MPE_R (TGACTGGAGTTCAGACGTGTG) primers in a final volume of 100µl. Thermocycling parameters were 2 minutes at 98°C, followed by 12 cycles of 20 seconds at 98°C; 45 seconds at 55°C and 30 seconds at 72°C, with a final elongation of 10 minutes at 72°C. A clean-up step is performed with 1 x volume Agencourt AMPure XP magnetic beads. The elution volume is 30 µL. Amplified DNA is individually controlled (sizing and estimation of the concentration) by electrophoresis on a AATI Fragment Analyzer™ (Advanced Analytical Technologies, Ankeny, IA, USA) device with the DNF-474 High Sensitivity Fragment Analysis Kit. Until 24 samples (corresponding to 24 hexamer barcodes on the PE-P5 and PE-P7 adapter) are pooled. A clean-up step is performed with 1 x volume of Agencourt AMPure XP magnetic beads. The elution volume is 94µL.

Enrichment and capture by hybridization followed the Standard Protocol, with a hybridization and washing temperature of 65°C, following the indications of the User Manual of the MYBaits Hybridization Capture for Targeted NGS, Custom DNA-Seq. An on- beads PCR amplification is undertaken to enrich library fragments, extend the adapter sequence and incorporate an index to the P7 adapter. The PCR reaction is using the KAPA® HiFi HotStart ReadyMix PCR Kit (KAPABiosystems, Boston, MA, Part # KR0370) in a final volume of 50 µl with:

- 15 pmol of SOL-PE-PCR_F primer (*59*)
  aatgatacggcgaccaccgagatctacactctttccctacacgacgctcttc
- 15 pmol of SOL-MPE-INDX_R indexed primers
  CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGT

This primer carries 6 bases of the official TruSeq Illumina Index. Thermocycling parameters were 2 minutes at 98°C, followed by 18 cycles of 20 seconds at 98°C; 30 seconds at 62°C and 30 seconds at 72°C, with a final elongation of 5 minutes at 72°C. The reaction volume is 50 µL. A clean-up step is performed with 1.3 x volume Agencourt AMPure XP magnetic beads. The elution volume is 20 µL. Indexed libraries are individually controlled (sizing and estimation of

the concentration) by electrophoresis on an AATI Fragment Analyzer™ device with the DNF-474 High Sensitivity Fragment Analysis Kit. The different indexed libraries, corresponding to all the captured barcoded DNA samples, are equally mixed. The final pooled library is quantified by qPCR with the KAPA Library Quantification Kit (Part # KK4824).

Supplementary Text 2: Extended results

Characterizing sex-linked genes

Whole-genome sequencing data (WGS) of *P. dactylifera* and two capture subsets, SexPhoenix (SP) targeting sex-linked sequences and PalmExons (PE) targeting exomic sequences in *P. dactylifera*, were mapped on the annotated Khalas DPV01 female genome. Female genomes have homologous XX sex-linked sequences that allow both X and Y reads to map on the same locus, which facilitates detection of sex-linked genes (*16*). The average mapping rate for the WGS (percentage of properly paired reads, Table S1), and for the SP and PE capture kits (Table S2), was 94,28%, 99.46% and 99.51%, respectively. Due to high relatedness between samples, three females and three males were discarded from the WGS data (fig. S1-A, S2-A, table S1), and one male for the sequence capture data (fig. S1-B, S2-B, table S2). The datasets thus consisted of 17 females and 9 males for the WGS data, with more than 450,000 SNPs and indels for the analysis of sex-linkage covering 97.61% of the DPV01 annotated genes (table S3). For both capture datasets, 9 individuals of each sex were retained for the analyses, revealing 4,693 variant positions in the SP data which represents 1.60% of the annotated genes and 180,035 SNPs in the PE data representing 77.06% of the genes (table S3). PE hence covers a large part of the cds and is fit for investigating genomic regions on their coding content. The capture sequences of *K. elegans* had 83.95% of reads properly paired for SexPhoenix (SP) and 99.20% of reads properly paired for PalmExons (PE) when mapped on the DPV01 female *P. dactylifera* genome. Most *K. elegans* samples were somewhat related but this relatedness affected females and males equally (fig. S1-C, S2-C, table S2) so all individuals were kept for subsequent analyses. SDpop was run on 14 females and 11 males of *K. elegans* with 161,384 variant positions for PE covering 58% of the DPV01 annotated genes and 8,395 variant positions covering 2.5% of the annotated genes for SP (table S3).

Upon analysis of the WGS data with SDpop, 91% of the polymorphic sites were inferred to be autosomal (table S4) while the proportions of haploid and sex-linked (XY gametologous and X-hemizygote) sites were inferred to be less than 1%. In the sequence capture datasets, less than 1% of sites were inferred as sex-linked in PalmExons (table S4), consistent with the proportions of the WGS data (table S4). As expected by design of the capture kits, the proportion of sex-linked sites within SexPhoenix baits is a lot higher: 11.9% (table S4). A higher proportion of paralog sequences has been inferred for *K. elegans* than *P. dactylifera* (table S4), as expected when the mapped reads belong to a different species than the reference genome. The SDpop model that best fits the genomic data of *K. elegans* according to the Bayesian Information Criterion (BIC) is the XY sex chromosome system (table S4). With heterogametic male and homogametic females, *K. elegans* hence shares a XY sex chromosome system with *P. dactylifera*.

X hemizygous genes

A few genes had posterior scores higher than 0.8 for X-hemizygosity, but upon verification, none of these had significant differences in female and male read coverage contrary to what is expected of genes with two copies in female genomes and one in male's. In conclusion, no X-hemizygous genes were detected (Data S2).

Y-specific genes in P. dactylifera

Among the genes previously described as Y-specific in *P. dactylifera* (9), one, LOG3like (LOC103701078), was inferred as XY in our analysis (Data S1). Further investigation confirmed that this is due to paralogy of the Y-specific gene and an autosomal gene, as previously found (9), as males had more reads than females and the reconstructed X and Y haplotypes aligned to two different loci in the date palm NCBI's nucleotide database: LOC120104231 on chromosome 17 in the BC4 assembly for the X sequence, and mRNA from the Y chromosome (MH685630) for the Y sequence. The mean age of the duplication between the LOG autosomal and Y-linked copies is estimated to 53.76 My (fig. S3) and may predate the sex-linkage. Concerning the other two Y-specific genes identified previously, we confirmed by remapping previously unmapped reads (i.e. that did not align to the female DPV01 genome) to the Barhee BC4 genome that GPAT3 (LOC120108105) and CYP703 (LOC120107104) both had no female reads for *P. dactylifera* (Data S1). Mapping unmapped reads to the Barhee BC4 male genome revealed that two Y-specific genes of *P. dactylifera* : GPAT3 (LOC120108105) and CYP703 (LOC120107104) are XY in *K. elegans*. The third Y-specific gene of *P. dactylifera* : LOG3like (LOC103701078) mentioned earlier was inferred as autosomal in *K. elegans* (Data S1).

