nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-021-01452-w

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

Benjamin A. Sandkam[1]✉, Pedro Almeida[2], Iulia Darolti[1], Benjamin L. S. Furman[1], Wouter van der Bijl[1], Jake Morris[2], Godfrey R. Bourne[3], Felix Breden[4] and Judith E. Mank[1]

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 uses a different complex reproductive strategy and morphs differ dramatically in colour, body size and mating behaviour. 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 colour 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 behaviour. 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.

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 various 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 uses distinct reproductive tactics, with associated differences in body size, colour 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, using a sneaker copulation strategy. The three melanzona morphs are similar in body size and have a coloured horizontal stripe (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 yr (~150 generations)[13,14,20]. This suggests that balancing selection, probably resulting from a combination of sexual and natural selection, is acting to maintain these five morphs. Most importantly, multigeneration pedigrees show that morph phenotype is always passed perfectly from father to son[13], indicating the five *P. parae* morphs are controlled by five different Y chromosomes. This system therefore offers the potential for a unique insight into the adaptive potential of Y chromosomes and the role of these regions of the genome in male phenotypes.

Y chromosomes are formed once recombination with the X is halted[1] and the loss of recombination on the Y leads to a complex cascade of non-adaptive processes that lead to the rapid buildup of heterochromatin and loss of gene activity[21–23]. However, the process of Y degeneration is not linear[24] and although poecilid species closely related to *P. parae* share the same homologous sex chromosome as *P. reticulata*[25] (guppies), the extent of Y chromosome degeneration differs markedly across the clade. Although the Y chromosome in *P. reticulata* and *P. wingei* contains only a small area of limited degeneration[25–28], the entirety of the Y chromosome of *P. picta* is highly degenerate[25]. *P. parae* is a sister species of *P. picta*

[1]Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada. [2]Department of Genetics, Evolution and Environment, University College London, London, UK. [3]Department of Biology, University of Missouri-St. Louis, St. Louis, MO, USA. [4]Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia, Canada. ✉e-mail: sandkam@zoology.ubc.ca

(diverging ~14.8 million years ago (Ma); ref. [29]), however, *P. picta* males are markedly different from *P. parae* and do not resemble any of the five *P. parae* morphs[18,20,30–32], suggesting remarkable diversity was generated on the *P. parae* Y chromosome after recombination was halted with the X chromosome. Work on model systems has indeed shown that Y chromosomes can accumulate new genetic material[2–5], yet these differences occur over long periods of time and are only evident when comparing Ys across species. Non-model systems, such as *P. parae*, provide a unique opportunity to explore the limits and role of non-recombining regions in generating diversity.

Because *P. parae* is very difficult to breed in the laboratory, we collected tissue from natural populations in South America where all five male morphs co-occur and used linked-read sequencing on multiple females and males of all five morphs. Linked-read sequencing allowed us to overcome many of the technical difficulties that traditional short-read methods run into when working on regions rich in repeats and transposable elements, which are often associated with sex chromosomes. We first confirmed that *P. parae* shares the same sex chromosome system as its close relatives[25,26]. We went on to find that patterns of X–Y divergence are the same for all five Y chromosomes and match the X–Y divergence we observed in *P. picta*, suggesting that the morphs emerged after Y chromosome recombination was stopped in a common ancestor of *P. parae* and *P. picta*. Comparing the five Y chromosomes to each other, we find that while the Ys of the three minor morphs (red, yellow and blue melanzona) that differ only in colour are highly similar, the Ys of the three major morphs (parae, immaculata and melanzona) that differ in reproductive strategy, body size and mating behaviour are dramatically diverged from one another and carry substantial amounts of unique genetic material. Taken together, our results reveal the surprising ability of the Y chromosome to not only overcome the constraints of recombination loss but to generate extreme diversity, resulting in five discrete Y chromosomes that correspond to complex reproductive strategies.

Results

We collected 40 individual *P. parae* from natural populations in Guyana in December 2016, including eight red melanzona, four blue melanzona, five yellow melanzona, five immaculata, seven parae morph males and 11 females. Twenty-nine samples with sufficiently high-molecular-weight DNA were individually sequenced with 10x Genomics linked-reads (including all morphs and several females). We generated a de novo genome assembly for each of these samples. The remaining 11 lower-molecular-weight samples were individually sequenced with Illumina sequencing paired-end reads (see Supplementary Tables 2 and 3 for sequencing and assembly details).

The *P. parae* Y chromosome is highly diverged from the X and shared with *P. picta*. Degeneration of the Y chromosome results in reduced male coverage when mapped to a female reference genome. Therefore, the ratio of male to female mapped reads can be used to identify regions where the Y and X chromosomes differ substantially from each other[25,33–35]. To do this, we used our best female de novo *P. parae* genome, on the basis of N50 and other assembly statistics (see Supplementary Table 3). We then determined chromosomal position of the scaffolds using the reference-assisted chromosome assembly (RACA) pipeline, which combines phylogenetic and sequencing data to place scaffolds along chromosomes[36]. Next we mapped reads from all 40 samples to this female assembly and calculated male:female (M:F) coverage, first for each of the five morphs independently and then all morphs together.

As we previously found in *P. picta*[25], chromosome 8 (syntenic to *P. reticulata* chromosome 12) showed a clear signal of reduced read coverage in males (Fig. 1a,b), indicating an XY sex determination system. Y divergence is evident across nearly the entire chromosome and is largely identical to the pattern we previously observed in *P. picta* (Fig. 1c and Extended Data Fig. 1). This suggests that these species inherited a highly degenerate Y chromosome from their common ancestor, well before the origin of the *P. parae* male morphs.

Short sequences representing all the possible substrings of length *k* that are contained in a genome are referred to as *k*-mers and *k*-mer comparisons between male and female genomes have been used to identify Y chromosome sequence using Y-specific *k*-mers (Y-mers) in a wide range of organisms[37–39], including guppies[40] and *P. picta*[25]. We first compared all males, representing all five morphs, to all females. We found 27,950,090 Y-mers (of 31 base pairs (bp)) that were present in at least two males but absent from all females. However, only 59 of these Y-mers were present in all 23 males (Fig. 2). To evaluate the number of false positive Y-mers identified by our approach we used the same pipeline on females since all female *k*-mers should be present in males. We found only 251,472 *k*-mers to be present in at least two females but absent from all males (zero of these were found in all six females). This suggests that only ~0.9% of the 27,950,090 Y-mers identified are false positives.

We next used Y-mer analysis to further test whether recombination was halted on the Y in the common ancestor of *P. parae* and *P. picta*. Of the 646,745 Y-mers that we previously identified in at least one *P. picta* male and no females[25], 790 *P. picta* Y-mers matched Y-mers we identified in *P. parae*, consistent with a shared history of suppressed recombination. Additionally, these shared Y-mers were present in males of all morphs (Extended Data Fig. 2), discussed in more detail below. To evaluate whether the number of Y-mers shared between species was due to chance, we compared the female-specific *k*-mers that we previously identified in *P. picta*[25] to the female-specific *k*-mers from *P. parae* we identified above. We found that only 80 *P. picta* female-specific *k*-mers were also female-specific *k*-mers in *P. parae*. Therefore, there were significantly more Y-mers shared between *P. picta* and *P. parae* than would be expected by chance ($X^2 = 1{,}258$, d.f. = 1, $P < 0.0001$), supporting a shared history of recombination suppression. These shared Y-mers, combined with the striking similarity in M:F read mapping (Extended Data Fig. 1) provide compelling evidence that most of Y chromosome recombination suppression occurred in a common ancestor of *P. picta* and *P. parae*.

The *P. parae* Y chromosomes are highly diverged from each other. We next compared Y-mers across individuals, generating a phylogeny on the presence/absence of Y-mers in all the *P. parae* individuals with *P. picta* as an outgroup (Fig. 2). Clear clades were recovered for each of the major morphs (immaculata, parae and melanzona) while the three minor morphs of melanzona (red, yellow, blue) were very similar to one another. The monophyly of morphs further indicates that the Y-mers we recovered are not random false positives and that morphs have distinct Y chromosomes.

The phylogenic relationships of individuals (Fig. 2) closely match the relative Y-mer comparisons across morphs (Fig. 3). We found 64,515 Y-mers in every immaculata male that were not in any parae or melanzona males (that is, immaculata-mers), 87,629 melanzona-mers and 1,435 parae-mers, suggesting that the melanzona and immaculata Y chromosomes may contain more unique sequence compared to the parae Y. Moreover, we found 10,673 Y-mers in all melanzona and parae males that did not occur in any immaculata males (Fig. 2), suggesting that the parae Y shares greater sequence similarity to the melanzona Y.

We calculated our false positive rate by randomly permuting our male samples into groups regardless of morph and determining Y-mers present in all males of each group that were absent from all other males. We found no unique Y-mers in groups of five or more random males and just 31 unique Y-mers in groups of four random males, demonstrating that the false positive rate of our

