



UNIL | Université de Lausanne

Unicentre

CH-1015 Lausanne

<http://serval.unil.ch>

Year : 2021

Genomics of clownfish adaptive radiation

Marcionetti Anna

Marcionetti Anna, 2021, Genomics of clownfish adaptive radiation

Originally published at : Thesis, University of Lausanne

Posted at the University of Lausanne Open Archive <http://serval.unil.ch>

Document URN : urn:nbn:ch:serval-BIB_7360DFA238AF8

Droits d'auteur

L'Université de Lausanne attire expressément l'attention des utilisateurs sur le fait que tous les documents publiés dans l'Archive SERVAL sont protégés par le droit d'auteur, conformément à la loi fédérale sur le droit d'auteur et les droits voisins (LDA). A ce titre, il est indispensable d'obtenir le consentement préalable de l'auteur et/ou de l'éditeur avant toute utilisation d'une oeuvre ou d'une partie d'une oeuvre ne relevant pas d'une utilisation à des fins personnelles au sens de la LDA (art. 19, al. 1 lettre a). A défaut, tout contrevenant s'expose aux sanctions prévues par cette loi. Nous déclinons toute responsabilité en la matière.

Copyright

The University of Lausanne expressly draws the attention of users to the fact that all documents published in the SERVAL Archive are protected by copyright in accordance with federal law on copyright and similar rights (LDA). Accordingly it is indispensable to obtain prior consent from the author and/or publisher before any use of a work or part of a work for purposes other than personal use within the meaning of LDA (art. 19, para. 1 letter a). Failure to do so will expose offenders to the sanctions laid down by this law. We accept no liability in this respect.



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Département de Biologie Computationnelle

Genomics of clownfish adaptive radiation

Thèse de doctorat ès sciences de la vie (PhD)

présentée à la

Faculté de biologie et de médecine
de l'Université de Lausanne

par

Anna MARCIONETTI

Master en Science Moléculaires du Vivant de l'Université de Lausanne

Jury

Prof. Jan Roelof van der Meer, Président

Prof. Nicolas Salamin, Directeur de thèse

Prof. Christophe Dessimoz, Expert

Prof. Walter Salzburger, Expert

Lausanne

2021



UNIL | Université de Lausanne

Faculté de biologie
et de médecine

Ecole Doctorale

Doctorat ès sciences de la vie

Imprimatur

Vu le rapport présenté par le jury d'examen, composé de

Président·e	Monsieur	Prof.	Jan Roelof	van der Meer
Directeur·trice de thèse	Monsieur	Prof.	Nicolas	Salamin
Expert·e·s	Monsieur	Prof.	Christophe	Dessimoz
	Monsieur	Prof.	Walter	Salzburger

le Conseil de Faculté autorise l'impression de la thèse de

Madame Anna Marcionetti

Maîtrise universitaire ès Sciences en sciences moléculaires du vivant, Université de Lausanne

intitulée

Genomics of clownfish adaptive radiation

Date de l'examen : 19 novembre 2021

Date d'émission de l'imprimatur : Lausanne, le 29 novembre 2021

pour le Doyen
de la Faculté de biologie et de médecine

Prof. Niko GELDNER
Directeur de l'Ecole Doctorale



“It is interesting to contemplate an entangled bank, clothed with many plants of many kinds, with birds singing on the bushes, with various insects flitting about, and with worms crawling through the damp earth, and to reflect that these elaborately constructed forms, so different from each other, and dependent on each other in so complex manner, have all been produced by laws acting around us. [...] There is a grandeur in this view of life, with its several powers, having been originally breathed into a few forms or into one; and that, whilst this planet has gone cycling on according to the fixed law of gravity, from so simple a beginning endless forms beautiful and most wonderful have been, and are being, evolved.”

Darwin, 1859

TABLE OF CONTENTS

Abstract	1
Résumé	3
Acknowledgements	5
General Introduction	7
Chapter 1. First draft genome of an iconic clownfish species (<i>Amphiprion frenatus</i>) . .	26
Chapter 2. Insights into the genomics of clownfish adaptive radiation: genetic basis of the mutualism with sea anemones	47
Chapter 3. Insights into the genomics of clownfish adaptive radiation: the genomic substrate of the diversification	78
Chapter 4. Genomics of the diversification of the clownfish skunk complex (<i>Amphiprion akallopisos</i> , <i>A. sandaracinos</i> and <i>A. perideraion</i>)	128
Discussion and perspectives	176
Appendices	193
Appendix 1	194
Appendix 2	195
Appendix 3	196
Appendix 4	197
Appendix 5	198

ABSTRACT

How species diversify, creating the astonishing biodiversity observed on Earth, has been a central question since Darwin's *On the Origin of Species*. Thanks to the sequencing revolution, this question can be approached today from a genomic perspective, examining how intrinsic genomic architecture and extrinsic biological and ecological factors interplay to shape the diversification of organisms. In this sense, clownfishes, which experienced an adaptive radiation following the acquisition of mutualism with sea anemones, represent a fascinating system. Thus, in this thesis, I combine comparative and population genomics approaches to study the genomic architecture underlying the diversification of this group. In the first two sections, I generate genomic resources for ten closely related but ecologically divergent clownfish species and the damselfish *Pomacentrus moluccensis*. Using this data, I question the genetic mechanisms underlying the acquisition of mutualism. I identify several candidate genes that experienced positive selection at the basis of clownfish radiation and show functions associated with sea anemones toxins discharge, thus likely involved in the evolution of clownfishes' ability to live unharmed within their otherwise-toxic hosts. In the last two sections, I dive into the diversification process of clownfishes. Through comparative genomics approaches, I show that the group experienced bursts of transposable elements, overall accelerated molecular evolution, and ancestral hybridization events, which likely facilitated the radiation of the group by generating the genomic variations necessary for natural selection to act on. I identify genes undergoing differential selective pressures linked with ecological divergence, suggesting that parallel evolution is shaping clownfish diversification, and I pinpoint candidate genes involved in the evolution of the particular size-based hierarchical social structure observed in the group. I finally focus on the mechanisms underlying the evolution of a clownfish clade, the skunk complex. Through population genomics approaches, I demonstrate that gene flow occurred throughout the diversification of the group. Indeed, the species experienced moderate ancestral gene flow, which lessened but still persists in sympatry. Moreover, contrary to what was previously suggested, I demonstrate that *A. sandaracinos* did not originate from hybrid speciation. I finally pinpoint candidate regions of introgression between species that likely played a role in the diversification of the complex. Overall, my work provides the first insights into the genomic mechanisms underlying clownfish adaptive radiation.

RÉSUMÉ

Comprendre comment les espèces se diversifient et créent l'étonnante biodiversité observée sur Terre sont des questions centrales depuis l'écriture de *l'Origine des espèces* par Darwin. Grâce à la révolution du séquençage, ces questions peuvent être abordées aujourd'hui en examinant comment l'architecture génomique et les facteurs biologiques et écologiques interagissent et mènent à la diversification des organismes. En ce sens, les poissons-clowns, qui ont connu une radiation adaptative suite à l'acquisition du mutualisme avec les anémones de mer, représentent un système fascinant. Ainsi, dans cette thèse, j'étudie l'architecture génomique qui sous-tend la diversification de ce groupe. Dans les deux premières sections, je génère des ressources génomiques pour dix espèces de poissons-clowns ainsi que pour l'espèce de demoiselle *Pomacentrus moluccensis*. À partir de ces données, je questionne les mécanismes génétiques qui sous-tendent l'acquisition du mutualisme. J'identifie plusieurs gènes ayant subi une sélection positive à la base du rayonnement des poissons-clowns. Ces gènes ont des fonctions associées à la décharge des toxines des anémones, suggérant donc une implication dans l'évolution de la capacité des poissons-clowns à vivre au sein de leurs hôtes normalement toxiques. Dans les deux dernières sections, je me plonge dans le processus de diversification des poissons-clowns. Grâce à des approches de génomique comparative, je montre non seulement que ce groupe compte une importante quantité d'éléments transposables au sein de son génome, mais qu'il a également subi une évolution moléculaire accélérée ainsi que des événements d'hybridation ancestrale. La combinaison de ces différents éléments a probablement facilité leur diversification en générant les variations génomiques nécessaires à l'action de la sélection naturelle. J'identifie également des gènes ayant subi des pressions de sélection différentielles en lien avec des divergences écologiques, suggérant donc un processus d'évolution parallèle impliquée dans la diversification des poissons-clowns. Finalement, je me suis concentrée sur les mécanismes liés à l'évolution d'un clade de poisson-clown - le clade "skunk". Je démontre qu'un flux de gènes modéré s'est produit tout au long de la diversification de ce groupe, et - bien que son amplitude ait diminué avec le temps - il persiste encore en sympatrie. De plus, j'identifie des régions candidates d'introgession entre espèces qui ont probablement joué un rôle dans la diversification du complexe. Dans l'ensemble, mon travail fournit les premières informations concernant les mécanismes génomiques impliqués dans la radiation adaptative des poissons-clowns.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deep gratitude to my supervisor Prof. Nicolas Salamin for giving me the opportunity of pursuing a Ph.D. For your support, guidance, knowledge, but also kindness and understanding, thank you from the bottom of my heart! I could not imagine having a better advisor and mentor for my Ph.D.

A massive thank you as well to all my incredible colleagues, past and present: Alberto, Arnaud, Baptiste, Benjamin, Daniele, Darina, Guanpeng, Iakov, Jonathan, Joris, Kana, Linda, Martha, Oriane, Pablo, Rocio, Sara, Sarah, Téó, Thibault, Victor, Virginie, Xavier. All of you contributed to creating a positive, pleasant, and enriching work environment. Thank you for the mutual help, the excellent mood, and the laughter. A special thanks to Martha, who has always believed in me. I would not have obtained this opportunity without you; thank you! A particular thank you also to Baptiste, Sara, and Sarah, who have been there to support (and bear) me when this work seemed interminable and giving up popped into my head. Thank you for your understanding, smiles, words, positive vibes and, simply, your important friendship!

I would also like to thank the researchers I had the chance to collaborate with during these years. In particular, Prof. Vincent Laudet, Dr. Natacha Roux, Dr. Pauline Salis and Dr. Mathieu Perret. Thank you also to the experts from my thesis committee: Prof. Christophe Dessimoz, Prof. Walter Salzburger, and Prof. Jan Roelof van der Meer.

A huge thank you also goes to my family, those people who are still and will always be with me whatever happened and whatever will happen. So, many thanks to: my mother Maria Luisa, my sister Beatrice, my brother in law Athos, and my almost-sister Semra. Grazie mille di avermi supportato e sopportato! Murat, with whom I spent most of the years of my Ph.D. Merci, je n'y serais pas arrivée sans toi! Dudo (Fabio) et Amel, with whom I have started this adventure. I could not be more grateful for your friendship! Merci d'avoir toujours été là pour moi! Letizia, who initially was my roommate but became so much more! Grazie zietta!

Last but not least, I thank all my amazing friends. Making a list of all of you would be too difficult, and I do not want this document to become even longer than it is already. So, thanks to all of you for the moments shared at the lake, at the mountain, at my or your place, at Zelig or the Great Escape for the *apéro* (and for the *babyfoot-time*), ... *Bref*, thanks to all of you!

GENERAL INTRODUCTION

Speciation and the buildup of biodiversity

Life today, as Darwin already noted, is extraordinarily diverse. The most precise estimate suggests the number of eukaryotic species on Earth is 8.7 million, with more than 80% of them still undiscovered (Mora et al., 2011). Therefore, it is not surprising that since Darwin's publication of *On the Origin of Species*, the question of how such diversity is created has always generated enthusiasm and passionate debate. To this day, understanding what shapes diversity is one of the most challenging research questions in ecology, evolution, genetics, and paleontology (Pennisi, 2005). Hence, the study of speciation – the process by which populations evolve to become distinct species – is crucial to identify the factors that promote biodiversity.

New species generally originate when substantial gene flow among populations stops and the isolated populations diverge, eventually reaching reproductive isolation. Historically, the geographical context in which this divergence occurs (allopatry, parapatry, or sympatry) was generally used to classify the different modes of speciation (Fitzpatrick et al., 2009). In the case of allopatric speciation, geographical isolation is primarily causing the interruption of gene flow among populations. Contrarily, in parapatric and sympatric speciation, extrinsic barriers to gene flow are respectively partial or absent. As a result, the existence of sympatric speciation has been widely questioned and remains a controversial topic (Foote, 2018), although empirical studies have provided evidence for this process (Foote, 2018). In this geographical context, allopatric speciation has been thus considered the most prevalent mode of speciation (Coyne & Orr, 2004).

Nevertheless, the continuum from absolute geographical isolation to complete sympatry is *de facto* only a poor proxy of the degree of occurring gene flow between populations (Fitzpatrick et al., 2009). Indeed, ecology, genetics, and the demographic history of populations are also crucial aspects influencing the degree of gene flow and the evolution of new species. Consequently, an effort to reclassify the modes of speciation based on the mechanisms underlying the evolution of reproductive isolation emerged since the early 2000s (Schluter, 2001; Rundle & Nosil, 2005; Schluter, 2009). In this sense, theoretical and empirical studies widely explored modes of speciation driven by ecologically based divergent selection (i.e., ecological speciation), by uniform selection (i.e., “mutation-order” speciation), but also by genetic drift, sexual selection, or hybridization (i.e., hybrid speciation) (reviewed

in Sobel et al., 2010; Gavrilets, 2014). Particular attention was nevertheless given to ecological speciation (Rundle and Nosil, 2005; Schluter, 2001, 2009; Nosil, 2012), the process underlying adaptive radiation, which is in turn considered of fundamental importance and potentially responsible for much of the diversity of life (Simpson, 1953; Schluter, 2000).

Adaptive radiations

Adaptive radiation is defined as the process by which a single ancestral species diversifies into many descendants adapted to a wide range of ecological conditions (Schluter, 2000). The process starts with an ecological opportunity, where the ancestral species has access to new, abundant, and underused resources (Losos, 2010; Yoder et al., 2010). Ecological opportunity can be caused by colonization of isolated areas, extinctions of competitors, or evolution of new traits (i.e., key innovation), allowing the interaction with the environment in novel ways (Losos, 2010; Yoder et al., 2010). Strong divergent natural selection acts in response to the difference in resources, rapidly driving the adaptation of populations to different ecological niches that ultimately lead to their speciation¹ (Schluter, 2000; Rundle & Price, 2009). This process results in the generation of species with distinct phenotypes specialized to different ecological niches (i.e., divergent evolution), but also the repeated and independent formation of similar phenotypes in species inhabiting ecologically similar habitats (i.e., convergent evolution, Schluter, 2000; Rundle & Price, 2009).

An example of adaptive radiation includes Darwin's finches on the Galapagos archipelago, which evolved specialized beaks morphologies adapted to different food resources (Grant & Grant, 2008). Similarly, in *Anolis* lizards, the divergence into ecomorphs adapted to distinct foraging niches occurred independently on different islands, resulting in phenotypically similar ecomorphs inhabiting equivalent niches (Losos et al., 1998). Furthermore, benthic and limnetic stickleback fish evolved repeatedly in postglacial lakes (Schluter & Nagel, 1995; Rundle et al., 2000), convergence in wing-color mimetics was observed in the highly diverse *Heliconius* butterfly (Turner, 1976), and in cichlids, convergent ecomorphs arose within (Muschick et al., 2012) and across (Kocher et al., 1993) East African Lakes.

Because of this dual nature, with convergence nested within divergence, adaptive radiation represents a natural framework to study not only the roles of ecology, selection, and

¹ Ecological speciation, which can be facilitated by hybridization (Seehausen, 2000, see below), or through sexual selection (Wagner et al., 2012)

adaptation in speciation and diversification but also the repeatability and predictability of evolution (Rosenblum et al., 2014, Kingman et al., 2021). The presence of phenotypic convergence interrogates the molecular processes underlying it and, notably, whether common genetic mechanisms are involved (i.e., parallel evolution, as defined in Rosenblum et al., 2014). Parallel evolution is likely in rapidly speciating lineages as they possess similar genomic backgrounds (Rosenblum et al., 2014). In such a case, the independently-evolved convergent phenotypes constitute evolutionary replicates that may facilitate the identification of loci involved in the adaptation to similar ecological conditions. For instance, parallel evolution was observed in sticklebacks, where the repeated loss of lateral plates in similar ecomorphs resulted from selective pressures on the same genes (Cresko et al., 2004; Colosimo et al., 2005; Chan et al., 2010). Similarly, the divergence in opsin genes in Lake Victoria cichlids (Seehausen et al., 2008) promoted recurrent adaptation to different light environments, while the *optix* gene drove the repeated convergent evolution of wing patterns mimicry in *Heliconius* (Reed et al., 2011).

Another characteristic of adaptive radiation is its speed: diversification and speciation occur rapidly. Several lineage-specific traits and properties may contribute to high speciation rates, such as sexual selection, its interaction with ecological opportunity, or the acquisition of key innovations (Schluter, 2000; Losos, 2010; Wagner, 2012). Nevertheless, these properties necessitate the availability of important genetic variation, which can be slow to achieve through only *de novo* mutations (Hedrick, 2013). In 2004, Seehausen suggested that such genetic variation (and thus, rapidity) may be generated by hybridization events at the onset or within the adaptive radiation. In the *hybrid swarm* hypothesis, hybridization between distantly related species facilitates the beginning of the diversification by establishing a hybrid population with new combinations of alleles, which favor the adaptation to multiple new niches (Seehausen, 2004; Marques et al., 2019). In the *syngameon* hypothesis, hybridization occurs within members of the radiation and allows for further ecological diversification by generating new adaptive traits combinations (Seehausen, 2004). While these hypotheses were initially supported by the observation of cytonuclear discordance (Seehausen, 2004), the accessibility to whole-genome data (see below) provided further evidence of the role played by hybridization in fueling and shaping the adaptive radiations of Darwin's finches, cichlids, and *Heliconius* butterflies (e.g., Dasmahapatra et al., 2012; Lamichhaney et al., 2015; Meier et al., 2017; Malinsky et al., 2018; Edelman et al., 2019; Svardal et al., 2020; Kozak et al., 2021).

The study of diversification processes in the genomic era

The next-generation sequencing revolution provided access to genomic data of natural populations of non-model organisms and resulted in an increased number of studies tackling the question of how species diversify from a genomic perspective (Seehausen et al., 2014; Wolf & Ellegren, 2017; Campbell et al., 2018; Marques et al., 2019). Two main approaches have been employed to gather insights into the genomics of speciation and diversification: population genomics of incipient or closely related species, and phylogenomics. The first principally aims at identifying barrier loci responsible for divergence and reproductive isolation, while the second mainly investigates earlier events that shaped the diversification of lineages.

Population genomics studies in various taxa resulted in the unexpected observation that species' divergence along the genome is highly heterogeneous (e.g., Turner et al., 2005; Ellegren et al., 2012; Jones et al., 2012; Nosil et al., 2012; Renaut et al., 2013; Martin et al., 2013; Soria-Carrasco et al., 2014). Regions of increased genomic divergence were initially interpreted as highly differentiated regions persisting despite the presence of gene flow homogenizing the divergence across the genome. Thus, these regions supposedly contained divergently selected barrier loci related to the reproductive isolation between the species (e.g., Turner et al., 2005; Ellegren et al., 2012; Soria-Carrasco et al., 2014). However, it was rapidly noticed that the heterogeneous differentiation landscape could also arise without gene flow (Nosil et al., 2012; Renaut et al., 2013; Martin et al., 2013) through processes such as mutation, demographic perturbation, or variation in recombination rates combined with linked selection (Seehausen et al., 2014; Wolf & Ellegren, 2017; Campbell et al., 2018). Therefore, in such cases, regions of high differentiation are not necessarily linked with divergent selection, adaptation, and speciation. Consequently, to discern the mechanisms generating the genomic divergence landscape and gain meaningful insights into processes underlying speciation and diversification, statistical measures of gene flow, such as the ABBA-BABA tests (Durand et al., 2011; Martin et al., 2015) or demographic inference approaches (Nielsen & Wakeley, 2001; Excoffier et al., 2013), were developed to complement genome scan analyses. These approaches confirmed, for instance, the presence of speciation-with-gene-flow in *Heliconius* butterfly (Dasmahapatra et al., 2012; Martin et al., 2013) and cichlid fish (Malinsky et al., 2015), allowing the identification of adaptive loci involved in population divergence and speciation.

Population genomics, together with comparative phylogenomics, were also employed to understand the genomics underlying adaptive radiation processes, with a particular focus in examining the genomic features that may promote the rapid diversification (e.g., Sticklebacks: Jones et al., 2012; *Heliconius*: Dasmahapatra et al., 2012; Supple et al., 2013; Edelman et al., 2019; Kozak et al., 2021; Cichlids: Brawand et al., 2014; Meier et al., 2017; Faber-Hammond et al., 2019; McGee et al., 2020; Svardal et al., 2020; Darwin's finches: Lamichhaney et al., 2015; *Anolis*: Feiner, 2016). These studies revealed that ancestral hybridization at the onset of the radiation (*hybrid swarm* hypothesis; Meier et al., 2017; Svardal et al., 2020), introgressive hybridization (*syngameon* hypothesis; Lamichhaney et al., 2015; Dasmahapatra et al., 2012; Supple et al., 2013; Edelman et al., 2019), but also shared standing genetic variation (Jones et al., 2012) characterize the diversification of these lineages, and likely facilitated rapid speciation. Nevertheless, other features such as structural inversions (Jones et al., 2012), accelerated coding and non-coding sequence evolution (Brawand et al., 2014), insertions and deletions (McGee et al., 2020), and high heterozygosity (Ronco et al., 2021) also distinguish the genomes of these groups, possibly playing a role in their diversification.

Altogether, the integration of genomic information in the study of speciation and diversification contributed to our knowledge of the genetics of speciation and established the presence (often through gene flow) and the reuse of ancient genetic variants in adaptation, speciation, and diversification processes (Seehausen et al., 2014; Berner & Salzburger, 2015; Marques et al., 2019). However, it also showed the need to complement the existing work with further studies across the full breadth of the Tree of life to achieve a comprehensive understanding of how genomic properties, ecology, geography, and demographic history interplay to shape species diversification (Seehausen et al., 2014; Wolf & Ellegren, 2017; Campbell et al., 2018). In this sense, clownfish represent a fascinating study system, as it constitutes a rare example of adaptive radiation in the marine environment (Litsios et al., 2012).

The clownfishes study system

Clownfish (or anemonefish, family Pomacentridae) is an iconic group of coral reef fishes including 28 recognized species and two natural hybrids (Fautin & Allen, 1997; Ollerton et al., 2007). Historically, they have been divided into two sister genera, *Amphiprion* and *Premnas* (Allen, 1991), but *Premnas* has been recently recovered within *Amphiprion* (Tang et al., 2021). Their distribution spans the tropical belt of the Indo-Pacific Ocean, with

the highest abundance observed in the Coral Triangle region, where up to nine species live in sympatry (Fautin & Allen, 1997; Elliott and Mariscal, 2001).

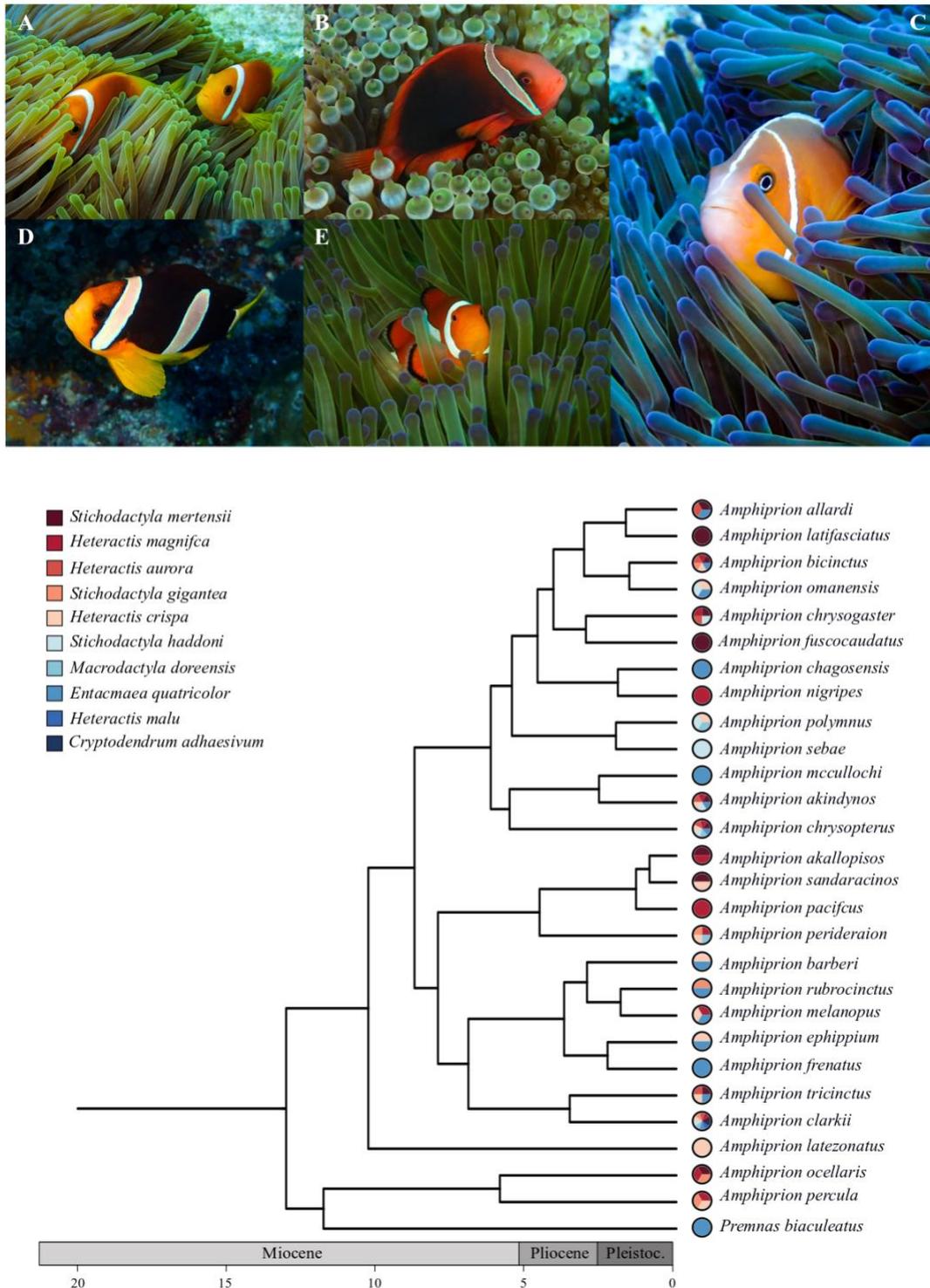


Figure 1: Photos of clownfish species (top) and the dated phylogeny of the group (Schmid et al., unpublished results). The pictures represent *Amphiprion nigripes* (A), *A. frenatus* (B), *A. perideraion* (C), *A. clarkii* (D), and *A. ocellaris* (E). Photos by Sara Heim.

Clownfishes display an outstanding lifespan, with estimations of about 30 years for *A. percula* (Buston & García, 2007) and the observation of individuals living longer than 20 years for *A. ocellaris* and *A. melanopus* (Sahm et al., 2019). This lifespan is twice as long as any other damselfish and six times greater than the expected longevity for fishes of their size (Buston & García, 2007). Another distinctive characteristic of the group is their mutualistic relationship with sea anemones (discussed in detail below; Fautin & Allen, 1997). Within their host sea anemone, clownfishes live in sized-based hierarchical social structures and are protandrous (male to female transition) hermaphrodites (Fricke, 1979; Buston, 2003). Each colony of clownfishes consists of a breeding pair and a variable number of juveniles. The breeding pair is composed of the female and the male, respectively the largest and the second-largest individuals. The non-breeders (if present) are progressively smaller as the hierarchy is descended (Fricke, 1979; Buston, 2003). The removal of the dominant female induces the sex change of the male (which become then female), which in turn triggers the differentiation of the biggest juvenile into a mature male to form the new breeding pair (Fricke, 1979; Casas et al., 2016). Thus, the size-based hierarchy represents a queue to achieve dominant status and reproduction, and the well-defined size differences between individuals are maintained by the precise regulation of the subordinate growth (Buston, 2003), often through aggressive behavior (Iwata et al., 2008, 2010).

The breeding pair lays hundreds of eggs each two to three weeks near their host anemone and provides parental care to them (Buston & Elith, 2011). Following hatching, clownfish have a life-cycle comparable to other coral reef fishes, with a dispersive larval phase followed by the settlement to the reef when larval metamorphosis is completed (Leis & McCormick, 2002; Roux et al., 2019). The return and settlement to the reef are achieved by combining visual, chemical, and acoustic cues (reviewed in Barth et al., 2015), and clownfishes recognize and locate their potential hosts by sensing chemicals released by the sea anemones (Miyagawa, 1989; Dixon, 2011).

The duration of the dispersive larval phase is widely used as a proxy for the potential dispersal distance of coral reef species (Shanks et al., 2003; Shanks, 2009). The dispersive larval phase in clownfishes is short and lasts only 10-15 days (Fautin & Allen, 1997; Roux et al., 2019, 2020), which typically corresponds to a dispersal range of 5-100 km (Salinas-de-Léon et al., 2012)². The short dispersal of clownfishes is likely accountable for the recurrent

² Although dispersal as long as 400 km has been documented for *A. omanensis*, probably due to the presence of strong oceanic currents (Simpson et al., 2014).

natal philopatry observed in the group, with related individuals often sharing the same host but avoiding extensive inbreeding through the hierarchical breeding system (Salles et al., 2016). Nevertheless, range expansion and colonization of new areas are observed in clownfishes despite the short pelagic larval phase (Litsios et al., 2014; Huyghe & Kochzius, 2017), and they are potentially achieved through a stepping stone model over several generations (Huyghe & Kochzius, 2017).

The mutualism with sea anemones

Clownfishes maintain an obligate symbiosis with sea anemones, and they benefit from this relationship with shelter from predators (Buston, 2003) and reproductive benefits gained through the protection of eggs (Saenz-Agudelo, 2011). On the other hand, sea anemones receive substantial benefits from hosting clownfishes - including oxygenation, increased availability of nutrients, and defense against predators (Holbrook & Schmitt, 2005) - but can also live solitarily (Fautin & Allen, 1997). The ten sea anemones species hosting clownfish belong to three distantly related families (*Thalassianthidae*, *Actinidae*, *Stichodactilidae*; Fautin & Allen, 1997), indicating that in these organisms the symbiosis evolved at least three times independently (Titus et al., 2019; Nguyen et al., 2020). Contrarily, from the clownfish side, mutualism was established only once (Litsios et al., 2012). While some other species can occasionally make contact with sea anemones, clownfishes are the only fishes that obligatorily live within them (Randall & Fautin, 2002).

Sea anemones are sessile organisms that have evolved a collection of toxins used for protection and hunting, which can be lethal to fishes (Nedosyko et al., 2014). These toxins are secreted in the sea anemone mucus (Mebs, 2009) and released from specialized cells (i.e., cnidocytes containing nematocysts) following chemical and mechanical stimuli (Anderson & Bouchard, 2009). Clownfish must thus have evolved some defense to live unharmed within sea anemones, and several studies have sought to understand the mechanisms underlying this protection (e.g., reviewed in Mebs, 2009). The general conclusion emerging from this research is that clownfish possess a protective mucus coat, which acts as a molecular camouflage and prevents sea anemones from recognizing them, thus avoiding nematocyst discharge. For instance, the mucus of the clownfish *A. sebae* contains glycoproteins similar to the one produced by its host, *Stichodactyla haddoni*, but different from the non-mutualistic fish *Terapon jarbua* (Balamurugan et al., 2014). Similarly, compared to other reef fishes, *A. ocellaris* mucus significantly lacks *N*-acetylneuraminic acid (Abdullah & Saad, 2015), which

stimulates nematocysts discharge (Ozacmak et al., 2001; Anderson & Bouchard, 2009). Nevertheless, while it appears that compounds within the mucus of clownfish inhibit the release of toxins, the exact molecular mechanisms behind this protection remain unresolved to this day.

Although all clownfishes live within sea anemones, some species require an acclimation procedure when first encountering the host anemones (Mebs, 2009). This observation led to the hypothesis that clownfishes were not innately immune to sea anemones, but they had to cover themselves with their hosts' mucus as camouflage (see Mebs et al., 2009). However, differences between species exist, with, for instance, *A. clarkii* settling immediately into their hosts (Lubbock, 1981; Miyagawa & Hidaka, 1980; Miyagawa, 1989). The differences likely reflect the disparities in the number of mutualistic hosts clownfishes can interact with. Indeed, some species are specialists and can inhabit a unique sea anemone species, while others are generalists and can cooperate with up to ten sea anemone hosts (Figure 1; Fautin & Allen, 1997; Ollerton et al., 2007). These observations suggest the presence of some species-specific mechanisms of defense that complement the general protection mechanisms shared by all clownfish species (Mebs, 2009). Nevertheless, while differences in host usage likely drove the group's diversification (discussed in detail below; Litsios et al., 2012), the mechanisms underlying the specific protections for different sea anemones species are not yet determined.

