Parallel Genomic Changes Drive Repeated Evolution of Placentas in Live-Bearing Fish

Henri van Kruistum,*[,1,2] Reindert Nijland,[3] David N. Reznick,[4] Martien A.M. Groenen,[1] Hendrik-Jan Megens,[†,1,5] and Bart J.A. Pollux*[†,2]

[1]Animal Breeding and Genomics Group, Wageningen University, Wageningen, The Netherlands
[2]Experimental Zoology Group, Wageningen University, Wageningen, The Netherlands
[3]Marine Animal Ecology Group, Wageningen University, Wageningen, The Netherlands
[4]Department of Biology, University of California, Riverside, CA, USA
[5]Aquaculture and Fisheries Group, Wageningen University, Wageningen, The Netherlands
[†]Shared last author.
***Corresponding authors:** E-mails: henri.vankruistum@wur.nl; bart.pollux@wur.nl.
Associate editor: Emma Teeling

Abstract

The evolutionary origin of complex organs challenges empirical study because most organs evolved hundreds of millions of years ago. The placenta of live-bearing fish in the family Poeciliidae represents a unique opportunity to study the evolutionary origin of complex organs, because in this family a placenta evolved at least nine times independently. It is currently unknown whether this repeated evolution is accompanied by similar, repeated, genomic changes in placental species. Here, we compare whole genomes of 26 poeciliid species representing six out of nine independent origins of placentation. Evolutionary rate analysis revealed that the evolution of the placenta coincides with convergent shifts in the evolutionary rate of 78 protein-coding genes, mainly observed in transporter- and vesicle-located genes. Furthermore, differences in sequence conservation showed that placental evolution coincided with similar changes in 76 noncoding regulatory elements, occurring primarily around genes that regulate development. The unexpected high occurrence of GATA simple repeats in the regulatory elements suggests an important function for GATA repeats in developmental gene regulation. The distinction in molecular evolution observed, with protein-coding parallel changes more often found in metabolic and structural pathways, compared with regulatory change more frequently found in developmental pathways, offers a compelling model for complex trait evolution in general: changing the regulation of otherwise highly conserved developmental genes may allow for the evolution of complex traits.

Key words: **molecular evolution, comparative genomics, sequencing, placenta, Poeciliidae.**

Introduction

The emergence of complex organs is one of the most significant phenomena in the evolution of multicellular organisms. Characterizing the origin of this complexity is a challenge because we are most often confronted with the end-products of evolution as they appear in currently living organisms, with little knowledge on intermediate stages. Ultimately, the development of these organs is encoded in the genome. This same genome, however, poses a puzzle: while there is remarkable diversity in vertebrate morphology, the genes that regulate morphological development tend to be highly conserved (Gaunt 2002; Hoegg and Meyer 2005).

Over the past decades, developments in genome science have unraveled details in how cell differentiation, cell signaling, and cell migration shape organisms and their organs during ontogeny. As the developmental pathways leading to specific organismal traits are better understood, it is becoming clear that organs, especially in vertebrates, are not only highly conserved in morphology and physiology, but also in developmental pathways (Farrell et al. 2018). The deep conservation in developmental pathways deployed once organs have emerged in evolution, however, does question the genomic basis of convergence: if structures, such as organs, develop in parallel in another animal group, is that mirrored in convergence in underlying molecular pathways, or is the parallel evolution only superficial, based on different developmental triggers?

Studies of genomic changes associated with convergent phenotypic evolution have identified genomic changes in physiological and structural genes common to convergent lineages (Foote et al. 2015; Chikina et al. 2016). However, this alignment of convergent morphology with convergent changes in the genome does not include the developmental genes that govern morphology. The absence of the expected association between developmental genes and phenotypic evolution may be because evolution has been assessed via changes in amino acid sequences or copy number, whereas another cause for morphological evolution could lie in

Article

changes in the spatiotemporal expression patterns of developmental gene expression (Carroll 2008; Levin et al. 2016). Such changes in expression are instead controlled by the elements that regulate developmental genes.

An excellent model to study convergent evolution of a complex trait is found in the live-bearing fish family Poeciliidae. The Poeciliidae are a family of livebearing fish consisting of around 275 species (Parenti 1981; Van Der Laan et al. 2014). In this family, a placenta has evolved independently at least nine times from a nonplacental ancestor (Pollux et al. 2009; Furness et al. 2019). The evolution of the placenta in this family coincides with a shift from pre- to postfertilization nutrient provisioning to the offspring. Nonplacental or lecithotrophic species supply nutrients to their offspring before fertilization in form of egg yolk. Placental or matrotrophic species supply only a very small amount of yolk, with the majority of nutrients being supplied after fertilization by means of their placenta. Furthermore, species that have a placenta of intermediate complexity are also present in this family, where nutrients are provided both before and after fertilization (Jollie and Jollie 1964; Grove and Wourms 1994; Kwan et al. 2015).

On a morphological level, differences between placental and nonplacental poeciliid species are found in the follicular tissue surrounding the embryos. In nonplacental species, a thin follicle can be observed surrounding the embryos. In placental species however, this tissue is thicker and has extensive folds (Jollie and Jollie 1964; Grove and Wourms 1991, 1994; Kwan et al. 2015). Additionally, microvilli and vesicles can be observed in the follicular tissue of placental species. The variation in placental complexity can be characterized by quantifying the degree of postfertilization nutrient provisioning to offspring with the Matrotrophy Index (MI) (Reznick et al. 2002; Pollux et al. 2009). The MI is defined as the embryo mass at birth divided by the egg mass at fertilization, and is used as a proxy for placental complexity (Pollux et al. 2009). Morphological studies have shown that the MI correlates well with the placental complexity in multiple poeciliid species (Jollie and Jollie 1964; Grove and Wourms 1994; Kwan et al. 2015).

The placenta in the fish family Poeciliidae allows for a genomic study of complex trait evolution because of several reasons. First, the integration of estimates of MI with a DNA-based phylogeny for the family suggests that the degree of placentation can evolve very quickly, with estimations based on molecular data suggesting that a placenta can evolve in as little as 0.75 My (Reznick et al. 2002). Placenta evolution in the Poeciliidae is a fairly recent event, with the most recent placenta evolutions in the genus *Poeciliopsis* being estimated to have happened <5 Ma (Reznick et al. 2017). Second, in some cases, species with placentas have closely related sister species without a placenta, which allows for comparisons between species that differ in placentation (Reznick et al. 2002; Pollux et al. 2014), but have very similar genomes in general (van Kruistum et al. 2020). Third, the multiple independent evolutions of the poeciliid placenta allows investigating genomic changes between multiple instances of placenta evolution that happened in parallel, which will increase the reliability of our results.

Previous studies have identified species-specific genomic changes in pathways involved in metabolism and development in individual placental poeciliids (O'Neill et al. 2007; Jue et al. 2018; van Kruistum et al. 2019, 2020). However, these studies did not compare multiple instances of placenta evolution on a genome-wide scale, include closely related nonplacental species, or consider noncoding regions of the genome. The similar physiological and morphological details of placentation, and the similarities in intermediary paths towards placentation, suggest that throughout the Poeciliidae similar pathways are at the basis of conferring placentation in all these species. Moreover, the intermediate stages suggest a quantitative nature of the trait, meaning it is not expected that a single "switch" gene exists that regulates this trait. Here, we take advantage of the multiple independent placenta evolutions in our study system to test whether the convergent morphological evolution of the placenta is reflected in convergent molecular evolution on a genomic level.