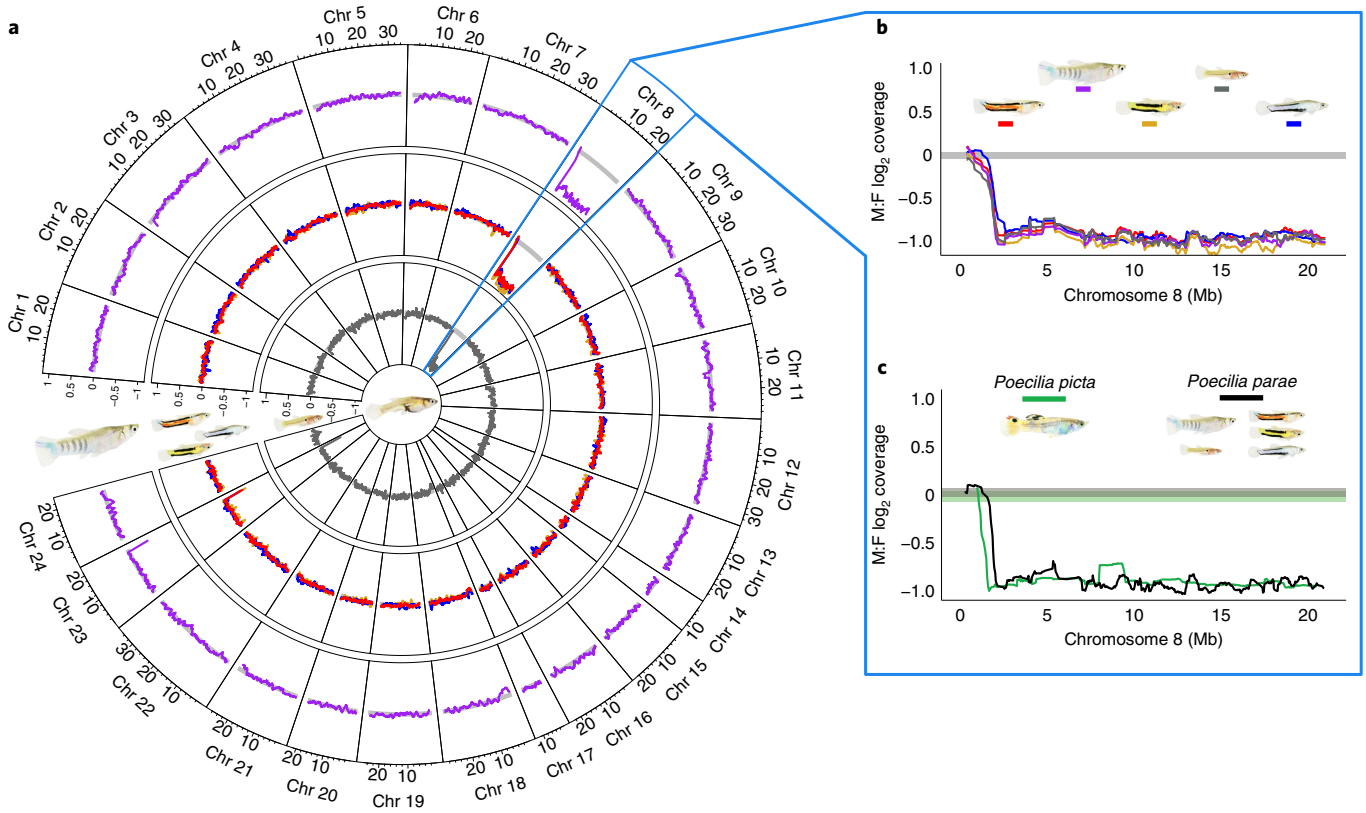


Fig. 1 | Coverage differences between the sexes (male:female $\log_2$) for female scaffolds placed by RACA on the reference *Xiphophorus hellerii* chromosomes. a, Average immaculata (inner ring), the three melanzona (middle ring) and parae morphs (outer ring) plotted across all chromosomes. Highlighted 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 that 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 that recombination was stopped in the common ancestor of *P. parae* and *P. picta* (14.8–18.5 Ma; ref. [29]). In each, horizontal grey-shaded areas represent the 95% confidence intervals based on bootstrap estimates across the autosomes. Chr, chromosome.

morph-specific Y-mer (morph-mer) approach is exceedingly low (Extended Data Fig. 3).

Mapping morph-mers confirms high diversity of Y chromosomes. The large number of morph-mers we identified could either indicate that the discrete morphs are the result of low divergence across large Y chromosomes or smaller complexes of highly diverged Y sequence. To resolve this, we mapped the morph-mers to the 21 de novo male genomes. We found that the Y chromosomes contain regions of highly diverged Y sequence, indicated by morph-mers disproportionately mapping to a few scaffolds and not being evenly dispersed. For example, a single melanzona scaffold (~110 kilobases (kb)) containing 27% of all melanzona-mers (23,773) and most morph-mers overlapped one another (Extended Data Fig. 4 and 5). This confirms that our morph-mer approach identified complexes of highly diverged Y sequence.

To compare the relative size of these diverged complexes across morphs, we identified all scaffolds that contained more than five morph-mers in each individual. The average amount of sequence contained within these morph-mer scaffolds was 1.3 megabases (Mb) for melanzona, 3.2 Mb for immaculata and just 0.1 Mb for parae individuals (Supplementary Table 4). This complements the relative number of Y-mers we found for each morph and together suggests that the amount of unique Y chromosome sequence differs across morphs, with the parae morph Y containing the least unique genetic material.

Read mapping confirms extensive divergence of Y chromosomes. To determine how divergent the five Y chromosomes are from one another, we mapped reads from all 39 samples to the best full de novo genome assembly of each male morph (195 total alignments). Most scaffolds contain autosomal sequence and coverage is not expected to differ by sex or morph. Meanwhile, scaffolds containing morph-specific sequence will have higher coverage by males of the same morph (for example, immaculata reads mapped to an immaculata assembly) than coverage by males of a different morph (for example, melanzona reads mapped to an immaculata assembly). Low female coverage of such scaffolds confirms that these regions are on the Y and are substantially diverged from the X.

As expected, when comparing coverage between males of the same morph as the reference assembly and the other morphs, we found average coverage was 1:1 when considering all scaffolds, yet scaffolds enriched for morph-mers (containing more than five) had much higher coverage by males of the same morph as the reference (Fig. 4). Surprisingly, we found no coverage by immaculata or parae reads for nearly half (40/100) of the scaffolds enriched for melanzona-mers. Similarly, 14 of the 93 scaffolds enriched for immaculata-mers (containing ~48 kb) had no coverage when we mapped melanzona and parae reads. Meanwhile, in agreement with our morph-mer analysis, all 12 of the scaffolds enriched for parae-mers had nearly equal coverage by melanzona and immaculata reads, once again suggesting that the parae Y contains very little unique Y sequence.

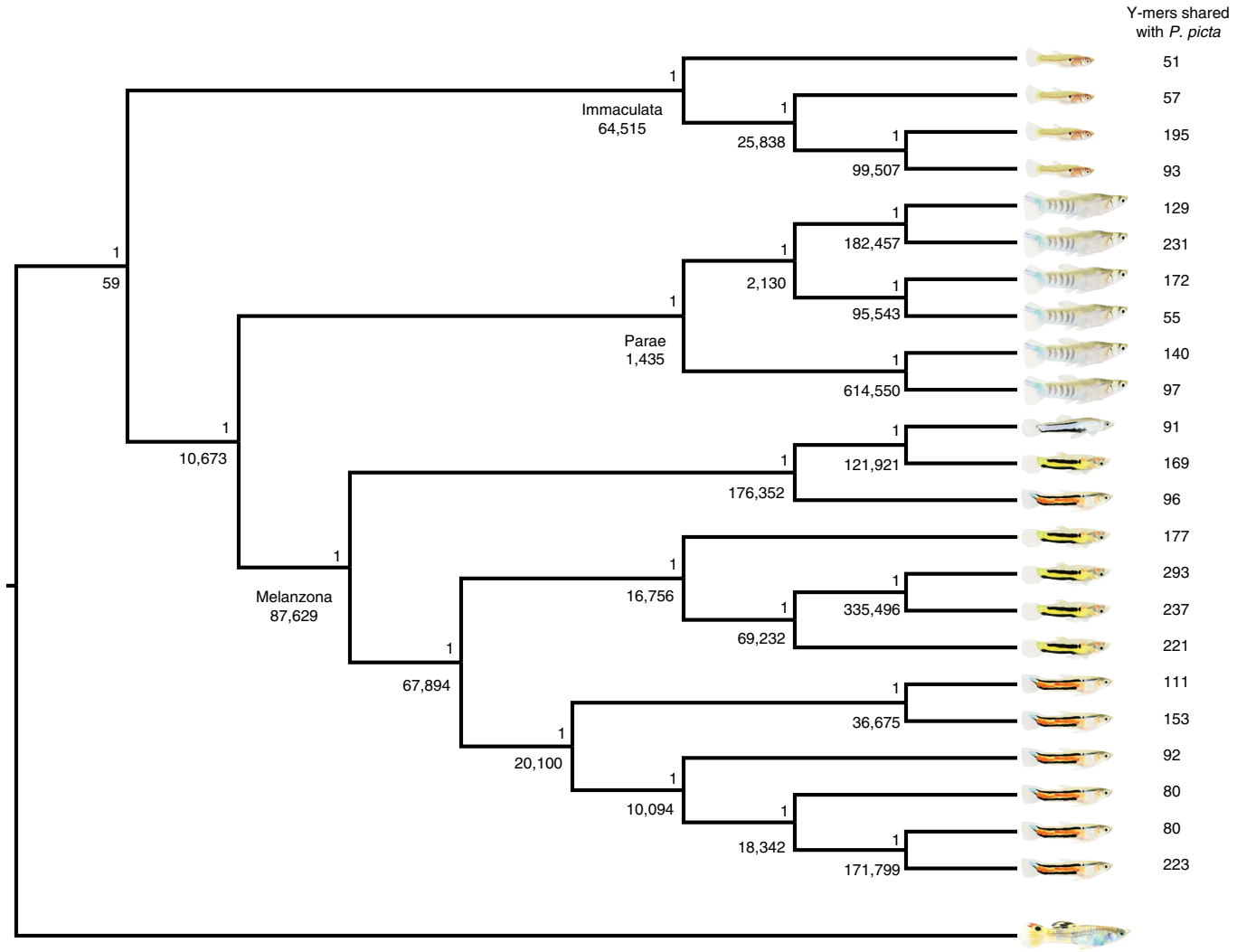


Fig. 2 | Bayesian Y chromosome phylogeny based on presence/absence of the 27,950,090 *P. parae* Y-mers and 1,646 *P. picta* Y-mers[25] in each individual and rooted on *P. picta*. The posterior probability is presented above each node; below the node is the number of *P. parae* Y-mers unique to all members 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.

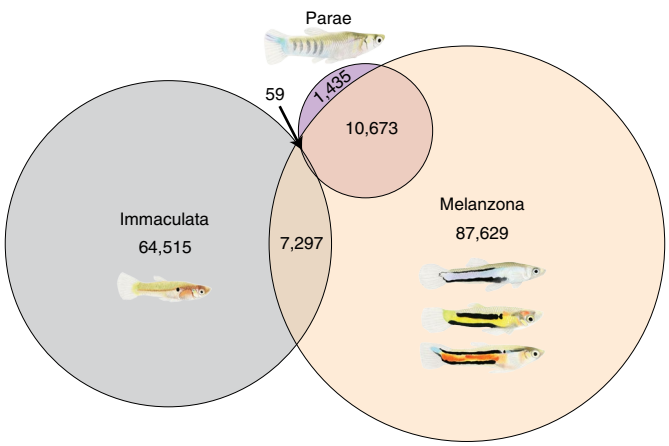


Fig. 3 | The distribution of the 27,950,090 *P. parae* Y-mers reveals strong differences across morphs. While there are very few Y-mers present in all morphs, each morph harbours unique Y-mers. The melanzona and parae morphs share more Y-mers with one another than either share with immaculata.

We also found the ratio of M:F coverage was much higher for morph-mer scaffolds, many of which had no female coverage, again confirming that these complexes of morph-specific sequence are located in non-recombining regions of the Y chromosome (Fig. 4).

Gene annotation of morph-mer scaffolds. We identified genes on scaffolds with more than five morph-mers. In total, we found seven genes on the scaffolds containing the 59 Y-mers present in all morphs (totalling 30,558,901 bp), 291 genes on the immaculata scaffolds of sample P09 (totalling 9,748,162 bp), 15 genes on the melanzona scaffolds of sample P01 (totalling 295,057 bp) and no genes on the parae morph scaffolds of P04 (totalling 127,542 bp) (Supplementary Tables 5, 6 and 7).

