



# Insights into the Genomics of Clownfish Adaptive Radiation: The Genomic Substrate of the Diversification

Anna Marcionetti  and Nicolas Salamin \*

Department of Computational Biology, Genopode, University of Lausanne, 1015 Lausanne, Switzerland

\*Corresponding author: E-mail: nicolas.salamin@unil.ch.

Accepted: 17 May 2023

## Abstract

Clownfishes are an iconic group of coral reef fishes that evolved a mutualistic interaction with sea anemones, which triggered the rapid diversification of the group. Following the emergence of this mutualism, clownfishes diversified into different ecological niches and developed convergent phenotypes associated with their host use. The genetic basis of the initial acquisition of the mutualism with host anemones has been described, but the genomic architecture underlying clownfish diversification once the mutualism was established and the extent to which clownfish phenotypic convergence originated through shared genetic mechanisms are still unknown. Here, we investigated these questions by performing comparative genomic analyses on the available genomic data of five pairs of closely related but ecologically divergent clownfish species. We found that clownfish diversification was characterized by bursts of transposable elements, an overall accelerated coding evolution, incomplete lineage sorting, and ancestral hybridization events. Additionally, we detected a signature of positive selection in 5.4% of the clownfish genes. Among them, five presented functions associated with social behavior and ecology, and they represent candidate genes involved in the evolution of the size-based hierarchical social structure so particular to clownfishes. Finally, we found genes with patterns of either relaxation or intensification of purifying selection and signals of positive selection linked with clownfish ecological divergence, suggesting some level of parallel evolution during the diversification of the group. Altogether, this work provides the first insights into the genomic substrate of clownfish adaptive radiation and integrates the growing collection of studies investigating the genomic mechanisms governing species diversification.

**Key words:** *Amphiprion*, anemonefish, convergent evolution, coral reef fish, hybridization, positive selection.

## Significance

We commonly associate the term clownfish with the orange clownfish *Amphiprion percula* from the film “*Finding Nemo*,” but clownfishes are more than this iconic species and form a group of 28 species that rapidly diversified once they acquired mutualism with sea anemones. While clownfishes are highly diverse morphologically and ecologically, we are still lacking an understanding of the role played by the genomic architecture of the group during the adaptive radiation. In this study, we investigated this question and found, for instance, that clownfish genomes carry the signal of ancestral hybridization events between species and that their evolution is accelerated. Our results provide the first insights into how the genomic architecture of clownfishes could relate to their morphological and ecological diversity.

## Introduction

Adaptive radiation is an outstanding process considered to play a central role in the buildup of the diversity of life on Earth (Simpson 1953; Schluter 2000). It is defined as the rapid diversification of an ancestral species into a multitude of

new forms that are adapted to diverse ecological niches (Schluter 2000). Famous examples of this process include Darwin’s finches, *Anolis* lizards, stickleback fishes, and cichlids (Schluter 2000). Within adaptive radiations, convergent evolution, defined here as the repeated evolution of

© The Author(s) 2023. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

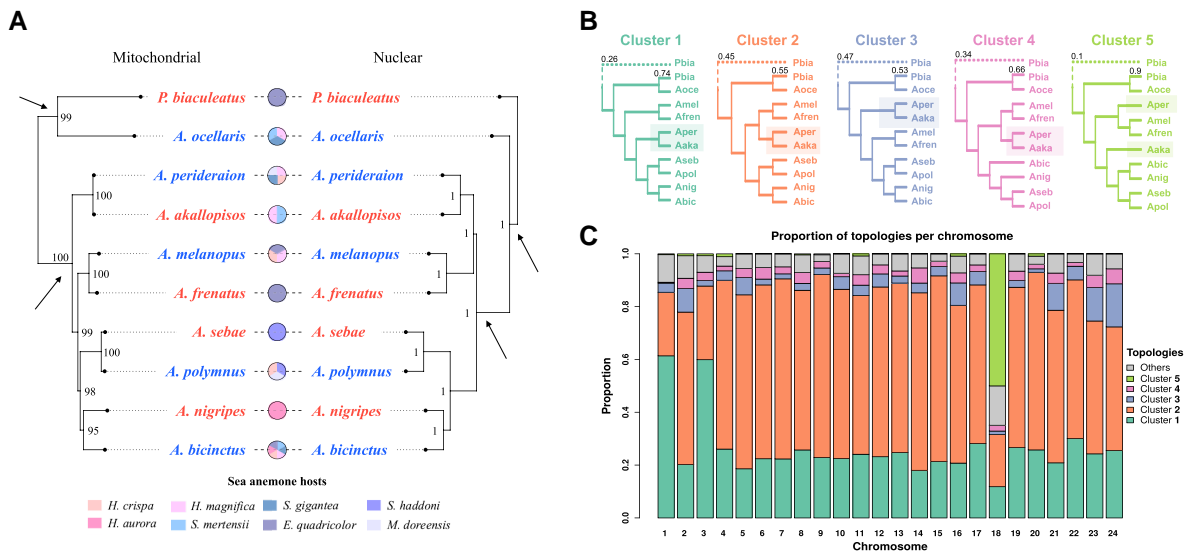
similar phenotypes into distinct lineages, is frequent (e.g., Blackledge and Gillespie 2004; Muschick et al. 2012; Vizueta et al. 2019), and it is generally explained as the result of independent adaptation to similar ecological conditions (e.g., Schluter 2000; Losos 2011). For decades, researchers have been investigating the causes and consequences of adaptive radiations (e.g., Schluter 2000; Seehausen 2004; Stroud and Losos 2016; Martin and Richards 2019) with the goal of broadening our understanding of the mechanisms governing species diversification, and assessing the predictability and repeatability of evolution, which is a central question in evolutionary biology (Rosenblum et al. 2014, Kingman et al. 2021).

With the advances in molecular genetics and sequencing technologies, the development, use, and availability of genomic tools for nonmodel systems have boomed, allowing researchers to incorporate the genomic component into the study of adaptively radiating groups. Several studies have started to investigate the intrinsic genomic factors that could predispose specific lineages to rapidly diversify (e.g., Jones et al. 2012; Brawand et al. 2014; Berner and Salzburger 2015; Feiner 2016; Tollis et al. 2018; Edelman et al. 2019; McGee et al. 2020; Ronco et al. 2021; Xiong et al. 2021). Altogether, they found a wide array of genomic characteristics, such as the presence of chromosomal duplication, bursts of transposable elements (TEs), expansions of gene families, increased genome-wide heterozygosity, and an accelerated evolution on coding and noncoding sequences. Nevertheless, genomic features that could consistently promote diversification in different lineages have not been identified yet (e.g., Brawand et al. 2014; Xiong et al. 2021, but see Ronco et al. 2021). A clear pattern that emerged as a crucial driver of diversification is, however, the access to ancient genetic variation via hybridization (Berner and Salzburger 2015; Marques et al. 2019). For instance, ancestral hybridization between distinct lineages has fueled the adaptive radiation of cichlids (Meier et al. 2017; Svoldal et al. 2020), while introgressive hybridization among members of the radiating lineages can facilitate ecological speciation (such as in *Heliconius* butterflies, Dasmahapatra et al. 2012). Additionally, the access to ancestral polymorphism through gene flow and/or hybridization has also been shown to contribute to the evolution of convergent phenotypes (e.g., Cresko et al. 2004, Dasmahapatra et al. 2012), as it facilitates the reuse of shared genetic and molecular mechanisms in different populations or species (i.e., parallel evolution as defined in Rosenblum et al. 2014).

Our knowledge of the processes leading to adaptive radiations—and the genomic components underlying them—has increased substantially in the past decades, and some common features have started to emerge.

Nevertheless, the genomic characterization of additional lineages is still needed to obtain a more general picture of the genomics of adaptive radiations. In this context, clownfishes (or anemonefish, genera *Amphiprion* and *Premnas*, with *Premnas* having been recently recovered within *Amphiprion*; Tang et al. 2021) are of interest to bring additional insights as a new study group. Clownfishes comprise 28 recognized species and two natural hybrids (Fautin and Allen 1997; Litsios and Salamin 2014; Gainsford et al. 2015), and they are characterized by their obligate mutualism with sea anemones. Inside the sea anemones, clownfishes live in a size-based social hierarchy (Fricke 1979; Buston 2003) and are protandrous sequential hermaphrodites (Fricke 1979). Variability in host use exists, with some clownfish species being specialists (i.e., inhabiting a single species of sea anemones) and others being generalists and associating with up to ten hosts (Fautin and Allen 1997; Gainsford et al. 2015). It has been proposed that this mutualism acted as the key innovation that initiated the adaptive radiation of the group (Litsios et al. 2012; Litsios et al. 2014). Following the acquisition of the interaction with sea anemones, the divergence in host usage likely drove the radiation, and within different clades, increasingly specialized species originated repeatedly and independently (Litsios et al. 2012). As a result, the diversifying species in different clades developed convergent phenotypes associated with host usage (i.e., generalists/specialists' gradient; Litsios et al. 2012). While the genetic basis of mutualism with sea anemones started to be elucidated (Marcionetti et al. 2019), the genomic architecture of clownfishes and how it relates to their ecological diversity has never been investigated before. It was further suggested, based on cytonuclear inconsistency observed on a small number of limited sets of genes, that past hybridization events promoted the radiation of clownfishes (Litsios and Salamin 2014). Additionally, hybridization is still occurring in the group—as shown by the presence of two natural hybrids (Litsios and Salamin 2014; Gainsford et al. 2015)—and may be facilitated by the occasional cohabitation of distinct species within the same sea anemone hosts (Gainsford et al. 2015). However, a thorough characterization of the role played by hybridization in shaping the genomic architecture of clownfishes has never been done, despite the importance of identifying these potential events to better understand the diversification of this group.