The adaptive radiation of the group

The mutualism with sea anemones acted like the trigger of clownfish adaptive radiation (Litsios et al., 2012). Indeed, as a key innovation, it provided ecological opportunity by opening to clownfish new habitats with additional exploitable resources (Yoder et al., 2010; Litsios et al., 2012). Clownfishes then diversified into several ecological niches associated with host and habitat usage, resulting in the variety of interactions (generalists-specialists gradient) observed today (Figure 1). During their diversification, clownfishes developed phenotypes correlated with their ecological niches, resulting in the phenotypic convergence³ of species inhabiting similar hosts and habitats (Litsios et al., 2012). This ecological sorting along the generalist to specialist axis allows different species to coexist in sympatry (Elliott & Mariscal, 2001). The primary radiation of clownfish occurred in the Indo-Australian Archipelago, where the group originated (Litsios et al., 2014). Following a colonization event along the eastern coast of Africa, a second geographically independent radiation occurred in

³ Mostly morphological convergence, Litsios et al., 2012

the Western Indian Ocean (Litsios et al., 2014). Indeed, the seven species that diversified there span the whole range of possible mutualistic interaction with sea anemones, from specialist to generalist species, and display the phenotypes associated with their ecological niche (Litsios et al., 2012; Litsios et al., 2014). Nevertheless, the genomics underlying this adaptive radiation are yet to be investigated.

Hybridization seemingly occurred during clownfish diversification (Litsios & Salamin, 2014) and potentially promoted the radiation of the group. Indeed, the cytonuclear inconsistency observed in clownfishes was associated with a substantial increase in diversification rate (Litsios & Salamin, 2014). Additionally, hybridization events are still happening in the group, as shown by the presence of two natural hybrids, *A. leucokranos* and *A. thiellei* (Ollerton et al., 2007; Litsios & Salamin, 2014). Hybridization may be facilitated by the occasional cohabitation of different species within the same sea anemone hosts (Elliott & Mariscal, 2001; Hattori, 2002; Hayashi et al., 2018; Songploy et al. 2021), which is mainly observed when host availability is limiting (Hattori, 2002; Camp et al., 2016). These observations open the question of the role that hybridization played and is still playing in clownfish diversification.

The objectives of this thesis

At a broad scale, the main goal of my thesis was to study, from a genomic perspective, the mechanisms underlying the adaptive radiation of clownfishes. After acquiring the needed genomic data, I investigated the genetics behind the acquisition of clownfish mutualism with sea anemones, thus triggering the group's diversification. By using comparative phylogenomics and population genomics approaches, I then focused on the radiation itself. I first investigated the genomic architecture of the diversification of the whole group and then examined the mechanisms underlying divergence and, ultimately, speciation in the clownfishes of the skunk complex. More specifically:

In **chapter 1**, I aimed to obtain the first genomic resource for clownfish. Using low-coverage PacBio long reads and high-coverage Illumina short reads, I assembled and annotated the genome of the tomato clownfish (*Amphiprion frenatus*). This reference genome and its annotation were of suitable quality, comparable to the genomic resources of other teleosts obtained at the time. This chapter is published in *Molecular Ecology Resources*.

In **chapter 2**, I aimed to extend the genomic resources and, subsequently, examine the genetic basis of clownfish mutualism with sea anemones. I obtained genome assembly and

annotation of nine additional clownfish species and the closely related outgroup *Pomacentrus moluccensis*. I then performed comparative genomic and molecular evolutionary analysis to identify genes likely involved in the evolution of mutualism. I identified 17 genes with signals of positive selection at the origin of the group. Among them, I pinpointed genes with functions associated with *N*-acetylated sugars, which are involved in the discharge of toxins by sea anemones. These genes thus represent candidate genes involved in the evolution of mutualism with sea anemones and the beginning of the clownfish adaptive radiation. This chapter is published in *Genome Biology and Evolution*.

In **chapter 3**, I explored the genomic architecture of clownfish adaptive radiation to identify genomic features that promoted their diversification. I found bursts of transposable elements, overall acceleration in coding sequence evolution, patterns of positive selection, and the presence of topological inconsistencies likely deriving from hybridization events. These features possibly facilitated the radiation of the group by generating genomic variation necessary for natural selection to act on. Additionally, I detected genes with patterns of differential selective pressures associated with clownfish ecological divergence, likely involved in the repeated adaptation to similar ecological niches, thus suggesting parallel evolution in the group.

In **chapter 4**, I examined the processes underlying the diversification of three clownfish species of the skunk complex by employing population genomics approaches. I rejected the presence of hybrid speciation in the group but discovered the species experienced moderate ancestral gene flow, which lessened but still persists in sympatry. I also identified regions of introgression and two highly divergent regions on chromosome 18 that potentially participated in the group's diversification. Altogether, these results provide the first insights into the skunk complex's diversification and show that hybridization events in the IAA are less pervasive than initially thought.

Finally, the first three annexes of this thesis contain two published articles I contributed to and a manuscript that I share co-first authorship, ready for submission. Annexes 4 and 5 contain additional projects that I set up during my thesis and on which I worked alongside two Master's students under my supervision. Given the considerable size of the annexes, I only integrated their abstracts into the manuscript, essentially to avoid excessive printing. Nevertheless, all annexes are available online and are easily accessible through the provided links.

Annex 1: Salis, P., Lorin, T., Lewis, V., Rey, C., **Marcionetti, A.**, Escande, M. L., *et al.* (2019). Developmental and comparative transcriptomic identification of iridophore contribution to white barring in clownfish. *Pigment cell & melanoma research*, 32(3), 391-402. This article explores the pigment cells origin of the white stripes in *A. ocellaris* and concludes that the pigments are iridophore. This study provides new clues to the genomic basis of color diversity and allows the identification of new iridophore genes in fish.

Annex 2: Salis, P., Roux, N., Huang, D., **Marcionetti, A.**, Mougnot, P., Reynaud, M., *et al.* (2021). Thyroid hormones regulate the formation and environmental plasticity of white bars in clownfishes. *Proceedings of the National Academy of Sciences*, 118(23). This article shows that thyroid hormones regulate white bar formation, and that a shift in hormone levels, associated with ecological difference, results in divergent color patterns in different sea anemones species in which the young fish is recruited.

Annex 3: Serrano-Serrano, M. L.*, **Marcionetti, A.***, Perret, M., & Salamin, N. (2019). Convergent changes in gene expression associated with repeated transitions between hummingbird and bee pollinated flowers. *in prep.* ***Shared first authorship.** This article explores the molecular mechanisms underlying the repeated evolution of pollination syndromes in the Gesneriaceae family and finds that convergent evolution of gene expression is involved in the build-up of the pollination syndromes.

Annex 4: Master Thesis of Ms. Virginie Ricci (2018). Identification of putative CNE potentially associated with clownfish adaptive radiation. *Master Thesis of Science In Molecular Life Sciences, University of Lausanne*. This Master thesis aimed at identifying and analyzing conserved non-coding regions in clownfish genomes. I worked alongside and supervised Ms. Virginie Ricci for the whole duration of the project.

Annex 5: Master Thesis of Ms. Sagane Dind (2019). Genetic basis of the symbiosis between clownfish and sea anemones: transcriptomic profile of the epidermal mucus of three clownfish species. *Master Thesis of Science in Behavior, Evolution and Conservation, specialization "Computational Ecology and Evolution", University of Lausanne*. This Master thesis aimed at investigating the presence of mRNA in clownfish mucus and characterizing it through sequencing. I worked alongside and supervised Ms. Sagane Dind for the whole duration of the project.

REFERENCES

- Abdullah, N. S., & Saad, S.** (2015). Rapid detection of *N*-acetylneuraminic acid from false clownfish using HPLC-FLD for symbiosis to host sea anemone. *Asian Journal of Applied Sciences*, 3(5).
- Allen G. R.** (1991) Damselfishes of the world. Mergus, Melle
- Anderson, P. A., & Bouchard, C.** (2009). The regulation of cnidocyte discharge. *Toxicon*, 54(8), 1046-1053.
- Balamurugan, J., Kumar, T. A., Kannan, R., & Pradeep, H. D.** (2014). Acclimation behaviour and biochemical changes during anemonefish (*Amphiprion sebae*) and sea anemone (*Stichodactyla haddoni*) symbiosis. *Symbiosis*, 64(3), 127-138.
- Barth, P., Berenshtein, I., Besson, M., Roux, N., Parmentier, E., Banaigs, B., & Lecchini, D.** (2015). From the ocean to a reef habitat: how do the larvae of coral reef fishes find their way home? A state of art on the latest advances. *Vie et milieu*, 65(2), 91-100.
- Berner, D., & Salzburger, W.** (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, 31(9), 491-499.
- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., ... & Di Palma, F.** (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513(7518), 375-381.
- Buston, P.** (2003). Size and growth modification in clownfish. *Nature*, 424(6945), 145-146.
- Buston, P. M., & Elith, J.** (2011). Determinants of reproductive success in dominant pairs of clownfish: a boosted regression tree analysis. *Journal of Animal Ecology*, 80(3), 528-538.
- Buston, P. M., & García, M. B.** (2007). An extraordinary life span estimate for the clown anemonefish *Amphiprion percula*. *Journal of Fish Biology*, 70(6), 1710-1719.
- Camp, E. F., Hobbs, J. P. A., De Brauwer, M., Dumbrell, A. J., & Smith, D. J.** (2016). Cohabitation promotes high diversity of clownfishes in the Coral Triangle. *Proceedings of the Royal Society B: Biological Sciences*, 283(1827), 20160277.
- Campbell, C. R., Poelstra, J. W., & Yoder, A. D.** (2018). What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, 124(4), 561-583.
- Casas, L., Saborido-Rey, F., Ryu, T., Michell, C., Ravasi, T., & Irigoien, X.** (2016). Sex change in clownfish: molecular insights from transcriptome analysis. *Scientific reports*, 6(1), 1-19.
- Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Shapiro, M. D., Brady, S. D., ... & Kingsley, D. M.** (2010). Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *science*, 327(5963), 302-305.
- Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., ... & Kingsley, D. M.** (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *science*, 307(5717), 1928-1933.
- Coyne, J. A., & Orr, H. A.** (2004). *Speciation* (Vol. 37). Sunderland, MA: Sinauer Associates.
- Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., ... & Postlethwait, J. H.** (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, 101(16), 6050-6055.

- Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., ... & Heliconius Genome Consortium.** (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405), 94.
- Dixson, D. L., Munday, P. L., Pratchett, M., & Jones, G. P.** (2011). Ontogenetic changes in responses to settlement cues by Anemonefish. *Coral Reefs*, 30(4), 903.
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M.** (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, 28(8), 2239-2252.
- Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., ... & Mallet, J.** (2019). Genomic architecture and introgression shape a butterfly radiation. *Science*, 366(6465), 594-599.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... & Wolf, J. B.** (2012). The genomic landscape of species divergence in Ficedula flycatchers. *Nature*, 491(7426), 756-760.
- Elliott, J. K., & Mariscal, R. N.** (2001). Coexistence of nine anemonefish species: differential host and habitat utilization, size and recruitment. *Marine Biology*, 138(1), 23-36.
- Elliott, J. K., Mariscal, R. N., & Roux, K. H.** (1994). Do anemonefishes use molecular mimicry to avoid being stung by host anemones?. *Journal of Experimental Marine Biology and Ecology*, 179(1), 99-113.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M.** (2013). Robust demographic inference from genomic and SNP data. *PLoS genetics*, 9(10), e1003905.
- Faber-Hammond, J. J., Bezault, E., Lunt, D. H., Joyce, D. A., & Renn, S. C.** (2019). The genomic substrate for adaptive radiation: copy number variation across 12 tribes of African cichlid species. *Genome biology and evolution*, 11(10), 2856-2874.
- Fautin, D. G., & Allen, G. R.** (1997). Anemone fishes and their host sea anemones. Western Australian Museum, Perth, 159 pp.
- Feiner, N.** (2016). Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards. *Proceedings of the Royal Society B: Biological Sciences*, 283(1840), 20161555.
- Fitzpatrick, B. M., Fordyce, J. A., & Gavrilets, S.** (2009). Pattern, process and geographic modes of speciation. *Journal of evolutionary biology*, 22(11), 2342-2347.
- Footo, A. D.** (2018). Sympatric speciation in the genomic era. *Trends in ecology & evolution*, 33(2), 85-95.
- Fricke, H. W.** (1979). Mating system, resource defence and sex change in the anemonefish Amphiprion akallopisos. *Zeitschrift für Tierpsychologie*, 50(3), 313-326.
- Gavrilets, S.** (2014). Models of speciation: where are we now?. *Journal of heredity*, 105(S1), 743-755.
- Grant, P. R., & Grant B. R.,** (2008) How and Why Species Multiply: The Radiation of Darwin's Finches. Princeton University Press
- Hattori, A.** (2002). Small and large anemonefishes can coexist using the same patchy resources on a coral reef, before habitat destruction. *Journal of Animal Ecology*, 71(5), 824-831.
- Hayashi, K., Tachihara, K., & Reimer, J. D.** (2018). Patterns of coexistence of six anemonefish species around subtropical Okinawa-jima Island, Japan. *Coral Reefs*, 37(4), 1027-1038.
- Hedrick, P. W.** (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular ecology*, 22(18), 4606-4618.
- Holbrook, S. J., & Schmitt, R. J.** (2005). Growth, reproduction and survival of a tropical sea anemone (Actiniaria): benefits of hosting anemonefish. *Coral reefs*, 24(1), 67-73.

- Huyghe, F., & Kochzius, M.** (2017). Highly restricted gene flow between disjunct populations of the skunk clownfish (*Amphiprion akallopisos*) in the Indian Ocean. *Marine Ecology*, 38(1), e12357.
- Iwata, E., Nagai, Y., & Sasaki, H.** (2010). Social rank modulates brain arginine vasotocin immunoreactivity in false clown anemonefish (*Amphiprion ocellaris*). *Fish physiology and biochemistry*, 36(3), 337-345.
- Iwata, E., Nagai, Y., Hyoudou, M., & Sasaki, H.** (2008). Social environment and sex differentiation in the false clown anemonefish, *Amphiprion ocellaris*. *Zoological Science*, 25(2), 123-128.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... & Kingsley, D. M.** (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55-61.
- Kingman, G. A. R., Vyas, D. N., Jones, F. C., Brady, S. D., Chen, H. I., Reid, K., ... & Veeramah, K. R.** (2021). Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, 7(25), eabg5285.
- Kocher, T. D., Conroy, J. A., McKaye, K. R., & Stauffer, J. R.** (1993). Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Molecular phylogenetics and evolution*, 2(2), 158-165.
- Kozak, K. M., Joron, M., McMillan, W. O., & Jiggins, C. D.** (2021). Rampant genome-wide admixture across the *Heliconius* radiation. *Genome Biology and Evolution*.
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martínez-Barrio, A., ... & Andersson, L.** (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371-375.
- Leis, J. M., & McCormick, M. I.** (2002). The biology, behavior, and ecology of the pelagic, larval stage of coral reef fishes. *Coral reef fishes: dynamics and diversity in a complex ecosystem*, 171-199.
- Litsios, G., & Salamin, N.** (2014). Hybridisation and diversification in the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, 14(1), 1-9.
- Litsios, G., Pearman, P. B., Lanterbecq, D., Tolou, N., & Salamin, N.** (2014). The radiation of the clownfishes has two geographical replicates. *Journal of biogeography*, 41(11), 2140-2149.
- Litsios, G., Sims, C. A., Wüest, R. O., Pearman, P. B., Zimmermann, N. E., & Salamin, N.** (2012). Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, 12(1), 1-15.
- Losos, J. B.** (2010). Adaptive radiation, ecological opportunity, and evolutionary determinism: American Society of Naturalists EO Wilson Award address. *The American Naturalist*, 175(6), 623-639.
- Losos, J. B., Jackman, T. R., Larson, A., de Queiroz, K., & Rodríguez-Schettino, L.** (1998). Contingency and determinism in replicated adaptive radiations of island lizards. *Science*, 279(5359), 2115-2118.
- Lubbock, R.** (1981). The clownfish/anemone symbiosis: a problem of cellular recognition. *Parasitology*, 82(1), 159-173.
- Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., ... & Turner, G. F.** (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, 350(6267), 1493-1498.
- Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R.** (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature ecology & evolution*, 2(12), 1940-1955.

- Marques, D. A., Meier, J. I., & Seehausen, O.** (2019). A combinatorial view on speciation and adaptive radiation. *Trends in ecology & evolution*, 34(6), 531-544.
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... & Jiggins, C. D.** (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome research*, 23(11), 1817-1828.
- Martin, S. H., Davey, J. W., & Jiggins, C. D.** (2015). Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular biology and evolution*, 32(1), 244-257.
- McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., ... & Seehausen, O.** (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, 586(7827), 75-79.
- Mebs, D.** (2009). Chemical biology of the mutualistic relationships of sea anemones with fish and crustaceans. *Toxicon*, 54(8), 1071-1074.
- Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O.** (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications*, 8(1), 1-11.
- Miyagawa, K.** (1989). Experimental analysis of the symbiosis between anemonefish and sea anemones. *Ethology*, 80(1-4), 19-46.
- Miyagawa, K., & Hidaka, T.** (1980). *Amphiprion clarkii* juvenile: innate protection against and chemical attraction by symbiotic sea anemones. *Proceedings of the Japan Academy, Series B*, 56(6), 356-361.
- Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B.** (2011). How many species are there on Earth and in the ocean?. *PLoS Biol*, 9(8), e1001127.
- Muschick, M., Indermaur, A., & Salzburger, W.** (2012). Convergent evolution within an adaptive radiation of cichlid fishes. *Current biology*, 22(24), 2362-2368.
- Nedosyko, A. M., Young, J. E., Edwards, J. W., & Burke da Silva, K.** (2014). Searching for a toxic key to unlock the mystery of anemonefish and anemone symbiosis. *PloS one*, 9(5), e98449.
- Nguyen, H. T. T., Dang, B. T., Glenner, H., & Geffen, A. J.** (2020). Cophylogenetic analysis of the relationship between anemonefish *Amphiprion* (Perciformes: Pomacentridae) and their symbiotic host anemones (Anthozoa: Actiniaria). *Marine Biology Research*, 16(2), 117-133.
- Nielsen, R., & Wakeley, J.** (2001). Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, 158(2), 885-896.
- Nosil, P.** (2012). *Ecological speciation*. Oxford University Press.
- Nosil, P., Harmon, L. J., & Seehausen, O.** (2009). Ecological explanations for (incomplete) speciation. *Trends in ecology & evolution*, 24(3), 145-156.
- Ollerton, J., McCollin, D., Fautin, D. G., & Allen, G. R.** (2007). Finding NEMO: nestedness engendered by mutualistic organization in anemonefish and their hosts. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609), 591-598.
- Ozacmak, V. H., Thorington, G. U., Fletcher, W. H., & Hessinger, D. A.** (2001). *N*-acetylneuraminic acid (NANA) stimulates in situ cyclic AMP production in tentacles of sea anemone (*Aiptasia pallida*): possible role in chemosensitization of nematocyst discharge. *Journal of Experimental Biology*, 204(11), 2011-2020.
- Pennisi, E.** (2005). What determines species diversity?. *Science*, 309(5731), 90-91.

- Randall, J., & Fautin, D.** (2002). Fishes other than anemonefishes that associate with sea anemones. *Coral Reefs*, 21(2), 188-190.
- Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., ... & McMillan, W. O.** (2011). Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, 333(6046), 1137-1141.
- Renaut, S., Grassa, C. J., Yeaman, S., Moyers, B. T., Lai, Z., Kane, N. C., ... & Rieseberg, L. H.** (2013). Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature communications*, 4(1), 1-8.
- Ronco, F., Matschiner, M., Böhne, A., Boila, A., Büscher, H. H., El Taher, A., ... & Salzburger, W.** (2021). Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*, 589(7840), 76-81.
- Rosenblum, E. B., Parent, C. E., & Brandt, E. E.** (2014). The molecular basis of phenotypic convergence. *Annual Review of Ecology, Evolution, and Systematics*, 45, 203-226.
- Roux, N., Salis, P., Lambert, A., Logeux, V., Soulat, O., Romans, P., ... & Laudet, V.** (2019). Staging and normal table of postembryonic development of the clownfish (*Amphiprion ocellaris*). *Developmental Dynamics*, 248(7), 545-568.
- Roux, N., Salis, P., Lee, S. H., Besseau, L., & Laudet, V.** (2020). Anemonefish, a model for eco-Evo-devo. *EvoDevo*, 11(1), 1-11.
- Rundell, R. J., & Price, T. D.** (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, 24(7), 394-399.
- Rundle, H. D., & Nosil, P.** (2005). Ecological speciation. *Ecology letters*, 8(3), 336-352.
- Rundle, H. D., Nagel, L., Boughman, J. W., & Schluter, D.** (2000). Natural selection and parallel speciation in sympatric sticklebacks. *Science*, 287(5451), 306-308.
- Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R., & Planes, S.** (2011). Detrimental effects of host anemone bleaching on anemonefish populations. *Coral Reefs*, 30(2), 497-506.
- Saenz-Agudelo, P., Jones, G. P., Thorrold, S. R., & Planes, S.** (2011). Detrimental effects of host anemone bleaching on anemonefish populations. *Coral Reefs*, 30(2), 497-506.
- Sahm, A., Almáida-Pagán, P., Bens, M., Mutalipassi, M., Lucas-Sánchez, A., de Costa Ruiz, J., ... & Cellerino, A.** (2019). Analysis of the coding sequences of clownfish reveals molecular convergence in the evolution of lifespan. *BMC evolutionary biology*, 19(1), 1-12.
- Salinas-de-León, P., Jones, T., & Bell, J. J.** (2012). Successful determination of larval dispersal distances and subsequent settlement for long-lived pelagic larvae. *PloS one*, 7(3), e32788.
- Salles, O. C., Pujol, B., Maynard, J. A., Almany, G. R., Berumen, M. L., Jones, G. P., ... & Planes, S.** (2016). First genealogy for a wild marine fish population reveals multigenerational philopatry. *Proceedings of the National Academy of Sciences*, 113(46), 13245-13250.
- Schlichter, D.** (1976). Macromolecular mimicry: substances released by sea anemones and their role in the protection of anemone fishes. In *Coelenterate ecology and behavior* (pp. 433-441). Springer, Boston, MA.
- Schluter, D.** (2000). *The ecology of adaptive radiation*. OUP Oxford.
- Schluter, D.** (2001). Ecology and the origin of species. *Trends in ecology & evolution*, 16(7), 372-380.
- Schluter, D.** (2009). Evidence for ecological speciation and its alternative. *Science*, 323(5915), 737-741.

- Schluter, D., & Nagel, L. M.** (1995). Parallel speciation by natural selection. *The American Naturalist*, *146*(2), 292-301.
- Seehausen, O.** (2004). Hybridization and adaptive radiation. *Trends in ecology & evolution*, *19*(4), 198-207.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... & Widmer, A.** (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*(3), 176-192.
- Seehausen, O., Terai, Y., Magalhaes, I. S., Carleton, K. L., Mrosso, H. D., Miyagi, R., ... & Okada, N.** (2008). Speciation through sensory drive in cichlid fish. *Nature*, *455*(7213), 620-626.
- Shanks, A. L.** (2009). Pelagic larval duration and dispersal distance revisited. *The biological bulletin*, *216*(3), 373-385.
- Shanks, A. L., Grantham, B. A., & Carr, M. H.** (2003). Propagule dispersal distance and the size and spacing of marine reserves. *Ecological applications*, *13*(sp1), 159-169.
- Simpson, G. G.** (1953). *The major features of evolution*. Columbia University Press.
- Simpson, S. D., Harrison, H. B., Claereboudt, M. R., & Planes, S.** (2014). Long-distance dispersal via ocean currents connects Omani clownfish populations throughout entire species range. *PLoS One*, *9*(9), e107610.
- Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W.** (2010). The biology of speciation. *Evolution: International Journal of organic evolution*, *64*(2), 295-315.
- Songploy, S., Chavanich, S., Mariasingarayan, Y., & Viyakarn, V.** (2021). The Sharing of the Same Host of Two Species of Anemonefish in the Gulf of Thailand, One of Which Is Possibly Introduced. *Diversity*, *13*(7), 304.
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., ... & Nosil, P.** (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, *344*(6185), 738-742.
- Supple, M. A., Hines, H. M., Dasmahapatra, K. K., Lewis, J. J., Nielsen, D. M., Lavoie, C., ... & Counterman, B. A.** (2013). Genomic architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies. *Genome research*, *23*(8), 1248-1257.
- Svardal, H., Quah, F. X., Malinsky, M., Ngatunga, B. P., Miska, E. A., Salzburger, W., ... & Durbin, R.** (2020). Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Molecular biology and evolution*, *37*(4), 1100-1113.
- Tang, K. L., Stiassny, M. L., Mayden, R. L., & DeSalle, R.** (2021). Systematics of Damsel-fishes. *Ichthyology & Herpetology*, *109*(1), 258-318.
- Titus, B. M., Benedict, C., Laroche, R., Gusmão, L. C., Van Deusen, V., Chiodo, T., ... & Rodríguez, E.** (2019). Phylogenetic relationships among the clownfish-hosting sea anemones. *Molecular phylogenetics and evolution*, *139*, 106526.
- Turner, J. R.** (1976). Adaptive radiation and convergence in subdivisions of the butterfly genus *Heliconius* (Lepidoptera: Nymphalidae). *Zoological Journal of the Linnean Society*, *58*(4), 297-308.
- Turner, T. L., Hahn, M. W., & Nuzhdin, S. V.** (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS biology*, *3*(9), e285.
- Wagner, C. E., Harmon, L. J., & Seehausen, O.** (2012). Ecological opportunity and sexual selection together predict adaptive radiation. *Nature*, *487*(7407), 366-369.

- Wolf, J. B., & Ellegren, H.** (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87-100.
- Yoder, J. B., Clancey, E., Des Roches, S., Eastman, J. M., Gentry, L., Godsoe, W., ... & Harmon, L. J.** (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of evolutionary biology*, 23(8), 1581-1596.

Chapter 1

First draft genome of an iconic clownfish species (*Amphiprion frenatus*)

From the article: **Marcionetti**, A., Rossier, V., Bertrand, J. A., Litsios, G., & Salamin, N. **2018**. First draft genome of an iconic clownfish species (*Amphiprion frenatus*). *Molecular ecology resources*, 18(5), 1092-1101. [DOI: 10.1111/1755-0998.12772](https://doi.org/10.1111/1755-0998.12772)

First draft genome of an iconic clownfish species (*Amphiprion frenatus*)

Anna Marcionetti^{1,2}, Victor Rossier^{1,2}, Joris A. M. Bertrand^{1,2}, Glenn Litsios^{1,2}, Nicolas Salamin^{1,2}

¹Department of Computational Biology, Biophore, University of Lausanne, Lausanne, Switzerland

²Swiss Institute of Bioinformatics, Lausanne, Switzerland

ABSTRACT

Clownfishes (or anemonefishes) form an iconic group of coral reef fishes, principally known for their mutualistic interaction with sea anemones. They are characterized by particular life history traits, such as a complex social structure and mating system involving sequential hermaphroditism, coupled with an exceptionally long lifespan. Additionally, clownfishes are considered to be one of the rare groups to have experienced an adaptive radiation in the marine environment.

Here, we assembled and annotated the first genome of a clownfish species, the tomato clownfish (*Amphiprion frenatus*). We obtained 17,801 assembled scaffolds, containing a total of 26,917 genes. The completeness of the assembly and annotation was satisfying, with 96.5% of the Actinopterygii Benchmarking Universal Single-Copy Orthologs (BUSCOs) being retrieved in *A. frenatus* assembly. The quality of the resulting assembly is comparable to other bony fish assemblies.

This resource is valuable for advancing studies of the particular life history traits of clownfishes, as well as being useful for population genetic studies and the development of new phylogenetic markers. It will also open the way to comparative genomics. Indeed, future genomic comparison among closely related fishes may provide means to identify genes related to the unique adaptations to different sea anemone hosts, as well as better characterize the genomic signatures of an adaptive radiation.

KEYWORDS

adaptation, anemonefish, fish, genomics, Indo-Pacific

1. INTRODUCTION

Clownfishes (or anemonefishes; subfamily Amphiprioninae, genera *Amphiprion* and *Premnas*) are an iconic and highly diverse group of coral reef fishes. They are part of the damselfish family (Pomacentridae), and they include 28 described species (Ollerton, McCollin, Fautin, & Allen, 2007). Their distribution spans the whole tropical belt of the Indo-West Pacific Ocean, but their highest species richness is situated in the Coral Triangle region, where up to nine clownfish species have been observed in sympatry (Elliott & Mariscal, 2001).

One distinctive characteristic of this group is the mutualistic interaction they maintain with sea anemones (Fautin & Allen, 1997). While all species of the clade are associated with sea anemones, there is a large variability in host usage within the group. Indeed, some species are strictly specialist and can interact with a unique species of sea anemones, while others are generalists and can cooperate with a large number of hosts (Ollerton et al., 2007). Studies have been conducted to understand both the process of host selection used by clownfishes (e.g., Arvedlund, McCormick, Fautin, & Bildsøe, 1999; Elliott, Elliott, & Mariscal, 1995; Elliott & Mariscal, 2001; Huebner, Dailey, Titus, Khalaf, & Chadwick, 2012) and the mechanisms granting them protection from sea anemones toxins (reviewed in Mebs, 2009). However, we do not have yet a full answer for these questions. In particular, the genomic bases of these mechanisms remain poorly understood.

Clownfishes are characterized also by particular life history traits and strategies compared to other damselfish and most other coral reef fishes. Indeed, they display an outstanding lifespan, with around 30 years estimated for *A. percula*. This lifespan is twice as long as any other damselfish and six times greater than the expected longevity for a fish of their size (Buston & García, 2007). Moreover, clownfishes live in complex social structures within the sea anemones and are protandrous hermaphrodites. Studies have been conducted to understand the maintenance of this social structure (e.g., Buston, 2003, 2004; Hattori, 2000; Mitchell, 2003) and the mechanisms involved in sex change (e.g., Casas et al., 2016; Kim, Jin, Lee, Kil, & Choi, 2010; Kim, Lee, Habibi, & Choi, 2013; Kim, Shin, Habibi, Lee, & Choi, 2012; Miura, Kobayashi, Bhandari, & Nakamura, 2013).

Litsios et al. (2012) proposed that the obligate mutualistic interaction of clownfishes with sea anemones acted as the main key innovation that triggered the adaptive radiation of the group. It was further shown that geographic isolation associated with a rather small

dispersal capacity and hybridization played a role in driving the burst of diversification and the adaptive process of this group (Litsios, Pearman, Lanterbecq, Tolou, & Salamin, 2014; Litsios & Salamin, 2014). Thus, the clownfishes potentially represent a new and interesting model system for the study of adaptive radiations and could be employed to validate the theoretical findings on the dynamics of this process (Gavrilets & Losos, 2009; Gavrilets & Vose, 2005).

Despite the many and different aspects of clownfishes that are being studied in different fields, the knowledge on their long-term evolution and its underlying genetic bases remains scarce. Yet, advances in next-generation sequencing technologies allow now to obtain genomic information also for non-model organisms. More precisely, the widely used Illumina short reads can be complemented with Pacific Biosciences (PacBio) long reads for hybrid assemblies (Deshpande, Fung, Pham, & Bafna, 2013; Koren et al., 2012; Miller et al., 2017). This dual strategy is fruitful as it allows to overcome the errors due to both the repeated regions of the genome that cannot be unambiguously assembled with short reads and the relatively higher error rate of long reads. Indeed, Illumina technology tends to be particularly sensitive to the first kind of error, whereas PacBio technology is expected to be more affected by the former one. Additionally, the sequencing of RNA can be used to improve the gene annotation in newly assembled genomes (Denton et al., 2014).

In this study, we aimed at obtaining the first draft genome of a clownfish species: the tomato clownfish (*Amphiprion frenatus*). This resource will provide new tools for future investigation of clownfish life history traits and the study of their mutualism with sea anemones. Additionally, new markers for phylogenetic and population genetics studies can be developed thanks to this draft reference genome. This resource also opens the way to comparative genomics among closely related fishes to identify genes related to the unique adaptations of clownfishes to their different sea anemone hosts. This genomic resource will provide the possibility to link these different fields of research and make a step forward in the understanding of clownfish ecology and evolution.

2. MATERIAL AND METHODS

2.1. *Amphiprion frenatus* samples

Samples from three individuals of *Amphiprion frenatus* were obtained from a local aquarium shop. The three individuals were not from the same breeding line, and because they

were acquired from an aquarium shop, the exact origin of the individuals was not available. The individuals of *A. frenatus* passed away beforehand at the aquarium shop, and samples from deceased fish were received. Thus, the three individuals did not undergo any manipulation or experimentation in the laboratory. The three individuals were used for short-reads Illumina sequencing, long-reads PacBio sequencing and RNA sequencing, respectively. The full liver sample obtained from one individual was used for RNA extraction, while the full muscle sample obtained from the second individual was used for long-reads library preparation. Fin tissue sample obtained from the third individual was used for short-reads library preparation, and the remaining sample is stored at the Department of Ecology and Evolution, University of Lausanne (sample ID: F4.6.1.3,7).

Although the use of the same individual for the generation of different sequencing data is normally preferred, the small amount of genomic DNA obtained for each individual did not allow us to use the same individual for the preparation of all the libraries. To overcome this issue, we corrected the obtained long reads with the short Illumina reads to account for both sequencing errors and intraspecific variation (see Section 2.4).

2.2 DNA extraction, library construction and Illumina sequencing

Genomic DNA (gDNA) was extracted from about 50 mg of fin tissue from sample F4.6.1.3,7 using DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The total amount of gDNA was measured using Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher Scientific, Waltham, USA). The integrity of the gDNA was verified with the Fragment Analyzer Automated CE System (Advanced Analytical Technologies, Fiorenzuola d'Arda, Italy). A total of 100 ng and 4 µg of gDNA were used for paired-end (PE) and mate-pair (MP) library preparation, respectively.