Only one gene was predicted on scaffolds that were completely unique to melanzona (*trim35*) and only two genes were predicted on scaffolds completely unique to immaculata (*trim39* and *nlrc3*). Members of the *Trim* gene family act throughout the body and are well known to rapidly evolve new functions[41,42]. Meanwhile, *nlrc3* has been shown to selectively block cellular proliferation and protein synthesis by inhibiting the mTOR signalling pathway[43], which could play a role in keeping immaculata the smallest morph. Several

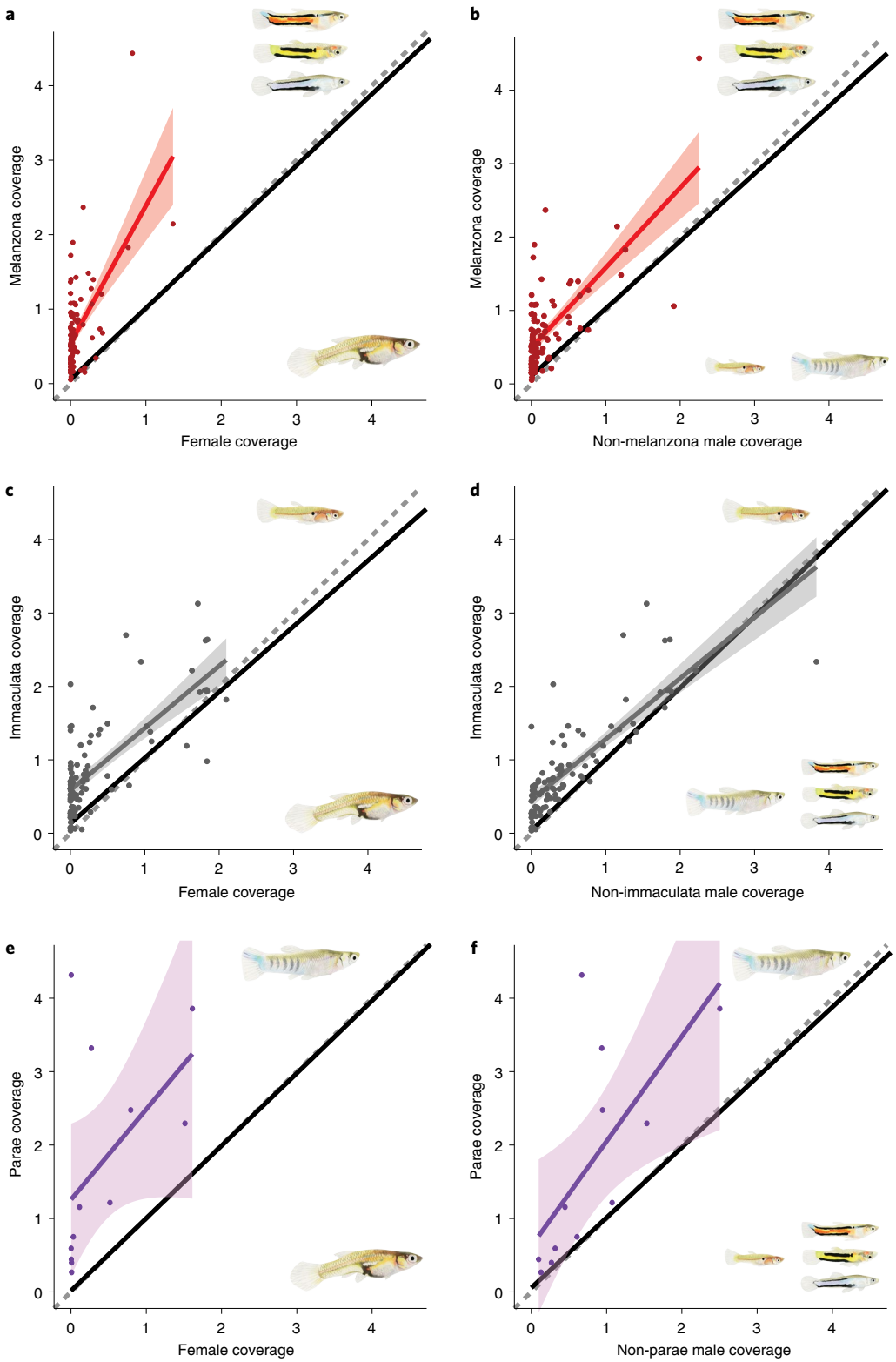


Fig. 4 | Comparison of mapping coverage for scaffolds enriched with melanzona, immaculata and parae morph-mers. a–f, Relative corrected scaffold coverage of 39 individuals when aligned to melanzona (**a**,**b**), immaculata (**c**,**d**) and parae (**e**,**f**) de novo genomes. Scaffolds containing morph-mers (coloured) had higher coverage by males than females (**a**,**c**,**e**) confirming that these scaffolds contain male-specific sequence. Scaffolds containing morph-mers also had higher coverage by males of the reference morph than did males of the other morphs (**b**,**d**,**f**), indicating that 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. The linear regression and standard error across all scaffolds are shown as a thick black line that is nearly 1:1 for all morphs (note: 95% confidence interval is presented but too small to distinguish from the regression line). The scaffolds containing morph-mers are shown as coloured points; the linear regression and standard error of morph-mer scaffolds are shown as a coloured line and shaded region respectively.

copies of the transcription factor *Tbx3* were present on male unique scaffolds in both melanzona and immaculata morphs. *Tbx* genes play key roles in development and act as developmental switches[44–46], raising the possibility that it could play a role in orchestrating the multi-tissue traits that differ across morphs.

We also found several copies of *texim* genes on male scaffolds of melanzona and immaculata that most closely match *texim2* and *texim3*. While the function of *texim2* and *texim3* are largely unknown they have been shown to be highly expressed in the brain and testis of closely related species[47]. In *Xiphophorus maculatus*, a close relative to *P. parae* (~45 Ma; ref. [29]), the transposable element *helitron* has moved *texim1* to the sex-determining region of the Y chromosome and duplicated it, resulting in three copies of *texim* that are expressed specifically in late-stage spermatogenesis[47]. The copies of *texim2* and *texim3* that we identified on the Y chromosome of *P. parae* are not the same as those in *X. maculatus*, as the Ys arose independently and the *X. maculatus* Y is not chromosome 8. Future analyses are needed to determine the roles of these and other genes in generating the morph-specific phenotypes.

Interspersed repeats and morph-mer scaffolds. We next assessed the accumulation of interspersed elements on the Y chromosomes in our samples. Repeatmasker revealed many transposable elements on scaffolds enriched for the Y-mers (present in all males) and morph-mer-enriched scaffolds (Supplementary Tables 8–12). These include 90 copies of the *helitron* transposable element on scaffolds enriched for melanzona-mers and 38 copies on scaffolds enriched for immaculata-mers.

Interspersed repeats indicate the current and/or previous presence of transposable elements in a region and can provide a measure of transposable element activity[48]. We found transposable element activity on autosomes and the X chromosome to be in line with other vertebrate genomes[49], as indicated by interspersed repeats comprising an average of 27.25% of the de novo female genomes (Supplementary Table 11). However, the male scaffolds enriched for morph-mers were composed of substantially higher proportions of interspersed repeats (melanzona $P<0.0001$, $t=6.798$, d.f. = 17; parae morph $P<0.0001$, $t=16.20$, d.f. = 10; immaculata $P=0.0322$, $t=2.587$, d.f. = 8) (Extended Data Fig. 6). The proportion of Y sequence comprised of interspersed repeats also differed by morph, with melanzona (59.28%) being higher than both parae morph (44.11%) ($P=0.0053$, $t=3.193$, d.f. = 17) and immaculata morph (40.66%) ($P=0.0140$, $t=2.780$, d.f. = 15) (Supplementary Table 11 and Extended Data Fig. 6).

Discussion

Recombination is widely regarded as one of the most important processes generating phenotypic diversity as it produces new allelic combinations on which selection can act[50]. The loss of recombination is classically assumed to result in reduced genetic diversity through sweeps and background selection[1,23,51–58] and have only limited potential for adaptive evolution. The power of these processes to deplete non-recombining regions of diversity is clearly evident in the Y chromosomes of many species[1].

Our results, based on both similarity of Y degeneration (Extended Data Fig. 1) and shared Y-mers, together with the recent reports of complete dosage compensation in both *P. parae*[59] and *P. picta*[25], are consistent with recombination suppression between the X and Y chromosome in the common ancestor of *P. parae* and *P. picta* (14.8–18.5 Ma; ref. [29]) and a highly degenerate Y present at the origin of *P. parae*. Importantly, none of the characteristics that differentiate the immaculata, parae or three melanzona morphs of *P. parae* is found in any close relatives. Therefore, although it is possible that the morphs evolved in an ancestor and were subsequently lost in all lineages except *P. parae*, the more parsimonious explanation is that the genetic basis of the extreme diversity in morphs evolved on a non-recombining, highly degenerate *P. parae* Y chromosome (Extended Data Fig. 1). Given that these morphs differ in a suite of complex traits, including body size, testis size, colour pattern and mating strategy, *P. parae* morph diversity is probably underpinned by either polygenic genetic architectures directly on the Y or new genetic elements on the Y that regulate larger regions of the genome, as observed in *Drosophila*[60]. Consistent with complex phenotypic differences between morphs, we identified substantial morph-specific genetic material (Figs. 2–4) that was also absent in females and therefore Y-linked.

Although male-specific regions of the genome may experience elevated mutation rates[61], it is also likely that the high *P. parae* Y diversity was generated through translocations and/or accumulation of interspersed repeats (for example, transposable element (TE) movement). Translocations have been shown to increase Y chromosome content[2,62]. In contrast, although TE movement has historically been considered to be a deleterious process, more recent reports have revealed that TEs can alter regions by removing or adding regulatory or coding sequence[44,63–66] and TEs may even act as substrate for new genes[67]. As predicted for non-recombining regions, we found a large number of TEs in the five *P. parae* Y chromosomes. In all morphs, we found that substantially more of the Y chromosomes were composed of interspersed repeats compared to the autosomes and X chromosome. Meanwhile the percentage of Y sequence composed of interspersed repeats differed between the morphs, further suggesting that TE activity may have played an important role in generating the diversity and divergence of these five non-recombining Y chromosomes.

Making and maintaining five morphs. We found substantial morph-specific genetic diversity on the Y chromosome of *P. parae*. Intriguingly, that diversity is maintained within populations, as evidenced by the stability in morph frequencies over repeated surveys spanning 50 yr or roughly 150 generations[13,14,20]. Even if alternative morphs have exactly equal fitness, populations are expected to eventually fix for one morph due to drift[68]. Maintaining alternative morphs within the same population relies on negative frequency-dependent selection, thus as one morph decreases in frequency its fitness increases, such as with the three male morphs of the side-blotched lizard[69].

Previous work suggests that *P. parae* morphs are also under negative frequency-dependent selection[14–17] and this could facilitate the establishment and maintenance of five distinct Y chromosomes within the same species. Most new mutations are expected to be lost through drift if they do not confer a high enough fitness advantage over alternative alleles[70] but mutations resulting in a new morph would be at the lowest frequencies and thus have the highest fitness, allowing them to rapidly stabilize in the population. Alternatively, it is possible that the morphs arose in separate populations and only later came into sympatry.

Genetic basis of male reproductive morphs. Autosomal non-recombining regions have been shown to be associated with alternative reproductive strategies in a range of species[7,9,71–75], yet the formation of the non-recombining regions underlying alternative reproductive strategies in *P. parae* differs from those previously described. For example, male morphs of the white-throated sparrow, which differ in pigmentation and social behaviour, are the result of a hybridization event which instantly brought together alternative sequence and halted recombination[72]. Importantly, because none of the characteristics of the *P. parae* morphs are found in any close relatives, it is unlikely that hybridization is the source of the Y chromosomes we describe. The alternative male morphs in the ruff are also controlled by an autosomal supergene that is composed of two alternative versions of an inverted region[10]. It has yet to be determined whether the diversity across these ruff supergenes pre-dates

the inversion or arose after recombination stopped as it did in *P. parae*. A large inverted region is also associated with social morphs in many ant species; this region likely formed in a common ancestor and has been maintained by balancing selection through repeated speciation events[12]. Although it is possible that the multiple male morphs in *P. parae* arose in an ancestor, this is less likely as they have not been observed in any related species to date (including four species that diverged after Y recombination was stopped). The new formation of the non-recombining regions associated with *P. parae* alternative reproductive strategies make this a powerful system for future work to explore the genetic basis of male reproductive morphs.