In this study, we investigated the genomic architecture of the clownfish adaptive radiation. We considered ten clownfish species for which genomes were publicly available (*Amphiprion akallopisos*, *A. bicinctus*, *A. frenatus*, *A. melanopus*, *A. nigripes*, *A. ocellaris*, *A. perideraion*, *A. polymnus*, *A. sebae*, and *Premnas biaculeatus*; Marcionetti et al. 2019). These species cover the entire



**Fig. 1.**—Cytonuclear incongruence (A), topologies observed along the nuclear genome (B), and their proportions (C). (A) The mitochondrial phylogenetic tree was obtained with RAxML from the whole mitochondrial genome alignments (total of 16,747 bp). The nuclear phylogenetic tree was obtained with ASTRAL-III using a 13,500 gene trees. Bootstrap support values (up to 100) and local posterior probability (up to 1) are reported at the nodes of the mitochondrial and nuclear phylogenetic trees, respectively. The black arrows illustrate the discordance between the two topologies. The possible interactions between clownfish and sea anemone species are shown, with generalist and specialist species reported in blue and red, respectively. (B) The five major topologies observed along the nuclear genome of clownfishes, obtained by hierarchical clustering based on the Robinson–Foulds distance between the trees reconstructed along the genome (supplementary fig. S3, Supplementary Material online). Aaka, Abic, Afre, Amel, Anig, Aoce, Aper, Apol, Aseb, and Pbia correspond to *A. akallopisos*, *A. bicinctus*, *A. frenatus*, *A. melanopus*, *A. nigripes*, *A. ocellaris*, *A. perideraion*, *A. polymnus*, *A. sebae*, and *P. biaculeatus*, respectively. The position of *P. biaculeatus* in the trees forming each cluster was variable, and the numbers on the top of the branches represent the proportion of trees showing the corresponding topology. Significant evidence of hybridization events was detected, with introgression primarily concerning the pair *A. akallopisos* + *A. perideraion* (highlighted; supplementary table S1 and fig. S4, Supplementary Material online). (C) Distribution of the five topologies across the chromosomes of *A. percula*. The additional topologies, reported in gray, are variable (supplementary figs. S2 and S3, Supplementary Material online).

divergence of the group, and they constitute five pairs of closely related species showing ecological and phenotypic divergence within pairs, but ecological and phenotypic convergence between them (Litsios et al. 2012; see fig. 1 in Marcionetti et al. 2019). By performing comparative genomic analyses, we first investigated the relationship between species all along the genome to assess the level of topological inconsistency and tested for hybridization events. By considering additional outgroup species, we then explored the presence of additional genomic features (burst in TEs, signature of accelerated evolution, excess of gene duplications, and patterns of positive selection) that were specific to clownfishes and thus potentially associated with the radiation of the group. Finally, we took advantage of the evolutionary replicates provided by the five pairs of closely related but ecologically divergent clownfish species to determine the extent and the mechanisms of parallel evolution in the group. We predicted that if shared genetic mechanisms were involved in the ecological divergence of multiples species, some genes would display similar evolutionary rates and selective pressures in all the concerned species or show topological inconsistencies in the case of introgressive hybridization.

## Results

### Mosaic Genomes in Clownfishes

The comparison of the mitochondrial and nuclear phylogenetic trees showed the presence of cytonuclear discordance at the deeper nodes of the tree (fig. 1A). Indeed, while at the mitochondrial level, *A. ocellaris* and *P. biaculeatus* were recovered as sister species, *P. biaculeatus* was basal to the other *Amphiprion* in the nuclear tree. Additionally, the pairs *A. melanopus* + *A. frenatus* and *A. perideraion* + *A. akallopisos* formed two sister groups in the nuclear phylogenetic tree, while the latter was basal to *A. melanopus* + *A. frenatus* in the mitochondrial phylogenetic tree (fig. 1A). This cytonuclear discordance was strongly supported, with bootstrap values higher than 0.95 and local posterior probabilities of 1 in the mitochondrial and nuclear trees, respectively (fig. 1A). Nevertheless, we observed a considerable level of gene tree incongruence at the nuclear level, as shown by the quartet support values obtained with ASTRAL (overall normalized quartet score of 0.79, supplementary fig. S1, Supplementary Material online).

Topological incongruence was also observed when looking at the phylogenetic trees reconstructed along the

nuclear genome (fig. 1B and supplementary fig. S2, Supplementary Material online). Indeed, the 5,936 reconstructed trees clustered into five major groups (fig. 1B and supplementary fig. S3, Supplementary Material online). Clusters 1 and 2 contained 27% and 57% of the trees, while the three others included far fewer (5%, 3%, and 2% of the trees for clusters 3, 4, and 5, respectively). The remaining trees (6%) showed a larger number of topologies and were not considered as a group (supplementary figs. S2 and S3, Supplementary Material online). Within each cluster, the branching of *P. biaculeatus* was variable (depicted in fig. 1B by dotted lines), with a slight majority of trees (53% to 74% depending on the cluster) placing *P. biaculeatus* as the sister species of *A. ocellaris*. An exception was observed for cluster 5, where *P. biaculeatus* was found at this position 90.8% of the time (fig. 1B).

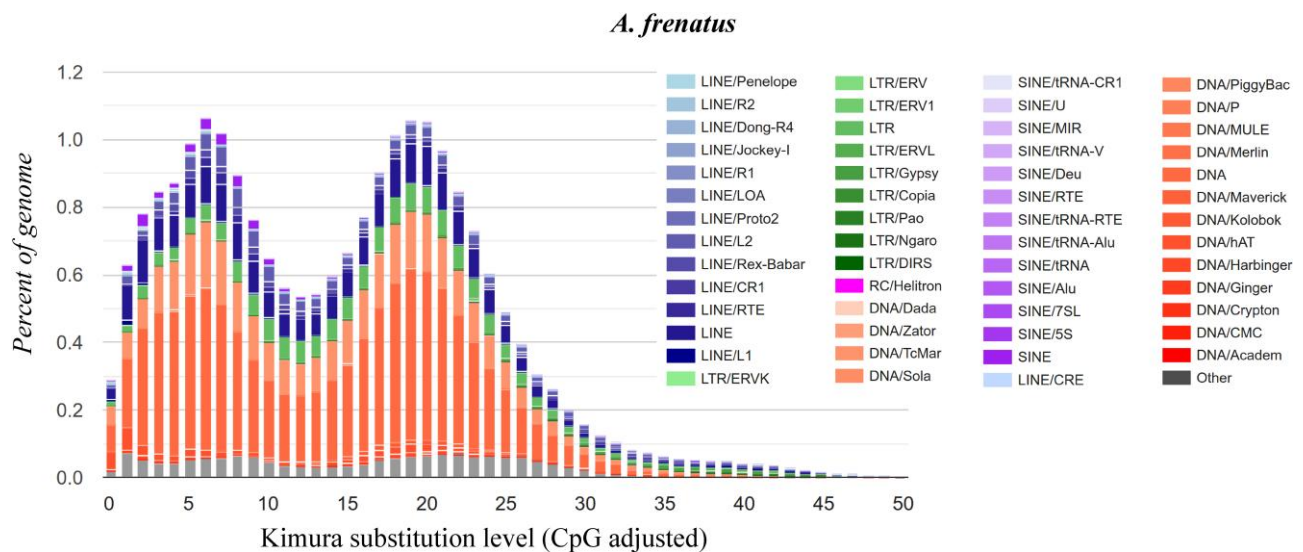
The topologies of the five clusters mainly differed in the branching of the *A. perideraion* + *A. akallopisos* pair. Indeed, *A. perideraion* + *A. akallopisos* and *A. melanopus* + *A. frenatus* were sister groups in cluster 2 (topology of the nuclear phylogenetic tree, fig. 1), while in cluster 1, the pair *A. perideraion* + *A. akallopisos* was sister to the *A. bicinctus* + *A. nigripes* + *A. sebae* + *A. polymnus* complex (fig. 1B). Cluster 3 was characterized by the *A. perideraion* + *A. akallopisos* pair being basal to the *A. melanopus* + *A. frenatus* pair (topology of the mitochondrial phylogenetic tree, fig. 1). In cluster 5, *A. perideraion* and *A. akallopisos* were not recovered as sister species, but *A. akallopisos* was basal to the *A. bicinctus* + *A. nigripes* + *A. sebae* + *A. polymnus* clade, while *A. perideraion* was sister to the *A. melanopus* + *A. frenatus* pair (fig. 1B). Cluster 4 was similar to cluster 2, but with *A. bicinctus* being basal to *A. nigripes* (fig. 2B). In accord with these topological inconsistencies, the HyDe analyses performed on the whole genome resulted in the detection of past hybridization events between the clades. Indeed, significant hybridization was detected in the *A. perideraion* + *A. akallopisos* pair when considering both the *A. bicinctus* + *A. nigripes* ( $Z$ -score = 222.62,  $P \cong 0.0$ ,  $\gamma = 0.629$ ) and *A. sebae* + *A. bicinctus* pairs ( $Z$ -score = 200.00,  $P \cong 0.0$ ,  $\gamma = 0.646$ ; supplementary table S1 and fig. S4, Supplementary Material online). Similarly, the computed Patterson's  $D$  statistics showed significant introgression in the *A. perideraion* + *A. akallopisos* pair from both the *A. bicinctus* + *A. nigripes* ( $D$  statistics of 0.24,  $P < 2.2e^{-16}$ ) and the *A. sebae* + *A. polymnus* clades ( $D$  statistics of 0.26,  $P < 2.2e^{-16}$ ). Results were consistent when *A. akallopisos*, and *A. perideraion* were considered separately (supplementary table S1, Supplementary Material online).

When looking at the distribution of the five topologies along the clownfish genome, we saw that the proportions of each topology in the 24 chromosomes were quite similar, with the topology of cluster 2 being the most frequent (fig. 1C). Nevertheless, we observed three exceptions on

chromosomes 1, 3, and 18. On chromosomes 1 and 3, the most abundant topology was the one of cluster 1 (around 60% of the windows; fig. 1C). This was reflected by a slightly higher Patterson's  $D$  statistics compared with the other chromosomes (supplementary table S1 and fig. S4, Supplementary Material online). Chromosome 18 showed a different scenario, with the most abundant topology being the one of cluster 5 (i.e., with *A. perideraion* and *A. akallopisos* not recovered as sister species, 50% of the windows), which was rarely observed elsewhere (fig. 1C). The HyDe analyses for this chromosome mirrored this result. Indeed, while  $\gamma$  values obtained for the *A. akallopisos* + *A. perideraion* pair were consistent with the whole genome results (supplementary table S1, Supplementary Material online), when considering the two species separately, the hybridization signal was higher for *A. akallopisos* (e.g.,  $Z$ -score = 25.5,  $P \cong 0.0$ ,  $\gamma = 0.943$ ) than for *A. perideraion* (e.g.,  $Z$ -score = 65.6,  $P \cong 0.0$ ,  $\gamma = 0.199$ ; supplementary table S1, Supplementary Material online). Similarly, the  $D$  statistics for introgression from *A. bicinctus* + *A. nigripes* and from *A. sebae* + *A. polymnus* clades in *A. akallopisos* were, respectively, of 0.68 ( $P < 2.2e^{-16}$ ) and 0.67 ( $P < 2.2e^{-16}$ ), while they were lower and similar to the rest of the genome for *A. perideraion* (supplementary table S1 and fig. S4, Supplementary Material online). Gene Ontology (GO) enrichment analysis on the 432 genes located in the regions showing support for topology 5 on chromosomes 18 resulted in 24 enriched terms ( $P < 0.01$ ; supplementary table S2, Supplementary Material online) associated with the morphogenesis of the epithelium (GO:1905332), fertilization (GO:0009566), axis elongation (GO:0003401), and retina vasculature development (GO:0061298).

