Evolution of sex chromosomes prior to speciation in the dioecious Phoenix species

E. Cherif, S. Zehdi-Azouzi, A. Crabos, K. Castillo, N. Chabrillange, J.-C. Pintaud, A. Salhi-Hannachi, S. Glémin, F. Aberlenc-Bertossi

▶ **To cite this version:**

E. Cherif, S. Zehdi-Azouzi, A. Crabos, K. Castillo, N. Chabrillange, et al.. Evolution of sex chromosomes prior to speciation in the dioecious Phoenix species. Journal of Evolutionary Biology, 2016, 29 (8), pp.1513-1522. 10.1111/jeb.12887 . hal-03038938

HAL Id: hal-03038938
https://hal.umontpellier.fr/hal-03038938v1

Submitted on 12 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 **Evolution of sex chromosomes is prior to speciation in the dioecious *Phoenix* species.**

2 Emira Cherif[*,†,1], Salwa Zehdi-Azouzi[†], Amandine Crabos[*], Karina Castillo[*], Nathalie
3 Chabrillange[*], Jean-Christophe Pintaud[‡], Amel Salhi-Hannachi[†], Sylvain Glémin[§,**] and
4 Frédérique Aberlenc-Bertossi[*]
5
6 [*]IRD/CIRAD F2F-palmiers Group, UMR DIADE, Centre IRD, 911 avenue Agropolis, 34394,
7 Montpellier, France;
8 [†]Laboratoire de Génétique Moléculaire, Immunologie et Biotechnologie, Faculté des sciences
9 de Tunis, Université Tunis El Manar, Campus Universitaire, 2092, El Manar, Tunisia;
10 [‡]Dynadiv, UMR DIADE, Centre IRD, 911 avenue Agropolis, 34394, Montpellier, France;
11 [§]Institut des Sciences de l'Evolution de Montpellier, Unité Mixte de Recherche 5554
12 (Université de Montpellier-CNRS-IRD-EPHE), 34095, Montpellier, France
13 ** Department of Ecology and Genetics, Evolutionary Biology Centre, Uppsala University,
14 SE-752 36 Uppsala, Sweden.
15
16
17
18
19
20
21
22
23
24
25

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

26

27

28 **Sex chromosome evolution in *Phoenix***

29

30 Key words:

31 sex-linked gene, sex chromosomes, recombination arrest, dioecy, speciation.

32

33 [1]Corresponding author:

34 Emira Cherif

35 IRD/CIRAD F2F-palmiers Group, UMR DIADE, Centre IRD, 911 avenue Agropolis, 34394,

36 Montpellier, France;

37 +33467416315

38 e-mail: emira.cherif@ird.fr ; emira.cherif@yahoo.fr

39

40

41

42

43

44

45

46

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

47

48 **Abstract**

49 Understanding the driving forces and molecular processes underlying dioecy and sex
50 chromosome evolution, leading from hermaphroditism to the occurrence of male and female
51 individuals, is of considerable interest in fundamental and applied research. The genus
52 *Phoenix*, belonging to the family Arecaceae, consists of only dioecious species. Phylogenetic
53 data suggests that the genus *Phoenix* diverged from a hermaphroditic ancestor shared with its
54 closest relatives. Here we investigated the evolution of suppressed recombination within the
55 genus *Phoenix* as a whole by extending the analysis of *P. dactylifera* sex-related loci to eight
56 other species within the genus. We also performed a phylogenetic analysis of a date palm sex-
57 linked *PdMYB1* gene in these species. We found that X and Y sex-linked alleles clustered in a
58 species-independent fashion. Our data show that sex chromosomes evolved before the
59 diversification of the extant dioecious species. Furthermore, the distribution of Y haplotypes
60 revealed two male ancestral paternal lineages which may have emerged prior to speciation.

61

62 **1. Introduction**

63 Dioecy, in angiosperms, may result from two sex-determining mutations, a recessive male-
64 sterility mutation and a mutation at a linked locus causing the loss of female functions [1]. If
65 both these partially linked mutations establish polymorphisms [1, 2], closer linkage between
66 the two sex-determination loci is favored by selection, to maintain the correct combination of
67 mutations, and avoid sterile recombinants. This may explain the suppressed recombination
68 that characterizes sex chromosomes. The time since recombination stopped in the sex-
69 determining region defines the age of the sex chromosome system, and it is also of interest to
70 know whether a single recombination suppression occurred, or multiple events, resulting in

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

71 several evolutionary strata having evolved, as has occurred in mammals, *Silene latifolia*, and
72 *Papaya carica* [3, 4]. Dioecy and sex chromosomes have evolved repeatedly and
73 independently in different plant taxa [5, 6, 7]. However, only a few systems have been
74 described in detail. Understanding sex chromosome emergence during the evolution of
75 dioecy, leading from hermaphroditism to the occurrence of male and female sterile
76 individuals, is therefore of major fundamental interest, with many potential agronomic
77 applications.
78 The genus *Phoenix* (Arecaceae, Coryphoideae, Phoeniceae) includes fourteen dioecious
79 species, distributed from the Atlantic islands throughout the Mediterranean region, Africa,
80 Middle East, and as far as southern Asia to the northwestern Pacific [8,9,10].
81 DNA sequence divergence from other palm genera is high [11, 12], and it has been suggested
82 that the genus *Phoenix* might possess an ancient sex chromosome system [20]. The tribe
83 Phoeniceae is sister to the predominantly hermaphroditic tribe Trachycarpeae [13], but is
84 distinguished by several morphological differences [10], and the divergence time is estimated
85 to be around 49 ± 16 mya [14]. Assuming a single origin of dioecy in *Phoenix*, this date gives
86 an upper bound to the age of the sex-linked non-recombining region. Interspecific
87 relationships within *Phoenix* were studied by Pintaud *et al*. [15] based on two chloroplast loci
88 (psbZ-trnfM and rpl16-rps3), recovering five phylogenetic lineages, namely *P. loureiroi-*
89 *acaulis-pusilla*, *P. roebelenii-paludosa*, *P. caespitosa*, *P. reclinata* and a larger lineage
90 consisting of *P. dactylifera, P. atlantica, P. theophrasti, P. sylvestris* and *P. rupicola*.
91 Sexual dimorphism in the genus *Phoenix* has been dated back to the Eocene period (between
92 33.9 and 55.8 million years ago) [10, 16] on the basis of fossil records of *Phoenix* male
93 flowers. Dioecy could thus be very ancient within the genus. In *P. dactylifera*, sex
94 differentiation results from the arrest of male or female organ development in the initial
95 bisexual flower buds [17] and the species has an XY sex chromosome system [18,19,20]. A

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

96 non-recombining XY-like region was inferred in the date palm genome, based on 3
97 microsatellite loci showing alleles confined to males, and two different Y haplogroups were
98 found [20]. Recently, Mathew *et al*. [21] constructed a genetic map of date palm and localized
99 the sex segregating region to LG12. The physical length of this region is estimated to be
100 13Mb, about 2% of the genome [21].

101 All known species in the genus *Phoenix* are dioecious. Dioecy is probably an ancestral
102 character in the genus. However, it is important to test explicitly whether sex chromosomes
103 evolved before speciation within the genus, to exclude the possibility that dioecy evolved in
104 more than one lineage, and to test whether suppressed recombination might have evolved in
105 only certain lineages. If X and Y alleles of different species cluster together, rather than by
106 clustering within their respective species, sex linkage must have evolved before speciation.
107 Conversely, if these alleles cluster according to species, then sex linkage must have evolved
108 after speciation [22].

109 We used sex-linked markers identified in *P. dactylifera* [20] to study eight other species, and
110 found that one sex-linked MYB gene, PdMYB1, was present in seven of the studied species.
111 Our results provide strong evidence that sex evolved before the appearance of the extant
112 species of the genus *Phoenix*.