Conclusion

The role of recombination in shaping co-adapted allele complexes has long been an enigma, given that recombination is a key mechanism in generating diverse allelic combinations, yet recombination also acts to break up such combinations. Our results suggest that substantial diversity can be generated without the power of recombination and the Y chromosome retains remarkable adaptive potential with regard to male phenotypic evolution. Our work indicates that the five Y-linked male morphs of *P. parae* emerged and diverged after recombination was halted, resulting in five unique Y chromosomes within one species. Future work identifying the mechanisms by which morphs are determined by these five Y chromosomes will provide much-needed insight to determining which evolutionary forces have led to and shaped these amazing complexes and their co-evolution with the rest of the genome, which is shared across all morphs.

Methods

Field collections and DNA isolation. To ensure we accounted for natural diversity in the five Y chromosomes of *P. parae*, and because this species is extremely difficult to breed in captivity, we collected all samples ($n = 40$) from three large native populations around Georgetown, Guyana in 2016 (see Supplementary Table 2 and Sandkam et al.[18] for description of populations) (Environmental Protection Agency of Guyana Permit 120616 SP: 015). Individuals were rapidly killed in MS-222, whole-tail tissue was dissected into ethanol and immediately placed in liquid nitrogen to maintain integrity of high-molecular-weight DNA. Tissue samples were brought back to the laboratory and kept at −80 °C until high-molecular-weight DNA extraction. All samples were collected in accordance with national and institutional ethical guidelines (Canadian Council on Animal Care, University of British Columbia).

High-molecular-weight DNA was extracted from 25 mg of tail tissue of each sample following a modified protocol from 10x Genomics described in Almeida et al.[28]. Briefly, nuclei were isolated by gently homogenizing tissue with a pestle in cold nuclei isolation buffer from a Nuclei PURE Prep Kit (Sigma). Nuclei were pelleted and supernatant removed before being digested by incubating in 70 µl of PBS, 10 µl of proteinase K (Qiagen) and 70 µl of digestion buffer (20 mM EDTA, 2 mM Tris-HCl, 10 mM *N*-laurylsarcosine) for 2 h at room temperature on a tube rotator. Tween 20 was added (0.1% final concentration) and DNA was bound to SPRIselect magnetic beads (Beckman Coulter) for 20 min. Beads were bound to a magnetic rack and washed twice with 70% ethanol before eluting DNA. Samples were visually screened for integrity of high-molecular-weight DNA on an agarose gel. Of the 40 samples extracted, 29 passed initial screening and were used for individual 10x chromium linked-read sequencing (six female, seven red, five yellow, one blue, four immaculata and six parae morph). The remaining 11 samples (five female, one red, three blue, one immaculata and one parae morph) were individually sequenced on an Illumina HiSeqX as 2 × 150 bp reads using the v.2.5 sequencing chemistry with 300-bp inserts and trimmed with trimmomatic (v.0.36)[76]. To ensure high coverage of the Y, all 40 samples were individually sequenced to a predicted coverage of 40X (see Supplementary Table 2 for number of reads after filtering), which would result in predicted 20X coverage of the haploid Y.

10x chromium linked-read sequencing and assembly. The 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 (v.0.36)[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 v.2.1.1 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 RACA pipeline[36]. Briefly, scaffolds were aligned with LASTZ v.1.04 (ref. [78]) against high-quality chromosome level genome assemblies of a close relative (*X. helleri* v.4.0; GenBank accession GCA_003331165) and an outgroup (*Oryzias latipes* v.1; 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, 150-bp reads from the five females sequenced with paired-end Illumina (300-bp inserts) were aligned to the target assembly with Bowtie2 v.2.2.9 (ref. [80]) reporting concordant mappings only (--no-discordant option). For long-insert data, synthetic 150-bp 'pseudo-mate-pair' reads were generated from the de novo scaffolds of the six 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 2.3-kb insert library and a 15-kb 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.

To identify which chromosome is the sex chromosome and determine the extent of X–Y divergence, we mapped reads from all 40 individuals to the female scaffolds that had RACA-generated chromosome annotations using the aln function of bwa (v.0.6.1)[81]. Alignments were filtered for uniquely mapped reads and average scaffold coverage was calculated using soap.coverage v.2.7.7 (ref. [82]). To account for differences across individuals in sequencing library size, we divided the coverage of each scaffold by the average coverage across all scaffolds for each individual. M:F-fold change in coverage of each scaffold was calculated for all males and each of the five morphs as $\log_2$(average male coverage) – $\log_2$(average female coverage)[83]. On observing chromosome 8 was the sex chromosome, we calculated the 95% CI for M:F coverage by bootstrapping across all scaffolds which RACA placed on the autosomes (1,000 replicates; mean of 20 scaffolds without replacement).

Identifying morph-specific sequence by *k*-mers. To locate morph-specific sequence, we identified morph-specific *k*-mers (morph-mers) and then mapped these to the respective de novo genome assemblies. First Jellyfish v.2.2.3 (ref. [81]) was used to identify all 31-bp *k*-mers from the 'megabubble' output of each of the 22 male de novo genome assemblies from supernova. We next identified the putative Y-linked *k*-mers in each sample by removing all *k*-mers present in any of the females. For female *k*-mer identification we conservatively took *k*-mers from both the megabubble outputs and the raw Illumina reads of all 11 female individuals, which we used to identify every 31-bp *k*-mer present more than three times (this maximized the chance of including *k*-mers from unassembled regions of female genomes but minimized *k*-mers from sequencing errors[40]). The *k*-mers present in females all occur either on autosomes or the X chromosome; therefore, by removing the female *k*-mers from all *k*-mers identified in males we are left with putative Y-linked *k*-mers that we call Y-mers[25,40]. To exclude *k*-mers representing autosomal single nucleotide polymorphisms (SNPs) unique to a single male, we required Y-mers be present in at least two individuals. Since all female sequence is theoretically present in males, we validated our method by identifying all *k*-mers from the six female megabubbles that were not present in the male Illumina reads and found how many were present in at least two female individuals.

We then combined all the Y-mers we found with those we previously identified in *P. picta*[25] and used the presence of Y-mers as character states to build a phylogeny of all individuals and *P. picta*. Two runs of MrBayes v.3.2.2 (ref. [84]) were run for 100,000 generations with Y-mers treated akin to restriction sites (model F81 with rates set to equal and default priors). The s.d. of the split frequencies between runs reached 0 indicating that both runs converged on identical and robust trees.

Monophyletic clades were recovered for each of the major morphs (immaculata, parae and melanzona). We then identified unique Y-mers in each clade (Y-mers present in every individual of that clade but not present outside that clade). This approach provided us with all morph-mers (Y-mers present in every individual of a given morph but not present in any individual of the other morphs).

These morph-mers reveal two insights: (1) at a gross level they provide a sense of how much Y sequence is shared within, versus across, morphs and (2) mapping these morph-mers to the respective de novo genomes allows us to identify regions of morph-specific sequence[39,40,85,86]. To find these regions of morph-specific sequence we first mapped the corresponding set of 31-bp morph-mers to each of the 22 de novo male genomes (pseudohap style of Supernova output) using bowtie2 (ref. [80]) allowing for no mismatches, gaps or trimming. We found morph-mers disproportionately map to scaffolds, indicating that they came from regions of highly diverged morph-specific sequence rather than evenly dispersed lowly diverged sequence (Extended Data Figs. 3 and 4).

To verify our pipeline was identifying true Y-specific alignments, we attempted to align the Y-mers and all of the morph-mers to each of the six female de novo genome assemblies using bowtie2 (ref. [80]) (see above) and found no alignments could be made. We next verified that our morph-mers were targeting morph-specific sequence by attempting to align each of the morph-mer datasets to individuals of the opposite morphs and again found that no alignments could be made.

Coverage analysis. To independently verify that the scaffolds identified by our *k*-mer approach contained highly diverged sequence, we aligned each of the 39 individuals to the individual with the best de novo genome of each morph (on the basis of assembly size, scaffold N50 and contig N50; Supplementary Table 3). Alignments were generated with the mem function of bwa (v.0.7.17; ref. [81]). The function samtools (v.1.10; ref. [87]) was used to remove unmapped reads and secondary alignments with the fixmate function and duplicates were removed with markdup. Function bamqc was then used to assess distribution of map quality. For each individual, the average coverage of each scaffold by reads with mapq ≥60 was determined using the depth function of samtools (v.1.10; ref. [87]). Y chromosomes are notorious for high incidence of transposable elements and repeats[1]; this highly conservative filtering decreased false alignments to these regions. To account for differences across individuals in sequencing library size, we took the coverage of each scaffold divided by that individual's coverage across all scaffolds. The average raw scaffold coverage across all individuals was 29.97X; therefore, any scaffold with a corrected coverage <0.025 (raw coverage <1X) was considered to have a coverage of 0.

Gene annotation. To identify genes on morph-specific scaffolds we followed the pipeline described in Almeida et al.[28]. Briefly, we took a very conservative approach by annotating only the scaffolds with more than five morph-mers from each of the de novo references used for the coverage analysis (one of each morph). The chance of a scaffold containing any particular 31-bp *k*-mer depends on the length of the scaffold and can be calculated roughly as $0.25^{31} \times$ scaffold length. The longest male scaffold we recovered was 19,887,348 and therefore had the greatest chance of containing any given Y-mer; 4.31×10^{-12}. The most abundant morph-mers were melanzona-mers (87,629), therefore the likelihood of the largest scaffold containing one melanzona-mer by chance was $4.31 \times 10^{-12} \times 87{,}629 = 3.78 \times 10^{-7}$ and the likelihood it contains five melanzona-mers purely by chance was $\sim 7.71 \times 10^{-33}$. If we conservatively assume that all scaffolds have the same probability of containing a Y-mer by chance and the male with the most scaffolds had 25,416 scaffolds—there was a likelihood of 1.96×10^{-28} that a scaffold was incorrectly identified.

We then annotated these scaffolds with MAKER v.2.31.10 (ref. [88]). We ran the MAKER pipeline twice: first on the basis of a guppy-specific repeat library, protein sequence, expressed sequence tags (ESTs) and RNA sequence data (later used to train ab initio software) and a second time combining evidence data from the first run and ab initio predictions. We created a repeat library for these scaffolds using de novo repeats identified by RepeatModeler v.1.0.10 (ref. [89]) which we then combined with Actinopterygii-specific repeats to use with RepeatMasker v.4.0.7 (ref. [90]). Annotated protein sequences were downloaded from Ensembl (release 95)[91] for eight fish species: *Danio rerio* (GRCz11), *Gasterosteus aculeatus* (BROADS1), *Oryzias latipes* (ASM223467v1), *P. latipinna* (1.0), *P. mexicana* (1.0), *P. reticulata* (1.0), *Takifugu rubripes* (FUGU5) and *X. maculatus* (5.0). For ESTs, we used 10,664 tags isolated from guppy embryos and male testis[92]. Furthermore, to support gene predictions we also used two publicly available libraries of RNA-Seq data collected from guppy male testis and male embryos[93] and assembled with StringTie 1.3.3b (ref. [94]). As a basis for the construction of gene models, we combined ab initio predictions from Augustus v.3.2.3 (ref. [95]), trained via BUSCO v.3.0.2 (ref. [96]) and SNAP v.2006-07-28 (ref. [97]). To train Augustus and SNAP, we ran the MAKER pipeline a first time to create a profile using the protein and EST evidence along with RNA-Seq data. Both Augustus and SNAP were then trained from this initial evidence-based annotation. Functional inference for genes and transcripts was performed using the translated CDS features of each coding transcript. Gene names and protein functions were retrieved using BLASTp to search the Uniprot/Swissprot, InterProscan v.5 and GenBank databases.

