Fixation of allelic gene expression landscapes and expression bias pattern shape the transcriptome of the clonal Amazon molly

Yuan Lu,[1] David Bierbach,[2] Jenny Ormanns,[3] Wesley C. Warren,[4] Ronald B. Walter,[1,6] and Manfred Schartl[1,5]

[1]*Xiphophorus Genetic Stock Center, Department of Chemistry and Biochemistry, Texas State University, San Marcos, Texas 78666, USA;* [2]*Department of Biology and Ecology of Fishes, Leibniz Institute of Freshwater Ecology and Inland Fisheries, Humboldt Universität zu Berlin, 10099 Berlin, Germany;* [3]*Biochemistry and Cell Biology, Biozentrum, University of Würzburg, 97074 Würzburg, Germany;* [4]*Bond Life Science Center, University of Missouri, Columbia, Missouri 65211, USA;* [5]*Developmental Biochemistry, Biozentrum, University of Würzburg, 97074 Würzburg, Germany;* [6]*Department of Life Sciences, Texas A&M University, Corpus Christi, Texas 78412, USA*

The Amazon molly is a unique clonal fish species that originated from an interspecies hybrid between *Poecilia* species *P. mexicana* and *P. latipinna*. It reproduces by gynogenesis, which eliminates paternal genomic contribution to offspring. An earlier study showed that Amazon molly shows biallelic expression for a large portion of the genome, leading to two main questions: (1) Are the allelic expression patterns from the initial hybridization event stabilized or changed during establishment of the asexual species and its further evolution? (2) Is allelic expression biased toward one parental allele a stochastic or adaptive process? To answer these questions, the allelic expression of *P. formosa* siblings was assessed to investigate intra- and inter-cohort allelic expression variability. For comparison, interspecies hybrids between *P. mexicana* and *P. latipinna* were produced in the laboratory to represent the *P. formosa* ancestor. We have identified inter-cohort and intra-cohort variation in parental allelic expression. The existence of inter-cohort divergence suggests functional *P. formosa* allelic expression patterns do not simply reflect the atavistic situation of the first interspecies hybrid but potentially result from long-term selection of transcriptional fitness. In addition, clonal fish show a transcriptional trend representing minimal intra-clonal variability in allelic expression patterns compared to the corresponding hybrids. The intra-clonal similarity in gene expression translates to sophisticated genetic functional regulation at the individuum level. These findings suggest the parental alleles inherited by *P. formosa* form tightly regulated genetic networks that lead to a stable transcriptomic landscape within clonal individuals.

[Supplemental material is available for this article.]

The Amazon molly, *Poecilia formosa,* is a small freshwater fish species representing a paradigm for vertebrate asexual reproduction. As with other asexual fishes, amphibia, and reptiles, *P. formosa* is an all-female species. It practices gynogenesis to produce offspring, whereby sperm from males of sympatric sexual *Poecilia* species triggers embryogenesis of diploid eggs without contributing sperm DNA to the offspring's genome. Therefore, all daughters are clones of their mothers (Schartl et al. 1991; Vrijenhoek 1994).

The advantage of an all-female lineage is a twofold higher reproduction rate than their sexual counterpart because such asexual lineages do not produce males, which do not contribute to population growth (i.e., cost of sex) (Maynard Smith 1978). This advantage allows asexual populations to grow quicker than sexual populations (Loewe and Lamatsch 2008; Stöck et al. 2010). Genetic theory predicts the Amazon molly, like other asexual lineages, should be evolutionarily short-lived (Lynch and Gabriel 1990). This hypothesis is attributed to the absence of meiotic recombination, which creates genetic diversity (i.e., "Red Queen" hypothesis) and allows for purging of deleterious genetic variation (i.e., Muller's ratchet), resulting in decreased fitness (Van Valen 1973; Bell 2019). These disadvantages are considered to outweigh the advantages of an all-female lineage. Thus, clonality should eventually lead to extinction over relatively short evolutionary times (Lynch et al. 1995; Neiman et al. 2010; Lively and Morran 2014). Such relatively transient existences of clonal lineages would explain the rarity of asexuality. Despite such theoretical projections, *P. formosa* is older than predicted (Lampert and Schartl 2008; Loewe and Lamatsch 2008; Stöck et al. 2010; Warren et al. 2018) and a successful colonizer in its natural habitats. Age estimations of *P. formosa* revealed this species has existed for about 100,000 yr, or 500,000 generations considering the generation time of 3–4 mo. This is severalfold beyond predictions from theoretical models based on Muller's ratchet (Loewe and Lamatsch 2008). To explain the persistence of the Amazon molly beyond its predicted time of extinction, it was first pointed out that *P. formosa* arose from the hybridization of two distantly related sexual molly species, *P. mexicana* and *P. latipinna* (Schartl et al. 1995). Owing to its ameiotic mode of reproduction, *P. formosa* has

Corresponding authors: phch1@biozentrum.uni-wuerzburg.de, y_l54@txstate.edu

Article published online before print. Article, supplemental material, and publication date are at https://www.genome.org/cgi/doi/10.1101/gr.268870.120. Freely available online through the *Genome Research* Open Access option.

31:372–379 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/21; www.genome.org

conserved the genomic features of an interspecies F_1 hybrid (Warren et al. 2018), thus called a "frozen hybrid genome," and benefits from heterosis/hybrid vigor. Second, many genetically different clonal lineages coexist in nature because of mutation (Schartl et al. 1991). The elevated genome-wide heterozygosity, notably, exceeds that of sympatric sexual species (Warren et al. 2018). Clones are well able to recognize sisters (own clone line) and non-sisters (different clone line), and the Amazon molly further shows levels of aggressiveness that are comparable to males of closely related sexual species (Laskowski et al. 2016; Doran et al. 2019). Thus, competition between clones is expected to eliminate those with decreased fitness and lead to survival of only the fittest clone. These attributes have been proposed as likely explanations for persistence of the Amazon molly beyond its predicted extinction time.

Although Amazon molly is of clearly known interspecific F_1 origin from known parental species, no Amazon molly has been recreated in the laboratory, despite many attempts, suggesting that establishment of Amazon molly is more complex than just mixing two distantly related genomes (Lampert et al. 2007; Stöck et al. 2010; Warren et al. 2018). The "rare formation" hypothesis has been forwarded, suggesting asexual vertebrate species are not rare because of their inferiority, but result from the rarely met very specific genomic combinations that may allow successful survival and reproduction (Stöck et al. 2010). Interspecies hybrids between the two ancestral species *P. mexicana* and *P. latipinna* can be produced under laboratory conditions because they do not suffer considerably from hybrid incompatibility. They are healthy and fertile and are in all aspects under laboratory conditions comparable to the parental species. The F_1 hybrids, particularly when *P. mexicana* was the maternal parent, sired predominantly triploid offspring when crossed to males of sexual *Poecilia* species. Production of unreduced oocytes suggested the F_1 hybrid is preadapted to gynogenesis (Lampert et al. 2007). However, laboratory hybrids lacked the mechanism of sperm exclusion to proceed to completion of gynogenesis. This sperm exclusion mechanism in Amazon molly occasionally fails, and paternal introgression occurs, causing the generation of triploid offspring (Lamatsch et al. 2000, 2009). These results suggest the key for gynogenesis may be the sperm exclusion mechanism contributed by the rare genomic situation to generate the first Amazon molly.