In this study, we use both publicly available and new genome assemblies to construct a large-scale comparative framework of genome evolution in the Poeciliidae, consisting of 26 species, of which eight species have a placenta (fig. 1). These eight placental species include six independent origins of placentation. Although it may seem from our data that independent losses of placentation are also a plausible option in the genus *Poeciliopsis*, studies that include more species from this genus show that the placental species we include do in fact represent three independent instances of placenta evolution (Reznick et al. 2002, 2017). By comparing the genomes of these 26 species, we are able to investigate 1) which genomic changes are associated with placenta evolution in the Poeciliidae, 2) whether similar genomic changes occur in each of the six origins of placentation, and 3) whether mutation in coding or noncoding genomic regions are most important for placenta evolution. Our study will provide new insights in genome evolution during the evolution of complex traits.

Results

Reconstructing a Molecular Phylogeny

We reconstructed a Maximum Likelihood phylogeny of the Poeciliidae using a concatenated alignment consisting of 1) whole mitochondrial genomes acquired from the whole genome sequencing data and 2) the complete coding sequence from 1,010 nuclear genes (fig. 1). Because of the difference in nuclear divergence between mitochondrial and nuclear DNA, these parts of the data were partitioned separately (supplementary fig. 1, Supplementary Material online). For the same reason, the three codon positions were also partitioned separately for both nuclear and mitochondrial genes. The resulting phylogeny was used as the basis for subsequent comparative analyses.

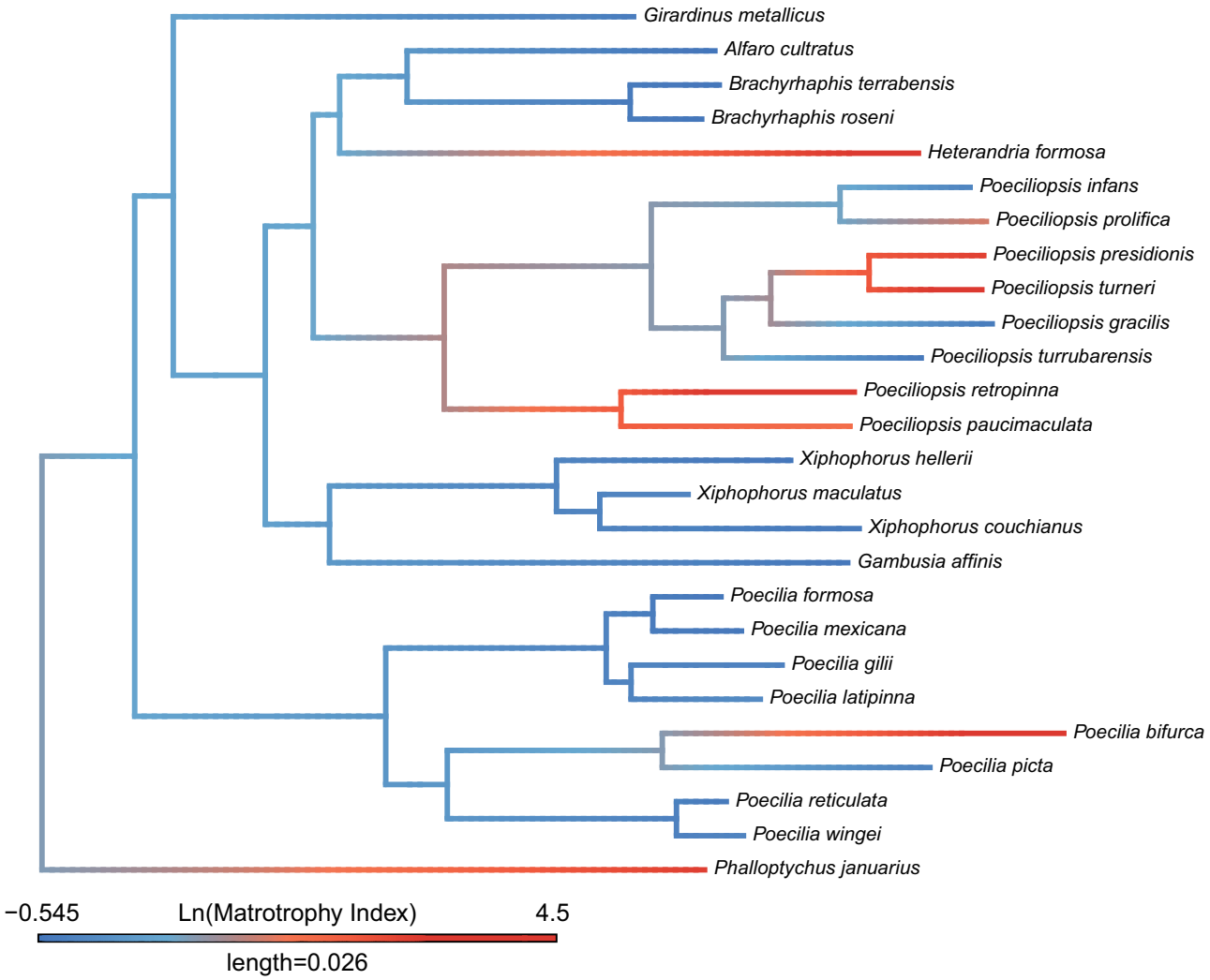

Fig. 1. Molecular phylogeny of investigated species. The colors of the branches represent the natural logarithm of the Matrotrophy Index (MI), either the observed MI for terminal branches or the estimated MI as determined by ancestral trait reconstruction.

Convergent Shifts in Evolutionary Rate for Genes in Placental Poeciliid Species

To test for the presence of convergent genomic changes in protein-coding regions, we applied an evolutionary rate analysis based on the method used by Chikina et al. (2016), modified to test for correlation between evolutionary rate and the continuous variable MI, instead of testing for a significant difference between two discrete classes. This analysis tests whether the relative rate of gene evolution is correlated with the MI across the phylogeny. The relative evolutionary rate is inferred from the number of amino acid substitutions across each branch of the phylogeny for a certain protein-coding gene and is normalized for branch- as well as gene-specific evolutionary rates (see Materials and Methods). These relative rates are then tested for correlation with observed or estimated MI values using Spearman's correlation test. Both the observed and estimated MI values can be found in supplementary table 1, Supplementary Material online. We created a null distribution by generating simulated data sets in which the MI values were randomly assigned to individual branches (orange in fig. 2). We infer genome-wide convergence in the rate of evolution by comparing the distribution of *P* values for the null and observed distributions. Analysis of the simulated data revealed a uniform *P*-value distribution, as expected, when no convergent evolution is present. However, the results for placental branches were heavily skewed towards lower *P* values (fig. 2), indicating that more genes show a good correlation between MI and evolutionary rate than would be expected by chance ($P = 1.36e\text{-}61$, Kolmogorov–Smirnov test).

After correcting for multiple testing, we identified 78 genes showing a significantly higher or lower evolutionary rate correlated with MI (q-value <0.1, supplementary table 2, Supplementary Material online). About 76 of these correlations were positive, indicating a higher evolutionary rate in placental species, and only two genes showed a negative correlation. The near-absence of slowly evolving genes in placental species may be due to the relatively short branches in the phylogeny: if the average number of mutations per branch is already low, the power to identify genes evolving at lower than average rates may be limited. Because of this, we focused on genes showing accelerated evolution in placental lineages for subsequent analyses.

