1 **Title**

2 **Extreme Y chromosome polymorphism corresponds to five male reproductive morphs of**

3 **a freshwater fish**

4 **Author List**

5 Benjamin A Sandkam[a*], Pedro Almeida[b], Iulia Darolti[a], Benjamin Furman[a], Wouter van der Bijl[a], Jake

6 Morris[b], Godfrey Bourne[c], Felix Breden[d], Judith E. Mank[a]

7 **Affiliations**

8 a. Department of Zoology, University of British Columbia, Vancouver, BC V6T 1Z4, Canada

9 b. Department of Genetics, Evolution and Environment, University College London, London WC1E 6BT,

10 United Kingdom

11 c. Department of Biology, University of Missouri-St. Louis, St. Louis, MO 63121, USA

12 d. Department of Biological Sciences, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

13

14 *** Corresponding Author:** Benjamin A Sandkam

15 **Email:** sandkam@zoology.ubc.ca

16

17 ORCIDs: 0000-0002-5043-9295, 0000-0002-7082-3577, 0000-0002-7366-1868, 0000-0002-0137-2610,

18 0000-0001-9762-6314, 0000-0002-2450-513X

19

1

Abstract

Loss of recombination between sex chromosomes often depletes Y chromosomes of functional content and genetic variation, which might limit their potential to generate adaptive diversity. Males of the freshwater fish *Poecilia parae* occur as one of five discrete morphs, all of which shoal together in natural populations where morph frequency has been stable for over 50 years. Each morph utilizes a different complex reproductive strategy, and morphs differ dramatically in color, body size, and mating behavior. Morph phenotype is passed perfectly from father to son, indicating there are five Y haplotypes segregating in the species, which encode the complex male morph characteristics. Here, we examine Y diversity in natural populations of *P. parae*. Using linked-read sequencing on multiple *P. parae* females and males of all five morphs, we find that the genetic architecture of the male morphs evolved on the Y chromosome after recombination suppression had occurred with the X. Comparing Y chromosomes between each of the morphs, we show that although the Ys of the three minor morphs that differ in color are highly similar, there are substantial amounts of unique genetic material and divergence between the Ys of the three major morphs that differ in reproductive strategy, body size and mating behavior. Altogether, our results suggest that the Y chromosome is able to overcome the constraints of recombination loss to generate extreme diversity, resulting in five discrete Y chromosomes that control complex reproductive strategies.

Introduction

Sex chromosomes form when recombination is halted between the X and Y chromosomes. The loss of recombination results in a host of evolutionary processes that quickly deplete Y chromosomes of functional content and genetic variation, severely limiting the scope for adaptive evolution[1]. Y chromosomes can counter this loss to some degree through a variety of mechanisms[2-5], however the adaptive potential of Y chromosomes is generally thought to be much lower than the remainder of the genome. Typically, this results in relatively low levels of Y chromosome diversity within species. The adaptive potential of non-recombining regions has far broader implications beyond just Y chromosomes given the increasing realization that supergenes, linked regions containing alleles at multiple loci underlying complex phenotypes, are key to many adaptive traits[6-12]. Many supergenes are lethal when homozygous and therefore also non-recombining[8,10]. Therefore, the processes that constrain Y chromosome evolution also affect much broader areas of the genome.

Poecilia parae is a small freshwater fish found in coastal streams of South America. Remarkably, males of this species occur as one of five distinct morphs, parae, immaculata, melanzona red, melanzona yellow, and melanzona blue, each of which utilizes distinct reproductive tactics, with associated differences in body size, color, and mating behaviours[13-19] (summarized in Supplementary Table 1). The parae morph has the largest body, vertical black stripes, an orange tail-stripe and is highly aggressive, chasing away rival males and aggressively copulating with females by force. Immaculata resembles a juvenile female. Although immaculata has the smallest body size of all morphs, it has the largest relative testes and produces the most sperm, employing a sneaker copulation strategy. The three melanzona morphs are similar in body size and have a colored horizontal stripe (either red, yellow or blue), which they present to females during courtship displays. We refer to the three minor morphs of melanzona collectively as the melanzona morph.

All five morphs co-occur in the same populations, and the relative frequency of morphs within populations (on average: ~35% immaculata, ~35% parae, and ~30% melanzona) is highly stable over repeated surveys spanning 50 years (~150 generations)[13,14,20]. This suggests that balancing selection, likely resulting from a combination of sexual and natural selection, is acting to maintain these five adapted

3

67   morphs. Most importantly, multigeneration pedigrees show that morph phenotype is always passed

68   perfectly from father to son[13], indicating the five *P. parae* morphs are controlled by five different Y

69   chromosomes. This system therefore offers the potential for a unique insight into the adaptive potential of

70   Y chromosomes, and the role of these regions of the genome in male phenotypes.

71   Y chromosomes are formed once recombination with the X is halted[1], and the loss of recombination on

72   the Y leads to a complex cascade of non-adaptive processes that lead to the rapid buildup of

73   heterochromatin and loss of gene activity[21-23]. However, the process of Y degeneration is not linear[24], and

74   although poecilid species closely related to *P. parae* share the same homologous sex chromosome as

75   *Poecilia reticulata*[25] (guppies), the extent of Y chromosome degeneration differs markedly across the

76   clade. Although the Y chromosome in *P. reticulata* and *Poecilia wingei* contains only a small area of

77   limited degeneration[25-28], the entirety of the Y chromosome of *Poecilia picta* is highly degenerate[25]. *P.*

78   *parae* is a sister species of *P. picta* (diverging ~14.8 mya[29]), however *P. picta* males are markedly

79   different from *P. parae* and do not resemble any of the five *P. parae* morphs[18,20,30-32], suggesting

80   remarkable diversity was generated on the *P. parae* Y chromosome after recombination was halted with

81   the X chromosome. Work on model systems has indeed shown Y chromosomes can accumulate new

82   genetic material[2-5], yet these differences occur over long periods of time and are only evident when

83   comparing Ys across species. Non-model systems, such as *P. parae*, provide a unique opportunity to

84   explore the limits and role of non-recombining regions in generating diversity.

85   Because *P. parae* is very difficult to breed in the lab, we collected tissue from natural populations in South

86   America where all five male morphs co-occur, and used linked-read sequencing on multiple females and

87   males of all five morphs. Linked-read sequencing allowed us to overcome many of the technical

88   difficulties that traditional short-read methods run into when working on regions rich in repeats and

89   transposable elements, which are often associated with sex chromosomes. We first confirmed that *P.*

90   *parae* shares the same sex chromosome system as its close relatives[25,26]. We went on to find patterns of

91   X-Y divergence are the same for all five Y chromosomes and matches the X-Y divergence we observed

92   in *P. picta*, suggesting that the morphs emerged after Y chromosome recombination was stopped in a

93   common ancestor of *P. parae* and *P. picta*. Comparing the five Y chromosomes to each other, we find

94   that while the Ys of the three minor morphs (red, yellow and blue melanzona) that differ only in color are

4

95   highly similar, the Ys of the three major morphs (parae, immaculata, and melanzona) that differ in

96   reproductive strategy, body size and mating behavior are significantly diverged from one another and

97   carry substantial amounts of unique genetic material. Taken together, our results reveal the surprising

98   ability of the Y chromosome to not only overcome the constraints of recombination loss, but to generate

99   extreme diversity, resulting in five discrete Y chromosomes that correspond to complex reproductive

100   strategies.

101

102   **Results**

103   We collected 40 individual *P. parae* from natural populations in Guyana in December 2016, including

104   eight red melanzona, four blue melanzona, five yellow melanzona, five immaculata, seven parae morph

105   males, and 11 females. 29 samples with sufficiently high molecular weight DNA were individually

106   sequenced with 10X Genomics linked-reads (including all morphs and several females). We generated a

107   *de novo* genome assembly for each of these samples. The remaining 11 lower molecular weight samples

108   were individually sequenced with Illumina sequencing paired end reads (see Supplementary Tables 2 and

109   3 for sequencing and assembly details).

110

111   *The P. parae Y chromosome is highly diverged from the X and shared with P. picta*

112   Degeneration of the Y chromosome results in reduced male coverage when mapped to a female

113   reference genome. Therefore, the ratio of male to female mapped reads can be used to identify regions

114   where the Y and X chromosomes differ substantially from each other[25,33-35]. To do this, we used our best

115   female *de novo P. parae* genome, based on N50 and other assembly statistics (see Supplementary Table

116   3). We then determined chromosomal position of the scaffolds using the reference-assisted chromosome

117   assembly (RACA) pipeline, which combines phylogenetic and sequencing data to place scaffolds along

118   chromosomes[36]. Next we mapped reads from all 40 samples to this female assembly and calculated

119   male:female coverage, first for each of the five morphs independently, and then all morphs together.

120   As we previously found in *P. picta*[25], chromosome 8 (syntenic to *P. reticulata* chromosome 12) showed a

121   clear signal of reduced read coverage in males (Figure 1a-b), indicating an XY sex determination system.

122   Y divergence is evident across nearly the entire chromosome and is largely identical to the pattern we

123   previously observed in *P. picta* (Figure 1c and Extended Data Fig. 1). This suggests that these species

124   inherited a highly degenerate Y chromosome from their common ancestor, well before the origin of the *P.*

125   *parae* male morphs.

126   Short sequences representing all the possible substrings of length *k* that are contained in a genome are

127   referred to as *k*-mers, and *k*-mer comparisons between male and female genomes has been used to

128   identify Y chromosome sequence using Y-specific *k*-mers (Y-mers) in a wide range of organisms[37-39],

129   including guppies[40] and *P. picta*[25]. We first compared all males, representing all five morphs, to all

130   females. We found a total of 27,950,090 Y-mers (of 31bp) that were present in at least two males but

131   absent from all females. However, only 59 of these Y-mers were present in all 23 males (Figure 2). To

132   evaluate the number of false positive Y-mers identified by our approach, we used the same pipeline on

133   females since all female *k*-mers should be present in males. We found only 251,472 *k*-mers to be present

134   in at least two females but absent from all males (0 of these were found in all 6 females). This suggests

135   that only about 0.9% of the 27,950,090 Y-mers identified are false positives.

136   We next used Y-mer analysis to further test whether recombination was halted on the Y in the common

137   ancestor of *P. parae* and *P. picta*. Of the 646,745 Y-mers that we previously identified in at least one *P.*

138   *picta* male and no females[25], 790 *P. picta* Y-mers matched Y-mers we identified in *P. parae*, consistent

139   with a shared history of suppressed recombination. Additionally, these shared Y-mers were present in

140   males of all morphs (Extended Data Fig. 2), discussed in more detail below. To evaluate whether the

141   number of Y-mers shared between species was due to chance we compared the female specific *k*-mers

142   that we previously identified in *P. picta*[25] to the female specific *k*-mers from *P. parae* we identified above.

143   We found only 80 *P. picta* female-specific *k*-mers were also female specific *k*-mers in *P. parae*. Therefore,

144   there were significantly more Y-mers shared between *P. picta* and *P. parae* than would be expected by

145   chance (X^2 = 1258, DF = 1, P < 0.0001), supporting a shared history of recombination suppression. These

146   shared Y-mers, combined with the striking similarity in male:female read mapping (Extended Data Fig. 1)

6

147   provide compelling evidence that the vast majority of Y chromosome recombination suppression occurred

148   in a common ancestor of *P. picta* and *P. parae.*

149

150   ***The P. parae Y chromosomes are highly diverged from each other***

151   We next compared Y-mers across individuals, generating a phylogeny on the presence/absence of Y-

152   mers in all the *P. parae* individuals with *P. picta* as an outgroup (Figure 2). Clear clades were recovered

153   for each of the major morphs (immaculata, parae and melanzona) while the three minor morphs of

154   melanzona (red, yellow, blue) were very similar to one another. The monophyly of morphs further

155   indicates that the Y-mers we recovered are not random false positives, and that morphs have distinct Y

156   chromosomes.

157   The phylogenic relationships of individuals (Figure 2) closely match the relative Y-mer comparisons

158   across morphs (Figure 3). We found 64,515 Y-mers in every immaculata male that were not in any parae

159   or melanzona males (i.e. immaculata-mers), 87,629 melanzona-mers, and 1,435 parae-mers, suggesting

160   that the melanzona and immaculata Y chromosomes may contain more unique sequence compared to

161   the parae Y. Moreover, we found 10,673 Y-mers in all melanzona and parae males that did not occur in

162   any immaculata males (Figure 2), suggesting that the parae Y shares greater sequence similarity to the

163   melanzona Y.

164   We calculated our false positive rate by randomly permuting our male samples into groups regardless of

165   morph and determining Y-mers present in all males of each group that were absent from all other males.

166   We found no unique Y-mers in groups of five or more random males, and just 31 unique Y-mers in groups

167   of four random males, demonstrating the false positive rate of our morph-specific Y-mer (morph-mer)

168   approach is exceedingly low (Extended Data Fig. 3).

169

170   ***Mapping morph-mers confirms high diversity of Y chromosomes***

171   The large number of morph-mers we identified could either indicate that the discrete morphs are the

172   result of low divergence across large Y chromosomes, or smaller complexes of highly diverged Y

7

173   sequence. To resolve this, we mapped the morph-mers to the 21 *de novo* male genomes. We found the Y

174   chromosomes contain regions of highly diverged Y sequence, indicated by morph-mers disproportionately

175   mapping to a few scaffolds, and not being evenly dispersed. For example, a single melanzona scaffold

176   (~110kb) containing 27% of all melanzona-mers (23,773), and most morph-mers overlapped one another

177   (Extended Data Fig. 4 and 5). This confirms our morph-mer approach identified complexes of highly

178   diverged Y sequence.

179   To compare the relative size of these diverged complexes across morphs, we identified all scaffolds that

180   contained >5 morph-mers in each individual. The average amount of sequence contained within these

181   morph-mer scaffolds was 1.3 Mb for melanzona, 3.2 Mb for immaculata, and just 0.1 Mb for parae

182   individuals (Supplementary Table 4). This complements the relative number of Y-mers we found for each

183   morph and together suggests the amount of unique Y chromosome sequence differs across morphs, with

184   the parae morph Y containing the smallest amount of unique genetic material.

185

Read mapping confirms extensive divergence of Y chromosomes

187   To determine how divergent the five Y chromosomes are from one another, we mapped reads from all 39

188   samples to the best full *de novo* genome assembly of each male morph (195 total alignments). Most

189   scaffolds contain autosomal sequence, and coverage is not expected to differ by sex or morph.

190   Meanwhile, scaffolds containing morph specific sequence will have higher coverage by males of the

191   same morph (e.g. immaculata reads mapped to an immaculata assembly) than coverage by males of a

192   different morph (e.g. melanzona reads mapped to an immaculata assembly). Low female coverage of

193   such scaffolds confirms these regions are on the Y and are substantially diverged from the X.

194   As expected, when comparing coverage between males of the same morph as the reference assembly

195   and the other morphs we found average coverage was 1:1 when considering all scaffolds, yet scaffolds

196   enriched for morph-mers (containing >5) had much higher coverage by males of the same morph as the

197   reference (Figure 4). Surprisingly, we found no coverage by immaculata or parae reads for nearly half

198   (40/100) of the scaffolds enriched for melanzona-mers. Similarly, 14 of the 93 scaffolds enriched for

199   immaculata-mers (containing ~48kb) had no coverage when we mapped melanzona and parae reads.

8

200 Meanwhile, in agreement with our morph-mer analysis, all 12 of the scaffolds enriched for parae-mers

201 had nearly equal coverage by melanzona and immaculata reads, once again suggesting that the parae Y

202 contains very little unique Y sequence.

203 We also found the ratio of male:female coverage was much higher for morph-mer scaffolds, many of

204 which had no female coverage, again confirming these complexes of morph specific sequence are

205 located in non-recombining regions of the Y chromosome (Figure 4).

206

207 ***Gene annotation of morph-mer scaffolds***

208 We identified genes on scaffolds with >5 morph-mers. In total, we found 7 genes on the scaffolds

209 containing the 59 Y-mers present in all morphs (totaling 30,558,901 bp), 291 genes on the immaculata

210 scaffolds of sample P09 (totaling 9,748,162bp), 15 genes on the melanzona scaffolds of sample P01

211 (totaling 295,057 bp), and no genes on the parae morph scaffolds of P04 (totaling 127,542 bp)

212 (Supplementary Tables 5, 6 and 7).

213 Only one gene was predicted on scaffolds that were completely unique to melanzona (*trim35)*, and only

214 two genes were predicted on scaffolds completely unique to immaculata (*trim39* and *nlrc3*). Members of

215 the *Trim* gene family act throughout the body and are well known to rapidly evolve novel functions[41,42].