Short-insert (350 bp) PE library was prepared at the Lausanne Genomic Technologies Facility (LGTF, Switzerland) using TruSeq Nano DNA LT Library Preparation Kit (Illumina). Long-insert (3 kb) MP library was prepared at Fasteris SA (Geneva, Switzerland) using the Nextera Mate Pair Library Preparation Kit from Illumina. The concentration, purity and size of the libraries obtained were verified using Agilent Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). The PE library was sequenced on two lanes of Illumina HiSeq2000 at the LGTF (run type: paired-end reads, read length of 100). The MP library was sequenced on half lane of Illumina HiSeq2500 at Fasteris (run type: paired-end reads, read length of 125 bp).

2.3. DNA extraction, library construction and Pacific Biosciences (PacBio) sequencing

High-molecular-weight gDNA was extracted from a second individual of *A. frenatus*, from about 100 mg of muscle tissue using QIAGEN Genomic-tip 100/G (Qiagen, Hilden, Germany) following the manufacturer's instructions. The total amount of gDNA was measured using Qubit dsDNA HS Assay Kit, and the integrity of the gDNA was verified with Fragment Analyzer Automated CE System. The construction of the SMRTbell sequencing libraries and the sequencing were performed at the LGTF from a starting material of 10 µg of gDNA. The SMRTbell libraries were sequenced on eight SMRT cells (Pacific Bioscience) using C2 chemistry on the PacBio RS (Pacific Biosciences) sequencing platform.

2.4. Preprocessing of sequenced reads

Reads quality has a major impact on the quality of the resulting assembly, and the use of error-corrected reads increases dramatically the size of the contigs (Salzberg et al., 2012). Two different PE reads correction strategies were therefore performed. The first consisted in correcting raw reads, without prior processing, with ALLPATHS-LG module for fragment read error correction using default parameters (release 44837; Gnerre et al., 2011).

The second strategy consisted of three steps. We removed PE reads that failed the chastity filtering of the Casava pipeline with `casava_filter_se.pl` (version 0.1-1, from <http://brianknaus.com/software/srtoolbox/>). Remaining PE reads were trimmed using Sickle (version 1.29; Joshi & Fass, 2011), with the following parameters: `-qual-threshold 30 -length-threshold 80`. Substitutions due to sequencing errors in the trimmed PE reads were corrected with Quake (version 0.3.5; Kelley, Schatz, & Salzberg, 2010). The k -mers frequency needed by Quake was obtained with Jellyfish (version 1.1; Marçais & Kingsford, 2011). A k -mer size of $k=18$ was selected according to Quake documentation, which suggests the use of $k=\log(200*\text{GenomeSize})/\log(4)$. The genome size for the calculation was obtained from the Animal Genome Size Database (Gregory, 2017), in which the reported genome sizes for the *Amphiprion* genus ranged from 792 to 1,200 Mb. The genome size of the *A. frenatus* individual was subsequently estimated from the genomic data, by dividing the number of error-free 18-mers by their peak coverage depth. The expected number of chromosomes in clownfish was also reported in the Animal Genome Size Database (Gregory, 2017), and it is of $2n = 48$ for *A. clarkii*.

The MP reads were processed at Fasteris SA (Geneva). Because MP libraries can have a relatively low total diversity, the data set was screened for paired-end reads sharing the exact

same sequences on the first 30 bases of both ends. This can be expected due to PCR duplicates, and only one of the copies was kept to obtain unique pairs. The data set was additionally screened to remove reads containing empty inserts. The linker sequences were searched and trimmed in the unique and non-empty pairs. The software Sickle was used to remove the remaining low-quality bases (parameters: `-qual-threshold 25 -length-threshold 80`). The quality of the resulting MP was verified with FASTQC (version 0.11.2; Andrews, 2010).

PacBio long reads were corrected with proofread (version 2.12; Hackl, Hedrich, Schultz, & Förster, 2014) using trimmed and error-corrected PE reads. This method allows to increase SMRT sequencing accuracy, which is substantially lower compared to Illumina technologies (Goodwin, McPherson, & McCombie, 2016). Because two different individuals were sequenced, proofread also corrects the possible polymorphism based on the Illumina-sequenced individual. This will remove possible errors due to the sequencing of different individuals for the genome assembly (Zhu et al., 2016).

2.5. Nuclear genome assembly

Trimmed MP and PE reads resulting from the two strategies of read correction were assembled using both Platanus (version 1.2.1; Kajitani et al., 2014) and SOAPdenovo2 (version 2.04.240; Luo et al., 2012). One of the advantages of Platanus is its automatic optimization of all parameters, including k -mer size. In SOAPdenovo2, assemblies were performed with a k -mer size of $k=35$ and $k=63$. The two values were chosen to span a large range, with the lower being comparable to the starting k -mer size of Platanus and the larger being close to the best k proposed by KmerGenie (release 1.6982; Chikhi & Medvedev, 2013).

Scaffolding and gap-closing were performed within the Platanus or SOAPdenovo2 pipelines. For scaffolding, both short-insert and long-insert libraries were used. The best genome assembly was selected by investigating assembly statistics (N50, maximum scaffold length, number of scaffolds, gap number). The best genome assembly was reached using the reads corrected with ALLPATHS-LG modules and assembled with Platanus. Because of the substantially better quality of Platanus assemblies over the SOAPdenovo2 ones, we decided not to perform SOAPdenovo2 assemblies by progressively increasing k -mer sizes.

We further closed gaps in the resulting best assembly using the corrected PacBio long reads with PBJelly2 (version 14.1; English et al., 2012). The script FakeQuals.py was used to

set a quality score of 40 to each base in each scaffold. The mapping of long reads on the genome in PBJelly2 was performed by blasr (version 1.3.1; Chaisson & Tesler, 2012), with the parameters set as following: -minPctIdentity 70 -SdpTupleSize 11 -nCandidates 20. The other parameters were left as default. Scaffolds smaller than 1 kb were removed from the final assembly.

2.6. RNA extraction, library construction, sequencing and read processing

Liver sample from an additional individual of *A. frenatus* was obtained from a local aquarium shop for RNA sequencing to improve gene annotation of genome assembly (Denton et al., 2014). RNA was extracted with RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The total amount of RNA in each sample and its quality were measured using the Fragment Analyzer Automated CE System.

A strand-specific cDNA library was prepared using TruSeq Stranded mRNA Sample Prep Kit (Illumina) from an initial amount of total RNA of 1 µg, following the manufacturer's instruction. The concentration, purity and size of the library were tested using the Fragment Analyzer Automated CE System. The library was sequenced on one lane of Illumina HiSeq2000 at the LGTF (run type: paired-end reads, read length of 100). Obtained PE reads were trimmed with Sickle, with the following parameters: -qual-threshold 20 -length-threshold 20.

2.7. Nuclear genome validation

We investigated the quality of the assembled genome by evaluating the mapping rates of the PE and MP libraries using BWA (version 0.7.12; Li & Durbin, 2009), with default parameters. PE reads were subsampled, and only the reads from a single Illumina lane were used. Prior to BWA mapping, MP reads were reversed to obtain the forward-reverse orientation with a homemade script. The RNA-Seq reads from *A. frenatus* were mapped to the genome with HiSat2 (version 2.0.2, default parameters; Kim, Langmead, & Salzberg, 2015). Mapping statistics were summarized with BamTools stats (version 2.3.0; Barnett, Garrison, Quinlan, Strömberg, & Marth, 2011). Insert sizes and read orientation were checked with Picard (version 2.2.1, "CollectInsertSizeMetrics" tool, <http://picard.sourceforge.net>).

The composition of the short scaffolds (<1 kb) removed from the final assembly was assessed using BLASTN (version 2.3.30, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) against RefSeq database (Release 80, E-value cut-off of 10^{-4}).

To further assess the quality of the assembly, a microsynteny analysis against *Oreochromis niloticus* genome (GCA_000188235.2; Brawand et al., 2014) was performed with SynChro (Drillon, Carbone, & Fischer, 2014). We allowed for 5 to 10 intervening genes between gene pairs, as performed in Dibattista *et al.* (2016). Finally, the completeness of the genome assembly was assessed with CEGMA (version 2.3; Parra, Bradnam, & Korf, 2007).

2.8. Nuclear genome annotation

Interspersed repeats and low-complexity DNA sequences in the genome were identified with RepeatModeler (version 1.08, engine ncbi) and soft-masked with RepeatMasker (version 4.0.5; Smit, Hubley, & Green, 2015). *Ab initio* gene prediction was carried out with BRAKER1 (version 1.9; Hoff, Lange, Lomsadze, Borodovsky, & Stanke, 2015). RNA-Seq data of *A. frenatus* previously mapped with HiSat2 (see Section 2.7) were used within BRAKER1 to improve *ab initio* gene prediction. RNA-Seq data were subsequently assembled into transcripts with Cufflink (version 2.2.1, default parameters; Trapnell et al., 2010). These transcripts, together with the *ab initio* gene predictions and the proteomes of *Danio rerio* (GCA_000002035.3), *O. niloticus* (GCA_000188235.2) and *Stegastes partitus* (GCA_000690725.1), were used to provide evidence for the inference of gene structures. The different pieces of evidence were aligned on each genome and synthesized into coherent gene models with MAKER2 (version 2.31.8; Holt & Yandell, 2011). The quality of the annotation was assessed by investigating the annotation edit distance (AED), which is calculated by MAKER2.

The completeness of the resulting gene models was assessed by comparing the length of the predicted proteins with the *O. niloticus* protein length. We performed BLASTP searches against *O. niloticus* proteome (total of 47,713 proteins, E-value cut-off of 10^{-6}). We assumed that the best blast hit was orthologous and calculated the difference in protein length. We also calculated the “query” (*A. frenatus*) and “target” (*O. niloticus*) coverage, as defined in https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Oreochromis_niloticus/102/ (see also Supplementary Figure S1).

Functional annotation was performed with BLASTP searches against the SwissProt database (subset: metazoans proteins, downloaded in June 2016, total of 104,439 proteins), with an E-value cut-off of 10^{-6} . We also blasted (BLASTP, E-value cut-off of 10^{-6}) *A. frenatus* proteins against RefSeq database (subset Actinopterygii sequences, downloaded on June 2016, total of 175,995 sequences), which is less accurate than SwissProt but more

comprehensive. To provide further functional annotations, we used InterProScan (version 5.16.55.0; Jones et al., 2014) to predict protein domains based on homologies with the PFAM database (release 28, 16,230 families; Finn et al., 2016). Gene Ontologies (GO) were annotated to each predicted protein by retrieving the GO associated with its best SwissProt hit. Additionally, GO associated with protein domains were annotated in the InterProScan pipeline (option -goterms).

The completeness of the genome annotation was investigated with BUSCO (version 2, data sets: Metazoan and Actinopterygii, mode: proteins; Simão, Waterhouse, Ioannidis, Kriventseva, & Zdobnov, 2015). Additionally, we calculated the “query” and “target” coverage for *A. frenatus* proteins and their SwissProt hits (Supplementary Figure S1). “Query” coverage and “target” coverage were compared to *O. niloticus*, *Maylandia zebra*, and *D. rerio* coverages retrieved from the NCBI *Oreochromis niloticus* Annotation release 102 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Oreochromis_niloticus/102/). The *Chaetodon austriacus* proteome was downloaded from <http://caus.reefgenomics.org> in January 2017. As for *A. frenatus*, protein sequences of *C. austriacus* were blasted against SwissProt metazoan, and “query” and “target” coverages were calculated.

2.9. Mitochondrial genome reconstruction and annotation

We reconstructed the entire mitochondrial genome from a random subsample of 20 million of the PE reads filtered with ALLPATH-LG. To do so, we followed the baiting and iterative mapping approach implemented in MITObim (version 1.9; Hahn, Bachmann, & Cheverus, 2013). We checked for the consistency of the outputs of two reconstruction methods. First, we used as reference a previously published complete mitochondrial genome of *A. frenatus* that we retrieved from GenBank (GB KJ833752; Li, Chen, Kang, & Liu, 2015). Alternatively, we also worked using a conspecific barcode sequence (i.e., COI gene) as a seed to initiate the process (GB FJ582759; Steinke, Zemlak, & Hebert, 2009). The circularity of the sequence was manually inferred, and the reads of the pool were mapped back onto the resulting mitochondrial genome to check for the reconstruction success and to assess coverage using Geneious (version 10.2.2; Kearse *et al.*, 2012). We used MitoAnnotator and the MitoFish online database to annotate the inferred mitochondrial genome (Iwasaki *et al.*, 2013).

3. RESULTS AND DISCUSSION

3.1. Nuclear genome sequencing and assembly

We obtained 534.9 million raw PE reads, corresponding to 108 Gb and 126X coverage. The ALLPATH-LG error correction led to a total of around 105 Gb, while Quake strategy led to a smaller number of bases in total (76 Gb). This difference is due to the strict Phred score threshold that was set to 30 during the trimming, which caused the removal of most of the reads from the data set. For MP data, we obtained 123.4 million raw pairs, which decreased to 57.1 million reads after filtering (9.6X). For PacBio long reads, we obtained 552,529 reads after correction with Illumina PE covering 1.8 Gb (2.2X). A summary of the sequencing results is provided in Table 1.

Table 1. *Amphiprion frenatus* genome sequencing statistics. Statistics are reported for both raw and pre-processed data.

Paired-end Library (350 bp insert)					
	# Pairs	# Orphan	Mean Length	Total # bases	Coverage ¹
Raw data	534 892 676	0	101 bp	108 048 320 552	126.1 X
ALLPATH-LG	507 172 071	22 941 362	101 bp	104 765 835 904	122.2 X
QUAKE	335 811 384	88 162 294	99.9 bp	75 932 521 405	88.6 X
Mate pairs library (3 kb insert)					
	# Pairs	# Orphan	Mean Length	Total # bases	Coverage ^a
Raw Data	123 437 124	0	125 bp	30 859 281 000	36.0 X
Unique Reads	96 476 505	0	125 bp	24 119 126 250	28.1 X
Final	57 140 128	3 245 451	70.3 bp	8 260 918 652	9.6 X
PacBio reads					
	# Reads	Mean Length	Median Length	Total # bases	Coverage ^a
Raw Data	660 691	3 480 bp	2 751 bp	2 299 043 897	2.7 X
Corrected reads	552 529	3 634 bp	3 009 bp	1 898 588 929	2.2 X

^a: coverage is calculated with an estimated genome size of *A. frenatus* of 857 Mb.

The frequency of *k*-mers in the ALLPATH-LG module estimated a genome size of 857 Mb, while in Quake strategy, the estimated genome size was 820 Mb. The C-values for *Amphiprion frenatus* are not known, but available C-values for *A. perideraion* range from 0.81 (792 Mb) to 1.22 (1.2 Gb; from Animal Genome Size Database, Gregory, 2017).

We selected the best genome assembly by investigating assembly statistics (N50, maximum scaffold length, number of scaffolds, gap number). The best assembly was achieved with ALLPATH-LG corrected PE and assembled with Platanus (Supplementary Table S1). After further gap-closing with PacBio long reads, the final assembly included 17,801 scaffolds (>1 kb), which covered a total length of 803.3 Mb (Table 2). Although the number of scaffolds is still important, 95% of the assembly is contained in less than 5,000 scaffolds (Figure 1), and the N50 and N90 statistics are 244.5 and 48.1 kb, respectively. The longest scaffold measures 1.7 Mb, and the assembly contains 1.5% of gaps (Table 2). These values are comparable to other published bony fish genomes (Austin, Tan, Croft, Hammer, & Gan, 2015; DiBattista et al., 2016; Nakamura et al., 2013). For example, the genome of the Pacific bluefin tuna (*Thunnus orientalis*) is composed of 16,802 scaffolds (>2 kb), with a N50 of 137 kb and the longest scaffold of 1 Mb (Nakamura et al., 2013). Similarly, the draft genome assembly of the blacktail butterflyfish (*Chaetodon austriacus*) is composed of 13,967 scaffolds (>200 bp), with a N50 of 150.2 kb and 6.85% of gaps (Dibattista et al., 2016).

Table 2. *Amphiprion frenatus* genome assembly statistics.

	Contigs	Scaffolds
Total Assembly Size	790 913 538	803 326 750
# Sequences	102 763	17 801
Longest Sequence (bp)	152 672	1 727 223
Average length (bp)	7 696	45 128
GC content	39.6 %	39.6 %
Non-ATGC characters (%)	0	1.5 %
Number of gaps	0	84 962
Sequences >= 10 Kb	25 615	5 646
Sequences >= 1 Mb	0	10
N50 index (count) ^a	14 928 (15 134)	244 530 (1 001)
N90 index (count) ^a	3 644 (55 579)	48 151 (3 637)

3.2. Nuclear genome validation

The overall mapping rates for PE, MP and RNA-Seq PE data were 99.4%, 98.2% and 92.3%, respectively (Table 3). The distribution of insert sizes estimated from the mapping was similar to the distribution obtained during the library preparation (Table 3 and Supplementary figure S2). Some larger inserts were estimated for RNA-Seq data and are explained by the presence of introns in the genome. The high mapping rates and expected

insert sizes reflect an overall good assembly. This is especially true for RNA-Seq data, which was obtained independently and was not used during genome assembly.

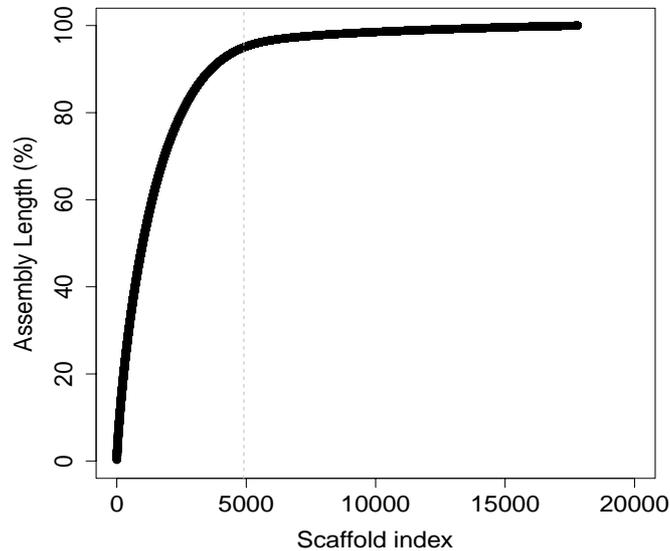


Figure 1. Cumulative length of *Amphiprion frenatus* assembly. Scaffolds are sorted from the longest to the shortest along the horizontal axis. The vertical dotted line indicates the number of scaffolds containing 95% of the assembly.

We omitted around 1.5 million scaffolds of small size (<1 kb; 21% of the assembly) from the final assembly, the majority of which (89.5%) had no matches in the RefSeq database.

We used CEGMA to assess the completeness of the assembly, which resulted in 99% of the core genes being either completely or partially represented in our assembly (Supplementary Table S2). The microsynteny analysis of *A. frenatus* and *O. niloticus* genome gave 2,383 syntenic blocks containing a total of 13,821 (5 intervening genes) and 13,847 (10 intervening genes) genes (Supplementary Table S3).

3.3. Nuclear genome annotation

The amount of repetitive elements in our *A. frenatus* genome was 27.83%. With a combined approach of *ab initio* gene prediction and evidence-based homology, we identified 26,917 genes coding for 31,054 predicted proteins (Supplementary Table S4). All the genes were predicted in a total of 6,497 scaffolds composing the 93% of the total assembly length.

The quality of the models is satisfying, with an average and median annotation edit distance (AED) of 0.19 and 0.14, respectively (Supplementary Figure S3).

Table 3. Mapping rates for paired-end (PE), mate-pairs (MP) and RNA-Seq data. PE and MP reads were mapped with BWA. RNA-Seq data was mapped with HiSat.

	# Reads	# Mapped reads	# Mapped reads (with pair) ^a	# Concordantly mapped ^b	Mapping rate	Average insert size
PE	508 471 016	505 167 148	503 236 407	472 532 563	99.4 %	395.9 bp
MP	114 789 629	112 791 964	111 019 771	88 384 076	98.3 %	3 163 bp
RNA-Seq	377 879 448	348 711 457	332 467 926	328 485 994	92.3 %	278.9

^a: number of reads were both pairs mapped

^b: concordantly mapped: pairs mapping at the expect insert size and with the right orientation

The lengths of *A. frenatus* predicted proteins were compared with the corresponding *O. niloticus* best hits. A total of 28,964 predicted proteins aligned with 22,110 *O. niloticus* targets, and around half (56.3%) had less than 50 amino acid length differences with the target proteins (Supplementary Figure S4, left panel). Additionally, for 20,411 *A. frenatus* proteins, the “query” coverage was higher than 90%. Similarly, the “target” coverage was higher than 90% in 17,419 cases (Supplementary Figure S4, right panel).

The majority of the genes (86.5%) returned a match to SwissProt metazoan proteins. This number further increased to 94.9% when we blasted our data against the RefSeq database. Protein domain annotation was possible for 25,002 genes with 5,397 domains and 2,999 gene ontologies associated with these domains. A total of 17,788 gene ontologies were also mapped to 25,862 proteins (Supplementary Table S5).

The largest number of genes annotated with RefSeq is explained by a lower divergence between *A. frenatus* and the Actinopterygii species selected from the RefSeq database. Indeed, most of the best SwissProt database hits were obtained with human sequences (Supplementary Figure S5). This lower divergence also explains the higher identity for the matches obtained with the RefSeq database (82.1% of average identity) compared with the SwissProt database (61.5% of average identity). Similarly, only 4,607 proteins had identity higher than 80% with proteins from SwissProt, while this number increased to 19,322 for RefSeq (Supplementary Figure S6). When comparing the completeness of *A. frenatus* gene models with other Actinopterygii species, we obtained results similar to *O. niloticus*, *M. zebra* and *D. rerio*, with 14,441 *A. frenatus* proteins having a “query” coverage larger than 90% and

12,605 proteins having a “target” coverage higher than 90%. Similar results were obtained for the *C. austriacus* genome (Figure 2).

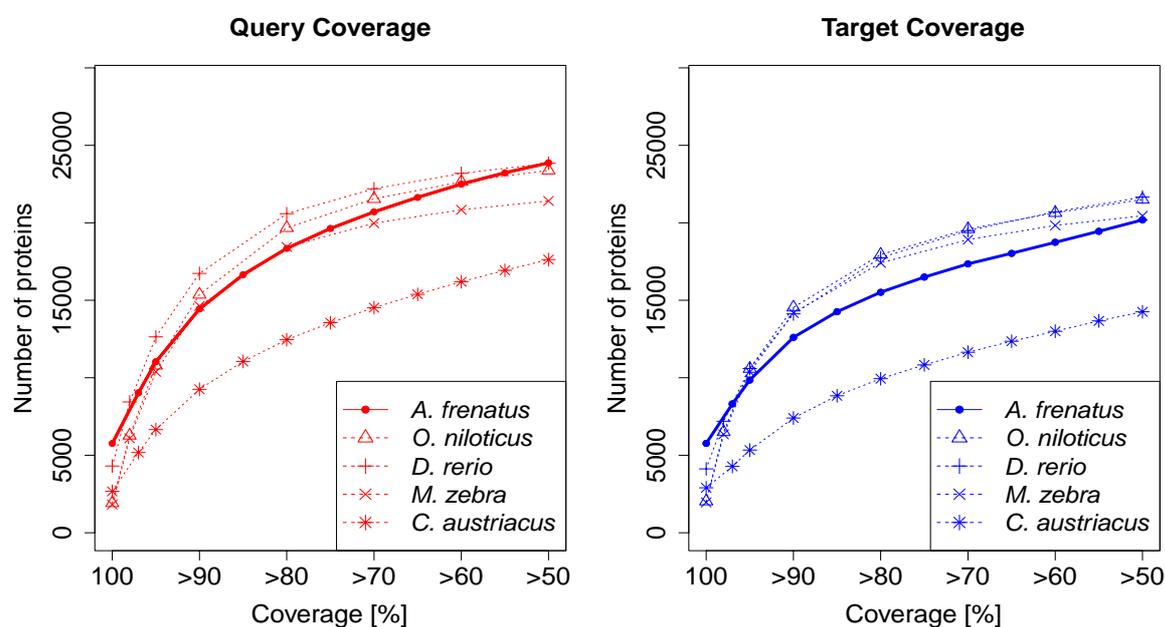


Figure 2. Query (left panel) and target (right panel) coverage for *Amphiprion frenatus*, *Oreochromis niloticus*, *Danio rerio*, *Maylandia zebra* and *Chaetodon austriacus* proteins and their best SwissProt hit proteins.

BUSCO analyses were performed to assess the completeness of *A. frenatus* assembly and annotation. For metazoan BUSCOs, 3.4% of the genes were missing, while for Actinopterygii BUSCOs, 3.5% of the genes were missing (Table 4).

Table 4. BUSCO results for completeness of *Amphiprion frenatus* genome assembly and annotation.

	Actinopterygii		Metazoan	
	Counts	Percent	Counts	Percent
Complete, single-copy	3542	77.3 %	821	83.9 %
Complete, duplicated	737	16.1 %	104	10.6 %
Total Complete	4279	93.3 %	925	94.6 %
Fragmented	150	3.3 %	19	1.9 %
Missing	155	3.4 %	34	3.5 %

3.4. Mitochondrial genome reconstruction

We successfully reconstructed the complete mitochondrial genome of *A. frenatus*. The two methods used gave highly congruent results with each other. The mapping of the reads

onto the inferred sequences led to a mean coverage of 20X (4X to 35X) and confirmed that the sequence could be unambiguously reconstructed. The inferred consensus sequence had a total length of 16,740 bp, which is slightly shorter than the 16,774 bp of the two available *A. frenatus* mitochondrial sequences (GB KJ833752, Li et al., 2015; GB LC089039). Its H-strand nucleotide composition is A: 29.6%, T: 25.7%, C: 29.3% and G: 15.4%, and its GC content is 44.7%. This circular genome has a structure that is typical of fish mitochondrial genomes. It contains 13 protein-coding genes, 22 transfer RNA (tRNAs) genes, 2 ribosomal RNAs (rRNAs), 1 control region (D-loop) plus another 33-bp short noncoding region (OL) located between the tRNA-Asn and the tRNA-Cys (see Supplementary Table S6 and Figure S7 for details). Pairwise differentiation between the three mitochondrial genomes available ranged from 0.77% to 2.0%, suggesting an interesting amount of intraspecific variation in *A. frenatus*.

4. CONCLUSION

Here, we presented the first nuclear genomic resource for a clownfish species. Despite the fragmented nature of our assembly, the overall quality and completeness of the tomato clownfish nuclear genome are satisfying and comparable to other recent bony fish genome assemblies.

The genome that we present here, along with further sequencing of additional species and possible sequencing refinement, provides a new resource for future investigations of clownfish adaptive radiation and their particular life history traits. It will also enable a deeper understanding of the origin of the mutualistic interactions with sea anemones by opening the way for comparative genomic analyses, which could allow the identification of the genomic bases of clownfish adaptive radiation. Additionally, this resource will allow the design of new phylogenetic or population genomic markers that can be useful to study clownfish and damselfish evolution.

SUPPLEMENTARY MATERIAL

Supplementary data are available at *Molecular Ecology Resources* [online](#). To facilitate the access, [Supplementary Information \(Figures and Tables\)](#)¹ are also available directly on OneDrive.

ACKNOWLEDGMENTS

We would like to thank the Vital-IT infrastructure from the Swiss Institute of Bioinformatics for the computational resources and the Lausanne Genomic Technologies Facility for the sequencing. We acknowledge Sacha Laurent and Anna Marcionetti for their help during genome assembly and Dessislava Savova Bianchi for the RNA and DNA extractions. Funding: University of Lausanne funds, Swiss National Science Foundation, Grant Number: 31003A-163428.

AUTHOR CONTRIBUTIONS

Nicolas Salamin, Glenn Litsios and Anna Marcionetti designed the research. Anna Marcionetti and Glenn Litsios obtained the genomic data. Anna Marcionetti performed the genome assembly and validation and participated in the genome annotation. Victor Rossier participated in the genome annotation. Joris Bertrand performed mitochondrial genome assembly and annotation. All authors contributed to the writing of the paper.

DATA ACCESSIBILITY

Raw Illumina and PacBio reads are available in the Sequence Read Archive (SRA), NCBI database (SRA Accession no.: [SRP132439](#)). The assembled nuclear genome, mitogenome and their annotation are available in the DRYAD repository ([DOI: 10.5061/dryad.nv1sv](#)).

REFERENCES

- Andrews, S.** (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Arvedlund, M., McCormick, M. I., Fautin, D. G., & Bildsøe, M.** (1999). Host recognition and possible imprinting in the anemonefish *Amphiprion melanopus* (Pisces: Pomacentridae). In *Marine ecology progress Sseries*, pp. 207-218.

¹https://unils-my.sharepoint.com/:b/g/personal/anna_marcionetti_unil_ch/ERJSjhu8LIZChG_C1iS-4qQBHXSVWTbn-Uly1ty_QaXEta?e=ueRL55

- Austin, C. M., Tan, M. H., Croft, L. J., Hammer, M. P., & Gan, H. M.** (2015). Whole Genome Sequencing of the Asian Arowana (*Scleropages formosus*) Provides Insights into the Evolution of Ray-Finned Fishes. *Genome Biology and Evolution*, 7(10), 2885–2895.
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T.** (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, 27(12), 1691-1692.
- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., ... Di Palma, F.** (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513(7518), 375–381.
- Buston, P. M.** (2003). Social hierarchies: size and growth modification in clownfish. *Nature*, 424(6945), 145-146.
- Buston, P. M.** (2004). Territory inheritance in clownfish. *Proceedings of the Royal Society B: Biological Sciences*, 271(Suppl 4), S252–S254.
- Buston, P. M., & García, M. B.** (2007). An extraordinary life span estimate for the clown anemonefish *Amphiprion percula*. *Journal of Fish Biology*, 70(6), 1710-1719.
- Casas, L., Saborido-Rey, F., Ryu, T., Michell, C., Ravasi, T., & Irigoien, X.** (2016). Sex Change in Clownfish: Molecular Insights from Transcriptome Analysis. *Scientific Reports*, 6, 35461.
- Chaisson, M. J., & Tesler, G.** (2012). Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. *BMC Bioinformatics*, 13, 238.
- Chikhi, R., & Medvedev, P.** (2013). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, 30(1), 31-37.
- Denton, J. F., Lugo-Martinez, J., Tucker, A. E., Schrider, D. R., Warren, W. C., & Hahn, M. W.** (2014). Extensive Error in the Number of Genes Inferred from Draft Genome Assemblies. *PLoS Computational Biology*, 10(12), e1003998.
- Deshpande, V., Fung, E. D., Pham, S., & Bafna, V.** (2013). Cerulean: A hybrid assembly using high throughput short and long reads. In *International Workshop on Algorithms in Bioinformatics* (pp. 349-363). Springer, Berlin, Heidelberg.
- DiBattista, J. D., Wang, X., Saenz-Agudelo, P., Piatek, M. J., Aranda, M., & Berumen, M. L.** (2016). Draft genome of an iconic Red Sea reef fish, the blacktail butterflyfish (*Chaetodon austriacus*): current status and its characteristics. *Molecular ecology resources*. doi:10.1111/1755-0998.12588
- Drillon, G., Carbone, A., & Fischer, G.** (2014). SynChro: a fast and easy tool to reconstruct and visualize synteny blocks along eukaryotic chromosomes. *PLoS One*, 9(3), e92621.
- Elliott, J. K., Elliott, J. M., & Mariscal, R. N.** (1995). Host selection, location, and association behaviors of anemonefishes in field settlement experiments. *Marine Biology*, 122(3), 377-389.
- Elliott, J. K., & Mariscal, R. N.** (2001). Coexistence of nine anemonefish species: differential host and habitat utilization, size and recruitment. *Marine Biology*, 138(1), 23-36.
- English, A. C., Richards, S., Han, Y., Wang, M., Vee, V., Qu, J., ... Gibbs, R. A.** (2012). Mind the Gap: Upgrading Genomes with Pacific Biosciences RS Long-Read Sequencing Technology. *PLoS ONE*, 7(11), e47768.
- Fautin, D., & Allen, G. R.** (1997). Anemone fishes and their host sea anemones. Western Australian Museum, Perth, 159 pp.

- Finn, R. D., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., ... Bateman, A.** (2016). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, *44* (Database issue), D279–D285.
- Gavrilets, S., & Losos, J. B.** (2009). Adaptive radiation: contrasting theory with data. *Science*, *323*(5915), 732–737.
- Gavrilets, S., & Vose, A.** (2005). Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(50), 18040–18045.
- Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., ... Jaffe, D. B.** (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(4), 1513–1518.
- Goodwin, S., McPherson, J. D., & McCombie, W. R.** (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*, *17*(6), 333–351.
- Gregory, T. R.** (2017). Animal genome size database. <http://www.genomesize.com>.
- Hackl, T., Hedrich, R., Schultz, J., & Förster, F.** (2014). *proofread*: large-scale high-accuracy PacBio correction through iterative short read consensus. *Bioinformatics*, *30*(21), 3004–3011.
- Hahn, C., Bachmann, L., & Chevreur, B.** (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Research*, *41*(13), e129.
- Hattori, A.** (2000). Social and mating systems of the protandrous anemonefish *Amphiprion perideraion* under the influence of a larger congener. *Austral Ecology*, *25*(2), 187–192.
- Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M., & Stanke, M.** (2015). BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics*, *32*(5), 767–769.
- Holt, C., & Yandell, M.** (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics*, *12*, 491.
- Huebner, L. K., Dailey, B., Titus, B. M., Khalaf, M., & Chadwick, N. E.** (2012). Host preference and habitat segregation among Red Sea anemonefish: effects of sea anemone traits and fish life stages. *Marine Ecology Progress Series*, *464*, 1–15.
- Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T. P., ... Nishida, M.** (2013). MitoFish and MitoAnnotator: A Mitochondrial Genome Database of Fish with an Accurate and Automatic Annotation Pipeline. *Molecular Biology and Evolution*, *30*(11), 2531–2540.
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... & Pesseat, S.** (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236–1240.
- Joshi, N. A., & Fass, J. N.** (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files [Software].
- Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., ... & Kohara, Y.** (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome research*, *24*(8), 1384–1395.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A.** (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647–1649.