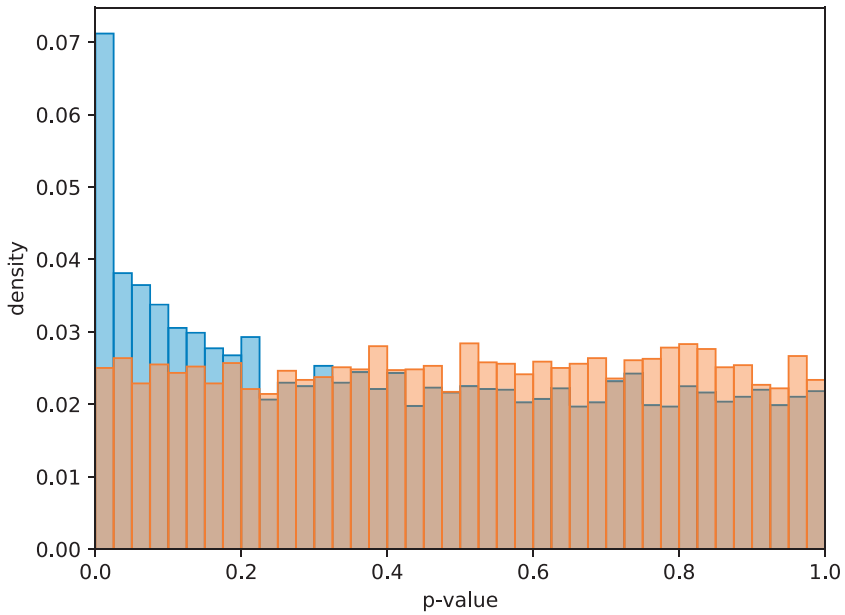

Fig. 2. In blue, distribution of *P* values for the test that the spearman correlation between MI and relative evolutionary rate of each branch is significantly different from zero. In orange, the same test, but MI values are shuffled randomly across phylogenetic branches.

GO enrichment analysis of accelerated genes revealed an enrichment of genes involved in metabolite transport. Overrepresented GO terms included "carboxylic acid transport" and "base amino acid transport" (supplementary table 3, Supplementary Material online). For cellular components, the gene set was enriched in vesicle-located genes.

The *slc7a7* gene is an example of a transporter gene showing evidence of accelerated evolution in placental poeciliids. This gene codes for an amino acid transporter and is involved in nitric oxide synthesis in human umbilical vein (Arancibia-Garavilla et al. 2003). Almost all placental poeciliids show a faster evolutionary rate than expected for this gene, with *Heterandria formosa* and *Poeciliopsis prolifica* showing exceptionally high evolutionary rates (fig. 3). These results suggest that the evolution of the placenta in the Poeciliidae is associated with consistent changes in nutrient transport systems. Four more examples of genes showing evidence of accelerated evolution in placental poeciliids are shown in supplementary figure 3, Supplementary Material online.

To gain more insight in the expression of our candidate genes in placental tissue, we used a previously published RNA-seq data set of placental tissue of *Poeciliopsis retropinna* to see whether our candidate genes are expressed within placental tissue of this species (Guernsey et al. 2020). Using this data set, we could confirm expression of 44 of 78 candidate genes (supplementary table 2, Supplementary Material online). All amino acid transporters in our candidate gene list show expression in placental tissue of *P. retropinna*.

Convergent Shifts in Evolutionary Rate Are Not Due to Positive Selection

We tested the hypothesis that the accelerated evolutionary rate of our candidate genes was due to positive selection by quantifying the synonymous to nonsynonymous mutation ratio (d*N*/d*S*, see Materials and Methods). After correction for multiple testing, five out of 76 placenta-accelerated genes show evidence of positive selection in at least one branch leading to a placental species in the form of an elevated d*N*/d*S* mutation ratio (supplementary table 2, Supplementary Material online). This is not a significantly higher proportion compared with the genome-wide proportion of genes evolving under positive selection in at least one of the placental branches (665 out of 14,468 genes, $P = 0.41$, Fisher's exact test). Additionally, evidence for positive selection was only apparent in one placental branch in the phylogeny for these five genes. A comparison of d*N* and d*S* between phylogenetic branches leading to placental species and other branches shows that for the majority of placenta-accelerated genes, both d*N* and d*S* are higher for placental branches than for other branches (supplementary table 2, Supplementary Material online). This leads to an increased evolutionary rate while not necessarily increasing the d*N*/d*S* ratio.

Another hypothesis is that positive selection manifests on other genes than those for which we observe an accelerated evolutionary rate among placental species, but is still convergent among placental species. To test this hypothesis, we tested all the 14,468 orthologous gene sets that were previously identified for positive selection using a branch-site model (see Materials and Methods), with the phylogenetic branches leading to six highly placental poeciliids as foreground branches. As a control, we performed the same analysis using six nonplacental species as foreground branches for which we do not expect any convergent molecular evolution. If placental species have more similar genes evolving under positive selection, the *P*-value distribution would be skewed towards lower *P* values compared with the control group. However, the placental group did not show an excess of low *P* values compared with the control group

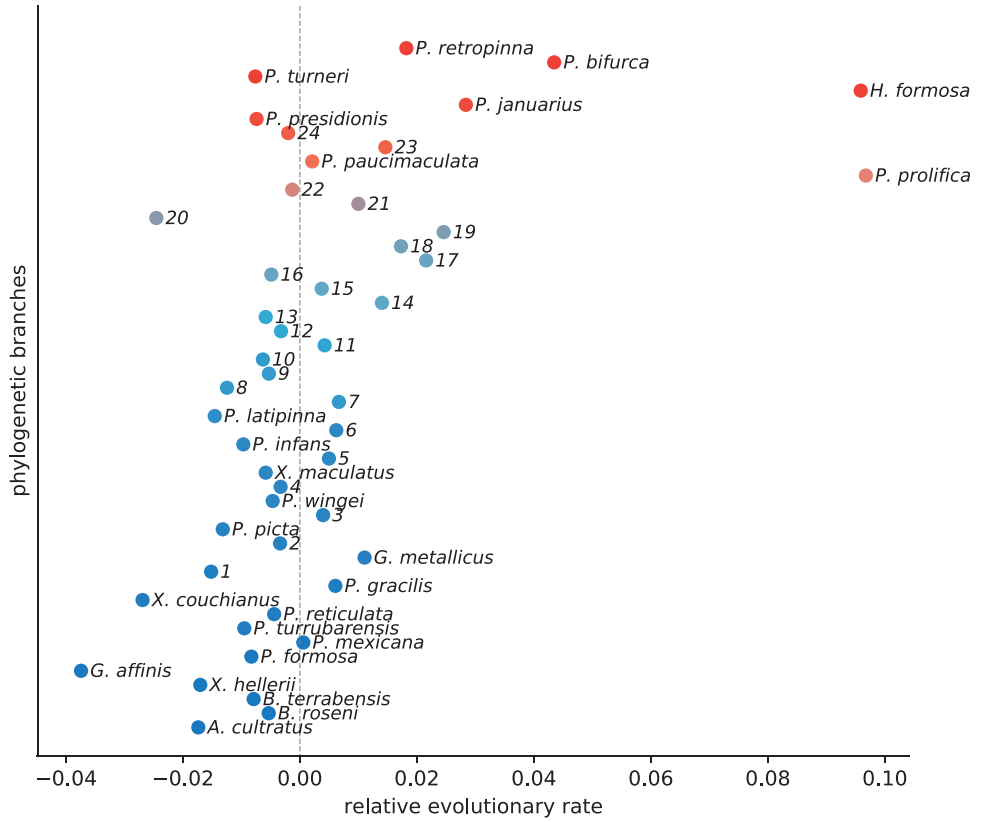

Fig. 3. Relative evolutionary rate of the slc7a7 gene in all branches of the investigated phylogeny, sorted from high to low MI. The relative rate is defined as the deviation from the expected evolutionary rate, given branch- and gene-specific average rates. Points labeled 1–24 represent ancestral branches in the phylogeny, which have estimated MI values based on ancestral trade reconstruction. See supplementary figure 2, Supplementary Material online, for the poeciliid phylogeny with labeled ancestral branches corresponding to these points.