113 2. Material and Methods

114 (a) Plant material

115 Nine of the 14 *Phoenix* species were studied, mainly from natural populations (including 64
116 males and 70 females). Three species with large samples in our study are widespread: *Phoenix*
117 *dactylifera* (34), *P. reclinata* (10), *P. sylvestris* (18), while three species have restricted
118 distribution: *P. atlantica* (17), *P. canariensis* (21) and *P. roebelenii* (24) were included
119 (Figure 1), and three other species only small samples, *P. acaulis*, *P. rupicola* (one male and

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

120 one female) and *P. loureroi* (three males and three females) (Figure 1); these three species
121 were excluded from the statistical analyses (electronic supplementary material, Table 1).

122 **(b) DNA extraction**

123 Leaf samples were freeze-dried for 72 h with an Alpha1-4LD Plus lyophilizer (Fisher
124 Scientific, Illkirch, France) and ground with a Tissue Lyser System (Qiagen). DNA extraction
125 was carried out using the Dneasy plant mini kit (Qiagen) according to the manufacturer's
126 instructions. All samples were adjusted to 10 ng.μl^{-1} concentration for subsequent analyses.

127 **(c) PCR assays**

128 To test the transferability of date palm sex-linked markers to the other *Phoenix* species, the
129 three SSR loci (mPdIRD50, mPdIRD52 and mPdIRD80) designed by Cherif *et al*. [20] were
130 tested across eight newly studied species and new genotypes of *P.dactylifera*.
131 PCR reactions were performed in an Eppendorf thermocycler (AG, Hamburg, Germany).
132 The reaction volume was 20 µl and contained 10 ng of genomic DNA, 2 µl of 10X reaction
133 buffer with 2 mM $MgCl_2$, 200 µM dNTPs, 1 U Taq polymerase, 10 pmol of the
134 fluorochrome-marked forward primer and of the reverse primer and MilliQ water. The PCR
135 reactions were carried out with following parameters: 5 min denaturation at 95°C, followed
136 by 35 cycles of 95°C for 30 s, 56°C for 30 s and 72°C for 45 s, and then a final elongation
137 step at 72°C for 10 min. PCR products were analysed using an ABI 3130XL Genetic
138 Analyzer (Applied BioSystems, Foster City, CA, USA). Allele sizes were scored with
139 GeneMapper software v4.0 (Applied BioSystems).

140 **(d) Cloning and sequencing of *PdMYB1* sequences**

141 In order to verify that the size conservation of sex-linked SSR across *Phoenix* species is not
142 due to size homoplasy, the mPdIRD80 locus was sequenced. In addition, potential genes
143 around the three date palm sex-linked markers [20] were searched in the published genome

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

144 assembly [19]. A *MYB* gene called *PdMYB1* was identified by BLASTn on GenBank database
145 near the mPdIRD80 locus.
146 Primers PdMYB1R1 or PdMYB1R2 and PdMYB1F5 or mPdIRD80F (electronic
147 supplementary material, Table 2) were used to amplify the target *MYB* gene from seven
148 *Phoenix* species (*P.dactylifera, P.reclinata, P.atlantica, P.sylvestris, P.roebelenii, P.rupicola,*
149 *P.canariensis*) (electronic supplementary material, Figure B). All sequences were amplified
150 from genomic DNA, using Taq polymerase (GoTaq G2 DNA Polymerase, Promega).
151 Reactions were performed in 20µL, with following final concentrations: 1X Green Buffer,
152 dNTPs 0.2mM each, primers 0.5µM each, GoTaq 1U/tube and 50ng DNA/tube. The PCR
153 conditions were as follows: one incubation at 95°C for 2 min; 30 cycles of: denaturing at
154 95°C for 30s, annealing at 57°C for 30s, and elongation at 72°C for 2 min; and a final
155 extension at 72°C for 5 min. PCR products were separated on 1% agarose gel and stained
156 with ethidium bromide. PCR products matching the target size under UV light were purified
157 using a PCR Clean-Up Kit (Wizard SV Gel and PCR Clean-Up System, Promega) and cloned
158 into pGEM-T Easy vector (pGEM-T Easy Vector System I, Promega) according to the
159 manufacturer's conditions. The ligation products were transformed into JM109-competent
160 cells (Bacterial Strain JM109, Glycerol Stock, Promega) and positive colonies were
161 confirmed by PCR using previous primers. Plasmids of positive colonies were isolated
162 (Wizard Plus SV Minipreps DNA Purification System Promega) and adjusted to a
163 concentration of 75ng/µl for sequencing using SP6 and T7 primers at Eurofins Genomics
164 (Germany). Sequence results in fasta format were assembled and analyzed using SeqMan Pro
165 software (Lasergene). X and Y alleles were respectively obtained from homozygous XX
166 female plants and from XY male plants. To identify western and eastern Y alleles, western
167 and eastern male alleles of *P. dactylifera* [20] were used as a reference.

168 **(e) Genetic analyses**

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

169 Much of our analysis used the GenAlEx 6.5 program [23, 24], including analyses of allele
170 frequencies and allele size distributions in male and female individuals within *Phoenix*
171 species, estimates of genetic variability, including values of Ho, He, F*is* and R_{st} values used to
172 characterize subdivision between males and females (electronic supplementary material,
173 Table 3, Table A, Table 4 and 5).

174 For the remainder of the study, each *Phoenix* species was considered as a population and the
175 subsequent genetic analyses focused on male and female, treated as separate groups in tests
176 for subdivision and differences in heterozygote frequencies. R_{st} was estimated according to
177 the stepwise mutation model [25] (Table 1), and the AMOVA procedure implemented in
178 GenAlEx 6.5 [23, 24] used microsatellite distance matrix data; significance was determined
179 by running 10 000 permutations (electronic supplementary material, Figure 1). Principal
180 component analysis (PCA) (electronic supplementary material, Table A) was performed using
181 the *dudi.pca* function, implemented in ade4 package for the R software v2.15.3 [26]
182 (electronic supplementary material, Figure A).

183 Expected heterozygote frequencies assuming Hardy–Weinberg genotype frequencies (He)
184 were compared with those observed (Ho) within males and females of *Phoenix* species using
185 the GenAlEx 6.5 program [23, 24].

186 Two different *Fis* estimates, namely Weir and Cockerham's estimate [27] and Robertson and
187 Hill's estimate [28] (electronic supplementary material, Table 5), were used to test whether
188 there was an excess of heterozygotes in the two groups (male and female) within each
189 *Phoenix* species. The second estimate has a lower variance under the null hypothesis [28]. To
190 determine the significance of *Fis*, P-values were calculated by the Markov chain method [29]
191 (electronic supplementary material, Table 6) using the Genepop 4.2 program [30, 31].

192 To infer X and Y haplotypes within *Phoenix* species, the alleles of the three sex-linked loci
193 (mPdIRDP80, mPdIRDP50 and mPdIRDP52) that were most frequently associated were

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

194 determined for each genotype in each species, using the EM algorithm implemented in the
195 PowerMarker program [32]. Additionally, the EM algorithm was used to resolve the genotype
196 of each male or female individual into the two most probable haplotypes (electronic
197 supplementary material, Note 1).
198 The Y-SSR haplotypes of the genus were used to construct a median joining haplotype
199 network [33] using the Network 4.6.1.1 program. We assumed a single-repeat mutation
200 model, and repeat variants were given a weighting of 10. To highlight the clustering of Y and
201 X alleles within the genus, a multiple alignment of *PdMYB1* genomic sequences from the
202 studied *Phoenix* species was performed using CLC Sequence Viewer v7.02 software (CLC
203 BIO) (electronic supplementary material, Figure C). The alignment was then used as an input
204 in MEGA6.0 [34] and a tree was produced, using coding and non-coding sequence, (gaps
205 excluded) by a Maximum Likelihood method under the Kimura 2-parameter model [35].
206 The latter model was chosen after comparison of the BIC scores (Bayesian Information
207 Criterion) between different models (electronic supplementary material, Table 7 and Figure
208 D). The same results were obtained by other Neighbor-joining and maximum parsimony.
209 The ortholog of *PdMYB1* in the *Elaeis guineensis* genome was used as an outgroup in
210 phylogeny reconstruction.