Identifying transposable element activity. To compare activity of transposable elements in morph-linked Y scaffolds to the rest of the genome (autosomes and the X chromosome) we identified the proportion of sequence comprised of interspersed repeats. Sequences identified as interspersed repeats are both current transposable elements and the flanking sequence that is left behind when transposable elements leave a region and thus act as a measure of transposable element activity within a region[48]. To build a repeat library for all *P. parae* repeats we first used RepeatModeler v.1.0.10 to create repeat libraries from the best female de novo genome and from the morph-linked scaffolds (scaffolds with more than five morph-mers) from each male. We then combined individual de novo repeat libraries with the Actinopterygii-specific repeats to create the full *P. parae* repeat library. This combined repeat library was used to identify interspersed repeats in the de novo genome from each female and the morph-linked scaffolds of each male using RepeatMasker v.4.0.7. The percentage of total interspersed repeats was compared with a one-way analysis of variance in Prism v.9.0.0 which revealed significant differences ($P < 0.0001$, $F(3,25) = 17.98$). Unpaired *t*-test were then used to follow up the significant main effect.

Confirming diversity is unique to the Y chromosome. To further confirm that the extreme divergence across morphs is unique to the Y chromosome (and not cryptic subpopulations) we built phylogenies of each of the autosomes. To do this we aligned reads from each individual to the female scaffolds that had RACA-generated chromosome annotations using the mem function of bwa (v.0.6.1)[81]. Duplicates were marked with samtools (v.1.10)[87]. Variants were called using the call function of bcftools (v.1.11) before generating consensus sequences of the longest scaffold from each autosome using the consensus function of bcftools with the -I flag (providing IUPAC ambiguity codes for polymorphic sites). For each autosome, the sequences from all 40 individuals were aligned using the MAFFT[98] plugin for Geneious Prime with default parameters. Approximate-maximum-likelihood trees were generated using FastTree (v.2.1.12)[99] plugin for Geneious with default parameters (Extended Data Fig. 7).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All of the data generated for this study have been made available in the NCBI repository under the BioProject accession number PRJNA714257.

Code availability

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

Received: 25 September 2020; Accepted: 23 March 2021;
Published online: 06 May 2021

References

1. Bachtrog, D. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat. Rev. Genet.* **14**, 113–124 (2013).
2. Tobler, R., Nolte, V. & Schlotterer, C. High rate of translocation-based gene birth on the *Drosophila* Y chromosome. *Proc. Natl Acad. Sci. USA* **114**, 11721–11726 (2017).
3. Mahajan, S. & Bachtrog, D. Convergent evolution of Y chromosome gene content in flies. *Nat. Commun.* **8**, 785 (2017).
4. Bachtrog, D., Mahajan, S. & Bracewell, R. Massive gene amplification on a recently formed *Drosophila* Y chromosome. *Nat. Ecol. Evol.* **3**, 1587–1597 (2019).
5. Hall, A. B. et al. Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *Proc. Natl Acad. Sci. USA* **113**, E2114–E2123 (2016).
6. Todesco, M. et al. Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602–607 (2020).
7. Schwander, T., Libbrecht, R. & Keller, L. Supergenes and complex phenotypes. *Curr. Biol.* **24**, 288–294 (2014).
8. Wang, J. et al. A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* **493**, 664–668 (2013).
9. Lamichhaney, S. et al. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2016).
10. Kupper, C. et al. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016).
11. Branco, S. et al. Multiple convergent supergene evolution events in mating-type chromosomes. *Nat. Commun.* **9**, 2000 (2018).
12. Yan, Z. et al. Evolution of a supergene that regulates a trans-species social polymorphism. *Nat. Ecol. Evol.* **4**, 240–249 (2020).
13. Lindholm, A. K., Brooks, R. & Breden, F. Extreme polymorphism in a Y-linked sexually selected trait. *Heredity* **92**, 156–162 (2004).
14. Hurtado-Gonzales, J. L. & Uy, J. A. Alternative mating strategies may favour the persistence of a genetically based colour polymorphism in a pentamorphic fish. *Anim. Behav.* **77**, 1187–1194 (2009).
15. Hurtado-Gonzales, J. L. & Uy, J. A. Intrasexual competition facilitates the evolution of alternative mating strategies in a colour polymorphic fish. *BMC Evol. Biol.* **10**, 391 (2010).
16. Hurtado-Gonzales, J. L., Baldassarre, D. T. & Uy, J. A. Interaction between female mating preferences and predation may explain the maintenance of rare males in the pentamorphic fish *Poecilia parae*. *J. Evol. Biol.* **23**, 1293–1301 (2010).
17. Hurtado-Gonzales, J. L., Loew, E. R. & Uy, J. A. Variation in the visual habitat may mediate the maintenance of color polymorphism in a poeciliid fish. *PLoS ONE* **9**, e101497 (2014).
18. Sandkam, B. A., Young, C. M., Breden, F. M., Bourne, G. R. & Breden, F. Color vision varies more among populations than among species of live-bearing fish from South America. *BMC Evol. Biol.* **15**, 225 (2015).

19. Bourne, G. R., Breden, F. & Allen, T. C. Females prefer carotenoid colored males as mates in the pentamorphic livebearing fish, *Poecilia parae*. *Naturwissenschaften* **90**, 402–405 (2003).
20. Liley, N. R. *Reproductive Isolation in Some Sympatric Species of Fishes*. PhD thesis, Oxford Univ. (1963).
21. Bachtrog, D. et al. Are all sex chromosomes created equal? *Trends Genet.* **27**, 350–357 (2011).
22. Rice, W. R. Evolution of the Y sex chromosome in animals. *BioScience* **46**, 331–343 (1996).
23. Wright, A. E., Dean, R., Zimmer, F. & Mank, J. E. How to make a sex chromosome. *Nat. Commun.* **7**, 12087 (2016).
24. Furman, B. L. S. et al. Sex chromosome evolution: so many exceptions to the rules. *Genome Biol. Evol.* **12**, 750–763 (2020).
25. Darolti, I. et al. Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers. *Proc. Natl Acad. Sci. USA* **116**, 19031–19036 (2019).
26. Wright, A. et al. Convergent recombination suppression suggests a role of sexual selection in guppy sex chromosome formation. *Nat. Commun.* **8**, 14251 (2017).
27. Darolti, I., Wright, A. E. & Mank, J. E. Guppy Y chromosome integrity maintained by incomplete recombination suppression. *Genome Biol. Evol.* **12**, 965–977 (2020).
28. Almeida, P. et al. Divergence and remarkable diversity of the Y chromosome in guppies. *Mol. Biol. Evol.* **38**, 619–633 (2021).
29. Rabosky, D. L. et al. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature* **559**, 392–395 (2018).
30. Reznick, D. N., Miles, D. B. & Winslow, S. Life history of *Poecilia picta* (Poeciliidae) from the Island of Trinidad. *Copeia* **1992**, 782–790 (1992).
31. Haskins, C. P. & Haskins, E. F. The role of sexual selection as an isolating mechanism in three species of poeciliid fishes. *Evolution* **3**, 160–169 (1949).
32. Liley, N. R. Ethological isolating mechanisms in four sympatric species of poeciliid fishes. *Behav. Suppl.* **14**, 1–197 (1966).
33. Vicoso, B. & Bachtrog, D. Reversal of an ancient sex chromosome to an autosome in *Drosophila*. *Nature* **499**, 332–335 (2013).
34. Vicoso, B. & Bachtrog, D. Numerous transitions of sex chromosomes in Diptera. *PLoS Biol.* **13**, e1002078 (2015).
35. Vicoso, B., Emerson, J. J., Zektser, Y., Mahajan, S. & Bachtrog, D. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol.* **11**, e1001643 (2013).
36. Kim, J. et al. Reference-assisted chromosome assembly. *Proc. Natl Acad. Sci. USA* **110**, 1785–1790 (2013).
37. Pucholt, P., Wright, A. E., Conze, L. L., Mank, J. E. & Berlin, S. Recent sex chromosome divergence despite ancient dioecy in the willow *Salix viminalis*. *Mol. Biol. Evol.* **34**, 1991–2001 (2017).
38. Akagi, T., Henry, I. M., Tao, R. & Comai, L. A Y-chromosome-encoded small RNA acts as a sex determinant in persimmons. *Science* **346**, 646–650 (2014).
39. Torres, M. F. et al. Genus-wide sequencing supports a two-locus model for sex-determination in *Phoenix*. *Nat. Commun.* **9**, 3969 (2018).
40. Morris, J., Darolti, I., Bloch, N. I., Wright, A. E. & Mank, J. E. Shared and species-specific patterns of nascent Y chromosome evolution in two guppy species. *Genes* **9**, 238 (2018).
41. Napolitano, L. M. & Meroni, G. TRIM family: pleiotropy and diversification through homomultimer and heteromultimer formation. *IUBMB Life* **64**, 64–71 (2012).
42. Sardiello, M., Cairo, S., Fontanella, B., Ballabio, A. & Meroni, G. Genomic analysis of the TRIM family reveals two groups of genes with distinct evolutionary properties. *BMC Evol. Biol.* **8**, 225 (2008).
43. Karki, R. et al. NLRC3 is an inhibitory sensor of PI3K-mTOR pathways in cancer. *Nature* **540**, 583–587 (2016).
44. Sandkam, B. A. et al. *Tbx2a* modulates switching of RH2 and LWS opsin gene expression. *Mol. Biol. Evol.* **37**, 2002–2014 (2020).
45. Showell, C., Christine, K. S., Mandel, E. M. & Conlon, F. L. Developmental expression patterns of *Tbx1*, *Tbx2*, *Tbx5*, and *Tbx20* in *Xenopus tropicalis*. *Dev. Dyn.* **235**, 1623–1630 (2006).
46. Gibson-Brown, J. J., S, I. A., Silver, L. M. & Papaioannou, V. E. Expression of T-box genes Tbx2-Tbx5 during chick organogenesis. *Mech. Dev.* **74**, 165–169 (1998).
47. Tomaszkiewicz, M., Chalopin, D., Schartl, M., Galiana, D. & Volff, J.-N. A multicopy Y-chromosomal SGNH hydrolase gene expressed in the testis of the platyfish has been captured and mobilized by a *Helitron* transposon. *BMC Genet.* **15**, 44 (2014).
48. Smit, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**, 657–663 (1999).
49. Sotero-Caio, C. G., Platt, R. N. 2nd, Suh, A. & Ray, D. A. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol. Evol.* **9**, 161–177 (2017).
50. Felsenstein, J. The evolutionary advantage of recombination. *Genetics* **78**, 737–756 (1974).
51. Lenormand, T., Fyon, F., Sun, E. & Roze, D. Sex chromosome degeneration by regulatory evolution. *Curr. Biol.* **30**, 3001–3006 (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 (2008).
54. Bachtrog, D. & Charlesworth, B. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**, 323–326 (2002).
55. Mank, J. E. Small but mighty: the evolutionary dynamics of W and Y sex chromosomes. *Chromosome Res.* **20**, 21–33, (2012).
56. Kaiser, V. B. & Charlesworth, B. Muller's ratchet and the degeneration of the *Drosophila miranda* neo-Y chromosome. *Genetics* **185**, 339–348 (2010).
57. Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).
58. Ritz, K. R., Noor, M. A. F. & Singh, N. D. Variation in recombination rate: adaptive or not? *Trends Genet.* **33**, 364–374 (2017).
59. Metzger, D. C. H., Sandkam, B. A., Darolti, I. & Mank, J. E. Rapid evolution of complete dosage compensation in *Poecilia*. Preprint at *bioRxiv* https://doi.org/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 (2008).
61. Ellegren, H. Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc. Biol. Sci.* **274**, 1–10 (2007).
62. Tennessen, J. A. et al. Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. *PLoS Biol.* **16**, e2006062 (2018).
63. Bourque, G. et al. Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
64. Carleton, K. L. et al. Movement of transposable elements contributes to cichlid diversity. *Mol. Ecol.* **29**, 4956–4969 (2020).
65. Brawand, D. et al. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375–381 (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 Genom.* **19**, 339 (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 (2016).
68. Sinervo, B. & Calsbeek, R. The developmental, physiological, neural, and genetical causes and consequences of frequency-dependent selection in the wild. *Annu. Rev. Ecol. Evol. Syst.* **37**, 581–610 (2006).
69. Sinervo, B. & Lively, C. M. The rock-paper-scissors game and the evolution of alternative male strategies. *Nature* **380**, 240–243 (1996).
70. Hartl, D. L. & Clark, A. G. *Principles of Population Genetics* (Sinauer Associates, 1997).
71. Lank, D. B., Smith, C. M., Hanotte, O., Burke, T. & Cooke, F. Genetic polymorphism for alternative mating behaviour in lekking male ruff *Philomachus pugnax*. *Nature* **378**, 59–62 (1995).
72. Tuttle, E. M. et al. Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* **26**, 344–350 (2016).
73. Hunt, B. G. Supergene evolution: recombination finds a way. *Curr. Biol.* **30**, 73–76, (2020).
74. Charlesworth, D. The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol. Appl.* **9**, 74–90 (2016).
75. Joron, M. et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
76. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
77. Elyanow, R., Wu, H.-T. & Raphael, B. J. Identifying structural variants using linked-read sequencing data. *Bioinformatics* **34**, 353–360 (2018).
78. Harris, R. S. *Improved Pairwise Alignment of Genomic DNA*. PhD thesis, Pennsylvania State Univ. (2007).
79. Kent, W. J., Baertsch, R., Hinrichs, A., Miller, W. & Haussler, D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl Acad. Sci. USA* **100**, 11484–11489 (2003).
80. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357 (2012).
81. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
82. Luo, R. et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**, 2047–217X–1–18 (2012).
83. Palmer, D. H., Rogers, T. F., Dean, R. & Wright, A. E. How to identify sex chromosomes and their turnover. *Mol. Ecol.* **28**, 4709–4724 (2019).
84. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).