(supplementary fig. 4, Supplementary Material online), indicating that convergent placenta evolution does not manifest in a molecular signal that can be picked up by this test for positive selection. It seems that the genes evolving under positive selection show no similarity between different placental species, whereas changes that are captured by the evolutionary rate analysis do show a clear sign of convergence.

No Excess of Convergent Amino Acid Substitutions in Placental Species Compared with Control Groups

If certain amino acid substitutions are necessary for placenta evolution, these substitutions would have to be observed across all phylogenetic branches where a placenta has evolved. We tested whether we could observe an excess of these convergent substitutions among six placental branches in the phylogeny in the 14,468 orthologous gene sets that we have identified across the Poeciliidae. For this, we compared the observed sequences of placental species with the inferred ancestral sequences. A convergent event was defined as an amino acid substitution at the same position in the protein for at least four placental species.

In total, we found 117 genes with one or more convergent amino acid substitutions in placental species. As a control, we performed the same analysis on 100 random combinations of six branches in the phylogeny, for which we do not expect convergent molecular evolution. We did not find an excess of convergent amino acid substitutions in placental species, compared with the control distribution (fig. 4). This shows that although convergent amino acid substitutions happen somewhat frequently, convergent amino acid substitutions linked to convergent phenotypic evolution do not happen at a frequency that allows it to be detected in these species, if it occurs at all.

Convergent Differences in the Conservation of Noncoding Elements between Placental and Nonplacental Poeciliid Species

To also test for convergent changes outside protein-coding regions, we generated two multigenome alignments: one with six placental poeciliids (MI $>$ 10), and one with six nonplacental poeciliids (MI $<$ 1). To gain insight in more ancient conservation in our study system, we included a third alignment with five teleost fish species, four of which are outside of the family Poeciliidae (see Materials and Methods for species). For all of these alignments, conserved elements were called independently using PhastCons (Hubisz et al. 2011). All of these conserved element predictions were aligned to the *P. retropinna* genome, allowing for a direct comparison of conserved elements in the different groups. Then, we identified conserved elements for placental poeciliids that showed no trace of conservation in the nonplacental species, as well as conserved elements for nonplacental poeciliids that

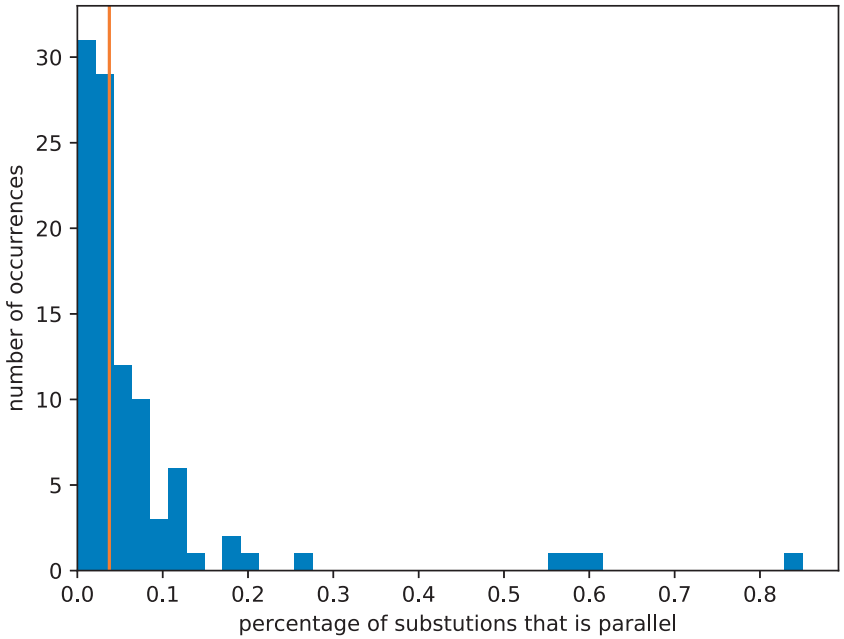

FIG. 4. Percentage of convergent amino acid substitutions compared with the total number of substitutions in hundred random sets of six branches in the poeciliid phylogeny. Orange line: percentage of convergent amino acid substitutions compared with the total number of substitutions in the branches leading to six extensive matrotrophs in the Poeciliidae. For each test for convergent substitutions, we examined 14,468 genes.

showed no trace of conservation in placental species. This means that for these elements, genomic sequences are identical or almost identical across all species in one group (average nucleotide diversity of 0.11 substitutions per site), whereas for the other group, there are a substantial number of mutations in these regions (average nucleotide diversity of 0.43 substitutions per site). This does not necessarily mean that these mutations have to occur in all species for the nonconserved group: a large number of mutations in the majority of the species will still lead to a low conservation score, even if the remaining species show high similarity to the conserved group. In total, we found 76 of these Differentially Conserved Elements (DCEs). Fifty of these elements were more conserved in the placental species, and 26 were more conserved in the nonplacental species. Our DCEs were highly enriched for simple sequence repeats (SSRs), especially GATA repeats: out of the 76 identified elements, 53 contained an SSR, and in 31 cases this was a GATA repeat. The proportion of DCEs containing an SSR is significantly higher than expected given the proportion of conserved elements containing an SSR across the whole genome ($P = 1.26e\text{-}85$, Fisher's exact test).

To infer which genes these DCEs potentially regulate, we extracted each gene from the *P. retropinna* genome that lies directly downstream from a DCE (maximum distance 50 kb). Additionally, we extracted genes for which a DCE lies in an intron or within the UTR of this gene. This resulted in a set of 85 genes that are potentially differently regulated in placental poeciliids compared with their nonplacental counterparts (supplementary table 5, Supplementary Material online). The Gene Ontology (GO) enrichment analysis revealed that this gene set was enriched for genes involved in several developmental processes, such as "anatomical structure development" or "system development" (supplementary table 6, Supplementary Material online). By contrast, a control analysis consisting of the same workflow to test for differences between two sets of nonplacental poeciliid species resulted in only 22 potentially differently regulated genes between these two groups of species, for which no significantly enriched GO terms were found (supplementary table 7, Supplementary Material online). To test for a possible influence of different conserved element density across the genome, we also tested 25 sets of 76 random conserved elements for GO term enrichments, using the same strategy. This yielded a few enrichments for some of these sets, but none of them had such a clear overrepresentation of developmental genes as our candidate gene set (supplementary table 8, Supplementary Material online).