211 **3. Results and Discussion**

212 *Phoenix* species were chosen to represent the main evolutionary lineages on the basis of
213 previously established morphological, anatomical and molecular phylogenies [8, 15, 36]. The
214 geographical distribution of the nine *Phoenix* species analyzed is shown in Figure 1 [8, 9, 10].
215 The species studied are characterized by a wide and continuous distribution from the Canary
216 and Cape Verde islands in the Atlantic Ocean, reaching as far as Taiwan through Africa,
217 Madagascar, the Middle East, Pakistan and India. *P. canariensis* and *P. atlantica* are
218 respectively endemic to the Canary and the Cape Verde islands. *P. roebelenii*, *P. rupicola* and

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

219 *P. acaulis* have restricted ranges, respectively limited to the north of Laos and Vietnam and
220 the north of India and Nepal. *P. dactylifera* has the largest distribution area, which overlaps
221 that of *P. canariensis, P. reclinata* in the Horn of Africa and *P. sylvestris* on the Indian
222 subcontinent.

223 **(a) Conservation of the sex-linked loci within overall the studied *Phoenix* species**

224 Three sex-linked SSR loci, mPdIRD50, mPdIRD52 and mPdIRD80, were identified in
225 *P. dactylifera* by Cherif *et al*. [20]. We mapped mPdIRD50 and mPdIRD80 on the sex linkage
226 group (LG12) [21]. The mPdIRD52 has not mapped on the available genetic map. The three
227 sex-linked loci were successfully amplified in all individuals of the eight additional species
228 analysed.

229 **(i) Allele frequencies**

230 To further test whether sex linkage is conserved between *P. dactylifera* and the other *Phoenix*
231 species, Y- and X-linked allele frequencies for each locus were investigated.

232 **mPdIRDP52 locus**

233 The mPdIRDP52 locus yielded a total of 18 alleles. Four were Y-linked (mPdIRDP52_188,
234 mPdIRDP52_189, mPdIRDP52_191 and mPdIRDP52_193) exclusively distributed within
235 males. *P. roebelenii* was found to have a private Y-allele (mPdIRDP52_189) in addition to
236 the mPdIRDP52_191 allele (Figure 2). The fourteen remaining alleles were X-linked.

237 **mPdIRDP50 locus**

238 The mPdIRDP50 locus also had Y-linked alleles shared between males of the seven *Phoenix*
239 species (Figure 2). Only *P. roebelenii* had a private allele (mPdIRDP50_177). In Eastern
240 accessions of *P. dactylifera*, we observed Y-linked duplications of the mPdIRDP52 and
241 mPdIRDP50 loci. As in *P. dactylifera, P. sylvestris* and *P. acaulis* displayed duplicated male
242 alleles (Figure 2). The observed identity of mPdIRDP50 allele sizes (Figure 2) seen in eastern

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

243 accessions of *P. dactylifera, P. sylvestris*, *P. loureroi* and in *P. acaulis* suggests an ancestral
244 origin of these alleles.

245 **mPdIRDP80 locus**

246 The mPdIRDP80 locus had a total of seven alleles in the entire sample (Figure 2). The
247 mPdIRDP80_192 and mPdIRDP80_308 alleles were both strictly associated with the male
248 phenotype in the studied species, confirming the linkage to the Y chromosome. The other five
249 alleles (mPdIRDP80_290, mPdIRDP80_292, mPdIRDP80_296, mPdIRDP80_299,
250 mPdIRDP80_312) were shared between males and females and therefore presumed to be X-
251 linked alleles. *P. roebelenii* and *P. acaulis* species had private X alleles, namely the
252 mPdIRDP80_296 allele for *P. roebelenii* and the mPdIRDP80_312 allele for *P. acaulis*
253 (Figure 3).

254 The mPdIRDP80 locus showed a high level of conservation among the studied *Phoenix*
255 species, especially for the Y-linked alleles of which only two forms were observed. To further
256 compare X and Y sequences in *Phoenix* spp. and to exclude a possible convergence of SSR
257 allele sizes, we sequenced the mPdIRDP80 locus and performed multiple alignments of X and
258 Y sequences. The alignments of all X sequences showed a shared identity ranging from 95%
259 and 100%. The corresponding western and eastern Y alleles in *P. dactylifera* [20] revealed
260 that the amplified region is identical in the entire male sampling (identity ≥ 99%) (electronic
261 supplementary material, Figure 2, 3 and 4).

262 Taken together, these results indicate that the sex-linked region bearing the mPdIRDP80,
263 mPdIRDP50 and mPdIRDP52 loci is highly conserved throughout the genus.

264 Consistent with this conclusion, R_{st} values [25] (Table 1 and electronic supplementary
265 material, Figure 1), indicate significant genetic differentiation between males and females
266 within each species studied, except for *P. reclinata* ($R_{st} < 0$; this value may be biased due to

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

267 the different numbers of males and females studied, see the electronic supplementary
268 material, Table 4). Moreover, principal component analysis (PCA) yielded separate clusters
269 of males and females (Figure 3). Thus, as for *P. dactylifera*, these three loci showed sex
270 linkage in the additional studied species.

271 **(ii) Heterozygosity at the sex-linked loci**

272 All males of each *Phoenix* species analysed were heterozygous (Y_{allele}/X_{allele}) for all three *P.*
273 *dactylifera* sex-linked microsatellite loci (F*is* ranged from -0.04 to -1, see electronic
274 supplementary material, Table 5), with significant differences from the Hardy–Weinberg
275 genotype frequencies, expected under random mating, F*is* = 0. In contrast, females of each
276 species mostly had significantly positive F*is* values (electronic supplementary material, Table
277 5), in line with previous reports for X-linked SNPs in date palm females [19].

278 Overall, these results demonstrate that the XY chromosome system previously observed in *P.*
279 *dactylifera* is also present in the eight other *Phoenix* species studied here. The findings of this
280 comparative genetic study reveal a common set of orthologous sex-linked markers shared
281 between the additional studied species (*P. canariensis*, *P. atlantica*, *P. reclinata*, *P. sylvestris*,
282 *P. roebelenii*, *P. rupicola*, *P. loureroi* and *P. acaulis*) and *Phoenix dactylifera*.

283 The Y-linked alleles of the *Phoenix* species studied were assigned to 11 Y haplotypes none of
284 which were found among the 39 X-haplotypes (Figure 4). Therefore, recombination does not
285 occur between the Y and X chromosome regions carrying the mPdIRDP80, mPdIRDP50 and
286 mPdIRDP52 loci in any of the species, strongly suggesting complete sex-linkage of these loci,
287 as in *P. dactylifera* [20]. Moreover, the mPdIRDP80 and mPdIRDP50 microsatellite loci are
288 located in the LG12 sex segregating region in the genetic map of date palm corroborating
289 somehow the small size of sex segregating region [21]. Nevertheless, mPdIRDP52 locus was
290 mapped neither in LG12 nor in any other linkage group suggesting that Y region carrying

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

291 mPdIRDP80, mPdIRDP50 and mPdIRDP52 loci is probably larger than proposed by Mathew
292 *et al.* [21].
293 We next investigate whether this sex-linked region has evolved once or several times
294 independently within each species.