- Kelley, D. R., Schatz, M. C., & Salzberg, S. L.** (2010). Quake: quality-aware detection and correction of sequencing errors. *Genome biology*, *11*(11), R116.
- Kim, N. N., Jin, D. H., Lee, J., Kil, G. S., & Choi, C. Y.** (2010). Upregulation of estrogen receptor subtypes and vitellogenin mRNA in cinnamon clownfish *Amphiprion melanopus* during the sex change process: profiles on effects of 17 β -estradiol. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *157*(2), 198-204.
- Kim, D., Langmead, B., & Salzberg, S. L.** (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, *12*(4), 357-360.
- Kim, N. N., Lee, J., Habibi, H. R., & Choi, C. Y.** (2013). Molecular cloning and expression of caspase-3 in the protandrous cinnamon clownfish, *Amphiprion melanopus*, during sex change. *Fish physiology and biochemistry*, *39*(3), 417-429.
- Kim, N. N., Shin, H. S., Habibi, H. R., Lee, J., & Choi, C. Y.** (2012). Expression profiles of three types of GnRH during sex-change in the protandrous cinnamon clownfish, *Amphiprion melanopus*: Effects of exogenous GnRHs. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *161*(2), 124-133.
- Koren, S., Schatz, M. C., Walenz, B. P., Martin, J., Howard, J., Ganapathy, G., ... Phillippy, A. M.** (2012). Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature Biotechnology*, *30*(7), 693–700.
- Li, J., Chen, X., Kang, B., & Liu, M.** (2015). Mitochondrial DNA Genomes Organization and Phylogenetic Relationships Analysis of Eight Anemonefishes (Pomacentridae: Amphiprioninae). *PLoS ONE*, *10*(4), e0123894.
- Li, H., & Durbin, R.** (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754-1760.
- Litsios, G., Pearman, P. B., Lanterbecq, D., Tolou, N., & Salamin, N.** (2014). The radiation of the clownfishes has two geographical replicates. *Journal of biogeography*, *41*(11), 2140-2149.
- Litsios, G., & Salamin, N.** (2014). Hybridisation and diversification in the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, *14*, 245.
- Litsios, G., Sims, C. A., Wüest, R. O., Pearman, P. B., Zimmermann, N. E., & Salamin, N.** (2012). Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, *12*, 212.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., ... Wang, J.** (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience*, *1*, 18.
- Marçais, G., & Kingsford, C.** (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*, *27*(6), 764–770.
- Mebs, D.** (2009). Chemical biology of the mutualistic relationships of sea anemones with fish and crustaceans. *Toxicon*, *54*(8), 1071-1074.
- Miller, J. R., Zhou, P., Mudge, J., Gurtowski, J., Lee, H., Ramaraj, T., ... Silverstein, K. A. T.** (2017). Hybrid assembly with long and short reads improves discovery of gene family expansions. *BMC Genomics*, *18*, 541.

- Mitchell, J. S.** (2003). Social correlates of reproductive success in false clown anemonefish: subordinate group members do not pay-to-stay. *Evolutionary Ecology Research*, 5(1), 89-104.
- Miura, S., Kobayashi, Y., Bhandari, R. K., & Nakamura, M.** (2013). Estrogen favors the differentiation of ovarian tissues in the ambisexual gonads of anemonefish *Amphiprion clarkii*. *Journal of Experimental Zoology Part A: Ecological Genetics and Physiology*, 319(10), 560-568.
- Nakamura, Y., Mori, K., Saitoh, K., Oshima, K., Mekuchi, M., Sugaya, T., ... Inouye, K.** (2013). Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. *Proceedings of the National Academy of Sciences of the United States of America*, 110(27), 11061–11066.
- Ollerton, J., McCollin, D., Fautin, D. G., & Allen, G. R.** (2007). Finding NEMO: nestedness engendered by mutualistic organization in anemonefish and their hosts. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609), 591–598.
- Parra, G., Bradnam, K., & Korf, I.** (2007). CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*, 23(9), 1061-1067.
- Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., ... Yorke, J. A.** (2012). GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, 22(3), 557–567.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M.** (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212.
- Smit, A. F. A., Hubley, R., & Green, P.** (2015). RepeatMasker Open-4.0. 2013–2015. *Institute for Systems Biology*. <http://repeatmasker.org>.
- Steinke, D., Zemlak, T. S., & Hebert, P. D. N.** (2009). Barcoding Nemo: DNA-Based Identifications for the Ornamental Fish Trade. *PLoS ONE*, 4(7), e6300.
- Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., ... Pachter, L.** (2010). Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology*, 28(5), 511–515.
- Zhu, W., Wang, L., Dong, Z., Chen, X., Song, F., Liu, N., ... & Fu, J.** (2016). Comparative transcriptome analysis identifies candidate genes related to skin color differentiation in red tilapia. *Scientific reports*, 6(1), 1-12.

Chapter 2

Insights into the genomics of clownfish adaptive radiation: genetic basis of the mutualism with sea anemones

From the article: **Marcionetti, A., Rossier, V., Roux, N., Salis, P., Laudet, V., & Salamin, N. 2019.** Insights into the genomics of clownfish adaptive radiation: Genetic basis of the mutualism with sea anemones. *Genome biology and evolution*, 11(3), 869-882. [DOI:10.1093/gbe/evz042](https://doi.org/10.1093/gbe/evz042)

Insights into the genomics of clownfish adaptive radiation: genetic basis of the mutualism with sea anemones

Anna Marcionetti¹, Victor Rossier¹, Natacha Roux², Pauline Salis², Vincent Laudet², Nicolas Salamin¹

¹ *Department of Computational Biology, Genopode, University of Lausanne, 1015 Lausanne, Switzerland*

² *Observatoire Océanologique de Banyuls-sur-Mer, UMR CNRS 7232 BIOM ; Sorbonne University ; 1, Avenue Pierre Fabre, 66660 Banyuls-sur-Mer, France*

ABSTRACT

Clownfishes are an iconic group of coral reef fishes, especially known for their mutualism with sea anemones. This mutualism is particularly interesting as it likely acted as the key innovation that triggered clownfish adaptive radiation. Indeed, after the acquisition of mutualism, clownfishes diversified into multiple ecological niches linked with host and habitat use. However, despite the importance of this mutualism, the genetic mechanisms allowing clownfishes to interact with sea anemones are still unclear.

Here, we used comparative genomics and molecular evolutionary analyses to investigate the genetic basis of clownfish mutualism with sea anemones. We assembled and annotated the genome of nine clownfish species and one closely related outgroup. Orthologous genes inferred between these species and additional publicly available teleost genomes resulted in almost 16,000 genes that were tested for positively selected substitutions potentially involved in the adaptation of clownfish to live in sea anemones. We identified 17 genes with a signal of positive selection at the origin of clownfish radiation. Two of them (Versican core protein and Protein O-GlcNAcase) show particularly interesting functions associated with *N*-acetylated sugars, which are known to be involved in sea anemone discharge of toxins.

This study provides the first insights into the genetic mechanisms of clownfish mutualism with sea anemones. Indeed, we identified the first candidate genes likely to be associated with clownfish protection from sea anemones, and thus the evolution of their mutualism. Additionally, the genomic resources acquired represent a valuable resource for further investigation of the genomic basis of clownfish adaptive radiation.

KEYWORDS

Anemonefish; *Amphiprion*; Coral Reef fish; Positive selection; Key-innovation

1. INTRODUCTION

The spectacular diversity of life on Earth that Darwin sought to explain in *On the Origin of Species* (Darwin, 1859) emerged through a variety of complex biological processes. One of these is adaptive radiation, during which a single ancestral species diversifies into many descendants adapted to a wide range of ecological conditions. It is considered of crucial importance and potentially responsible for much of the diversity of life (Simpson, 1953; Schluter, 2000). However, the process of adaptive radiation is an extremely complex process influenced by a variety of ecological, genetic, and developmental factors, and since decades researchers have been trying to understand the causes, consequences, and mechanisms of this process (Simpson, 1953; Givnish & Sytsma, 1997; Schluter, 2000; Givnish, 2015; Soulebeau et al., 2015).

Current theories postulate that adaptive radiation starts with ecological opportunity, in which an ancestral species occupies an environment with abundant and underused resources (Yoder et al., 2010; Stroud & Losos, 2016). Divergent natural selection among these different resources should subsequently drive the adaptive diversification of the ancestral species through ecological speciation (Rundell & Price, 2009). This starting ecological opportunity is seen in empirical studies, with clades diversifying after the colonization of isolated areas (e.g. Galapagos finches: Grant & Grant, 2008; African Rift Lake cichlids: Seehausen, 2006; Caribbean *Anolis* lizards: Losos, 2009), following the appearance of new habitat and resources (e.g. grasses and grazing horses in MacFadden, 2005), after an extinction event (e.g. Erwin, 2007), or following the evolution of traits (i.e., key innovation) allowing the interaction with the environment in a novel way (e.g. the evolution of flight in bats in Simmons et al., 2008; the evolution of the pharyngeal jaw apparatus in cichlids and labrid fishes in Mabuchi et al., 2007; the evolution of antifreeze glycoproteins in Antarctic notothenioid fishes in Near et al., 2012).

The importance of ecological opportunity was also emphasized by modeling approaches aiming at identifying the general patterns that should be observed during adaptive radiations (Gavrilets & Vose, 2005; Gavrilets & Losos, 2009). Other general patterns predicted by these studies include patterns of evolutionary rates, geographical components of speciation,

selection intensity, and genomic architecture (Gavrilets & Vose, 2005; Gavrilets & Losos, 2009). Until recently, however, empirical studies describing adaptive radiations were not able to fully assess the predictions made by those models, as the necessary deep genomic data were missing. This data starts to be available for iconic clades such as cichlids (Brawand et al., 2014), sticklebacks (Jones et al., 2012), *Heliconius* butterflies (Dasmahapatra et al., 2012; Supple et al., 2013), and Darwin's finches (Lamichhaney et al., 2015). These studies revealed the first insight on the genomic mechanisms of adaptive radiations, with for example, the reuse of standing variation having an important role in the evolution of sticklebacks and cichlids (Jones et al., 2012; Brawand et al., 2014), and introgressive hybridization playing a role in *Heliconius* and Darwin's finches diversification (Dasmahapatra et al., 2012; Lamichhaney et al., 2015).

Despite these empirical studies, modeling approaches and acquired genomic data, much remains to be understood about the general mechanisms of adaptive radiations. This is particularly true for marine ecosystems, where described cases of adaptive radiations remain scarce (e.g., Notothenioid fish in Antarctica, Near et al., 2012) as barriers to dispersal are uncommon, making ecological speciation less likely than in more isolated landscapes (Puebla, 2009). Therefore, to obtain a wider overview of the processes underlying adaptive radiations, it is essential to step back from classical textbook examples of adaptive radiations and gather data from less studied clades occurring in different ecosystems. One interesting case of recently described adaptive radiation in marine environments is represented by clownfishes (family Pomacentridae, genera *Amphiprion* and *Premnas*, Litsios et al., 2012).

Clownfishes are an iconic group of coral reef fishes distributed in the tropical belt of the Indo-Pacific Ocean, and it includes 28 currently recognized species and 2 natural hybrids (Fautin & Allen, 1997; Ollerton et al., 2007; Gainsford et al., 2015). A distinctive characteristic of this group is the mutualistic interaction they maintain with sea anemones (Fautin & Allen, 1997; Figure 1). This mutualism is particularly important as it was proposed to act as the key innovation that triggered clownfish adaptive radiation (Litsios et al., 2012). Indeed, after the acquisition of mutualism, clownfishes diversified into multiple ecological niches linked with both host and habitat use (Litsios et al., 2012).

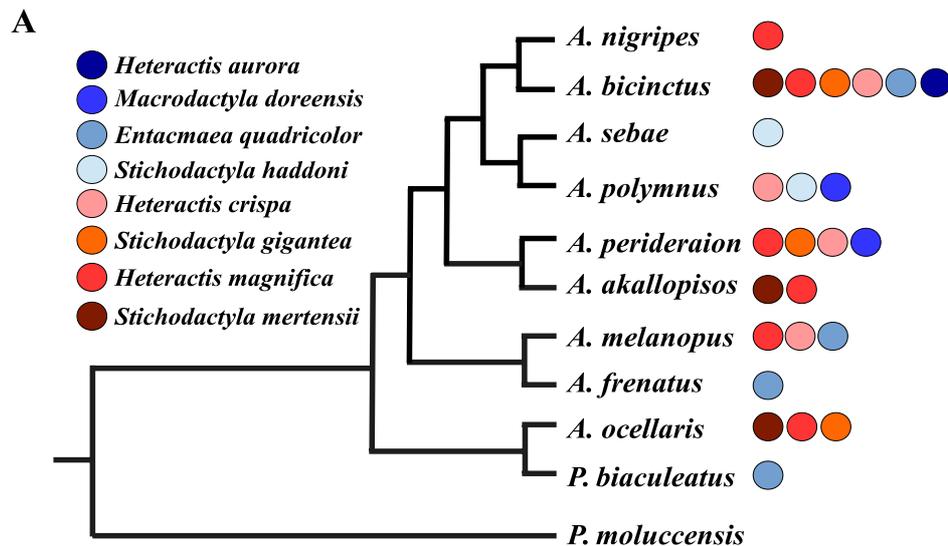


Figure 1. (A) Phylogenetic relationship of the nine selected clownfish species, *Amphiprion frenatus* (available from Marcionetti et al., 2018), and the outgroup species *Pomacentrus moluccensis*. Circles represent the sea anemones species with whom each clownfish can interact (Fautin & Allen, 1997). Closely related species with divergent host usages were selected. (B) and (C) show, respectively, *A. nigripes* and *A. ocellaris* in their host sea anemone *Heteractis magnifica*.

We can today take advantage of next-generation sequencing technologies to obtain genomes of different clownfish species to better understand the mechanisms of clownfish adaptation to sea anemones. By considering mutualism as a new and advantageous phenotype that evolved in clownfishes, we can investigate the role of selection on the genetic basis of the adaptation. Indeed, phenotypic evolution may occur through alterations of the structure of protein-coding genes, which can be fixed by positive selection if they confer an advantage (as, e.g., in Spady et al., 2005; Hoekstra et al., 2006; Protas et al., 2006; Lynch, 2007). In this study, we obtained genomic data for several clownfish species and tested the genetic mechanisms underlying clownfish protection from sea anemone toxins using comparative genomic and molecular evolution analyses. We hypothesized that this protection could be granted by positively selected substitutions modifying the original function of protein-coding

genes in a way that ultimately prevent the release of sea anemone toxins or provide immunity to these toxins. These mechanisms resulted in the mutualism with sea anemones, which acted as the probable key innovation that triggered clownfish adaptive radiation. Thus, this study will not only improve our understanding of the genetic mechanisms involved at the beginning of an adaptive radiation, but it will also provide data for further investigation of the diversification process in marine environments.

2. MATERIAL AND METHODS

2.1. Species selection, DNA extraction, library preparation

We selected nine clownfish species (*Premnas biaculeatus*, *Amphiprion ocellaris*, *A. perideraion*, *A. akallopisos*, *A. polymnus*, *A. sebae*, *A. melanopus*, *A. bicinctus*, *A. nigripes*) spanning the whole clownfish divergence and the whole distribution range of the group. Genomic data from one additional species (*A. frenatus*) were already available (Marcionetti et al., 2018). This total of ten species forms five pairs of closely related but ecologically divergent species in their host and habitat usage (Figure 1). The lemon damselfish (*Pomacentrus moluccensis*) was selected as a closely related outgroup species whose estimated divergence with clownfishes ranged from 21.5 to 38.5 Ma depending on the study (Litsios et al., 2012; Sanciangco et al., 2016).

One individual of each clownfish species and *P. moluccensis* was obtained from a local aquarium shop. Because all individuals were acquired from an aquarium shop, their exact origin is not available. All individuals passed away beforehand at the aquarium shop, and samples from deceased fish were received. Thus, all the individuals sampled did not undergo any manipulation or experimentation in the laboratory. All remaining samples are stored at the Department of Computational Biology, University of Lausanne (Switzerland).

For each species, genomic DNA (gDNA) was extracted from 50 mg of fin tissue using DNeasy Blood & Tissue Kit (Quiagen, Hilden, Germany) and following the manufacturer's instructions. Short-insert (350 bp) paired-end (PE) libraries were prepared from 100 ng of gDNA at the Lausanne Genomic Technologies Facility (LGTF, Switzerland), using TruSeq Nano DNA LT Library Preparation Kit (Illumina). PE libraries of *A. ocellaris* and *P. moluccensis* were sequenced on two lanes of Illumina HiSeq2000 at the LGTF, while PE libraries for the other species were each sequenced on one lane. For *A. ocellaris*, a long-insert (3 kb) mate pairs (MP) library was prepared from 4 µg of gDNA at Fasteris SA (Geneva,

Switzerland) using the Nextera Mate Pair Library Preparation Kit from Illumina. This MP library was sequenced on a half lane of Illumina HiSeq2500 at Fasteris.

2.2. Whole-Genome Assemblies

Because we needed to acquire genomic data for ten different species, we investigated an alternative strategy for genome assembly that allowed for reduced coverage and library types, as well as decreased computational time and memory usage during the assembly process. This strategy consisted of using an available reference genome of a species as the substrate to reconstruct the genome of a second species. Such an approach is conceivable only if the divergence between the considered species is low, and if large genomic rearrangements did not occur since the split of those species. Because clownfishes are a fast-diversifying group with most of the diversification occurring 5 Ma (Litsios et al., 2012), we did not expect to observe high divergence and large genomic rearrangements within the group. Thus, we investigated the feasibility of such a reference-based approach in clownfishes by assembling *A. ocellaris* genome with both *de novo* and reference-based strategies, and by comparing the results. Similar methods taking advantage of reference genomes from closely related species for the assembly of new species are also reported in the literature (Buza et al., 2015; Lischer & Shimizu, 2017), with for instance the genomes of *Arabidopsis thaliana* (Schneeberger et al., 2011) or *Tetraodon tetrix* (Wang et al., 2014) being obtained successfully by using a reference to guide their assemblies.

The processing of sequenced reads for all species and the *de novo* genome assembly of *A. ocellaris* were performed as reported in Marcionetti et al. (2018; more details in Supplementary Material & Methods, Supplementary Material online). Reference-based assembly of *A. ocellaris* was performed using half of the original coverage (one Illumina lane, around 50X) and employing *A. frenatus* genome as the reference. For this, we mapped processed PE reads of *A. ocellaris* against the assembly of *A. frenatus* using Stampy (version 1.0.28; Lunter & Goodson, 2011), setting the expected substitution rate parameter to 0.05 to allow the mapping of reads including substitutions. We retrieved the consensus sequences with SAMtools (version 1.3; Li et al., 2009) and we closed gaps with GapCloser (from SOAPdenovo2, version 2.04.240; Luo et al., 2012). The remaining species were also assembled following this reference-guided assembly strategy and using the entire set of processed reads (total of one Illumina lane per species).

2.3. Validation of the Reference-Based Assembly Strategy

To validate the reference-based approach, we compared assembly statistics and mapping rates of the *de novo* and reference-guided assemblies of *A. ocellaris*. Because it is difficult to perform synteny analysis with fragmented assemblies, we used SynMap2 (Haug-Baltzell et al., 2017) to investigate the synteny and collinearity between the recently available *A. percula* genome (Lehmann et al., 2018) and the two *A. ocellaris* assemblies. We reordered *A. ocellaris* scaffolds according to the alignment regions of *A. percula* genome and we plotted the synteny in R (R Core Team, 2013).

To confirm that the reference-guided assembly method resulted in the correct reconstruction of species sequences, we reconstructed a phylogeny containing additional publicly available clownfish samples. Only eight nuclear gene sequences were available for these additional samples (BMP-4, Glyt, Hox6, RAG1, RH, S7, SVEP1, Zic1; GenBank ID in Supplementary Table S1, Supplementary Material online). We extracted these genes from the obtained assemblies based on the functional annotation of the genomes. We aligned the genes using Mafft (version 7.305; Katoh & Standley, 2013) and we concatenated the alignments within Geneious (version 10.0.5; Kearse et al., 2012). We constructed the gene trees for each separate alignment and for the concatenate one with PhyML (version 3.3; GTR + Γ model, bootstrap 100, Guindon et al., 2010). The trees were plotted with Dendroscope (version 1.4; Huson et al., 2007) and they were visually examined for inconsistency in topology.

2.4. Genome Quality Investigation and Genome Annotation

We assessed the quality of all the obtained assemblies (the *de novo* *A. ocellaris* assembly and all the reference-guided assemblies) and we structurally and functionally annotated them as performed in Marcionetti et al. (2018, more details in Supplementary Material & Methods, Supplementary Material online). The completeness of the genome annotation was investigated with BUSCO (version 1.0; data set: vertebrates; Simão et al., 2015). For each species, we calculated the sequence coverage (proportion of the sequence covered by mapped reads) and average depth (average number of reads mapping to the gene) with bedtools coverage (version 2.22.1; Quinlan & Hall, 2010).

2.5. Orthology inference, HOG filtering and classification

We inferred orthologous genes between the ten clownfish species, *P. moluccensis* and 12 publicly available Actinopterygii species (*Astyanax mexicanus*, *Danio rerio*, *Gadus*

morhua, *Gasterosteus aculeatus*, *Lepisosteus oculatus*, *Oreochromis niloticus*, *Oryzias latipes*, *Poecilia formosa*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Xiphophorus maculatus*, and *Stegastes partitus*, Supplementary Table S2 and Figure S1, Supplementary Material online). The use of additional Actinopterygii species was necessary for the positive selection analysis. Indeed, the power in detecting patterns of positive selection is increasing with increasing taxa (Anisimova et al., 2002). Orthology inference was performed with OMA standalone (v.1.0.6, Altenhoff et al., 2013) on the proteomes of the 23 species, using the species tree represented in Supplementary Figure S1, Supplementary Material online, to guide the clustering of orthologous pairs. For each species and gene, the longest protein isoform was used for orthology inference. The resulting Hierarchical Orthologous Groups (HOGs) were filtered to keep only HOGs containing both clownfish and outgroup species, with a minimum number of species required set to six species. Additionally, only HOGs containing sequences for *P. moluccensis* were kept, as this species corresponds to the most closely related species to clownfish, and it is necessary for specifically aiming at the ancestral branch of clownfish group (Figure 2).

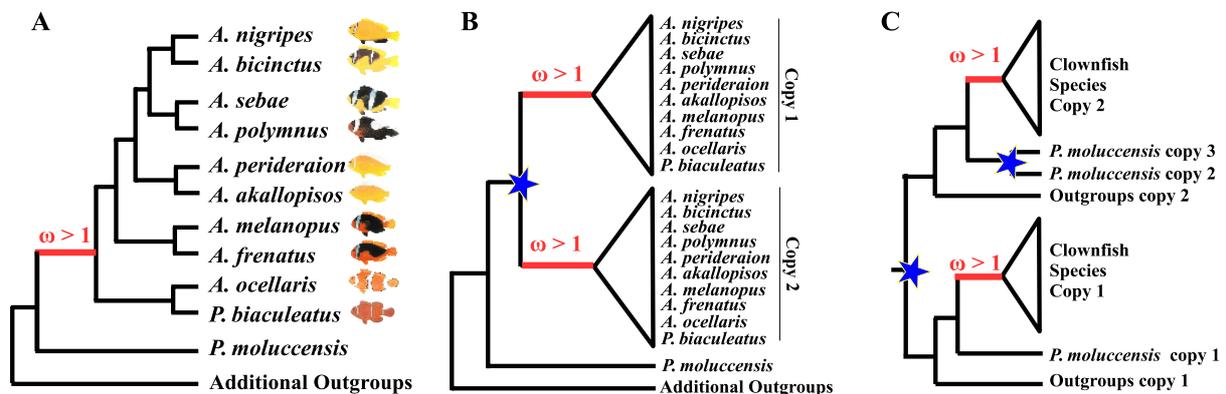


Figure 2. Examples of gene trees for 1-to-1 OG (A), clownfish-specific duplicated genes (B) and overall multicopy HOGs (C). Mutualism with sea anemones appeared on the ancestral basis of clownfishes (in red in A). Genes were tested for positive selection ($\omega > 1$) on branches specific to all clownfishes (in red in A, B & C). Gene duplication events are visualized with blue stars.

HOGs were classified as single-copy orthologs (1-to-1 OG), clownfish-specific duplicated genes (i.e., genes with potential duplications event on the branch leading to clownfish), and overall multicopy orthologs. Single-copy orthologs were obtained by selecting HOGs with one sequence per species at different taxonomic levels. We defined “level 1” as all species being kept, “level 2” where *L. oculatus* was removed, “level 3” where

L. oculatus, *D. rerio*, and *A. mexicanus* were removed and “level 4” where *L. oculatus*, *D. rerio*, *A. mexicanus*, and *G. morhua* were removed (Supplementary Figure S1, Supplementary Material online). HOGs were classified as clownfish-specific duplicated genes when the minimal number of gene copies in clownfishes was higher than the maximum number of gene copies in the outgroup species. This strategy allows for possible incomplete annotation of both clownfish and outgroup genomes to be accounted for. A minimum number of two outgroups was required for all analyses, and the four different taxonomic levels (Supplementary Figure S1, Supplementary Material online) were considered. To identify potential false positives, we investigated the coverage (proportion of sequence covered by mapped reads, and the number of mapped reads) and length of clownfish-specific duplicated genes. The remaining HOGs were classified as overall multicopy orthologs.

2.6. Positive selection analysis

All HOGs resulting from orthology inference were composed by the longest protein isoforms of each gene and species. For each HOG, we performed protein alignments with MAFFT (version 7.305; G-INS-i strategy; Katoh & Standley, 2013), with the option “--allowshift”. Codon alignments were inferred from protein alignments with PAL2NAL (Suyama et al., 2006). Because positive selection analyses are sensitive to alignment errors (Fletcher & Yang, 2010), we filtered the alignments to keep only highly confident homologous regions. For this, we followed a stringent filtering approach proposed in the Selectome database (Moretti et al., 2014). Details are available in Supplementary Material & Methods, Supplementary Material online. The strict filtering strategy also allows reducing false positives potentially arising from the use of different isoforms for different species in each HOG, as mentioned in Villanueva-Cañas et al. (2013). Gene trees were obtained with PhyML (version 3.3; Guindon et al., 2010) from the unfiltered codon alignments. For each HOG, the gene tree was reconstructed with both HKY85 and GTR substitution models (100 bootstrap). The best model was selected with a likelihood ratio test (df=4).

For 1-to-1 OGs, positive selection was tested with CodeML implemented in the PAML package (version 4.9; Yang, 2007), using the filtered codon alignments and obtained gene trees. We tested for positive selection at the onset of the clownfish radiation with the “branch-site model”, by setting the branch leading to the clownfish as foreground branch and all other branches as the background (Figure 2A). The null model (with foreground ω constrained to be smaller or equal to 1) was compared with the alternative model (with estimation of

foreground ω) with a likelihood ratio test ($df=1$). We corrected for multiple-testing with the Benjamini-Hochberg method implemented in the q-value package in R (FDR threshold of 0.1; Dabney et al., 2010). Additional information is reported in Supplementary Material & Methods, Supplementary Material online.

For clownfish-specific duplicated genes and overall multicopy HOGs, positive selection was tested with the method aBSREL implemented in HyPhy (version 2.3.7; Smith et al., 2015). The analysis was run in an exploratory way, testing for positive selection at each branch (Figure 2B and C). Although this approach reduces the power due to multiple testing, it was preferred as we do not know *a priori* which copy of the genes may be positively selected. We corrected for multiple-testing with the Benjamini-Hochberg method implemented in the q-value package in R (FDR threshold of 0.1; Dabney et al., 2010).

Positively selected HOGs were annotated by retrieving the SwissProt ID annotation of genes forming the HOGs. We ensured that all genes of different species forming the HOGs were annotated with the same function. Gene trees were plotted with FigTree (version 1.4.2; Rambaut, 2014).

2.7. Comparison of gene trees vs species tree approaches

The tree topology has an effect on the inference of positive selection (Diekmann & Pereira-Leal, 2015), and the use of either gene trees or the species tree may lead to different results if topology incongruence is present. We investigated the effect of using gene trees or the species tree in the positive selection analysis by randomly selecting 5,000 1-to-1 OGs and inferring positive selection using the species tree as input tree. We investigated the level of topology incongruence in the randomly selected data set by calculating the unweighted Robinson-Foulds (uRF) distance between the species tree and the gene tree using the python library DendroPy (Sukumaran & Holder, 2010) and compared it with the results of positive-selection measured as the number of significant results, both before (p -values < 0.05) and after (q -values < 0.05) multiple-testing correction. More information is available in Supplementary Material & Methods, Supplementary Material online.

2.8. Power and Type I error in positive selection analyses

We investigated the power to detect positive selection on the branch leading to clownfishes by simulating data using the software evolver in the PAML package (version 4.9; Yang, 2007), and by testing positive selection on the simulated data with CodeML. We

simulated codon alignments (alignment length: 5,000, 1,000 and 550 codons) under the branch-site model, with ω varying both among sites and branches, to match the model used in the positive selection analyses. We generated trees following the species tree topology, and with branch lengths randomly drawn from the branch lengths distributions obtained from all gene trees of analyzed HOGs. Different selection strengths were simulated, with ω values ranging from 2 to 900. To assess the level of Type I errors in the analysis, we also simulated codon alignments without positive selection ($\omega = 0.5$ and $\omega = 1$ on the foreground branch). For each alignment length, randomly generated tree, and ω value we simulated four sets of sequences (Supplementary Table S3, Supplementary Material online).

Simulated codon alignments were tested for positive selection with CodeML (PAML version 4.9; Yang, 2007), applying the same pipeline developed for the test of positive selection on 1-to-1 OG. We investigated the power to detect positive selection and the number of false positives (Type I errors) by recording the number of significant LRT (p-value < 0.05) between the null model and the alternative model. More information on this analysis is available in Supplementary Material & Methods, Supplementary Material online.

3. RESULTS

3.1. Genome Assemblies, Quality Assessment and Annotations

For all species, paired-end (PE) sequencing and reads processing with ALLPATH-LG module (Gnerre et al., 2011) performed well. This resulted in an average coverage of 125.8X for *A. ocellaris* (sequenced on two Illumina lanes), while an average coverage between 36.5X (*A. sebae*) and 54.7X (*A. polymnus*) was obtained for the other species (sequenced on a single Illumina lane; Supplementary Table S4, Supplementary Material online). The sequencing of long-insert mate-pairs (MP) for *A. ocellaris* resulted in a low level of unique reads (31.8%), which corresponds to a final genomic coverage of 3.5X (Supplementary Table S4, Supplementary Material online).

The higher coverage and different library types for *A. ocellaris* were necessary because a classical *de novo* approach was also used to assemble the genome of this species. The best *de novo* assembly for *A. ocellaris* was obtained with ALLPATH-LG processed reads assembled with PLATANUS (total assembly size of 744 Mb, 27,951 scaffolds, N50 of 136 kb; Table 1 and Supplementary Table S5A, Supplementary Material online). The fragmentation of the assembly is mainly due to the low number of unique MP, which

prevented an optimal scaffolding. Reference-guided assemblies for *A. ocellaris* (obtained with only half of the original PE coverage and without the use of MP) and for the additional species were less fragmented. This is because they were constructed based on the genome of *A. frenatus*, and therefore statistics for these assemblies mainly reflect the ones of *A. frenatus* genome (Marcionetti et al., 2018; Table 1 and Supplementary Table S5B, Supplementary Material online).

Table 1. Genome assembly and annotation statistics for the nine assembled clownfish species and *Pomacentrus moluccensis*. For *A. ocellaris*, statistics of both *de novo* and reference-guided assemblies are reported. Reference-guided assemblies were obtained using *A. frenatus* (Marcionetti et al. 2018) as reference genome. N50 index indicates the shortest scaffold length above which 50% of the genome is assembled. CEGMA and BUSCOs genes represent the completeness of the genome assemblies and annotations, respectively.

	De Novo assembly	Reference-Guided assembly				
	<i>A. ocellaris</i>	<i>A. ocellaris</i>	<i>A. bicinctus</i>	<i>A. nigripes</i>	<i>A. polymnus</i>	<i>A. sebae</i>
Total Assembly Size (Mb)	744	798	799	800	800	799
Number of scaffolds	27951	16543	16953	16995	17050	16941
N50 (bp)	136417	246482	246127	246124	246119	245870
Non-ATGC characters (%)	4.6	3.6	2.8	2.7	2.7	2.9
Paired-ends mapping rate (%)	95.3	98.2	98.9	98.9	99.0	99.0
Number of genes	24383	29913	28891	28558	28640	28727
Number of proteins	27606	33845	33219	32905	33128	33271
Functional annotated proteins (%)	94.0	92.7	93.2	93.1	92.9	92.9
CEGMA genes in assembly (%)	97.2	99.6	99.6	99.6	99.6	100
BUSCOs genes in annotation (%)	87	93	94	95	95	95

	Reference-guided assembly				
	<i>A. akallopisos</i>	<i>A. perideraion</i>	<i>A. melanopus</i>	<i>P. biaculeatus</i>	<i>P. moluccensis</i>
Total Assembly Size (Mb)	801	801	803	797	794
Number of scaffolds	17172	17212	17399	16164	15505
N50 (bp)	246052	246037	245703	247121	246470
Non-ATGC characters (%)	1.9	2.0	1.4	2.9	7.9
Paired-ends mapping rate (%)	99.0	99.0	97.4	96.9	81.2
Number of genes	28730	29014	29408	28170	28885
Number of proteins	33120	33320	33768	32385	32027
Functional annotated proteins (%)	93.1	92.9	92.7	93.8	94.0
CEGMA genes in assembly (%)	99.6	99.6	99.6	99	99.6
BUSCOs genes in annotation (%)	95	94	94	95	89

The completeness of the obtained assemblies was assessed with CEGMA. As for *A. frenatus* genome, reference-guided assemblies resulted in 99% to 100% of the core genes being either completely or partially represented in the assembly of the different species. Because of the larger fragmentation, this number is slightly decreased in *A. ocellaris de novo* assembly, with only 97.2% of the genes being retrieved (Table 1 and Supplementary Table S6, Supplementary Material online).