Interspecific hybrids, one of which was the "prima Eva" of the Amazon molly, benefit from synergistic genetic interactions but are expected to suffer from negative epistatic interactions between genes from different parental genomes. One possibility of reducing the impact of hybrid incompatibilities and resolving the conflict between the allospecific genomes in an F_1 hybrid is allelic expression bias. An earlier study (Warren et al. 2018) revealed that in *P. formosa* 5% of the genes show allele-specific expression from one of the ancestral parental genomes. Therefore, in this study, we aim to compare allelic expression divergence intra-cohort or inter-cohort of clonal Amazon molly and laboratory-produced F_1 interspecies hybrid to answer two critical questions related to the genomic conditions from which *P. formosa* originated and to its molecular evolution: (1) Are the allelic expression patterns (i.e., relationships of gene expression from both parental alleles) from the initial hybridization event stabilized or altered during establishment of the asexual species and upon its further evolution? (2) Is expression variation and allelic expression biased toward one parental allele or is it a stochastic process? To answer these questions, global allelic expression was compared both within clonal *P. formosa* individuals and between *P. formosa* and interspecies F_1 hybrids produced in the laboratory.

Results

Morphological similarity of *P. formosa* and interspecies F_1 hybrids

Previous work has unequivocally shown that *P. formosa* originated from a hybridization event between *P. mexicana* and *P. latipinna*, in which *P. mexicana* and *P. latipinna* served as the maternal and paternal species, respectively. The wild-caught *P. formosa* and the laboratory-produced F_1 interspecies hybrid between the two ancestral parental species do not show noticeable morphological differences (Fig. 1), and no malformations or other gross characteristics of hybrid dysgenesis were observed. We performed RNA-seq of both clonal and F_1 fish (Supplemental Table S1). To confirm the maternal lineage of *P. formosa*, sequencing reads were mapped to both *P. mexicana* and *P. latipinna* mitochondrion genomes. Mitochondrial gene expression for both *P. formosa* and F_1 interspecies hybrids was predominant (99%) from *P. mexicana*, confirming that mitochondria were inherited from female *P. mexicana* in both cohorts (Supplemental Fig. S1).

Intra-cohort and inter-cohort allelic expression comparison

Clonal *P. formosa* and F_1 hybrids showed equal expression of parental alleles for most of the genes (Fig. 2). However, on average, 12.95% of genes in *P. formosa* (13.06% in brain, 6.84% in liver, and 18.93% in ovary) and 11.17% of genes in F_1 hybrids (10.47% in brain, 6.03% in liver, and 17.00% in ovary) displayed <40%, or more than 60% expression from one parental allele [χ^2 test, $|\mathrm{Log}_2$ (*P. latipinna* expression/*P. mexicana* expression)|>0.27; *P*-value < 0.05] (Fig. 2). Consistently, *P. formosa* showed a larger number of genes with parental allele-biased expression than F_1 hybrids.

To compare allelic expression patterns between clonal and F_1 cohorts, a gene was determined to show "consistent allelic expression" if the parental allelic expression patterns within each cohort are the same (Fig. 3; Supplemental Figs. S2–S6). In all assessed organs, clonal *P. formosa* displayed more genes showing consistent expression patterns than interspecies F_1 hybrids, regardless of allelic expression bias (Fig. 3; Supplemental Tables S2, S3; Supplemental Figs. S2–S6).

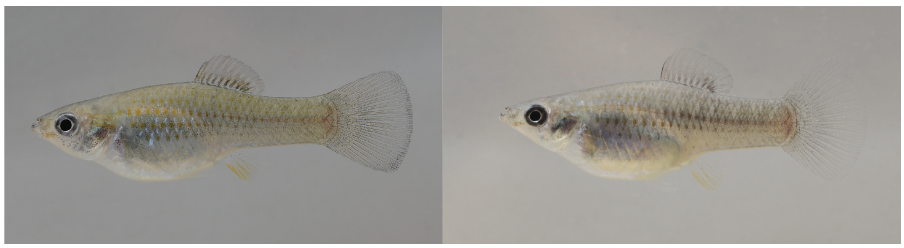

Figure 1. Clonal *P. formosa* and artificial interspecies hybrid *P. formosa* caught from the wild (*left*) and an artificial interspecies hybrid between female *P. mexicana* and male *P. latipinna* (*right*) do not show any noticeable phenotypical differences.

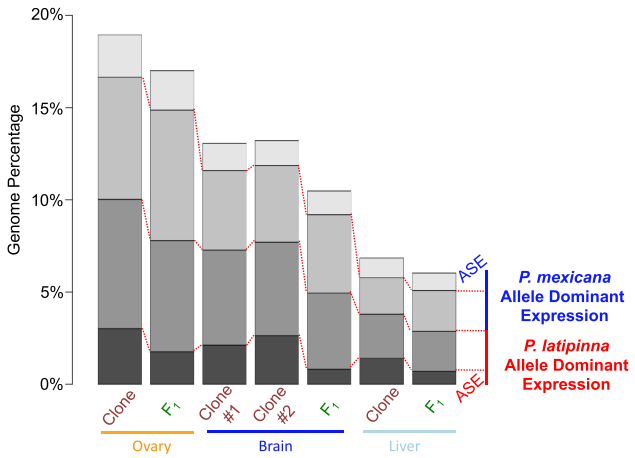

Figure 2. Percentage of genome showing biased parental allelic expression. Stacked bar graphs show the percentage of the genome that shows biased or monoallelic parental gene expression: (ASE) allele specifically expressed genes.