216 Meanwhile, *nlrc3* has been shown to selectively block cellular proliferation and protein synthesis by

217 inhibiting the mTOR signalling pathway[43], which could play a role in keeping immaculata the smallest

218 morph. Several copies of the transcription factor *Tbx3* were present on male unique scaffolds in both

219 melanzona and immaculata morphs. *Tbx* genes play key roles in development and act as developmental

220 switches[44-46], raising the possibility that it could play a role in orchestrating the multi-tissue traits that differ

221 across morphs.

222 We also found several copies of *texim* genes on male scaffolds of melanzona and immaculata that most

223 closely match *texim2* and *texim3*. While the function of *texim2* and *texim3* are largely unknown they have

224 been shown to be highly expressed in the brain and testis of closely related species[47]. In *Xiphophorus*

225 *maculatus*, a close relative to *P. parae* (~45 mya[29]), the transposable element *helitron* has moved *texim1*

226 to the sex determining region of the Y chromosome and duplicated it resulting in three copies of *texim*

9

227 that are expressed specifically in late stage spermatogenesis[47]. The copies of *texim2* and *texim3* that we

228 identified on the Y chromosome of *P. parae* are not the same as those in *X. maculatus* as the Ys arose

229 independently and the *X. maculatus* Y is not chromosome 8. Future analyses are needed to determine

230 the roles of these and other genes in generating the morph-specific phenotypes.

231

232 ***Interspersed repeats and morph-mer scaffolds***

233 We next assessed the accumulation of interspersed elements on the Y chromosomes in our samples.

234 Repeatmasker revealed a large number of transposable elements on scaffolds enriched for the Y-mers

235 (present in all males) and morph-mer-enriched scaffolds (Supplementary Tables 8-12). These include 90

236 copies of the *helitron* transposable element on scaffolds enriched for melanzona-mers, and 38 copies on

237 scaffolds enriched for immaculata-mers.

238 Interspersed repeats indicate the current and/or previous presence of transposable elements in a region

239 and can provide a measure of transposable element activity[48]. We found transposable element activity on

240 autosomes and the X chromosome to be in line with other vertebrate genomes[49], as indicated by

241 interspersed repeats comprising an average of 27.25% of the *de novo* female genomes (Supplementary

242 Table 11). However, the male scaffolds enriched for morph-mers were composed of substantially higher

243 proportions of interspersed repeats (melanzona: $P<0.0001$, $t=6.798$, $df=17$; parae morph: $P<0.0001$,

244 $t=16.20$, $df=10$; immaculata: $P=0.0322$, $t=2.587$, $df=8$) (Extended Data Fig. 6). The proportion of Y

245 sequence comprised of interspersed repeats also differed by morph, with Melanzona (59.28%) being

246 higher than both parae morph (44.11%) ($P=0.0053$, $t=3.193$, $df=17$) and immaculata morph (40.66%)

247 ($P=0.0140$, $t=2.780$, $df=15$) (Supplementary Table 11, Extended Data Fig. 6).

248

249 **Discussion**

250 Recombination is widely regarded as one of the most important processes generating phenotypic

251 diversity as it produces novel allelic combinations upon which selection can act[50]. The loss of

252 recombination is classically assumed to result in reduced genetic diversity through sweeps and

10

253 background selection[1,23,51-58] and have only limited potential for adaptive evolution. The power of these

254 processes to deplete non-recombining regions of diversity is clearly evident in the Y chromosomes of

255 many species[1].

256 Our results, based on both similarity of Y degeneration (Extended Data Fig. 1) and shared Y-mers,

257 together with the recent reports of complete dosage compensation in both *P. parae*[59] and *P. picta*[25], are

258 consistent with recombination suppression between the X and Y chromosome in the common ancestor of

259 *P. parae* and *P. picta* (14.8-18.5 mya[29]) and a highly degenerate Y present at the origin of *P. parae*.

260 Importantly, none of the characteristics that differentiate the immaculata, parae or three melanzona

261 morphs of *P. parae* are found in any close relatives. Therefore, although it is possible that the morphs

262 evolved in an ancestor and were subsequently lost in all lineages except *P. parae*, the more parsimonious

263 explanation is that the genetic basis of the extreme diversity in morphs evolved on a non-recombining,

264 highly degenerate *P. parae* Y chromosome (Extended Data Fig. 1). Given that these morphs differ in a

265 suite of complex traits, including body size, testis size, color pattern, and mating strategy, *P. parae* morph

266 diversity is likely underpinned by either polygenic genetic architectures directly on the Y, or novel genetic

267 elements on the Y that regulate larger regions of the genome, as observed in *Drosophila*[60]. Consistent

268 with complex phenotypic differences between morphs, we identified substantial morph-specific genetic

269 material (Figures 2-4) that was also absent in females and therefore Y-linked.

270 Although male-specific regions of the genome may experience elevated mutation rates[61], it is also likely

271 that the high *P. parae* Y diversity was generated through translocations and/or accumulation of

272 interspersed repeats (eg. transposable element (TE) movement). Translocations have been shown to

273 increase Y-chromosome content[2,62]. In contrast, although TE movement has historically been considered

274 to be a deleterious process, more recent reports have revealed that TEs can alter regions by removing or

275 adding regulatory or coding sequence[44,63-66], and TEs may even act as substrate for novel genes[67]. As

276 predicted for non-recombining regions, we found a large number of TEs in the five *P. parae* Y

277 chromosomes. In all morphs we found substantially more of the Y chromosomes were composed of

278 interspersed repeats compared to the autosomes and X chromosome. Meanwhile the percentage of Y

279 sequence composed of interspersed repeats differed between the morphs, further suggesting that TE

11

280 activity may have played an important role in generating the diversity and divergence of these five non-

281 recombining Y chromosomes.

282

283 ***Making and Maintaining Five Morphs***

284 We found substantial morph specific genetic diversity on the Y chromosome of *P. parae*. Intriguingly, that

285 diversity is maintained within populations, as evidenced by the stability in morph frequencies over

286 repeated surveys spanning 50 years, or roughly 150 generations[13,14,20]. Even if alternative morphs have

287 exactly equal fitness, populations are expected to eventually fix for one morph due to drift[68]. Maintaining

288 alternative morphs within the same population relies on negative frequency dependent selection, thus as

289 one morph decreases in frequency its fitness increases, such as with the three male morphs of the side

290 blotched lizard[69].

291 Previous work suggests that *P. parae* morphs are also under negative frequency dependent selection[14-17],

292 and this could facilitate the establishment and maintenance of five distinct Y chromosomes within the

293 same species. Most new mutations are expected to be lost through drift if they do not confer a high

294 enough fitness advantage over alternative alleles[70], but mutations resulting in a new morph would be at

295 the lowest frequencies and thus have the highest fitness, allowing them to rapidly stabilize in the

296 population. Alternatively, it is possible that the morphs arose in separate populations and only later came

297 into sympatry.

298

299 ***Genetic Basis of Male Reproductive Morphs***

300 Autosomal non-recombining regions have been shown to be associated with alternative reproductive

301 strategies in a range of species[7,9,71-75], yet the formation of the nonrecombining regions underlying

302 alternative reproductive strategies in *P. parae* differs from those previously described. For example, male

303 morphs of the white throated sparrow, which differ in pigmentation and social behavior, are the result of a

304 hybridization event which instantly brought together alternative sequence and halted recombination[72].

305 Importantly, because none of the characteristics of the *P. parae* morphs are found in any close relatives,

12

306 it is unlikely that hybridization is the source of the Y chromosomes we describe. The alternative male

307 morphs in the ruff are also controlled by an autosomal supergene which is composed of two alternative

308 versions of an inverted region[10]. It has yet to be determined whether the diversity across these ruff

309 supergenes pre-dates the inversion or arose after recombination stopped as it did in *P. parae*. A large

310 inverted region is also associated with social morphs in many ant species, this region formed in a

311 common ancestor and has been maintained by balancing selection through repeated speciation events[12].

312 Although it is possible that the multiple male morphs in *P. parae* arose in an ancestor, this is less likely as

313 they have not been observed in any related species to date (including four species that diverged after Y

314 recombination was stopped). The novel formation of the non-recombining regions associated with *P.*

315 *parae* alternative reproductive strategies make this a powerful system for future work to explore the

316 genetic basis of male reproductive morphs.

317

318 **Conclusion**

319 The role of recombination in shaping co-adapted allele complexes has long been an enigma, given that

320 recombination is a key mechanism in generating diverse allelic combinations, yet recombination also acts

321 to break up such combinations. Our results suggest that substantial diversity can be generated without

322 the power of recombination, and the Y chromosome retains remarkable adaptive potential with regard to

323 male phenotypic evolution. Our work indicates that the five Y-linked male morphs of *P. parae* emerged

324 and diverged after recombination was halted, resulting in five unique Y chromosomes within one species.

325 Future work identifying the mechanisms by which morphs are determined by these five Y chromosomes

326 will provide much needed insight to determining which evolutionary forces have led to and shaped these

327 amazing complexes and their co-evolution with the rest of the genome, which is shared across all

328 morphs.

329

330 **Methods**

331

13

332   ***Field Collections and DNA Isolation***

333   To ensure we accounted for natural diversity in the five Y chromosomes of *P. parae*, and because this

334   species is extremely difficult to breed in captivity, we collected all samples (N=40) from three large native

335   populations around Georgetown, Guyana in 2016 (see Supplementary Table 2 and Sandkam, et al. [18] for

336   description of populations) (Environmental Protection Agency of Guyana Permit 120616 SP: 015).

337   Individuals were rapidly sacrificed in MS-222, whole-tail tissue was dissected into EtOH and immediately

338   placed in liquid nitrogen to maintain integrity of high molecular weight DNA. Tissue samples were brought

339   back to the lab and kept at -80º C until high molecular weight DNA extraction. All samples were collected

340   in accordance with national and institutional ethical guidelines (Canadian Council on Animal Care,

341   University of British Columbia).

342   High molecular weight DNA was extracted from 25mg tail tissue of each sample following a modified

343   protocol from 10X Genomics described in Almeida *et al*[28]. Briefly, nuclei were isolated by gently

344   homogenizing tissue with a pestle in cold Nuclei Isolation Buffer from a Nuclei PURE Prep Kit (Sigma).

345   Nuclei were pelleted and supernatant removed before being digested by incubating in 70 µl PBS, 10 µl

346   Proteinase K (Qiagen), and 70 µl Digestion buffer (20mM EDTA, 2mM Tris-HCL, 10mM N-

347   Laurylsarcosine) for two hours at room temperature on a tube rotator. Tween 20 was added (0.1% final

348   concentration) and DNA was bound to SPRIselect magnetic beads (Beckman Coulter) for 20 min. Beads

349   were bound to a magnetic rack and washed twice with 70% Ethanol before eluting DNA. Samples were

350   visually screened for integrity of high molecular weight DNA on an agarose gel. Of the 40 samples

351   extracted, 29 passed initial screening and were used for individual 10X Chromium linked-read sequencing

352   (six female, seven red, five yellow, one blue, four immaculata, and six parae morph). The remaining 11

353   samples (five female, one red, three blue, one immaculata, and one parae morph) were individually

354   sequenced on an Illumina HiSeqX as 2 x 150 bp reads using the v2.5 sequencing chemistry with 300bp

355   inserts and trimmed with trimmomatic (v0.36)[76]. To ensure high coverage of the Y, all 40 samples were

356   individually sequenced to a predicted coverage of 40X (see Supplementary Table 2 for number of reads

357   after filtering), which would result in predicted 20X coverage of the haploid Y.

358

14

10X Chromium Linked-read Sequencing and Assembly

10X Chromium linked-read sequencing was performed at the SciLifeLab, Uppsala Sweden. The 10X Chromium pipeline adds unique tags to each piece of high molecular weight DNA before sequencing on an Illumina platform. These tagged reads were either used directly in the 10X assembly pipeline, or tags were removed using the *basic* function of Longranger v.2.2.2 (10X Genomics), trimmed with trimmomatic (v0.36)[76] and treated as normal Illumina reads[77] for coverage analyses (Supplementary Table 2).

Scaffold level *de novo* genomes were assembled for each of the 29 linked-read samples using the Supernova v2.1.1 software package (10X Genomics) (see Supplementary Table 3 for assembly statistics).

A female chromosome level assembly was created by assigning scaffolds to chromosomal positions. The female with the best *de novo* assembly (largest assembly size and scaffold N50) was used for the Reference Assisted Chromosome Assembly (RACA) pipeline[36]. Briefly, scaffolds were aligned with LASTZ v1.04[78] against high-quality chromosome level genome assemblies of a close relative (*Xiphophorus helleri* v4.0; GenBank accession GCA_003331165) and an outgroup (*Oryzias latipes* v1; GenBank accession GCA_002234675). Alignments were then run through the UCSC chains and nets pipeline from the kentUtils software suite[79] before passing to the RACA pipeline. RACA uses alignments of short-insert and long-insert paired reads that bridge scaffolds to further order and confirm scaffold arrangement. For short-insert data, 150bp reads from the five females sequenced with paired-end Illumina (300bp inserts) were aligned to the target assembly with Bowtie2 v2.2.9[80] reporting concordant mappings only (--no-discordant option). For long-insert data, synthetic 150bp 'pseudo-mate-pair' reads were generated from the *de novo* scaffolds of the 6 female *P. parae* samples sequenced with Chromium linked-reads. To increase the likelihood that bridge reads spanned scaffolds, we generated two long-insert pseudo-mate-pair libraries for each of the six *de novo* female genomes, a 2.3kb insert library and a 15kb insert library, and aligned these to the target assembly. RACA then used the information from both the phylogenetically weighted genome pairwise alignments, and the read mapping data to order the target scaffolds into longer predicted chromosome fragments.

385 To identify which chromosome is the sex chromosome and determine the extent of X-Y divergence, we

386 mapped reads from all 40 individuals to the female scaffolds that had RACA generated chromosome

387 annotations using the *aln* function of *bwa* (v0.6.1)[81]. Alignments were filtered for uniquely mapped reads

388 and average scaffold coverage was calculated using soap.coverage v2.7.7

389 (http://soap.genomics.org.cn/). To account for differences across individuals in sequencing library size,

390 we divided the coverage of each scaffold by the average coverage across all scaffolds for each individual.

391 Male to female (M:F) fold change in coverage of each scaffold was calculated for all males, and each of

392 the five morphs as $\log_2$(average male coverage) – $\log_2$(average female coverage)[82]. Upon observing

393 chromosome 8 was the sex chromosome, we calculated the 95% CI for M:F coverage by bootstrapping

394 across all scaffolds which RACA placed on the autosomes (1000 replicates; mean of 20 scaffolds without

395 replacement).

396

397 ***Identifying Morph specific sequence by k-mers***

398 To locate morph specific sequence, we identified morph specific *k*-mers (morph-mers) and then mapped

399 these to the respective *de novo* genome assemblies. First Jellyfish v2.2.3[81] was used to identify all 31bp

400 *k*-mers from the 'megabubble' output of each of the 22 male *de novo* genome assemblies from

401 supernova. We next identified the putative Y-linked *k*-mers in each sample by removing all *k*-mers present

402 in any of the females. For female *k*-mer identification we conservatively took *k*-mers from both the

403 megabubble outputs and the raw Illumina reads of all 11 female individuals, which we used to identify

404 every 31bp *k*-mer present >3 times (this maximized the chance of including *k*-mers from unassembled

405 regions of female genomes but minimized *k*-mers from sequencing errors[40]). The *k*-mers present in

406 females all occur either on autosomes or the X chromosome, therefore by removing the female *k*-mers

407 from all *k*-mers identified in males we are left with putative Y-linked *k*-mers that we call Y-mers[25,40]. To

408 exclude *k*-mers representing autosomal SNPs unique to a single male we required Y-mers be present in

409 at least two individuals. Since all female sequence is theoretically present in males, we validated our

410 method by identifying all *k*-mers from the six female megabubbles that were not present in the male

411 Illumina reads and found how many were present in at least two female individuals.

16

412   We then combined all the Y-mers we found with those we previously identified in *P. picta*[25] and used the

413   presence of Y-mers as character states to build a phylogeny of all individuals and *P. picta*. Two runs of

414   MrBayes v3.2.2[83] were run for 100,000 generations with Y-mers treated akin to restriction sites (model

415   F81 with rates set to equal and default priors). The SD of the split frequencies between runs reached 0

416   indicating both runs converged on identical and robust trees.

417   Monophyletic clades were recovered for each of the major morphs (immaculata, parae and melanzona).

418   We then identified unique Y-mers in each clade (Y-mers present in every individual of that clade but not

419   present outside that clade). This approach provided us with all morph-mers (Y-mers present in every

420   individual of a given morph but not present in any individual of the other morphs).