A shared sex-linked region

A quarter of the sex-linked genes of *P. dactylifera* and *K. elegans* were positioned on the LG12 (NC_052403.1) of the Barhee BC4 genome (Data S1) between 9.4 and 12.4Mb. Most of the other sex-linked genes were positioned in unplaced scaffolds. To explore the synteny between the genomes of date palm and oil palm, the sex-linked genes of *P. dactylifer*a and *K. elegans* were positioned on the *Elaeis guineensis* genome (Fig. 2).

The majority of the XY genes were found placed in the scaffold KE606796.1/DVFH01001701.1 assembled in the region between 43 and 46.1Mb of the *E. guineensis* chromosome 10 (GK000085.1) of EGPMv6 genome (Fig. 2). Thus, 78% of genes inferred as sex-linked in either *P. dactylifera* or *K. elegans* were positioned in this region (Figure 2-C). This reveals a very strong synteny between the 9.4-12.4Mb region of the date palm chromosome 12 including several unplaced scaffolds and the 43- 46.1 Mb region of the oil palm chromosome 10. More sex-linked genes are positioned on a single genomic region of *E. guineensis* genome than of *P. dactylifera*'s, likely due to the difficulty of assembling sex chromosome non recombining regions. In *K. elegans*, three sex-linked genes were positioned less than 400kb away on *E. guineensis* chromosome 3 and *P. dactylifera* chromosome 1 (Data S1). LOC103715767, LOC103721482 and LOC103695590 are therefore most likely positioned in the non recombining region of *K. elegans* genome.

Sex-linked gene phylogenies for P. dactylifera and K. elegans

To estimate synonymous divergence (dS) and gene phylogenies, the X and Y haplotype sequences of 40 common sex-linked genes were inferred using SDpop for *K. elegans* and *P. dactylifera*. For a better estimation of divergence, gene conversion between the X and Y copies was investigated. Gene conversion was detected in only 3 genes for *P. dactylifera* and 3 genes in *K. elegans* among the 40 genes commonly inferred as XY gametologous (fig. S5, table S5) covering on average 27% of the gene. The sequence stretches of the X and Y haplotypes that could have possibly undergone gene conversion were removed for the subsequent analysis to avoid underestimating the age of recombination arrest (fig. S5, table S5).

We compared the synonymous divergence (dS) of sex-linked genes with the dS of ten autosomal single copy genes. The dS between *E. guineensis* and the X or the Y sequences of *P. dactylifera* revealed a similar distribution to the dS between autosomal single copy orthologs of *P. dactylifera* and *E. guineensis* (fig. S4). Hence sex-linked genes diverge at a similar rate as autosomal genes in both *K. elegans* and *P. dactylifera*. We therefore assume the rate of divergence has not increased or decreased following recombination arrest and did not significantly impact age inference.

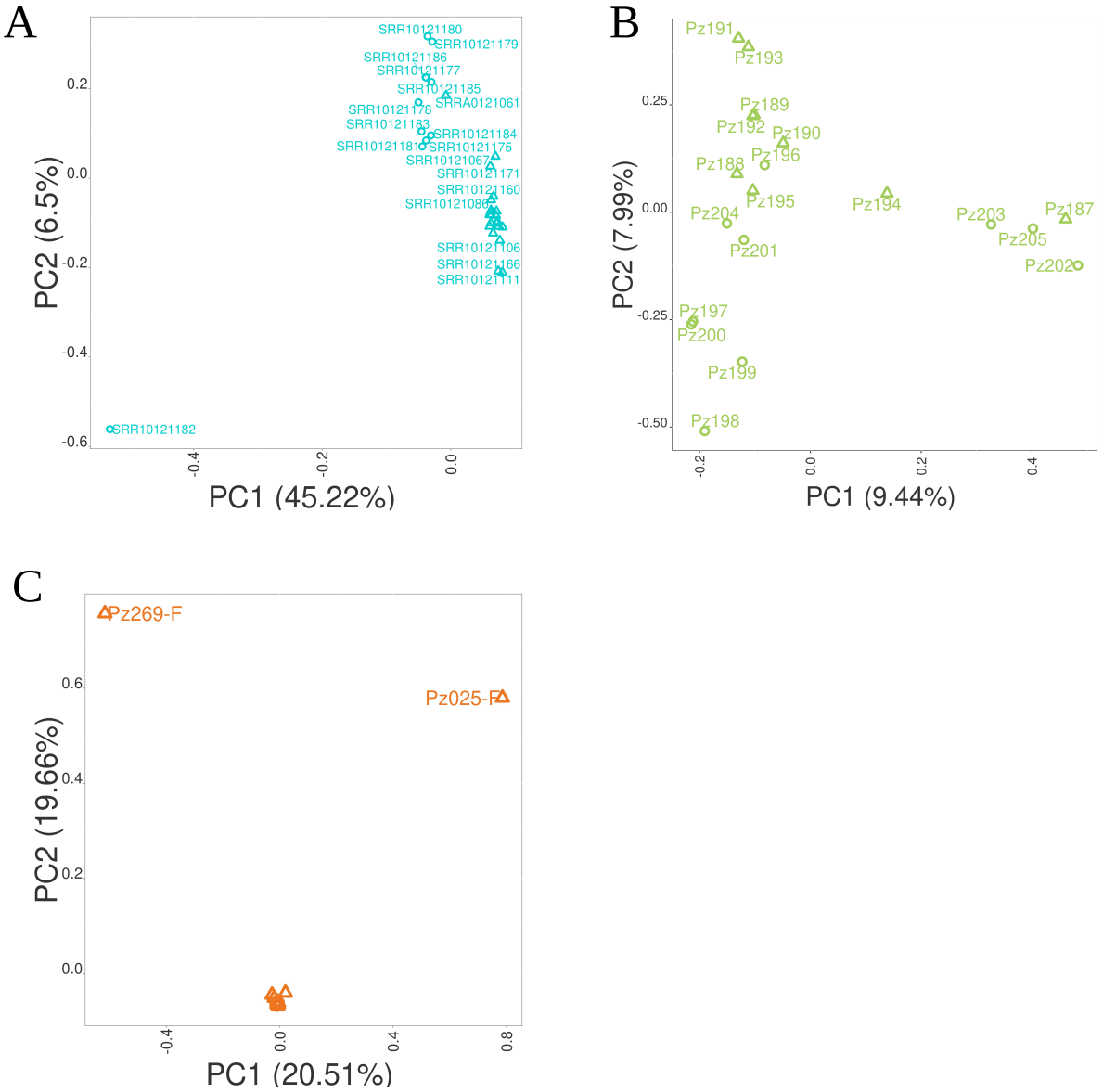


Fig. S1.

PCA of the genetic variation of studied populations. Females are represented by triangles, males by circles, one ambiguous *K. elegans* individual by a square. **(A)** PCA of the genetic variation of the *P. dactylifera* whole genome individuals. (B) PCA of the genetic variation captured with PalmExons kit among the *P. dactylifera* individuals. (C) PCA of the genetic variation captured with PalmExons kit among the *K. elegans* individuals.