Comparison between clonal and F_1 cohorts revealed a majority of the transcriptome showed consistent allelic expression patterns between the two cohorts (see Supplemental Table S1, "clonal.no.bias_f1.no.bias" and "clonal.bias_f1.bias_same"). In contrast, there were 11.0–20.8% of genes showed different allelic expression patterns between the two cohorts in different organs (see Supplemental Table S1, "clonal.no.bias_f1.bias," "clonal.no.bias_f1.inconsistent," "clonal.bias_f1.no.bias," "clonal.bias_f1.bias_reversed," "clonal.biased_f1.inconsistent," "clonal.inconsistent_f1.no.bias," and "clonal.inconsistent_f1.bias"). Among these genes, 2%–6% displayed allelic usage differences between the clonal *P. formosa* and the interspecies F_1 hybrids among the three organs assessed: 0.4%–1.9% of genes showed equal contribution by parental alleles in clonal fish but unequal contribution in the interspecies F_1; 1.5%–3.5% of genes showed unequal contribution to gene expression by parental alleles in clonal fish but equal contribution in interspecies F_1; and 0.1%–0.3% of genes displayed conflicting preference in allelic usage (i.e., *P. mexicana* alleles biased in clonal fish, but *P. latipinna* allele biased in F_1, or vice versa) (Supplemental Table S2). Those genes that showed divergent allelic expression bias in both *P. formosa* and interspecies hybrids have a consistent expression pattern in all individuals of each cohort, suggesting these genes are "fixed" for their expression patterns.

To investigate different allelic expression patterns between clonal *P. formosa* and interspecies F_1 hybrids, we selected genes that showed fixed but divergent allelic expression patterns (i.e., unequally expressed in F_1, equally expressed in clonal; equally expressed in F_1, unequally expressed in clonal; and unequally expressed in both F_1 and clonal, but reversed pattern) and studied their particular expression patterns between the two populations to infer transcriptional adaptation following the initial interspecies hybridization event. In brain, liver, and ovary, 747, 434, and 1208 genes fit these criteria, respectively (Supplemental Tables S4–S6). Not only did *P. formosa* display more parental allele-biased gene expression, but there were also more genes dominantly expressed from *P. latipinna* alleles (Supplemental Figs. S7–S9; Supplemental Table S3), although both cohorts showed similar numbers of heterozygous loci (Supplemental Fig. S10). Similar results were observed for other clones and in different organs (Supplemental Fig. S9). These observations suggest that *P. formosa* harbors more genes that are dominantly expressed from *P. latipinna*, especially allele specifically expressed (ASE) genes. Among the genes showing biased allelic expression patterns, 17.5% in *P. formosa*, but only 9.9% in F_1 were ASE exclusively expressing *P. latipinna* alleles. For genes specifically expressing *P. mexicana* alleles, both cohorts were more similar: 13.0% in *P. formosa* and 13.5% in F_1 (Fig. 2). *P. latipinna* ASE genes are not distributed randomly within the genome but show enrichment in specific contigs (i.e., monoallelic expression of the *P. latipinna* allele, biallelic expression in F_1 hybrids) (Fig. 4; Supplemental Figs. S11, S12). PCR analyses from genomic DNA of *P. formosa* confirmed these observations were not caused by loss of the corresponding region from the *P. mexicana* derived locus, but instead are likely a result of transcriptional silencing of *P. mexicana* alleles (Supplemental Table S7).

Of the genes showing divergent allelic expression patterns in brain ($n = 747$), liver ($n = 434$), or ovary ($n = 1208$), 153 show divergent allelic expression patterns in all assessed organs. Although all these genes display different allelic expression patterns between clonal *P. formosa* and F_1 interspecies hybrids, they largely show the same allelic bias in brain, liver, and ovary of both *P. formosa* and the F_1 hybrids, with only eight of the 153 genes showing different allelic bias in the different organs (Fig. 5; Supplemental Tables S8, S9).

Discussion

In this study, we aimed to answer two major questions: (1) Are the *P. formosa* allelic expression patterns stabilized from the initial hybridization event? (2) Is biased allelic expression toward one or the other parental allele stochastic or results of a deterministic process?

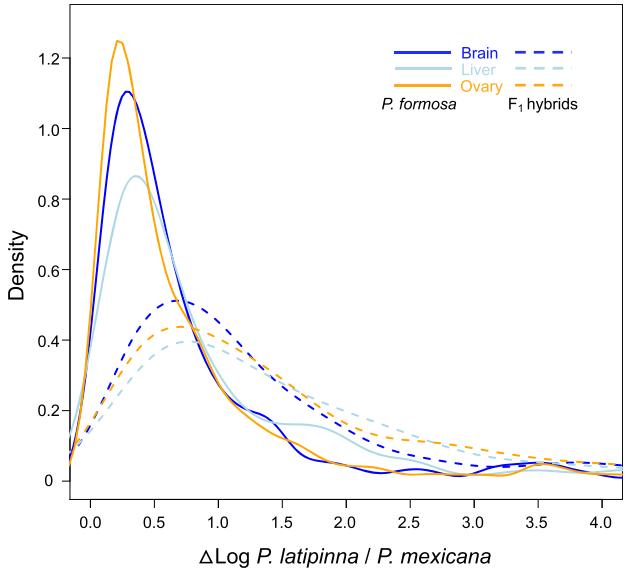

Figure 3. Comparison between clonal and F_1 hybrid fish allelic expression pattern differences. Genes showing consistently biased allelic expression patterns within the clonal fish cohort, but inconsistent allelic expression patterns within the F_1 hybrid fish cohort, or the reverse case, were used to calculate the maximum allelic expression pattern difference [(Log *P. latipinna*/*P. mexicana*)$_{max}$ – (Log *P. latipinna*/*P. mexicana*)$_{min}$] for both clonal and F_1 hybrid fish in all three organs. Density curves of allelic expression pattern differences were plotted to represent distribution of these values in both fish cohorts.