295 **(b) The PdMYB1 gene and the timing of recombination arrest in the *Phoenix* genus**

296 To estimate the time since recombination stopped, estimates of sequence divergence between
297 X and Y alleles are needed (ideally for multiple genes, to test whether different strata exist).
298 The microsatellite loci do not provide such an estimate. One gene is, however, now available.
299 The *PdMYB1* gene (see Methods) is homologous to the predicted *P. dactylifera* transcription
300 factor *MYB86-like* (GenBank accession number XM_008777432.1) and maps in the LG12
301 (PDK_30s6550963) sex-linked region of the date palm genetic map [21] and the mPdIRDP80
302 microsatellite locus is localized within its promoter sequence (Figure 5A). We cloned and
303 sequenced the ~ 2 Kb region including the *PdMYB1* gene and mPdIRDP80 (electronic
304 supplementary material, Figure 5). The *PdMYB1* has three exons of 135, 128 and 963 pb, and
305 two introns of 191 and 102 pb, and the mPdIRDP80 sequence is 176 pb upstream of the first
306 exon (Figure 5 A). Sequencing of the mPdIRDP80 locus confirmed the observed sizes of X
307 and Y alleles from capillary gel genotyping (Figure 5 A). We cloned the *PdMYB1* gene in the
308 seven related *Phoenix*, and identified their X/Y pairs based on the sizes of the mPdIRDP80 X
309 and Y alleles. All males had different X and Y sequences while all females had only the X
310 sequences.
311 We reconstructed the phylogeny of the X and Y copies of the *PdMYB1*gene from the newly
312 studied *Phoenix* species, and found distinct Y and X alleles clusters, rather than a clustering
313 by species of origin (Figure 5 B). The same result was also obtained with the distribution of X
314 and Y mPdIRDP80, mPdIRDP50 and mPdIRDP52 SSRs alleles among the *Phoenix* species
315 by PCA (see the electronic supplementary material, Figure 6). These results provide evidence

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

316 that recombination between X and Y alleles of the *PdMYB1*gene and the sex-linked loci
317 stopped before the *Phoenix* spp. split. Similarly, phylogenetic analysis suggested that
318 recombination stopped before the speciation of the dioecious species *S. latifolia*, *S. dioica*,
319 and *S. diclinis*, whereas sequences of the SIX1/SIY1 gene may have undergone suppressed
320 recombination independently in these species [3].

321 **(c) Y chromosome evolution in the genus *Phoenix***

322 The network based on the Y haplotypes revealed at least two major Y chromosome lineages
323 (Figure 6). The first includes four species, two from the Atlantic islands (*P. canariensis* and
324 *P. atlantica*), one from sub-Saharan Africa (*P. reclinata*) plus the western *P. dactylifera*
325 lineage (North Africa). The second lineage includes three species, the eastern *P. dactylifera*
326 lineage, from Pakistan and the Middle East, and two species from Asia (*P. sylvestris and P.*
327 *roebelenii*) (Figure 6). In the eastern lineage, the *P. roebelenii* group appears divergent from
328 the *P. sylvestris* and eastern *P. dactylifera* groups and has private Y haplotypes. The two male
329 lineages within *P. dactylifera* were previously identified [20]. The observation that these
330 haplotypes are also found in other *Phoenix* species suggests a split into two male paternal
331 lineages from an ancestral Y population before speciation, with the two Y-haplotypes
332 transmitted to the evolving species and becoming fixed in different species. However, the
333 presence of haplotypes within *P. dactylifera* is surprising. It could be due to ancestral
334 polymorphism maintained on the Y chromosome or to more recent introgressions but this still
335 need to be investigated.

336 **4. Conclusion**

337 According to fossils of *Phoenix* male flowers, dioecy could have emerged in the genus during
338 the Eocene (~ 49.5 mya) [10, 38].
339 Our data bring evidence that recombination arrest evolved prior to speciation in the
340 genus *Phoenix*. Furthermore, the entire extant *Phoenix* species are dioecious which allows to

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

341 conclude that sex chromosomes and dioecy evolved only once in this lineage. The time at
342 which recombination stopped in the *Phoeniceae* can be estimated by analyzing X/Y
343 divergence of other sex-linked genes especially from the non-recombining regions of the date
344 palm sex-linkage group [21].

345

346 **Data accessibility**

347 *The datasets supporting this article have been uploaded as part of the Supplementary*
348 *Material.*

349 **Competing interests**

350 *'We have no competing interests.'*

351 **Author's contributions**

352 FAB conceived and planned the project, EC and KC, performed the in silico research of SSR,
353 EC, NC, JCP performed the DNA preparations, EC performed the genotyping, AC performed
354 the cloning and sequencing of the MYB gene, EC performed the genetic and statistical
355 analyses, EC, SZ, AC, JCP, SS, SG, ASH and FAB, discussed the results, EC and FAB wrote
356 the paper. All authors gave final approval for publication.

357 **Acknowledgements**

358 We thank Ali Zouba, Adbelmadjid Rhouma, Karim Kadri, Ahmed Othmani (CRRAO,
359 Tunisia), Abdourahman Daher, Sabira Abdoulkader (CERD, Djibouti), , Claudio Littardi
360 (CRSP, Italy), Jose Plumed (Botanical garden of Valencia, Spain), Marco Ballardini
361 (CRAFSO, Italy), Muriel Gros-Balthazar (IRD, France), Robert Krueger (USDA, USA),
362 Emmanuel Spick (Jardin des Plantes de Montpellier, France), Bill Baker (Kew Gardens, UK)
363 and Alain Rival (Cirad, France) for providing plant material; Sylvain Santoni and Muriel

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

364 Latreille (INRA, France) for technical assistance; and James Tregear for corrections to the
365 manuscript.

366 **Funding**

367 This work was supported by the AUF-Mersi Project, the 'Ministère de l'Enseignement
368 Supérieur et de la Recherche' of Tunisia and the Qatar National Research Fund (NPRP-EP X-
369 014-4-001).

370 **Literature Cited**

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

373 2. Charlesworth D, Mank JE. 2010 The birds and the bees and the flowers and the trees:
374 lessons from genetic mapping of sex determination in plants and animals. *Genetics*. **186**, 9–
375 31. (doi: 10.1534/genetics.110.117697)

376 3. Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D,
377 Negrutiu I, Charlesworth D, Monéger F. 2004 A gradual process of recombination restriction
378 in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol.* **3**, 1–4.
379 (doi: 10.1371/journal.pbio.0030004)

380 4. Wang J, Na JK, Yu Q. Gschwend AR, Han J, Zeng F *et al*. 2012. Sequencing papaya X and
381 Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl.*
382 *Acad. Sci.* **109,** 13710–13715. (doi: 10.1073/pnas.1207833109)

383 5. Vyskot B, Hobza R. 2004 Gender in plants: sex chromosomes are emerging from the fog.
384 *Trends Genet*. **20**, 432–438. (doi: 10.1016/j.tig.2004.06.006)

385 6. Ming R, Wang J, Moore PH, Paterson AH. 2007 Sex chromosomes in flowering plants. *Am.*
386 *J. Bot.* **94**,141–50. (doi:10.3732/ajb.94.2.141)

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

387 7. Janousek B, Mrackova M. 2010 Sex chromosomes and sex determination pathway
388 dynamics in plant and animal models. *Biol. J. Linn. Soc.* **100**, 737–52. (doi: 10.1111/j.1095-
389 8312.2010.01470.x)
390 8. Barrow S. 1998 A monograph of *Phoenix* L. (Palmae: Coryphoideae). *Kew Bull.* **53**, 513–
391 575.
392 9. Henderson S, Billotte N, Pintaud JC. 2006 Genetic isolation of Cape Verde Islands *Phoenix*
393 *atlantica* (Arecaceae) revealed by microsatellite markers. *Conserv. Genet.* **7**, 213–223. (doi:
394 10.1007/s10592-006-9128-7)
395 10. Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE. 2008 Genera
396 Palmarum: the evolution and classification of palms. Kew, UK: Kew Publishing.744 p.
397 11. Asmussen CB, Chase MW. 2001 Coding and noncoding plastid DNA in palm
398 systematics. *Am. J. Bot.* **88**, 1103–1117**.**
399 12. Hahn WJ. 2002 A molecular phylogenetic study of the Palmae (Arecaceae) based on atpB,
400 rbcL, and 18S nrDNA sequences. *Syst. Biol.* **51**, 92–112. (doi:
401 10.1080/106351502753475899)
402 13. Asmussen CB, Dransfield J, Deickmann V, Barfod AS, Pintaud JC, Baker WJ. 2006 A
403 new subfamily classification of the palm family (Arecaceae), evidence from plastid DNA
404 phylogeny. *Biol. J. Linn. Soc.* **151**, 15–38. (doi: 10.1111/j.1095-8339.2006.00521.x)
405 14. Baker WJ, Couvreur TLP. 2013 Global biogeography and diversification of palms sheds
406 light on the evolution of tropical lineages. I. Historical biogeography. *J.Biogeogr.* **40**, 274–
407 285. (doi: 10.1111/j.1365-2699.2012.02795.x)
408 15. Pintaud JC, Ludeña B, Aberlenc-Bertossi F, Zehdi S, Gros-Balthazard M, Ivorra S, Terral
409 JF, Newton C, Tengberg M, Abdoulkader S *et al.* 2013 Biogeography of the date palm
410 (*Phoenix dactylifera* L.,Arecaceae): Insights on the origin and on the structure of modern
411 diversity. *Acta Hort.* **994**, 19–38.