## TE Content

We found that approximately 23–25% of clownfish genomes consisted of TEs, with TE proportion and composition similar to what was observed in *A. percula* and *Oreochromis niloticus* genomes (supplementary fig. S5 and table S3, Supplementary Material online). We detected two major TE bursts in the clownfish genomes (Kimura distance  $K$ -value of 0.05–0.06 and 0.18–0.19; fig. 2 and supplementary fig. S6, Supplementary Material online). The most recent burst was also present in *Pomacentrus moluccensis*, but it was less pronounced. Indeed, at  $K$ -values of 0.05 and 0.06, the average TE content in clownfishes was 1.07% (standard deviation (SD) = 0.06) and 1.14% (SD = 0.05) but only 0.45% and 0.53% in *P. moluccensis* genome (supplementary fig. S6, Supplementary Material online). When comparing the Kimura distance to the neutral genomic divergence between clownfish and the outgroup (see species divergence, supplementary fig. S7, Supplementary Material online), we found that the older



**Fig. 2.**—Kimura distance-based copy divergence analysis of TEs in *A. frenatus*. The different colors represent the different TE superfamilies. TEs were clustered according to Kimura distances ( $K$ -values) to their corresponding consensus sequence. Copies clustering on low  $K$ -values did not largely diverge from the consensus sequence and likely coincided to recent events, while sequences with higher  $K$ -values likely corresponded to older divergence. Peaks in the graph indicate TE bursts. Similar results were obtained for the additional nine clownfish (supplementary fig. S6, Supplementary Material online).

TE burst ( $K$ -value of 0.18–0.21) happened at the time of the split of *O. niloticus* with the common ancestor of the Pomacentridae. The more recent burst ( $K$ -value of 0.05–0.06) was situated around the split of *P. moluccensis* from the common ancestor of clownfishes.

#### Gene Duplication Rate and Positive Selection on Multicopy Genes

We tested whether clownfishes showed an increased rate of gene duplication that could be associated with their diversification. We found that the duplication rate in the common ancestor of clownfishes was similar to the one observed before their split from *P. moluccensis* and to the one in the common ancestor of the Pomacentridae (around 12 duplicated genes/percent of divergence; supplementary Information S1 and fig. S7, Supplementary Material online). Furthermore, we did not detect any multicopy gene with evidence of positive selection throughout the whole clownfish clade (supplementary Information S1, Supplementary Material online).

#### Overall Accelerated Evolution and Positively Selected Single-Copy Genes in Clownfish

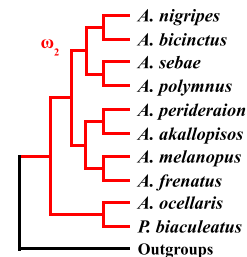
We investigated the rate of protein-coding sequence evolution in clownfishes and found a significantly higher  $\omega$  value in clownfishes ( $M = 0.14$ ,  $SD = 0.03$ ) compared with the outgroups ( $M = 0.09$ ,  $SD = 0.02$ ;  $t(73) = 11.0$ ,  $P < 0.001$ ), which suggests an overall accelerated evolution in this group.

After correcting for multiple testing, we identified 732 genes that showed a significant signal of positive selection in the whole clade (5.4% of the genes tested, fig. 3A). The positively selected genes were distributed across the 24 chromosomes (between 21 and 50 significant genes per chromosome; supplementary fig. S8, Supplementary Material online). After normalizing by the total number of genes mapped to each chromosome, the percentage of positively selected genes was homogeneous across the chromosomes (supplementary fig. S8, Supplementary Material online). The proportion of sites under positive selection and the estimated  $\omega$  for those sites varied across positively selected genes (supplementary fig. S9, Supplementary Material online). On average, the percentage of sites with signatures of positive selection (site class 2a and 2b, fig. 3A) was 2.7%, with an average foreground  $\omega$  of 19.2.

The data simulated under purifying selection or neutral evolution scenarios showed that the positive selection analyses did not detect any false positives (supplementary fig. S10, Supplementary Material online). The power to detect positive selection increased with the increasing strength of selection (i.e., increasing  $\omega$ , supplementary fig. S10, Supplementary Material online). While the power was only 5% when we simulated an  $\omega$  of 2, it increased to 54%, 85%, and 97.5% for a  $\omega$  of 5, 10, and 20, respectively (supplementary fig. S10, Supplementary Material online). Positive selection analyses also showed consistent results when using the gene trees or the species tree (supplementary table S4, Supplementary Material online).

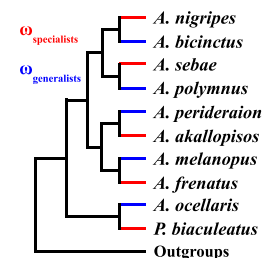
**A**

Sites	Class 0		Class 1		Class 2a			Class 2b		
	Prop.	Backg. Foreg.	Prop.	Backg. Foreg.	Prop.	Backg.	Foreg.	Prop.	Backg.	Foreg.
<b>H1</b>	$p_0$	$0 < \omega_0 < 1$	$p_1$	$\omega_1 = 1$	$\frac{(1 - p_0 - p_1)p_0}{p_0 + p_1}$	$0 < \omega_0 < 1$	$\omega_2 \geq 1$	$\frac{(1 - p_1 - p_0)p_1}{p_0 + p_1}$	$\omega_1 = 1$	$\omega_2 \geq 1$
<b>H0</b>	$p_0$	$0 < \omega_0 < 1$	$p_1$	$\omega_1 = 1$	$\frac{(1 - p_1 - p_0)p_1}{p_0 + p_1}$	$0 < \omega_0 < 1$	$\omega_2 = 1$	$\frac{(1 - p_1 - p_0)p_1}{p_0 + p_1}$	$\omega_1 = 1$	$\omega_2 = 1$



**B**

Sites	Class 0		Class 1		Class 2	
	Prop.	$\omega$	Prop.	$\omega$	Prop.	$\omega$
<b>Clade Model C</b>	$p_0$	$0 < \omega_0 < 1$	$p_1$	$\omega_1 = 1$	$1 - p_0 - p_1$	$\omega_{\text{outgroup}}, \omega_{\text{specialists}}, \omega_{\text{generalists}} > 0$
<b>M2a_rel</b>	$p_0$	$0 < \omega_0 < 1$	$p_1$	$\omega_1 = 1$	$1 - p_0 - p_1$	$\omega_{\text{outgroup}} (= \omega_{\text{specialists}} = \omega_{\text{generalists}}) > 0$



**Fig. 3.**—Schematic trees and models used for the positive selection analysis in the whole clownfish clade (A) and the analysis of selection linked with host and habitat divergence (B). (A) In branch–site model A, the null model with fixed  $\omega_2$  (H0) was compared with the alternative model (H1), where  $\omega_2$  was estimated. A total of 732 genes were significant and showed patterns of positive selection. (B) The null model M2a\_rel with equal  $\omega$  across branches was compared with Clade Model C, where  $\omega$  could vary in the three groups (i.e., specialists’ branches in red, generalists’ branches in blue, and outgroups in black). A total of 3,991 were significant and showed different selective pressures linked with habitat divergence. For details on the trees used in the analyses, see [Supplementary Information S2](#).

**Table 1**

Positively Selected Genes in the Whole Clownfish Clade Showing Particularly Interesting Functions

OG ID	SwissProt ID	Protein name	logL(H0)	logL(H1)	LRT $q$ -values	% PS sites	$\omega$
OG20045_2b	Q90334	Isotocin receptor (ITR)	−4555.44	−4547.56	4.60e <sup>−03</sup>	0.3	51.142
OG14285	P01170	Somatostatin 2 (SST2)	−1784.81	−1768.73	4.31e <sup>−03</sup>	2.7	61.387
OG16291	Q9Y5X5	Neuropeptide FF receptor 2 (NPFFR2)	−5198.93	−5191.55	5.55e <sup>−03</sup>	1	10.922
OG8036	Q8HZK2	Dual oxidase 2 (DUOX2)	−32014.46	−32005.58	4.31e <sup>−03</sup>	0.4	14.285
OG8544	O18315	Rhodopsin (RHO)	−4389.29	−4374.97	1.12e <sup>−02</sup>	1	26.142

For each gene, the corresponding SwissProt ID and gene names are reported. Information on the log-likelihood of the null model (H0, no positive selection) and alternative model (H1, positive selection in clownfishes, fig. 3A) is reported. The  $q$ -values of the LRT after multiple-testing correction, the proportion of sites under positive selection on the tested branches (% PS sites), and the corresponding  $\omega$  values are reported for each gene.

GO enrichment analysis performed on the positively selected genes resulted in 30 enriched GO terms ([supplementary table S5, Supplementary Material online](#)). We found GOs linked to sexual reproduction (GO:0019953), detection of abiotic stimulus (GO:0009582), cellular response to interferon-gamma (GO:0071346), and cuticle development (GO:0042335). Within the positively selected genes participating in the enrichment of these GO terms ([supplementary table S6, Supplementary Material online](#)), we found genes encoding for neuropeptide FF receptor 2 (NPFFR2, OG16291; [table 1](#)), the dual oxidase protein (DUOX, OG8036; [table 1](#)), and the rhodopsin (RHO, OG8544; [table 1](#)). Among the genes with the strongest signal of

selection ([supplementary table S7, Supplementary Material online](#)), we also found the ones encoding for the isotocin receptor (ITR, OG20045\_2b; [table 1](#)) and for the somatostatin 2 (SST2, OG14285; [table 1](#)).

**Parallel Evolution Associated with Host and Habitat Divergence**

We did not detect any gene with a higher evolutionary rate in all generalists or in all specialists ([Supplementary Information S3](#)). Similarly, we did not identify regions of topological inconsistency where specialist or generalist species branched together (fig. 1B), which would have

suggested recent introgressive hybridization. Nevertheless, we detected 3,991 genes that showed a different  $\omega$  in specialists, generalists, and outgroups (clade model C better than null model M2a\_rel;  $P$  value corrected for multiple testing; fig. 3B). Most of these genes (97.4%) presented signatures of purifying selection or neutral evolution ( $\omega \leq 1$ ) in all specialists, generalists, and background branches. However, there were twice as many genes where the  $\omega_{\text{specialists}}$  was estimated to be at least ten times lower than the  $\omega_{\text{generalists}}$  (1,929 genes) than the reverse (955 genes). We also found a total of 155 genes showing relaxation or intensification of purifying selection in either specialists or generalists (table 2 and supplementary table S8, Supplementary Material online). The estimates of  $\omega_{\text{generalists}}$ ,  $\omega_{\text{specialists}}$ , and  $\omega_{\text{background}}$  for these genes are shown in the supplementary figure S11, Supplementary Material online. GO enrichment analysis on the genes showing intensified purifying selection in specialists resulted in 8 enriched GOs, including terms associated with the reproductive process (GO:0001541, ovarian follicle development; GO:0007283, spermatogenesis; supplementary table S9, Supplementary Material online). Similarly, the enrichment of one GO term was found for genes with patterns of intensified purifying selection in generalists (GO:0046676, negative regulation of insulin secretion; supplementary table S9, Supplementary Material online).