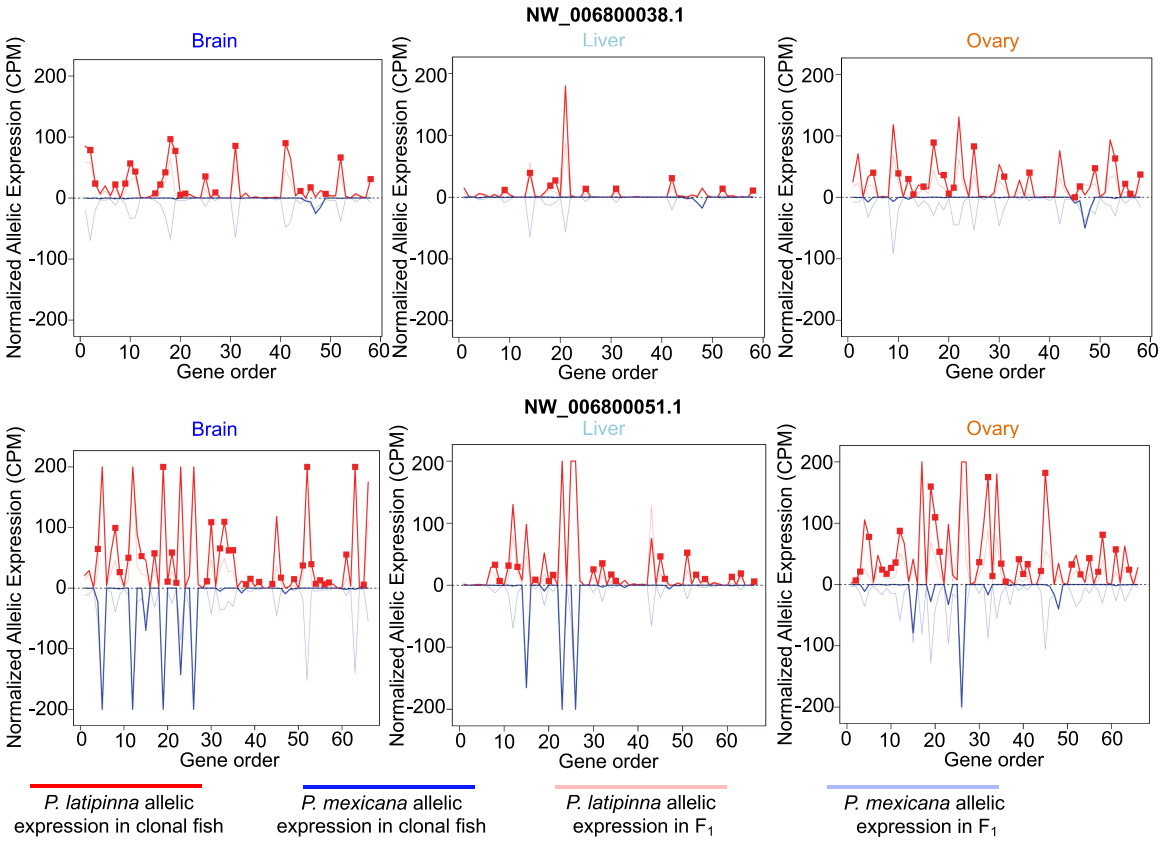

Figure 4. Clustering of genes showing *P. latipinna*-biased expression in *P. formosa.* Allelic expression of two contigs that are enriched of *P. latipinna*-biased genes in all organs are plotted. The *x*-axis is the order of gene on contig, and the *y*-axis represents normalized allelic expression of *P. latipinna* (*above* the *x*-axis), *P. mexicana* (*below* the *x*-axis) alleles. Solid lines of different colors (red: *P. latipinna* allele; blue: *P. mexicana* allele) represent both parental allelic expression in the *P. formosa*, and the faint lines of different colors (red: *P. latipinna* allele; blue: *P. mexicana* allele) represent allelic expression in the artificial interspecies hybrid. Red squares designate the *P. latipinna* alleles that show dominant expression in clonal fish but equal expression as *P. mexicana* alleles in the F_1 hybrids.

To answer these questions, we assessed allelic expression and intracohort allelic expression pattern consistency of *P. formosa*, and between one clone of *P. formosa*, and F_1 hybrids produced from known parental species that gave rise to *P. formosa* more than 100,000 yr ago (Stöck et al. 2010).

To address the first question, we compared clonal *P. formosa* and interspecies hybrids. Their differences in allelic expression patterns were observed at both the cohort level and the individuum level. At the cohort level in *P. formosa*, a higher fraction of the genome showed parental allele-biased expression in all assessed organs (Fig. 2). At the individuum level, allelic expression patterns are more consistent among *P. formosa* than F_1 hybrids (Fig. 3; Supplemental Figs. S2–S6). Because all *P. formosa* were raised under identical conditions and they have identical genomes, the variation in allelic expression patterns should not be a result of extrinsic factors. Stochastic processes and epigenetic mechanisms may be involved. It will be interesting to evaluate if the observed individual expression signatures play a role in individuality and behavioral personality (Bierbach et al. 2017). Although the laboratory-raised F_1 interspecies hybrids are only an approximation of the Amazon molly ancestor, they served to model the genomic composition of them. The extant allelic expression pattern we observed in *P. formosa* may be a final or transient result of selection and fixation, and the differences in allelic expression patterns between clonal and F_1 fish may indicate that the transcriptional patterns have diverged since the initial hybridization event. Functional analyses of those genes that appear to have changed their expression profile during evolution revealed enrichment for signaling pathways centered on PI3K and NF-kB genes (i.e., *pik3r4* and *nfkb1*), including Rac signaling, eIF signaling (Regulation of eIF4 and p70S6K signaling, and eIF2 signaling), and estrogen-dependent cell proliferation (Estrogen-Dependent Proliferation Signaling) (Supplemental Table S11), the latter being intriguing because *P. formosa* is an all-female lineage. Thus, suggesting functional changes in these pathways may partially lead to establishment of a functional *P. formosa* transcriptional landscape. Therefore, for the first question, we can conclude that allelic expression patterns have evolved to become different between the extant *P. formosa* and their hypothetical single common ancestor.

For the second question, the data suggests that although *P. formosa* had more genes showing parental allele-biased expression patterns than F_1, the expression patterns are relatively consistent within the *P. formosa* clone. In addition, clones share the feature of a higher percentage of the genome showing allele-biased expression toward *P. latipinna* and more *P. latipinna* ASE genes (i.e., monoallelic expression) (Fig. 2). In contrast, such observation was not made within F_1 hybrids. Although we only included F_1 hybrids from one set of parental fish, our previous study using different F_1 interspecies hybrid (i.e., *Xiphophorus maculatus* x *Xiphophorus couchianus*) showed allelic expression patterns in interspecies hybrids,

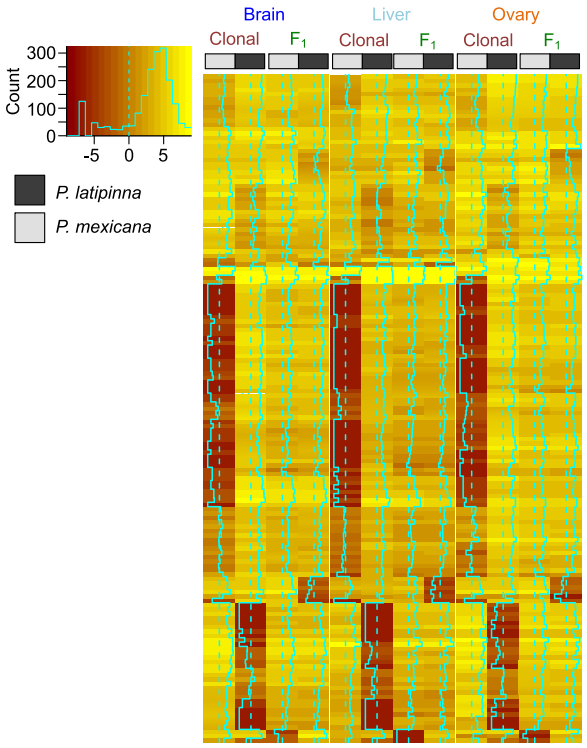

Figure 5. Differences in allelic expression pattern between clonal *P. formosa* and interspecies F_1 hybrid. A total of 153 genes were identified to display allelic expression pattern differences between the clonal and interspecies F_1 hybrid progeny in all three organs assessed. The heatmap represents allelic expression in different organs and cohorts. Colored blocks represent mean Log_2 allelic expression. Cyan dashed lines in the color blocks mark the center value (0) of the heatmap, and the solid lines that are close to the *center* dashed line display the allelic expression value. If the line is on the *left* of the dashed line, the given allele is lowly expressed. If the line is on the *right* of the dashed line, the given allele is relatively highly expressed. Color key displays the relationship between colors and values of allelic expression, with the histogram showing the summary of allelic expression levels for all genes.