A notable example of a difference in gene regulation between placental and nonplacental species is found within a cluster of four ultraconserved genes (*foxa2*, *pax1*, *nkx2-2a*, *nkx2-4*) that play a role in embryonic development. Synteny of this gene cluster is conserved across both teleost fish and mammals. A DCE was found between the *pax1* and *nkx2-2a* genes, consisting of a GATA simple repeat that was found in all placental species, but is not conserved in nonplacental species (fig. 5). The absence in conservation of this element in both the nonplacental species, as well as the ancient conservation track suggests that this element has emerged in all placental species, although the partial presence of this element in some nonplacental species indicates that this element did probably not emerge in the individual branches leading to placental species, but rather emerged deeper in the Poeciliid phylogeny and subsequently acquired mutations

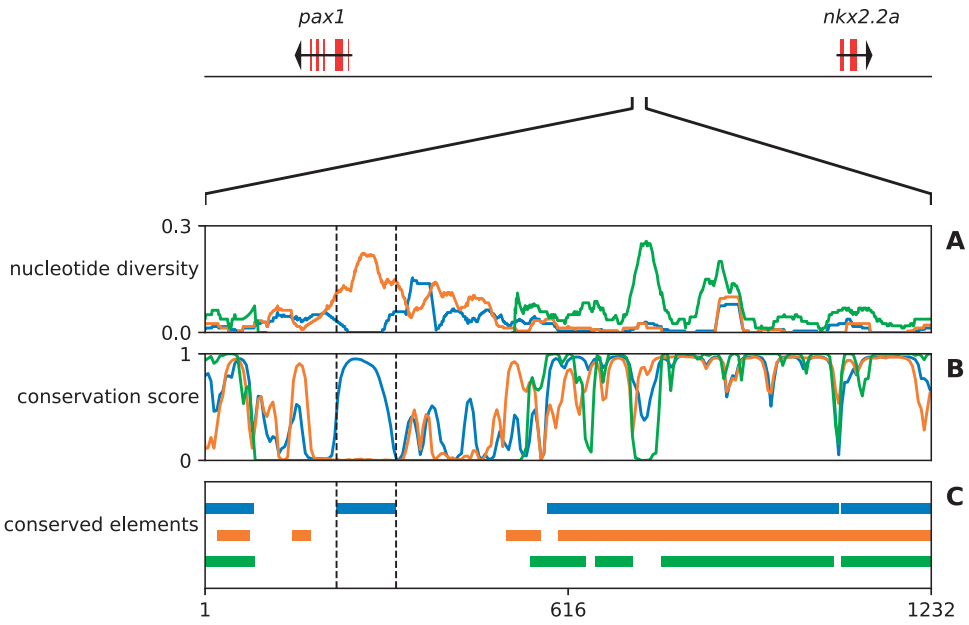

Fig. 5. A placenta-specific boundary element near an ancient conserved element. Color codes: blue: placental poeciliids alignment; orange: nonplacental poeciliids alignment; green: "ancient" teleost alignment. (*A*) Average nucleotide diversity for each alignment, 20-bp sliding window. (*B*) Per-base conservation scores, representing the estimated likelihood of the base belonging to a conserved element according to PhastCons. (*C*) Conserved elements predicted by PhastCons.

only in nonplacental species. It has been shown that GATA repeat elements act as regulatory boundaries in both human and fruit fly, and a similar mechanism seems to be present in the Poeciliidae (Kumar et al. 2013).

Discussion

Our results support the hypothesis that similar genomic changes underlie the repeated evolutionary origins of the placenta in the Poeciliidae. We find convergence in evolutionary rate shifts of protein-coding genes for placental species, as well as in differential conservation of noncoding elements for placental species as compared with their nonplacental relatives. In both cases, we find a significantly higher incidence of these events compared with control analyses.

Our phylogenetic analysis based on a maximum likelihood strategy revealed a phylogeny that is identical to previous work (Pollux et al. 2014). However, some of the basal branches showed suboptimal bootstrapping support. An alternative analysis based on a coalescent method resulted in a tree topology that is somewhat different at these basal nodes (supplementary fig. 5, Supplementary Material online). It seems that the earliest splits in the poeciliid phylogeny are hard to solve. However, running the downstream analyses on 100 random bootstrap trees showed that these slight differences in topology have very little effect on the results of the evolutionary rate analysis.

Many of the genes showing accelerated evolution in placental poeciliids are involved in vesicle functioning. This result is in agreement with a recent study showing that the placenta in the genus *Poeciliopsis* is based on a secretory system (Guernsey et al. 2020), a mechanism that likely applies to other genera in the Poeciliidae given the morphological similarities in placenta structure and development. The presence of several plasma membrane transporters in our accelerated gene set, such as the amino acid transporters *slc3a2*, *slc7a7*, and *slc38a3*, suggests that additionally some nutrients are secreted through transporter proteins instead of vesicles. Investigation of a previously published RNA-seq data set of placental tissue of *P. retropinna* (Guernsey et al. 2020) shows that these genes are indeed expressed in placental tissue of this species (supplementary table 2, Supplementary Material online).

Despite the clear signature of convergence in the evolutionary rate of these genes, we did not find more genes evolving under positive selection in placental species than would be expected by chance. Prior studies employing an evolutionary rate approach to find convergent evolution also found only a small proportion of their candidate genes to evolve under positive selection (Chikina et al. 2016; Partha et al. 2017). An increase in evolutionary rate without evidence for positive selection is usually attributed to relaxation from evolutionary constraint. In our case, this seems unlikely, as many of our candidate genes have vital functions in all fish species, and placenta evolution would presumably not affect this. A possible explanation for this discrepancy could be low power of tests for positive selection when selection is weak and phylogenetic branches are relatively short, as is the case for our phylogeny (Gharib and Robinson-Rechavi 2013). Alternatively, the observed changes in amino acid residues could be a consequence of positive selection acting outside of the coding region, with the increase in evolutionary rate inside the coding region being due to linkage disequilibrium rather than positive selection on certain amino acid residues.

Further research, for instance on population genomic data, may be able to detect these selective forces.

The observed changes in the size and location of conserved elements between placental and nonplacental poeciliids suggest convergent regulatory change in these species. However, the conservation differences between placental and nonplacental species in our DCEs are not of the nature that the element is completely absent in the nonconserved group. Often, the element is still present or partially present in some of the species of the nonconserved group, but has acquired several mutations for others. This indicates that these elements likely did not emerge in the phylogenetic branch leading to highly placental species, but deeper in the phylogeny of the Poeciliidae. Then, the element was only conserved in either placental or nonplacental species, while acquiring mutations in the other group. This suggests that some preconditions for placenta evolution may have already taken place before the evolution of a complex placenta in this family, which is not surprising given the continuous nature of the trait.

The overrepresentation of GATA repeats in our DCEs suggests a role of these elements in the gene regulation of these species. GATA repeats are known to function as enhancer blockers in both human and fruit fly (Kumar et al. 2013), can modulate promotor activity (Krishnan et al. 2017), and are associated with chromatin structure (Subramanian et al. 2003). GATA elements appear to have a specific length distribution, with GATA10-12 being the most abundant in humans (Kumar et al. 2010). In the Poeciliidae, we generally observe a prevalence of GATA20-40, longer than reported for other vertebrates (Kumar et al. 2010) (supplementary fig. 6, Supplementary Material online). The length of these repeats further suggests a functional significance, as without selection maintaining their length these would be highly unstable (Kim and Mirkin 2013). Indeed, other four-base SSRs do not show a preference for longer repeats (supplementary fig. 7, Supplementary Material online). Besides this apparent functional significance, SSRs are exceptionally variable, having mutation rates that are orders of magnitude higher than nonrepetitive DNA (Gemayel et al. 2012). Because of this combination of having a functional role and being highly variable, GATA repeats could provide a potential mechanism for placenta evolution in the Poeciliidae by changing the regulation of developmental genes.