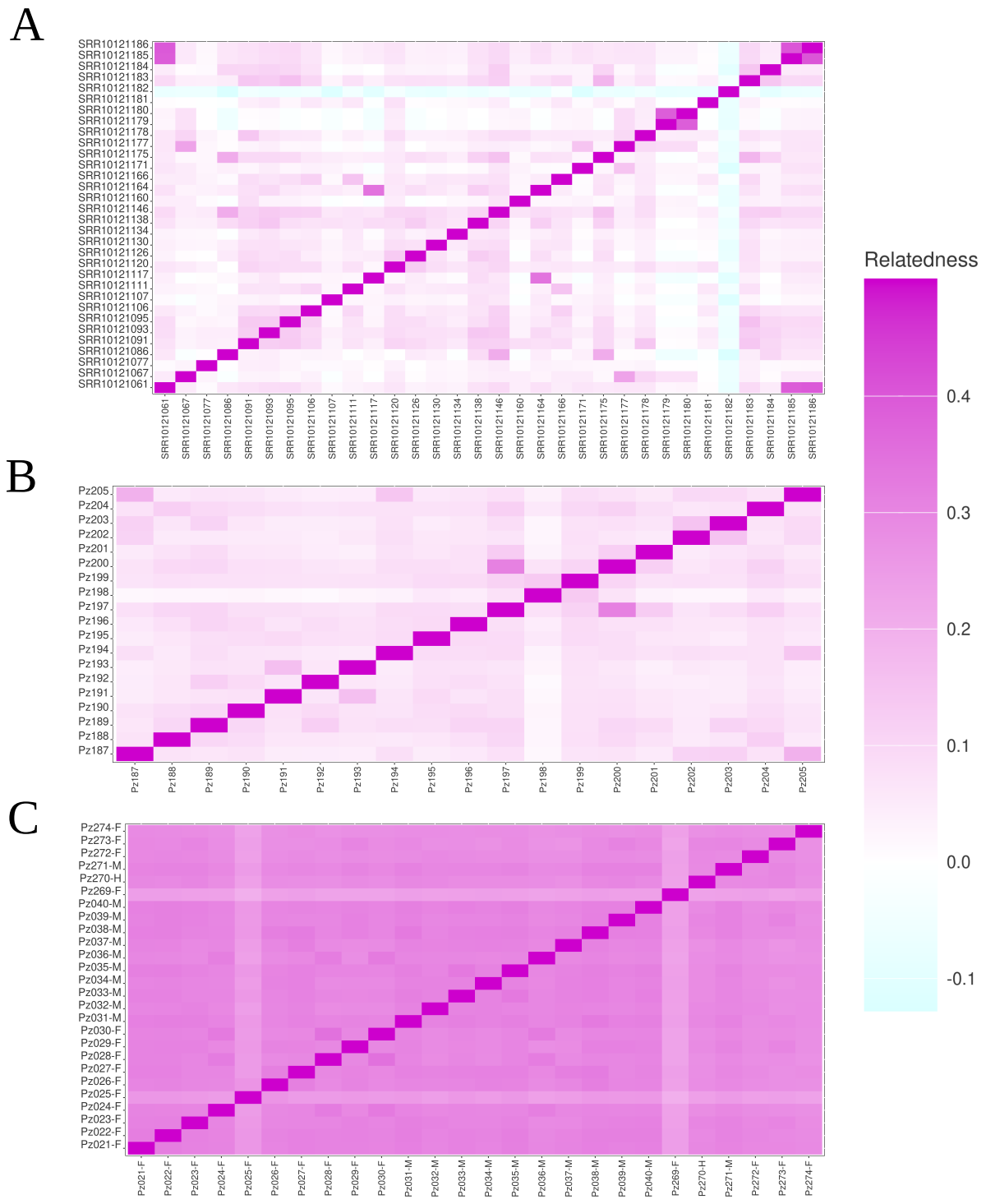
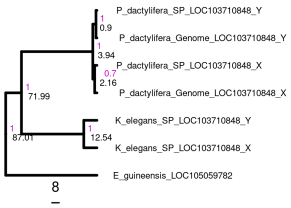
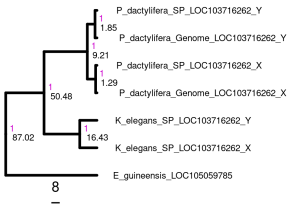


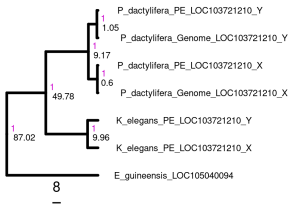
Fig. S2.

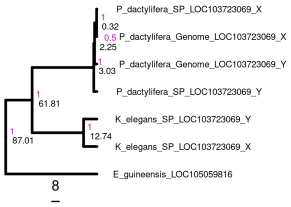
Heatmap of estimated relatedness between individuals. A relatedness of 0 is represented by a white square and indicates unrelated individuals from the same population. A relatedness of 0.5

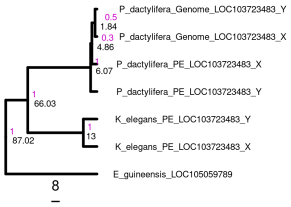
(the maximum value) is represented with a magenta square and is characteristic of clones. A cyan square indicates negative relatedness as one would expect between different populations. Gradients between these colors represent intermediate values. **(A)** Heatmap of the relatedness between *P. dactylifera* whole genome individuals. (B) Heatmap of the relatedness between *P. dactylifera* individuals in the data captured with PalmExons kit. (C) Heatmap of the relatedness between *K. elegans* individuals in the data captured with PalmExons kit.

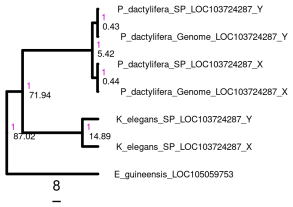

P_dactylifera_SP_LOC103710848_Y
P_dactylifera_Genome_LOC103710848_Y
P_dactylifera_SP_LOC103710848_X
P_dactylifera_Genome_LOC103710848_X
K_elegans_SP_LOC103710848_Y
K_elegans_SP_LOC103710848_X
E_guineensis_LOC105059782
1
0.9
3.94
0.7
2.16
71.99
12.54
87.01
8

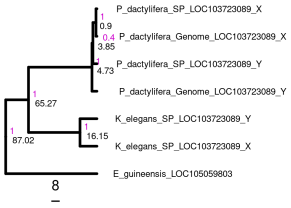

P_dactylifera_SP_LOC103716262_Y
P_dactylifera_Genome_LOC103716262_Y
P_dactylifera_SP_LOC103716262_X
P_dactylifera_Genome_LOC103716262_X
K_elegans_SP_LOC103716262_Y
K_elegans_SP_LOC103716262_X
E_guineensis_LOC105059785
1
1.85
9.21
1.29
50.48
16.43
87.02
8