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

412 16. Ogg J. 2004 Status of divisions of the international geologic time scale. *Lethaia.* **37**, 183–
413 199. (doi: 10.1080/00241160410006492)
414 17. Daher A, Adam H, Chabrillange N, Collin M, Mohamed N, Tregear JW, Aberlenc-Bertossi
415 F. 2010 Cell cycle arrest characterizes the transition from a bisexual floral bud to a unisexual
416 flower in *Phoenix dactylifera*. *Ann. Bot.* **106**, 255–266. (doi: 10.1093/aob/mcq103)
417 18. Siljak-Yakovlev S, Cerbah M, Sarr A, Benmalek S, Bounaga N, Coba de la Pena T, Brown
418 S. 1996 Chromosomal sex determination and heterochromatin structure in date palm. *Sex.*
419 *Plant Reprod.* **9**, 127–132. (doi: 10.1007/BF02221391)
420 19. Al-Dous EK, George B, Al-Mahmoud ME, Al-Jaber MY, Wang H, Salameh YM,
421 Al-Azwani EK, Chaluvadi S, Pontaroli AC, DeBarry J *et al*. 2011 De novo genome
422 sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat. Biotechnol.*
423 **29**, 521–527. (doi:10.1038/nbt.1860)
424 20. Cherif E, Zehdi S, Castillo K, Chabrillange N, Abdoulkader S, Pintaud JC, Santoni S,
425 Salhi-Hannachi A, Glémin S, Aberlenc-Bertossi, F. 2013 Male-specific DNA markers provide
426 genetic evidence of an XY chromosome system, a recombination arrest and allow the tracing
427 of paternal lineages in date palm. *New Phytol.* **197**, 409–415. (doi: 10.1111/nph.12069)
428 21. Mathew LS, Spannagl M, Al-Malki A, George B, Torres MF, Al-Dous EK, Al-Azwani
429 EK, Hussein E, Mathew S, Mayer KFX *et al*. 2014 A first genetic map of date palm (*Phoenix*
430 *dactylifera*) reveals long-range genome structure conservation in the palms. *BMC*
431 *Genomics.* **15**, 285. (doi:10.1186/1471-2164-15-285)
432 22. Lawson-Handley LJ, Ceplitis H, Ellegren H. 2004 Evolutionary strata on the chicken Z
433 chromosome: implications for sex chromosome evolution. *Genetics.* **167**, 367–376.
434 (doi: 10.1534/genetics.167.1.367)

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

435 23. Peakall R, Smouse PE. 2006 GENALEX 6, genetic analysis in Excel. Population genetic
436 software for teaching and research. *Mol. Ecol Notes*. **6**, 288–295. (doi: 10.1111/j.1471-
437 8286.2005.01155.x)
438 24. Peakall, R, Smouse PE. 2012 GenAlEx 6.5: genetic analysis in Excel. Population genetic
439 software for teaching and research-an update. *Bioinformatics*. **28**, 2537–2539.
440 (doi:10.1093/bioinformatics/bts460)
441 25. Slatkin M. 1995 A measure of population subdivision based on microsatellite allele
442 frequencies. *Genetics* **139**, 457–462.
443 26. R Development Core Team. 2011 "R: A language and environment for statistical
444 computing." R Foundation for Statistical Computing, Vienna, Austria,URL: .
445 27. Weir BS, Cockerham CC. 1984 Estimating F-statistics for the analysis of population
446 structure. *Evolution*. **38**, 1358–1370. (doi: 10.2307/2408641)
447 28. Robertson A, Hill WG. 1984 Deviations from Hardy–Weinberg proportions: sampling
448 variances and use in estimation of inbreeding coefficients. *Genetics*. **107**, 703–718.
449 29. Guo SW, Thompson EA. 1992 Performing the exact test of Hardy–Weinberg proportion
450 for multiple alleles. *Biometrics*. **48**, 361–372. (doi: 10.2307/2532296)
451 30. Raymond M, Rousset F. 1995 GENEPOP (version 1.2): population genetics software for
452 exact tests and ecumenicism. *J. Hered.* **86**, 248–249.
453 31. Rousset F. 2008 Genepop’007: a complete reimplementation of the Genepop software for
454 Windows and Linux. *Mol. Ecol. Resour.* **8**, 103–106. (doi: 10.1111/j.1471-8286.2007.01931.x)
455 32. Excoffier L, Slatkin M. 1995 Maximum-likelihood estimation of molecular haplotype
456 frequencies in a diploid population. *Mol. Biol. Evol.* **12**, 921–927.
457 33. Bandelt HJ, Forster P, Röhl A. 1999 Median-joining networks for inferring intraspecific
458 phylogenies. *Mol. Biol. Evol.* **16**, 37–480.

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

459 34. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013 MEGA6: Molecular
460 Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729.
461 (doi: 10.1093/molbev/mst197)
462 35. Kimura M. 1980 A simple method for estimating evolutionary rate of base substitutions
463 through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16,** 111–120.
464 (doi:10.1007/BF01731581)

465 36. Pintaud JC, Zehdi S, Couvreur T, Barrow S, Henderson S, Aberlenc-Bertossi F, Tregear J,
466 Billotte N. 2010 Species delimitation in the genus *Phoenix* (Arecaceae) based on SSR
467 markers, with emphasis on the identity of the date palm (*Phoenix dactylifera* L.). In: Seberg
468 O, Petersen G, Barfod AS, Davis JI, editors. Diversity, phylogeny and evolution in the
469 monocotyledons: Aarhus University Press. pp. 267–286.
470 37. Billotte N, Marseillac N, Brottier P, Noyer JL, Jacquemoud-Collet JP, Moreau C,
471 Couvreur T, Chevallier MH, Pintaud JC, Risterucci AM. 2004 Nuclear microsatellite markers
472 for the date palm (*Phoenix dactylifera* L.): characterization and utility across the genus
473 *Phoenix* and in other palm genera. *Mol. Ecol. Notes.* **4**, 256–258. (doi: 10.1111/j.1471-
474 8286.2004.00634.x)
475 38. Rivera D, Obón C, García-Arteaga J, Egea T, Alcaraz F, Laguna E *et al*. 2014
476 Carpological analysis of *Phoenix* (Arecaceae): contributions to the taxonomy and
477 evolutionary history of the genus. *Bot J Linn Soc.* **175**, 74–122. (doi: 10.1111/boj.12164)
478 39. Allen, SE. 2015 Fossil Palm Flowers from the Eocene of the Rocky Mountain Region
479 with Affinities to *Phoenix* L.(Arecaceae: Coryphoideae). *Int. J. Plant Sci.* **176**, 586–596.
480 (doi: 10.1086/681605)

481 **Figure Legends**

482 **Figure 1. Geographical distribution of the studied *Phoenix* species.**

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

483 **Figure 2. Dispersion of *Phoenix* species in relation to sex on the first biplot of the**
484 **Principal Component Analysis.**
485 **DM**: *P.dactylifera* male; **DF**: *P.dactylifera* female; **CM**: *P.canariensis* male; **CF**:
486 *P.canariensis* female; **SM**: *P.sylvestris* male; **SF**: *P.sylvestris* female; **RoM**: *P.roebelenii*
487 male; **RoF**: *P.roebelenii* female; **AM**: *P.atlantica* male; **Af**: *P.atlantica* female; **ReM**:
488 *P.reclinata* male; **ReF**: *P.reclinata* female.