unlike Amazon molly, are equally represented by both parental alleles (Lu et al. 2015). Therefore, we conclude allelic expression patterns of an individuum are not a result of stochastic processes but caused by fixed characters. We noted the two independent *P. formosa* clones showed 18.9% of the transcriptome expressed differently for allelic expression and intra-cohort allelic expression consistency (Supplemental Table S10; Supplemental Fig. S13). Although such inter-clonal differences are likely caused by a combination of genetic, age, and environmental factors in this study, we can estimate the genetic contribution to such divergences is smaller between clones than between *P. formosa* and F_1 hybrids (i.e., 20.5%) (Supplemental Table S2). It has been shown that extant *P. formosa* populations comprise many clonal lineages, which arose by genome divergence owing to natural mutations. The elevated genome-wide heterozygosity of *P. formosa* exceeds that of sympatric sexual species (Warren et al. 2018). Therefore, interclonal expression divergence is not unexpected.

An unexpected finding is that clonal fish had more genes showing *P. latipinna* allele-biased expression. It has been observed that in rare instances, exclusion mechanisms of the male sperm DNA trigger for gynogenesis is faulty, and some genetic material of the sexual host species remains in the clonal *P. formosa* lineage. We can exclude that *P. latipinna* bias is a result of paternal introgression because the *P. formosa* used in this study were collected from habitats where the clones occur in exclusive sympatry with *P. mexicana*, meaning *P. mexicana* males were the ones that "mate" with *P. formosa*. The collection site is far from the natural range of *P. latipinna* in Mexico. Also, *P. latipinna* was not used as the sperm donor in the laboratory. Therefore, *P. mexicana* alleles would be expected to be overexpressed if there is paternal genome introgression, not *P. latipinna*. Second, such bias is not a result of purging of "unfavored" alleles because (1) the recombination required to do so is absent in *P. formosa* embryogenesis, and (2) PCR analyses showed the silenced alleles are still present in *P. formosa* genome (Supplemental Table S7). Therefore, the observed expression bias to *P. latipinna* alleles is caused by transcriptional or epigenetic regulation and may result from selection.

Although we generated an interspecies F_1 hybrid from the two ancestral species *P. mexicana* and *P. latipinna,* recreation of Amazon molly has not been successful (Lampert et al. 2007). The inability to do so might be because the original hybridization was a special event requiring specific allelic expression patterns and particular mutations. For this initial hybrid to successfully produce further viable and fertile offspring, producing unreduced oocytes and establishing mechanisms in excluding sperm genetic material are two prerequisites. Our prior trials in producing laboratory hybrids using *P. mexicana* as a maternal parent and *P. latipinna* as a paternal parent, but not the reciprocal cross direction, showed 50% of the F_1 produce diploid oocytes, suggesting the hybrids are preadapted to the parthenogenesis (Lampert et al. 2007). However, offspring of such F_1 are triploids, indicating a lack of the capacity to reject sperm DNA. We also found that Amazon molly rarely produces triploid offspring, suggesting the sperm rejection mechanism can fail in rare cases (Lamatsch et al. 2000, 2009). This evidence indicates that sperm exclusion is a sophisticatedly controlled mechanism; establishing such a mechanism may be caused by a rare incidence triggered by genomic shock in the original hybrid. Therefore, one future direction of the Amazon molly should focus on expanding the cohorts of *P. latipinna* and *P. mexicana* hybrid F_1 to allow such a rare incident to take place, and enable a population level comparison between Amazon molly and laboratory-produced F_1 to identify loci contributing to the stabilization of Amazon molly genome.

The consistency of higher allelic expression suggests *P. formosa* inherited sophisticated *cis*- (e.g., regulatory sequence) and *trans*- (e.g., transcription factor) regulation mechanisms of gene expression. We established a model to explain the genetics underlying the observed "tighter" gene expression regulation in clonal lineages (Fig. 6). Upon interspecies hybridization, coadapted *cis*- (e.g., regulatory sequence and target gene) and *trans*-regulators (e.g., transcription factor and target gene) in each parental species are conserved in the hybrid. However, feedback interactions of a target gene with its own *cis*-element may be interfered with by the presence of a similar product from the other parental allele. Similarly, *trans*-regulators can interfere with expression of the other divergent allele. Such effects are minimized or eliminated within *P. formosa* siblings owing to the emergence of common regulators that control both parental alleles or divergence of parental alleles through mutation or epigenetic alterations. Although our current data set is not informative in explaining the molecular mechanism that established Amazon molly, it provided a collection of molecular traits that can be used to answer this question in a future study. Comparing the evolution of both parental alleles of Amazon molly and the assessment of gene expression under