85. Carvalho, A. B. & Clark, A. G. Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res.* **23**, 1894–1907 (2013).
86. Carvalho, A. B., Vicoso, B., Russo, C. A., Swenor, B. & Clark, A. G. Birth of a new gene on the Y chromosome of *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **112**, 12450–12455 (2015).
87. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
88. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinform.* **12**, 491 (2011).
89. Smit, A., Hubley, R. *RepeatModeler Open-1.0*. (ISB, 2013-2015); http://www.repeatmasker.org
90. Smit, A., Hubley, R. & Green, P. *RepeatMasker Open-4.0*. (ISB, 2013-2015); http://www.repeatmasker.org
91. Howe, K. L. et al. Ensembl Genomes 2020-enabling non-vertebrate genomic research. *Nucleic Acids Res.* **48**, D689–D695 (2020).
92. Dreyer, C. et al. ESTs and EST-linked polymorphisms for genetic mapping and phylogenetic reconstruction in the guppy, *Poecilia reticulata*. *BMC Genom.* **8**, 269 (2007).
93. Sharma, E. et al. Transcriptome assemblies for studying sex-biased gene expression in the guppy, *Poecilia reticulata*. *BMC Genom.* **15**, 400 (2014).
94. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).
95. Stanke, M. et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
96. Seppey, M., Manni, M. & Zdobnov, E. M. in *Gene Prediction* (ed. Kollmar, M.) 227–245 (Humana, 2019).
97. Korf, I. Gene finding in novel genomes. *BMC Bioinform.* **5**, 59 (2004).
98. Katoh, K., Asimenos, G. & Toh, H. Multiple alignment of DNA sequences with MAFFT. *Methods Mol. Biol.* **537**, 39–64 (2009).
99. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2–approximately maximum-likelihood trees for large alignments. *PLoS ONE* **5**, e9490 (2010).

Acknowledgements

We thank the members of the Mank lab and N. Prior for stimulating conversations and excellent feedback on early drafts of the manuscript. This was supported by the Natural Sciences and Engineering Research Council of Canada through a Banting Postdoctoral Fellowship (to B.A.S.), the European Research Council (grant nos. 260233 and 680951 to J.E.M.) and a Canada 150 Research Chair (to J.E.M.). Field work was conducted under Permit 120616 SP: 015 from the Environmental Protection Agency of Guyana. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, Sweden. The CEIBA Biological Center partially subsidized our expenses during field collection in Guyana. We thank C. Lacy for the fish illustrations.

Author contributions

B.A.S. and J.E.M. designed the research. B.A.S., J.E.M, F.B. and G.R.B. conducted field work. B.A.S., P.A., I.D., B.L.S.F., W.v.d.B. and J.M. conducted bioinformatic analyses. B.A.S., P.A., I.D., 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.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41559-021-01452-w.

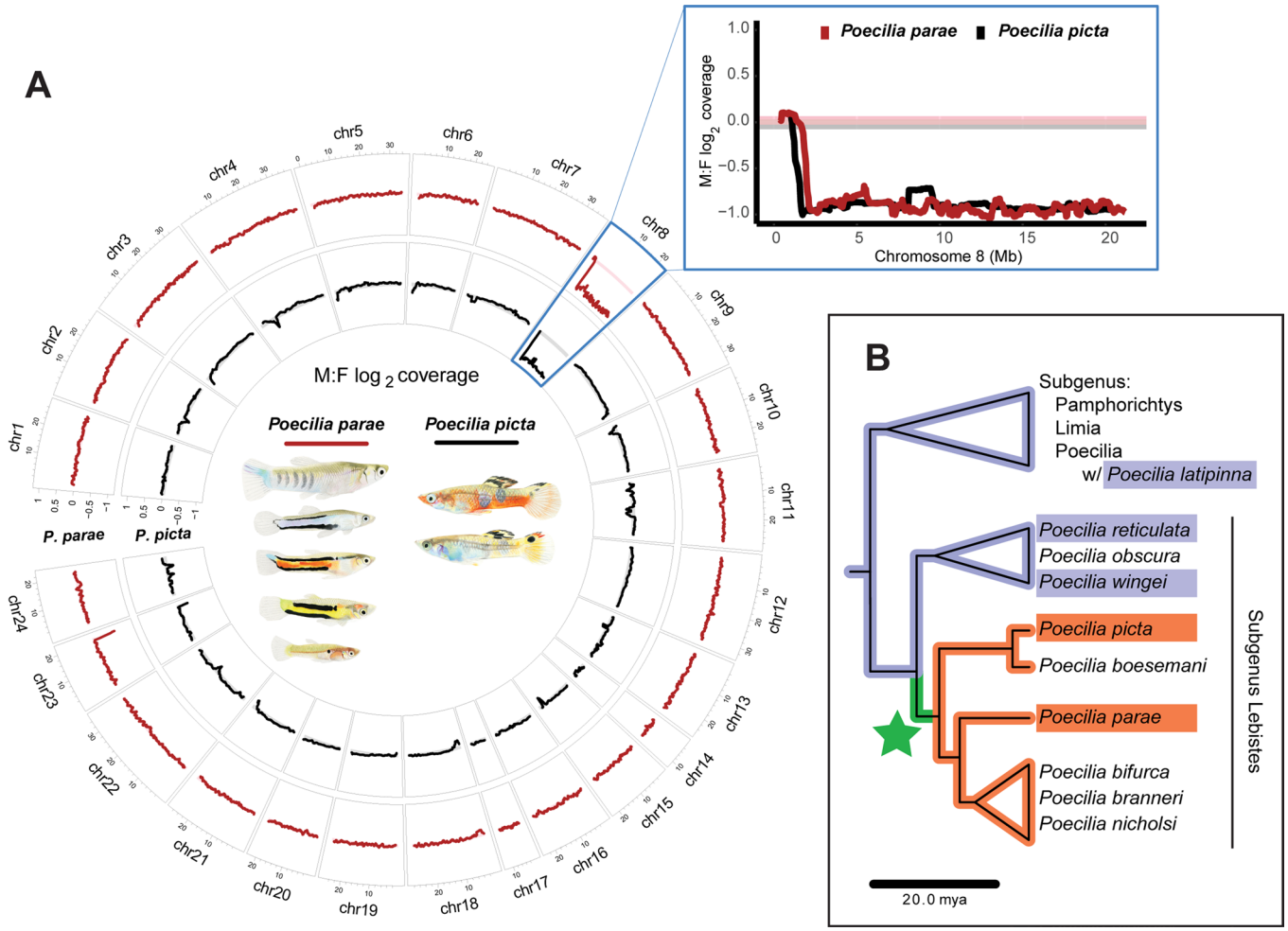
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41559-021-01452-w.

Correspondence and requests for materials should be addressed to B.A.S.