489 **Figure 3. Allelic distribution of mPdIRD50, mPdIRD52 and mPdIRD80 loci within the**
490 **studied *Phoenix* species.**
491 Left side corresponds to male genotypes and right side corresponds to female genotypes. Each
492 dot represents an allele. Alleles shared between male and female individuals (X) are
493 represented in red and male specific alleles (Y) are represented in blue. Duplicated male
494 alleles (Y) are represented in light-blue. Female individuals have only alleles shared between
495 male and female while male individuals have shared alleles and male specific alleles.

496 **Figure 4. Haplotypes distribution within studied *Phoenix* species analysed as a whole**
497 **(*P. dactylifera*, *P. reclinata*, *P.sylvestris*, *P. atlantica*, *P. canariensis*, *P. roebelenii* and *P.***
498 ***acaulis*).** Light and dark red histograms represent respectively X female and X male
499 haplotypes. Blue histograms correspond to Y haplotypes.

500 **Figure 5. Structure of *Phoenix PdMYB1* gene and molecular phylogenetic analysis by**
501 **Maximum Likelihood method.**
502 **A.** The structure of the *PdMYB1* gene was predicted based on multiple alignments including
503 the *P.dactylifera* transcription factor *MYB*86-like and the orthologous sequences of other
504 species. Blue, yellow and grey rectangles indicate respectively the position of mPdIRDP80
505 motif, exons and introns of the *PdMYB1* gene.

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

506 **B.** The tree with the highest log likelihood (-3368.6438) is shown. The percentage of trees in
507 which the associated taxa clustered together is shown next to the branches (1000
508 replications). Initial tree (s) for the heuristic search were obtained automatically by
509 applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated
510 using the Maximum Composite Likelihood (MCL) approach, and then selecting the
511 topology with superior log likelihood value. The tree is drawn to scale, with branch lengths
512 measured in the number of substitutions per site. The analysis involved 19 nucleotide
513 sequences. All positions with less than 95% site coverage were eliminated. That is, fewer
514 than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position.
515 There were a total of 1494 positions in the final dataset. Corresponding orthologous of the
516 *Elaeis guineensis PdMYB1* gene was used as the outgroup. X and Y represent the X- and
517 Y-specific alleles, respectively.

518 **Figure 6. Date palm Y-SSR haplotype Network.**

519 Red dots indicate the median vectors corresponding to the hypothesised haplotypes required
520 to connect the existing ones.

521 **Table 1.** Genetic structuring within the studied *Phoenix* species generated by each sex-linked
522 microsatellite between the two sexes.

523

524

525

526

527

528

bioRxiv preprint doi: https://doi.org/10.1101/033365; this version posted December 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

529 **Table 1.** Genetic structuring within the studied *Phoenix* species generated by each sex-linked
530 microsatellite between the two sexes

Locus		R_{st}*
Sex-Linked	mPdIRD P52	**0.326**
	mPdIRD P50	**0.351**
	mPdIRD P80	**0.317**

531 * The R_{st} index measures the genetic differentiation generated by the sex-linked loci. We observe that
532 the values of R_{st} obtained with the sex-linked loci set (in bold) are significantly higher than zero and
533 clearly show a genetic structuring of the population in relation to sex.

534

535

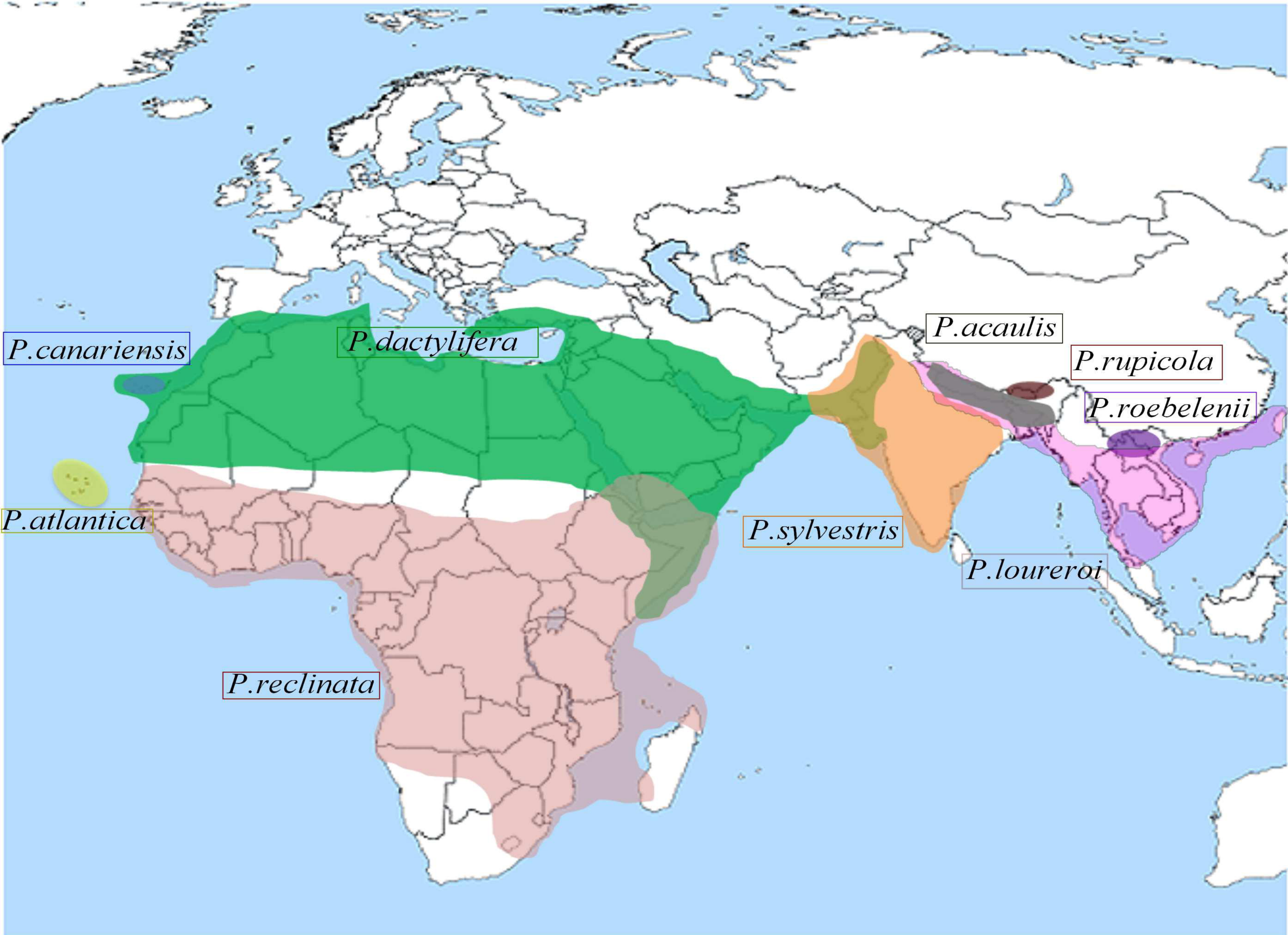

Figure 1. Geographical distribution of the studied *Phoenix* species.