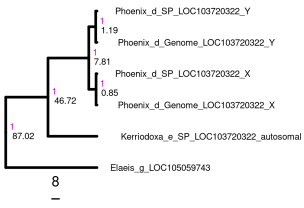

P_dactylifera_PE_LOC103721210_Y
P_dactylifera_Genome_LOC103721210_Y
P_dactylifera_PE_LOC103721210_X
P_dactylifera_Genome_LOC103721210_X
K_elegans_PE_LOC103721210_Y
K_elegans_PE_LOC103721210_X
E_guineensis_LOC105040094
1
1.05
9.17
0.6
49.78
9.96
87.02
8

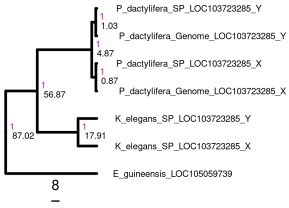

P_dactylifera_SP_LOC103723069_X
P_dactylifera_Genome_LOC103723069_X
P_dactylifera_Genome_LOC103723069_Y
P_dactylifera_SP_LOC103723069_Y
K_elegans_SP_LOC103723069_Y
K_elegans_SP_LOC103723069_X
E_guineensis_LOC105059816
1
0.32
0.5
2.25
3.03
61.81
12.74
87.01
8

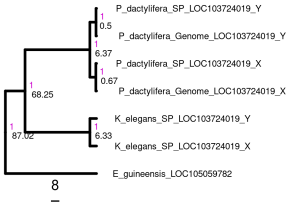

P_dactylifera_Genome_LOC103723483_Y
P_dactylifera_Genome_LOC103723483_X
P_dactylifera_PE_LOC103723483_X
P_dactylifera_PE_LOC103723483_Y
K_elegans_PE_LOC103723483_Y
K_elegans_PE_LOC103723483_X
E_guineensis_LOC105059789
0.5
1.84
0.3
4.86
6.07
66.03
13
87.02
8

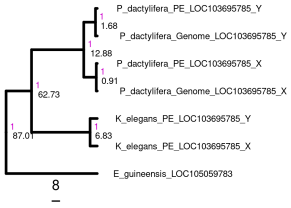

P_dactylifera_SP_LOC103724287_Y
P_dactylifera_Genome_LOC103724287_Y
P_dactylifera_SP_LOC103724287_X
P_dactylifera_Genome_LOC103724287_X
K_elegans_SP_LOC103724287_Y
K_elegans_SP_LOC103724287_X
E_guineensis_LOC105059753
1
0.43
5.42
0.44
71.94
14.89
87.02
8

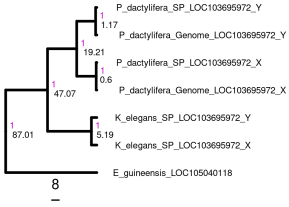

P_dactylifera_SP_LOC103723089_X
P_dactylifera_Genome_LOC103723089_X
P_dactylifera_SP_LOC103723089_Y
P_dactylifera_Genome_LOC103723089_Y
K_elegans_SP_LOC103723089_Y
K_elegans_SP_LOC103723089_X
E_guineensis_LOC105059803
1
0.9
0.4
3.85
4.73
65.27
16.15
87.02
8

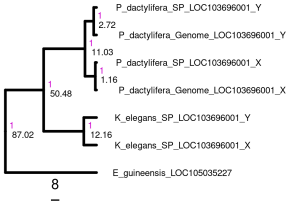

Phoenix_d_SP_LOC103720322_Y
Phoenix_d_Genome_LOC103720322_Y
Phoenix_d_SP_LOC103720322_X
Phoenix_d_Genome_LOC103720322_X
Kerriodoxa_e_SP_LOC103720322_autosomal
Elaeis_g_LOC105059743
1
1.19
7.81
0.85
46.72
87.02
8

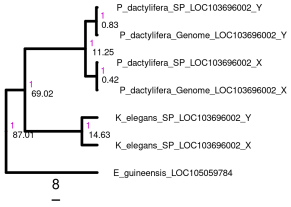

P_dactylifera_SP_LOC103723285_Y
1.03
P_dactylifera_Genome_LOC103723285_Y
4.87
P_dactylifera_SP_LOC103723285_X
0.87
P_dactylifera_Genome_LOC103723285_X
56.87
K_elegans_SP_LOC103723285_Y
87.02
17.91
K_elegans_SP_LOC103723285_X
E_guineensis_LOC105059739
8

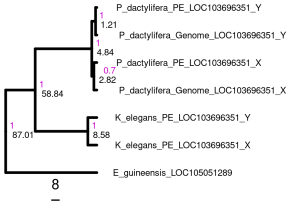

P_dactylifera_SP_LOC103724019_Y
0.5
P_dactylifera_Genome_LOC103724019_Y
6.37
P_dactylifera_SP_LOC103724019_X
0.67
P_dactylifera_Genome_LOC103724019_X
68.25
K_elegans_SP_LOC103724019_Y
87.02
6.33
K_elegans_SP_LOC103724019_X
E_guineensis_LOC105059782
8

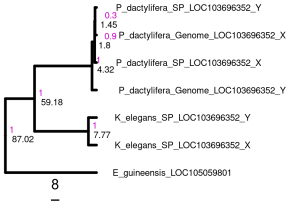

P_dactylifera_PE_LOC103695785_Y
1.68
P_dactylifera_Genome_LOC103695785_Y
12.88
P_dactylifera_PE_LOC103695785_X
0.91
P_dactylifera_Genome_LOC103695785_X
62.73
K_elegans_PE_LOC103695785_Y
87.01
6.83
K_elegans_PE_LOC103695785_X
E_guineensis_LOC105059783
8

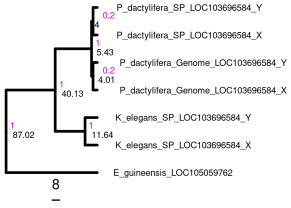

P_dactylifera_SP_LOC103695972_Y
1.17
P_dactylifera_Genome_LOC103695972_Y
19.21
P_dactylifera_SP_LOC103695972_X
0.6
P_dactylifera_Genome_LOC103695972_X
47.07
K_elegans_SP_LOC103695972_Y
87.01
5.19
K_elegans_SP_LOC103695972_X
E_guineensis_LOC105040118
8

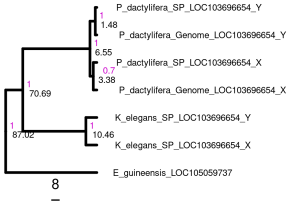

P_dactylifera_SP_LOC103696001_Y
2.72
P_dactylifera_Genome_LOC103696001_Y
11.03
P_dactylifera_SP_LOC103696001_X
1.16
P_dactylifera_Genome_LOC103696001_X
50.48
K_elegans_SP_LOC103696001_Y
87.02
12.16
K_elegans_SP_LOC103696001_X
E_guineensis_LOC105035227
8

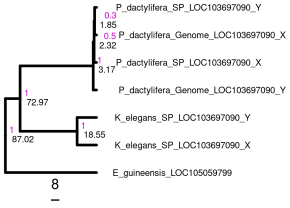

P_dactylifera_SP_LOC103696002_Y
0.83
P_dactylifera_Genome_LOC103696002_Y
11.25
P_dactylifera_SP_LOC103696002_X
0.42
P_dactylifera_Genome_LOC103696002_X
69.02
K_elegans_SP_LOC103696002_Y
87.01
14.63
K_elegans_SP_LOC103696002_X
E_guineensis_LOC105059784
8