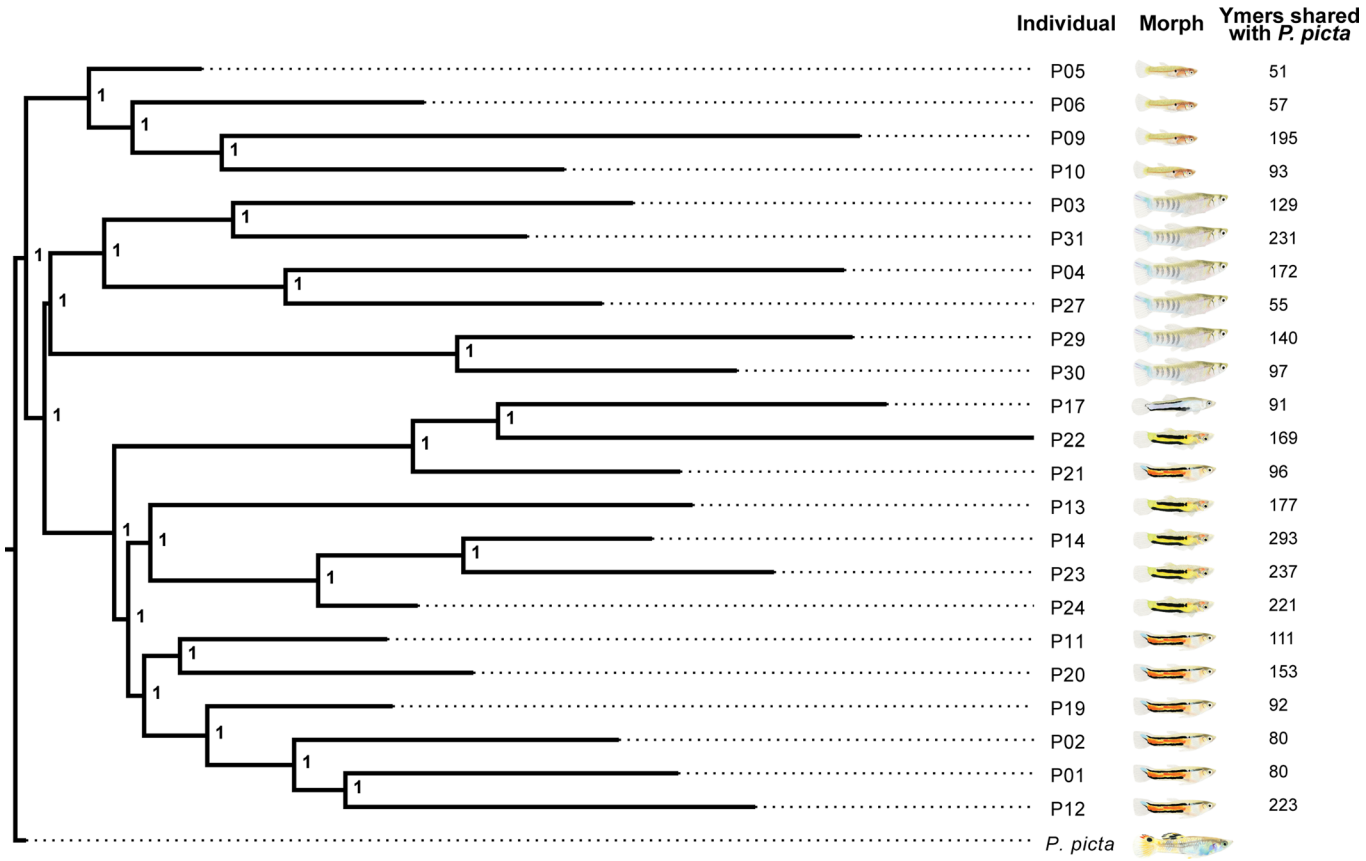
Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

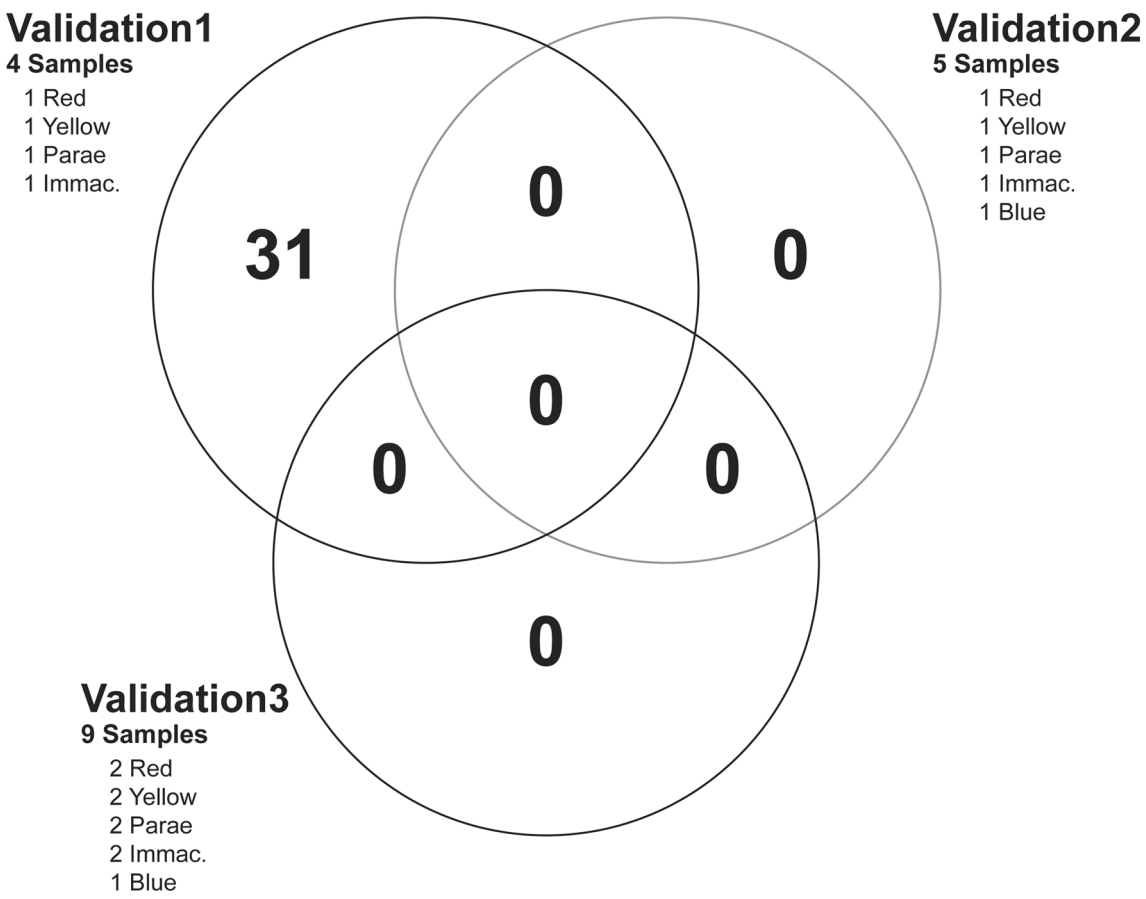
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



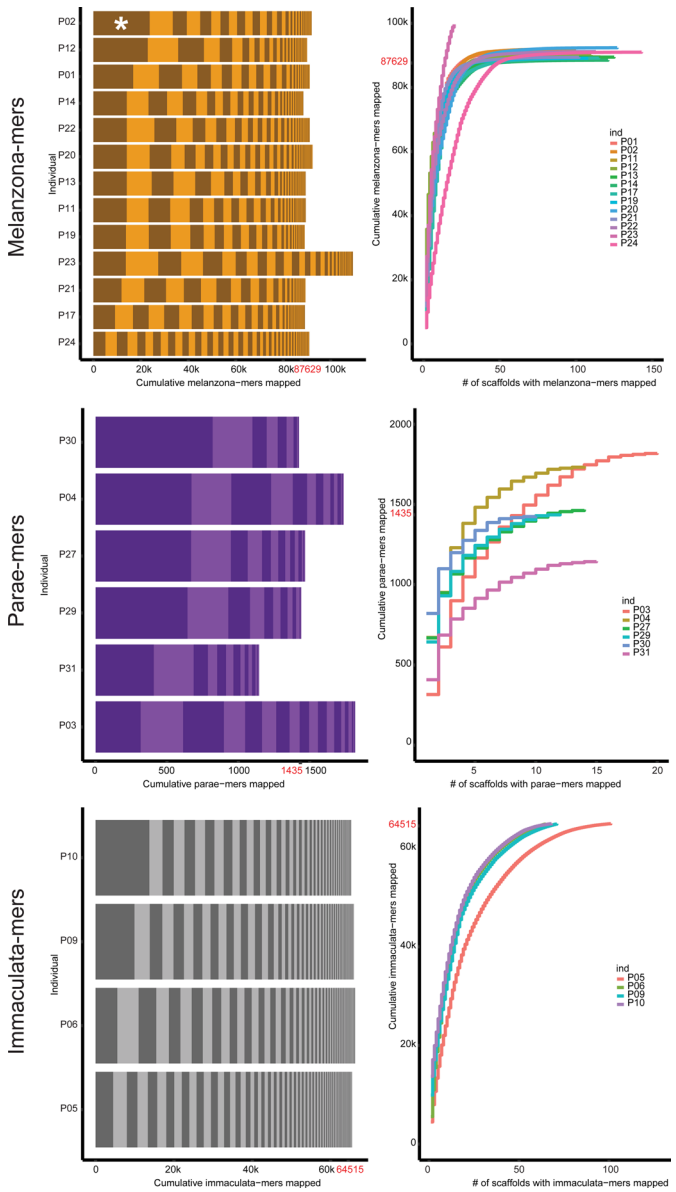
Extended Data Fig. 1 | Divergence between X and Y in *Poecilia parae* and the sister species *Poecilia picta* indicate recombination was stopped before the five morphs controlled by the Y chromosome evolved in *Poecilia parae*. a, M:F $\log_2$ coverage of RACA anchored scaffolds for all five morphs of *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. (Note, a version of the phylogeny in **b** is also presented in ref. [59]).



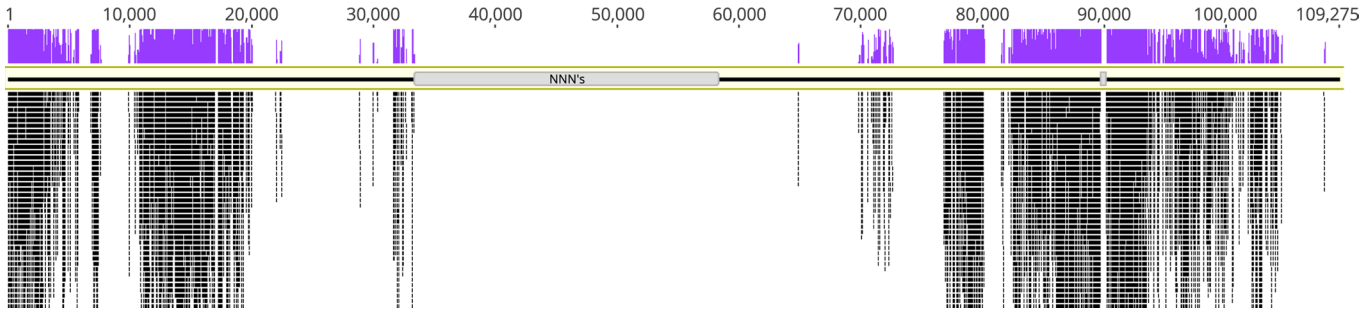
Extended Data Fig. 2 | Bayesian phylogeny built on presence/absence of the 27,950,090 *P. parae* Y-mers and the 646,754 *P. picta* Y-mers in each individual and rooted on *P. picta* (as in Fig. 2). The posterior probability is presented at each node. The number of Y-mers each individual shares with *P. 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*.