Genes with patterns of positive selection specific to specialists ( $\omega_{\text{specialists}} > 1.5$ ;  $\omega_{\text{generalists}}$  and  $\omega_{\text{background}} \leq 1$ ) and generalists ( $\omega_{\text{generalists}} > 1.5$ ;  $\omega_{\text{specialists}}$  and  $\omega_{\text{background}} \leq 1$ ) were also detected (table 2 and supplementary table S10 and fig. S12, Supplementary Material online). GO enrichment analysis on genes positively selected in specialists resulted in two enriched ontologies: GO:0031398 (positive regulation of protein ubiquitination) and GO:0071456 (cellular response to hypoxia; supplementary table S9, Supplementary Material online). Similarly, two enriched GO terms were detected for genes positively selected in generalists: GO:1903076 (regulation of protein localization to the plasma membrane) and GO:0006906 (vesicle fusion; supplementary table S9, Supplementary Material online).

**Table 2**

Number of Genes Showing Patterns of Selection Linked with Host and Habitat Divergence

	Specialists	Generalists
Relaxation from purifying selection	4	16
Intensification of purifying selection	82	53
Positive selection	26	39

Information on genes under purifying selection is available in supplementary table S8, Supplementary Material online, while positively selected genes are reported in supplementary table S10, Supplementary Material online. The distribution of the  $\omega$  values for each category is reported in supplementary tables S11 (for purifying selection) and S12 (for positive selection), Supplementary Material online.

## Discussion

In this study, we took advantage of the genomic data available for five pairs of closely related but ecologically divergent clownfish species to investigate the genomic architecture of clownfish adaptive radiation. We found that clownfishes possess mosaic genomes resulting from past hybridization events, suggesting that hybridization might have acted as a driver also in the diversification of this group (Berner and Salzburger 2015; Marques et al. 2019). We also saw that clownfish radiation was associated with bursts of TEs, an overall accelerated coding evolution, and signals of positive selection in 5.4% of the analyzed genes. Among them, five had functions linked to social behavior and ecology (further discussed below), and they represent candidate genes involved in the evolution of the size-based hierarchical social structure so particular to clownfishes. Finally, while we did not observe evidence of introgressive hybridization linked with host usage, we found genes with patterns of relaxation or intensification of purifying selection and genes with signals of positive selection linked with the ecological divergence, suggesting some level of parallel evolution during the diversification of the group.

### Incomplete Lineage Sorting (ILS) and Past Hybridization Events in Clownfish Diversification

We detected cytonuclear incongruence and topological inconsistencies along clownfish genomes. The main topological inconsistencies were observed in the deep nodes of the tree, while the five pairs of closely related species mainly branched together. Nevertheless, we observed a clear exception for *P. biaculeatus*, which mostly branched as basal to the *Amphiprion* clade but was also frequently recovered as sister to *A. ocellaris* (fig. 1C). This disparity mirrors the conflicting phylogenies reported in the literature (e.g., Frédérich et al. 2013, Litsios and Salamin 2014, na Ayudhaya et al. 2017) and possibly originated from high levels of ILS, as suggested by the low-quartet support value of the node obtained with ASTRAL (supplementary fig. S1, Supplementary Material online).

While ILS also occurred throughout the diversification of clownfishes (supplementary fig. S1, Supplementary Material online), we observed significant evidence of past hybridization events in the group. Indeed, the pair *A. perideraion* + *A. akallopisos* showed significant levels of introgressive hybridization from both *A. bicinctus* + *A. nigripes* and *A. sebae* + *A. polymnus* clades, without a clear difference in the estimated Patterson's  $D$  statistics for the two clades. As the whole genome  $D$  statistics were consistent also when considering *A. akallopisos* and *A. perideraion* separately, these results suggest that hybridization events occurred between the ancestor of the *A. akallopisos* + *A. perideraion* pair and the one of the other clades.

A distinctive pattern was observed on chromosome 18, where the most frequent topology splits *A. perideraion* from *A. akallopisos* (fig. 1C). *A. akallopisos* was recovered as sister to *A. nigripes* + *A. bicinctus* + *A. sebae* + *A. polymnus* and showed an increased level of introgressive hybridization (supplementary table S1, Supplementary Material online). This topology clustered in two large regions and was rarely observed elsewhere, suggesting that a stronger signal of introgression persisted on chromosome 18 of *A. akallopisos* despite extensive backcrossing through mechanisms that disrupt recombination, such as chromosome inversions (Stevison et al. 2011). Genomic inversions that break recombination, creating clusters of loci controlling ecologically important traits, are observed in the case of supergenes (e.g., Joron et al. 2011; Küpper et al. 2016; Branco et al. 2018). Here, however, we cannot easily link the functions of genes in these regions (supplementary table S2, Supplementary Material online) to important ecological traits without further studies. It is worth mentioning that the evolution of sex chromosomes may also result in similar chromosome patterns (e.g., Natri et al. 2019). However, we do not believe that this is relevant in clownfishes as the species are sequential hermaphrodites and do not possess sex chromosomes (Fricke 1979; Arai 2011), and genes involved in the sex change are scattered throughout the genome (Casas et al. 2018).

By considering 10 clownfish species out of the existing 28, we might be underestimating the extent of genetic exchange between species, and hybridization events (both ancestral and recent) might be even more pervasive. Nevertheless, we showed that ancestral hybridization has occurred during clownfish radiation, and we open the road for further investigation on the interesting patterns observed on clownfishes' chromosome 18. These hybridization events possibly facilitated the diversification of the group (Litsios and Salamin 2014), as observed in other radiating lineages (Berner and Salzburger 2015; Marques et al. 2019).

#### Extended Burst in TE at the Basis of Clownfish Radiation

The proportion of TEs present in clownfishes is high (20–25% of the genome) and is comparable to what was observed in East African cichlids (Brawand et al. 2014; Shao et al. 2019) and in the *A. percula* genome (Lehmann et al. 2019). This result confirms a reliable annotation of the TEs despite the higher fragmentation of the assemblies used in this study.

The large TE proportion in clownfishes originated mainly from two bursts of transpositions. The older burst happened in the ancestor of *O. niloticus* and the Pomacentridae, and it was also observed in the cichlids (Shao et al. 2019) and *P. moluccensis* (supplementary fig. S6, Supplementary Material online). The more recent burst

occurred around the split of *P. moluccensis* from the common ancestor of clownfishes. This increase in transpositions was also observed in *P. moluccensis* but was less intense and rapidly decreased, suggesting that the burst started in the *P. moluccensis*–clownfish ancestor and, after the split of the species, only continued in clownfishes.

The movement of TEs can lead to insertions, deletions, and chromosomal rearrangements, and it has also been shown to promote speciation and species diversification (e.g., Lönnig and Saedler 2002; Feiner 2016, Auvinet et al. 2018). Nevertheless, the extent to which TEs might facilitate adaptive radiations is still unclear and might depend on the lineages considered. For instance, a correlation between species diversity and TE activity was observed in *Anolis* lizards (Feiner 2016) but not in cichlids (Ronco et al. 2021; Xiong et al. 2021). Here, we found that clownfish radiation is associated with an increased burst in TEs, which thus potentially played a role in the diversification of the group. Nevertheless, further studies are needed to assess a causality between this burst and the increased diversification rate of the group.

#### Accelerated Coding Evolution Potentially Linked with Mutualism

We detected a significantly increased evolutionary rate in a subset of randomly sampled clownfish genes, suggesting a global accelerated coding evolution in the group. Interestingly, accelerated morphological evolution in clownfish compared with damselfish has also been previously reported (Litsios et al. 2012). While accelerated evolution in genes involved in morphological and developmental processes was observed in Eastern African cichlids (Brawand et al. 2014), in clownfishes, this acceleration seems to be generalized at the whole-genomic level.

Overall differences in evolutionary rates between lineages are widely observed (e.g., Welch et al. 2008; Bromham 2009; Thomas et al. 2010) and are primarily attributed to differences in generation time, where organisms with shorter generation times likely evolve faster (Bromham 2009). However, the generation time in clownfishes (5 years estimated on *A. percula*, Buston and García 2007) is longer than in the considered outgroup species (i.e., from a few months to 4 years, supplementary table S11, Supplementary Material online), excluding it as the driver of the acceleration in coding evolution. The effective population size ( $N_e$ ) also affects the evolutionary rate through the action of genetic drift (Lynch and Walsh 2007), and lineages with smaller  $N_e$  show an increased rate of evolution (e.g., Woolfit and Bromham 2003, 2005). Although estimates of  $N_e$  for the species in the study were not available, sequential hermaphroditism can potentially reduce it (Coscia et al. 2016; Benvenuto et al. 2017; but not in Waples et al. 2018), and clownfishes are the only sex-



changing species considered here (data from FishBase; Froese and Pauly 2000).

Coevolution between species can also modify the rate of molecular evolution. Antagonistically coevolving species are expected to show an increased rate of molecular evolution (the Red Queen Hypothesis, Van Valen 1973; Paterson et al. 2010), while species in mutualistic relationships should show a decrease in their rate of evolution (the Red King Effect theory, Bergstrom and Lachmann 2003). Nevertheless, similarly to what was detected here for clownfishes, an increased evolutionary rate was observed in obligate symbiotic organisms (e.g., Lutzoni and Pagel 1997; Yoshizawa and Johnson 2003; Bromham et al. 2013) and mutualistic ants (Rubin and Moreau 2016). Demographic history (e.g.,  $N_e$  and generation time) could explain the increased rate in symbiotic organisms, but not in the mutualistic ants (Rubin and Moreau 2016). In this less intimate interaction, the increased rate of evolution was suggested to result from the presence of multiple dynamic environments to which the species have to adapt (theirs and the ones of their symbionts; Rubin and Moreau 2016), likely leading to selective pressure similar to those experienced by antagonistically coevolving species (Van Valen 1971). This hypothesis could also hold for clownfish mutualism, especially given the presence of generalist species living in different hosts (see also below). Thus, while we cannot exclude an effect of demographic history without further analyses, the increased evolutionary rate observed in clownfishes may also result from the acquisition of mutualism with sea anemones.