421   These morph-mers reveal two insights: (1) at a gross level they provide a sense of how much Y sequence

422   is shared within versus across morphs and (2) mapping these morph-mers to the respective *de novo*

423   genomes allows us to identify regions of morph specific sequence[39,40,84,85]. To find these regions of morph

424   specific sequence we first mapped the corresponding set of 31bp morph-mers to each of the 22 *de novo*

425   male genomes (pseudohap style of Supernova output) using bowtie2[80] allowing for no mismatches, gaps,

426   or trimming. We found morph-mers disproportionately map to scaffolds, indicating they came from regions

427   of highly diverged morph-specific sequence rather than evenly dispersed lowly diverged sequence

428   (Extended Data Fig. 3 and 4).

429   To verify our pipeline was identifying true Y-specific alignments we attempted to align the Y-mers and all

430   of the morph-mers to each of the six female *de novo* genome assemblies using bowtie2[80] (see above)

431   and found no alignments could be made. We next verified that our morph-mers were targeting morph

432   specific sequence by attempting to align each of the morph-mer datasets to individuals of the opposite

433   morphs and again found no alignments could be made.

434

435   ***Coverage analysis***

436   To independently verify that the scaffolds identified by our *k*-mer approach contained highly diverged

437   sequence, we aligned each of the 39 individuals to the individual with the best *de novo* genome of each

438   morph (based on assembly size, scaffold N50, and contig N50, Supplementary Table 3). Alignments were

17

439 generated with the *mem* function of *bwa* (v0.7.17)[81]. *samtools* (v1.10)[86] was used to remove unmapped

440 reads and secondary alignments with the *fixmate* function, and duplicates were removed with *markdup*.

441 *bamqc* was then used to assess distribution of map quality. For each individual, the average coverage of

442 each scaffold by reads with mapq $\geq$60 was determined using the *depth* function of *samtools* (v1.10)[86]. Y

443 chromosomes are notorious for high incidence of transposable elements and repeats[1], this highly

444 conservative filtering decreased false alignments to these regions. To account for differences across

445 individuals in sequencing library size, we took the coverage of each scaffold divided by that individual's

446 coverage across all scaffolds. The average raw scaffold coverage across all individuals was 29.97X,

447 therefore any scaffold with a corrected coverage < 0.025 (raw coverage <1X) was considered to have a

448 coverage of 0.

449

450 ***Gene Annotation***

451 To identify genes on morph specific scaffolds we followed the pipeline described in Almeida *et al*[28].

452 Briefly, we took a very conservative approach by annotating only the scaffolds with >5 morph-mers from

453 each of the *de novo* references used for the coverage analysis (one of each morph). The chance of a

454 scaffold containing any particular 31bp *k*-mer depends on the length of the scaffold and can be calculated

455 roughly as 0.25^{31} x scaffold length. The longest male scaffold we recovered was 19,887,348 and

456 therefore had the greatest chance of containing any given Y-mer; $4.31*10^{-12}$. The most abundant morph-

457 mers were melanzona-mers (87,629), therefore the likelihood of the largest scaffold containing one

458 melanzona-mer by chance was $4.31*10^{-12}$ x 87,629 = $3.78*10^{-7}$ and the likelihood it contains 5

459 melanzona-mers purely by chance was roughly $7.71*10^{-33}$. If we conservatively assume that all scaffolds

460 have the same probability of containing a Y-mer by chance and the male with the most scaffolds had

461 25,416 scaffolds – there was a likelihood of $1.96*10^{-28}$ that a scaffold was incorrectly identified.

462 We then annotated these scaffolds with MAKER v2.31.10[87]. We ran the MAKER pipeline twice: first

463 based on a guppy-specific repeat library, protein sequence, EST and RNA sequence data (later used to

464 train *ab-initio* software) and a second time combining evidence data from the first run and *ab-initio*

465 predictions. We created a repeat library for these scaffolds using *de novo* repeats identified by

18

466   RepeatModeler v1.0.10[88] which we then combined with Actinopterygii-specific repeats to use with

467   RepeatMasker v4.0.7[89]. Annotated protein sequences were downloaded from Ensembl (release 95)[90] for

468   8 fish species: *Danio rerio* (GRCz11), *Gasterosteus aculeatus* (BROADS1), *Oryzias latipes*

469   (ASM223467v1), *Poecilia latipinna* (1.0), *Poecilia mexicana* (1.0), *Poecilia reticulata* (1.0), *Takifugu*

470   *rubripes* (FUGU5) and *Xiphophorus maculatus* (5.0). For ESTs, we used 10,664 tags isolated from guppy

471   embryos and male testis[91]. Furthermore, to support gene predictions we also used two publicly available

472   libraries of RNA-seq data collected from guppy male testis and male embryos[92] and assembled with

473   StringTie 1.3.3b[93]. As basis for the construction of gene models, we combined *ab-initio* predictions from

474   Augustus v3.2.3[94], trained via BUSCO v3.0.2[95], and SNAP v2006-07-28[96]. To train Augustus and SNAP,

475   we ran the MAKER pipeline a first time to create a profile using the protein and EST evidence along with

476   RNA-seq data. Both Augustus and SNAP were then trained from this initial evidence-based annotation.

477   Functional inference for genes and transcripts was performed using the translated CDS features of each

478   coding transcript. Gene names and protein functions were retrieved using BLASTp to search the

479   Uniprot/Swissprot, InterProscan v5 and GenBank databases.

480

481   ***Identifying Transposable Element Activity***

482   To compare activity of transposable elements in morph-linked Y scaffolds to the rest of the genome

483   (autosomes and the X chromosome) we identified the proportion of sequence comprised of interspersed

484   repeats. Sequences identified as interspersed repeats are both current transposable elements and the

485   flanking sequence that is left behind when transposable elements leave a region, and thus act as a

486   measure of transposable element activity within a region[48]. To build a repeat library for all *P. parae*

487   repeats we first used RepeatModeler v1.0.10 to create repeat libraries from the best female *de novo*

488   genome, and from the morph-linked scaffolds (scaffolds with >5 morph-mers) from each male. We then

489   combined individual *de novo* repeat libraries with the Actinopterygii-specific repeats to create the full *P.*

490   *parae* repeat library. This combined repeat library was used to identify interspersed repeats in the *de*

491   *novo* genome from each female and the morph-linked scaffolds of each male using RepeatMasker v4.0.7.

492   The percentage of total interspersed repeats was compared with a one-way ANOVA in Prism v.9.0.0

19

493   which revealed significant differences ($P<0.0001$, $F(3,25)=17.98$). Unpaired t-test were then used to

494   follow up the significant main effect.

495

496   ***Confirming diversity is unique to Y chromosome***

497   To further confirm the extreme divergence across morphs is unique to the Y chromosome (and not cryptic

498   subpopulations) we built phylogenies of each of the autosomes. To do this we aligned reads from each

499   individual to the female scaffolds that had RACA generated chromosome annotations using the *mem*

500   function of *bwa* (v0.6.1)[81]. Duplicates were marked with *samtools* (v1.10)[86]. Variants were called using the

501   *call* function of *bcftools* (v1.11) before generating consensus sequences of the longest scaffold from each

502   autosome using the consensus function of *bcftools* with the -I flag (providing IUPAC ambiguity codes for

503   polymorphic sites). For each autosome, the sequence from all 40 individuals were aligned using the

504   MAFFT[97] plugin for Geneious Prime with default parameters. Approximate-maximum-likelihood trees

505   were generated using FastTree (v2.1.12)[98] plugin for Geneious with default parameters (Extended Data

506   Fig. 7).

507

508   **Data Availability**

509   All of the data generated for this study are archived in the Sequence Read Archive (under BioProject ID

510   _) at the National Centre for Biotechnology Information (www.ncbi.nlm.nih.gov/sra).

511   **Code Availability**

512   All scripts and pipelines are available at https://github.com/manklab/Poecilia_parae_Y_Diversity

513

20

514   **References**

515

516   1   Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-
517   chromosome degeneration. *Nat. Rev. Genet.* **14**, 113-124, doi:10.1038/nrg3366 (2013).
518   2   Tobler, R., Nolte, V. & Schlotterer, C. High rate of translocation-based gene birth on the
519   Drosophila Y chromosome. *P Natl Acad Sci USA* **114**, 11721-11726,
520   doi:10.1073/pnas.1706502114 (2017).
521   3   Mahajan, S. & Bachtrog, D. Convergent evolution of Y chromosome gene content in
522   flies. *Nat Commun* **8**, 785, doi:10.1038/s41467-017-00653-x (2017).
523   4   Bachtrog, D., Mahajan, S. & Bracewell, R. Massive gene amplification on a recently
524   formed Drosophila Y chromosome. *Nat Ecol Evol* **3**, 1587-1597, doi:10.1038/s41559-
525   019-1009-9 (2019).
526   5   Hall, A. B. *et al.* Radical remodeling of the Y chromosome in a recent radiation of malaria
527   mosquitoes. *P Natl Acad Sci USA* **113**, E2114-2123, doi:10.1073/pnas.1525164113
528   (2016).
529   6   Todesco, M. *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers.
530   *Nature*, doi:10.1038/s41586-020-2467-6 (2020).
531   7   Schwander, T., Libbrecht, R. & Keller, L. Supergenes and Complex Phenotypes. *Curr. Biol.*
532   **24**, R288-R294, doi:10.1016/j.cub.2014.01.056 (2014).
533   8   Wang, J. *et al.* A Y-like social chromosome causes alternative colony organization in fire
534   ants. *Nature* **493**, 664-668, doi:10.1038/nature11832 (2013).
535   9   Lamichhaney, S. *et al.* Structural genomic changes underlie alternative reproductive
536   strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84-88, doi:10.1038/ng.3430
537   (2016).
538   10   Kupper, C. *et al.* A supergene determines highly divergent male reproductive morphs in
539   the ruff. *Nat. Genet.* **48**, 79-83, doi:10.1038/ng.3443 (2016).
540   11   Branco, S. *et al.* Multiple convergent supergene evolution events in mating-type
541   chromosomes. *Nature Communications* **9**, 2000, doi:10.1038/s41467-018-04380-9
542   (2018).
543   12   Yan, Z. *et al.* Evolution of a supergene that regulates a trans-species social
544   polymorphism. *Nat Ecol Evol* **4**, 240-249, doi:10.1038/s41559-019-1081-1 (2020).
545   13   Lindholm, A. K., Brooks, R. & Breden, F. Extreme polymorphism in a Y-linked sexually
546   selected trait. *Heredity (Edinb)* **92**, 156-162, doi:10.1038/sj.hdy.6800386 (2004).
547   14   Hurtado-Gonzales, J. L. & Uy, J. A. Alternative mating strategies may favour the
548   persistence of a genetically based colour polymorphism in a pentamorphic fish. *Anim.*
549   *Behav.* **77**, 1187-1194, doi:10.1016/j.anbehav.20 (2009).
550   15   Hurtado-Gonzales, J. L. & Uy, J. A. Intrasexual competition facilitates the evolution of
551   alternative mating strategies in a colour polymorphic fish. *BMC Evol. Biol.* **10**, 391,
552   doi:10.1186/1471-2148-10-391 (2010).
553   16   Hurtado-Gonzales, J. L., Baldassarre, D. T. & Uy, J. A. Interaction between female mating
554   preferences and predation may explain the maintenance of rare males in the

21

555   pentamorphic fish *Poecilia parae*. *J. Evol. Biol.* **23**, 1293-1301, doi:10.1111/j.1420-
556   9101.2010.01995.x (2010).

557   17   Hurtado-Gonzales, J. L., Loew, E. R. & Uy, J. A. Variation in the visual habitat may
558   mediate the maintenance of color polymorphism in a poeciliid fish. *PLoS One* **9**,
559   e101497, doi:10.1371/journal.pone.0101497 (2014).

560   18   Sandkam, B. A., Young, C. M., Breden, F. M., Bourne, G. R. & Breden, F. Color vision
561   varies more among populations than among species of live-bearing fish from South
562   America. *BMC Evol. Biol.* **15**, 225, doi:10.1186/s12862-015-0501-3 (2015).

563   19   Bourne, G. R., Breden, F. & Allen, T. C. Females prefer carotenoid colored males as
564   mates in the pentamorphic livebearing fish, *Poecilia parae*. *Naturwissenschaften* **90**,
565   402-405, doi:10.1007/s00114-003-0444-1 (2003).

566   20   Liley, N. R. *Reproductive Isolation in some sympatric species of fishes* Doctor of
567   Philosophy thesis, Oxford, (1963).

568   21   Bachtrog, D. *et al.* Are all sex chromosomes created equal? *Trends Genet.* **27**, 350-357,
569   doi:10.1016/j.tig.2011.05.005 (2011).

570   22   Rice, W. R. Evolution of the Y Sex Chromosome in Animals. *Bioscience* **46**, 331-343,
571   doi:10.2307/1312947 (1996).

572   23   Wright, A. E., Dean, R., Zimmer, F. & Mank, J. E. How to make a sex chromosome. *Nat*
573   *Commun* **7**, 12087, doi:10.1038/ncomms12087 (2016).

574   24   Furman, B. L. S. *et al.* Sex Chromosome Evolution: So Many Exceptions to the Rules.
575   *Genome Biol Evol* **12**, 750-763, doi:10.1093/gbe/evaa081 (2020).

576   25   Darolti, I. *et al.* Extreme heterogeneity in sex chromosome differentiation and dosage
577   compensation in livebearers. *Proceedings of the National Academy of Sciences* **116**,
578   19031-19036, doi:10.1073/pnas.1905298116 (2019).

579   26   Wright, A. *et al.* Convergent recombination suppression suggests a role of sexual
580   selection in guppy sex chromosome formation. *Nature Communications* **8**, 14251 (2017).

581   27   Darolti, I., Wright, A. E. & Mank, J. E. Guppy Y Chromosome Integrity Maintained by
582   Incomplete Recombination Suppression. *Genome Biol Evol* **12**, 965-977,
583   doi:10.1093/gbe/evaa099 (2020).

584   28   Almeida, P. *et al.* Divergence and Remarkable Diversity of the Y Chromosome in
585   Guppies. *bioRxiv*, doi:10.1101/2020.07.13.200196 (2020).

586   29   Rabosky, D. L. *et al.* An inverse latitudinal gradient in speciation rate for marine fishes.
587   *Nature* **559**, 392-395, doi:10.1038/s41586-018-0273-1 (2018).

588   30   Reznick, D. N., Miles, D. B. & Winslow, S. Life History of *Poecilia picta* (Poeciliidae) from
589   the Island of Trinidad. *Copeia* **1992**, 782-790, doi:10.2307/1446155 (1992).

590   31   Haskins, C. P. & Haskins, E. F. The role of sexual selection as an isolating mechanism in
591   three species of Poeciliid fishes. *Evolution* **3**, 160-169 (1949).

592   32   Liley, N. R. Ethological Isolating Mechanisms in Four Sympatric Species of Poeciliid
593   Fishes. *Behaviour Supplement* **13**, 1-197 (1966).

594   33   Vicoso, B. & Bachtrog, D. Reversal of an ancient sex chromosome to an autosome in
595   Drosophila. *Nature* **499**, 332-335, doi:10.1038/nature12235 (2013).

596   34   Vicoso, B. & Bachtrog, D. Numerous transitions of sex chromosomes in Diptera. *PLoS*
597   *Biol.* **13**, e1002078, doi:10.1371/journal.pbio.1002078 (2015).