To assess the correct reconstruction of the genomic sequence for each species, we investigated the mapping statistics of PE against the assembled genomes. Here as well, slightly better results were obtained for reference-guided assemblies compared with the *A. ocellaris de novo* assembly. Indeed, depending on the species, between 97% and 99% of reads mapped against the corresponding reference-guided assembly, while only 95% of PE reads of *A. ocellaris* mapped against its *de novo* assembly (Table 1 and Supplementary Table S7, Supplementary Material online). Additionally, to validate the reference-guided assembly strategy, we performed synteny analysis of the *de novo* and reference-guided assemblies of *A. ocellaris* and the recently available *A. percula* genome (Lehmann et al., 2018). As expected, we found that overall, the synteny and collinearity pattern is consistent between the two assembly strategies and *A. percula* genome (Supplementary Figures S2 and S3, Supplementary Material online).

Structural annotation of *A. ocellaris de novo* assembly resulted in 24,383 predicted genes. This number is increased in reference-based assemblies, for which the number of predicted genes ranged from 28,170 to 29,913 depending on the species (Table 1 and Supplementary Table S8, Supplementary Material online). The number of annotated genes in two recent assemblies of *A. percula* (Lehmann et al., 2018) and *A. ocellaris* (Tan et al., 2018) genomes were 26,597 and 27,420, respectively. This suggests that several gene predictions are missing in our *de novo* assembly of *A. ocellaris*, but not in our reference-based assemblies. Evidence for this is also provided by BUSCO analyses, which showed that 13% of BUSCO genes were missing in the *A. ocellaris de novo* assembly, while only 5% to 6% of genes were missing in the reference-guided assemblies of clownfishes (Table 1 and Supplementary Table S9, Supplementary Material online). The missing gene predictions in the *de novo A. ocellaris* assembly are due to the increased fragmentation of this assembly compared with the reference-based assemblies (Table 1).

For all assemblies, most of the predicted proteins (92% to 94%) were functionally annotated (Table 1 and Supplementary Table S10, Supplementary Material online), with

proteins in the reference-based assemblies showing an overall good coverage with proteins from the SwissProt database (Supplementary Figure S4, Supplementary Material online, in red). This coverage was reduced for proteins predicted in the *A. ocellaris de novo* assembly (Supplementary Figure S4, Supplementary Material online, in blue), suggesting a lower quality of gene structure prediction for the *de novo* assembly.

The phylogeny reconstructed based on all the publicly available clownfish sequences and sequences extracted from the assembled genomes resulted in the expected topology (Supplementary Figure S5, Supplementary Material online). Most of the assembled individuals branched with individuals of the same species. Three exceptions were observed for *A. ocellaris*, *A. akallopisos*, and *A. melanopus*. However, these inconsistencies are mainly due to a lack of resolution, as suggested by the low support of these nodes.

Taken together, these results indicate that the genome of *A. ocellaris* obtained by reference-guided assembly is at least as good as the one obtained with the *de novo* assembly strategy. Thus, through a reference-based approach, we managed to obtain overall good quality assemblies for all the species while reducing the sequencing and computational costs. *Amphiprion ocellaris de novo* assembly was not considered for further analysis.

3.2. Orthology inference, HOG filtering and classification

Orthology inference performed with OMA on Actinopterygii proteomes (10 clownfish species and 13 outgroup species, Supplementary Figure S1, Supplementary Material online) resulted in a total of 35,976 Hierarchical Orthologous Groups (HOGs). To investigate the level of selective pressure on genes at the origin of clownfishes, HOGs composed by both clownfish species and outgroup Actinopterygii species are necessary (Figure 2). For this reason, we discarded 14,903 HOGs that were formed by either only clownfish sequences (i.e., clownfish-specific HOGs) or by only outgroup sequences (i.e., outgroup-specific HOGs). These discarded HOGs were mainly composed of inaccurately predicted proteins, as suggested by them being composed by only few species with overall shorter sequences compared with the remaining HOGs (Supplementary Figure S6, Supplementary Material online). In addition, 5,133 HOGs were discarded because they were formed by fewer than six species or because they did not contain any sequence from *P. moluccensis*, which is necessary to specifically target our estimation of positive selection on the ancestral branch of clownfishes. This filtering resulted in a total of 15,940 HOGs being retained for positive selection analysis.

Out of the 15,940 HOGs, 13,215 were single-copy when considering the four taxonomic levels (Supplementary Figure S1, Supplementary Material online). As HOGs may be formed by several 1-to-1 OG (i.e., single-copy OG at a given taxonomic level) when considering the different taxonomic level, these 13,215 HOGs corresponded to a total of 13,500 1-to-1 OG. Only 19 HOGs were found specifically duplicated in clownfishes when considering the four taxonomic levels (i.e., clownfish-specific duplicated genes), while the remaining 2,706 HOGs were classified as overall multicopy genes. Most of the genes in the 23 Actinopterygii genomes were part of these 15,940 HOGs tested for signature of positive selection at the basis of the clownfishes (Supplementary Table S11, Supplementary Material online).

3.3. Positive Selection on Single-Copy Genes

We tested for positive selection at the basis of the clownfishes clade on the 13,500 1-to-1 OG. After correction for multiple testing, we found a total of 13 genes that evolved under positive selection in the branch leading to clownfishes (Table 2). The functions of the positively selected genes are diverse, and they are reported in Table 3. Examples of positively selected genes include genes involved in cell adhesion, such as protocadherin-15 (HOG4335_1a), vezatin (HOG16495), and Cadherin-related family member 2 (HOG4262). Other examples include the Versican Core Protein (HOG1437), which is involved in hyaluronic acid binding, and the Protein O-GlcNAcase (HOG16500), which plays a role in the N-acetylglucosamine metabolic process.

The use of either gene trees or the species tree for the positive selection analysis on a subset of the data produced similar results. Before multiple testing correction, 86 genes were found consistently positively selected (i.e., significant in both species tree and gene trees analyses). Twelve additional genes were found positively selected when using the gene trees, and 16 when using the species tree (Supplementary Table S12, Supplementary Material online). However, these differences are no longer present after multiple-testing correction, which resulted in seven genes consistently being detected as positively selected with both species and gene trees (Supplementary Table S12, Supplementary Material online). Thus, the use of either gene or species trees does not affect the results of the analysis after correcting for multiple testing.

Table 2. Results for the Positive Selection Analysis on 1-to-1 OG. The 13 positively selected genes are reported here, with information on the log-likelihood of the null model (no positive selection) and alternative model (positive selection on the branch leading to clownfishes). Likelihood-Ratio test (LRT) p-values, multiple-testing corrected q-values, the proportion of sites under positive selection in the tested branch (ω classes 2a and 2b) and the corresponding ω values are reported for each gene.

HOG name	logL (Null Model)	logL (Alternative Model)	LRT p-values	q-values	Positively selected sites (%)	ω
HOG11195	-41547.05	-41526.90	2.19E-010	2.28E-007	0.8	233.5
HOG16495	-13655.66	-13642.13	1.96E-007	1.53E-004	0.5	999.0
HOG1437	-16835.23	-16825.39	9.19E-006	4.79E-003	0.3	248.0
HOG9295	-4960.03	-4950.14	8.71E-006	4.79E-003	1.0	102.8
HOG5827_3b	-2138.88	-2129.50	1.48E-005	6.61E-003	1.1	999.0
HOG11468	-14064.81	-14055.92	2.47E-005	7.85E-003	0.4	760.6
HOG4335_1a	-23361.23	-23352.35	2.51E-005	7.85E-003	0.5	340.4
HOG11290	-10498.06	-10489.15	2.42E-005	7.85E-003	0.4	999.0
HOG14257	-23503.25	-23495.89	1.24E-004	3.53E-002	0.1	999.0
HOG16500	-11287.91	-11280.90	1.79E-004	3.87E-002	0.2	999.0
HOG21171	-69291.90	-69284.86	1.75E-004	3.87E-002	1.3	25.3
HOG4262	-31942.98	-31935.94	1.76E-004	3.87E-002	2.0	27.8
HOG16343	-6212.65	-6205.67	1.86E-004	3.87E-002	0.5	359.5

Table 3. Annotation of the Positively Selected 1-to-1 OG

HOG name	SwissProt ID	SwissProt Name
HOG11195	P0C5E4	Phosphatidylinositol phosphatase PTPRQ
HOG16495	Q5RFL7	Vezatin
HOG1437	Q90953	Versican core protein
HOG9295	Q3UHZ5	Leiomodrin-2
HOG5827_3b	Q803L0	Protein lin-28 homolog A
HOG11468	Q9D805	Calpain-9
HOG4335_1a	Q0ZM14	Protocadherin-15
HOG11290	Q92581	Sodium/hydrogen exchanger 6
HOG14257	Q8WXG6	MAP kinase-activating death domain protein
HOG16500	Q9EQQ9	Protein O-GlcNAcase
HOG21171	Q9TU53	Cubilin
HOG4262	Q9BYE9	Cadherin-related family member 2
HOG16343	P37892	Carboxypeptidase E

The simulations showed that the positive selection analysis performed on data simulated under neutral or purifying selection scenarios resulted in no false positive detected, and this independently of the simulated sequence length (Supplementary Figure S7, Supplementary Material online). The power to detect positive selection is increased when the strength of selection is larger (i.e., increasing ω ; Supplementary Figure S8, Supplementary Material

online) until it reaches a maximum of 75% for large ω ($\omega > 200$, Supplementary Figure S8, Supplementary Material online). This pattern is observed also for shorter simulated sequences, although the maximum power for large ω is reduced.

Transcriptomic analysis (see Supplementary Material & Methods online) provided evidence of expression of at least seven positively selected 1-to-1 OG in *A. ocellaris* epidermis (TPM>2, Supplementary Table S14, Supplementary Material online), which is the layer of interaction with sea anemones tentacles. Taken together, all these results provide a set of candidate genes that may be linked with the acquisition of the particular life-history traits of clownfishes, such as the mutualism with sea anemones.

Table 4. Results for the positive selection on clownfish-specific duplicated genes. We report the nodes with inferred positive selection, the Likelihood Ratio Test (LRT) statistic for selection, the corrected p-value and the value of the inferred ω classes, with the proportion of sites in each class. The reported nodes correspond to nodes from the inferred gene trees (Figure S9).

	Node	LRT	Corrected p-value	ω1	ω2
HOG4655	Node172	26.5139	0.0001	0.0681 (97%)	46.2 (2.9%)
	Node119	24.3395	0.0004	1.00 (98%)	10000 (2.2%)
	Node70	19.7309	0.0041	0.00 (100%)	10000 (0.21%)
HOG5344	Node89	23.3766	0.0006	0.00 (85%)	15.9 (15%)
	Node204	20.9201	0.002	0.0401 (87%)	11.1 (13%)
	Node142	17.5192	0.0109	0.00 (95%)	10000 (5.1%)
	AMPSE31855	15.4114	0.0314	0.00 (92%)	92.5 (7.6%)
	Node120	14.6422	0.046	0.00 (98%)	10000 (2.4%)
HOG5488	ENSDARG00000098394	20.6963	0.001	0.184 (81%)	111 (19%)
	Node63	20.8642	0.001	0.00 (67%)	9410 (33%)
	Node32	15.7701	0.0121	0.00 (92%)	10000 (8.4%)
HOG19886	Node7	21.4722	0.001	0.484 (93%)	47.7 (7.1%)
	Node26	17.837	0.0061	0.439 (95%)	21.3 (4.9%)

Table 5. Annotation of the Positively Selected Clownfish-Specific Duplicated Genes.

HOG name	SwissProt ID	SwissProt Name
HOG4655	P33267	Cytochrome P450 2F2
HOG5344	P30568	Glutathione S-transferase A
HOG5488	P04437	T-cell receptor alpha chain V
HOG19886	P30122	Bile salt-activated lipase

3.4. Positive Selection on Duplicated Genes

For the overall multicopy genes (i.e., genes with duplications not specific to clownfishes), no evidence of positive selection on gene copies specific to clownfish was found. Out of the 19 clownfish-specific duplicated HOGs, we found four genes with a signature of positive selection in at least one gene copy specific to clownfishes (Table 4 and Supplementary Figure S9, Supplementary Material online). All these positively selected clownfish-specific duplicated genes were annotated with SwissProt IDs (Table 5). One of these positively selected genes is the T-cell receptor alpha (HOG5488), which plays a role in immunity responses. Two other genes, the Glutathione S-transferase (HOG5344) and Cytochrome P450 (HOG4655), are involved in the detoxification of various endogenous and exogenous substances. Transcriptomic analysis (see Supplementary Material & Methods online) showed evidence of expression of Glutathione S-transferase (HOG5344) in *A. ocellaris* epidermis (TPM>2, Supplementary Table S14, Supplementary Material online), supporting a potential role of this gene in the interaction with sea anemones.

4. DISCUSSION

The knowledge on the genomic mechanisms underlying adaptive radiations is still scarce, and this is particularly true when the radiations occurred in a marine ecosystem. In this study, we acquired genomic data for nine clownfish species and one closely related outgroup, in addition to the previously available genome of *A. frenatus* (Marcionetti et al., 2018). These are valuable resources that may be further exploited for advancing our understanding of the genomic patterns observed in adaptive radiations.

In this study, these genomic data sets were exploited to obtain the first insights on the genetic mechanisms underlying the clownfish protection from sea anemone toxins, which resulted in the mutualism that acted as the probable key innovation that triggered clownfish adaptive radiation. Out of the almost 16,000 genes tested, we only found a total of 17 genes showing a signal of positive selection at the origin of clownfishes. Even if a causal link cannot be confirmed without further experimental validation, some of these positively selected genes show functions that are likely to be associated with the protection from sea anemone toxins.

4.1. Genomic Resources for Clownfishes and *P. moluccensis*

To reduce sequencing and computational effort, genomes assemblies for the clownfish species and *P. moluccensis* were obtained using a reference-based approach. Similar approaches were successfully used in the literature (Buza et al., 2015; Lischer & Shimizu, 2017), with for instance the genomes of *Arabidopsis thaliana* (Schneeberger et al., 2011) or *Tetraodon tetrix* (Wang et al., 2014) being obtained by using a reference to guide their assemblies. These methods may nevertheless raise concerns about the validity of the final genomic sequences obtained, especially in the case of nonconserved synteny and collinearity between the reference and the newly assembled species.

Teleost genomes have been found to be evolutionary stable, with genetic content of chromosomes being conserved over nearly 200 Myr of evolution (Schartl et al., 2013). Almost complete synteny and large blocks of collinearity were also observed between the sea bass (*Dicentrarchus labrax*) and three teleost genomes: *Oreochromis niloticus*, *Gasterosteus aculeatus*, and *Tetraodon nigroviridis* (Tine et al., 2014). The divergence time between *D. labrax* and these three species is > 100 Ma (126.8 Ma for *O. niloticus*, 104.8 for *T. nigroviridis* and *G. aculeatus*; Sanciangco et al., 2016). Non-conserved synteny and non-collinearity were therefore not expected to be a concern here, especially considering that clownfishes started to diversify between 12.1 (Santini et al., 2009) and 18.9 Ma (Litsios et al., 2012).

The observed synteny and collinearity between the two *A. ocellaris* assemblies (i.e., *de novo* and reference-guided) and the available genome of *A. percula* (Lehmann et al., 2018, Supplementary Figures S2 and S3, Supplementary Material online) confirmed this expectation. This clearly indicates that the use of *A. frenatus* as reference did not introduce a striking bias in the reconstructed genomic sequences of clownfishes. Evidence for this is also given by the good mapping statistics of paired-end reads (and mate-reads for *A. ocellaris*) for all reference-based assemblies (Supplementary Table S7, Supplementary Material online), which imply that most reads mapped with the expected insertion size and orientation on the assembled genomes. Therefore, the use of *A. frenatus* assembly as reference resulted in all the assemblies having an overall quality that is comparable to the reference used (Marcionetti et al., 2018) but achieved with only half of the original coverage and only one library type.

Although we verified the validity of the obtained genomes, we should keep in mind that the reference-guided assemblies may still miss characteristics that are specific to newly

assembled species but not found in the used reference. For instance, species-specific gene duplications or losses may be omitted when looking exclusively at the resulting assembled genomes. However, these features may be identified by taking advantage of the gene coverage, in a similar way to what is done for copy-number variation detection (e.g., Yoon et al., 2009; Trost et al., 2018). Here, the distribution of the gene coverage was overall normally distributed, with the mean centered on the expected average coverage (Supplementary Figure S10, Supplementary Material online), suggesting the absence of high levels of species-specific duplication or losses. Species-specific features are in any case out of the scope of this study, as we investigated here what is common to all clownfish species.

4.2. Candidate Genes Involved in Clownfish Protection from Sea Anemones Toxins

Evolutionary mechanisms that may result in the appearance of new advantageous traits (such as the protection from toxins) include positive selection on protein-coding genes, where mutations altering the function of genes are fixed in the population because they are favorable. Examples of this process have already been reported (e.g. Spady et al., 2005; Hoekstra et al., 2006; Protas et al., 2006; Lynch, 2007). By contrast, purifying selection is the mechanism preventing the fixation of deleterious mutations, as those mutations are detrimental for the organism. Therefore, the appearance of an advantageous trait by positive selection in an ancestral species may be followed by a switch in the selective pressure, with this trait undergoing purifying selection in the descendant species (i.e., if this trait is still advantageous for them). Examples of this scenario with pattern of positive selection in internal branches of a phylogeny, followed by a switch to purifying selection are found in primates (Perry et al., 2012; Daub et al., 2017), grasses (Schwerdt et al., 2015), seagrasses (Wissler et al., 2011), and rust fungi (Silva et al., 2015).

This scenario of a switch in selective pressure in the internal branches was tested in this study as it fits well with the appearance of clownfish-specific life history traits, such as their mutualism with sea anemones. For this, the presence of the outgroup *P. moluccensis* was necessary, as it allowed us to specifically aim for the ancestral branch of clownfishes. Thus, after the acquisition of the advantageous traits such as the ability to live unharmed in sea anemones on this specific branch, these traits must have been conserved (i.e., underwent purifying selection) across the whole clownfish group.

A total of 17 genes (either single copy or duplicated genes) were found to have evolved under positive selection at the origin of clownfishes and showed a later switch to purifying

selection in the other branches of the clade. Simulations showed that the level of false positive results that we can expect in our data sets is very low, which suggests that we can have a high confidence in these results. In addition to the mutualism with sea anemones, these positively selected genes that are specific to the evolution of clownfishes may be associated with other clownfish-specific traits, such as their outstanding estimated lifespan (Buston & García, 2007) or their hierarchical social structure (Buston, 2003). Similarly, although none of the positively selected genes are documented as involved in the evolution of coloration in teleosts (Lorin et al., 2018), we cannot exclude their potential role in the evolution of clownfish' particular coloration.

One of the detected positively selected genes is the HOG1437, which is annotated as coding for the Versican Core Protein. This protein plays a role in intercellular signaling, in connecting cells with the extracellular matrix, and it may also take part in the regulation of cell motility, growth, and differentiation. Additionally, it is binding hyaluronic acid (Bignami et al., 1989; Perides et al., 1989), a glycosaminoglycan distributed widely throughout connective, epithelial, and neural tissues. Glycosaminoglycans are polysaccharides consisting of repeating amino-sugar units, such as *N*-acetylglucosamine (GlcNAc). Another gene found positive selected is the HOG16500, annotated as coding for Protein O-GlcNAcase, whose function is to cleave GlcNAc from O-glycosylated proteins (Toleman et al., 2006).

These observations are interesting since *N*-acetylated sugars (such as GlcNAc) have been shown to trigger the discharge of sea anemones cnidocytes, leading to the release of toxins (Anderson & Bouchard, 2009). Chemoreceptors of *N*-acetylated sugars are located in cells surrounding cnidocytes, which change in morphology in response to stimulation of these receptors by *N*-acetylated sugars. These structural modifications alter the mechanical properties of the hair bundles and tune them to the frequencies of vibrations emitted by swimming prey, resulting in an increase in the baseline discharge of cnidocytes when the anemone touches the prey (Thorington & Hessinger, 1988a; Mire-Thibodeaux & Watson, 1994). One *N*-acetylated sugar shown to trigger cnidocytes discharges is the *N*-acetylneuraminic acid (NANA; Ozacmak et al., 2001). This compound was found to be significantly lacking in *A. ocellaris* mucus (Abdullah and Saad, 2015). GlcNAc is another *N*-acetylated sugar that may be recognized by *N*-acetylated chemoreceptors, and thus trigger the discharge of sea anemone toxins. This is supported by studies showing that hyaluronic acid, which is composed by GlcNAc, was the only polysaccharide able to strongly excite cnidocytes and trigger their discharge (Lubbock, 1979; Thorington & Hessinger, 1988b).

The two positively selected genes HOG1437 (Versican Core Protein) and HOG16500 (Protein O-GlcNAcase) display therefore interesting functions associated with *N*-acetylated sugars. The Versican Core Protein is observed to be expressed in *A. ocellaris* epidermis (Supplementary Table S14, Supplementary Material online), that is, the layer of interaction with sea anemones tentacles. A low signal of Protein O-GlcNAcase expression was also detected in *A. ocellaris* epidermis (Supplementary Table S14, Supplementary Material online). With this evidence, we hypothesize that these genes might play a role in the masking (GlcNAc binding by Versican Core Protein) or removal (cleavage by Protein O-GlcNAcase) of *N*-acetylated sugars. This would therefore help decrease or prevent the stimulation of the chemoreceptors for *N*-acetylated sugars, thus preventing or decreasing cnidocytes discharge and the release of toxins. Clownfishes might thus not necessarily be fully resistant to toxins released by cnidocytes, but they could have evolved a system that prevents these toxins to be discharged (as previously suggested in Lubbock, 1980, 1981).

Sea anemones toxicity is not only due to the discharge of cnidocytes but also by the presence of secreted toxins in sea anemone mucus such as cytolytic toxins. A resistance against sea anemone cytolytic toxins was effectively observed in some clownfish species (Mebs, 1994), suggesting that this resistance may be mediated through specific mechanisms such as immune response (Mebs, 2009). Clownfish-specific duplicated genes involved in immunity response as the T-cell receptor alpha (HOG5488), or involved in detoxification such as Cytochrome P450 (HOG4655; Manikandan & Nagini, 2018) and Glutathione S-transferases (HOG5344; Sheehan et al., 2001) are found positively selected at the origin of clownfishes. These genes are part of gene families having a large number of different roles (Sheehan et al., 2001; Manikandan & Nagini, 2018), thus making it difficult to define their precise function. In addition, genes involved in immune responses are often seen as subject to positive selection (Schlenke & Begun, 2003; Jiggins & Kim, 2007), and have been seen to evolve faster than non-immune genes (McTaggart et al., 2012). For these reasons, direct links between these positively selected genes and a potential role in the protection from sea anemones secreted toxins cannot be drawn without further experimental evidence.

Furthermore, as only some clownfish species showed resistance to cytolytic toxins (Mebs, 1994), this resistance could have appeared later in the evolution of clownfishes and be specific to only some species.

In addition to positive selection on protein-coding genes (i.e., coding changes), the acquisition of new phenotypes may also occur through regulatory changes that alter gene expression profiles (e.g., Wittkopp et al., 2003; Shapiro et al., 2004). However, the identification and analysis of noncoding elements such as transcription factor binding sites in nonmodel organisms remain challenging. Therefore, although not analyzed here, we may expect that regulatory sequence evolution has acted in concert with the coding changes (i.e., positive selection on coding genes) identified in this study in the built up of clownfish mutualism with sea anemones.

5. CONCLUSIONS

In this study, we acquired genomic data for nine clownfish species and one closely related outgroup. These data are a valuable resource that may be further exploited for advancing our understanding of the genomic patterns observed in adaptive radiations.

Using these newly assembled genomes, we investigated here the mechanisms underlying clownfish protection from sea anemone toxins, which resulted in the acquisition of the mutualism that likely acted as the key innovation triggering clownfish adaptive radiation. We identified 17 genes with a signal of positive selection at the origin of clownfishes. Some of these genes showed interesting functions associated with *N*-acetylated sugars, which are known to be involved in sea anemones' discharge of toxins. Although further experimental validations are necessary to find a causal link between these genes and the ability to interact with sea anemones, this study provides the first genomic approach to try to disentangle the mechanisms behind the mutualism between sea anemones and clownfishes.

SUPPLEMENTARY MATERIAL

Supplementary data are available at *Genome Biology and Evolution* [online](#). To facilitate the access, [Supplementary Information and Figures](#)¹ and [Supplementary Tables](#)² are also available directly on OneDrive.

¹https://unils-my.sharepoint.com/:b:/g/personal/anna_marcionetti_unil_ch/EUg0Bx-fhf9DjcIp7AcT5NgBgtOHs7e-oKu6Stp96J6SLA?e=1I9vrB

²https://unils-my.sharepoint.com/:x:/g/personal/anna_marcionetti_unil_ch/Ee0wQS1uCPREthTwMI3E3GUB8CYc1zFGR1WMRcU92eIKig?e=RjyVb1

ACKNOWLEDGMENTS

We would like to thank the Vital-IT infrastructure from the Swiss Institute of Bioinformatics for the computational resources and the Lausanne Genomic Technology Facility for the sequencing. Funding: University of Lausanne funds, Swiss National Science Foundation, Grant Number: 31003A-163428.

AUTHOR CONTRIBUTIONS

Nicolas Salamin and Anna Marcionetti designed the research. Anna Marcionetti obtained the genome data, performed the genome assembly, validation and annotation. Anna Marcionetti performed the transcriptomic and comparative genomic analyses. Victor Rossier participated in the genome annotation and positive selection analyses. Natacha Roux and Vincent Laudet performed the experimental part of the transcriptomic analyses. All authors contributed to the writing of the paper.

DATA ACCESSIBILITY

Raw Illumina reads are available in the Sequence Read Archive in the NCBI database (BioProject ID: [PRJNA515163](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA515163)). The assembled genomes and their annotation are available in Zenodo repository ([DOI: 10.5281/zenodo.2540241](https://doi.org/10.5281/zenodo.2540241)).

REFERENCES

- Abdullah, N. S., Saad, S.** (2015). Rapid Detection of N-Acetylneuraminic Acid from False Clownfish using HPLC-FLD for Symbiosis to Host Sea Anemone. *Asian Journal of Applied Sciences*, 03(5), 858-864
- Altenhoff, A. M., Gil, M., Gonnet, G. H., & Dessimoz, C.** (2013). Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS One*, 8(1), e53786.
- Anderson, P. A., & Bouchard, C.** (2009). The regulation of cnidocyte discharge. *Toxicon*, 54(8), 1046-1053.
- Anisimova, M., Bielawski, J. P., & Yang, Z.** (2002). Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular biology and evolution*, 19(6), 950-958.
- Balamurugan, J., Kumar, T. A., Kannan, R., & Pradeep, H. D.** (2014). Acclimation behaviour and biochemical changes during anemonefish (*Amphiprion sebae*) and sea anemone (*Stichodactyla haddoni*) symbiosis. *Symbiosis*, 64(3), 127-138.
- Bignami, A., Lane, W. S., Andrews, D., & Dahl, D.** (1989). Structural similarity of hyaluronate binding proteins in brain and cartilage. *Brain research bulletin*, 22(1), 67-70.

- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., ... & Turner-Maier, J.** (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, *513*(7518), 375.
- Buston, P. M.** (2003). Social hierarchies: size and growth modification in clownfish. *Nature*, *424*(6945), 145.
- Buston, P. M., & García, M. B.** (2007). An extraordinary life span estimate for the clown anemonefish *Amphiprion percula*. *Journal of Fish Biology*, *70*(6), 1710-1719.
- Buza, K., Wilczynski, B., & Dojer, N.** (2015). RECORD: reference-assisted genome assembly for closely related genomes. *International journal of genomics*, 2015.
- Dabney, A., Storey, J. D., & Warnes, G. R.** (2010). qvalue: Q-value estimation for false discovery rate control. *R package version*, 1(0).
- Darwin, C.** (1859). *The Origin of Species by Means of Natural Election, Or the Preservation of Favored Races in the Struggle for Life*. Murray, London
- Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., ... & Salazar, C.** (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94.
- Daub, J. T., Moretti, S., Davydov, I. I., Excoffier, L., & Robinson-Rechavi, M.** (2017). Detection of pathways affected by positive selection in primate lineages ancestral to humans. *Molecular biology and evolution*, *34*(6), 1391-1402.
- Diekmann, Y., & Pereira-Leal, J. B.** (2015). Gene tree affects inference of sites under selection by the branch-site test of positive selection. *Evolutionary Bioinformatics*, *11*, EBO-S30902.
- Erwin, D. H.** (2007). Increasing returns, ecological feedback and the Early Triassic recovery. *Palaeoworld*, *16*(1-3), 9-15.
- Fautin, D. G., & Allen, G. R.** (1997). *Anemonefishes and their host sea anemones*. Western Australian Museum: Perth.
- Fletcher, W., & Yang, Z.** (2010). The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Molecular biology and evolution*, *27*(10), 2257-2267.
- Gainsford, A., Van Herwerden, L., & Jones, G. P.** (2015). Hierarchical behaviour, habitat use and species size differences shape evolutionary outcomes of hybridization in a coral reef fish. *Journal of evolutionary biology*, *28*(1), 205-222.
- Gavrilets, S., & Losos, J. B.** (2009). Adaptive radiation: contrasting theory with data. *Science*, *323*(5915), 732-737.
- Gavrilets, S., & Vose, A.** (2005). Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences*, *102*(50), 18040-18045.
- Givnish, T. J.** (2015). Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. *New Phytologist*, *207*(2), 297-303.
- Givnish, T. J., & Sytsma, K. J.** (1997). *Molecular evolution and adaptive radiation*. (Cambridge University Press, New York).
- Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., ... & Berlin, A. M.** (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences*, *108*(4), 1513-1518.

- Grant, P. R., & Grant, B. R.** (2008). *How and why species multiply: the radiation of Darwin's finches.* (Princeton University Press, Princeton, NJ).
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O.** (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307-321.
- Haug-Baltzell, A., Stephens, S. A., Davey, S., Scheidegger, C. E., & Lyons, E.** (2017). SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics*, 33(14), 2197-2198.
- Hoekstra, H. E., Hirschmann, R. J., Bunday, R. A., Insel, P. A., & Crossland, J. P.** (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, 313(5783), 101-104.
- Huson, D. H., Richter, D. C., Rausch, C., DeZulian, T., Franz, M., & Rupp, R.** (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC bioinformatics*, 8(1), 460.
- Jiggins, F. M., & Kim, K. W.** (2007). A screen for immunity genes evolving under positive selection in *Drosophila*. *Journal of evolutionary biology*, 20(3), 965-970.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... & Birney, E.** (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55.
- Katoh, K., & Standley, D. M.** (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Thierer, T.** (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., ... & Grant, B. R.** (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371.
- Lehmann, R., Lightfoot, D. J., Schunter, C., Michell, C. T., Ohyanagi, H., Mineta, K., ... & Gojobori, T.** (2018). Finding Nemo's Genes: A chromosome-scale reference assembly of the genome of the orange clownfish *Amphiprion percula*. *Molecular ecology resources*.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R.** (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Lischer, H. E., & Shimizu, K. K.** (2017). Reference-guided de novo assembly approach improves genome reconstruction for related species. *BMC bioinformatics*, 18(1), 474.
- Litsios, G., Sims, C. A., Wüest, R. O., Pearman, P. B., Zimmermann, N. E., & Salamin, N.** (2012). Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, 12(1), 212.
- Lorin, T., Brunet, F. G., Laudet, V., & Voff, J. N.** (2018). Teleost Fish-Specific Preferential Retention of Pigmentation Gene-Containing Families After Whole Genome Duplications in Vertebrates. *G3: Genes, Genomes, Genetics*, g3-200201.
- Losos, J. B.** (2009). *Lizards in an evolutionary tree: ecology and adaptive radiation of Anoles.* (University of California Press, Berkeley, CA).
- Lubbock, R.** (1979). Chemical recognition and nematocyte excitation in a sea anemone. *Journal of Experimental Biology*, 83(1), 283-292.