### Genes Under Positive Selection and Associated with the Social Structure of Clownfishes

We found that at least 5.4% of the genes were positively selected throughout clownfish diversification. While the simulations demonstrated the absence of false positive, they also showed a low power to detect weak selection. Thus, this percentage of positively selected genes might be underestimated. By investigating the genes with the highest strength of selection, we identified five genes with particularly interesting functions linked with social behavior and ecology that we further discuss.

First, we detected positive selection on the gene encoding the somatostatin 2 (SST2, OG14285, table 1). Somatostatins are a family of peptide hormones that influence organismal growth by inhibiting the production and release of the growth hormone (Sheridan and Hagemester 2010). The variation in somatic growth rate resulting from social status changes in the fish *Astatotilapia burtoni* was associated with a shift in the volume of somatostatin-containing neurons (Hofmann and Fernald 2000). Similarly, in *A. burtoni*, somatostatin 1 regulates aggressive behavior in dominant males (Trainor and

Hofmann 2006), and its expression in the hypothalamus is associated with the control of somatic growth depending on social status (Trainor and Hofmann 2007). Second, we observed positive selection on the gene encoding the isotocin receptor (ITR, OG20045\_2b, table 1). In the cichlid *Neolamprologus pulcher*, isotocin is involved in modulating social behavior, increasing responsiveness to social information (e.g., Reddon et al. 2012, 2015). For instance, exogenous administration of isotocin followed by an aggressive challenge on the fish resulted in an increased submissive behavior (Reddon et al. 2012). We also detected positive selection on the gene *NPFFR2* (OG16291, table 1), encoding for the neuropeptide FF receptor 2, which plays a role in processes related to feeding (including food intake, appetite control, and gastrointestinal motility) in *Lateolabrax maculatus* (Li et al. 2019).

Within sea anemones, clownfishes are organized in a size-based dominance hierarchy, with the female and the male being, respectively, the largest and the second-largest individuals. The nonbreeders (if present) are gradually smaller as the hierarchy is descended (Fricke 1979). These differences in sizes between individuals are maintained by a precise regulation of the growth of subordinates (Buston 2003), often achieved through aggressive behavior (Iwata et al. 2008). It was also hypothesized that, in *A. percula*, subordinates reduce their food intake to avoid surpassing size thresholds, which could lead to conflicts with dominants (Chausson et al. 2018). Thus, positive selection in somatostatin 2, isotocin receptor, and *NPFFR2* genes possibly contributed to the evolution of the size-based dominance hierarchy in clownfishes through the modulation of both growth and aggressive/submissive behaviors.

We also observed positive selection in the *RHO* gene (OG8544, table 1), encoding the rhodopsin. This protein is a photoreceptor necessary for image-forming vision at a low light intensity, and modifications in its gene sequence results in changes in the absorbed wavelength (Bowmaker 2008). In cichlids, positive selection on rhodopsin was associated with shifts in the wavelength absorbance of fish at different depths, promoting ecological divergence in the group (Spady et al. 2005). Divergent positive selection on this gene was also observed among lake and river cichlid species (Schott et al. 2014). Clownfishes live at depths between 1 and 40 meters, with the depth depending on the habitat of their host sea anemones (Fautin and Allen 1997). Although a causative effect of the positively selected sites on rhodopsin absorbance on clownfishes should be formally tested before drawing robust conclusions, this gene represents an interesting candidate gene that may have played a role in clownfish adaptation to different depth.

Finally, we detected positive selection in the *DUOX2* gene (OG8036, table 1), encoding the dual oxidase protein involved in synthesizing thyroid hormones (Chopra et al.

2019), which regulates the white stripe formation in *A. percula* (Salis et al. 2021). They also observed that shifts in *DUOX2* expression and thyroid hormone levels due to ecological differences resulted in the divergent formation of stripes and color patterns in *A. percula* (Salis et al. 2021). The function of the striped patterns in clownfish is still unknown but could be associated with the sea anemone ecology (Salis et al. 2021) or with species recognition, as observed in other teleosts (e.g., Kelley et al. 2013). Positive selection on *DUOX2* in clownfishes may thus be associated with divergence in the formation of white stripes in the group.

These genes show interesting functions associated with clownfish social behavior and ecology and thus represent candidate genes involved in the evolution of the sized-based hierarchical social structure so particular to this group. Nevertheless, further functional studies are needed to validate their specific function and confirm their role in clownfish diversification.

#### No Increase in Gene Duplication Rate at the Basis of Clownfish Radiation

The highest duplication rate was observed in the common ancestor of the Pomacentridae and *O. niloticus* (supplementary fig. S7, Supplementary Material online) and was comparable to the estimated duplication rate in the cichlids' ancestor (Brawand et al. 2014). The duplication rate detected in clownfish was similar to the one observed in nonradiating teleosts (fig. 2 in Brawand et al. 2014). While the overall duplication rate may be underestimated in a phylogenetic duplication analysis (PDA) approach (as performed here) compared with read depth or array comparative genomic hybridization (aCGH) methods, the relative difference in the rate between branches remains consistent among the analyses (see Brawand et al. 2014), which reinforce the validity of our findings.

Although we did not observe an increased gene duplication rate in clownfishes, we tested if the duplication events were, nevertheless, followed by positive selection in the whole group. Indeed, gene duplications allow for the divergent evolution of the resulting gene copies, permitting functional innovation of the proteins and/or expression patterns (Lynch and Conery 2000; Kondrashov 2012). We did not detect any duplicated gene positively selected in all clownfishes and thus associated with the diversification of the whole clade. It is worth mentioning that the method we employed here tests for signals of positive selection in all branches and may result in a reduced power due to multiple testing (Smith et al. 2015). Nevertheless, it was important to perform the analysis in an exploratory way, as we did not know beforehand which copy of the genes was potentially positively selected. Anyhow, if some duplicated genes were indeed positively selected in all clownfishes, the

intensity of positive selection was not strong enough to permit their detection.

#### Parallel Evolution in Clownfish Diversification

We did not detect extensive topological inconsistencies potentially arising from hybridization events among clownfish specialist or generalist species (fig. 1B), indicating that parallel evolution through recent introgressive hybridization is not occurring in the species considered in this study. However, we identified genes that consistently experienced similar selective pressures in all specialist or generalist clownfish species. Indeed, we detected genes that were specifically positively selected in all specialists or generalist species and thus might have played a role in the parallel adaptation to similar ecological niches and the evolution of morphological convergence in clownfishes (Litsios et al. 2012). We also detected genes with parallel patterns of relaxation or intensification of purifying selection in specialist or generalist species. The selective pressures on these genes may reflect parallel outcomes of the adaptation to similar ecological niches. These results suggest the presence of some level of parallel evolution during the diversification of clownfishes. Nevertheless, a clear link between the function of these genes and their potential role in—or how they are affected by—clownfish adaptation to similar ecological niches cannot be drawn without a better characterization of clownfish functional traits.

It is worth mentioning that, in general, we detected stronger purifying selection in specialist species compared with generalists. This observation is in accord with theoretical expectations postulating that, overall, generalist species experience relaxed selection because they use multiple environments, leading to a decrease in the efficiency of selection (e.g., Kawecki 1994; Bono et al. 2020). Additionally, by considering the rationale behind the increased evolutionary rate reported for mutualistic ants (Rubin and Moreau 2016), generalist species might show a higher rate of sequence evolution as their environments are more changing than those of specialists (i.e., going in the direction of the Red Queen Hypothesis, Van Valen 1971).

#### Material and Methods

Whole-genome assemblies and annotations for the ten clownfish species (*A. akallopisos*, *A. bicinctus*, *A. frenatus*, *A. melanopus*, *A. nigripes*, *A. ocellaris*, *A. perideraion*, *A. polymnus*, *A. sebae*, and *P. biaculeatus*) and the lemon damselfish (*P. moluccensis*) were taken from public repositories (Marcionetti et al. 2018: DRYAD Repository: <https://doi.org/10.5061/dryad.nv1sv>; Marcionetti et al. 2019: Zenodo Repository <https://doi.org/10.5281/zenodo.2540241>). We classified clownfish species depending on the number of interacting sea anemone species, resulting in either specialist (up to two sea anemone hosts) or

generalist (more than two sea anemone hosts) species (supplementary table S10, Supplementary Material online). Although the differential host and habitat use is more complex than this dichotomy, our classification separated clownfish species on the two principal axes of variation describing the mutualistic interaction and the morphological differentiation of the species (Litsios et al. 2012).

### Mitochondrial Genome Reconstruction

The mitochondrial genome of *A. frenatus* was available from Marcionetti et al. (2018; DRYAD Repository: <https://doi.org/10.5061/dryad.nv1sv>). We performed mitochondrial genome reconstruction of the nine additional clownfish species and the outgroup *P. moluccensis* as in Marcionetti et al. (2018). Briefly, we retrieved the sequenced reads from the SRA database (NCBI, BioProject ID: PRJNA515163), we randomly subsampled 20 million reads for each species, and we assembled them using MITObim (v.1.9; Hahn et al. 2013). We employed two different reconstruction methods, using either available mitochondrial genomes or barcode sequences to initiate the assembly (NCBI accession IDs are reported in supplementary table S13, Supplementary Material online). We confirmed the consistency of the two reconstruction methods and inferred the circularity of the sequences with Geneious (v.10.2.2; Kearse et al. 2012).

### Ortholog Inference, Codon Alignments, and Gene Tree Reconstruction

Orthologous genes between *P. moluccensis*, the 10 clownfish species, and 12 additional Actinopterygii species (*Astyanax mexicanus*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Lepisosteus oculatus*, *O. niloticus*, *Oryzias latipes*, *Poecilia formosa*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Xiphophorus maculatus*, and *Stegastes partitus*) were obtained as reported in Marcionetti et al. (2019). We performed orthologous inference with OMA standalone (v.1.0.6, Altenhoff et al. 2013). We filtered the results to keep only hierarchical orthologous groups (HOGs) composed of both clownfish and outgroup species (total of 15,940 HOGs). We then classified HOGs as single-copy orthologous genes (1-to-1 OGs, total of 13,500) or multicopy HOGs (total of 2,725). All HOGs were functionally annotated by appending the function of the genes composing them.

For each HOG, we obtained protein alignments with MAFFT (v.7.130; Katoh and Standley 2016), using the G-INS-i strategy and controlling for overalignment with the *-allowshift* option. We obtained codon alignments from protein alignments using PAL2NAL (Suyama et al. 2006). Because alignment errors can affect positive selection analyses (Fletcher and Yang 2010), we filtered alignments to keep only high-confidence homologous regions,

following the approach used in the Selectome database (Moretti et al. 2014).