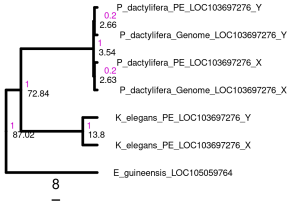

P_dactylifera_PE_LOC103696351_Y
1.21
P_dactylifera_Genome_LOC103696351_Y
4.84
P_dactylifera_PE_LOC103696351_X
0.7
2.82
P_dactylifera_Genome_LOC103696351_X
58.84
K_elegans_PE_LOC103696351_Y
87.01
8.58
K_elegans_PE_LOC103696351_X
E_guineensis_LOC105051289
8

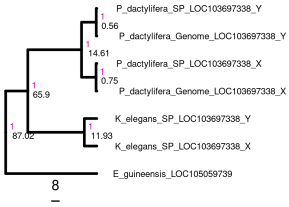

P_dactylifera_SP_LOC103696352_Y
0.3
1.45
0.9
P_dactylifera_Genome_LOC103696352_X
1.8
4.32
P_dactylifera_SP_LOC103696352_X
P_dactylifera_Genome_LOC103696352_Y
59.18
K_elegans_SP_LOC103696352_Y
87.02
7.77
K_elegans_SP_LOC103696352_X
E_guineensis_LOC105059801
8

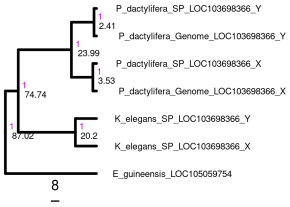

P_dactylifera_SP_LOC103696584_Y
0.2
4
P_dactylifera_SP_LOC103696584_X
5.43
P_dactylifera_Genome_LOC103696584_Y
0.2
4.01
P_dactylifera_Genome_LOC103696584_X
40.13
K_elegans_SP_LOC103696584_Y
87.02
11.64
K_elegans_SP_LOC103696584_X
E_guineensis_LOC105059762
8

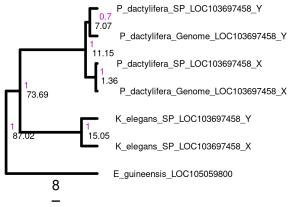

P_dactylifera_SP_LOC103696654_Y
1.48
P_dactylifera_Genome_LOC103696654_Y
6.55
P_dactylifera_SP_LOC103696654_X
0.7
3.38
P_dactylifera_Genome_LOC103696654_X
70.69
K_elegans_SP_LOC103696654_Y
87.02
10.46
K_elegans_SP_LOC103696654_X
E_guineensis_LOC105059737
8

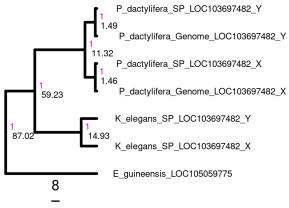

P_dactylifera_SP_LOC103697090_Y
0.3
1.85
0.5
P_dactylifera_Genome_LOC103697090_X
2.32
P_dactylifera_SP_LOC103697090_X
3.17
P_dactylifera_Genome_LOC103697090_Y
72.97
K_elegans_SP_LOC103697090_Y
87.02
18.55
K_elegans_SP_LOC103697090_X
E_guineensis_LOC105059799
8

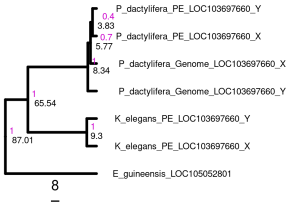

P_dactylifera_PE_LOC103697276_Y
0.2
2.66
P_dactylifera_Genome_LOC103697276_Y
3.54
P_dactylifera_PE_LOC103697276_X
0.2
2.63
P_dactylifera_Genome_LOC103697276_X
72.84
K_elegans_PE_LOC103697276_Y
87.02
13.8
K_elegans_PE_LOC103697276_X
E_guineensis_LOC105059764
8

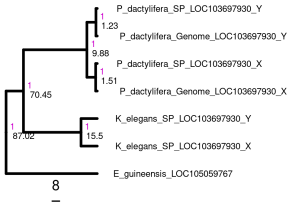

P_dactylifera_SP_LOC103697338_Y
0.56
P_dactylifera_Genome_LOC103697338_Y
14.61
P_dactylifera_SP_LOC103697338_X
0.75
P_dactylifera_Genome_LOC103697338_X
65.9
K_elegans_SP_LOC103697338_Y
87.02
11.93
K_elegans_SP_LOC103697338_X
E_guineensis_LOC105059739
8


P_dactylifera_SP_LOC103698366_Y
2.41
P_dactylifera_Genome_LOC103698366_Y
23.99
P_dactylifera_SP_LOC103698366_X
3.53
P_dactylifera_Genome_LOC103698366_X
74.74
K_elegans_SP_LOC103698366_Y
87.02
20.2
K_elegans_SP_LOC103698366_X
E_guineensis_LOC105059754
8


P_dactylifera_SP_LOC103697458_Y
0.7
7.07
P_dactylifera_Genome_LOC103697458_Y
11.15
P_dactylifera_SP_LOC103697458_X
1.36
P_dactylifera_Genome_LOC103697458_X
73.69
K_elegans_SP_LOC103697458_Y
87.02
15.05
K_elegans_SP_LOC103697458_X
E_guineensis_LOC105059800
8


P_dactylifera_SP_LOC103697482_Y
1.49
P_dactylifera_Genome_LOC103697482_Y
11.32
P_dactylifera_SP_LOC103697482_X
1.46
P_dactylifera_Genome_LOC103697482_X
59.23
K_elegans_SP_LOC103697482_Y
87.02
14.93
K_elegans_SP_LOC103697482_X
E_guineensis_LOC105059775
8


P_dactylifera_PE_LOC103697660_Y
0.4
3.83
0.7
P_dactylifera_PE_LOC103697660_X
5.77
P_dactylifera_Genome_LOC103697660_X
8.34
P_dactylifera_Genome_LOC103697660_Y
65.54
K_elegans_PE_LOC103697660_Y
87.01
9.3
K_elegans_PE_LOC103697660_X
E_guineensis_LOC105052801
8


P_dactylifera_SP_LOC103697930_Y
1.23
P_dactylifera_Genome_LOC103697930_Y
9.88
P_dactylifera_SP_LOC103697930_X
1.51
P_dactylifera_Genome_LOC103697930_X
70.45
K_elegans_SP_LOC103697930_Y
87.02
15.5
K_elegans_SP_LOC103697930_X
E_guineensis_LOC105059767
8