- Lubbock, R.** (1980). Why are clownfishes not stung by sea anemones?. *Proc. R. Soc. Lond. B*, 207(1166), 35-61.
- Lubbock, R.** (1981). The clownfish/anemone symbiosis: a problem of cellular recognition. *Parasitology*, 82(1), 159-173.
- Lunter, G., & Goodson, M.** (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome research*, 21(6), 936-939.
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., ... & Tang, J.** (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, 1(1), 18.
- Lynch, V. J.** (2007). Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A 2 genes. *BMC evolutionary biology*, 7(1), 2.
- Mabuchi, K., Miya, M., Azuma, Y., & Nishida, M.** (2007). Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. *BMC Evolutionary Biology*, 7(1), 10.
- MacFadden, B. J.** (2005). Fossil horses--evidence for evolution. *Science*, 307(5716), 1728-1730.
- Manikandan, P., & Nagini, S.** (2018). Cytochrome P450 Structure, Function and Clinical Significance: A Review. *Current drug targets*, 19(1), 38-54.
- Marcionetti, A., Rossier, V., Bertrand, J. A., Litsios, G., & Salamin, N.** (2018). First draft genome of an iconic clownfish species (*Amphiprion frenatus*). *Molecular ecology resources*.
- McTaggart, S. J., Obbard, D. J., Conlon, C., & Little, T. J.** (2012). Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex*. *BMC evolutionary biology*, 12(1), 63.
- Mebs, D.** (1994). Anemonefish symbiosis: vulnerability and resistance of fish to the toxin of the sea anemone. *Toxicon*, 32(9), 1059-1068.
- Mebs, D.** (2009). Chemical biology of the mutualistic relationships of sea anemones with fish and crustaceans. *Toxicon*, 54(8), 1071-1074.
- Mire-Thibodeaux, P., & Watson, G. M.** (1994). Morphodynamic hair bundles arising from sensory cell/supporting cell complexes frequency-tune nematocyst discharge in sea anemones. *Journal of Experimental Zoology*, 268(4), 282-292.
- Miyagawa, K.** (1989). Experimental analysis of the symbiosis between anemonefish and sea anemones. *Ethology*, 80(1-4), 19-46.
- Miyagawa, K., & Hidaka, T.** (1980). *Amphiprion clarkii* juvenile: innate protection against and chemical attraction by symbiotic sea anemones. *Proceedings of the Japan Academy, Series B*, 56(6), 356-361.
- Moretti, S., Laurency, B., Gharib, W. H., Castella, B., Kuzniar, A., Schabauer, H., ... & Robinson-Rechavi, M.** (2014). Selectome update: quality control and computational improvements to a database of positive selection. *Nucleic acids research*, 42(D1), D917-D921.
- Near, T. J., Dornburg, A., Kuhn, K. L., Eastman, J. T., Pennington, J. N., Patarnello, T., ... & Jones, C. D.** (2012). Ancient climate change, antifreeze, and the evolutionary diversification of Antarctic fishes. *Proceedings of the National Academy of Sciences*, 109(9), 3434-3439.
- Nedosyko, A. M., Young, J. E., Edwards, J. W., & da Silva, K. B.** (2014). Searching for a toxic key to unlock the mystery of anemonefish and anemone symbiosis. *PloS one*, 9(5), e98449.
- Ollerton, J., McCollin, D., Fautin, D. G., & Allen, G. R.** (2007). Finding NEMO: nestedness engendered by mutualistic organization in anemonefish and their hosts. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1609), 591-598.

- Ozacmak, V. H., Thorington, G. U., Fletcher, W. H., & Hessinger, D. A.** (2001). N-acetylneuraminic acid (NANA) stimulates in situ cyclic AMP production in tentacles of sea anemone (*Aiptasia pallida*): possible role in chemosensitization of nematocyst discharge. *Journal of Experimental Biology*, 204(11), 2011-2020.
- Perides, G., Lane, W. S., Andrews, D., Dahl, D., & Bignami, A.** (1989). Isolation and partial characterization of a glial hyaluronate-binding protein. *Journal of Biological Chemistry*, 264(10), 5981-5987.
- Perry, G. H., Melsted, P., Marioni, J. C., Wang, Y., Bainer, R., Pickrell, J. K., ... & Pritchard, J. K.** (2012). Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome research*, 22(4), 602-610.
- Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., ... & Tabin, C. J.** (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature genetics*, 38(1), 107.
- Puebla, O.** (2009). Ecological speciation in marine v. freshwater fishes. *Journal of Fish Biology*, 75(5), 960-996.
- Quinlan, A. R., & Hall, I. M.** (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842.
- Rambaut, A.** (2014). FigTree version 1.4.2. Computer program distributed by the author. <http://tree.bio.ed.ac.uk/software/figtree/>
- Rundell, R. J., & Price, T. D.** (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, 24(7), 394-399.
- Sanciango, M. D., Carpenter, K. E., & Betancur-R, R.** (2016). Phylogenetic placement of enigmatic percomorph families (Teleostei: Percomorphaceae). *Molecular phylogenetics and evolution*, 94, 565-576.
- Santini, F., Harmon, L. J., Carnevale, G., & Alfaro, M. E.** (2009). Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes. *BMC evolutionary biology*, 9(1), 194.
- Schartl, M., Walter, R. B., Shen, Y., Garcia, T., Catchen, J., Amores, A., ... & Bisazza, A.** (2013). The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nature genetics*, 45(5), 567.
- Schlenke, T. A., & Begun, D. J.** (2003). Natural selection drives *Drosophila* immune system evolution. *Genetics*, 164(4), 1471-1480.
- Schlichter, D.** (1976). Macromolecular mimicry: substances released by sea anemones and their role in the protection of anemone fishes. In *Coelenterate ecology and behavior* (pp. 433-441). Springer, Boston, MA.
- Schluter, D.** (2000). *The ecology of adaptive radiation*. (Oxford University Press, Oxford).
- Schneeberger, K., Ossowski, S., Ott, F., Klein, J. D., Wang, X., Lanz, C., ... & Henz, S. R.** (2011). Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proceedings of the National Academy of Sciences*, 108(25), 10249-10254.
- Schwerdt J. G., MacKenzie K., Oehme D. et al** (2015) Evolutionary dynamics of the cellulose synthase gene superfamily in grasses. *Plant Physiol* 168:968–983
- Seehausen, O.** (2006). African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1597), 1987-1998.

- Shapiro, M. D., Marks, M. E., Peichel, C. L., Blackman, B. K., Nereng, K. S., Jónsson, B., ... & Kingsley, D. M.** (2004). Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*, *428*(6984), 717.
- Sheehan, D., Meade, G., & Foley, V. M.** (2001). Structure, function and evolution of glutathione transferases: implications for classification of non-mammalian members of an ancient enzyme superfamily. *Biochemical Journal*, *360*(1), 1-16.
- Silva, D. N., Duplessis, S., Talhinhos, P., Azinheira, H., Paulo, O. S., & Batista, D.** (2015). Genomic patterns of positive selection at the origin of rust fungi. *PLoS one*, *10*(12), e0143959.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M.** (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, *31*(19), 3210-3212.
- Simmons, N. B., Seymour, K. L., Habersetzer, J., & Gunnell, G. F.** (2008). Primitive Early Eocene bat from Wyoming and the evolution of flight and echolocation. *Nature*, *451*(7180), 818.
- Simpson G. G.** (1953). *The major features of evolution*. (Columbia University Press, New York).
- Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L.** (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular biology and evolution*, *32*(5), 1342-1353.
- Soulebeau, A., Aubriot, X., Gaudeul, M., Rouhan, G., Hennequin, S., Haevermans, T., ... & Jabbour, F.** (2015). The hypothesis of adaptive radiation in evolutionary biology: hard facts about a hazy concept. *Organisms Diversity & Evolution*, *15*(4), 747-761.
- Spady, T. C., Seehausen, O., Loew, E. R., Jordan, R. C., Kocher, T. D., & Carleton, K. L.** (2005). Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Molecular Biology and Evolution*, *22*(6), 1412-1422.
- Stroud, J. T., & Losos, J. B.** (2016). Ecological opportunity and adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, *47*, 507-532.
- Sukumaran, J. & Holder, M. T.** (2010). DendroPy: A Python library for phylogenetic computing. *Bioinformatics* *26*: 1569-1571.
- Supple, M. A., Hines, H. M., Dasmahapatra, K. K., Lewis, J. J., Nielsen, D. M., Lavoie, C., ... & Counterman, B. A.** (2013). Genomic architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies. *Genome research*, gr-150615.
- Suyama, M., Torrents, D., & Bork, P.** (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, *34*(suppl_2), W609-W612.
- Tan, M. H., Austin, C. M., Hammer, M. P., Lee, Y. P., Croft, L. J., & Gan, H. M.** (2018). Finding Nemo: hybrid assembly with Oxford Nanopore and Illumina reads greatly improves the clownfish (*Amphiprion ocellaris*) genome assembly. *GigaScience*, *7*(3), gix137.
- Team, R. C.** (2013). R: A language and environment for statistical computing
- Thorington, G. U., & Hessinger, D. A.** (1988a). Control of discharge: factors affecting discharge of cnidae. In *The biology of nematocysts* (pp. 233-253).
- Thorington, G. U., & Hessinger, D. A.** (1988b). Control of cnida discharge: I. Evidence for two classes of chemoreceptor. *The Biological Bulletin*, *174*(2), 163-171.

- Tine, M., Kuhl, H., Gagnaire, P. A., Louro, B., Desmarais, E., Martins, R. S., ... & Dieterich, R.** (2014). European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature communications*, *5*, 5770.
- Toleman, C., Paterson, A. J., Shin, R., & Kudlow, J. E.** (2006). Streptozotocin inhibits O-GlcNAcase via the production of a transition state analog. *Biochemical and biophysical research communications*, *340*(2), 526-534.
- Trost, B., Walker, S., Wang, Z., Thiruvahindrapuram, B., MacDonald, J. R., Sung, W. W., ... & Reuter, M. S.** (2018). A comprehensive workflow for read depth-based identification of copy-number variation from whole-genome sequence data. *The American Journal of Human Genetics*, *102*(1), 142-155.
- Villanueva-Canas, J. L., Laurie, S., & Alba, M. M.** (2013). Improving genome-wide scans of positive selection by using protein isoforms of similar length. *Genome biology and evolution*, *5*(2), 457-467.
- Wang, B., Ekblom, R., Bunikis, I., Siitari, H., & Höglund, J.** (2014). Whole genome sequencing of the black grouse (*Tetrao tetrix*): reference guided assembly suggests faster-Z and MHC evolution. *BMC genomics*, *15*(1), 180.
- Wissler, L., Codoñer, F. M., Gu, J., Reusch, T. B., Olsen, J. L., Procaccini, G., & Bornberg-Bauer, E.** (2011). Back to the sea twice: identifying candidate plant genes for molecular evolution to marine life. *BMC evolutionary biology*, *11*(1), 8.
- Wittkopp, P. J., Williams, B. L., Selegue, J. E., & Carroll, S. B.** (2003). *Drosophila* pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proceedings of the National Academy of Sciences*, *100*(4), 1808-1813.
- Yang, Z.** (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, *24*(8), 1586-1591.
- Yoder, J. B., Clancey, E., Des Roches, S., Eastman, J. M., Gentry, L., Godsoe, W., ... & Sarver, B. A. J.** (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of evolutionary biology*, *23*(8), 1581-1596.
- Yoon, S., Xuan, Z., Makarov, V., Ye, K., & Sebat, J.** (2009). Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome research*.

Chapter 3

Insights into the genomics of clownfish adaptive radiation: the genomic substrate of the diversification

Insights into the genomics of clownfish adaptive radiation: the genomic substrate of the diversification

ABSTRACT

Clownfishes are an iconic group of coral reef fish that evolved a mutualistic interaction with sea anemones, which triggered the rapid diversification of the group. Here, we investigated the genomic architecture underlying this process. First, we determined which genomic characteristics were associated with the adaptive radiation of the group. Secondly, we assessed the presence and the mechanisms of parallel evolution in clownfishes.

We took advantage of the available genomic data of five pairs of closely related but ecologically divergent clownfish species and performed comparative genomic analyses. We found that clownfish genomes show two bursts of transposable elements, overall accelerated coding evolution, and topology inconsistencies potentially resulting from hybridization events. These characteristics possibly facilitated the rapid diversification of the group. We also detected a signature of positive selection throughout the radiation in 5.4 % of the analyzed genes. Among them, five presented functions associated with social behavior and ecology, and they could have potentially played a role in the evolution of clownfishes' particular size-based hierarchical social structure. Finally, we found genes with patterns of relaxation and 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, these results provide the first insights into the genomic architecture of clownfish adaptive radiation. This work integrates the growing collection of studies investigating the genomic mechanisms governing species diversification, which brings us a step closer to understanding how biodiversity on Earth is created.

1. INTRODUCTION

Adaptive radiation is defined as the rapid diversification of an ancestral population into several ecologically different species, associated with adaptive morphological or physiological divergence (Schluter, 2000). This process is considered to play a central role in creating the spectacular diversity of life on Earth (Simpson, 1953; Schluter, 2000). For decades, researchers have been investigating the causes and consequences of adaptive radiations (e.g., Givnish & Sytsma, 1997; Schluter, 2000; Seehausen, 2004; Glor, 2012; Yoder et al., 2010; Givnish, 2015; Soulebeau et al., 2015; Stroud & Losos, 2016; Martin & Richards, 2019), with the ultimate goal to broaden our understanding of the mechanisms governing species diversification and, ultimately, the buildup of biodiversity.

A widespread phenomenon in adaptive radiations is convergent evolution, which describes the repeated evolution of similar phenotypes into distinct lineages (e.g., Schluter & Nagel, 1995; Rundle et al., 2000; Blackledge et al., 2004; Muschick et al., 2012; Vizueta et al., 2019) and is generally explained as the results of independent adaptation to similar ecological conditions (Schluter, 2000; Brakefield, 2006; Losos, 2011). While textbook examples of convergent evolution typically include the development of wings in birds and bats or the evolution of echolocation in bats and dolphins (Roberts, 1986; Vater & Kössl, 2004; Liu et al., 2010), in rapidly radiating lineages, convergent traits can also be observed between phylogenetically closer species. For instance, equivalent ecotypes of *Anolis* lizards (Losos, 2009) or *Tetragnatha* spiders (Blackledge & Gillespie, 2004) have arisen in the Greater Antilles and Hawaiian Islands, respectively. Similarly, benthic and limnetic sticklebacks fish have evolved repeatedly in postglacial lakes (Schluter & Nagel, 1995; Rundle et al., 2000), and the occurrence of convergent forms of cichlids was observed both within the East African Lake Tanganyika (Muschick et al., 2012) and across the Tanganyika and the Malawi Lakes (Kocher et al., 1993). Thus, the adaptive radiation process provides an interesting setup to investigate not only the mechanisms behind species diversification, but also the repeatability and predictability of evolution, a central question in evolutionary biology (Rosenblum et al., 2014, Kingman et al., 2021).

With the advances in molecular genetic and sequencing technologies, the development, use, and availability of genomic tools for non-model systems have boomed (Abzhanov et al., 2008). Consequently, several studies have started investigating the intrinsic genomic factors that may promote adaptive radiations. Recent examples have included cichlids (Brawand et al.,

2014; Faber-Hammond et al., 2019; McGee et al., 2020; Xiong et al., 2021), threespine sticklebacks (Jones et al., 2012; Verta & Jones, 2019), *Anolis* lizards (Feiner, 2016), Darwin's finches (Lamichhaney et al., 2015), *Heliconius* butterflies (Dasmahapatra et al., 2012; Supple et al., 2013; Edelman et al., 2019) or *Dysdera* spiders (Vizueta et al., 2019). These studies found that a wide array of genome-wide changes, including chromosomal duplications, expansions of gene families, bursts of transposable elements (TEs), and accelerated evolution on coding and non-coding sequences, could predispose particular lineages to radiate adaptively (Jones et al., 2012; Brawand et al., 2014; Fan & Meyer, 2014; Feiner, 2016; Berner & Salzburger, 2015; Faber-Hammond et al., 2019; Verta & Jones, 2019; Xiong et al., 2021). Besides these overall genomic features potentially linked with the rapid diversification, these studies provided evidence of the key role played by ancient polymorphism and hybridization events (both ancestral or between diverging lineages) in shaping adaptive radiations (e.g., Dasmahapatra et al., 2012; Berner & Salzburger, 2015; Lamichhaney et al., 2015; Meier et al., 2017; Malinsky et al., 2018; Edelman et al., 2019; Svardal et al., 2020; Kozak et al., 2021).

The question of whether convergent phenotypes - widely observed, for instance, in adaptive radiations - are originated by shared genetic and molecular mechanisms (i.e., “parallel evolution” as defined in Rosenblum et al., 2014) has also started to be elucidated (Rosenblum et al., 2014; Elmer & Meyer, 2011; Sackton & Clark, 2019). Parallel evolution between different organisms was detected at diverse hierarchical levels, with, for instance, mutations at identical amino acid residues (Protas et al., 2006; Zhen et al., 2012; Projecto-Garcia et al., 2013), at distinct sites within the same locus (Kingsley et al., 2009; Rosenblum et al., 2010; Linnen et al., 2013) or in different genes within the same pathway (Arendt & Reznick, 2008). These cases arose from mutations that occurred independently in different species (see Table 1 in Stern, 2013). However, parallel evolution can also be achieved by the appearance of shared ancestral (and polymorphic) alleles - like in the convergent loss of lateral plates in the adaptive radiation of sticklebacks (Cresko et al., 2004; Colosimo et al., 2005) - or by introgressive hybridization - as in the convergent acquisition of wing patterns during the diversification of *Heliconius* butterflies (Dasmahapatra et al., 2012; Lewis et al., 2019).

The study of adaptively radiating lineages has broadened our knowledge of the mechanisms behind species diversification. However, much remains to be understood on the overall genomic patterns associated with adaptive radiations, the mechanisms underlying the acquisition of convergent phenotypes, and the extent of parallel evolution. An interesting group to extend our understanding of these processes are clownfishes (or anemonefishes). These

iconic coral reef fishes (genera *Amphiprion* and *Premnas*) consist of 28 recognized species and two natural hybrids (Fautin & Allen, 1997; Ollerton et al., 2007; Gainsford et al., 2015). One distinctive characteristic of this group is the mutualistic interaction they maintain with sea anemones. Within the sea anemones, clownfishes live in a size-based social hierarchy (Fricke, 1979; Ochi, 1989; Buston, 2003) and are sequential hermaphrodites (Fricke & Fricke, 1977; Moyer & Nakazono, 1978; Fricke, 1979). While all species are associated with sea anemones, there is a large variability in host usage within the group. Some species are strictly specialists and can interact with a single species of sea anemones, while others are generalists and can inhabit up to ten hosts (Fautin & Allen, 1997; Ollerton et al., 2007; Gainsford et al., 2015). This mutualism acted as the key innovation that triggered the adaptive radiation of the group (Litsios et al., 2012). 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 primary radiation of the clownfishes happened in the Indo-Australian Archipelago, a second replicated and geographically independent radiation showing the same gradient of host usage occurred in the Western Indian Ocean (Litsios et al., 2014).

In this study, we investigated the genomic architecture of the clownfish adaptive radiation. Our first aim was to explore the genomic characteristics associated with the radiation of the clownfishes. In particular, we examined if bursts of TEs, increased gene duplications, signatures of accelerated evolution, and evidence of hybridization events - previously observed in other radiating lineages - were observed in this group. We expected to detect at least some of these features, as they create genomic variations necessary for natural selection to act, likely facilitating the diversification. We performed comparative genomic analyses of ten publicly available clownfish species covering the entire divergence and distribution range of the group (i.e., *P. biaculeatus*, *A. ocellaris*, *A. perideraion*, *A. akallopisos*, *A. polymnus*, *A. sebae*, *A. melanopus*, *A. bicinctus*, *A. nigripes*, and *A. frenatus*; Marcionetti et al., 2019). These species represent 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 Figure 1 in Marcionetti et al., 2019).

Our second aim was to determine the extent and the mechanisms of parallel evolution in the radiation of clownfishes. By taking advantage of these evolutionary replicates, we

investigated whether the observed convergence at the phenotypic level was mirrored by convergent changes at the genetic level. The rationale behind this second aim was that genes involved in the ecological divergence of clownfishes might evolve at different rates or show different selective pressures associated with host usage. Suppose the same genes are involved in the ecological divergence of multiple species (i.e., parallel evolution). In that case, these genes should display similar evolutionary rates and selective pressures in all the concerned specialist and generalist species, or they might show topological inconsistencies in the case of adaptive introgressive hybridization (as in Dasmahapatra et al., 2012; Supple et al., 2013).

2. MATERIAL AND METHODS

Whole-genome assemblies and annotations for the ten clownfish species (*Premnas biaculeatus*, *Amphiprion ocellaris*, *A. perideraion*, *A. akallopisos*, *A. polymnus*, *A. sebae*, *A. melanopus*, *A. bicinctus*, *A. nigripes*, and *A. frenatus*) and the lemon damselfish (*Pomacentrus 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 anemones hosts) or generalist (more than two sea anemones hosts) species (Supplementary Table S1). Although the differential host and habitat use is more complex than this dichotomy, our classification separates clownfish species on the two principal axes of mutualistic interaction variation and its correlated morphological differentiation (Litsios et al., 2012).

2.1. 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](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA515163)). We randomly sub-sampled 20 million reads for each species and 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. The NCBI accession IDs for the sequences used in both approaches are reported in Supplementary Table S2. We confirmed the consistency of the two reconstruction methods with Geneious (v.10.2.2; Kearse et al., 2012). We manually inferred

the circularity of the sequences, and we mapped the reads of the pool back onto the resulting mitochondrial genome to verify the reconstruction and assess the coverage using Geneious (v.10.2.2; Kearse et al., 2012).

2.2. Orthology Inference between clownfishes and publicly available Actinopterygii genomes, codon alignments and gene tree reconstruction

Orthologous genes between the ten clownfish species, *P. moluccensis*, and 12 publicly available Actinopterygii species (*Astyanax mexicanus*, *Danio rerio*, *Gadus morhua*, *Gasterosteus aculeatus*, *Lepisosteus oculatus*, *Oreochromis 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. We then classified HOGs as single-copy orthologous genes (1-to-1 OGs) or multicopy HOGs (Marcionetti et al. 2019). All HOGs were functionally annotated by appending the function of the genes composing them. The use of additional Actinopterygii species was necessary for positive selection analyses (see below) because the power to detect patterns of positive selection increases with increasing divergence and species (Anisimova et al., 2002). We also used the outgroup species to investigate the history of gene duplication in the clownfishes (see *Section 2.6*).

We performed HOGs codon alignments as reported in Marcionetti et al. (2019). For each HOG, we produced protein alignments with MAFFT (v.7.130; Katoh & Standley, 2016), using the G-INS-i strategy and controlling for over-alignment with the `-allowshift` option. Codon alignments were obtained from protein alignments using PAL2NAL (Suyama et al., 2006). Because positive selection analyses are sensitive to alignment errors (Fletcher and Yang 2010), alignments were filtered to keep only high-confidence homologous regions following the approach done for the Selectome database (Moretti et al., 2014).

For each HOG, we reconstructed the gene tree from the unfiltered codon alignments with 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 ($df=4$). Unfiltered alignments were preferred to filtered alignments as the filtering steps frequently worsen single-gene phylogenetic inference (Tan et al., 2015).

2.3. Mitochondrial and Nuclear Phylogenetic Trees

We investigated the relationship between the species by reconstructing 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 & Standley, 2013). We visually checked the alignments to avoid poorly aligned regions, and we reconstructed the mitochondrial phylogenetic tree with RaxML (v.8.2.12; Stamatakis, 2014) under the GTR+ Γ model, performing 100 bootstrap replicates. We reconstructed the nuclear phylogeny on the concatenated alignment of the 13,500 1-to-1 OGs. In the case of missing genes in some species, gaps were introduced to maintain the correct concatenation. The nuclear phylogenetic tree was reconstructed with RaxML (v.8.2.12; Stamatakis, 2014) under the GTR+ Γ model (single model for the whole concatenated alignment), performing 100 bootstrap replicates. We did not use multispecies coalescent approaches because our goal was not to infer an accurate species tree but rather to investigate topological inconsistency in clownfishes (see below). Thus, it was only used as a reference topology.

We plotted the mitochondrial and nuclear phylogenies 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.

2.4. Topology inconsistency along the genome

We investigated the presence of topology inconsistencies reflecting potential hybridization events or incomplete lineage sorting across the nuclear genome of clownfishes. Because the genome assemblies of all species were based on *A. frenatus* reference (see Marcionetti et al., 2019), we assumed orthology between them. We considered scaffolds larger than 100 kb to reduce the number of scaffolds (2,508 scaffolds kept out of 17,801) while keeping most genomic information (80% of the total assembly length; Supplementary Table S3). In addition, we considered only scaffolds present in all clownfish species and *P. moluccensis*, the latter being used to root the phylogenetic trees.

Scaffolds were aligned using MAFFT (-auto parameter; v.7.130; Katoh & Standley, 2013). We filtered the resulting alignments with trimAl (--gappyout; v.1.4.1; Capella-Gutiérrez et al., 2009) to remove poorly aligned and gaps-rich regions. Alignment statistics before and after the filtering are available in Supplementary Table S3. We split the filtered alignments to obtain non-overlapping windows of 10 kb, 50 kb, or 100 kb. We only kept the last window of each alignment if its length was at least 60 % of the considered window length. We

reconstructed phylogenetic trees for each window using PhyML (GTR+ Γ model, 100 bootstraps; v3.3; Guindon et al. 2010). We rooted the trees with *P. moluccensis*, which we removed before further analyses. We checked 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 Figure S1).

We visually investigated the tree topologies with DensiTree (v.2.2.5; Bouckaert, 2010), as implemented in the phangorn R package (v.2.4.0; Schliep, 2011). 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). Groups of similar trees were identified by hierarchical clustering, using the function *findGroves* in the treespace R package (cutoff set to 380, see cluster dendrogram Figure 4A). Results obtained with the different window sizes (10 kb, 50 kb, and 100 kb) were consistent, and we only considered windows of 100 kb. We visualized the MDS plot with ggplot2 (v.3.0.0; Wickham, 2016), and we plotted trees with the ggtree R package (v.1.14.6; Yu et al., 2017). We used the R package ape (v.5.2; Paradis and Schliep, 2019) to further explore the presence of topologies linked with host and habitat divergence (i.e., we searched for trees where specialist or generalist species branched together),

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 Gene Ontology (GO) enrichment analysis for specific regions of the genome (see below), contrasting the GO annotations of the genes in these regions against those of all genes present. We used the topGO package (v.2.26.0; Alexa & 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-values < 0.01 considered significant. We present here raw p-values instead of p-values corrected for multiple testing, following recommendations from the topGO manual.

2.5. Transposable elements (TEs) annotation and analyses

We identified *de novo* transposable element (TE) families for the ten clownfish species and *P. moluccensis* with RepeatModeler (v.1.0.11; engine ncbi; Hubley & Smit, <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 obtained 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.fishtedb.org/> (Shao et al., 2018). We annotated the TEs in clownfish and *P. moluccensis* genomes with RepeatMasker (v.4.0.7, <http://www.repeatmasker.org/>; Smit et al., 2015), using these TE libraries. We obtained information of TEs 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 TEs 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; Kimura, 1980). We obtained the Kimura distance between each annotated TE copy and the consensus sequence of the respective TE family with the scripts `calcDivergenceFromAlign.pl` and `createRepeatLandscape.pl` provided in the RepeatMasker util directory.

2.6. 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 the 2,725 multicopy HOGs and filtered out the gene copies of the additional outgroup species not considered here. We employed a phylogenetic duplication analysis (PDA) approach (similarly to Brawand et al., 2014), counting the number of gene duplication events observed at each branch of the phylogenetic trees of these species. Within each HOG, a gene duplication event was counted when the number of gene copies in all the species descending from that branch was higher than the maximum number of gene copies in all other species. This approach is conservative, as it does not consider parallel gene duplications (i.e., duplications events happening independently in different taxa) or gene losses. However, it reduces in the deeper branches the errors caused by misassembled genes. Indeed, fragmented gene annotations or separate assemblies of divergent alleles may be mistaken for paralogous genes in terminal branches. This bias is decreased in internal branches as an increasing number of species must display the higher number of gene copies. Because of this,

we only considered duplication events in the internal branches.

The number of duplication events in each branch was normalized to account for the divergence between species. We estimated the neutral genomic divergence between the species using ca. 7.5 million fourfold degenerate sites that we obtained from the codon alignments of 1-to-1 OGs (see *Section 2.2*). We extracted the fourfold degenerate sites in each alignment, and we concatenated them into a single alignment before reconstructing the phylogenetic tree with RaxML (GTR+ Γ model, 100 bootstraps; v.8.2.12; Stamatakis, 2014). The number of duplication events detected on each branch was then divided by the corresponding branch length to obtain the duplication rate (i.e., the number of duplications normalized by the neutral divergence between the species).

2.7. Overall rate of evolution of clownfish genes

We explored the overall rate of evolution of clownfish genes by estimating the ratio of non-synonymous over synonymous substitutions (ω or Ka/Ks or dN/dS) in clownfish compared to the outgroups species. Values of ω smaller, equal, or larger than one correspond respectively to purifying selection, neutral evolution, and positive selection. With a homemade script, we randomly selected 20 1-to-1 OGs and concatenated their alignments. In the case of missing genes in some species, gaps were introduced to maintain the correct concatenation. We repeated the procedure 50 times to obtain 50 concatenated alignments of 20 randomly selected genes. For each alignment, we reconstructed the gene tree using PhyML (GTR model, 100 bootstraps v3.3; Guindon et al. 2010). We labeled the clownfish clade in the obtained trees and estimated the ω 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 ω estimates obtained for the clownfishes and those obtained for the outgroups with a Welch Two Sample *t*-test.

2.8. Gene trees vs species tree for selection and evolutionary rate analyses

The tree topology affects the inference of selection (Diekmann and Pereira-Leal, 2015, see *Section 2.9 and 2.10*) and can bias the analyses of the evolutionary rate associated with host usage (see *Section 2.11*). When topological incongruence exists, the use of either gene trees or the species tree may lead to inconsistent results. The species topology does not always represent the correct gene genealogy across the whole genome (see *Section 2.4*), while the gene tree reconstruction can result in inaccurate topologies due to sampling errors (Planet, 2006). This is particularly the case in highly conserved protein-coding genes, where the number of variable

sites is limited. Additionally, topological accuracy in gene trees also decreases with alignment errors (Ogden & Rosenberg, 2006).

We performed Shimodaira-Hasegawa tests (SH-test; Shimodaira and Hasegawa, 1999) between the species tree and all gene trees to assess the level of topological incongruence. This test assesses whether a set of selected trees are equally good explanations of the data (null hypothesis) or if a set of trees are significantly better (alternative hypothesis). We computed the likelihood of the species and gene trees given the alignments with the *pml* and *optim.pml* functions implemented in the phangorn R package (v.2.4.0; Schliep, 2011), and we performed the SH-test with the *SH.test* function of the same library. We classified the OGs into three categories: “better gene trees” (GT; 206 OGs), “better species trees” (ST; 2,021 OGs), and “non-significant” (NS; 11,273 OGs). We verified that genes in the GT category were not located in genomic regions showing alternative topologies (see *Section 2.4*). We visually inspected these alignments and computed the following alignment statistics: alignment length, number of gaps, and number of variable sites. As OGs in this category showed overall shorter and less accurate alignments (Supplementary Table S4), their gene trees probably resulted from alignment errors, and we did not consider them further.

The analysis of the evolutionary rate (see *Section 2.11*) was performed on the gene trees, but only for genes in the NS category. We did not consider OGs in the ST category, as using gene trees for these OGs would not have been accurate (i.e., gene trees are not a good explanation of the data). We performed the selection analyses using the species tree (see *Section 2.9* and *Section 2.10*), considering the OGs in the NS and ST categories. We nevertheless verified the effect of using either the species or gene trees (Diekmann and Pereira-Leal, 2015) by randomly selecting 200 1-to-1 OGs and repeating the positive selection analyses following the same procedures (*Section 2.9*) but using the gene tree. We compared the results by visually investigating the log-likelihood obtained for each model. We examined whether the same OGs were inferred as positively selected after multiple-testing corrections of the *p-values*, and we verified that similar estimates of ω were obtained with the species and gene trees.

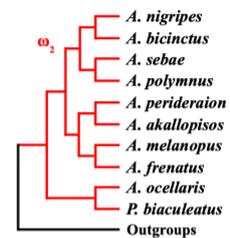
2.9. Positive selection analyses on the whole clownfish group

We tested for the presence of genes under positive selection in the whole clownfish clade. For single copy genes (1-to-1 OGs), we used the branch-site model implemented in *codeml* (PAML, v.4.9; Yang, 2007) to analyze the filtered codon alignments and the species tree (see *Section 2.8*). All branches of the clownfish group were set as foreground branches, while the

remaining branches were assigned to the background (Figure 1A). In the null model, the foreground ω was constrained to be smaller or equal to 1, while in the alternative model, the foreground ω was estimated but was forced to be larger than 1 (Figure 1A). For each 1-to-1 OG, the best model was determined with a likelihood ratio test (LRT; $df=1$). We corrected the resulting 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., 2013; 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 & Dos Reis, 2010), we fitted both the null and the alternative models three times.

A

Sites	Class 0		Class 1		Class 2a			Class 2b		
	Prop.	Backg. Foreg.	Prop.	Backg. Foreg.	Prop.	Backg.	Foreg.	Prop.	Backg.	Foreg.
H1	p_0	$0 < \omega_0 < 1$	p_1	$\omega_1 = 1$	$\frac{(1-p_0-p_1)p_0}{p_0+p_1}$	$0 < \omega_0 < 1$	$\omega_2 \geq 1$	$\frac{(1-p_1-p_0)p_1}{p_0+p_1}$	$\omega_1 = 1$	$\omega_2 \geq 1$
H0	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.	ω	Prop.	ω	Prop.	ω
Clade Model C	p_0	$0 < \omega_0 < 1$	p_1	$\omega_1 = 1$	$1 - p_0 - p_1$	$\omega_{\text{outgroup}} \cdot \omega_{\text{specialists}} \cdot \omega_{\text{generalists}} > 0$
M2a_rel	p_0	$0 < \omega_0 < 1$	p_1	$\omega_1 = 1$	$1 - p_0 - p_1$	$\omega_{\text{outgroup}} (= \omega_{\text{specialists}} = \omega_{\text{generalists}}) > 0$

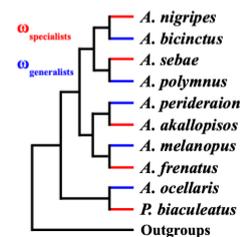


Figure 1. Models used for the positive selection analysis in the whole clownfish clade (A) and the analysis of selection linked with host and habitat divergence (B). A) In the branch-site model A, the null model with fixed ω_2 (H0) was compared to the alternative model (H1), where ω_2 was estimated. B) The null model M2a_rel with equal ω across branches was compared to the Clade Model C, where ω could vary in the three groups (i.e., specialist branches in red, generalist branches in blue and outgroups in black).