We reconstructed the gene tree of each HOG from the unfiltered codon alignments using PhyML (v3.3; Guindon et al. 2010), applying both the HKY85 and GTR substitution models (100 bootstraps). We selected the best model with a likelihood ratio test (LRT) ( $df=4$ ). Unfiltered alignments were preferred to filtered alignments as the filtering steps frequently worsen phylogenetic inference (Tan et al. 2015).

### Mitochondrial and Nuclear Phylogenetic Trees

We aligned the mitochondrial genomes of the ten clownfish species and the outgroup *P. moluccensis* with MAFFT (default parameters; v.7.450; Katoh and Standley 2016). We visually checked the alignments to avoid poorly aligned regions and reconstructed the mitochondrial phylogenetic tree with RAxML (GTR+ $\Gamma$  model, 100 bootstrap replicates; v.8.2.12; Stamatakis 2014). The nuclear phylogenetic tree was reconstructed with ASTRAL-III (v.5.7.8, Zhang et al. 2018) from the gene trees of 13,500 1-to-1 OGs (see above). Because contracting very low support branches in gene trees improves ASTRAL-III accuracy (Zhang et al. 2018), branches with bootstrap support lower than 10% were set to a length of zero using the Newick Utilities (Junier and Zdobnow 2010). We plotted the mitochondrial and nuclear phylogenetic trees with the *cophylo* command of the R package phytools (v.0.6.44; Revell 2012). The *P. moluccensis* individual was used to root the phylogenetic trees and was then removed from the plot.

### Topology Inconsistency along the Genome

We considered scaffolds larger than 100 kb to reduce their number (2,508 scaffolds kept out of 17,801), while keeping most of the genomic information (80% of the total assembly length; supplementary table S14, Supplementary Material online). We considered only scaffolds present in all clownfish species and *P. moluccensis*, with the latter being used to root the phylogenetic trees.

We aligned scaffolds with MAFFT (*-auto* parameter; v.7.130; Katoh and Standley 2016), and we filtered the alignments with trimAl (*-gappout*; v.1.4.1; Capella-Gutiérrez et al. 2009) to remove poorly aligned and gap-rich regions (supplementary table S14, Supplementary Material online). We split the filtered alignments to obtain nonoverlapping windows of 10, 50, or 100 kb. We reconstructed phylogenetic trees for each window using PhyML (GTR+ $\Gamma$  model, 100 bootstraps; v3.3; Guindon et al. 2010). We validated the support obtained for the trees by plotting the distribution of the bootstrap values of all nodes of all the trees and by calculating the average bootstrap support for each window (supplementary fig. S13, Supplementary Material online).

We visually investigated the tree topologies with DensiTree (v.2.2.5; Bouckaert 2010). We summarized the topologies using the treespace R package (v.1.1.3; Jombart et al. 2017) by calculating the Robinson–Foulds distances (method = RF) before applying metric multidimensional scaling (MDS). We identified groups of similar trees by hierarchical clustering, using the function *findGroves* in the treespace R package (cutoff set to 380, see cluster dendrogram in [supplementary fig. S3, Supplementary Material](#) online). We plotted trees with the ggtree R package (v.1.14.6; Yu et al. 2017). Results obtained with the different window sizes (10, 50, and 100 kb) were consistent, and we only considered windows of 100 kb (total of 5,936 trees).

We mapped the scaffolds of *A. frenatus* against the chromosome-level assembly of *A. percula* (Lehmann et al. 2019; downloaded from Ensembl, Assembly AmpOce1.0, GCA\_002776465.1) with blast (blastn v.2.7.1; <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) to identify the genomic locations of the alternative topologies. We linked the scaffolds to the best chromosome hit and transferred the information of topological support for each window on the corresponding chromosome. We performed GO enrichment analysis for specific regions of the genome (see below) using the topGO package (v.2.26.0; Alexa and Rahnenfuhrer 2016) available in Bioconductor (<http://www.bioconductor.org>), setting a minimum node size of 10. Fisher's exact tests with the weight01 algorithms were applied to examine the significance of enrichment, with  $P < 0.01$  considered significant. Following recommendations from the topGO manual, we present here raw  $P$  values rather than  $P$  values corrected for multiple testing.

### Detection of Hybridization

We tested for hybridization events in the diversification of clownfishes using the software HyDe (Blischak et al. 2018). HyDe analyses calculate the probability of hybridization of three taxa in respect to an outgroup (here, *P. moluccensis*) using phylogenetic invariants. As we wanted to test for hybridization at deep nodes of the tree (where topological inconsistencies in the branching of clades was observed, see results), we considered the pairs of species as populations. We excluded the pair *P. biaculeatus* and *A. ocellaris* as these species were consistently basal to other clownfishes. We tested the 12 possible triple comparisons among clades using the script *run\_hyde.py* on the concatenated alignments of all scaffolds (i.e., "whole genome") and on the concatenation of the alignments per chromosomes (i.e., "by chromosome"). We only considered hypothesis tests significant at the  $\alpha < 0.05$  level (after a Bonferroni correction) with an estimate of  $\gamma$  between 0.1 and 0.9. For the clades showing significant levels of hybridization, we then run *individual\_hyde.py* to test for

hybridization at the species level. Additionally, for the significant results, we computed Patterson's  $D$  statistics (Patterson et al. 2012) and tested its significance by using a  $\chi^2$  test to assess if the proportion of ABBA and BABA sites was significantly different.

### TE Annotation and Analyses

We identified de novo families of TE for the ten clownfish species and *P. moluccensis* with RepeatModeler (v.1.0.11; engine ncbi; <http://www.repeatmasker.org/RepeatModeler/>), and we classified them with RepeatClassifier (within RepeatModeler) and TEClass (v.2.1.3, Abrusán et al. 2009). We complemented the TE libraries with those of publicly available teleosts (*A. mexicanus*, *D. rerio*, *G. aculeatus*, *G. morhua*, *O. latipes*, *O. niloticus*, and *P. formosa*) downloaded from <http://www.fishedb.org/> (Shao et al. 2018). We used these libraries to annotate the TEs in clownfish and *P. moluccensis* genomes with RepeatMasker (v.4.0.7, <http://www.repeatmasker.org/>; Smit et al. 2015). We obtained information of the TE content in *A. percula* (Lehmann et al. 2019), *O. niloticus* (Brawand et al. 2014), *T. nigroviridis*, *G. aculeatus*, and *D. rerio* (Gao et al. 2016) for comparison with the TE content in clownfishes. We investigated the transposition history in clownfishes by performing a copy divergence analysis of the TE superfamilies based on the Kimura 2-parameter distance ( $K$ -values) that was obtained with the scripts *calcDivergenceFromAlign.pl* and *createRepeatLandscape.pl* provided in the RepeatMasker util directory.

### Gene Duplication Analyses

We investigated the gene duplication events occurring during the diversification of clownfishes and the other fish species *P. moluccensis*, *S. partitus*, *O. niloticus*, *G. aculeatus*, and *T. nigroviridis*. We retrieved 2,725 multicopy HOGs and filtered out the gene copies of the additional outgroup species not considered here. We employed a PDA approach (as in Brawand et al. 2014), counting the number of gene duplication events observed at each branch of the phylogenetic tree of these species (see Supplementary Information S1 for details). We estimated the neutral genomic divergence between the species with RAXML (GTR + $\Gamma$  model, 100 bootstraps; v.8.2.12; Stamatakis 2014), using approximately 7.5 million 4-fold degenerate sites that we obtained from the codon alignments of 1-to-1 OGs. The number of duplication events on each branch was then normalized by the neutral divergence between species ([Supplementary Information S1](#)).

### Overall Rate of Evolution of Clownfish Genes

We explored the overall rate of evolution of clownfish genes by estimating the ratio of nonsynonymous over synonymous substitutions ( $\omega$  or Ka/Ks or dN/dS) in clownfish compared with the outgroup species. Values of  $\omega$  smaller,

equal, or larger than 1 correspond, respectively, to purifying selection, neutral evolution, and positive selection. We randomly picked 20 1-to-1 OGs and concatenated their alignments. We repeated the procedure 50 times to obtain 50 concatenated alignments of 20 randomly picked genes. For each alignment, we reconstructed the gene tree using PhyML (GTR model, 100 bootstraps v3.3; Guindon et al. 2010). We estimated the  $\omega$  ratio in clownfishes and outgroups using the branch model implemented in *codeml* (PAML, v.4.9; Yang 2007). We tested for a significant difference between the  $\omega$  estimates obtained for the clownfishes and those obtained for the outgroups with a Welch two sample *t*-test.

### Positive Selection Analyses on the Whole Clownfish Group

We tested for the presence of positive selection in the whole clownfish clade using the branch–site model implemented in *codeml* (PAML, v.4.9; Yang 2007) on the filtered codon alignments of 13,500 1-to-1 OGs (fig. 3A and [supplementary Information S2, Supplementary Material](#) online). All branches of the clownfish group were set as foreground branches, while the remaining branches were assigned to the background (fig. 3A). In the null model, the foreground  $\omega$  was constrained to be smaller or equal to 1, while in the alternative model, the foreground  $\omega$  was estimated but was forced to be larger than 1 (fig. 3A). For each 1-to-1 OG, the best model was determined with a LRT ( $df = 1$ ). We corrected *P* values for multiple testing with the Benjamin–Hochberg method implemented in the *qvalue* package in R, following the approach used in the Selectome database (Moretti et al. 2014; false discovery rate (FDR) level of 0.1, the “robust” option, and the “bootstrap” method). To account for convergence issues encountered in the likelihood optimization in branch–site tests (Yang and Dos Reis 2010), we fitted both the null and the alternative models three times.

To investigate the type I error and power to detect positive selection, we simulated data with *evolver* (PAML; v.4.9; Yang 2007), following the approach used in Marcionetti et al. (2018). We simulated alignments of 1,000 codons under the branch–site model. We generated 20 trees with the species tree topology by assigning branch lengths randomly drawn from the branch length distribution of all the analyzed 1-to-1 OG gene trees. For each tree, we simulated six sets of alignments with either no ( $\omega_2 = 0.5$  and  $\omega_2 = 1$ ) or an increasing level ( $\omega_2$  of 2, 5, 10, and 20) of positive selection. We used the same pipeline as described above to estimate positive selection. We examined the type I errors (number of false positives) and the power to detect positive selection by looking at the number of significant LRT ( $P < 0.05$ ).

We retrieved the position of the genes under positive selection in the *A. percula* chromosomes (see *Topology*

*inconsistency along the genome*). We performed GO enrichment analysis on the positively selected genes with the TopGO package (v.2.26.0; Alexa and Rahnenfuhrer 2016), following the same procedure described above.