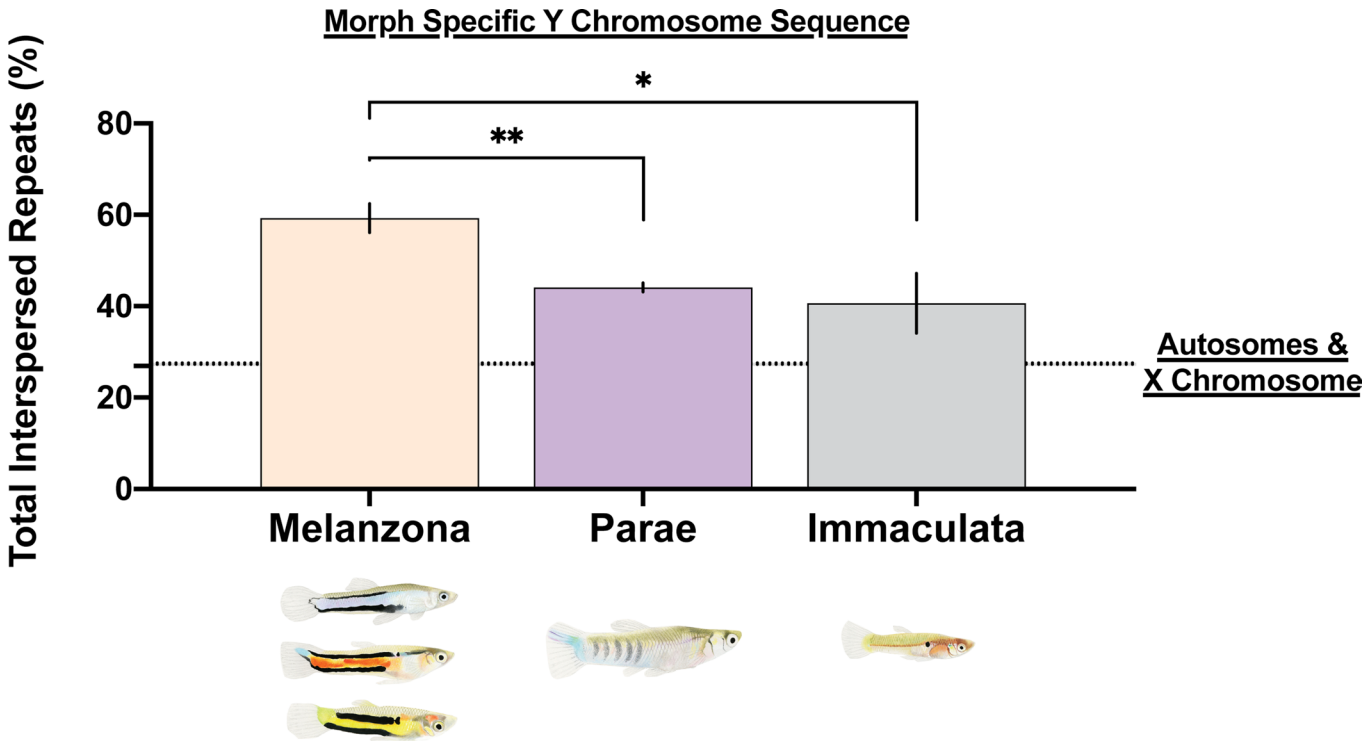
Extended Data Fig. 3 | Validation of morph-mer identification pipeline using random sets of individuals from each of the different morphs. Different samples were used for each set except for blue where the 1 sample was used in validation set 2 and validation set 3. There were a low number of Y-mers unique 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.



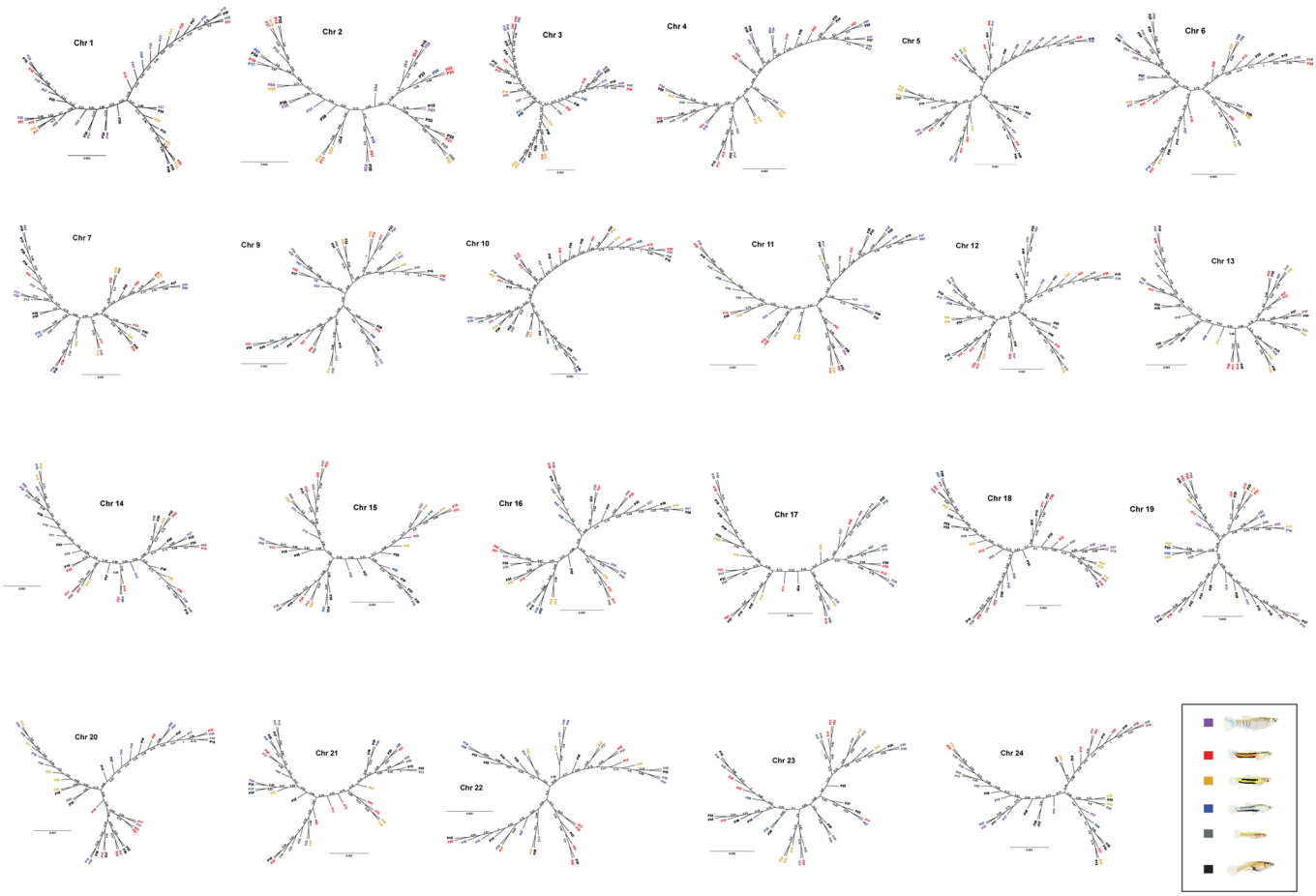
Extended Data Fig. 4 | Mapping distribution for each set of morph-mers mapped to *de novo* scaffolds of males of that morph with no mismatches, gaps, or trimming. There was a low incidence of individual morph-mers mapping to more than one scaffold (0 of the 59 Y-mers were contained in more than one scaffold 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 Extended Data Fig. 5.



Extended Data Fig. 5 | Melanzona-mers aligned to scaffold 104666 of sample P02 with no mismatches, gaps, or trimming. Each 31 bp 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.



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 (* $P<0.05$, ** $P<0.01$).



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 subpopulations. Trees were built using the consensus sequence of the longest scaffold from each chromosome (as identified by RACA). Tips denote sample name, colour indicates morph, and numbers on branches indicate FastTree support value.

nature research

Corresponding author(s): Benjamin A Sandkam

Last updated by author(s): Mar 8, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Longranger v.2.2.2 (10X Genomics), trimmomatic (v0.36), Supernova v2.1.1 software package (10X Genomics), Reference Assisted Chromosome Assembly (RACA) pipeline, LASTZ v1.04, Bowtie2 v2.2.9, soap.coverage v2.7.7, bwa v0.6.1, Jellyfish v2.2.3, MrBayes v3.2.2, bowtie2, samtools v1.10, MAKER v2.31.10, RepeatModeler v1.0.10, RepeatMasker v4.0.7, Augustus v3.2.3, BUSCO v3.0.2, SNAP v2006-07-28. All pipelines and custom scrips available at https://github.com/manklab/Poecilia_parae_Y_Diversity

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All of the data generated for this study have been made available in the NCBI repository under the BioProject accession number PRJNA714257.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☒ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Here we used a combination of linked read sequencing and straight Illumina reads to sequence the genomes of females and the five male morphs of Poecilia parae from natural populations in Guyana, South America. We then identify sequence on the Y chromosome and compare across male morphs to show that there are distinct Y chromosomes within one species.
Research sample	Wild samples of Poecilia parae were taken from natural populations around Georgetown, Guyana. P. parae is a small freshwater fish that lives in shallow streams at extremely high densities within their native range. All samples taken were adults, as evident by fully developed gonopodia and coloration in males, while adult females are substantially larger than males, lack a gonopodium and have a pronounced belly.
Sampling strategy	Fish were caught with dip nets in shallow water and sacrificed in an overdose of MS-222. Muscular tail tissue was placed in 95% EtOH and tubes were placed into liquid nitrogen to maintain high molecular weight DNA. Samples were flown to the lab and kept at -80C before DNA extraction.
Data collection	For each sample the unique identifier number, sex, morph, standard length, and photographs were taken by BAS, JEM, FB and GRB.
Timing and spatial scale	All samples were collected in November 2016.
Data exclusions	No data were excluded.
Reproducibility	Reproducibility was achieved by keeping all individuals separate during sequencing, thereby allowing for biologically independent comparisons. We then only identified sequence that was unique to all individuals of each morph.
Randomization	Randomization was largely not relevant to our study because we were looking for sequence that was unique to each morph. As a control to demonstrate our methods identified morph unique sequence, we did create random groups of individuals and showed that our methods did not have high numbers of false positives.
Blinding	Blinding was not relevant to our study as sequence level comparisons were made such that they were either present or absent and had to be exact matches.

Did the study involve field work? ☒ Yes ☐ No

Field work, collection and transport

Field conditions	Tissue collections were made in the field, where P. parae live in shallow freshwater streams. All sampling was performed in November 2016, during a mix of rain and sunny days.
Location	All populations sampled were streams 2-3m wide, and <1m deep. All locations were in and around Georgetown Guyana. 06º 48.045' N 058º 09.086W ; 06º 49.520'N 058º 08.637'W ; 06º 49.067' N 058º 06.764'W ; 06º 41.752' N 058º 12.066' W.
Access & import/export	All populations were easily accessible by public roadways. All collections were made under a permit from the Guyana Environmental Protection Agency (number: 120616 SP: 015).
Disturbance	Poecilia parae live in high density populations. Our study was deemed by the Guyana EPA to not disturb natural populations.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Study did not involve laboratory animals.
Wild animals	Study involved collection of 40 Poecilia parae individuals from wild populations. All individuals were adults. All individuals were rapidly sacrificed in an overdose of MS-222.
Field-collected samples	All field collected samples were sacrificed in an overdose of MS-222.
Ethics oversight	The Environmental Protection Agency of Guyana evaluated the ethical aspects of our work with live animals (Permit 120616 SP: 015).

Note that full information on the approval of the study protocol must also be provided in the manuscript.