We investigated the power and type I error in detecting positive selection by simulating data using *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 following the species tree topology and with the branch lengths randomly drawn from the branch length distributions obtained from all the gene trees of the analyzed 1-to-1 OGs. For each tree, we simulated four sets of alignments with either no ($\omega_2 = 0.5$ and $\omega_2 = 1$) or an increasing level (ω_2 of 2, 5, 10 and 20) of positive selection. We used the same pipeline

as described above to estimate positive selection. We investigated the power to detect positive selection and the number of false positives (type I errors) by recording the number of significant LRT (p -value < 0.05) between the null model and the alternative model.

We retrieved the position on *A. frenatus* scaffolds of each gene under positive selection. This position was then transferred on the *A. percula* reference genome (see *Section 2.4*) to associate these genes to *A. percula* chromosomes. We selected the genes in the upper 90% of the foreground ω distribution and examined their function on the UniProt database (UniProt Consortium, 2018). We did gene ontology (GO) enrichment analysis by contrasting the GO annotation of all the significant genes against all analyzed genes (13,294 1-to-1 OGs) using the TopGO package (v.2.26.0; Alexa & Rahnenfuhrer, 2016) available in Bioconductor (<http://www.bioconductor.org>). Fisher's exact tests with the weight01 algorithms were applied to examine the significance of enrichment, with p -values < 0.01 considered as significant. We present here *raw p-values* instead of p -values corrected for multiple testing, following recommendations from the TopGO manual. We extracted the genes associated with the significantly enriched GO terms and retrieved their functional information from the UniProt database (UniProt Consortium, 2018).

We investigated the presence of positive selection on some copies of the multicopy HOGs using the method aBSREL implemented in HyPhy (v.2.3.7; Smith et al., 2015). We performed the analysis in an exploratory way, testing for positive selection at each branch of the tree. Although this approach has a reduced power due to multiple testing, it was preferred as we did not know beforehand which copy of the genes could be positively selected. Thus, it allowed obtaining an overview of the selective pressures acting on the genes in all species. We corrected for multiple testing with the Benjamini-Hochberg method implemented in the *qvalue* package in R (FDR threshold of 0.1, “robust” option, and “bootstrap” method; Dabney et al., 2010). A similar analysis of multicopy genes was performed in Marcionetti et al. (2019), and we followed the same procedure here.

2.10. Selection signatures associated with hosts and habitats divergence

We tested for signatures of selection associated with hosts and habitats divergence using the clade model C implemented in *codeml* (PAML v.4.9; Yang, 2007; Figure 1B), using the filtered codon alignments and the species tree (see *Section 2.8*). Clade models allow differences in site-specific selective constraints among clades in the tree (Bielawski and Yang, 2004; Forsberg and Christiansen, 2003; Weadick and Chang, 2012). Here, we assigned clownfish

species to either a specialist or generalist category, and we labeled the terminal branches of clownfishes with this host usage classification (Supplementary Table S1 and Figure 1B). The remaining branches were assigned to the background category. We fitted the clade model C, estimating separate ω ratios for each category in the analysis, and we compared it to the null model M2a_rel, where ω is fixed among clades (Figure 1B), with a LRT ($df=2$). The resulting p -values were corrected for multiple testing with the Benjamini-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: OGs with patterns of purifying selection in both clownfish groups ($\omega_{\text{specialists}}$ and $\omega_{\text{generalists}} < 1$) and OGs with signatures of positive selection in all specialists and/or generalist species ($\omega_{\text{specialists}}$ or/and $\omega_{\text{generalists}} > 1$).

For the first group, we looked for genes showing a different pattern of purifying selection in either specialist or generalist species. For this, we estimated the distribution of $\omega_{\text{background}}$, $\omega_{\text{specialists}}$ and $\omega_{\text{generalists}}$ and tagged genes falling in the lower 10% of the $\omega_{\text{specialists}}$ distribution but in the upper 10% of the $\omega_{\text{generalists}}$ and $\omega_{\text{background}}$ distributions as those experiencing intensified purifying selection in specialists (Figure 2A). Likewise, genes experiencing intensified purifying selection in generalists were obtained by considering the genes in the lower 10% of the $\omega_{\text{generalists}}$ distribution and in the upper 10% of the $\omega_{\text{specialists}}$ and $\omega_{\text{background}}$ distributions (Figure 2B). Finally, genes experiencing relaxed purifying selection in either specialists or generalists species were retrieved by considering those genes falling in the upper 10% of respectively the $\omega_{\text{specialists}}$ or $\omega_{\text{generalists}}$ distributions, but in the lower 10% of the remaining ω distributions (Figure 2C-D) 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.

For the second category, significant results were classified into genes positively selected only in specialists ($\omega_{\text{specialists}} > 1.5$; $\omega_{\text{generalists}}$ and $\omega_{\text{background}} \leq 1$) and genes positively selected only in generalists ($\omega_{\text{generalists}} > 1.5$; $\omega_{\text{specialists}}$ and $\omega_{\text{background}} \leq 1$). 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.

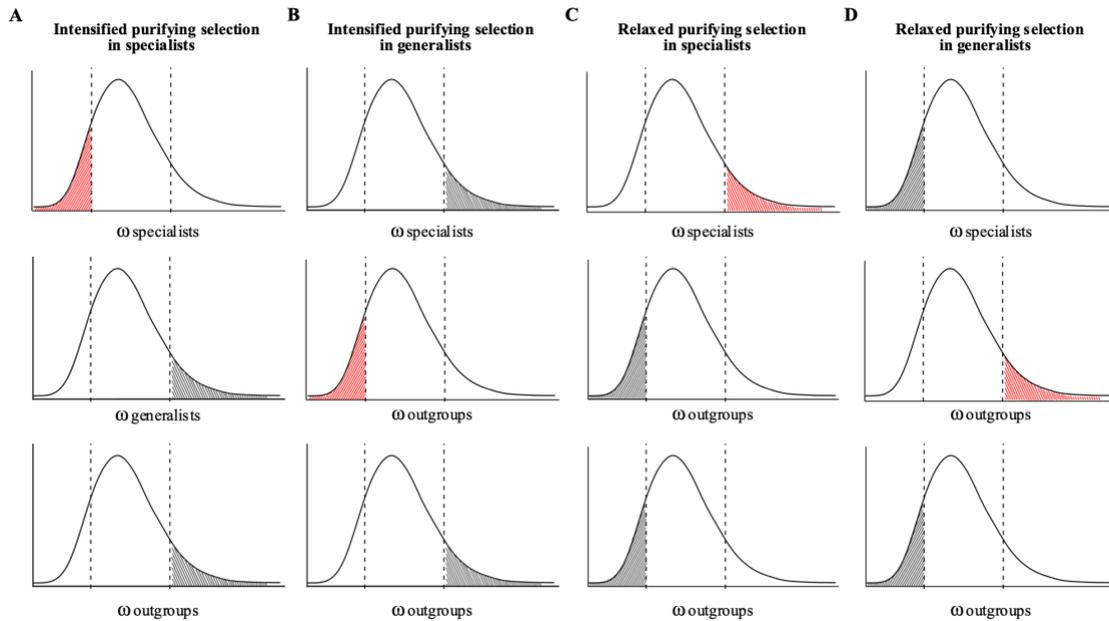


Figure 2. Illustration of the method used to define the categories of genes showing intensification or relaxation from purifying selection. Plots represent the distribution of ω estimates in specialists, generalists, and outgroups. Dotted lines represent the lower and upper 10% of the distribution. We defined genes with intensification of purifying selection in specialists (A) and generalists (B) and genes with relaxed purifying selection in specialists (C) and generalists (D).

We manually investigated the functional annotation of the genes identified with the procedure described above by performing GO enrichment analyses with the TopGO package (v.2.26.0; Alexa & Rahnenfuhrer, 2016) available in Bioconductor (<http://www.bioconductor.org>). We contrasted the GO annotation of the significant genes in each category against all analyzed genes (13,294 1-to-1 OGs) and followed the same procedure as for the positively selected genes in the whole clownfish group (see *Section 2.9*).

2.11. Evolutionary rate linked with host and habitat usage

We retrieved the reconstructed gene trees for each 1-to-1 OG (see *Section 2.8*), and we considered the branch length as a proxy of the evolutionary rate of the OGs. We investigated the presence of OGs with the 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. We considered the five pairs as replicates, and for each OG, we tested whether the difference was significantly different from zero with a one-sample Student's t-test. We corrected for multiple testing using the *p.adjust* function in R (method FDR). Additionally, for each pair

of closely related species, we computed the distribution of the difference in branch length between generalists and specialists. We defined genes with a higher evolutionary rate in specialists and generalists those genes in the lower and upper 5% of the distribution, respectively (see Figure 9A). We investigated the intersection of these genes in the five pairs of closely related species to obtain genes that have a parallel increase of evolutionary rate in specialist or generalist species. We plotted the results with the function *venn* (*venn* R package, v. 1.10; Dusa, 2021). We tested whether the number of genes shared between 2, 3 and 4 species pairs was significantly different between generalists and specialists by performing two-sample Student's t-tests for each category.

3. RESULTS

Information on the nuclear genome assembly and annotation for the ten clownfish species and *P. moluccensis* is available in Marcionetti et al. (2019). Details on the retrieved hierarchical orthologous groups (HOGs) between clownfishes and the 13 Actinopterygii outgroup are also reported in Marcionetti et al. (2019). We obtained a final set of 15,940 HOGs, among which 13,500 were single-copy orthologous genes (1-to-1 OGs), while the others were classified as multicopy orthologs. We used the genome assemblies and this set of genes to investigate the genomic architecture and the molecular evolution associated with the adaptive radiation of clownfishes (3.1. *Genomic substrate of clownfish adaptive radiation*) and to test for patterns of parallel evolution in specialist and generalist species (3.2. *Parallel evolution associated with host and habitat divergence*).

3.1. *Genomic substrate of clownfish adaptive radiation*

3.1.1 Mosaic genomes in clownfishes

We investigated the presence of topological inconsistencies between the mitochondrial and the nuclear genomes, as well as throughout the nuclear genome of clownfishes. Cytonuclear incongruences and topological disparities suggest that hybridizations or incomplete lineage sorting (ILS) occurred in the group. Additionally, inconsistencies where ecologically similar species branch together could indicate introgressive hybridization potentially involved in the buildup of the convergent phenotypes.

In both mitochondrial and nuclear phylogenetic trees, each species branched with the expected closely related pair (Figure. 3), which confirmed that the ten species formed five pairs

of phylogenetically close but ecologically distinct species. Nevertheless, we found cytonuclear discordance at deeper nodes in the tree. Indeed, the pairs *A. melanopus* - *A. frenatus* and *A. perideraion* - *A. akallopisos* formed two sister groups in the nuclear phylogeny, while the latter was basal to *A. melanopus* - *A. nigripes* in the mitochondrial phylogenetic tree (Figure 3). This cytonuclear discordance was highly supported, with bootstrap support for the node higher than 0.95 (Figure 3).

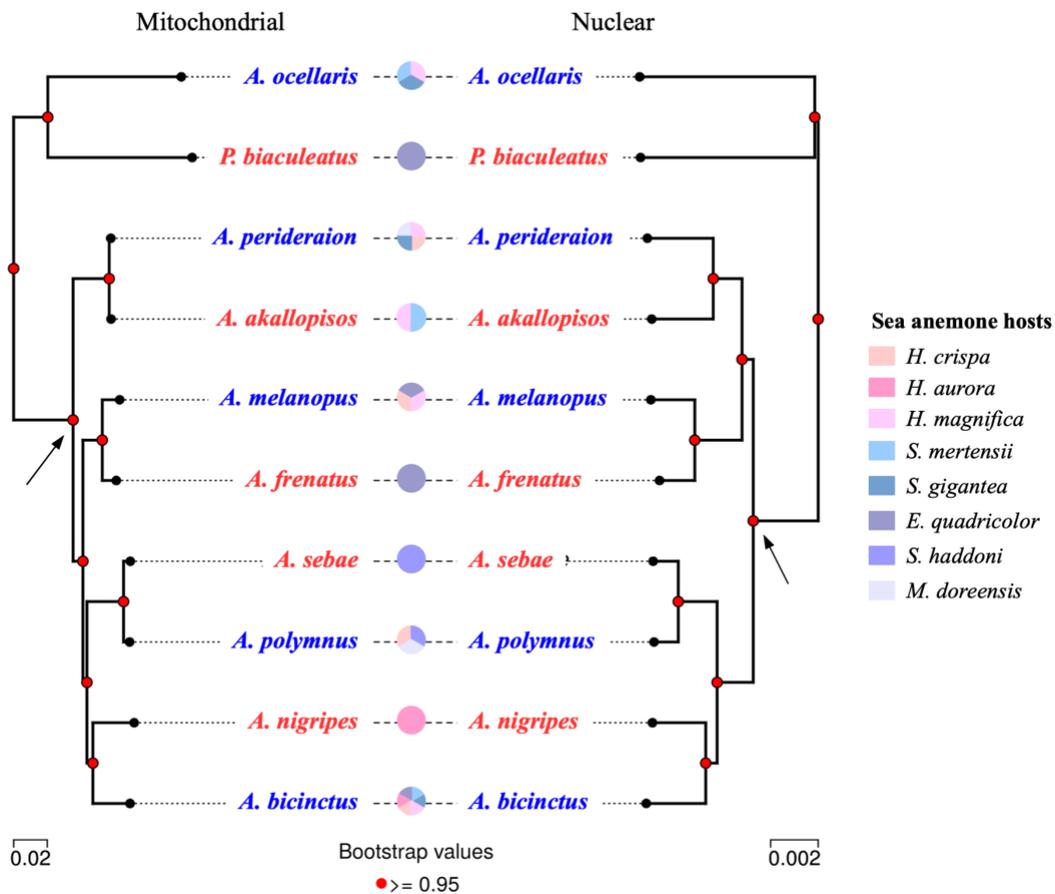


Figure 3: Mitochondrial and nuclear phylogenetic trees of the ten clownfish species. The mitochondrial phylogenetic tree was obtained with RAxML from the whole mitochondrial genome alignments (total of 16,747 bp). The nuclear phylogenetic tree was obtained with RaxML from the concatenated alignment of 13,500 orthologous genes (total of 22.5 Mbp). For both trees, 100 bootstrap replicates were performed. All nodes have bootstrap support higher than 0.95. The black arrows illustrate the topology discordance. Clownfish possible interactions with sea anemones are shown, with generalist and specialist species reported in blue and red, respectively.

For the analysis of topological inconsistencies along the genome, phylogenetic reconstruction for non-overlapping windows of 100 kb resulted in a total of 5,936 trees. Overall, the trees were well supported, with an average bootstrap value higher than 0.95 for 78 % of the

windows (99.7 % when considering average bootstrap support higher than 0.8; Supplementary Figure S1).

Visual investigation of the tree topologies showed the presence of topological disparities in deep nodes, while the five pairs of closely related species predominantly branched together (Supplementary Figure S2). We quantified the different topologies along the nuclear genome and found five major clusters of trees (Figure 4). Clusters 1 and 2 contained 27% and 57% of the 5,936 gene 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 thus we did not consider them as a cluster (Figure 4B, Supplementary Figure S3). All clusters contained trees with two different topologies that differed only by the branching of *P. biaculeatus* (depicted in Figure 4C by dotted lines), with a slight majority of trees placing *P. biaculeatus* as the sister species of *A. ocellaris* (between 53% and 74% of the gene trees), except in cluster 5, where *P. biaculeatus* was found sister 90.8% of the time (Figure 4C).

The main differences between the topologies of these five groups were given by the branching of the *A. akallopisos* – *A. perideraion* pair. Indeed, *A. akallopisos* – *A. perideraion* and *A. melanopus* – *A. frenatus* were sister groups in cluster 2. This topology was the most represented in the genome and corresponded to the nuclear phylogenetic tree (Figure 3) and the expected species topology. In cluster 1, the pair *A. akallopisos* – *A. perideraion* was basal to the *A. bicinctus* – *A. nigripes* – *A. polymnus* – *A. sebae* complex (Figure 4C). Cluster 3 was characterized by the *A. akallopisos* – *A. perideraion* pair being basal to the *A. melanopus* – *A. frenatus* pair, which is the topology of the mitochondrial phylogenetic tree (Figure 3). In cluster 5, *A. perideraion* and *A. akallopisos* were not branching as sister species, but *A. akallopisos* was basal to the *A. bicinctus* – *A. nigripes* – *A. polymnus* – *A. sebae* group, while *A. perideraion* was basal to the *A. frenatus* – *A. melanopus* pair (Figure 4C). Cluster 4 was similar to cluster 2, but with *A. bicinctus* being basal to *A. nigripes* (Figure 4C).

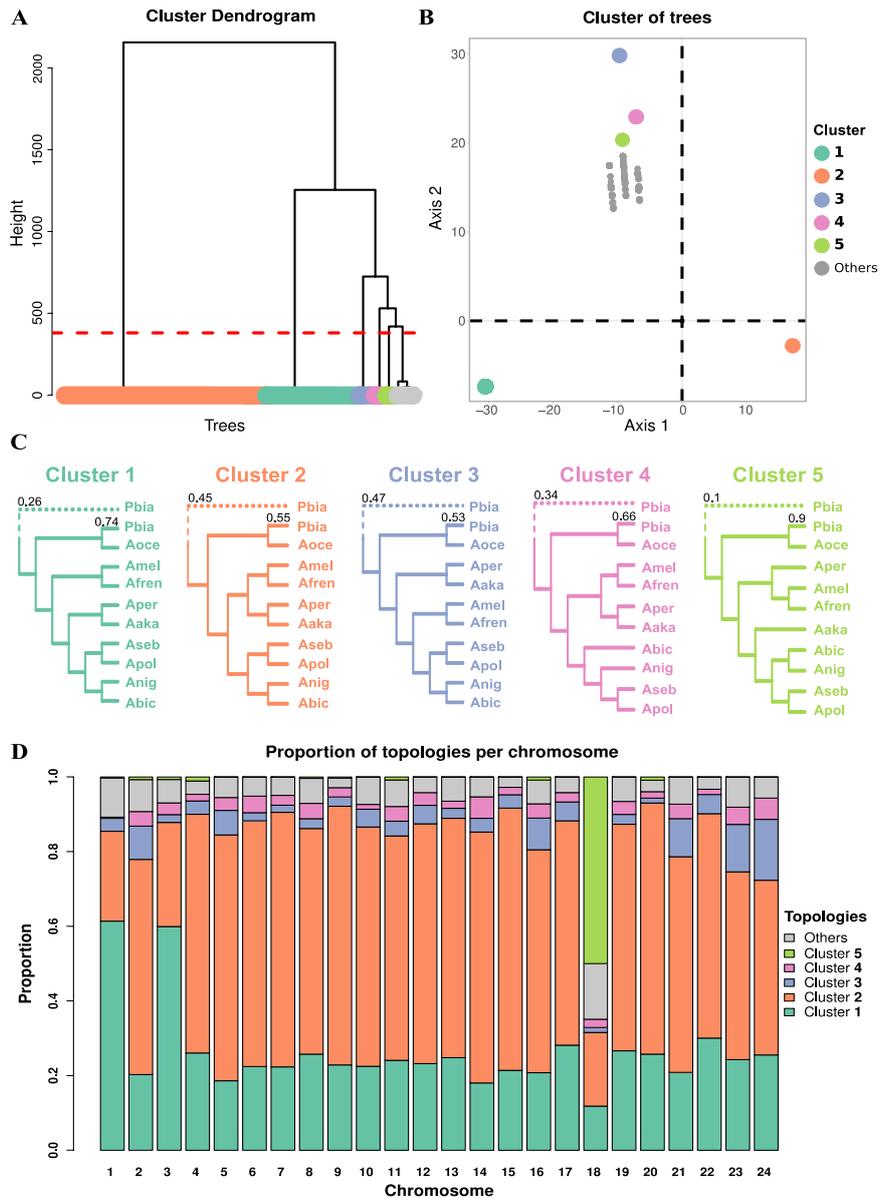


Figure 4: The five major topologies observed along the nuclear genome of clownfishes. A) Hierarchical clustering based on the Robinson-Foulds distance between the trees reconstructed along the genome. The red line corresponds to the cutoff set to define the five major clusters of trees. The last cluster (gray) is not considered because it is composed of different topologies (see DensiTree plot in Supplementary Figure S3). B) Multidimensional scaling plot of the five clusters of trees. C) The topologies of the five main clusters of trees represented in the genome. Pbia, Aoce, Amel, Afre, Aper, Aaka, Abic, Anig, Aseb, and Apol correspond to *P. biaculeatus*, *A. ocellaris*, *A. melanopus*, *A. frenatus*, *A. perideraion*, *A. akallopisos*, *A. bicinctus*, *A. nigripes*, *A. sebae*, and *A. polymnus*, 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. D) Distribution of the five topologies across the chromosomes of *A. percula*. The additional topologies, reported in gray, are variable (see DensiTree plot in Supplementary Figure S3).

The five topologies were distributed across the genome, with 1,053, 1774, 224, 165, and 55 scaffolds showing topologies of clusters 1 to 5, respectively. Figure 4D shows the distribution of the six topologies across the 24 chromosomes of the *A. percula* genome (Lehmann et al., 2019). The pattern was similar for most of the 24 chromosomes, with the expected species tree (cluster 2) most frequently observed. Nevertheless, we noticed three exceptions for chromosomes 1, 3, and 18 (Figure 4D). On chromosomes 1 and 3, the most abundant topology (around 60% of the windows) was the one of cluster 1 (Figure 4D). Chromosome 18 showed a completely different scenario, with the topology of cluster 5 being the most frequent. Indeed, 88% of the windows supporting this topology were located on chromosome 18, and, altogether, they accounted for 50 % of the topologies observed on this chromosome (Figure 4D). In these regions of chromosome 18, we identified 432 genes, of which 331 were functionally annotated. Gene Ontology (GO) enrichment analysis of these genes resulted in 24 enriched terms (p -value < 0.01; Supplementary Table S5). Among them, we found GO terms associated with the morphogenesis of the epithelium (GO:1905332), fertilization (GO:0009566), axis elongation (GO:0003401), and retina vasculature development (GO:0061298).

None of the five clusters of trees was summarized by topologies with specialist or generalist species branching together. Nonetheless, when we considered the remaining and more variable trees (represented in gray in Figure 4A, B, D), we observed 13 in which the specialist species *A. nigripes* and *A. sebae* branched as sister species (Supplementary Figure S4). In all the trees, the node splitting the two species had bootstrap support of 100 %. The windows showing these topologies were located on ten different chromosomes and contained 57 annotated genes (Supplementary Table S6). Among them, we found the genes encoding for olfactory receptors (*OR52N5*, Olfactory receptor 4S1; *OR4S1*, Olfactory receptor 4S1) and the gene *Has2*, encoding for the hyaluronan synthase 2.

3.1.2. Transposable elements content

We investigated transposable elements (TEs) content and transposition history in clownfishes to determine if TEs potentially played a role in the group's diversification. We found that approximately 23-25% of clownfish genomes consisted of TEs (Figure 5, left and Supplementary Table S7). More than 50% of TEs were DNA transposons, while the second most abundant families were LINE transposons (Figure 5, right). Similar TEs proportion and

composition were observed in the *A. percula* and *O. niloticus* genomes (Figure 5).

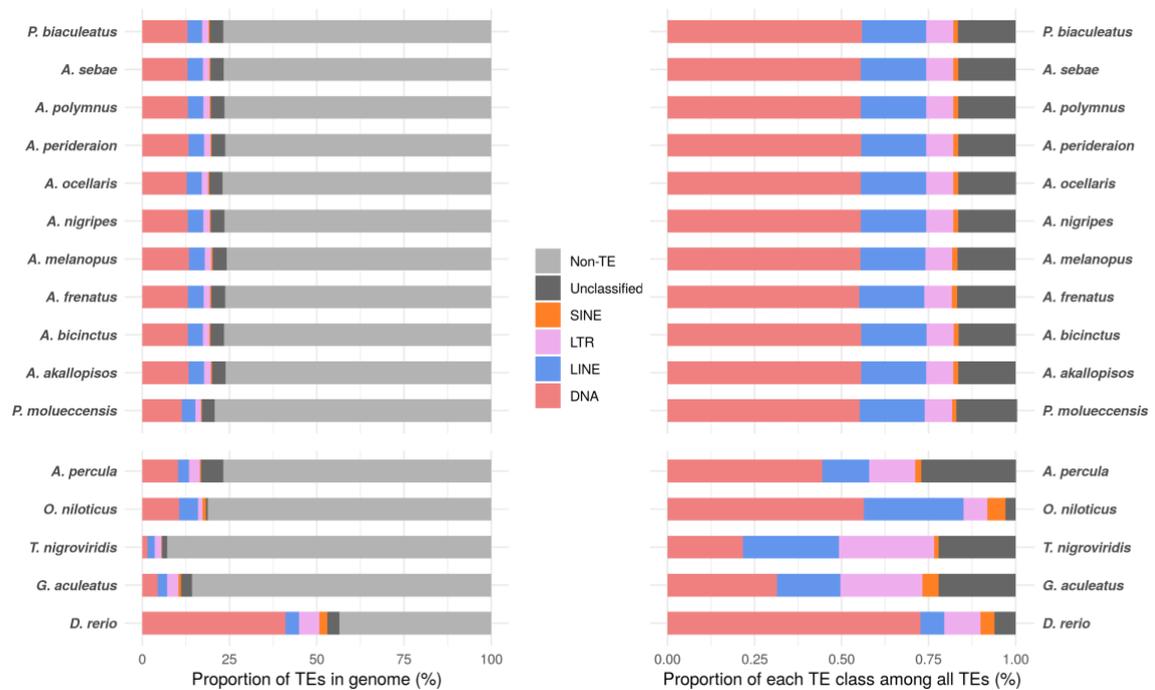


Figure 5. Overall proportion of TEs (left) and proportion of each major TE class (right) in clownfish, *P. moluccensis* and additional Actinopterygii genomes. Clownfishes and *P. moluccensis* genomes were annotated with RepeatMasker (<http://www.repeatmasker.org/>). Information on TE content for the additional species was retrieved from previous studies (Brawand et al., 2014; Gao et al., 2106; Lehmann et al., 2019).

We detected two major TE bursts in the clownfish genomes (Kimura distance *K-value* of 0.05-0.06 and 0.18-0.19; Figure 6 and Supplementary Figure S5). The most recent burst was also present in *P. moluccensis*, although it was less pronounced. Indeed, at *K-value* of 0.05 and 0.06, the average TEs percent in clownfish genome was respectively 1.07% (SD=0.06) and 1.14% (SD=0.05), while in *P. moluccensis*, TEs represented only 0.45% and 0.53% of the genome, respectively (Supplementary Figure S5). When comparing the Kimura distance to the neutral genomic divergence between clownfish and the outgroup (see species divergence, Figure 7), we observed that the older TE burst (*K-value* 0.18-0.21) happened just before *O. niloticus* split from 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.

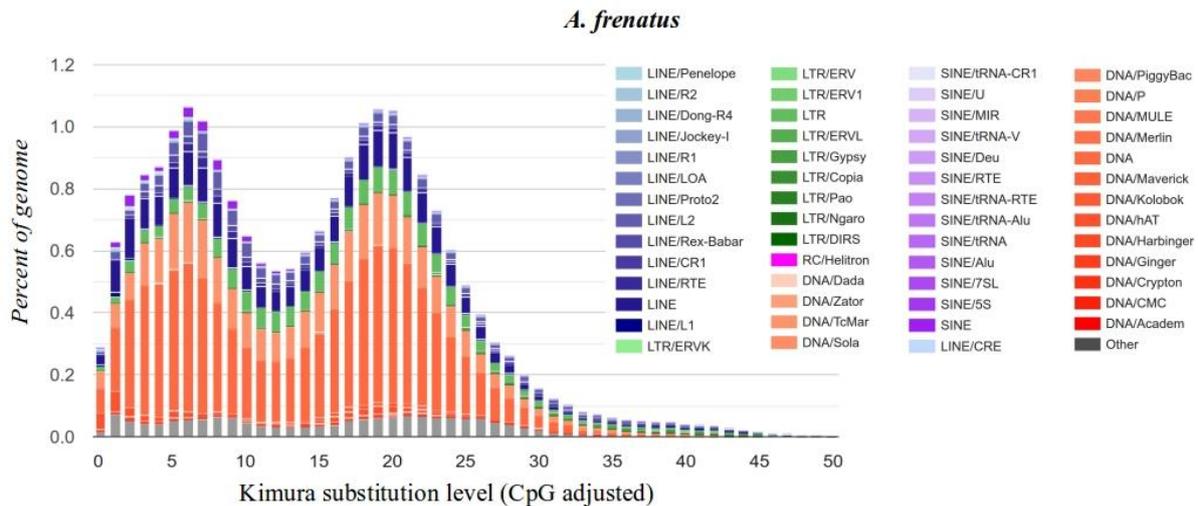


Figure 6. Kimura distance-based copy divergence analysis of transposable elements in *A. frenatus*. The different colors represent the different TE superfamilies. TEs were clustered according to Kimura distances to their corresponding consensus sequence. Copies clustering on the left side of the graph did not greatly diverge from the consensus sequence and potentially corresponded to recent events, while sequences on the right side likely corresponded to older divergence. Peaks in the graph indicate TE bursts. Similar results were obtained for the additional nine clownfish (Supplementary Figure S5).

3.1.3. Gene duplication rate and positive selection on multicopy HOGs

We explored whether clownfishes showed an increased gene duplication rate associated with their diversification. We compared the level of gene duplications in clownfishes and the outgroup species. We found a total of 1,747 HOGs with specific duplication events across the full phylogenetic tree (Figure 7 and Supplementary Figure S6). We investigated when the duplication occurred, and we detected 19 of them in the common ancestor of clownfishes (branch A in Figure 7 and Supplementary Figure S6). In contrast, we observed 90 duplications both in the common ancestor of the Pomacentridae (branch C in Figure 7 and Supplementary Figure S6) and before the split of *P. moluccensis* and the clownfishes (branch B in Figure 7 and Supplementary Figure S6). When normalizing for species divergence, the duplication rates were similar, with about 12 duplicated genes/percent of divergence (Figure 7). A higher duplication rate was observed in the common ancestor of the Pomacentridae and *O. niloticus* (40 duplications/percent of divergence; branch D in Figure 7).

Although clownfishes did not show an increased gene duplication rate, we tested if positive selection acted on the gene copies of the duplicated genes. We analyzed 2,725 HOGs and, while we found 116 genes with signals of positive selection in clownfish species, selection was only acting on a few clownfish branches. Thus, we did not detect positive selection

throughout the whole clade potentially associated with the group's diversification (schematic in Figure 1A). Similarly, we did not observe any HOG with signals of positive selection linked with host and habitat use and thus associated with the evolution of convergent phenotypes (schematic in Figure 1B).

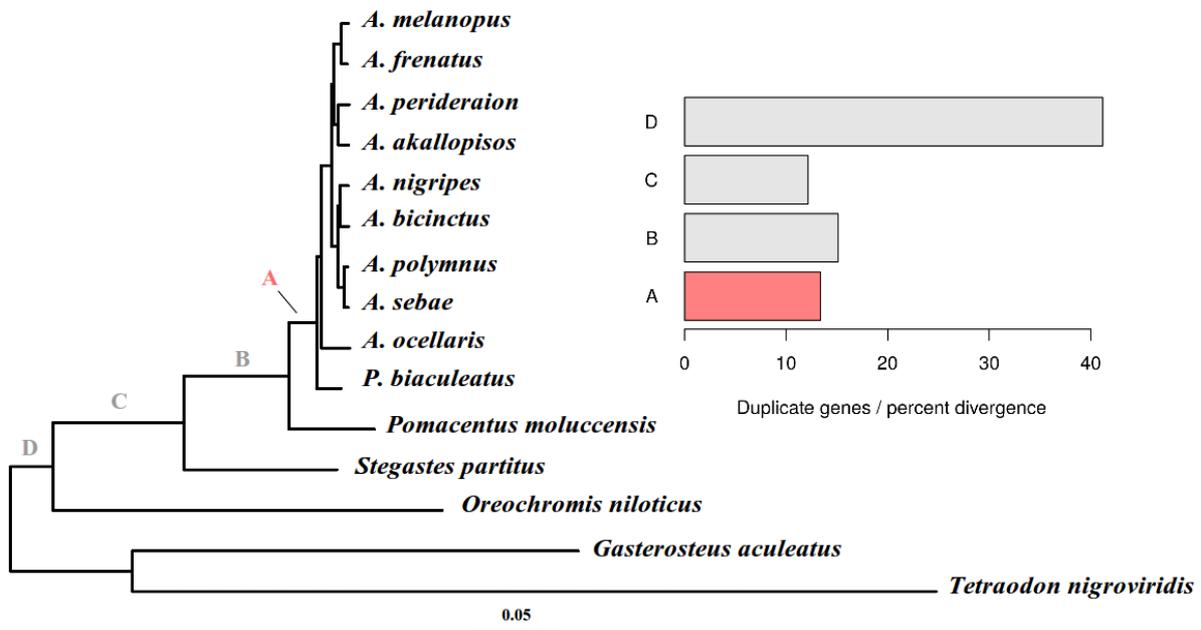


Figure 7: Divergence calculated as the neutral genomic divergence between clownfishes and six outgroup species and duplication rates in four internal branches (on the right). Results for duplication rate are reported for the internal branches labelled as A, B, C and D, and reported in the tree. Branch A corresponds to the ancestral branch of clownfishes. The duplication rate was calculated as the number of duplication events in the branch normalized by the branch length.