Finally, our results show that the evolution of the placenta in the Poeciliidae is accompanied by changes in both protein-coding and regulatory regions, suggesting that genomic changes in both categories are important for complex trait evolution. Notably, mutations in these two categories seem to be associated with different biological processes: protein-coding genes that show an evolutionary rate that correlates well with MI are mainly involved in metabolism and transport, whereas differentially conserved noncoding elements are mostly associated with development. This observed duality seems logical given that the amino acid sequences of developmental genes are usually highly conserved because they often act on many targets, making changes in the ensuing protein likely harmful due to pleiotropic effects. Although our study focuses on placenta evolution, the paradigm of protein-coding change in metabolic genes and regulatory change around developmental genes has been observed before and may be applicable to complex trait evolution in general (Partha et al. 2017; Sackton et al. 2019). As the genomes of more species become available this hypothesis can readily be tested in other evolutionary models.

Materials and Methods

Genome Assemblies

About 15 genomes were assembled for this study, and 12 publicly available genome assemblies were included (supplementary table 9, Supplementary Material online). Short-read assemblies were assembled from 30 to 50× coverage of 150-bp paired-end Illumina reads using SPAdes 3.13.0 (Bankevich et al. 2012) with default settings, using the assembly generated with a k-mer size of 77. This is the default k-mer size for 150-bp reads. Following assembly, contigs corresponding to heterozygous sections of the genome were removed using redundans v0.14a (Pryszcz and Gabaldón 2016), using settings –usebwa and –nogapclosing. For the *Phalloptychus januarius* assembly, the genome was assembled from 20× coverage of Oxford Nanopore long-reads (read N50 8.7 kb) using Flye version 2.5 (Kolmogorov et al. 2019) with default settings and setting the estimated genome size to 600 Mb. After assembly, the bases were polished by mapping 30× coverage of 150-bp paired-end Illumina reads to the assembly using BWA mem 0.7.17 (Li and Durbin 2009), before consensus calling with the wtdbg 2.5 consensus module (Ruan and Li 2019) with default settings.

Collecting Orthologous Genes

For assemblies of reference quality (see supplementary table 8, Supplementary Material online), predicted protein sets were collected based on their annotations and analyzed for orthology using ProteinOrtho v5.16b (Lechner et al. 2011) using default settings and blastp as the used program for alignment. Genes that displayed 1:1 orthology across all species were used for further analysis. Genes that displayed 1:1 orthology but were missing in one species were also used. This resulted in a set of 15,305 orthologous genes. Subsequently, the coding sequences of these genes were recovered from all short-read assemblies by aligning the coding sequence of the closest relative with a reference-quality genome to the assembly using exonerate version 2.2.0 (Slater and Birney 2005), using the cdna2genome model. Considering a 1:1 orthology in 12 reference genomes, a full-length match of the coding sequence in a single contig of the respective short-read assembly was assumed to be the 1:1 orthologous gene in this species too. Sequences with a premature stop codon were removed from the database. We continued analysis on genes of which we could recover the full coding sequence in at least three placental species, which was the case for 14,468 genes.

Construction of a Molecular Phylogeny

For the construction of a molecular phylogeny, we reconstructed the complete mitochondrial genome of all

investigated species using MITObim 1.9.1 (Hahn et al. 2013) using settings –mismatch 1, -start 1, and -end 30, using the published mitochondrion of *Poecilia reticulata* as reference. In addition, we made codon alignments of all orthologous genes recovered from all investigated species, as well as the non-poeciliid *Oryzias latipes* (1,010 genes). Columns with gaps were removed from this alignment using trimAl v1.4 (Capella-Gutiérrez et al. 2009), using the -nogaps flag. The resulting alignments were concatenated into a "supermatrix" alignment with a length of about 1.1 Mb. With this alignment, the phylogeny was reconstructed using RAxML 8.2.9 with the GTR+GAMMA model(Stamatakis 2014). *O. latipes* was used as an outgroup to root the tree. The RAxML analysis was done with eight partitions: three partitions for the three codon positions in both the mitochondrial coding sequence and the nuclear genes, one partition for mitochondrial noncoding RNA (tRNA and rRNA), and one partition for mitochondrial noncoding DNA (D-loop and some very small segments).

As an alternative analysis, we generated a phylogeny using a coalescent method. We made gene trees for each of the 1,010 previously mentioned genes, as well as the mitochondrial genes using RAxML 8.2.9 with the GTR+GAMMA model (Stamatakis 2014). Then, we combined these gene trees into a species tree using ASTRAL v5.7.1, using default settings (Zhang et al. 2018). As ASTRAL does not estimate terminal branch lengths, branch lengths for the resulting phylogeny were estimated using RAxML 8.2.9 using option "-f e" for branch length estimation for a given topology.

Evolutionary Rate Analysis

Evolutionary rate analysis was performed as in Chikina et al. (2016), slightly modified to test for a correlation with MI instead of a difference between two discrete classes. Amino acid alignments of previously identified orthologous genes were made using mafft v7.402 (Katoh et al. 2002). For each alignment, branch lengths were estimated for each branch across the reconstructed phylogeny using the AAML program of the PAML package (Yang 2007), using an empirical substitution model (Whelan and Goldman 2001). These raw branch lengths were converted into relative rates of evolution by normalizing for both the average rate of evolution of the investigated gene across all branches as well as the average rate of evolution of all genes within the investigated branch (as in Sato et al. [2005]). A resulting relative rate that is higher than zero corresponds to a gene that evolves faster than expected in the investigated branch, whereas a relative rate below zero corresponds to a gene that evolves slower than expected in the investigated branch. The relative rates were then used to test for the hypothesis that a gene evolves with a relative rate that correlates with the Matrotrophy Index (MI)—a proxy for placental complexity—using a Spearman ranked correlation test. For terminal notes in the phylogeny, this MI value was taken from Pollux et al. (2014). For ancestral nodes, the MI value was estimated using the phytools R package (Revell 2012). As a control, the same analysis was performed on the same data set, but with MI labels for each node randomly shuffled across the phylogeny. On a genome-wide scale, the hypothesis was tested that more genes show an evolutionary rate that correlates well with MI than expected by comparing the case and control *P*-value distributions using the Kolmogorov–Smirnov test. To find candidate genes, correction for multiple testing was performed using the *q*-value method, with a threshold value of $q = 0.1$ (Dabney et al. 2010).

Additionally, to assess robustness of results, we repeated the ancestral trait reconstruction and evolutionary rate analysis on 100 random bootstrap trees (supplementary table 10, Supplementary Material online). We confirmed our candidates for each of the 100 trees, and added the amount of trees that support our candidates in supplementary table 2, Supplementary Material online.