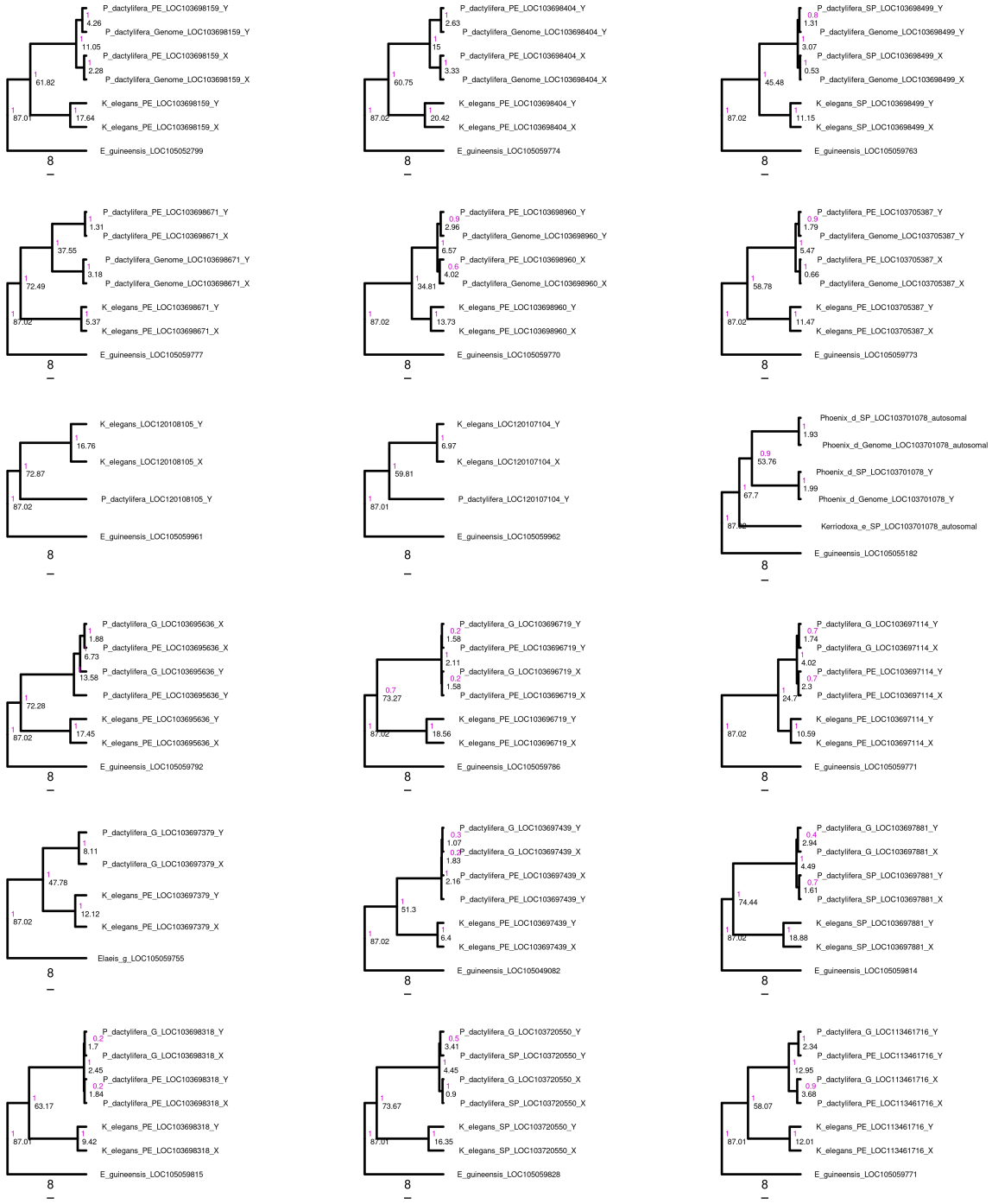


Fig. S3.

Phylogenies of sex-linked genes. Phylogenies of all sex-linked genes common to *K. elegans* and *P. dactylifera* and two known sex-linked genes in *P. dactylifera* : , LOGlike (LOC103701078) and guanosine deaminase (LOC103720322). Two Y-specific sequences of *P. dactylifera* were taken from NCBI : GPAT3 (LOC120108105, MH668906), CYP703 (LOC120107104, MH668888). The node values are the posterior probabilities of the node colored in magenta and the mean age of the node in black. The scales below the trees are in million years.

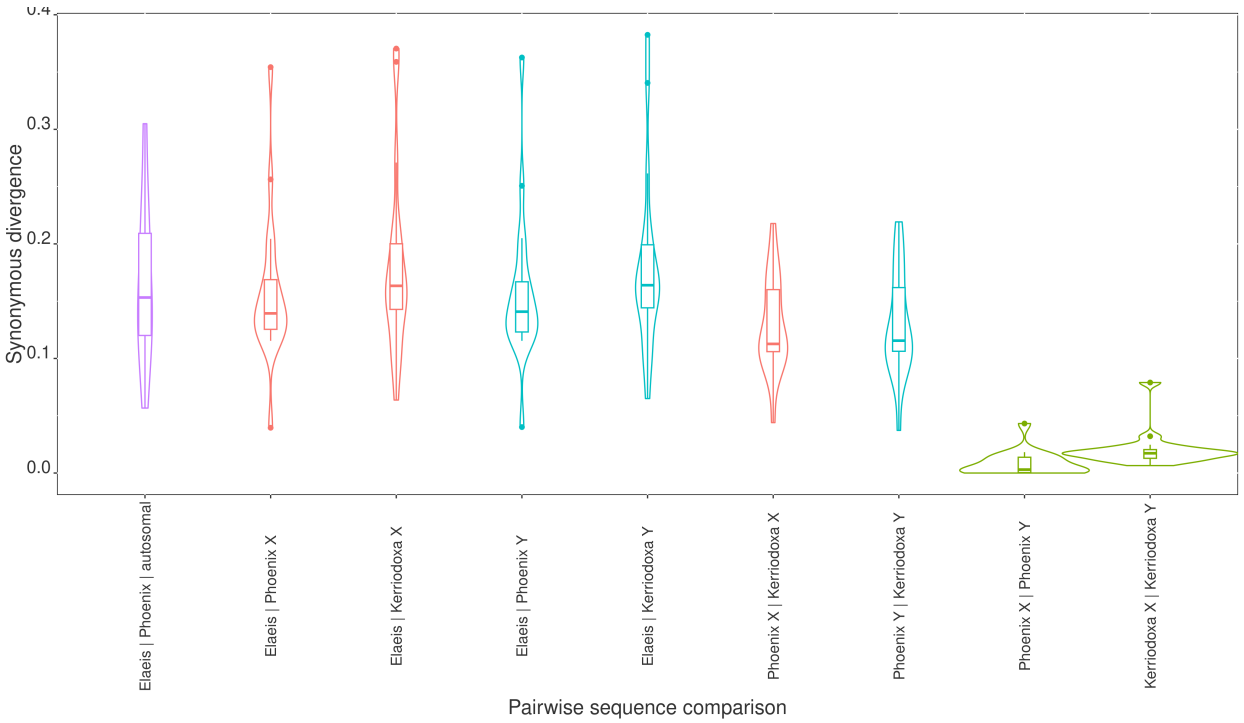


Fig. S4.