22

598 35 Vicoso, B., Emerson, J. J., Zektser, Y., Mahajan, S. & Bachtrog, D. Comparative sex
599   chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global
600   dosage compensation. *PLoS Biol.* **11**, e1001643, doi:10.1371/journal.pbio.1001643
601   (2013).
602 36 Kim, J. *et al.* Reference-assisted chromosome assembly. *Proceedings of the National*
603   *Academy of Sciences* **110**, 1785-1790, doi:10.1073/pnas.1220349110 (2013).
604 37 Pucholt, P., Wright, A. E., Conze, L. L., Mank, J. E. & Berlin, S. Recent Sex Chromosome
605   Divergence despite Ancient Dioecy in the Willow Salix viminalis. *Mol. Biol. Evol.* **34**,
606   1991-2001, doi:10.1093/molbev/msx144 (2017).
607 38 Akagi, T., Henry, I. M., Tao, R. & Comai, L. A Y-chromosome-encoded small RNA acts as a
608   sex determinant in persimmons. *Science* **346** (2014).
609 39 Torres, M. F. *et al.* Genus-wide sequencing supports a two-locus model for sex-
610   determination in Phoenix. *Nat Commun* **9**, 3969, doi:10.1038/s41467-018-06375-y
611   (2018).
612 40 Morris, J., Darolti, I., Bloch, N. I., Wright, A. E. & Mank, J. E. Shared and Species-Specific
613   Patterns of Nascent Y Chromosome Evolution in Two Guppy Species. *Genes (Basel)* **9**,
614   doi:10.3390/genes9050238 (2018).
615 41 Napolitano, L. M. & Meroni, G. TRIM family: Pleiotropy and diversification through
616   homomultimer and heteromultimer formation. *IUBMB Life* **64**, 64-71,
617   doi:10.1002/iub.580 (2012).
618 42 Sardiello, M., Cairo, S., Fontanella, B., Ballabio, A. & Meroni, G. Genomic analysis of the
619   TRIM family reveals two groups of genes with distinct evolutionary properties. *BMC*
620   *Evol. Biol.* **8**, 225, doi:10.1186/1471-2148-8-225 (2008).
621 43 Karki, R. *et al.* NLRC3 is an inhibitory sensor of PI3K-mTOR pathways in cancer. *Nature*
622   **540**, 583-587, doi:10.1038/nature20597 (2016).
623 44 Sandkam, B. A. *et al. Tbx2a* Modulates Switching of RH2 and LWS Opsin Gene
624   Expression. *Mol. Biol. Evol.* **37**, 2002-2014, doi:10.1093/molbev/msaa062 (2020).
625 45 Showell, C., Christine, K. S., Mandel, E. M. & Conlon, F. L. Developmental expression
626   patterns of *Tbx1*, *Tbx2*, *Tbx5*, and *Tbx20* in *Xenopus tropicalis*. *Dev. Dyn.* **235**, 1623-1630,
627   doi:10.1002/dvdy.20714 (2006).
628 46 Gibson-Brown, J. J., S, I. A., Silver, L. M. & Papaioannou, V. E. Expression of T-box genes
629   Tbx2-Tbx5 during chick organogenesis. *Mech Dev* **74**, 165-169 (1998).
630 47 Tomaszkiewicz, M., Chalopin, D., Schartl, M., Galiana, D. & Volff, J.-N. A multicopy Y-
631   chromosomal SGNH hydrolase gene expressed in the testis of the platyfish has been
632   captured and mobilized by a Helitron transposon. *BMC Genet.* **15** (2014).
633 48 Smit, A. F. Interspersed repeats and other mementos of transposable elements in
634   mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657-663, doi:10.1016/s0959-
635   437x(99)00031-3 (1999).
636 49 Sotero-Caio, C. G., Platt, R. N., 2nd, Suh, A. & Ray, D. A. Evolution and Diversity of
637   Transposable Elements in Vertebrate Genomes. *Genome Biol Evol* **9**, 161-177,
638   doi:10.1093/gbe/evw264 (2017).
639 50 Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737-756
640   (1974).

51   Lenormand, T., Fyon, F., Sun, E. & Roze, D. Sex Chromosome Degeneration by Regulatory Evolution. *Curr. Biol.*, doi:10.1016/j.cub.2020.05.052 (2020).

52   Hough, J., Wang, W., Barrett, S. C. H. & Wright, S. I. Hill-Robertson Interference Reduces Genetic Diversity on a Young Plant Y-Chromosome. *Genetics* **207**, 685 (2017).

53   Marais, G. A. *et al.* Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Curr. Biol.* **18**, 545-549, doi:10.1016/j.cub.2008.03.023 (2008).

54   Bachtrog, D. & Charlesworth, B. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**, 323-326, doi:10.1038/416323a (2002).

55   Mank, J. E. Small but mighty: the evolutionary dynamics of W and Y sex chromosomes. *Chromosome Res* **20**, 21-33, doi:10.1007/s10577-011-9251-2 (2012).

56   Kaiser, V. B. & Charlesworth, B. Muller's ratchet and the degeneration of the Drosophila miranda neo-Y chromosome. *Genetics* **185**, 339-348, doi:10.1534/genetics.109.112789 (2010).

57   Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89-92, doi:10.1038/nature05049 (2006).

58   Ritz, K. R., Noor, M. A. F. & Singh, N. D. Variation in Recombination Rate: Adaptive or Not? *Trends Genet.* **33**, 364-374, doi:10.1016/j.tig.2017.03.003 (2017).

59   Metzger, D. C. H., Sandkam, B. A., Darolti, I. & Mank, J. E. Rapid evolution of complete dosage compensation in *Poecilia*. *BioRxiv*, doi:10.1101/2021.02.12.431036 (2021).

60   Lemos, B., Araripe, L. O. & Hartl, D. L. Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* **319**, 91-93, doi:10.1126/science.1148861 (2008).

61   Ellegren, H. Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc Biol Sci* **274**, 1-10, doi:10.1098/rspb.2006.3720 (2007).

62   Tennessen, J. A. *et al.* Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biol.* **16**, e2006062, doi:10.1371/journal.pbio.2006062 (2018).

63   Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol* **19**, 199, doi:10.1186/s13059-018-1577-z (2018).

64   Carleton, K. L. *et al.* Movement of transposable elements contributes to cichlid diversity. doi:10.1101/2020.02.26.961987 (2020).

65   Brawand, D. *et al.* The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375-+, doi:10.1038/nature13726 (2014).

66   Auvinet, J. *et al.* Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: the case for the Antarctic teleost genus Trematomus. *BMC Genomics* **19**, 339, doi:10.1186/s12864-018-4714-x (2018).

67   Naville, M. *et al.* Not so bad after all: retroviruses and long terminal repeat retrotransposons as a source of new genes in vertebrates. *Clin. Microbiol. Infect.* **22**, 312-323, doi:10.1016/j.cmi.2016.02.001 (2016).

68   Sinervo, B. & Calsbeek, R. The Developmental, Physiological, Neural, and Genetical Causes and Consequences of Frequency-Dependent Selection in the Wild. *Annual Review of Ecology, Evolution, and Systematics* **37**, 581-610, doi:10.1146/annurev.ecolsys.37.091305.110128 (2006).

24

685 69 Sinervo, B. & Lively, C. M. The rock-paper-scissors game and the evolution of alternative
686   male strategies. *Nature* **380**, 240-243, doi:DOI 10.1038/380240a0 (1996).
687 70 Hartl, D. L. & Clark, A. G. *Principles of Population Genetics*. (Sinauer Associates, 1997).
688 71 Lank, D. B., Smith, C. M., Hanotte, O., Burke, T. & Cooke, F. Genetic polymorphism for
689   alternative mating behaviour in lekking male ruff *Philomachus pugnax*. *Nature* **378**, 59-
690   62, doi:DOI 10.1038/378059a0 (1995).
691 72 Tuttle, E. M. *et al.* Divergence and Functional Degradation of a Sex Chromosome-like
692   Supergene. *Curr. Biol.* **26**, 344-350, doi:10.1016/j.cub.2015.11.069 (2016).
693 73 Hunt, B. G. Supergene Evolution: Recombination Finds a Way. *Curr. Biol.* **30**, R73-R76,
694   doi:10.1016/j.cub.2019.12.006 (2020).
695 74 Charlesworth, D. The status of supergenes in the 21st century: recombination
696   suppression in Batesian mimicry and sex chromosomes and other complex adaptations.
697   *Evol Appl* **9**, 74-90, doi:10.1111/eva.12291 (2016).
698 75 Joron, M. *et al.* Chromosomal rearrangements maintain a polymorphic supergene
699   controlling butterfly mimicry. *Nature* **477**, 203-206, doi:10.1038/nature10341 (2011).
700 76 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
701   sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170
702   (2014).
703 77 Elyanow, R., Wu, H.-T. & Raphael, B. J. Identifying structural variants using linked-read
704   sequencing data. *Bioinformatics (Oxford, England)* **34**, 353-360,
705   doi:10.1093/bioinformatics/btx712 (2018).
706 78 Harris, R. S. *Improved pairwise alignment of genomic DNA* Doctor of Philosophy thesis,
707   Penn State, (2007).
708 79 Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron:
709   duplication, deletion, and rearrangement in the mouse and human genomes. *P Natl*
710   *Acad Sci USA* **100**, 11484-11489, doi:10.1073/pnas.1932072100 (2003).
711 80 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods*
712   **9**, 357 (2012).
713 81 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler
714   transform. *Bioinformatics* **26**, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
715 82 Palmer, D. H., Rogers, T. F., Dean, R. & Wright, A. E. How to identify sex chromosomes
716   and their turnover. *Mol. Ecol.* **28**, 4709-4724, doi:10.1111/mec.15245 (2019).
717 83 Ronquist, F. *et al.* MrBayes 3.2: efficient Bayesian phylogenetic inference and model
718   choice across a large model space. *Syst. Biol.* **61**, 539-542, doi:10.1093/sysbio/sys029
719   (2012).
720 84 Carvalho, A. B. & Clark, A. G. Efficient identification of Y chromosome sequences in the
721   human and Drosophila genomes. *Genome Res.* **23**, 1894-1907,
722   doi:10.1101/gr.156034.113 (2013).
723 85 Carvalho, A. B., Vicoso, B., Russo, C. A., Swenor, B. & Clark, A. G. Birth of a new gene on
724   the Y chromosome of *Drosophila melanogaster*. *P Natl Acad Sci USA* **112**, 12450-12455,
725   doi:10.1073/pnas.1516543112 (2015).
726 86 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
727   2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

25

728   87   Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database
729   management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491,
730   doi:10.1186/1471-2105-12-491 (2011).
731   88   RepeatModeler Open-1.0 (http://www.repeatmasker.org, 2015).
732   89   RepeatMasker Open-4.0 (http://www.repeatmasker.org, 2015).
733   90   Howe, K. L. *et al.* Ensembl Genomes 2020-enabling non-vertebrate genomic research.
734   *Nucleic Acids Res.* **48**, D689-D695, doi:10.1093/nar/gkz890 (2020).
735   91   Dreyer, C. *et al.* ESTs and EST-linked polymorphisms for genetic mapping and
736   phylogenetic reconstruction in the guppy, *Poecilia reticulata*. *BMC Genomics* **8**, 269,
737   doi:10.1186/1471-2164-8-269 (2007).
738   92   Sharma, E. *et al.* Transcriptome assemblies for studying sex-biased gene expression in
739   the guppy, *Poecilia reticulata*. *BMC Genomics* **15**, 400, doi:10.1186/1471-2164-15-400
740   (2014).
741   93   Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from
742   RNA-seq reads. *Nat. Biotechnol.* **33**, 290-295, doi:10.1038/nbt.3122 (2015).
743   94   Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids*
744   *Res.* **34**, W435-439, doi:10.1093/nar/gkl200 (2006).
745   95   Seppey, M., Manni, M. & Zdobnov, E. M. in *Gene Prediction*   227-245 (Springer, 2019).
746   96   Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59, doi:10.1186/1471-
747   2105-5-59 (2004).
748   97   Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT.
749   *Methods Mol Biol* **537**, 39-64, doi:10.1007/978-1-59745-251-9_3 (2009).
750   98   Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2--approximately maximum-likelihood
751   trees for large alignments. *PLoS One* **5**, e9490, doi:10.1371/journal.pone.0009490
752   (2010).

753

763

Author Contributions

B.A.S. and J.E.M. designed research; B.A.S., J.E.M, F.B., G.R.B. conducted field work; B.A.S., P.A., I.A.,

B.L.S.F., W.v.d.B., and J.M. conducted bioinformatic analyses; B.A.S., P.A., I.A., B.L.S.F., W.v.d.B., J.M.,

G.R.B., F.B. and J.E.M. wrote the paper.

Competing Interests

The authors declare no competing interests.

27
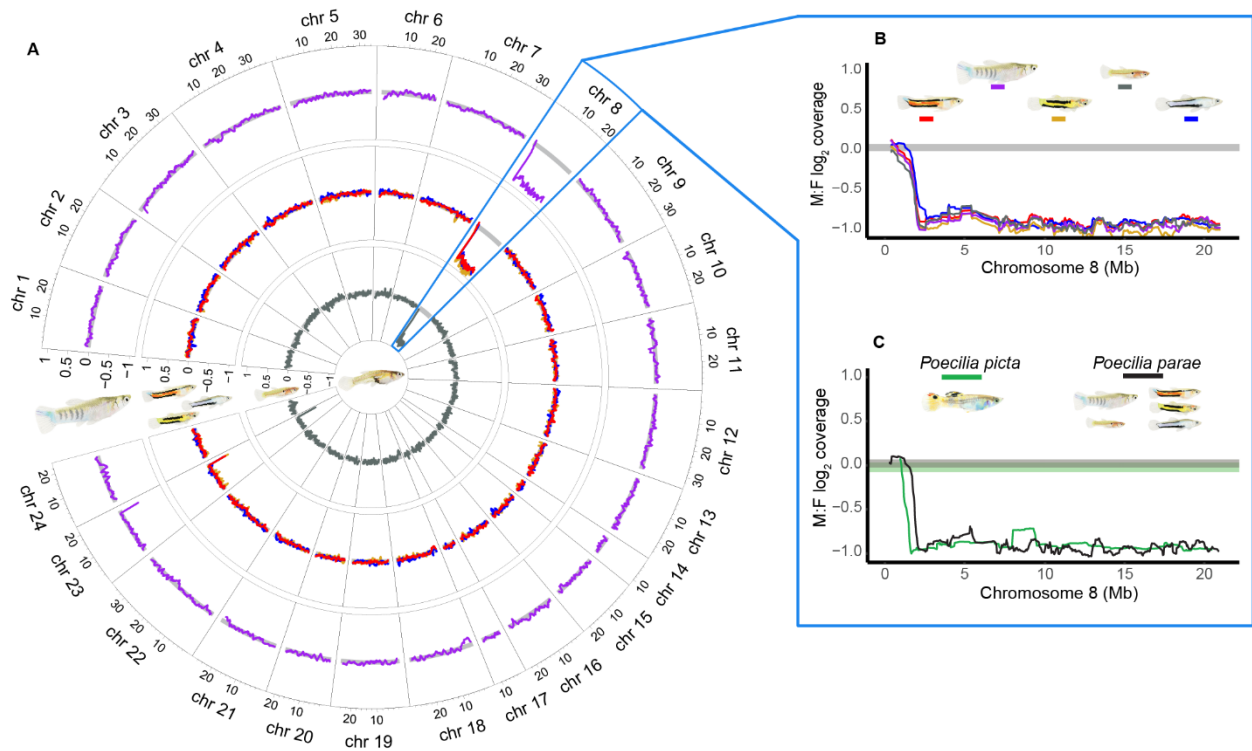
Figure 1. Coverage differences between the sexes (male:female log$_2$) for female scaffolds placed by RACA on the reference *Xiphophorus hellerii* chromosomes. (A) Average immaculata (inner ring), the three melanzona (middle ring) and parae morphs (outer ring) plotted across all chromosomes. Highlighted in blue is *X. hellerii* chromosome 8 which is syntenic to the guppy sex chromosome (*P. reticulata* chromosome 12). The decreased male coverage of chromosome 8 indicates this is also the sex chromosome in *P. parae*. (B) All five *P. parae* morphs share the same pattern of XY divergence, indicating a shared history of recombination suppression. (C) Pattern of *P. parae* XY divergence is the same as the sister species *P. picta*, indicating recombination was stopped in the common ancestor of *P. parae* and *P. picta* (14.8-18.5 mya[29]). In each, horizontal grey-shaded areas represent the 95% confidence intervals based on bootstrap estimates across the autosomes.