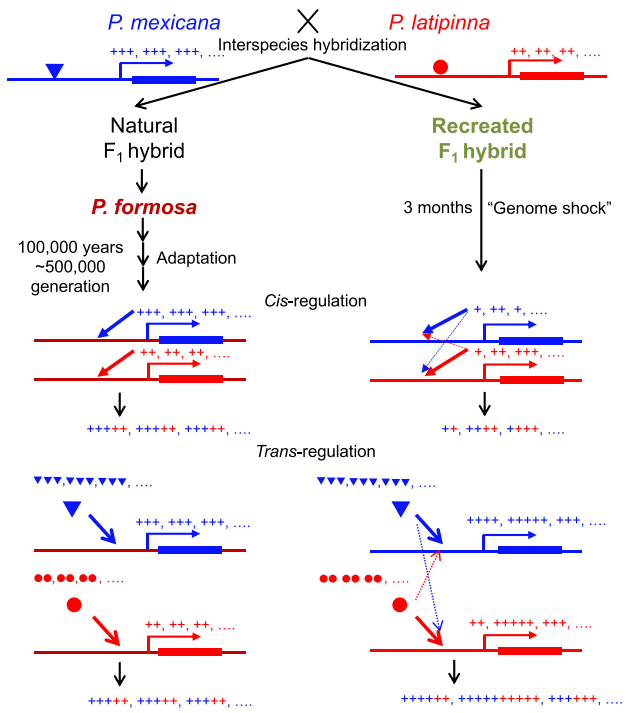

Figure 6. Schematic illustration of establishment of inter-allelic regulation. This figure illustrates novel inter-allelic expression upon hybridization that disrupts the coadapted parental regulation network and how parental allele expression is regulated in *P. formosa* to reach a consistent gene expression pattern within clonal fish. Colors represent different parental alleles, and symbols represent different parts of the gene expression regulation network: (thin line) genome sequence; (thick lines) gene; (triangle or circle) *trans*-element; (comma delimited "plus" sign) expression levels of different individuals; (thick solid arrows) adapted interactions; (thin dashed arrows) unadapted interactions. *Cis*- and *trans*-regulators and target genes are coadapted within each parental genome, respectively (i.e., *P. mexicana*, blue; *P. latipinna*, red). Upon hybridization (i.e., Recreated F_1 hybrid), cross-interaction between the *P. mexicana* product and *P. latipinna cis*-regulatory element disrupts the regulation on *P. latipinna* gene products and vice versa. In *P. formosa,* such cross-interaction is eliminated owing to the alteration of one parental allele of *cis*-element by mutation or epigenetic modulation, divergence of both alleles, or development of a common *cis*-element that adapts to both gene products. For *trans*-regulation, even *trans*-regulators, for example, transcription factors, are expressed at a similar level in different individuals; cross-interaction can disrupt regulation on target gene expression. *P. formosa* clonal offspring inherited the same regulatory element coevolved with both parental alleles. Such regulatory elements minimize unregulated inter-allelic effects.

positive selection can indeed provide insight into how fixation of the transcriptomic landscape is formed. Performing such analyses calls in the necessity of long continuity haploid genome assemblies for Amazon molly.

In summary, the allelic expression patterns of highly heterozygous genomes are fixed following an initial hybridization event. *P. formosa* shows a low level of intra-clonal transcriptional variability associated with consistency in gene expression regulation.

Methods

Research animals

We used two different clones of *P. formosa* (clone 1, $N = 4$ individuals sampled, clone 2, $N = 19$) as well as F_1 interspecies hybrids produced by mating a female *P. mexicana* and a male *P. latipinna* ($N = 6$) for our study. Founder fish for clone 1 of the laboratory strain of *P. formosa* were collected from the Canal Principal E at Ciudad Mante, Tamaulipas, Mexico, where *P. mexicana* is the sperm donor host species. A clonal lineage of this *P. formosa* collection was maintained in the aquarium (WLC#1588) for over about 70 generations before dissections of organs for transcriptome profiling was performed. Founder fish clone 2 of the laboratory strain of *P. formosa* was collected in 2001 near Tampico, Mexico. Founder fish of the *P. mexicana* (WLC#1353) and *P. latipinna* (WLC#1368) originated from Laguna Champaxan at Altamira, Tamaulipas, Mexico. F_1 interspecies hybrids were produced by mating a single virgin female *P. mexicana* and a male *P. latipinna* under regular aquarium breeding conditions.

P. formosa clone 1 ($N = 4$) and F_1 interspecies hybrids ($N = 6$) were from a single female and the same brood. They were raised under identical conditions in the fish room of the Biocenter of the University of Würzburg and were size and age (3 mo) matched. Animals were kept and sampled in accordance with the applicable EU and German national legislation governing animal experimentation, in particular, all experimental protocols were approved through an authorization (568/300-1870/13) of the Veterinary Office of the District Government of Lower Franconia, Germany, in accordance with the German Animal Protection Law (TierSchG).

P. formosa clone 2 ($N = 19$) samples stem from four females that were age-matched sisters. They were raised at the Leibniz-Institute of Freshwater Ecology and Inland Fisheries (Berlin). Sampling took place at the age of 10 mo old, and experimental protocols were approved by Berlin's Landesamt für Gesundheit und Soziales (LaGeSo, permit number G0124/14).

Confirmation of parental allele heterozygosity

To show that both parental scaffolds are present in the *P. formosa* genome, primers were designed that amplify products of different sizes in *P. mexicana* and *P. latipinna* (Supplemental Table S7).

RNA isolation

Brain, liver, and ovary from four *P. formosa* (clone 1) and six interspecies hybrids and brain of 19 *P. formosa* from an independent clone (clone 2) were sampled for RNA isolation. Total RNA was isolated using TRIzol Reagent (Thermo Fisher Scientific) according to the supplier's recommendation. All samples were treated with DNase. Total RNA concentration was determined using a Qubit 2.0 fluorometer (Life Technologies). RNA quality was verified on an Agilent 2100 Bioanalyzer (Agilent Technologies) to confirm that RIN scores were above 8.0 before sequencing.

RNA sequencing and data processing

The poly(A) RNA of each sample was first enriched and subsequently forwarded to a single sequencing library construction. Libraries were sequenced using the BGI-Seq system (sequencing strategy: 2×100 bp). Adaptor sequences were first removed from sequencing reads by the BGI-Seq pipeline. Sequencing reads were further trimmed to remove low-quality base calls at the end of the sequencing read (Phred score ≥ 30 for the last base call, with a remaining sequencing read at least 35 nt long) and filtered to keep only sequencing reads with high base call quality (Phred score ≥ 30 for at least 80% of all base calls) using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).

Assessment of gene expression and allelic expression

To assess total gene expression, filtered short sequence reads from *P. formosa* and *P. latipinna*-*P. mexicana* F_1 brain, liver, and ovary were mapped to the *P. formosa* genome (GCF_000485575.1) using TopHat2 (Kim et al. 2013). Mapped reads were quantified as raw sequencing read counts by the Subreads package function "featureCounts" and then converted to counts per million (cpm) for each sample (Liao et al. 2014):

$$cpm_i = \frac{I_i}{\sum_{1\to n} I} \times 10^6,$$

where I_i is the row count of gene i of a genome containing n gene. A gene was determined to be expressed if at least one sample of the biological replicates reached a library size normalized read count (i.e., count per million sequencing reads [cpm]) of 1.