Synonymous divergence of sex-linked and autosomal nuclear genes. Violin plots and boxplots of the synonymous divergence (dS) of the 18 genes inferred as XY in SexPhoenix (SP) and 10 randomly selected autosomal single copy orthologs pertaining to the BUSCO *Liliopsidae* database. dS between X and Y sequences are colored in green, dS between X sequences from SP are colored in red, dS between Y sequences from SP are colored in cyan and dS between autosomal single copy orthologs are colored in purple.

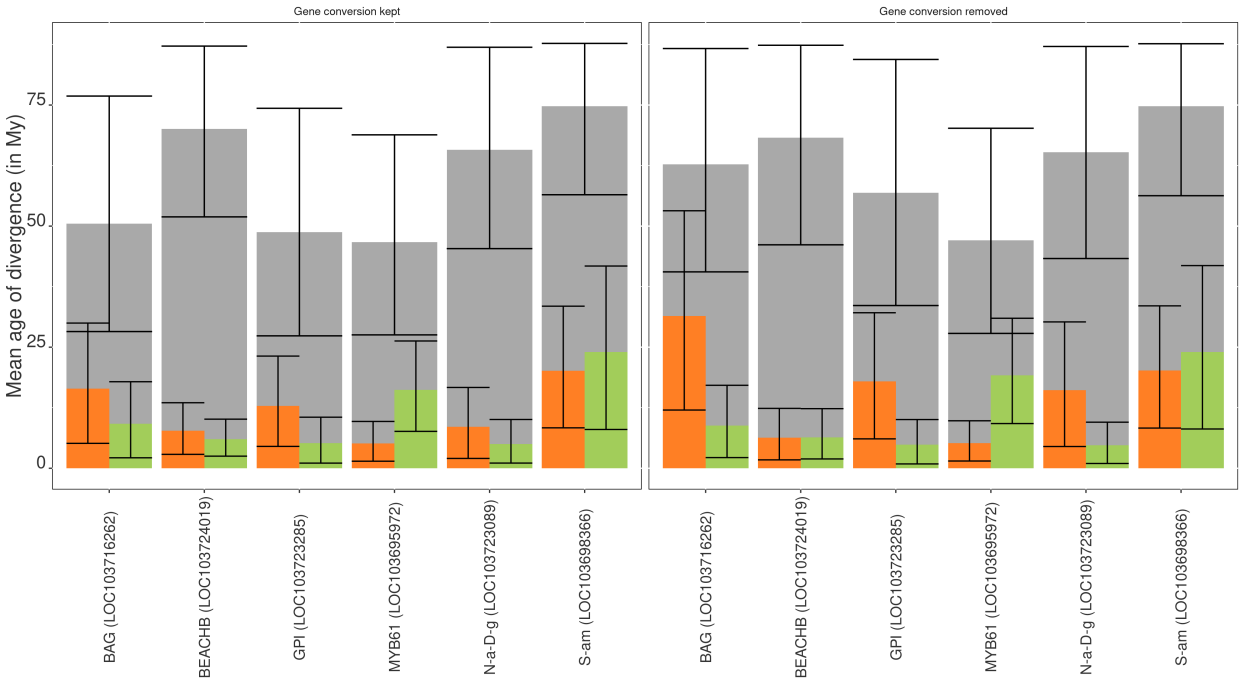


Fig. S5.

Effect of gene conversion between X and Y copies on node ages.

Mean age of divergence of X and Y sequences for common sex-linked genes with possible gene conversion. Gene fragments detected as having undergone gene conversion (see Table S5) were either retained in the analysis (left panel) or removed (right panel). *P. dactylifera* sex-linked sequence divergence estimates are colored in green, estimated for *K. elegans* in orange, and the estimates of the age of the split of the lineages between the two species in grey. Error bars correspond to the 95% probability density of the node age.

Sample name	Reads (R1 + R2)	Mapped properly paired	Discarded
SRR10121061	69906798	93.41%	Yes
SRR10121067	79422680	94.90%	Yes
SRR10121077	94044556	95.49%	No
SRR10121086	70564142	94.67%	Yes
SRR10121091	76352168	92.67%	No
SRR10121093	99377344	93.34%	No
SRR10121095	94123466	95.07%	No
SRR10121106	94660272	96.39%	No
SRR10121107	61350290	92.63%	No
SRR10121111	81614162	96.86%	No
SRR10121117	74139608	96.28%	Yes
SRR10121120	87950120	93.51%	No
SRR10121126	78984780	95.37%	No
SRR10121130	60753842	92.48%	No
SRR10121134	92117304	94.42%	No
SRR10121138	89963210	93.65%	No
SRR10121146	88845434	93.90%	No
SRR10121160	69178418	93.86%	No
SRR10121164	77056038	94.57%	No
SRR10121166	88559374	93.80%	No
SRR10121171	88773146	94.72%	No
SRR10121175	82661366	94.40%	No
SRR10121177	80185434	93.98%	No
SRR10121178	69218370	93.77%	No

Table S1.

Mapping of the samples of the *P. dactylifera* WGS dataset. For each sample, the number of reads (forward, R1, and backward, R2) and the proportion of properly paired mapped reads is given. Related individuals (see Fig. S2A) that were discarded from the analysis are indicated.

Sample	Species	Reads (R1 + R2)		Mapped properly paired		discarded
		SP	PE	SP	PE	
Pz187-F	*P. dactylifera*	4377416	18632428	99.67%	99.84%	No
Pz188-F	*P. dactylifera*	5008768	18696304	99.61%	99.85%	No
Pz189-F	*P. dactylifera*	2483626	9135350	99.49%	99.75%	No
Pz190-F	*P. dactylifera*	4385870	17747354	99.65%	99.83%	No
Pz191-F	*P. dactylifera*	2939824	10817454	99.62%	99.76%	No
Pz192-F	*P. dactylifera*	3127450	10421980	99.47%	99.41%	No
Pz193-F	*P. dactylifera*	1985378	6219258	99.51%	99.76%	No
Pz194-F	*P. dactylifera*	3061202	11576270	99.54%	99.63%	No
Pz195-F	*P. dactylifera*	2594946	9306702	99.53%	99.63%	No
Pz196-M	*P. dactylifera*	2627820	10596866	99.47%	99.69%	No
Pz197-M	*P. dactylifera*	3694942	16134226	99.51%	99.76%	No
Pz198-M	*P. dactylifera*	2281358	9850922	98.30%	96.05%	No
Pz199-M	*P. dactylifera*	2078630	9343366	99.49%	99.77%	No
Pz200-M	*P. dactylifera*	4513112	22881886	99.41%	99.29%	Yes
Pz201-M	*P. dactylifera*	2016912	11571982	99.44%	99.73%	No
Pz202-M	*P. dactylifera*	1772086	7567128	99.45%	99.70%	No
Pz203-M	*P. dactylifera*	1941712	10554778	99.45%	99.69%	No
Pz204-M	*P. dactylifera*	3704796	20085258	99.55%	99.83%	No
Pz205-M	*P. dactylifera*	3084618	14996024	99.57%	99.79%	No
Pz021-F	*K. elegans*	666270	16383618	84.20%	99.24%	No
Pz022-F	*K. elegans*	752908	17161294	83.45%	99.21%	No
Pz023-F	*K. elegans*	771894	17218782	82.95%	99.17%	No
Pz024-F	*K. elegans*	738372	15364196	83.78%	99.09%	No
Pz025-F	*K. elegans*	874352	19668318	83.90%	99.32%	No
Pz026-F	*K. elegans*	603370	13660796	83.50%	99.17%	No
Pz027-F	*K. elegans*	576916	13522682	84.68%	99.16%	No
Pz028-F	*K. elegans*	752862	18841052	85.51%	99.24%	No
Pz029-F	*K. elegans*	926680	24343416	85.12%	99.40%	No
Pz030-F	*K. elegans*	733968	17460034	84.45%	99.28%	No
Pz031-M	*K. elegans*	789534	18569286	85.17%	99.28%	No
Pz032-M	*K. elegans*	439660	9782804	83.12%	99.23%	No
Pz033-M	*K. elegans*	677530	15510204	84.48%	99.42%	No
Pz034-M	*K. elegans*	500752	11488268	83.08%	98.93%	No
Pz035-M	*K. elegans*	709868	17164758	85.20%	99.27%	No
Pz036-M	*K. elegans*	575470	12718802	83.73%	99.17%	No