Comparative Noncoding Element Analysis

To compare conserved noncoding elements between placental and nonplacental poeciliids, two multigenome alignments were made: one alignment consisting of the six placental species that have the highest MI values in the family (*Poeciliopsis retropinna*, *Poeciliopsis presidionis*, *Poeciliopsis turneri*, *Ph. januarius*, *Heterandria formosa*, *Po. bifurca*), and one with six nonplacental species that were chosen to follow a similar topology on the phylogeny as for the placental species (*Alfaro cultratus*, *Poeciliopsis turrubarensis*, *Poeciliopsis gracilis*, *Poeciliopsis infans*, *Po. picta*, *Xiphophorus hellerii*). A third multialignment consisting of five teleost fish (*Poeciliopsis retropinna*, *Oryzias latipes*, *Oreochromis niloticus*, *Gasterosteus aculeatus*, *Danio rerio*) was used to get insight on more ancient conservation, but it was not used for differential conservation analysis. The multialignments were made by a custom pipeline (available on https://git.wageningenur.nl/kruis015/whole_genome_alignment). All genomes were aligned to the chosen reference genome (*P. retropinna*) using the pairwise genome aligner MUMmer 4.0.0 (Marçais et al. 2018). After pairwise alignment, overlapping alignment blocks were merged based on reference coordinates and locally realigned using mafft v7.402(Katoh et al. 2002), yielding a reference-based multigenome alignment. For both multigenome alignments, conserved elements and base-specific conservation scores were called using PhastCons v1.5 (Hubisz et al. 2011). The nonconserved model was based on fourfold degenerate sites extracted from the respective alignment. For the nonplacental alignment, the sequence of the placental reference genome (*P. retropinna*) was not used for predicting conservation by using the –not-informative option of the PhastCons program. In this way, base-wise conservation scores between the placental and the nonplacental multialignment could be compared within the same genome, without using this genome for both predictions. After prediction of conserved elements, regions with a large difference in conservation between placental and nonplacental species were extracted from the reference genome. The requirements for this were (1) the predicted element based on one multialignment should not overlap with one based on the other multialignment, (2) the mean difference in conservation scores across all bases of the predicted element should be at least 0.75, (3) there should be a significant difference in base-wise

conservation scores between the two predictions based on a permutation test. As a control, the same analysis was performed, but instead of extracting differentially conserved elements between a placental and a nonplacental set of species, differences between two nonplacental sets of species were extracted. For this, the six nonplacental poeciliid multigenome alignment as mentioned before was used and compared with a multigenome alignment of six other nonplacental poeciliids (*Brachyrhaphis roseni*, *Gambusia affinis*, *Po. gilii*, *Po. mexicana*, *Po. reticulata*, *Xiphophorus maculatus*).

Simple Sequence Repeat Analysis

To investigate the presence of Simple Sequence Repeats (SSRs) in our conserved elements, we identified SSRs across the genome of *P. retropinna* using MISA v1.0 (Thiel et al. 2003). The minimum number of repeated elements for identification of an SSR was given as ten repeats for an element size of one, six repeats for an element size of two, and five repeats for an element size of three or more.

Gene Ontology Enrichment Analysis

Gene Ontology (GO) enrichment tests and network analysis were performed using STRINGdb (database version 11.0)(Szklarczyk et al. 2016). For the evolutionary rate analysis, the predicted protein sequences of genes evolving at significantly different rates in placental species were extracted from the *P. retropinna* genome. The STRING database was searched for the human orthologs of these sequences, followed by manual curation when multiple candidates were presented. Subsequently, GO enrichment tests and network analysis were performed using the STRING web application (database version 11.0).

To find genes potentially regulated by our candidate DCEs, the first gene for which the element lies upstream was selected on both sides of the element as potentially regulated gene. If the first gene next to the element was not in the orientation so that the element lies upstream of the gene, the gene was not selected. In addition, the gene was not selected if it was further than 50 kb away from the element. The identified genes were subjected to the same analysis in STRINGdb as done for the genes identified in the evolutionary rate analysis.

Detecting Positive Selection

We performed tests for positive selection on the 14,468 orthologous gene sets that were previously identified. For each gene, a codon alignment was made using PRANK v.170427 (Löytynoja 2014), using options -codon and -F. These alignments were used to test for positive selection with the codeml program that is part of PAML 4.9 (Yang 2007).

To test for positive selection that may arise during the evolution of the placenta, we used the so-called branch-site model to test for positive selection for every branch leading to all placental species (*P. retropinna*, *P. paucimaculata*, *P. presidionis*, *P. turneri*, *P. prolifica*, *Ph. januarius*, *H. formosa*, *Po. bifurca*). Each gene was tested for every placental branch separately. The hypothesis that genes evolving under positive selection in placental species are overrepresented in the set of genes that show accelerated evolution in placental species was tested using Fisher's exact test.

Additionally, convergence among positive selection when evolving a placenta was tested by using the same branch-site model, but now using the phylogenetic branches leading to six highly placental poeciliids (*P. retropinna*, *P. presidionis*, *P. turneri*, *Ph. januarius*, *H. formosa*, *Po. bifurca*) as foreground branches. As a control, the same analysis was performed to a group of six nonplacental species (*A. cultratus*, *P. turrubarensis*, *P. gracilis*, *P. infans*, *Po. picta*, *X. hellerii*) for which we do not expect genomic convergence. We then compared the distribution of *P* values between the placental group and the control group to see if there is a consistent enrichment of low *P* values when using the placental species as foreground branches using the Kolmogorov–Smirnov test.

Detecting Convergent Amino Acid Substitutions

To test whether placental species disproportionally show convergent amino acid substitutions, we reconstructed ancestral sequences of each of the 14,468 orthologous gene sets that were previously identified with the AAML program that is part of PAML 4.9 (Yang 2007), using an empirical amino acid substitution matrix (Whelan and Goldman 2001). For six highly placental species (*P. retropinna*, *P. presidionis*, *P. turneri*, *Ph. januarius*, *H. formosa*, *Po. bifurca*), we compared the observed amino acid sequence with that of its closest ancestor. However, we took an exception for the ancestor of *P. presidionis* and *P. turneri*, as these species represent a single origin of placentation in the Poeciliidae, and their common ancestor is hypothesized to have a placenta a well. Therefore, we compared both the sequences of *P. presidionis* and *P. turneri* with the common ancestor of these two species and the nonplacental *P. gracilis*. In these comparisons, amino acid substitutions that occur on the same position in at least four out of six comparisons were identified. These amino acid substitutions were noted as potential convergent events. Then, the same analysis was performed for 100 random combinations of six species, to get a background distribution of convergent amino acid substitutions when no morphological convergence is apparent.

Placental Expression Analysis

To confirm expression of candidate genes in placental tissue, we downloaded RNA-seq data gathered from follicular epithelium of the placental *P. retropinna* (Guernsey et al. 2020). These data were mapped to the *P. retropinna* genome using HISAT2 v2.1.0 (Kim et al. 2019), using default settings. To detect expression of candidate genes, the coverage of mapped reads was determined for all exonic positions of candidate genes using samtools depth (Li et al. 2009). An average read coverage of 2× across all exons was used as the cutoff for gene expression.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Samples were collected and exported by B.J.A.P. following the regulations of the Costa Rican government under permit number SINAC-CUS-PI-R-005-2017 or obtained from D.N.R. (University of California Riverside). This work was supported by a WIAS research grant from Wageningen Institute of Animal Sciences, awarded to H.-J.M. and B.J.A.P., and a VIDI grant from the Netherlands Organisation for Scientific Research (864.14.008) awarded to B.J.A.P. All animal-derived biological materials used this study were obtained according to local ethical regulations.

Data Availability

The data underlying this article are available in the European Nucleotide Archive and can be accessed via Bioproject PRJEB37697. For more detailed accession numbers for each species, see supplementary table 8, Supplementary Material online. The multigenome alignment pipeline used in this study is available at https://git.wageningenur.nl/kruis015/whole_genome_alignment.

References

Arancibia-Garavilla Y, Toledo F, Casanello P, Sobrevia L. 2003. Nitric oxide synthesis requires activity of the cationic and neutral amino acid transport system y+ L in human umbilical vein endothelium. *Exp Physiol*. 88(6):699–710.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 19(5):455–477.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972–1973.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1):25–36.

Chikina M, Robinson JD, Clark NL. 2016. Hundreds of genes experienced convergent shifts in selective pressure in marine mammals. *Mol Biol Evol*. 33(9):2182–2192.