806
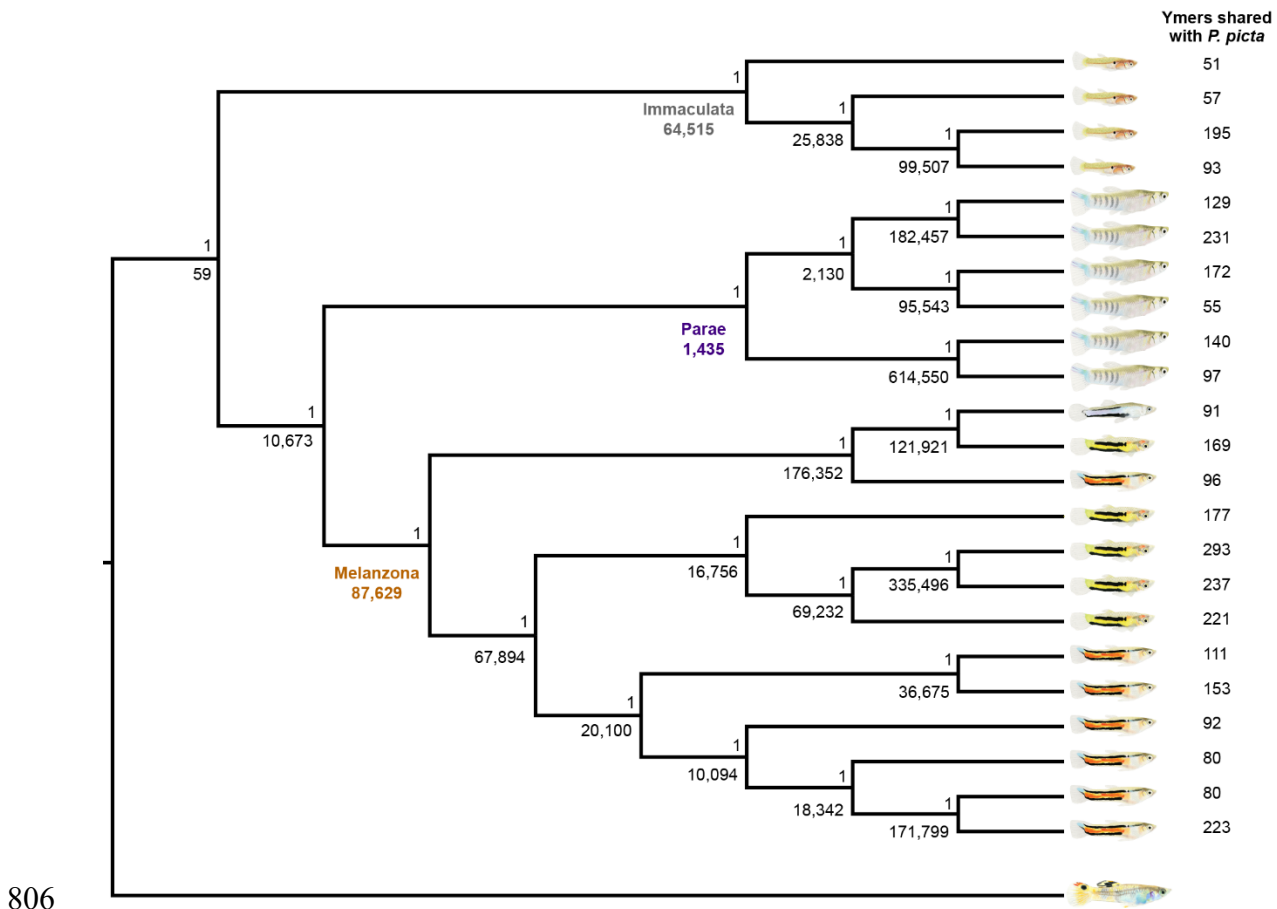
Figure 2. Bayesian Y chromosome phylogeny based on presence/absence of the 27,950,090 *P. parae* Y-mers and 1,646 *P. picta* Y-mers[25] in each individual and rooted on *P. picta*. The posterior probability is presented above each node, below the node is the number of *P. parae* Y-mers unique to all members of that clade. The three major morphs of *P. parae* (immaculata, parae and melanzona) formed distinct clades and the Y-mers unique to all members of these clades are called morph-mers.

807

808

809

810

811

812

813

814

815

816

817

29
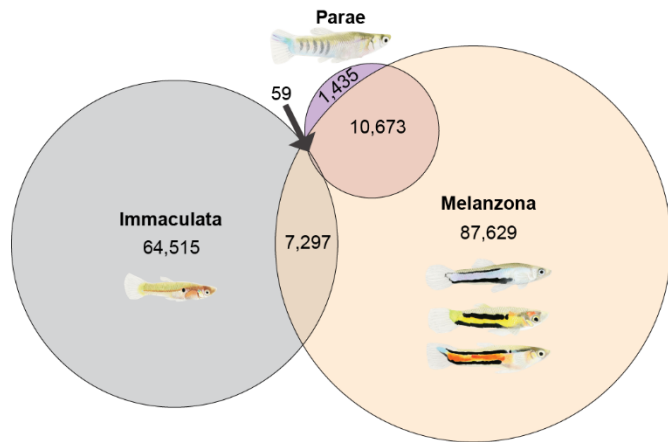
Figure 3. The distribution of the 27,950,090 *P. parae* Y-mers reveals strong differences across morphs. While there are very few Y-mers present in all morphs, each morph harbors unique Y-mers. The melanzona and parae morphs share more Y-mers with one another than either share with immaculata.

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

30

836
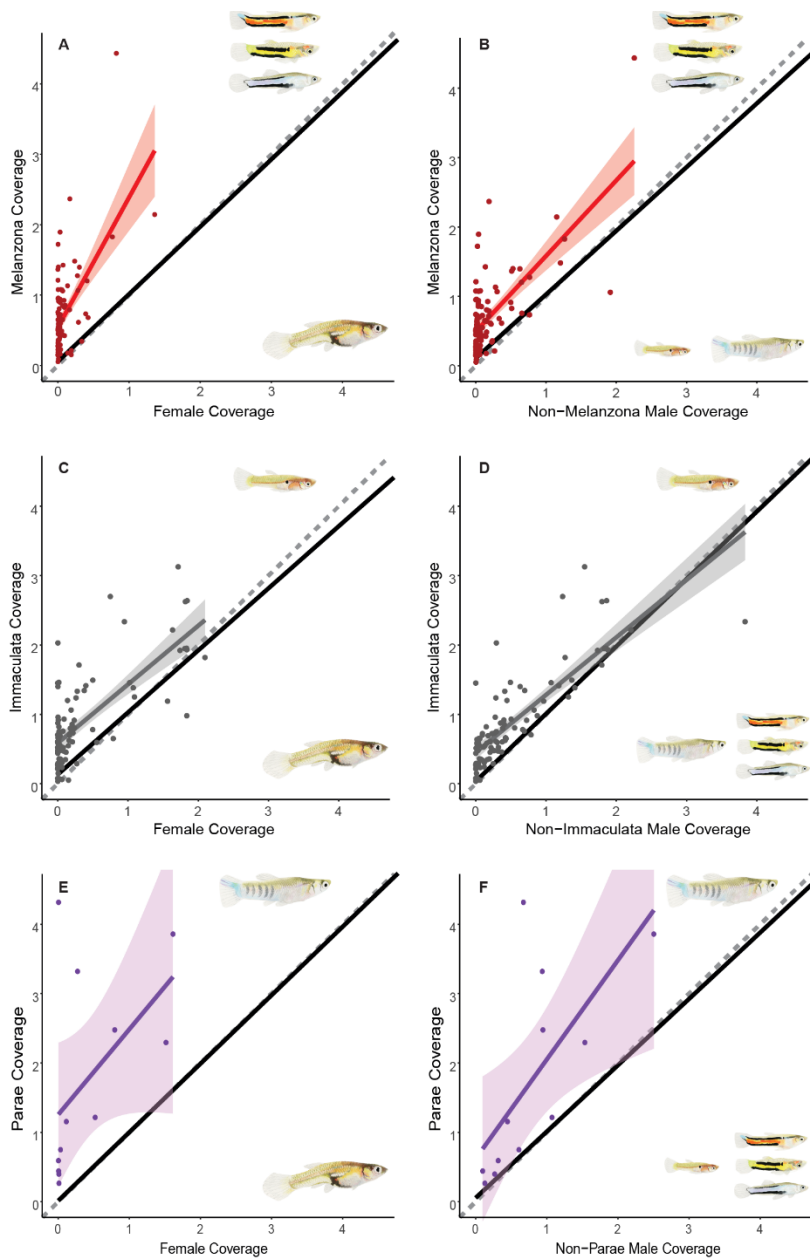
Figure 4. Relative corrected scaffold coverage of 39 individuals when aligned to melanzona (A-B),

immaculata (C-D), and parae (E-F) *de novo* genomes. Scaffolds containing morph-mers (colored) had

higher coverage by males than females (A,C,E) confirming these scaffolds contain male specific

sequence. Scaffolds containing morph-mers also had higher coverage by males of the reference morph

than males of the other morphs (B,D,F), indicating the Y chromosome sequence is substantially diverged

across morphs. In each, corrected scaffold coverage of focal morph is on the Y axis and corrected

scaffold coverage of the compared morph is on the X axis. The 1:1 line is denoted as a grey dashed line.

844   The linear regression and standard error across all scaffolds are shown as a thick black line that is nearly

845   1:1 for all morphs (note – 95% confidence interval is presented but too small to distinguish from the

846   regression line). The scaffolds containing morph-mers are shown as colored points, the linear regression

847   and standard error of morph-mer scaffolds are shown as a colored line and shaded region respectively.

Supplementary Information for

Extreme Y chromosome polymorphism corresponds to five extreme male reproductive morphs of a freshwater fish

Benjamin A Sandkam[*], Pedro Almeida, Iulia Darolti, Benjamin Furman, Wouter van der Bijl, Jake Morris, Godfrey Bourne, Felix Breden, Judith E. Mank

*Benjamin A Sandkam
Email: sandkam@zoology.ubc.ca
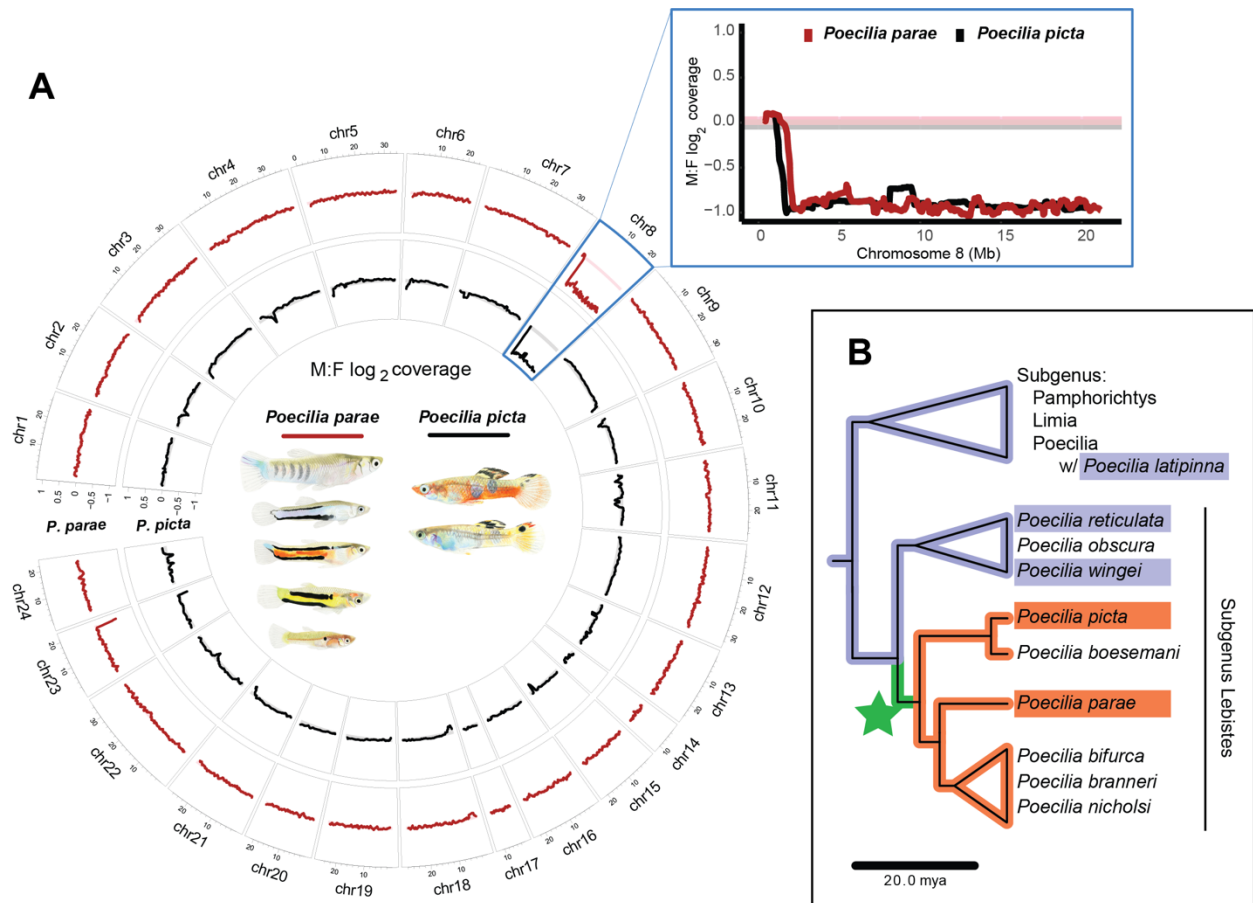
Extended Data Fig. 1 Divergence between X and Y in *Poecilia parae* and the sister species *Poecilia picta* indicate recombination was stopped before the five morphs controlled by the Y chromosome evolved in *Poecilia parae*. (A) M:F log$_2$ coverage of RACA anchored scaffolds for all five morphs of *P. parae* (red) and the close relative *P. picta* (black)[25]. Lines represent sliding window of 15 scaffolds. Shaded bars represent the 95% confidence interval based on bootstrapping coverage across the autosomes for *P. parae* (pink) and *P. picta* (grey). (B) Phylogeny from The Fish Tree of Life[29]. Orange indicate species for which chr 8 is known to be the Y and is highly diverged from the X. Blue indicates species for which chr 8 is known to be not degraded[25]. Green star denotes the branch on which X-Y recombination was arrested and the Y chromosome diverged. None of the male morphs of *P. parae* are found in other species, making the most parsimonious explanation that all five morphs arose after recombination stopped.

Individual	Morph	Ymers shared with *P. picta*
P05		51
P06		57
P09		195
P10		93
P03		129
P31		231
P04		172
P27		55
P29		140
P30		97
P17		91
P22		169
P21		96
P13		177
P14		293
P23		237
P24		221
P11		111
P20		153
P19		92
P02		80
P01		80
P12		223
P. picta		
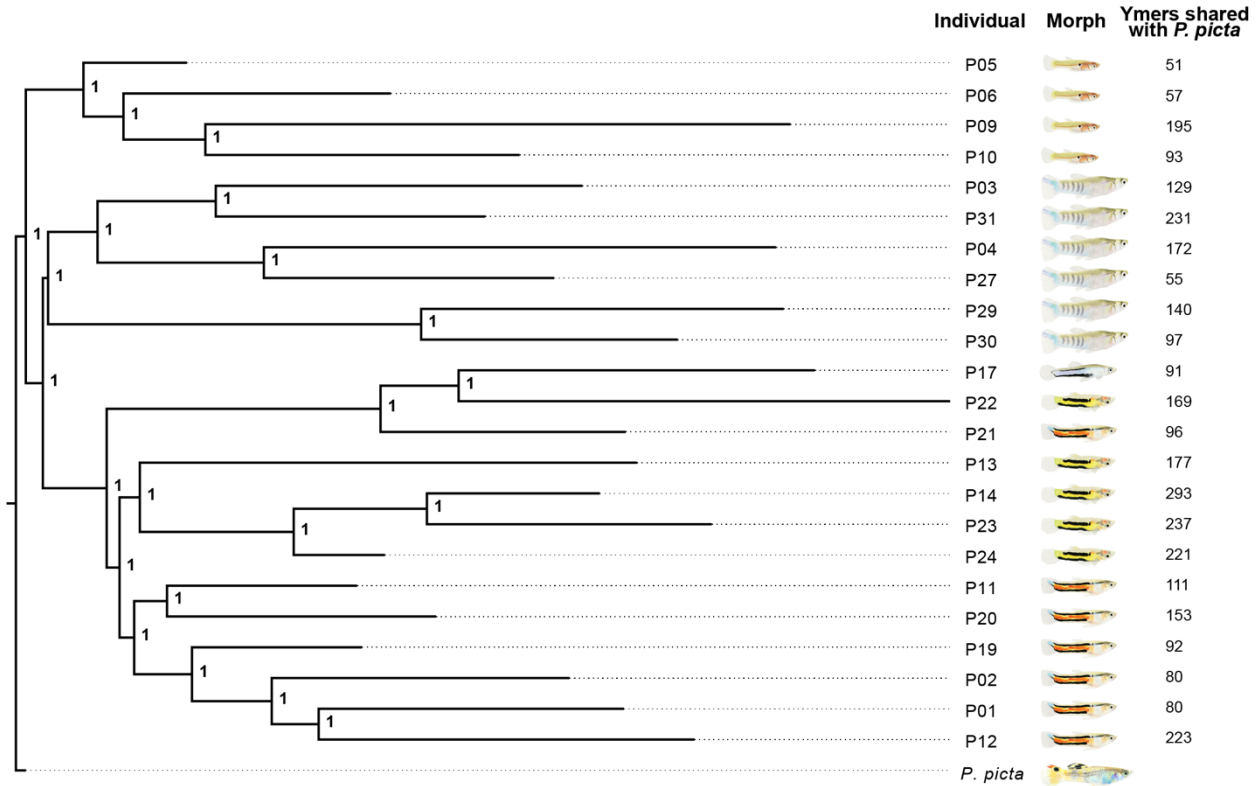
Extended Data Fig. 2 Bayesian phylogeny built on presence/absence of the 27,950,090 *P. parae* Y-mers and the 646,754 *P. picta* Y-mers in each individual and rooted on *P. picta* (as in Figure 2). The posterior probability is presented at each node. The number of Y-mers each individual shares with *P. picta* Y-mers is denoted to the right. *P. picta* Y-mers are distributed across all morphs indicating that they have been segregating on non-recombining regions of the Y chromosome since recombination was stopped in the common ancestor of *P. parae* and *P. picta*.