To assess allelic expression, *P. latipinna* (GCF_001443285.1) and *P. mexicana* (GCF_001443325.1) reference RNA sequences were downloaded from NCBI Assembly database (https://www.ncbi.nlm.nih.gov/assembly). Sequence homology between parental alleles (i.e., *P. latipinna* and *P. mexicana*) and a *P. formosa* gene were identified using BLASTN (-evalue 1×10^{-6}, -best_hit_score_edge 0.1, -best_hit_overhang 0.1, -num_alignments 1, -max_hsps 1) (Shen et al. 2013; Lu et al. 2015). When multiple representations of homology were observed, the parental allele that generated the longest sequence alignment was kept to represent one parental allele of a *P. formosa* gene. Of 25,338 coding genes annotated in the *P. formosa* genome that have a genome feature as "mRNA," 22,118 genes can be assigned as both *P. latipinna* and *P. mexicana* alleles. Among these orthologous pairs, 21,119 showed transcript length differences less than twofold and were kept for allelic expression profiling. Sequences of both parental alleles were combined into a single reference sequence file to represent a hybrid genetic background for both *P. formosa* and F_1 interspecies hybrid. In addition to the *P. formosa* and interspecies F_1 described in the research animal section, additional sequencing files of liver, skin, and gills from independent clones were downloaded, followed by the same data processing for allelic expression assessment (NCBI Sequence Read Archive [SRA; https://www.ncbi.nlm.nih.gov/sra] under accession numbers SRR629501, SRR629518, SRR629511, SRR629503, SRR629508, SRR629510). This clone was derived from the Rio Purification near Barretal, Tamaulipas, Mexico.

Short sequencing files generated from *P. formosa* and F_1 interspecies hybrid brain, liver, and ovary were mapped to the hybrid reference sequences using Bowtie 2 (Langmead et al. 2009). A custom Perl script was used to retrieve and quantify the short reads that only aligned to one of the parental alleles (Lu et al. 2015). The sequencing reads that mapped to the polymorphic sites between the two parental alleles were quantified and used to calculate the allelic composition of expressed genes. Total gene expression was further assigned to parental allele expression by normalizing the gene expression cpm values to the allelic composition and allele length as follows:

$$A^{mex}{}_i = cpm_i \times (R^{mex}{}_i/L^{mex}{}_i)/[(R^{mex}{}_i/L^{mex}{}_i) + (R^{lat}{}_i/L^{lat}{}_i)];$$

$$A^{lat}{}_i = cpm_i \times (R^{lat}{}_i/L^{lat}{}_i)/[(R^{mex}{}_i/L^{mex}{}_i) + (R^{lat}{}_i/L^{lat}{}_i)]$$

where A_i is *P. mexicana* or *P. latipinna* allelic expression of gene$_i$; R_i is the number of reads that only map to a *P. mexicana* or *P. latipinna* allele at interspecific polymorphic sites of gene$_i$; and L_i is the length of transcript length of *P. mexicana* or *P. latipinna* of gene$_i$. To identify loci that showed unequal expression from both parental alleles, the *P. latipinna* and *P. mexicana* allelic expression for each locus of clonal *P. formosa* and interspecies F_1 progeny were used. We aimed to assess how similar or dissimilar Amazon molly individuals are in allelic expression and how it is compared to laboratory-produced F_1 individuals. We used χ^2 to test if the expression ratio of both parental alleles per locus are different in each individual. This method has been described in earlier studies (Birmingham et al. 2009; Heap et al. 2010). To test against the null hypothesis that parental alleles contribute equally to the gene expression using a χ^2 test, expected allelic expression was calculated under the null hypothesis as each allele accounts for 50% of total expression as $A^{mex}{}_{i\ expected} = cpm_i \times 50\%$, $A^{lat}{}_{i\ expected} = cpm_i \times 50\%$. $A^{mex}{}_{i\ expected}$, $A^{lat}{}_{i\ expected}$ and $A^{mex}{}_i$, $A^{lat}{}_i$ was used to form a contingency table for the χ^2 test. Genes that showed a χ^2 test with P-value < 0.05, and Log_2 (relative allelic expression) ≥ 0.27 or ≤ -0.27 (equivalent to 20% expression differences between the two parental alleles) were forwarded as genes with unequal parental allele expression.

Similar analyses were performed on mitochondria genome expression using *P. latipinna* (NCBI Nucleotide database [https://www.ncbi.nlm.nih.gov/nuccore] under accession number KT175511.1) and *P. mexicana* (accession number KT175512.1) mitochondria genome sequences as references.

Quantification of number of alleles

Following short sequencing read mapping to the *P. formosa* reference genome, alignment files were processed using SAMtools (V1.3.1) (Li and Durbin 2009; Li et al. 2009) to produce mpileup for each sample, followed by identification of variants using VarScan (V2.3.7; minimum coverage = 10, P-value < 0.05) (Koboldt et al. 2012). All genetic variant results of both *P. formosa* and F_1 hybrids were pooled together to quantify the number of alternative alleles. Owing to the hybrid genetic background of both cohorts, all genotyped loci are expected to be heterozygous. Per locus, we quantified numbers of alternative alleles within *P. formosa* and F_1 hybrids, respectively. Loci with unequal numbers of alternative alleles were forwarded for further analyses. The maximum numbers of alternative alleles for both *P. formosa* and F_1 hybrids are two (i.e., two alleles in *P. formosa* and three alleles in F_1, vice versa). These loci were mapped to the genome assembly to test if they are within a gene model.

Identification of genomic regions showing allele-specific expression in *P. formosa*

Allelic gene expression profiling classified an expressed gene as *P. latipinna* allele overexpressed, *P. latipinna* allele underexpressed, or equally expressed by both parental alleles. Because the genome-wide allelic expression showed higher expression from the *P. latipinna* alleles in *P. formosa,* we sought to answer if this was caused by *P. latipinna* allele-specific expression and to identify loci showing such expression patterns. For each genome contig, the numbers of genes showing *P. latipinna* allele overexpression were determined. Contigs with more than nine genes and those with more than 80% of the genes showing *P. latipinna* allele overexpression were forwarded as *P. latipinna*-biased-expressing loci.

Functional analyses

Functional analyses were performed using Ingenuity Pathway Analyses (IPA) that compare the input gene list to an internal knowledge base. The knowledge base documented gene-function and gene-signaling pathway activities. Overrepresentation of function or pathway, using the genome as background, was determined by Fisher's exact test ($P < 0.05$).

Data access

All raw and processed sequencing data generated in this study have been submitted to the NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE164222. Custom R scripts in normalizing and calculating allelic expression are available in Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

This work was supported by overhead funds from the Deutsche Forschungsgemeinschaft, and the National Institutes of Health, National Cancer Institute, R15-CA-223964, Office of Research Infrastructure Programs R24-OD-011120.

Author contributions: Y.L., D.B., W.C.W., R.B.W., and M.S. designed the study; D.B., J.O., and M.S. collected samples; D.B., J.O., and M.S. conducted experiments; Y.L., D.B., and M.S. performed data analyses; and Y.L., D.B., W.C.W., R.B.W., and M.S. drafted the manuscript.