We investigated the presence of positive selection on some copies of the multicopy HOGs using aBSREL from HyPhy (v.2.3.7; Smith et al. 2015). We tested for positive selection at each branch of the tree, in an exploratory way. Although this method results in a decreased power due to multiple testing, it was chosen as we did not know beforehand which copy of the genes could be positively selected. We corrected for multiple testing with the Benjamin–Hochberg method implemented in the *qvalue* package in R (FDR threshold of 0.1, “robust” option, and “bootstrap” method; Dabney et al. 2010).

### Selection Signatures Associated with Host and Habitat Divergence

We tested the 1-to-1 OGs for signatures of selection associated with host and habitat divergence using the clade model C implemented in *codeml* (PAML v.4.9; Yang 2007; fig. 3B, [Supplementary Information S1](#)). Clade models test for differences in selective constraints among clades in a tree (Bielawski and Yang 2004; Weadick and Chang 2012). We labelled the terminal branches of clownfishes according to their host usage (e.g., specialists/generalists; [supplementary table S12, Supplementary Material](#) online and fig. 3B), and the remaining branches were assigned to the background category. We fitted the clade model C, estimating separate  $\omega$  ratios for each category, and we compared it with a LRT ( $df = 2$ ) to the null model M2a\_rel, where  $\omega$  is fixed among clades (fig. 3B). The resulting *P* values were corrected for multiple testing with the Benjamin–Hochberg method implemented in the *qvalue* R package (FDR threshold of 0.1, “robust” option, and “bootstrap” method; Dabney et al. 2010). To account for convergence issues potentially encountered in the likelihood optimization, we ran both models five times and kept the run resulting in the best likelihood.

Significant genes ( $q$ -value  $< 0.01$ ) were classified into two main groups. The first group consisted of genes under positive selection in specialist ( $\omega_{\text{specialists}} > 1.5$ ;  $\omega_{\text{generalists}}$  and  $\omega_{\text{background}} \leq 1$ ) or generalist ( $\omega_{\text{generalists}} > 1.5$ ;  $\omega_{\text{specialists}}$  and  $\omega_{\text{background}} \leq 1$ ) species. We selected  $\omega > 1.5$  as the threshold (instead of simply  $> 1$ ) to exclude genes with a weak signal of selection and avoid false positives. The second group was composed by genes showing different intensities of purifying selection in specialists and generalists ( $\omega_{\text{specialists}}$  and  $\omega_{\text{generalists}} < 1$ , but  $\omega_{\text{specialists}} \neq \omega_{\text{generalists}}$ ). We estimated the distribution of  $\omega_{\text{background}}$ ,  $\omega_{\text{specialists}}$ , and  $\omega_{\text{generalists}}$  across all genes, and we identified genes experiencing intensified purifying selection in either specialist or generalist species as the genes falling in the

lower 10% of the  $\omega_{\text{specialists}}$  or  $\omega_{\text{generalists}}$  distribution, respectively, but in the upper 10% of the remaining two other  $\omega$  distributions (supplementary fig. S14A and B, Supplementary Material online). Likewise, genes experiencing relaxed purifying selection in either specialist or generalist species were retrieved by considering those genes falling in the upper 10% of the  $\omega_{\text{specialists}}$  or  $\omega_{\text{generalists}}$  distributions, respectively, but in the lower 10% of the remaining two other  $\omega$  distributions (Supplementary supplementary fig. S14C and D, Supplementary Material online). Analyzing these results by considering the estimated  $\omega_{\text{background}}$  of each gene was necessary to differentiate between intensified purifying selection in one group and relaxed purifying selection in the other. We performed GO enrichment analyses on the identified genes with the TopGO package (v.2.26.0; Alexa and Rahnenfuhrer 2016), following the same procedure described above.

### Evolutionary Rate Linked with Host and Habitat Usage

We retrieved the reconstructed gene trees for each 1-to-1 OG (see Supplementary Information S1), and we considered the branch length as a proxy of the evolutionary rate of the OGs. We investigated the presence of OGs with an evolutionary rate linked with host and habitat usage by calculating the difference in branch length between generalists and specialists in each pair of closely related species (see Supplementary Information S3 for details).

### Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

### Acknowledgments

We thank B. Micheli and S. Schmid for the general support. We also thank the DCSR infrastructure of the University of Lausanne. Funding was from the University of Lausanne funds, Swiss National Science Foundation, Grant Number: 31003A-163428.

### Author Contributions

A.M. and N.S. designed the research, A.M. performed the analyses, A.M. and N.S. wrote the paper.

### Data Availability

Whole-genome assemblies and annotations for the ten clownfish species and the lemon damselfish were taken from public repositories (Marcionetti et al. 2018: DRYAD Repository: <https://doi.org/10.5061/dryad.nv1sv>; Marcionetti et al. 2019: Zenodo Repository <https://doi.org/10.5281/zenodo.2540241>). The mitochondrial genome of *A. frenatus* was available from Marcionetti et al. (2018; DRYAD

Repository: <https://doi.org/10.5061/dryad.nv1sv>). The mitochondrial genome assemblies for the additional species are available in the Zenodo repository (<https://doi.org/10.5281/zenodo.7863281>).

### Literature Cited

- Abrusán G, Grundmann N, DeMester L, Makalowski W. 2009. TEclass—a tool for automated classification of unknown eukaryotic transposable elements. *Bioinformatics* 25(10):1329–1330.
- Alexa A, Rahnenfuhrer J. 2016. topGO: Enrichment Analysis for Gene Ontology. R package version 2.26.0. *Bioconductor Improv.* 27:1–26.
- Altenhoff AM, Gil M, Gonnet GH, Dessimoz C. 2013. Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS One.* 8:e53786.
- Arai R. 2011. *Fish karyotypes: a check list.* Springer Science & Business Media.
- Auvinet J, et al. 2018. Mobilization of retrotransposons as a cause of chromosomal diversification and rapid speciation: the case for the Antarctic teleost genus *Trematomus*. *BMC Genomics.* 19(1):339.
- Benvenuto C, Coscia I, Chopelet J, Sala-Bozano M, Mariani S. 2017. Ecological and evolutionary consequences of alternative sex-change pathways in fish. *Sci Rep.* 7(1):9084.
- Bergstrom CT, Lachmann M. 2003. The Red King effect: when the slowest runner wins the coevolutionary race. *Proc Natl Acad Sci U S A.* 100(2):593–598.
- Berner D, Salzburger W. 2015. The genomics of organismal diversification illuminated by adaptive radiations. *Trends Genet.* 31(9):491–499.
- Bielawski JP, Yang Z. 2004. A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J Mol Evol.* 59(1):121–132.
- Blackledge TA, Gillespie RG. 2004. Convergent evolution of behavior in an adaptive radiation of Hawaiian web-building spiders. *Proc Natl Acad Sci U S A.* 101(46):16228–16233.
- Blischak PD, Chifman J, Wolfe AD, Kubatko LS. 2018. Hyde: a Python package for genome-scale hybridization detection. *Syst Biol.* 67(5): 821–829.
- Bono LM, Draghi JA, Turner PE. 2020. Evolvability costs of niche expansion. *Trends Genet.* 36(1):14–23.
- Bouckaert RR. 2010. Densitree: making sense of sets of phylogenetic trees. *Bioinformatics* 26(10):1372–1373.
- Bowmaker JK. 2008. Evolution of vertebrate visual pigments. *Vision Res.* 48(20):2022–2041.
- Branco S, et al. 2018. Multiple convergent supergene evolution events in mating-type chromosomes. *Nat Commun.* 9(1):1–13.
- Brawand D, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513(7518):375–381.
- Bromham L. 2009. Why do species vary in their rate of molecular evolution? *Biol Lett.* 5(3):401–404.
- Bromham L, Cowman PF, Lanfear R. 2013. Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC Evol Biol.* 13(1):1–11.
- Buston P. 2003. Size and growth modification in clownfish. *Nature* 424(6945):145–146.
- Buston PM, García MB. 2007. An extraordinary life span estimate for the clown anemonefish *Amphiprion percula*. *J Fish Biol.* 70(6): 1710–1719.
- 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.
- Casas L, Saenz-Agudelo P, Irigoien X. 2018. High-throughput sequencing and linkage mapping of a clownfish genome provide insights on the distribution of molecular players involved in sex change. *Sci Rep.* 8(1):1–12.