Pz037-M	*K. elegans*	623986	13381650	85.05%	98.97%	No
Pz038-M	*K. elegans*	691268	15951318	83.54%	99.24%	No
Pz039-M	*K. elegans*	942786	21623126	84.48%	99.29%	No
Pz040-M	*K. elegans*	756512	18298996	85.16%	99.39%	No
Pz269-F	*K. elegans*	578186	12884964	83.44%	98.83%	No
Pz270-H	*K. elegans*	595258	14863394	83.04%	99.06%	Yes
Pz271-M	*K. elegans*	755654	19055042	84.54%	99.39%	No
Pz272-F	*K. elegans*	508120	11612810	81.86%	99.03%	No
Pz273-F	*K. elegans*	1014112	26035728	82.67%	99.10%	No
Pz274-F	*K. elegans*	607980	12829890	82.60%	99.20%	No

Table S2.

Mapping and relatedness of the capture sequences of Sex Phoenix (SP) and Palm Exons (PE). The number of reads and proportion of reads mapped and properly-paired is given. Related individuals that were discarded from the analysis are indicated.

	P. dactylifera			*K. elegans*	
	Whole Genome	SP	PE	SP	PE
DPV01 genes covered (%)	97.61	1.60	77.06	2.54	58.31
Median covered length	99.95	86.08	52.14	60.00	70.56
[1st - 3rd quartile] (%)	[99.63 - 1]	[57.07 – 99.94]	[34.52 – 74.07]	[28.76 – 93.07]	[54.32 – 85.83]
Median depth per position (females)	11.03	39.01	22.43	19.89	32.50
[1st - 3rd quartile]	[10.07 – 11.74]	[11.62 – 93.24]	[8.51 – 48.42]	[1.86 – 82.25]	[9.50 – 87.91]
Median depth per position (males)	11.76	36.07	19.36	18.93	31.18
[1st - 3rd quartile]	[10.85 – 12.40]	[15.01 – 80.80]	[7.70 – 40.75]	[1.88 – 78.84]	[9.10 – 84.67]
Number of variant positions analyzed	466719	4693	180035	8395	161384

Table S3.

Depth and coverage of the DPV01 genome by our different datasets. Percentage of annotated genes covered, median, first and third quartile of the length proportion of the genes covered and of the mean depth per position per gene and number of variant positions analyzed.

SDpop model	BIC	Autosomal	Haploid	Paralog	hemizygote	gametolog
Phoenix dactylifera Whole Genome Sequencing						
XY	**11064321.61**	0.9124	0.0019	0.0791	0.0042	0.0023
Phoenix dactylifera SexPhoenix						
XY	**85066.94**	0.8375	0.0029	0.0403	0.0e-6	0.1193
Phoenix dactylifera PalmExons						
XY	**2839713.34**	0.9537	0.0047	0.0387	0.0001	0.0028
Kerriodoxa elegans SexPhoenix						
XY	**113544.76**	0.7420	0.0016	0.1208	8.0e-6	0.1356
ZW	123130.07	0.8010	0.0009	0.1212	0.0419	0.0349
No sex chromosomes	124038.66	0.8766	0.0014	0.1221		
Kerriodoxa elegans PalmExons						
XY	**2642686.41**	0.6481	0.0031	0.3408	0.0003	0.0077
ZW	123130.07	0.6558	0.0031	0.3409	0.0e-6	0.0001
No sex chromosomes	2649680.07	0.6560	0.0031	0.3409		

Table S4.

SDpop model analysis summary. All models available are documented for both *K. elegans* dataset whereas only XY is documented for *P. dactylifera* datasets. The Bayesian Information Criterion (BIC) is the lowest for the model that best explains the data. The inferred proportion of sites for each segregation type is given.

Gene	XY paired sequences	Simulated p-value	KA p-value	start position	end position
LOC103695972	P. dactylifera Genome	0.0143	0.03123	1013	1214
LOC103695972	P. dactylifera SP	0.0133	0.03123	1013	1214
LOC103698366	P. dactylifera Genome	0.0176	0.08899	73	716
LOC103698366	P. dactylifera SP	0.0059	0.04848	73	716
LOC103716262	K. elegans SP	0.0146	0.08210	250	699
LOC103723089	K. elegans SP	0.024	0.13872	448	1061
LOC103724019	P. dactylifera Genome	0.021	0.07904	3299	4064
LOC103724019	P. dactylifera SP	0.034	0.1082	3299	4064
LOC103723285	K. elegans SP	0.0031	0.02272	686	824

Table S5.

XY sequences for which gene conversion was detected. The sequences of the genes that underwent gene conversion between the X and Y copies are described here with the simulated and KA p-values as well as the starting and ending position of the alignment which underwent gene conversion.

Data S1. (separate file)

Summary of all genes either inferred as XY in at least one dataset or aligned on the EGPMv6 genome (*19*) on chromosome 10 between 43Mb and 46.1Mb. Each gene identification and annotation is documented as well as their position on the Barhee BC4 genome (*10*) and EGPMv6 (*19*). This table contains the mean female and male depth per base as well as the mean female reads ratio, the number of polymorphisms and mean number of individual genotyped per polymorphism as well as the autosomal and XY posterior probability of each of these genes.

Data S2. (separate file)

Summary of all genes inferred as X-hemizygous in at least one dataset. Each sheet contains the X-hemizygous genes of one dataset. Since no X-hemizygous genes were found in any SexPhoenix dataset, no sheet is present in the file for these datasets. Each gene identification and annotation is documented as well as their position on the Barhee BC4 genome (*10*) and EGPMv6 (*19*). This table contains the mean female and male depth per base as well as the mean female reads ratio, the number of polymorphisms and mean number of individual genotyped per polymorphism as well as the X-hemizygous posterior probability of each of these genes.