Validation1
4 Samples
1 Red
1 Yellow
1 Parae
1 Immac.
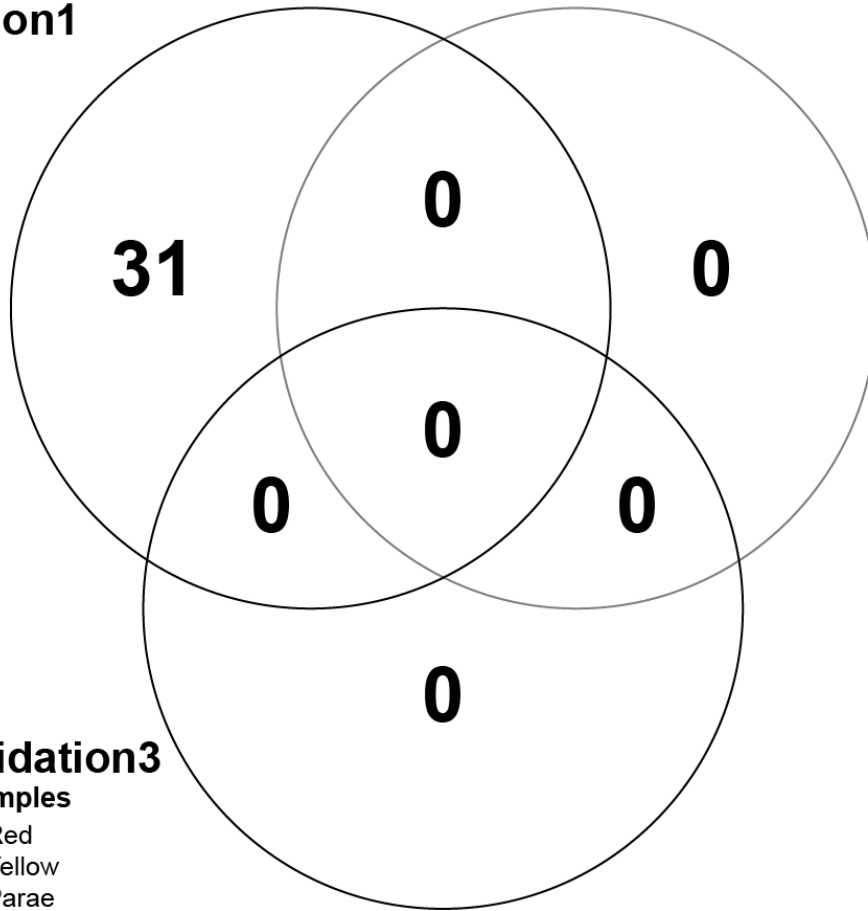
Validation2
5 Samples
1 Red
1 Yellow
1 Parae
1 Immac.
1 Blue

Validation3
9 Samples
2 Red
2 Yellow
2 Parae
2 Immac.
1 Blue

Extended Data Fig. 3 Validation of morph-mer identification pipeline using random sets of individuals from each of the different morphs. Different samples were used for each set except for blue where the 1 sample was used in validation set 2 and validation set 3. There were a low number of Y-mers unique to sets of four random individuals and zero Y-mers unique to sets with more than four individuals. This demonstrates the false positive rate of our morph-mer analysis was quite low because all major morphs had at least four individuals.
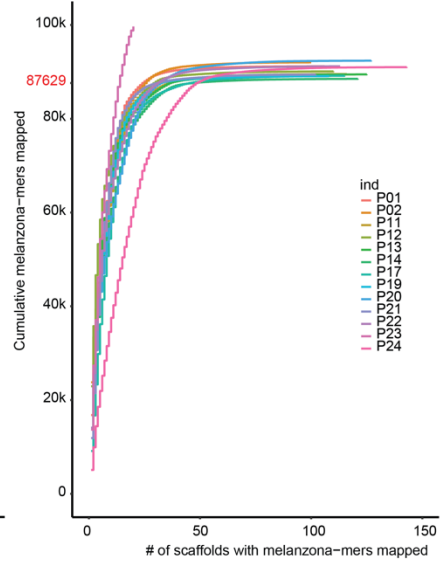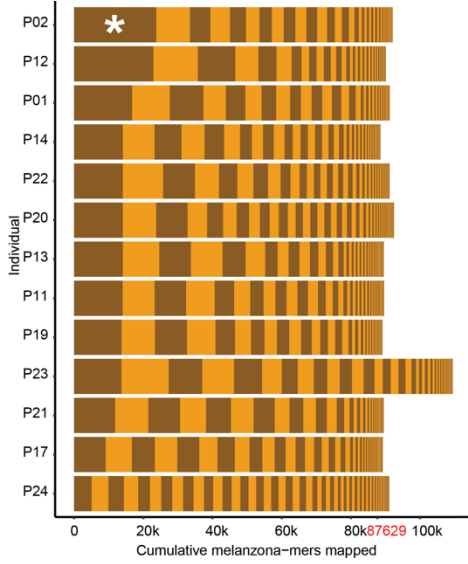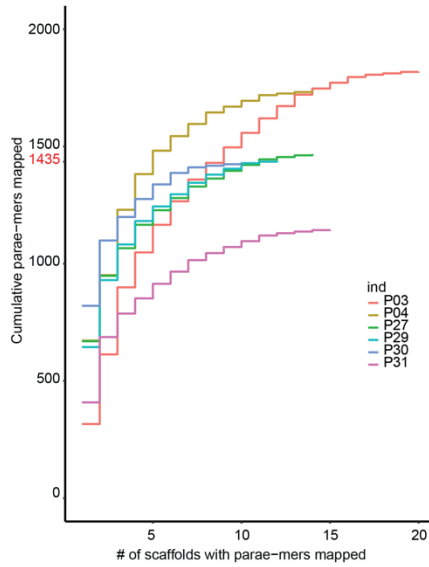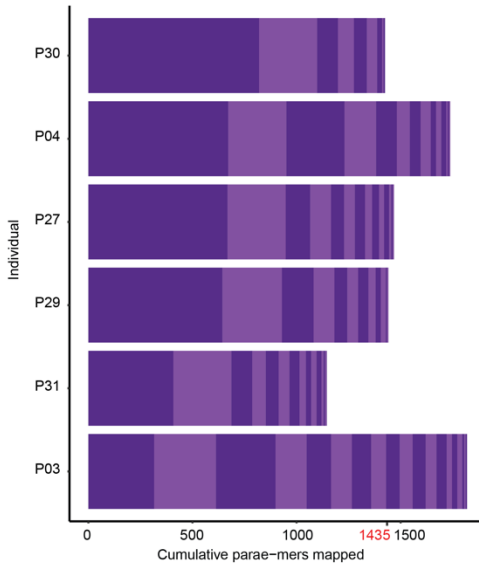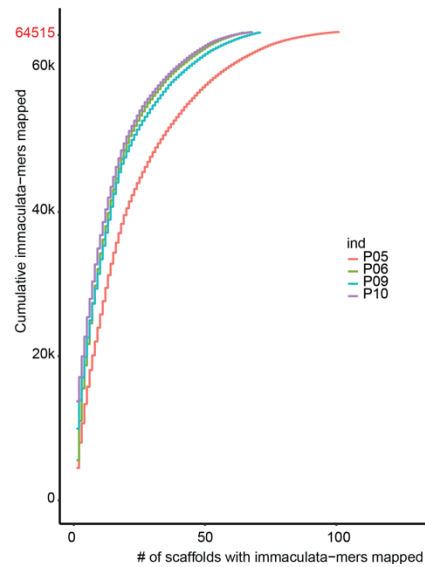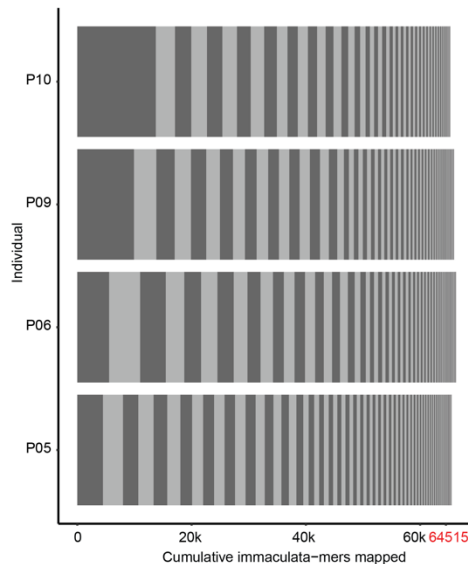
Extended Data Fig. 4 Mapping distribution for each set of morph-mers mapped to *de novo* scaffolds of males of that morph with no mismatches, gaps, or trimming. There was a low incidence of individual morph-mers mapping to more than one scaffold (0 of the 59 Y-mers were contained in more than one scaffold across all males; 0 of the 1,435 parae-mers were contained in more than one scaffold in parae males; 131 of the 87,629 melanzona-mers were contained in more than one scaffold in melanzona males; 138 of the 64,515 immaculata-mers were contained in more than one scaffold in immaculata males). Left: cumulative morph-mers mapped for each individual, each change in hue is a different scaffold. A large percentage of morph-mers generally map to just one or a few scaffolds indicating that our *k*-mer approach reveals regions of highly diverged morph specific sequence rather than single SNPs distributed throughout the genome. Right: cumulative morph-mers mapped presented as a function of the number of scaffolds. The strong deviation from 1:1 shows morph-mer mapping is non-random and further supports the morph-mers approach is identifying regions of morph specific sequence. The total number of unique morph-mers identified for that morph is indicated in red on the axis (note the variation in number of morph-mers mapped is due to some individuals having morph-mers map to multiple scaffolds). Astrix in P02 of the melanzona-mers indicates the example alignment scaffold with melanzona-mers presented in Supplementary Figure 5.
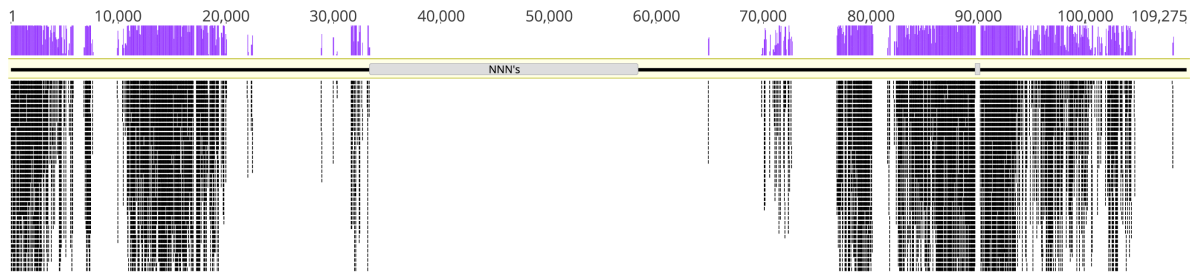
Extended Data Fig. 5 Melanzona-mers aligned to scaffold 104666 of sample P02 with no mismatches, gaps, or trimming. Each 31bp melanzona-mer is shown aligned below the reference sequence, and coverage is shown in purple above the reference sequence. Of the 87,629 unique melanzona-mers; 23,773 aligned to this scaffold. Regions of Ns are denoted on the reference genome in grey and explain a lack of melanzona-mers aligning to these regions. The strong clustering and overlapping nature of the melanzona-mers indicates sequence is highly diverged both from females and from the other morphs.

Morph Specific Y Chromosome Sequence
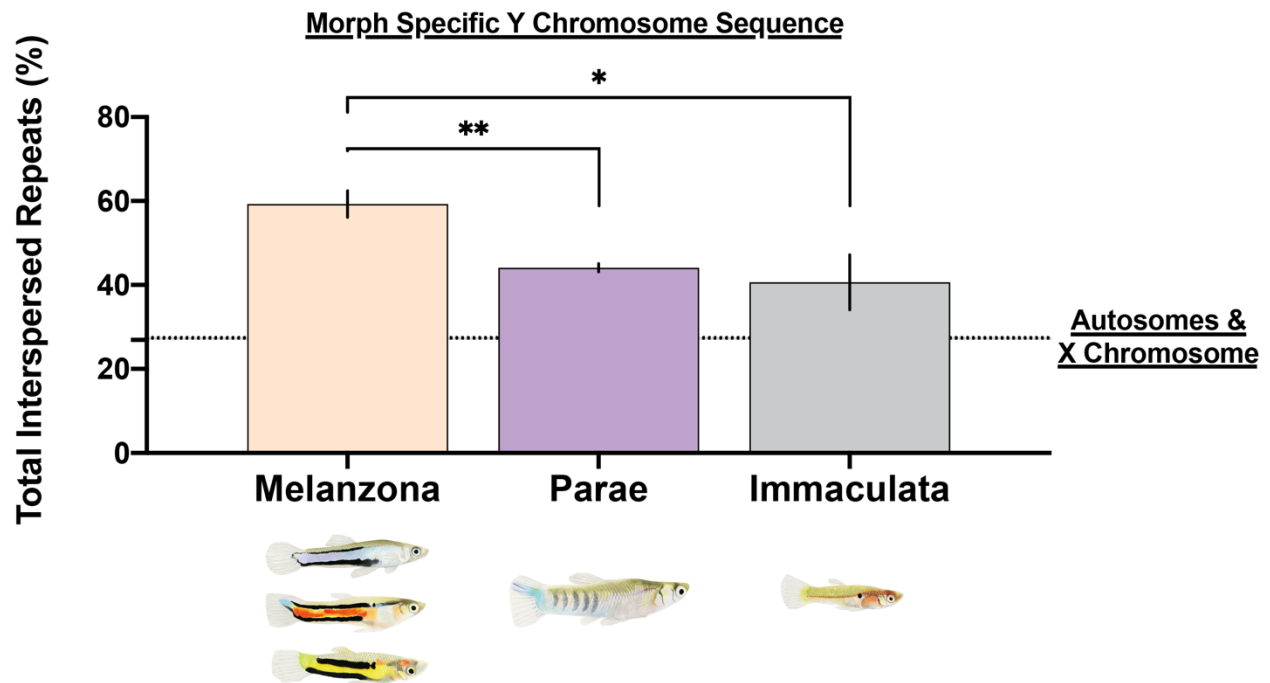
Extended Data Fig. 6 Morph specific Y chromosome sequence is composed of significantly more interspersed repeats than the autosomes and X chromosome. For males, only scaffolds containing >5 morph-specific Y-mers were evaluated, ensuring sequence is morph-specific and Y linked. To determine rates of autosomes and X chromosomes, female full *de novo* genomes were evaluated. Stars indicate significant differences between morphs (* $P<0.05$, ** $P<0.01$).