References

Bell G. 2019. *The masterpiece of nature: the evolution and genetics of sexuality*. Routledge, Oxfordshire, UK.

Bierbach D, Laskowski KL, Wolf M. 2017. Behavioural individuality in clonal fish arises despite near-identical rearing conditions. *Nat Commun* **8:** 15361. doi:10.1038/ncomms15361

Birmingham A, Selfors LM, Forster T, Wrobel D, Kennedy CJ, Shanks E, Santoyo-Lopez J, Dunican DJ, Long A, Kelleher D, et al. 2009. Statistical methods for analysis of high-throughput RNA interference screens. *Nat Methods* **6:** 569–575. doi:10.1038/nmeth.1351

Doran C, Bierbach D, Laskowski KL. 2019. Familiarity increases aggressiveness among clonal fish. *Anim Behav* **148:** 153–159. doi:10.1016/j.anbehav.2018.12.013

Heap GA, Yang JH, Downes K, Healy BC, Hunt KA, Bockett N, Franke L, Dubois PC, Mein CA, Dobson RJ, et al. 2010. Genome-wide analysis of allelic expression imbalance in human primary cells by high-throughput transcriptome resequencing. *Hum Mol Genet* **19:** 122–134. doi:10.1093/hmg/ddp473

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14:** R36. doi:10.1186/gb-2013-14-4-r36

Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, Miller CA, Mardis ER, Ding L, Wilson RK. 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* **22:** 568–576. doi:10.1101/gr.129684.111

Lamatsch DK, Nanda I, Epplen JT, Schmid M, Schartl M. 2000. Unusual triploid males in a microchromosome-carrying clone of the Amazon molly, *Poecilia formosa*. *Cytogenet Cell Genet* **91:** 148–156. doi:10.1159/000056836

Lamatsch DK, Lampert KP, Fischer P, Geiger M, Schlupp I, Schartl M. 2009. Diploid Amazon mollies (*Poecilia formosa*) show a higher fitness than triploids in clonal competition experiments. *Evol Ecol* **23:** 687–697. doi:10.1007/s10682-008-9264-2

Lampert KP, Schartl M. 2008. The origin and evolution of a unisexual hybrid: *Poecilia formosa*. *Philos Trans R Soc Lond B Biol Sci* **363:** 2901–2909. doi:10.1098/rstb.2008.0040

Lampert KP, Lamatsch DK, Fischer P, Epplen JT, Nanda I, Schmid M, Schartl M. 2007. Automictic reproduction in interspecific hybrids of poeciliid fish. *Curr Biol* **17:** 1948–1953. doi:10.1016/j.cub.2007.09.064

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25. doi:10.1186/gb-2009-10-3-r25

Laskowski KL, Wolf M, Bierbach D. 2016. The making of winners (and losers): how early dominance interactions determine adult social structure in a clonal fish. *Proc Biol Sci* **283:** 20160183. doi:10.1098/rspb.2016.0183

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25:** 1754–1760. doi:10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079. doi:10.1093/bioinformatics/btp352

Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30:** 923–930. doi:10.1093/bioinformatics/btt656

Lively CM, Morran LT. 2014. The ecology of sexual reproduction. *J Evol Biol* **27:** 1292–1303. doi:10.1111/jeb.12354

Loewe L, Lamatsch DK. 2008. Quantifying the threat of extinction from Muller's ratchet in the diploid Amazon molly (*Poecilia formosa*). *BMC Evol Biol* **8:** 88. doi:10.1186/1471-2148-8-88

Lu Y, Bowswell M, Bowswell W, Yang K, Schartl M, Walter RB. 2015. Molecular genetic response of *Xiphophorus maculatus–X. couchianus* interspecies hybrid skin to UVB exposure. *Comp Biochem Physiol C Toxicol Pharmacol* **178:** 86–92. doi:10.1016/j.cbpc.2015.07.011

Lynch M, Gabriel W. 1990. Mutation load and the survival of small populations. *Evolution* **44:** 1725–1737. doi:10.1111/j.1558-5646.1990.tb05244.x

Lynch M, Conery J, Bürger R. 1995. Mutational meltdowns in sexual populations. *Evolution* **49:** 1067–1080. doi:10.1111/j.1558-5646.1995.tb04434.x

Maynard Smith J. 1978. *The evolution of sex*. Cambridge University Press, Cambridge.

Neiman M, Hehman G, Miller JT, Logsdon JM Jr, Taylor DR. 2010. Accelerated mutation accumulation in asexual lineages of a freshwater snail. *Mol Biol Evol* **27:** 954–963. doi:10.1093/molbev/msp300

Schartl M, Schlupp I, Schartl A, Meyer MK, Nanda I, Schmid M, Epplen JT, Parzefall J. 1991. On the stability of dispensable constituents of the eukaryotic genome: stability of coding sequences versus truly hypervariable sequences in a clonal vertebrate, the Amazon molly, *Poecilia formosa*. *Proc Natl Acad Sci* **88:** 8759–8763. doi:10.1073/pnas.88.19.8759

Schartl M, Wilde B, Schlupp I, Parzefall J. 1995. Evolutionary origin of a parthenoform, the Amazon molly *Poecilia formosa*, on the basis of a molecular genealogy. *Evolution* **49:** 827–835. doi:10.1111/j.1558-5646.1995.tb02319.x

Shen Y, Garcia T, Pabuwal V, Boswell M, Pasquali A, Beldorth I, Warren W, Schartl M, Cresko WA, Walter RB. 2013. Alternative strategies for development of a reference transcriptome for quantification of allele specific expression in organisms having sparse genomic resources. *Comp Biochem Physiol Part D Genomics Proteomics* **8:** 11–16. doi:10.1016/j.cbd.2012.10.006

Stöck M, Lampert KP, Möller D, Schlupp I, Schartl M. 2010. Monophyletic origin of multiple clonal lineages in an asexual fish (*Poecilia formosa*). *Mol Ecol* **19:** 5204–5215. doi:10.1111/j.1365-294X.2010.04869.x

Van Valen L. 1973. A new evolutionary law. *Evol Theory* **1:** 1–30.

Vrijenhoek R. 1994. Unisexual fish: model systems for studying ecology and evolution. *Annu Rev Ecol* **25:** 71–96. doi:10.1146/annurev.es.25.110194.000443

Warren WC, García-Pírez R, Xu S, Lampert KP, Chalopin D, Stöck M, Loewe L, Lu Y, Kuderna L, Minx P, et al. 2018. Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nat Ecol Evol* **2:** 669–679. doi:10.1038/s41559-018-0473-y

Received July 15, 2020; accepted in revised form January 12, 2021.