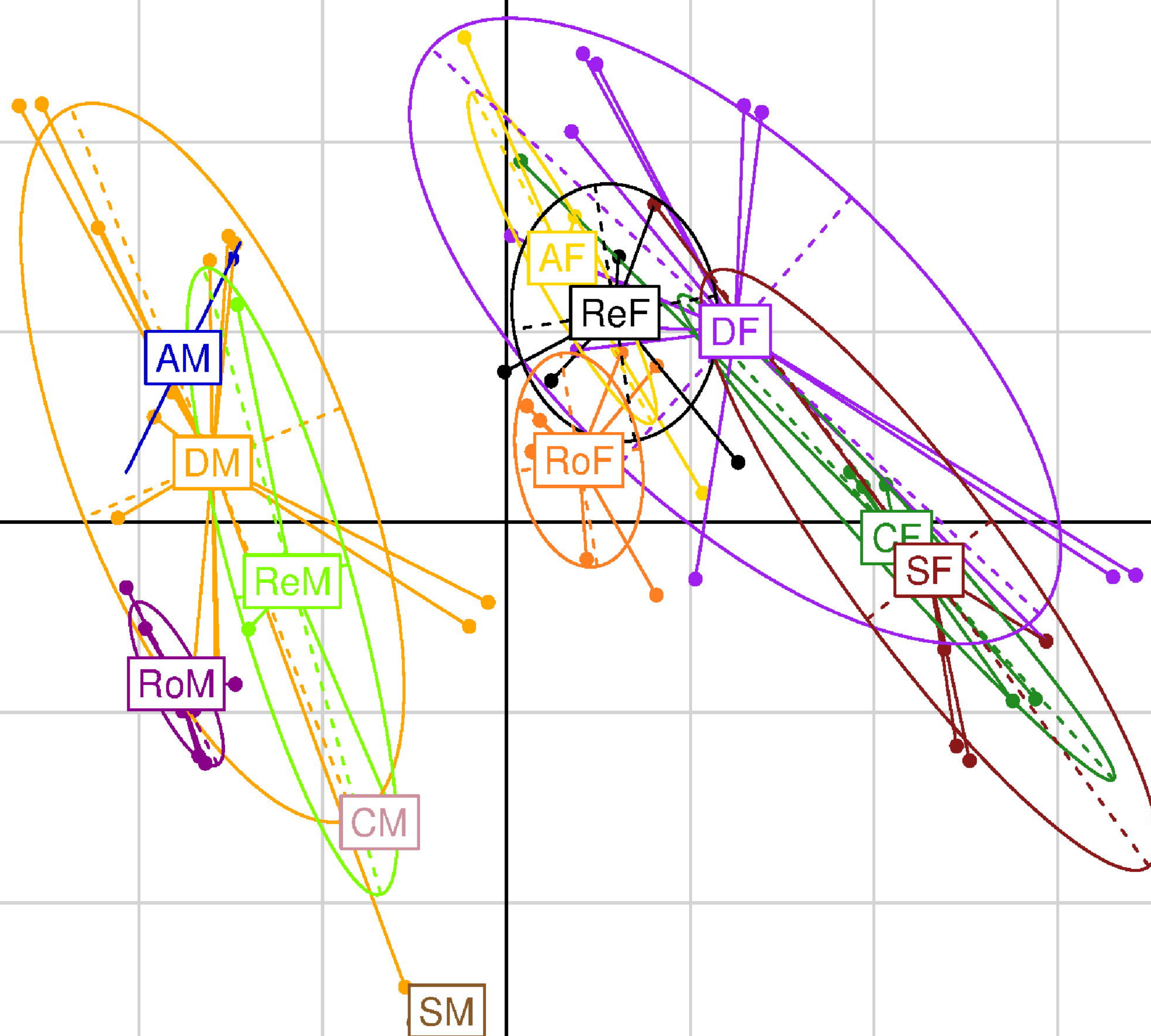

Figure 2. Dispersion of *Phoenix* species in relation to sex on the first biplot of the Principal Component Analysis. **DM**: *P.dactylifera* male; **DF**: *P.dactylifera* female; **CM**: *P.canariensis* male; **CF**: *P.canariensis* female; **SM**: *P.sylvestris* male; **SF**: *P.sylvestris* female; **RoM**: *P.roebelenii* male; **RoF**: *P.roebelenii* female; **AM**: *P.atlantica* male; **Af**: *P.atlantica* female; **ReM**: *P.reclinata* male; **ReF**: *P.reclinata* female.

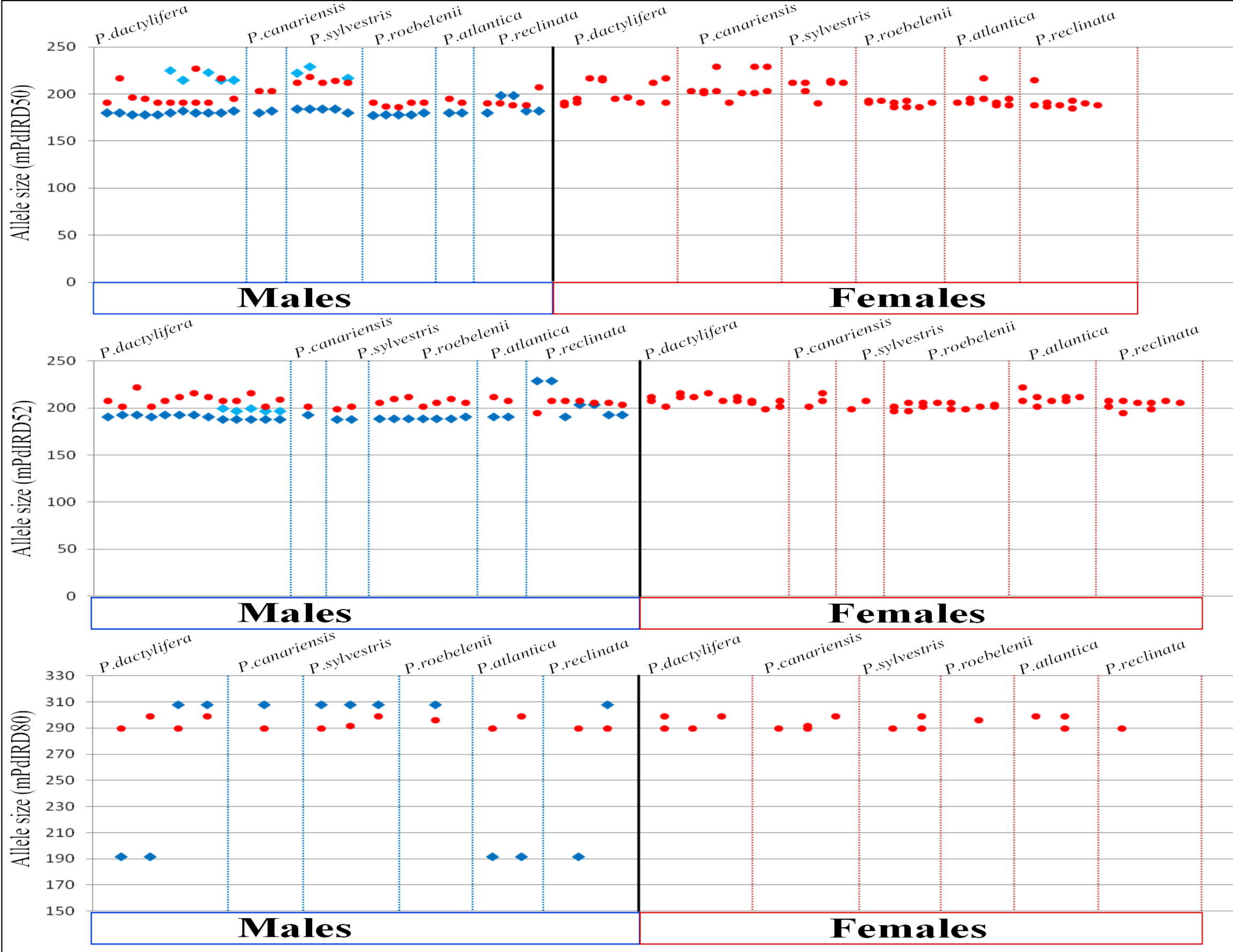

Figure 3. Allelic distribution of mPdIRD50, mPdIRD52 and mPdIRD80 loci within the studied *Phoenix* species.
Left side corresponds to male genotypes and right side cprresponds to female genotypes. Each dot represents an allele. Alleles Alleles shared between male and female individuals (X) are represented in red and male specific alleles (Y) are represented in blue. Duplicated male alleles (Y) are represented in light-blue. Female individuals have only alleles shared between males and female while male individuals have shared alleles and male specific alleles.