- Chausson J, Srinivasan M, Jones GP. 2018. Host anemone size as a determinant of social group size and structure in the orange clownfish (*Amphiprion percula*). *PeerJ* 6:e5841.
- Chopra K, Ishibashi S, Amaya E. 2019. Zebrafish duox mutations provide a model for human congenital hypothyroidism. *Biol Open*. 8:2.
- Coscia I, Choquet J, Waples RS, Mann BQ, Mariani S. 2016. Sex change and effective population size: implications for population genetic studies in marine fish. *Heredity (Edinb)*. 117(4):251–258.
- Cresko WA, et al. 2004. Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proc Natl Acad Sci U S A*. 101(16):6050–6055.
- Dabney A, Storey JD, Warnes GR. (2010). qvalue: Q-value estimation for false discovery rate control. R package version 1.38.0, <https://github.com/StoreyLab/qvalue>
- Dasmahapatra KK, et al. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487-(7405):94–98.
- Edelman NB, et al. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* 366(6465):594–599.
- Fautin DG, Allen GR. 1997. Anemone fishes and their host sea anemones. Perth: Western Australian Museum. p. 159.
- Feiner N. 2016. Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards. *Proc Royal Soc B Biol Sci*. 283(1840):20161555.
- Fletcher W, Yang Z. 2010. The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Mol Biol Evol*. 27(10):2257–2267.
- Frédérich B, Sorenson L, Santini F, Slater GJ, Alfaro ME. 2013. Iterative ecological radiation and convergence during the evolutionary history of damselfishes (Pomacentridae). *Am Nat*. 181(1):94–113.
- Fricke HW. 1979. Mating system, resource defence and sex change in the anemonefish *Amphiprion akallopisos*. *Zeitschrift für Tierpsychologie*. 50(3):313–326.
- Froese R, Pauly D. 2000. FishBase 2000: concepts designs and data sources (Vol. 1594). WorldFish.
- Gainsford A, Van Herwerden L, Jones GP. 2015. Hierarchical behaviour, habitat use and species size differences shape evolutionary outcomes of hybridization in a coral reef fish. *J Evol Biol*. 28(1):205–222.
- Gao B, et al. 2016. The contribution of transposable elements to size variations between four teleost genomes. *Mob DNA*. 7(1):4.
- Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 59(3):307–321.
- Hahn C, Bachmann L, Chevreaux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res*. 41(13):e129.
- Hofmann HA, Fernald RD. 2000. Social status controls somatostatin neuron size and growth. *J Neurosci*. 20(12):4740–4744.
- Iwata E, Nagai Y, Hyoudou M, Sasaki H. 2008. Social environment and sex differentiation in the false clown anemonefish, *Amphiprion ocellaris*. *Zool Sci*. 25(2):123–128.
- Jombart T, Kendall M, Almagro-Garcia J, Colijn C. 2017. Treespace: statistical exploration of landscapes of phylogenetic trees. *Mol Ecol Resour*. 17(6):1385–1392.
- Jones FC, et al. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55–61.
- Joron M, et al. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 477-(7363):203–206.
- Junier T, Zdobnov EM. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics* 26-(13):1669–1670.
- Katoh K, Standley DM. 2016. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* 32(13):1933–1942.
- Kawecki TJ. 1994. Accumulation of deleterious mutations and the evolutionary cost of being a generalist. *Am Nat*. 144(5):833–838.
- Kearse M, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.
- Kelley JL, Fitzpatrick JL, Merilaita S. 2013. Spots and stripes: ecology and colour pattern evolution in butterflyfishes. *Proc Royal Soc B Biol Sci*. 280(1757):20122730.
- Kingman GAR, Vyas DN, Jones FC, Brady SD, Chen HI, Reid K, Veeramah KR. 2021. Predicting future from past: the genomic basis of recurrent and rapid stickleback evolution. *Sci Adv*. 7(25):eabg5285.
- Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Royal Soc B Biol Sci*. 279(1749):5048–5057.
- Küpper C, et al. 2016. A supergene determines highly divergent male reproductive morphs in the ruff. *Nat Genet*. 48(1):79–83.
- Lehmann R, Lightfoot DJ, Schunter C, Michell CT, Ohyanagi H, Mineta K, Gojobori T. 2019. Finding Nemo's Genes: a chromosome-scale reference assembly of the genome of the orange clownfish *Amphiprion percula*. *Mol Ecol Resour*. 19(3):570–585.
- Li Q, et al. 2019. Evidence for the direct effect of the NPFF peptide on the expression of feeding-related factors in spotted sea bass (*Lateolabrax maculatus*). *Front Endocrinol*. 10:545.
- Litsios G, et al. 2012. Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evol Biol*. 12(1):212.
- Litsios G, Pearman PB, Lanterbecq D, Tolou N, Salamin N. 2014. The radiation of the clownfishes has two geographical replicates. *J Biogeogr*. 41(11):2140–2149.
- Litsios G, Salamin N. 2014. Hybridisation and diversification in the adaptive radiation of clownfishes. *BMC Evol Biol*. 14(1):245.
- Lönnig WE, Saedler H. 2002. Chromosome rearrangements and transposable elements. *Annu Rev Genet*. 36(1):389–410.
- Losos JB. 2011. Convergence, adaptation, and constraint. *Evolution* 65(7):1827–1840.
- Lutzoni F, Pagel M. 1997. Accelerated evolution as a consequence of transitions to mutualism. *Proc Natl Acad Sci*. 94(21):11422–11427.
- Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–1155.
- Lynch M, Walsh B. 2007. The origins of genome architecture (vol. 98). Sunderland, MA: Sinauer associates.
- Marcionetti A, et al. 2019. Insights into the genomics of clownfish adaptive radiation: genetic basis of the mutualism with sea anemones. *Genome Biol evol*. 11(3):869–882.
- Marcionetti A, Rossier V, Bertrand JA, Litsios G, Salamin N. 2018. First draft genome of an iconic clownfish species (*Amphiprion frenatus*). *Mol Ecol Resour*. 18(5):1092–1101.
- Marques DA, Meier JI, Seehausen O. 2019. A combinatorial view on speciation and adaptive radiation. *Trends Ecol Evol*. 34(6):531–544.
- Martin CH, Richards EJ. 2019. The paradox behind the pattern of rapid adaptive radiation: how can the speciation process sustain itself through an early burst? *Annu Rev Ecol Syst*. 50:569–593.
- McGee MD, et al. 2020. The ecological and genomic basis of explosive adaptive radiation. *Nature* 586(7827):75–79.
- Meier JI, et al. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat Commun*. 8(1):1–11.
- Moretti S, et al. 2014. Selectome update: quality control and computational improvements to a database of positive selection. *Nucleic Acids Res*. 42(D1):D917–D921.

- Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. *Curr Biol*. 22(24):2362–2368.
- na Ayudhaya PT, et al. 2017. Unveiling cryptic diversity of the anemonefish genera *Amphiprion* and *Premnas* (Perciformes: Pomacentridae) in Thailand with mitochondrial DNA barcodes. *Agric Nat Resour*. 51(3):198–205.
- Natri HM, Merilä J, Shikano T. 2019. The evolution of sex determination associated with a chromosomal inversion. *Nat Commun*. 10(1):1–13.
- Paterson S, et al. 2010. Antagonistic coevolution accelerates molecular evolution. *Nature* 464(7286):275–278.
- Patterson N, et al. 2012. Ancient admixture in human history. *Genetics* 192(3):1065–1093.
- Reddon AR, et al. 2015. Brain nonapeptide levels are related to social status and affiliative behaviour in a cooperatively breeding cichlid fish. *Royal Soc Open Sci*. 2(2):140072.
- Reddon AR, O'Connor CM, Marsh-Rollo SE, Balshine S. 2012. Effects of isotocin on social responses in a cooperatively breeding fish. *Animal Behav*. 84(4):753–760.
- Revell LJ. 2012. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3(2):217–223.
- Ronco F, et al. 2021. Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature* 589(7840):76–81.
- Rosenblum EB, Parent CE, Brandt EE. 2014. The molecular basis of phenotypic convergence. *Annu Rev Ecol Evol Syst*. 45:203–226.
- Rubin BE, Moreau CS. 2016. Comparative genomics reveals convergent rates of evolution in ant-plant mutualisms. *Nat Commun*. 7(1):1–11.
- Salis P, et al. 2021. Thyroid hormones regulate the formation and environmental plasticity of white bars in clownfishes. *Proc Natl Acad Sci U S A*. 118:23.
- Schluter D. 2000. *The ecology of adaptive radiation*. Oxford: Oxford University Press.
- Schott RK, Refvik SP, Hauser FE, López-Fernández H, Chang BS. 2014. Divergent positive selection in rhodopsin from lake and riverine cichlid fishes. *Mol Biol Evol*. 31(5):1149–1165.
- Seehausen O. 2004. Hybridization and adaptive radiation. *Trends Ecol Evol*. 19(4):198–207.
- Shao F, Han M, Peng Z. 2019. Evolution and diversity of transposable elements in fish genomes. *Sci Rep*. 9(1):15399.
- Shao F, Wang J, Xu H, Peng Z. 2018. FishTEDB: a collective database of transposable elements identified in the complete genomes of fish. *Database* 2018:bax106.
- Sheridan MA, Hagemeister AL. 2010. Somatostatin and somatostatin receptors in fish growth. *Gen Comp Endocrinol*. 167(3):360–365.
- Simpson GG. 1953. *The major features of evolution*. New York: Columbia University Press.
- Smit A, Hubley R, Green P. 2015. (2013) RepeatMasker Open-4.0. 2013. Institute of Systems Biology. <http://www.repeatmasker.org/>
- Smith MD, et al. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol*. 32(5):1342–1353.
- Spady TC, et al. 2005. Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Mol Biol Evol*. 22(6):1412–1422.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30(9):1312–1313.
- Stevison LS, Hoehn KB, Noor MA. 2011. Effects of inversions on within-and between-species recombination and divergence. *Genome Biol Evol*. 3:830–841.
- Stroud JT, Losos JB. 2016. Ecological opportunity and adaptive radiation. *Annu Rev Ecol Evol Syst*. 47:507–532.
- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(suppl\_2):W609–W612.
- Svardal H, et al. 2020. Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Mol Biol Evol*. 37(4):1100–1113.
- Tan G, et al. 2015. Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Syst Biol*. 64(5):778–791.
- Tang KL, Stiassny ML, Mayden RL, DeSalle R. 2021. Systematics of Damsel-fishes. *Ichthyol Herpetol*. 109(1):258–318.
- Thomas JA, Welch JJ, Lanfear R, Bromham L. 2010. A generation time effect on the rate of molecular evolution in invertebrates. *Mol Biol Evol*. 27(5):1173–1180.
- Tollis M, et al. 2018. Comparative genomics reveals accelerated evolution in conserved pathways during the diversification of anole lizards. *Genome Biol Evol*. 10(2):489–506.
- Trainor BC, Hofmann HA. 2006. Somatostatin regulates aggressive behavior in an African cichlid fish. *Endocrinology* 147(11):5119–5125.
- Trainor BC, Hofmann HA. 2007. Somatostatin and somatostatin receptor gene expression in dominant and subordinate males of an African cichlid fish. *Behav Brain Res*. 179(2):314–320.
- Van Valen L. 1973. A new evolutionary law. *Evol Theory*. 1:1–30.
- Vizueta J, Macías-Hernández N, Arnedo MA, Rozas J, Sánchez-Gracia A. 2019. Chance and predictability in evolution: the genomic basis of convergent dietary specializations in an adaptive radiation. *Mol Ecol*. 28(17):4028–4045.
- Waples RS, Mariani S, Benvenuto C. 2018. Consequences of sex change for effective population size. *Proc Royal Soc B Biol Sci*. 285-(1893):20181702.
- Weadick CJ, Chang BS. 2012. An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol Biol Evol*. 29(5):1297–1300.
- Welch JJ, Bininda-Emonds OR, Bromham L. 2008. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol Biol*. 8(1):1–12.
- Woolfit M, Bromham L. 2003. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Mol Biol Evol*. 20(9):1545–1555.
- Woolfit M, Bromham L. 2005. Population size and molecular evolution on islands. *Proc Royal Soc B Biol Sci*. 272(1578):2277–2282.
- Xiong P, et al. 2021. The comparative genomic landscape of adaptive radiation in crater lake cichlid fishes. *Mol Ecol*. 30(4):955–972.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24(8):1586–1591.
- Yang Z, Dos Reis M. 2010. Statistical properties of the branch-site test of positive selection. *Mol Biol Evol*. 28(3):1217–1228.
- Yoshizawa K, Johnson KP. 2003. Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. *Mol Phylogenet Evol*. 29(1):102–114.
- Yu G, Smith DK, Zhu H, Guan Y, Lam TTY. 2017. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 8(1):28–36.
- 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.

Associate editor: Prof. Judith Mank