Chr 1

Chr 2

Chr 3

Chr 4

Chr 5

Chr 6

Chr 7

Chr 9

Chr 14

Chr 15

Chr 16

Chr 17
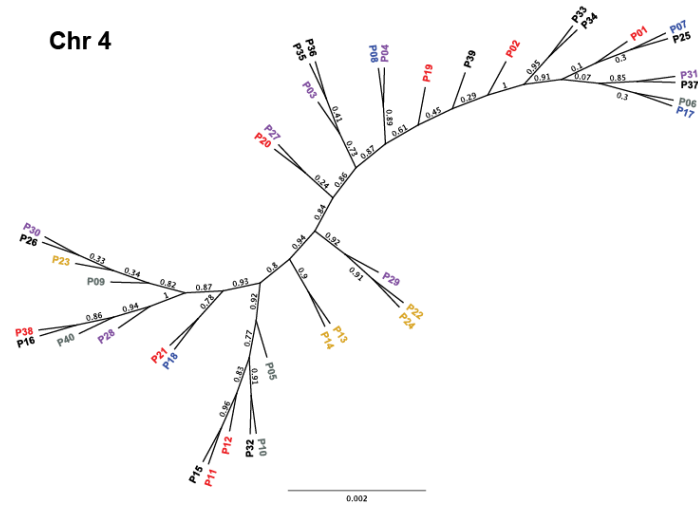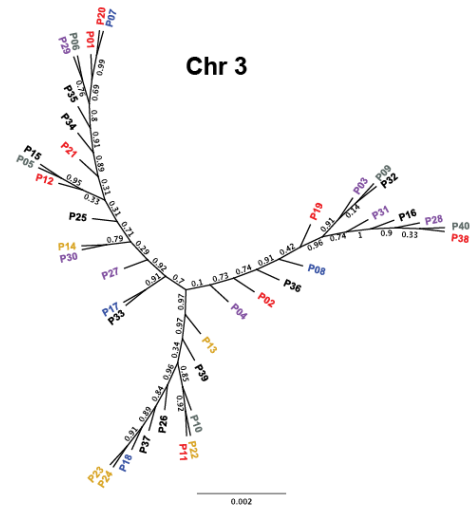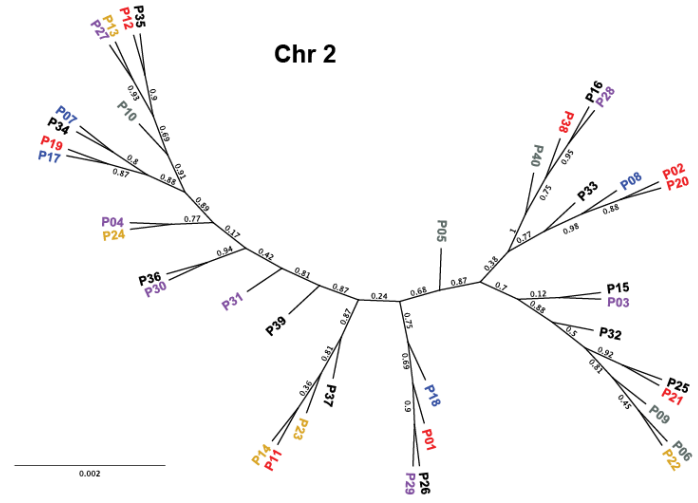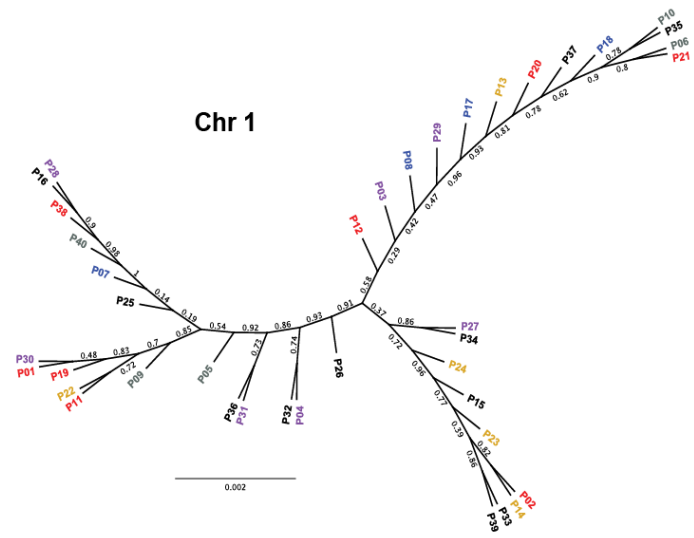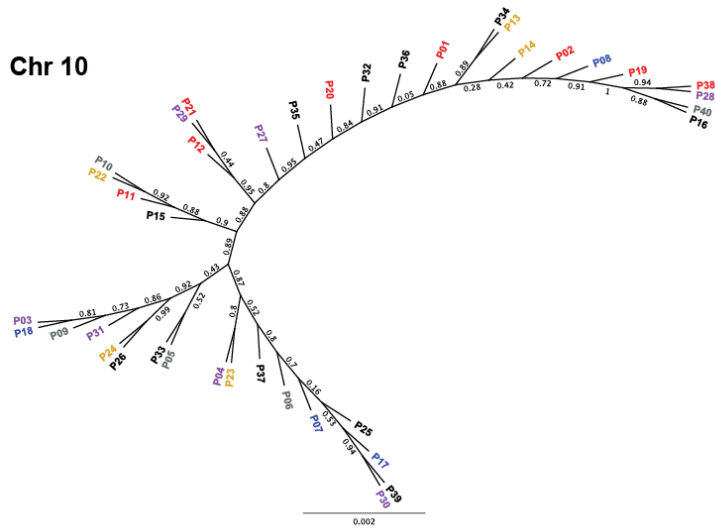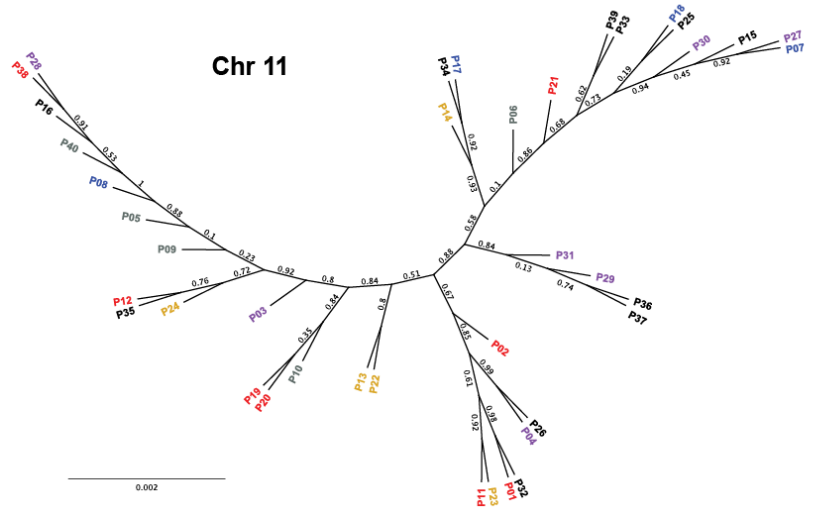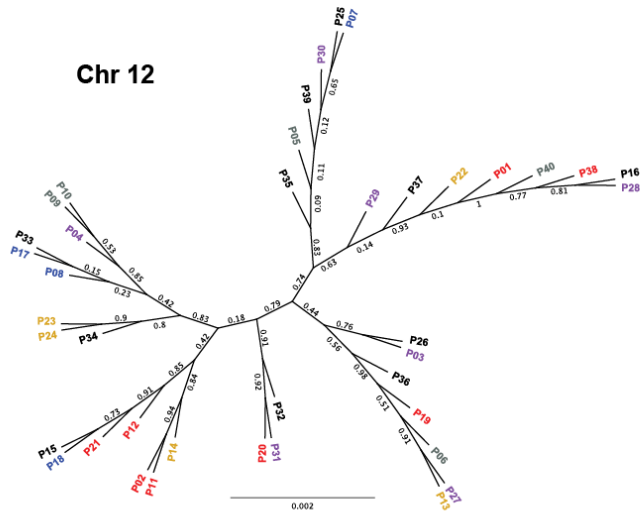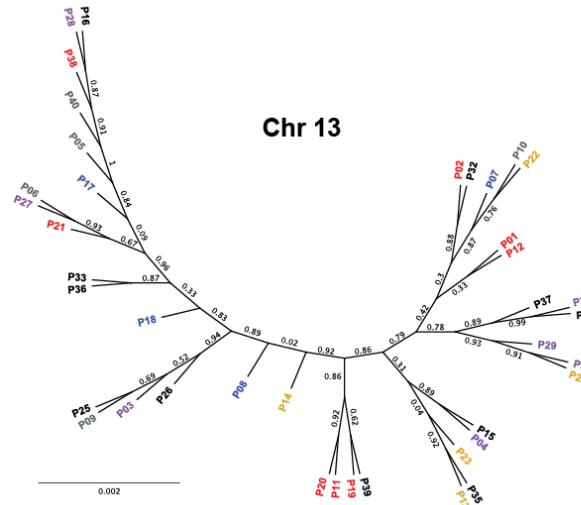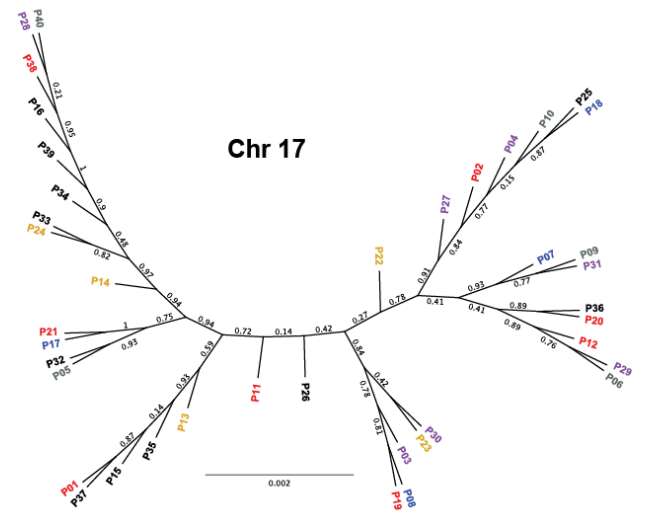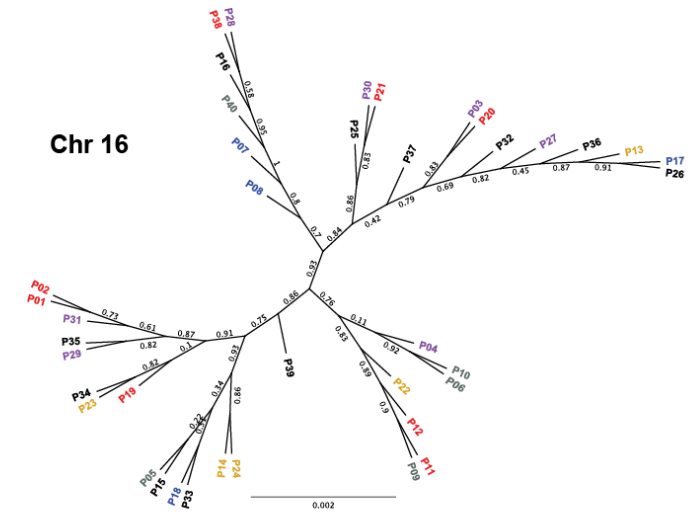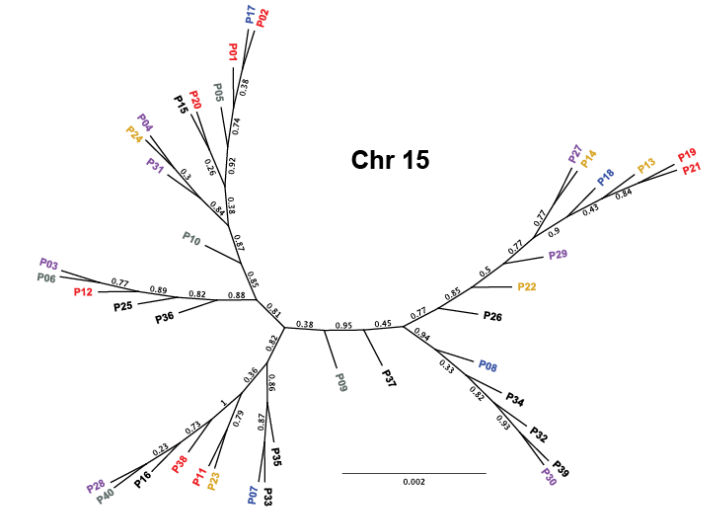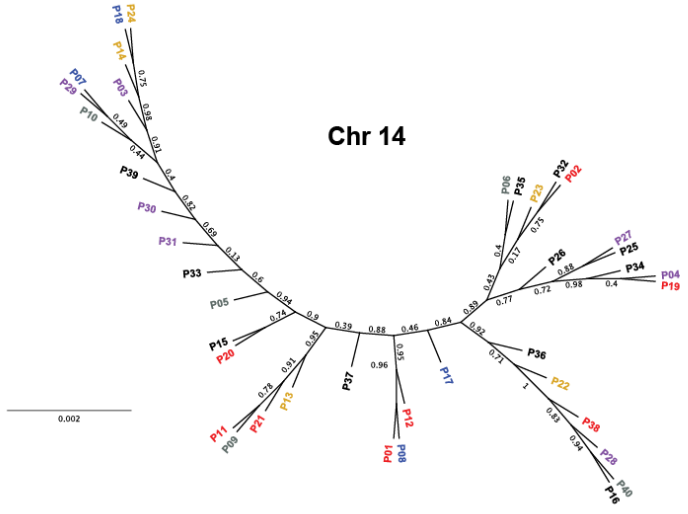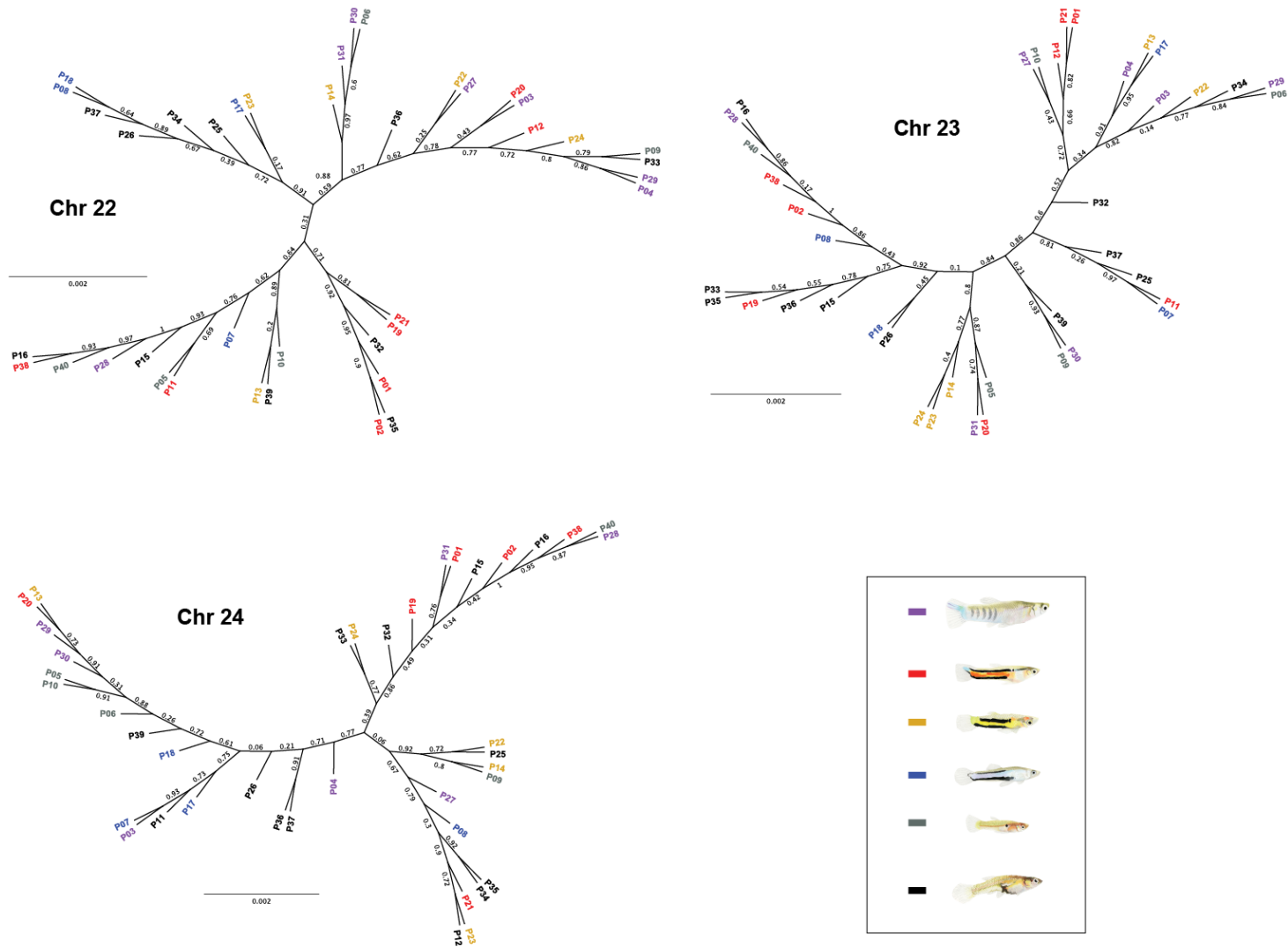
Extended Data Fig. 7 Unrooted approximately-maximum-likelihood trees (FastTree) of each of the autosomes confirm that the extreme divergence across morphs is specific to the Y chromosome, and not the result of cryptic sub-populations. Trees were built using the consensus sequence of the longest scaffold from each chromosome (as identified by RACA). Tips denote sample name, color indicates morph, and numbers on branches indicate FastTree support value.

Supplementary Table 1. Phenotypic characteristics that differ across the five male morphs of *Poecilia parae*.

	Parae	Immaculata	Melanzona Red	Melanzona Yellow	Melanzona Blue	Ref.
Physiology						
Body Size	Large	Small	Medium	Medium	Medium	1
Color Pattern	Vertical stripes	Resembles juvenile female	Horizontal stripes	Horizontal stripes	Horizontal stripes	1,2,3
Coloration	Silver/purple	None	Red	Yellow	Blue	1,2,3
Absolute Testis Mass	Medium (90% of immaculata)	Largest	Small (70% of immaculata)	Medium (90% of immaculata)	Small (70% of immaculata)	1
Testis/Body Mass	Intermediate	High	Low	Intermediate	Low	1
Sperm Midpiece	Long	Short	Long	Long	Long	1
Flagellum Length	Short	Long	Short	Intermediate	Short	1
Sperm per Ejaculate	Fewer	Most	Fewer	Fewer	Fewer	1
Sperm Length	Short	Long	Short	Long	Short	1
Behavior						
Interaction w/ other males	Highly aggressive	Non-aggressive	Aggressive	Aggressive	Non-aggressive	3
Courtship Behavior	Force copulate	Sneak copulate	Display	Display	Display	1

(1) Hurtado-Gonzales and Uy. 2009. *Anim. Behav*. **77**,1187-1194.   (2) Lindholm *et al.* 2004. *Heredity* **92**, 156-162.
(3) Hurtado-Gonzales and Uy. 2010. *BMC Evol Biol* **10**, 391.   (4) Liley. 1966. Behaviour Supplement **13**, 1-197.