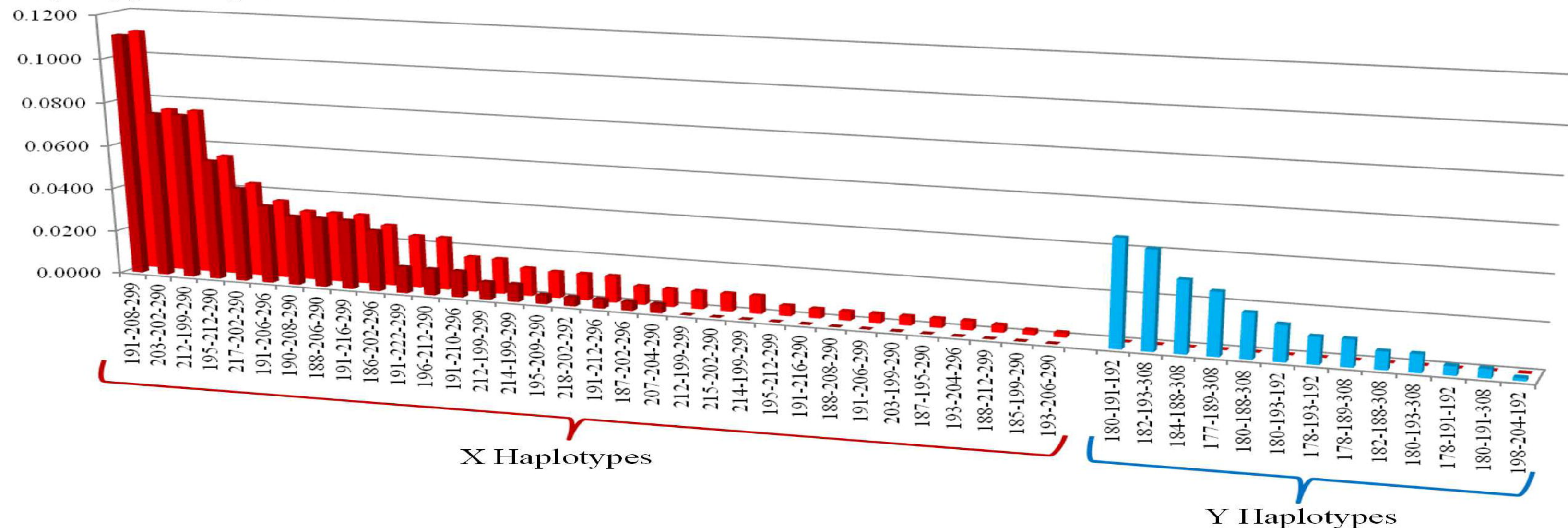

Figure 4. Haplotypes distribution within studied *Phoenix* species analysed as a whole (*P.dactylifera, P.reclinata,P.sylvestris P.atlantica, P.canariensis, P.roebelenii and P.acaulis*).

Light and dark red histograms represent respectively X female and X male haplotypes. Blue histograms correspond to Y haplotypes. Each haplotype was constructed by assigning a phase to the genotype obtained by the alleles of the three sex-linked SSRs (P50, P52 and P80).

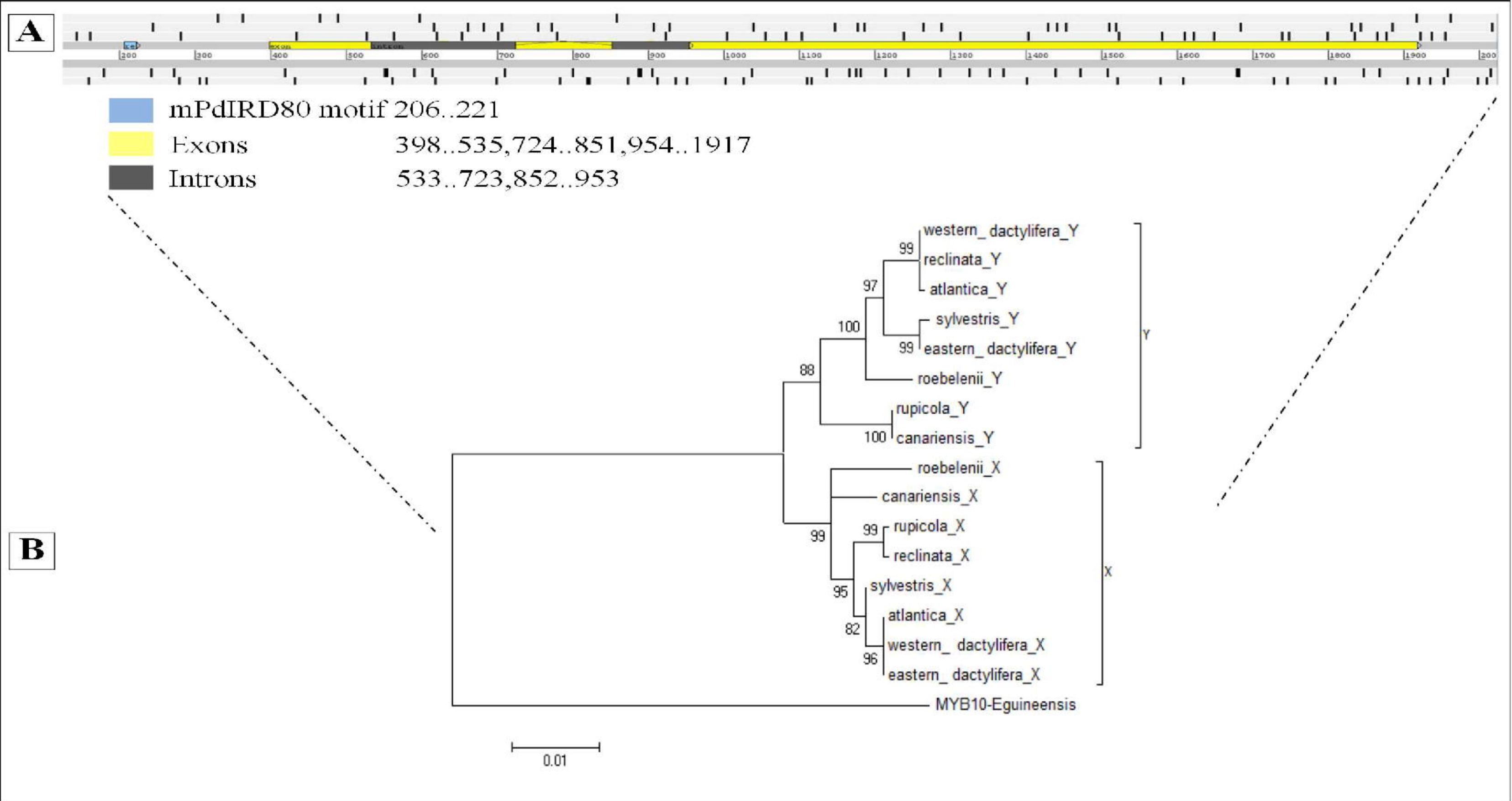

Figure 5. Structure of *Phoenix PdMYB1* gene and molecular phylogenetic analysis by Maximum Likelihood method.

A. The structure of the *PdMYB1* gene was predicted based on multiple alignments including the *P.dactylifera* transcription factor MYB86-like and the orthologous sequences of other species. Blue, yellow and grey rectangles indicate the position of mPdIRDP80 motif, exons and introns of the *PdMYB1* gene respectively.

B. The tree with the highest log likelihood (-3351.2862) is shown.The percentage of trees in which the associated taxa clustered together is shown next to the branches (1000 replications).Initial tree (s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to to matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 17 nucleotide sequences. All positions with less than 95% site coverage were eliminated. That is, fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position.There were a total of 1494 positions in the final dataset. Corresponding orthologous of the *Elaeis guineensis PdMYB1* gene was used as the outgroup. X and Y represent the X- and Y-specific alleles, respectively.

Figure 6. Date palm Y-SSR haplotype Network.
Red dots indicate the median vectors corresponding to the hypothesised haplotypes required to connect the existing ones.