Dabney A, Storey JD, Warnes G. 2010. qvalue: Q-value estimation for false discovery rate control. R package version 1.

Farrell JA, Wang Y, Riesenfeld SJ, Shekhar K, Regev A, Schier AF. 2018. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* 360(6392):eaar3131.

Foote AD, Liu Y, Thomas GWC, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V, et al. 2015. Convergent evolution of the genomes of marine mammals. *Nat Genet*. 47(3):272–275.

Furness AI, Pollux BJ, Meredith RW, Springer MS, Reznick DN. 2019. How conflict shapes evolution in poeciliid fishes. *Nat Commun*. 10(1):1–12.

Gaunt S. 2002. Conservation in the Hox code during morphological evolution. *Int J Dev Biol*. 38:549–552.

Gemayel R, Cho J, Boeynaems S, Verstrepen KJ. 2012. Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. *Genes*. 3(3):461–480.

Gharib WH, Robinson-Rechavi M. 2013. The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol Biol Evol*. 30(7):1675–1686.

Grove BD, Wourms JP. 1991. The follicular placenta of the viviparous fish, Heterandria formosa. I. Ultrastructure and development of the embryonic absorptive surface. *J Morphol*. 209(3):265–284.

Grove BD, Wourms JP. 1994. Follicular placenta of the viviparous fish, Heterandria formosa: II. Ultrastructure and development of the follicular epithelium. *J Morphol*. 220(2):167–184.

Guernsey MW, van Kruistum H, Reznick DN, Pollux BJ, Baker JC. 2020. Molecular signatures of placentation and secretion uncovered in *Poeciliopsis* maternal follicles. *Mol Biol Evol*. 37(9):2679–2690.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res*. 41(13):e129.

Hoegg S, Meyer A. 2005. Hox clusters as models for vertebrate genome evolution. *Trends Genet*. 21(8):421–424.

Hubisz MJ, Pollard KS, Siepel A. 2011. PHAST and RPHAST: phylogenetic analysis with space/time models. *Brief Bioinformatics* 12(1):41–51.

Jollie WP, Jollie LG. 1964. The fine structure of the ovarian follicle of the ovoviviparous poeciliid fish, Lebistes reticulatus. II. Formation of follicular pseudoplacenta. *J Morphol*. 114:503–525.

Jue NK, Foley RJ, Reznick DN, O'Neill RJ, O'Neill MJ. 2018. Tissue-specific transcriptome for *Poeciliopsis prolifica* reveals evidence for genetic adaptation related to the evolution of a placental fish. *G3 (Bethesda)* 8:2181–2192.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30(14):3059–3066.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 37(8):907–915.

Kim JC, Mirkin SM. 2013. The balancing act of DNA repeat expansions. *Curr Opin Genet Dev*. 23(3):280–288.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 37(5):540–546.

Krishnan J, Athar F, Rani TS, Mishra RK. 2017. Simple sequence repeats showing 'length preference' have regulatory functions in humans. *Gene* 628:156–161.

Kumar RP, Krishnan J, Singh NP, Singh L, Mishra RK. 2013. GATA simple sequence repeats function as enhancer blocker boundaries. *Nat Commun*. 4(1):1844.

Kumar RP, Senthilkumar R, Singh V, Mishra RK. 2010. Repeat performance: how do genome packaging and regulation depend on simple sequence repeats? *Bioessays* 32(2):165–174.

Kwan L, Fris M, Rodd FH, Rowe L, Tuhela L, Panhuis TM. 2015. An examination of the variation in maternal placentae across the genus Poeciliopsis (Poeciliidae). *J Morphol*. 276(6):707–720.

Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (co-) orthologs in large-scale analysis. *BMC Bioinformatics* 12:124.

Levin M, Anavy L, Cole AG, Winter E, Mostov N, Khair S, Senderovich N, Kovalev E, Silver DH, Feder M, et al. 2016. The mid-developmental transition and the evolution of animal body plans. *Nature* 531(7596):637–641.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.

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(16):2078–2079.

Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. Multiple sequence alignment methods. Totowa (NJ): Humana press. p. 155–170.

Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol*. 14(1):e1005944.

O'Neill MJ, Lawton BR, Mateos M, Carone DM, Ferreri GC, Hrbek T, Meredith RW, Reznick DN, O'Neill RJ. 2007. Ancient and continuing Darwinian selection on insulin-like growth factor II in placental fishes. *Proc Natl Acad Sci U S A*. 104(30):12404–12409.

Parenti LR. 1981. A phylogenetic and biogeographic analysis of cyprinodontiform fishes (Teleostei, Atherinomorpha). *Bull Am Mus Nat Hist*. 168(4):335–557.

Partha R, Chauhan BK, Ferreira Z, Robinson JD, Lathrop K, Nischal KK, Chikina M, Clark NL. 2017. Subterranean mammals show convergent regression in ocular genes and enhancers, along with adaptation to tunneling. *Elife* 6:e25884.

Pollux B, Meredith R, Springer M, Garland T, Reznick D. 2014. The evolution of the placenta drives a shift in sexual selection in livebearing fish. *Nature* 513(7517):233–236.

Pollux B, Pires M, Banet A, Reznick D. 2009. Evolution of placentas in the fish family Poeciliidae: an empirical study of macroevolution. *Annu Rev Ecol Evol Syst*. 40(1):271–289.

Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res*. 44(12):e113.

Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3(2): 217–223.

Reznick DN, Furness AI, Meredith RW, Springer MS. 2017. The origin and biogeographic diversification of fishes in the family Poeciliidae. *PLoS One* 12(3):e0172546.

Reznick DN, Mateos M, Springer MS. 2002. Independent origins and rapid evolution of the placenta in the fish genus *Poeciliopsis*. *Science* 298(5595):1018–1020.

Ruan J, Li H. 2019. Fast and accurate long-read assembly with wtdbg2. *Nat Methods*. 17.2(2020):155–158.

Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science* 364(6435):74–78.

Sato T, Yamanishi Y, Kanehisa M, Toh H. 2005. The inference of protein–protein interactions by co-evolutionary analysis is improved by excluding the information about the phylogenetic relationships. *Bioinformatics* 21(17):3482–3489.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* 6(1):31.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.

Subramanian S, Mishra RK, Singh L. 2003. Genome-wide analysis of Bkm sequences (GATA repeats): predominant association with sex chromosomes and potential role in higher order chromatin organization and function. *Bioinformatics* 19(6):681–685.

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P. 2016. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. *Nucleic Acids Res*. 45(D1):D362–D368.

Thiel T, Michalek W, Varshney R, Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (Hordeum vulgare L.). *Theor Appl Genet*. 106(3):411–422.

Van Der Laan R, Eschmeyer WN, Fricke R. 2014. Family-group names of recent fishes. *Zootaxa* 3882(1):1–230.

van Kruistum H, Guernsey MW, Baker JC, Kloet SL, Groenen MA, Pollux BJ, Megens H-J. 2020. The genomes of the livebearing fish species *Poeciliopsis retropinna* and *Poeciliopsis turrubarensis* reflect their different reproductive strategies. *Mol Biol Evol*. 24:1586–1591.

van Kruistum H, Van Den Heuvel J, Travis J, Kraaijeveld K, Zwaan BJ, Groenen MA, Megens H-J, Pollux BJ. 2019. The genome of the livebearing fish *Heterandria formosa* implicates a role of conserved vertebrate genes in the evolution of placental fish. *BMC Evol Biol*. 19(1):156.

Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 18(5):691–699.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.

Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinform*. *19*(6):15–30.