Supplementary Table 2. Sequencing method (10X Chromium linked reads or direct PE Illumina) and number of paired reads used for coverage analyses of all 40 wild caught individuals.

Morph	Sample	Population	Sequencing	# Paired Reads
Females	P 15	Cemetery	10X	566,474,152
	P 26	Seawall Trench	10X	586,091,672
	P 32	Cemetery	10X	642,872,944
	P 34	Seawall Trench	10X	599,429,008
	P 36	Cemetery	10X	587,807,726
	P 37	Seawall Trench	10X	586,019,978
	P 16	Cemetery	Illumina	755,768,880
	P 25	Seawall Trench	Illumina	861,217,302
	P 33	Cemetery	Illumina	591,002,416
	P 35	Cemetery	Illumina	813,848,774
	P 39	Seawall Trench	Illumina	541,464,112
Immaculata Morph	P 05	Seawall Trench	10X	665,696,764
	P 06	Seawall Trench	10X	687,641,508
	P 09	Seawall Trench	10X	666,405,666
	P 10	Seawall Trench	10X	628,472,300
	P 40	Seawall Trench	Illumina	492,568,288
Parae Morph	P 03	Seawall Trench	10X	636,937,278
	P 04	Seawall Trench	10X	673,606,290
	P 27	Seawall Trench	10X	682,154,596
	P 29	Seawall Trench	10X	697,480,156
	P 30	Seawall Trench	10X	598,808,486
	P 31	Seawall Trench	10X	634,574,988

	P 28	Cemetery	Illumina	799,757,386
Red Melanzona Morph	P 01	Cemetery	10X	605,264,908
	P 02	Cemetery	10X	599,899,768
	P 11	Seawall Trench	10X	663,454,532
	P 12	Cemetery	10X	679,401,488
	P 19	Cemetery	10X	619,323,522
	P 20	Seawall Trench	10X	614,834,422
	P 21	Cemetery	10X	564,829,682
	P 38	Seawall Trench	Illumina	789,618,098
Yellow Melanzona Morph	P 13	Seawall Trench	10X	632,280,752
	P 14	West Watuka	10X	622,311,946
	P 22	Seawall Trench	10X	649,735,580
	P 23	West Watuka	10X	674,089,582
	P 24	West Watuka	10X	587,826,556
Blue Melanzona Morph	P 17	Cemetery	10X	705,952,894
	P 07	Seawall Trench	Illumina	771,876,214
	P 08	Cemetery	Illumina	479,310,754
	P 18	Seawall Trench	Illumina	813,513,894

Supplementary Table 3. Assembly statistics for each of the 29 genomes *de novo* assembled by Supernova. Bold samples indicate those used for follow-up coverage analyses. Genome size is expected to be ~730Mb based on close relatives with known genome sizes (*Poecilia reticulata* 731Mb, *Xiphophorus helleri* 730Mb).

Morph	Sample	Assembly Size (Mb)	Scaffold N50 (kb)	No. Scaffolds > 10kb	No. Scaffolds > 1kb	Contig N50 (kb)	Phased N50 (kb)	Coverage
Red Melanzona	**P01**	**663.83**	**3201.95**	**2081**	**20376**	**38.36**	**2252.96**	**53X**
	P02	664.43	2150.00	2493	20516	38.78	1359.88	51X
	P11	562.45	44.07	17986	70941	23.36	70.98	52X
	P12	661.61	3683.18	2114	22269	37.97	2240.87	60X
	P19	582.84	47.86	17071	67134	23.08	100.55	48X
	P20	611.40	45.28	18734	63207	23.45	94.87	42X
	P21	665.11	1679.06	2677	20797	38.28	1266.47	51X
Yellow Melanzona	P13	654.30	1212.01	3205	23821	36.88	1120.80	53X
	P14	616.95	50.94	16711	58496	24.61	115.13	49X
	P22	665.85	61.08	15566	53875	27.00	137.91	46X
	P23	655.37	97.46	12443	43298	27.22	190.30	61X
	P24	234.10	17.09	13932	117930	13.32	5.82	45X
Blue Melanzona	P17	654.75	1995.50	2568	23989	33.84	1598.03	63X
Parae Morph	P03	655.50	93.43	12361	43771	28.68	165.34	51X
	P04	**658.87**	**599.09**	**4375**	**25416**	**34.50**	**599.94**	**55X**
	P27	667.40	274.95	6754	28937	30.94	312.62	59X
	P29	599.06	38.95	20291	68801	21.92	84.43	46X
	P30	341.91	20.87	17566	108816	15.66	40.56	39X
	P31	604.74	39.45	20427	65373	21.69	91.41	44X
Immaculata	P05	71.97	14.71	4740	138919	13.03	5.44	42X
	P06	623.35	46.83	18525	62105	24.07	105.78	53X

	P09	**651.69**	**3308.41**	**2291**	**22613**	**35.38**	**2284.14**	**61X**
	P10	650.86	105.21	11446	39921	28.88	183.48	52X
Female	P15	69.89	15.53	4441	140510	13.68	3.91	44X
	P26	96.86	13.50	6848	144112	11.90	6.34	38X
	P32	**600.79**	**40.57**	**19565**	**62451**	**23.02**	**75.27**	**48X**
	P34	508.35	34.52	18503	79658	20.75	92.35	37X
	P36	329.51	23.33	15804	103659	16.40	35.28	18X
	P37	382.74	23.43	18121	95193	17.16	41.07	21X

Supplementary Table 4. Amount of sequence in scaffolds with >5 morph-mers. Bold indicates samples used for annotation.

Morph	Sample	Number of Scaffolds	Scaffold N50 (kb)	Amount of sequence
Melanzona Morph	**P01**	**77**	**20.23**	**995,370**
	P02	80	25.71	1,122,565
	P11	85	14.89	865,095
	P12	81	27.33	1,153,044
	P13	92	66.15	1,610,202
	P14	91	15.42	926,825
	P17	104	3169.34	4,250,903
	P19	82	20.11	990,514
	P20	94	18.81	877,954
	P21	80	25.52	1,070,743
	P22	95	28.44	1,164,800
	P23	98	25.18	1,283,990
	P24	113	7.69	668,430
Parae Morph	P03	19	17.37	164,531
	P04	**14**	**9.57**	**127,542**
	P27	14	8.17	84,778
	P29	12	15.89	90,235
	P30	9	24.16	96,497
	P31	14	11.35	114,805
Immaculata Morph	P05	126	5.53	521,712
	P06	100	24.92	1,035,078
	P09	**100**	**5995.38**	**9,748,162**
	P10	81	155.60	1,388,117

Supplementary Table 5. Genes annotated on the scaffolds with the Y-mers that are present in every male.

Gene Name	Copies
CRYBB1	3
CRYBA4	1
GGTA1	1
Retrovirus-related Pol polyprotein from type-2 retrotransposable element R2DM	
Retrovirus-related Pol polyprotein from type-1 retrotransposable element R2	

Supplementary Table 6. *Melanzona morph*: Genes predicted on sample P01 (red melanzona) scaffolds containing >5 melanzona-mers and >1.5-fold melanzona male: female coverage. Scaffolds with >0.025X coverage by melanzona males but <0.025X coverage by females are considered 'Male Unique'. Scaffolds with >0.025X coverage by melanzona males but <0.025X coverage by non-melanzona males are considered 'Melanzona Unique'.

Gene	P01 Scaffold	Melanzona:Female Coverage	Melanzona:Non-Melanzona Coverage
TRIM35	111589_P01	Male Unique	Melanzona Unique
Tbx3	113310_P01	Male Unique	5.08
Tbx3	113310_P01	Male Unique	5.08
Texim2	114702_P01	Male Unique	14.57
KAT7	115621_P01	Male Unique	4.09
Retrovirus-related	111503_P01	Male Unique	3.15
Texim3	111503_P01	Male Unique	3.15
Translation initiation factor IF-2 (low E)	112167_P01	15.69	0.55
Texim2	103684_P01	13.63	10.75
Unknown	103684_P01	13.63	10.75
Unknown	111891_P01	8.69	4.48
LINE-1 type transposase	115646_P01	3.78	3.52
Texim2	112537_P01	1.67	2.12
Unknown	112537_P01	1.67	2.12
Amyloid-beta A4 precursor	116260_P01	1.91	0.97

Supplementary Table 7. *Immaculata morph*: Genes predicted on sample P09 scaffolds containing >5 immaculata-mers and >1.5 fold immaculata male: female coverage. Scaffolds with >0.025X coverage by immaculata males but <0.025X coverage by females are considered 'Male Unique'. Scaffolds with >0.025X coverage by immaculata males but <0.025X coverage by non-immaculata males are considered 'Immaculata Unique'.

Gene	P09 Scaffold	Immac:Female Coverage	Immac:Non-Immac Coverage
NLRC3	124510_P09	Male Unique	Immaculata Unique
Trim39	125211_P09	Male Unique	Immaculata Unique
Ty3 retrotransposon	141715_P09	Male Unique	2.01
MSI1	143420_P09	Male Unique	11.57
Tbx3	143386_P09	Male Unique	2.40
Trypsin-2	143386_P09	Male Unique	2.40
R2DM retrovirus	154564_P09	Male Unique	3.94
Tbx3	140269_P09	Male Unique	1.62
TBX3	87093_P09	Male Unique	0.94
GPI-anchored protein 58	102953_P09	4.14	8.25
NLRC3	139939_P09	4.06	7.46
unknown	139939_P09	4.06	7.46
NLRC3	124079_P09	4.13	3.01
Cetn3	137893_P09	1.57	1.15
Retrovirus PABLB	149075_P09	4.17	2.08
Texim2	131979_P09	3.90	1.31
unknown	126759_P09	5.56	1.27
unknown	148405_P09	3.56	1.95

Supplementary Table 8. Scaffolds containing >5 of the 59 Y-mers present in all males contain 8,547 transposable elements in their 30,558,901bp.

Class/family	Copies identified by repeat masker
DNA	264
DNA/CMC-EnSpm	297
DNA/Crypton-V	10
DNA/Dada	13
DNA/Ginger-1	13
DNA/IS3EU	62
DNA/Kolobok-T2	32
DNA/MULE-MuDR	27
DNA/Maverick	99
DNA/Merlin	31
DNA/P	5
DNA/PIF-Harbinger	63
DNA/PIF-ISL2EU	63
DNA/PiggyBac	12
DNA/Sola-1	2
DNA/TcMar	8
DNA/TcMar-ISRm11	34
DNA/TcMar-Tc1	4626
DNA/TcMar-Tc2	30
DNA/TcMar-Tigger	5
DNA/Zisupton	31
DNA/hAT	17
DNA/hAT-Ac	314
DNA/hAT-Blackjack	14
DNA/hAT-Charlie	225
DNA/hAT-Tip100	75

DNA/hAT-hAT5	35
DNA/hAT-hAT6	11
LINE/Dong-R4	120
LINE/I	54
LINE/L1	77
LINE/L1-Tx1	30
LINE/L2	776
LINE/Penelope	14
LINE/R2-Hero	2
LINE/RTE-BovB	361
LINE/Rex-Babar	399
LTR/Copia	6
LTR/ERV1	37
LTR/Gypsy	113
LTR/Ngaro	65
LTR/Pao	36
RC/Helitron	22
Retroposon	4
SINE	1
SINE/tRNA-Core-RTE	1
SINE/tRNA-V	7
SINE/tRNA-V-RTE	4

Supplementary Table 9. Scaffolds containing >5 melanzona-mers contain 392 transposable elements in their 995,370bp.

Class/family	Copies identified by repeat masker
DNA	13
DNA/CMC-EnSpm	22
DNA/Dada	1
DNA/IS3EU	11
DNA/Kolobok-T2	3
DNA/Maverick	10
DNA/Merlin	3
DNA/PIF-Harbinger	1
DNA/PIF-ISL2EU	1
DNA/TcMar-Tc1	3
DNA/Zisupton	2
DNA/hAT	6
DNA/hAT-Ac	27
DNA/hAT-Charlie	28
DNA/hAT-Tip100	2
LINE/I	2
LINE/L1	24
LINE/L1-Tx1	2
LINE/L2	38
LINE/RTE-BovB	27
LINE/Rex-Babar	1
LTR/Copia	3
LTR/ERV1	5
LTR/Gypsy	43
LTR/Pao	22
RC/Helitron	90
SINE/tRNA-Core-RTE	2

Supplementary Table 10. Scaffolds containing >5 immaculata-mers contain 2,565 transposable elements in their 9,748,162bp.

Class/Family	Copies identified by repeat masker
DNA	103
DNA/CMC-EnSpm	104
DNA/Crypton-H	1
DNA/Crypton-V	7
DNA/Dada	4
DNA/Ginger-1	3
DNA/IS3EU	47
DNA/Kolobok-T2	13
DNA/MULE-MuDR	8
DNA/MULE-NOF	1
DNA/Maverick	25
DNA/Merlin	25
DNA/P	1
DNA/PIF-Harbinger	20
DNA/Sola-1	1
DNA/TcMar	4
DNA/TcMar-ISRm11	9
DNA/TcMar-Tc1	1026
DNA/TcMar-Tc2	21
DNA/TcMar-Tigger	4
DNA/Zisupton	8
DNA/hAT	12
DNA/hAT-Ac	145
DNA/hAT-Blackjack	2
DNA/hAT-Charlie	68
DNA/hAT-Tip100	39
DNA/hAT-hAT5	18
DNA/hAT-hAT6	4
DNA/hAT-hobo	1
LINE/Dong-R4	37
LINE/I	21
LINE/L1	28
LINE/L1-Tx1	7

LINE/L2	340
LINE/Penelope	3
LINE/R2-Hero	1
LINE/RTE-BovB	112
LINE/Rex-Babar	112
LTR/Copia	1
LTR/ERV1	12
LTR/Gypsy	55
LTR/Ngaro	14
LTR/Pao	9
RC/Helitron	38
Retroposon	1
SINE/tRNA-Core-L2	45
SINE/tRNA-V	4
SINE/tRNA-V-RTE	1

Supplementary Table 11. Proportion of morph-linked Y scaffolds (scaffolds containing > 5 of the respective morph-mer dataset) made up of interspersed repeats. In females proportion of total *de novo* genome made up of interspersed repeats. Interspersed repeats indicate transposable element activity. Comparing proportion of interspersed repeats between morph-linked Y scaffolds and female genomes indicates increased effect of transposable element activity on the Y chromosomes.

Morph	Sample	Sequence in morph-specific Y scaffolds (bp)	% Sequence Comprised of Total Interspersed Repeats
Red Melanzona	P01	995,370	68.08
	P02	1,122,565	66.38
	P11	865,095	65.15
	P12	1,153,044	58.75
	P19	990,514	63.80
	P20	877,954	64.38
	P21	1,070,743	60.04
Yellow Melanzona	P13	1,610,202	42.22
	P14	926,825	67.07
	P22	1,164,800	61.72
	P23	1,283,990	59.16
	P24	668,430	65.33
Blue Melanzona	P17	4,250,903	28.59
Parae Morph	P03	164,531	46.20
	P04	127,542	40.99
	P27	84,778	45.32
	P29	90,235	44.59
	P30	96,497	41.21
	P31	114,805	46.36
Immaculata	P05	521,712	53.98
	P06	1,035,078	47.84
	P09	9,748,162	24.35
	P10	1,388,117	36.48
Female (Full genome)	P15	469,458,131	28.72
	P26	524,160,639	27.55
	P32	760,885,093	26.97
	P34	728,337,754	26.88
	P36	650,681,137	26.42
	P37	660,609,987	26.98

Supplementary Table 12. Composition of interspersed repeats on the Y linked scaffolds by TE type for each morph (mean $\pm$ standard error).

	Melanzona	Immaculata	Parae
LINE	12.37% $\pm$ 0.40	12.87% $\pm$ 1.31	8.12% $\pm$ 1.51
SINE	0	0.01% $\pm$ 0.01	0
Penelope	0	0	0
DNA transposons	3.89% $\pm$ 0.80	8.07% $\pm$ 2.29	3.94% $\pm$ 1.22
LTR elements	10.81% $\pm$ 0.69	6.41% $\pm$ 1.38	10.23% $\pm$ 1.19
Unclassified	72.92% $\pm$ 0.73	72.65% $\pm$ 1.23	77.71% $\pm$ 2.14