3.1.4. Overall accelerated evolution and positively selected single-copy genes in clownfish

We investigated the rate of coding sequence evolution in clownfishes compared to the outgroups. We found a significantly higher ω ratio in clownfishes ($M=0.14$, $SD=0.03$) compared to the outgroups ($M=0.09$, $SD=0.02$; $t(73)=11.0$, $p < .001$; Figure 8) in concatenated alignments of randomly selected OGs, which suggests an overall accelerated evolution in this group.

We investigated the presence of genes that were positively selected in clownfishes (schematic in Figure 1A) and thus, potentially involved in the group's diversification. After correcting for multiple testing, we found a total of 732 genes with a significant signal of positive selection in the whole clade (5.4 % of the genes tested). The positively selected genes were distributed across the 24 chromosomes (between 21 and 50 significant genes per chromosome;

Supplementary Figure S7). After normalizing by the total number of genes mapped to each chromosome, the percentage of positively selected genes was homogeneous across the chromosomes (Supplementary Figure S7). The proportion of sites under positive selection and the estimated ω for those sites varied across positively selected genes (Supplementary Figure S8). On average, the percentage of sites with signatures of positive selection (site class 2a and 2b, Figure 1) was 2.7%, with an average foreground ω of 19.2.

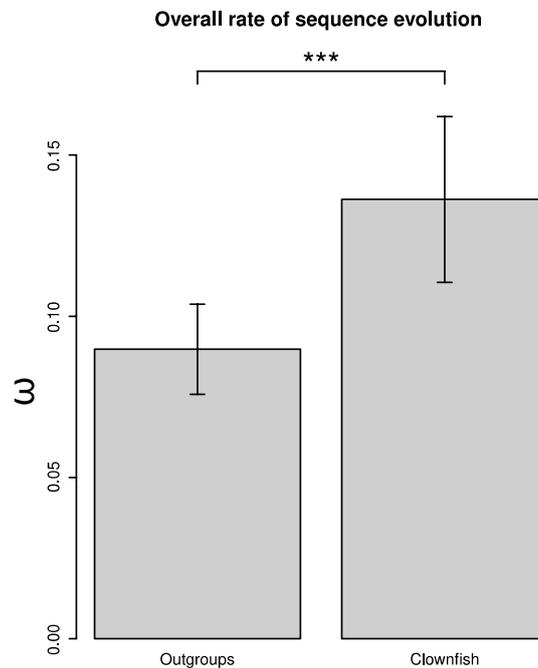


Figure 8: Overall rate of sequence evolution in outgroup species and clownfishes. The values of ω for outgroups and clownfishes were estimated with the branch model implemented in PAML, on 50 replicates of the concatenated alignments of 20 randomly picked genes. The significance of the difference in ω was tested with a two-sample Student's t-test.

Positive selection analyses on 200 randomly selected genes using either the gene trees or the species tree resulted in similar results. Indeed, after correcting for multiple testing, we detected three concordant positively selected genes independently of the tree used (Supplementary Table S9). The data simulated under neutral or purifying selection scenarios showed that positive selection analyses resulted in no false positives detected (Supplementary Figure S9). The power to detect positive selection increased with the increasing strength of selection (i.e., increasing ω , Supplementary Figure S9). While the power was only 5 % when we simulated an ω of 2, it increased to 54 %, 85 %, and 97.5 % for ω of 5, 10, and 20, respectively (Supplementary Figure S9).

Out of the 732 positively selected genes, 721 were functionally annotated. We investigated the function of the genes with the strongest signal of selection (Supplementary Table S9). Among these genes, we found the gene encoding for the isotocin receptor (ITR, OG20045_2b; Table 1) and the gene encoding for somatostatin 2 (SSTS, OG14285; Table 1). GO enrichment analysis performed on the positively selected genes resulted in 30 enriched GO terms (Supplementary Table S10). Among them, 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 S11), we found the genes encoding for the neuropeptide FF receptor 2 (NPFFR2, OG16291), the dual oxidase protein (DUOX, OG8036), and the rhodopsin (RHO, OG8544; Table 1).

Table 1. Positively selected genes in the whole clownfish clade showing particularly interesting functions. 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, Figure 1A) is reported. Likelihood-ratio test (LRT) multiple-testing corrected q values, the proportion of sites under positive selection on the tested branches (% PS sites) and the corresponding ω_2 values are reported for each gene.

OG ID	SwissProt ID	Protein Name	logL(H0)	logL(H1)	LRT q Values	% PS sites	ω_2
OG20045_2b	Q90334	Isotocin receptor (ITR)	-4555.44	-4547.56	4.60E-003	0.3	51.142
OG14285	P01170	Somatostatin-2 (SST2)	-1784.81	-1768.73	4.31E-003	2.7	61.387
OG16291	Q9Y5X5	Neuropeptide FF receptor 2 (NPFFR2)	-5198.93	-5191.55	5.55E-003	1.0	10.922
OG8036	Q8HZK2	Dual oxidase 2 (DUOX2)	-32014.46	-32005.58	4.31E-003	0.4	14.285
OG8544	O18315	Rhodopsin (RHO)	-4389.29	-4374.97	1.12E-002	1.0	26.142

3.2. Parallel evolution associated with host and habitat divergence

Genes involved in the ecological divergence of clownfishes might evolve at different rates or show different selective pressures in specialist and generalist species. Thus, we investigated the presence of genes showing concerted differences in the evolutionary rate and selective

pressures between all specialist and generalist species, which would suggest some level of parallel evolution in clownfish diversification.

3.2.1. Evolutionary rates linked with host and habitat divergence

We calculated the differences in evolutionary rate between generalist and specialist species for each gene. We considered the five species pairs as replicates and tested whether this difference was significantly different from zero, which would suggest a parallel change in the gene's evolutionary rate linked with ecological divergence. After multiple testing corrections, we did not detect any gene with significant differences in the rate of evolution between all generalist and all specialist species.

We identified genes with a higher evolutionary rate in generalists or specialists in each species pair (Figure 9A and Supplementary Figure S10) and investigated whether these genes were shared among species pairs (Figure 9B-C). The number of genes shared between 2, 3, 4, and 5 species pairs was similar in specialists and generalists (two-sample Student's t-tests, p -values > 0.05 , Supplementary Table S13; Figure 9B-C), and it quickly decreased with the increasing number of considered pairs (Figure 9B-C and Supplementary Figure S10). We did not detect any gene showing a higher evolutionary rate in all generalists (Figure 9B) or in all specialists (Figure 9C). When considering genes shared by 2, 3, or 4 species pairs, similar results were obtained for each combination of species pairs, independently from their phylogenetic relationship (Supplementary Figure S11).

3.2.2. Selection patterns linked with host and habitat divergence

We detected 3,991 genes that showed significantly different ω for specialists, generalists, and outgroups (clade model C better than null model M2a_rel; p -value corrected for multiple testing; Figure 1B). Most of these genes (3,889 or 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). These results suggest that, in general, purifying selection is acting more strongly on specialist clownfishes than on generalists.

We explored more in detail the presence of genes with patterns of intensification or relaxation from positive selection linked with host usage (Figure 2). We found a total of 155 genes showing relaxation or intensification of purifying selection in either specialist or

generalist species (Table 2). GO enrichment analysis of 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 S13). 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 S13).

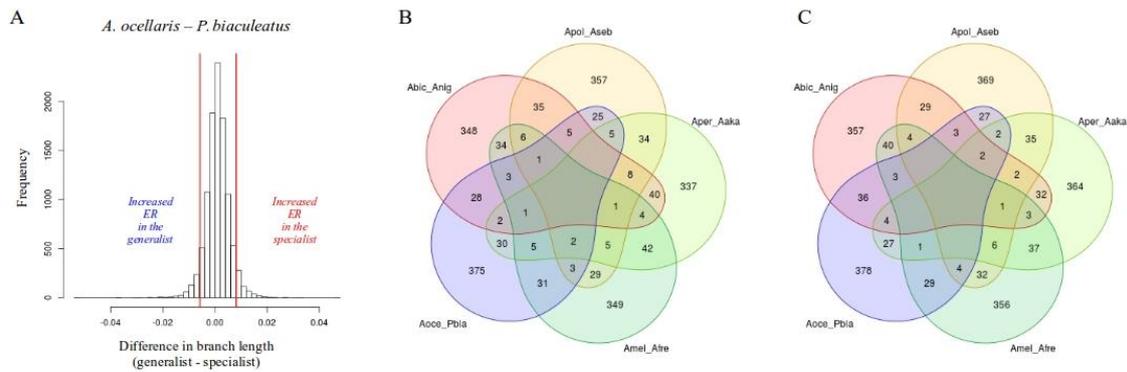


Figure 9: Evolutionary rate linked with host and habitat usage. A) Distribution of the branch length difference between the generalist *A. ocellaris* and the specialist *P. biaculeatus* obtained from the gene tree reconstructions for 11,273 OGs. The branch length was considered as a proxy for the evolutionary rate. We defined genes with a higher evolutionary rate in specialists and generalists those genes in the lower and upper 5% of the distribution, respectively. ER means evolutionary rate. The distributions of the additional species pairs are reported in Supplementary Figure S10). B) Number of genes with a higher evolutionary rate in generalists shared among the five species pairs. C) Number of genes with a higher evolutionary rate in specialists shared among the five species pairs. The species pairs correspond to: Abic_Anig: *A. bicinctus* and *A. nigripes*; Apol_Aseb: *A. polymnus* and *A. sebae*; Aper_Aaka: *A. perideraion* and *A. akallopisos*; Amel_Afre: *A. melanopus* and *A. frenatus*; Aoce_Pbia: *A. ocellaris* and *P. biaculeatus*.

Table 2: Number of genes showing patterns of selection linked with host and habitat divergence. Information on purifying selected genes is available in Supplementary Table S14, while positively selected genes are reported in Supplementary Table S15. The distribution of the ω values for each category is reported in Supplementary Figure S12 (for purifying selection) and Supplementary Figure S13 (for positive selection).

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

Genes with patterns of positive selection specific to specialists ($\omega_{\text{specialists}} > 1.5$; $\omega_{\text{generalists}}$ and $\omega_{\text{background}} \leq 1$; Supplementary Table S15 and Supplementary Figure S13) and generalists ($\omega_{\text{generalists}} > 1.5$; $\omega_{\text{specialists}}$ and $\omega_{\text{background}} \leq 1$; Supplementary Table S15 and Supplementary Figure S13) were also detected (Table 2). GO enrichment analysis on genes positively selected in specialists resulted in two enriched GO terms: GO:0031398 (positive regulation of protein ubiquitination) and GO:0071456 (cellular response to hypoxia; Supplementary Table S13). 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 13).

4. DISCUSSION

In this study, we took advantage of the genomic data available for five pairs of closely related but ecologically divergent clownfish species to first investigate the genomic architecture of clownfish adaptive radiation and, second, explore if the evolution of their convergent phenotypes resulted from parallel genomic changes. Similarly to what was observed in other lineages undergoing adaptive radiation (e.g., Jones et al., 2012; Brawand et al., 2014; Fan & Meyer, 2014; Feiner, 2016; Berner & Salzburger, 2015; Faber-Hammond et al., 2019; Xiong et al., 2021), we found that clownfish genomes show bursts of transposable elements, overall accelerated coding evolution and topology inconsistencies potentially resulting from hybridization events. These characteristics possibly facilitated the rapid diversification of the group by generating genomic variations on which natural selection can act. Additionally, we observed that 5.4% of the analyzed genes showed a positive selection signature throughout clownfish radiation. Among these genes, five have functions associated with social behavior and ecology (discussed below), and they could have played a role in the evolution of clownfishes' size-based hierarchical social structure. Finally, 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 group's diversification.

4.1. The overall genomic architecture of clownfish adaptive radiation

4.1.1 Burst in transposable element

The proportion of transposable elements (TEs) that we found in clownfishes was high (20-25% of the genome), similar to what was observed in East African cichlids (Brawand et al.,

2014; Shao et al., 2019) and in *A. percula* genome (Lehmann et al., 2019), which confirms a reliable annotation of the TEs despite the higher fragmentation of the assemblies used in this study. This large TEs proportion in clownfishes principally originated from two bursts of transpositions. The older burst happened in the ancestor of *O. niloticus* and the Pomacentridae, and it was indeed also observed in the cichlids (Shao et al., 2019) and *P. moluccensis* (Supplementary Figure S5). 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 (Supplementary Figure S5), suggesting that the burst started in the *P. moluccensis*-clownfishes ancestor and, after the split of the species, only continued in clownfishes.

Bursts in TEs transpositions are common in fishes and likely favor species diversity (Shao et al., 2019). Indeed, the movement of TEs can lead to insertions, deletions, and chromosomal rearrangements (Lönnig & Saedler, 2002; Fedoroff et al., 2012), potentially contributing to adaptation, speciation, and diversification processes (Syvanen, 1984; Rebollo et al., 2010, Auvinet et al., 2018). The high percentage of TEs observed in clownfishes, and the increased burst of transposition compared to *P. moluccensis* might therefore have facilitated the diversification of the group, as observed in other adaptive radiations such as the East African cichlid fishes (Brawand et al., 2014; Fan & Meyer, 2014), *Heliconius* butterflies (Dasmahapatra et al., 2012; Lavoie et al., 2013) and *Anolis* lizards (Feiner, 2016).

We did not detect any main difference in the TEs content or TEs transposition histories for the different clownfish species. Nevertheless, species-specific genomic features may be underestimated here as the assemblies were based on the *A. frenatus* reference genome (Marcionetti et al., 2019).

4.1.2 Accelerated coding evolution

We detected a significantly increased evolutionary rate in a subset of randomly sampled clownfish genes, suggesting a global accelerated coding evolution in the group. Accelerated evolution in genes involved in morphological and developmental processes was observed in Eastern African cichlids (Brawand et al., 2014) and potentially facilitated their rapid diversification (Brawand et al., 2014; Berner & Salzburger, 2015). Because accelerated evolution was widespread in clownfish, pinpointing specific biological processes with higher evolutionary rates and potentially involved in the diversification (as performed in Brawand et al., 2014) has not been possible.

Overall differences in evolutionary rates between lineages are widely observed (e.g., Ohta, 1993; Bromham et al., 1996; Bromham, 2002; Nabholz et al., 2008; Bromham, 2009; Welch et al., 2008; Thomas et al., 2010) and are primarily attributed to differences in generation-time, where organisms with shorter generation times likely evolve faster (i.e., the generation-time hypothesis, Bromham, 2009). However, the generation time in clownfishes (5 years estimated on *A. percula*, Buston & García, 2007) is higher than in the considered outgroup species (i.e., from a few months to 4 years, Supplementary Table S16), excluding it as the driver of the acceleration in coding evolution. The effective population also affects the evolutionary rate through the action of genetic drift (Lynch & Walsh, 2007), and lineages with smaller effective population sizes (N_e) show an increased rate of evolution (Woolfit & Bromham, 2003, 2005). Although N_e for the species in the study was not available, sequential hermaphroditism can potentially reduce it (Chopelet et al., 2009; 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 & Pauly, 2000). Thus, further studies investigating N_e are needed to evaluate its effect on clownfishes' observed accelerated coding evolution.

Antagonistically coevolving species are expected to show an increased rate of molecular evolution (the Red Queen Hypothesis, Van Valen, 1971; Pal et al., 2007; Paterson et al., 2010), while species in mutualistic relationships should show a decreased rate of evolution (the Red King Effect theory, Bergstrom & Lachmann, 2003). The results obtained for clownfish contradict this Red King Effect theory, as it does the increased evolutionary rate observed in obligate symbiotic organisms (Lutzoni & Pagel, 1997; Yoshizawa et al., 2003; Bromham et al., 2013) and mutualistic ants (Rubin & Moreau, 2016). The studies on the symbiotic organisms explain the increased rate with demographic history (e.g., N_e and generation time), which was excluded in the mutualistic ants (Rubin & Moreau, 2016). There, authors argue that in lowly-intimate mutualistic interactions, species must adapt to their own changing environments and those of their symbionts. Thus, mutualism may lead to selective pressure similar to those experienced by antagonistically coevolving species (Van Valen, 1971), resulting in an increased rate of evolution (Rubin & Moreau, 2016). This hypothesis could also hold for the 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.

4.1.3. Positive selection in clownfishes

We found that at least 5.4% of the genes were positively selected throughout clownfish diversification. This percentage could be underestimated, as simulations showed a low power in detecting weak selection and demonstrated the absence of false positives. Positive selection on ecologically important genes can drive the adaptation and diversification of organisms (e.g., Rundle & Nosil, 2005; Schluter & Conte, 2009), and patterns of positive selection were observed in the genomes of radiating lineages (e.g., Kapralov et al., 2013; Brawand et al., 2014; Cornetti et al., 2015; Nevado et al., 2016). While linking all positively selected genes with a potential role in clownfish diversification would be unreasonable, we identified five genes with particularly interesting functions that we discuss further.

First, we detected positive selection on the gene encoding the somatostatin 2 (SST2, OG14285, Table 1). Somatostatin is a diverse family of peptide hormones that influences organismal growth by inhibiting the production and release of the growth hormone (Very & Sheridan, 2002; Sheridan & Hagemester, 2010). The variation in somatic growth rate resulting from social status changes in *Astatotilapia burtoni* was associated with a shift in the volume of somatostatin-containing neurons (Hofmann & Fernald, 2000). Similarly, in *A. burtoni*, somatostatin 1 regulates aggressive behavior in dominant males (Trainor & Hofmann, 2006), and its expression in the hypothalamus is associated with the control of somatic growth depending on social status (Trainor & Hofmann, 2007). Secondly, we observed positive selection on the gene encoding the isotocin receptor (ITR, OG20045_2b, Table 1). In the cichlid *Neolamprologus pulcher*, isotocin plays an important role in modulating social behavior, increasing responsiveness to social information (Reddon et al., 2012, 2015, 2017). For instance, exogenous administration of isotocin resulted in increased submissive behavior when challenged aggressively (Reddon et al., 2012), and the size and number of isotocin neurons were significantly different between cooperatively and independently breeding species (Reddon et al., 2017). We also detected positive selection on the gene *NPF2R2* (OG16291, Table 1), encoding for the neuropeptide FF receptor 2, which plays a role in feeding-related processes (including appetite control, food intake, 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 non-breeders (if present) are progressively smaller as the hierarchy is descended (Fricke,

1979; Ochi, 1989). These well-defined size differences between individuals are maintained by the precise regulation of the subordinate growth (Buston, 2003), often through aggressive behavior (Iwata et al., 2008, 2010). It was also hypothesized that, in *A. percula*, subordinates reduce their food intake to avoid exceeding 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 on the *RHO* gene (OG8544, Table 1), encoding the rhodopsin. This protein is a photoreceptor required for image-forming vision at a low light intensity, and modifications in its gene sequence likely lead to 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; Sugawara et al., 2009). 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 & Allen, 1997). Although we did not detect explicit positively selected sites affecting rhodopsin absorbance, this gene may have played a role in clownfish adaptation to different depths.

Finally, we detected positive selection in the *duox* gene (OG8036, Table 1), encoding the dual oxidase protein involved in synthesizing thyroid hormones (Chopra et al., 2019), which regulates the white stripes formation in *A. percula* (Salis et al., 2021). They also observed that shifts in *duox* 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 functional role of 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 (Seehausen et al., 1999; Kelley et al., 2013). Positive selection on *duox* in clownfishes may thus be associated with divergence in the formation of white stripes in the group.

While these genes show interesting functions associated with clownfish social behavior and ecology, further investigation is needed to validate the link between the detected positively selected genes and their role in clownfish evolution and diversification. For instance, expression evidence for these genes is lacking (except for *duox*) and should be examined in the future.

4.1.4. Mosaic genomes in clownfishes

The reconstruction of clownfishes' mitochondrial and nuclear phylogenetic trees resulted in cytonuclear discordance. Nodes in both the mitochondrial and nuclear phylogenetic trees were strongly supported, suggesting that the inconsistency results from past hybridization events or incomplete lineage sorting (ILS). Cytonuclear incongruences in clownfish were previously detected and were associated with a burst in the diversification rate of the group (Litsios & Salamin, 2014). We also observed topological discordance throughout the nuclear genome, with phylogenetic trees inferred in the different genomic regions being well supported, confirming potential hybridization events and/or ILS in the clownfish diversification. The main topological inconsistencies were seen in the deep nodes of the trees, while the five pairs of closely related species mainly branched together. A first exception was detected for a few windows, in which the specialists *A. nigripes* and *A. sebae* branched as sister species, which we will further discuss below (Section 4.2). We observed a second exception for *P. biaculeatus*, which principally branched as sister species of *A. ocellaris* but was also frequently placed as basal to the *Amphiprion* clade (Figure 4C). This disparity mirrors the conflicting phylogenies reported in the literature (e.g., Frédérich et al., 2013, Mirande, 2017; na Ayudhaya et al., 2017, Lobato et al., 2014; DiBattista et al., 2016). While *Premnas* has been recently recovered within *Amphiprion* (Tang et al., 2021), these inconsistencies suggest a complex evolutionary history of *P. biaculeatus* - such as potential hybridization events with species outside the *Amphiprion* genus - that should be further investigated.

As the topological inconsistencies were observed in the deep nodes of the trees, hybridization events and/or ILS likely happened in ancestral clownfish species. Following the topology of the nuclear phylogenetic tree, the second most abundant topology was the one of cluster 1 (i.e., not the one reflecting the mitochondrial tree). This topology was predominant in chromosomes 1 and 3, suggesting the presence of mechanisms (such as structural inversions, decreased recombination, selection, and/or genetic drift) increasing its fixation (see also below). However, to identify such mechanisms and characterize the topological landscape of clownfish more accurately, the precise identification of topological breakpoints directly on the *A. percula* reference genome is necessary. Indeed, here we do not have the exact location of the inconsistencies, and we cannot discern whether they cluster in some regions or are scattered throughout the chromosomes.

We observed another particular topological inconsistency on chromosome 18, where the most frequent topology split the species pair of *A. perideraion* - *A. akallopisos* (topology of cluster 5, Figure 4C) and suggested past gene flow between *A. perideraion* and the *A. melanopus* - *A. frenatus* ancestor. This topology was almost exclusively observed on this chromosome and clustered in two large regions (see Chapter 4), potentially indicating that the introgression signal was removed from the rest of the genome by extensive backcrossing but persisted on chromosome 18 through the disruption of recombination. The recombination disruption was possibly achieved through the genomic inversions of the regions - originated in the ancestor of *A. melanopus* - *A. frenatus* and introgressed in *A. perideraion* or arisen in *A. perideraion* after the gene flow - that were fixed in *A. perideraion* through genetic drift or selection. Genomic inversions that break recombination, creating clusters of loci controlling ecologically important traits, consequently fixed by natural selection, are observed in the case of supergenes (e.g., Joron et al., 2011; Kunte et al., 2014; Zinzow-Kramer et al., 2015; Küpper et al., 2016; Branco et al., 2018). Here, however, we cannot establish a role of selection, as we cannot easily link the genes' functions in these regions to species' important ecological traits without further studies (Supplementary Table S5), and we did not observe an enrichment of positively selected genes on chromosome 18. Thus, we cannot exclude the role of genetic drift in fixing these regions in *A. perideraion*. It is worth mentioning that the evolution of sex chromosomes may also result in particular patterns as those observed on chromosome 18 (e.g., Natri et al., 2019). However, we do not believe that this is relevant in clownfishes, as these species are sequential hermaphrodites with no sex chromosomes (Fricke & Fricke, 1977; Moyer & Nakazono, 1978; Fricke, 1979; Arai, 2011), and genes involved in the sex change are scattered throughout the genome (Casas et al., 2018).

The patterns observed on chromosome 18, together with the additional topological inconsistencies, suggest ancestral hybridization events in the diversification of clownfishes. Gene flow spreading ancient genetic variation among species has been proposed to facilitate adaptive radiation (Berner & Salzburger, 2015; Marques et al., 2019). For instance, ancestral hybridization between distinct lineages has fueled the adaptive radiation of cichlids (i.e., the *hybrid swarm* hypothesis, Seehausen, 2004; Meier et al., 2017; Svardal et al., 2020), while introgressive hybridization among members of the radiating lineages (i.e., the *syngameon* hypothesis, Seehausen, 2004) has facilitated ecological speciation in *Heliconius* butterflies (Dasmahapatra et al., 2012; Pardo-Diaz et al., 2012) and Darwin's finches (Lamichhaney et al., 2015). While hybridization events may have also participated in clownfish diversification, we

cannot yet exclude that the mosaic genomes observed in clownfishes are, at least partially, the result of ILS. Further studies should focus on the formal testing of gene flow, for instance, with ABBA-BABA tests and related Patterson's *D* statistics (Green et al., 2010; Durand et al., 2011) as performed in Maier et al. (2017).

4.1.5. No increase in gene duplication rate at the basis of clownfish radiation

We observed the highest duplication rate in the common ancestor of the Pomacentridae and *O. niloticus* (branch D in Figure 7), comparable to the estimated duplication rate in the cichlids' ancestor (Figure 2 in Brawand et al., 2014). Contrarily, the duplication rate detected in clownfish was similar to the one observed in non-radiating teleosts (Figure 2 in Brawand et al., 2014). Thus, clownfish diversification does not seem to be characterized by an increase in gene duplication events, differently from what has been observed in cichlids adaptive radiation (Brawand et al., 2014; Machado et al., 2014). While the overall duplication rate may be underestimated in a phylogenetic duplication analysis approach (as performed here) compared to 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 & Conery, 2000; Taylor & Raes, 2004; 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.

4.2. Parallel evolution in clownfish diversification

In clownfish, we did not detect extensive topological inconsistencies potentially arising from hybridization events among specialist or generalist species (Figure 4C), indicating that parallel evolution through recent introgressive hybridization is not a major driver of clownfish

diversification. Nevertheless, we observed an exception in a few genomic windows, where the two specialists *A. nigripes* and *A. sebae* branched as sister species. The node of this split was fully supported in all windows. *A. sebae* is part of the primary radiation of clownfishes, which diversified in the Coral Triangle region (Litsios et al., 2014). After the diversification, this species experienced range expansion (Litsios et al., 2014), and its distribution now reaches the Northern Indian Ocean (including Sri Lanka and Maldives; Allen, 1991, Litsios et al., 2014). *A. nigripes* is endemic to the Maldives and Sri Lanka (Allen, 1991, Litsios et al., 2014) and belongs to the clownfish replicated radiation, which diversified independently in the Indian Ocean after a colonization event (Litsios et al., 2014). While there is no report of sympatric populations of the two species, their geographical distribution suggests that they co-occur in the Maldives and Sri Lanka, and gene flow between them is thus conceivable.

If introgression between the species occurred, extensive backcrossing likely removed its signal from the rest of the genome, and natural selection possibly fixed the regions showing topological inconsistency. Genes within these regions included olfactory receptors and the *Has2* gene, encoding for hyaluronan synthase 2, which catalyzes the addition of GlcNac to nascent hyaluronan polymers (Tian et al., 2013). Interestingly, it is well known that coral reef fish larvae use olfactory cues to identify suitable settlement sites (e.g., Lecchini et al., 2005a,b; Gerlach et al., 2007; Dixon et al., 2008). Additionally, genes with functions associated with GlcNac and hyaluronan were previously found positively selected at the basis of clownfish radiation, and they were hypothesized to play a role in the acquisition of mutualism by avoiding the release of sea anemones toxins. (Marcionetti et al., 2019). These genes might be involved in the ecological convergence of *A. nigripes* and *A. sandaracinos*, with introgressive hybridization playing a role in the parallel evolution of species, as observed in *Heliconius* butterflies (Dasmahapatra et al., 2012; Supple et al., 2013). Nevertheless, future studies that aim to better characterize the regions of topological inconsistency (using *A. percula* reference genome), better evaluate genes' functions and expression, and better define species' phenotypes are needed before drawing any conclusion.

While parallel evolution through recent introgressive hybridization does not seem frequent in clownfish diversification, parallel evolution could also result from differential rates of evolution or selective pressures in genes associated with ecological divergence (e.g., Cresko et al., 2004; Colosimo et al., 2005; Linnen et al., 2013; Projecto-Garcia et al., 2013). Regarding the evolutionary rate, we did not detect any gene evolving at different rates in all the species pairs. It is nevertheless worth noting that, when considering only some of the species pairs, we

observed genes with an increased evolutionary rate shared among specialists or generalists (Figure 9B-C). However, the number of shared genes rapidly decreased with the increasing number of species pairs considered. The decrease was similar in specialists and generalists and was independent of the phylogenetic relationship between species pairs. These observations suggest that the shared genes were likely obtained by random processes rather than being related to ecological divergence.

Contrary to the evolutionary rate, we observed genes undergoing similar selective pressures in all specialist or generalist clownfish species. Indeed, we detected 24 and 39 genes that were specifically positively selected in specialists and generalists, respectively. Because they show patterns of positive selection, these genes were potentially involved in the parallel adaptation to similar ecological niches, which resulted in the convergent morphological changes observed in clownfishes (Litsios et al., 2012). Genes with parallel patterns of relaxation or intensification of purifying selection in specialist or generalist species were also observed. 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, in general, 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 to 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 (Kawecki, 1994; Bono et al., 2020; Draghi, 2021). Additionally, by considering the rationale behind the increased evolutionary rate reported for mutualistic ants (Rubin & Moreau, 2016, see above), 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, see above).

5. CONCLUSIONS

Studies on well-described adaptive radiations have started investigating the intrinsic genomic factors that may promote the rapid diversification of lineages. Here, we examined the genomic features underlying clownfish diversification.

Like other adaptive radiations, our results showed that clownfish genomes show two bursts of transposable elements, overall accelerated coding evolution, and topology inconsistencies potentially resulting from hybridization events. These characteristics possibly promoted clownfish adaptive radiation. Additionally, we detected positively selected genes with interesting functions, which likely participated in the evolution of the size-based dominance hierarchy in the group. Finally, we observed genes that underwent differential selective pressures associated with ecological divergence, suggesting some extent of parallel evolution during the group's diversification.

This study provides cues for the further investigation of the genomics underlying clownfish adaptive radiation. For instance, along with the necessity to better characterize clownfish functional traits, future studies need to take advantage of the chromosomal-level assembly of *A. percula* and formally test for ancestral hybridization events in the clownfish diversification. Additionally, thanks to this new resource, other genomic features proposed to have facilitated rapid diversification - such as regulatory element changes and structural variants - should also be investigated. Estimating clownfishes' effective population sizes is also essential to understand the mechanisms underlying the observed acceleration in coding sequence evolution in the group. Nevertheless, our results provide the first genomic insights into clownfish adaptive radiation and integrate the growing collection of studies investigating the genomic mechanisms governing species diversification.

SUPPLEMENTARY MATERIAL

Given its considerable size, the supplementary material was not directly integrated into the manuscript, essentially to avoid excessive printing. The supplementary material is nevertheless easily accessible online. [Supplementary Information and Figures](#)¹ and [Supplementary Tables](#)² are available on OneDrive.

¹https://unils-my.sharepoint.com/:b/g/personal/anna_marcionetti_unil_ch/Eec8ZJQ7d4dGjD6fm62kjkMBm4YnFrq3_vxyvIli42uytA?e=IEgpe9

²https://unils-my.sharepoint.com/:x/g/personal/anna_marcionetti_unil_ch/EabKtF-V0GdAmT9MY1nluYYBPolpZON5CMujwkJyWG3OWw?e=fdcPOJ

REFERENCES

- 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.
- Abzhanov, A., Extavour, C. G., Groover, A., Hodges, S. A., Hoekstra, H. E., Kramer, E. M., & Monteiro, A.** (2008). Are we there yet? Tracking the development of new model systems. *Trends in Genetics*, *24*(7), 353-360.
- Alexa, A., & Rahnenführer, J.** (2016). topGO: Enrichment Analysis for Gene Ontology. R package version 2.26.0. *Bioconductor Improv*, *27*,1-26
- Allen, G. R.** (1991). Damsel-fishes of the world. Mergus Publishers, Melle, Germany.
- Altenhoff, A. M., Gil, M., Gonnet, G. H., & Dessimoz, C.** (2013). Inferring hierarchical orthologous groups from orthologous gene pairs. *PLoS One*, *8*(1).
- Anisimova, M., Bielawski, J. P., & Yang, Z.** (2002). Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular biology and evolution*, *19*(6), 950-958.
- Arai, R.** (2011). *Fish karyotypes: a check list*. Springer Science & Business Media.
- Arendt, J., & Reznick, D.** (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation?. *Trends in ecology & evolution*, *23*(1), 26-32.
- Auvinet, J., Graça, P., Belkadi, L., Petit, L., Bonnivard, E., Dettai, A., ... & Higuete, D.** (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), 1-18.
- Benvenuto, C., Coscia, I., Chopelet, J., Sala-Bozano, M., & Mariani, S.** (2017). Ecological and evolutionary consequences of alternative sex-change pathways in fish. *Scientific Reports*, *7*(1), 1-12.
- Bergstrom, C. T., & Lachmann, M.** (2003). The Red King effect: when the slowest runner wins the coevolutionary race. *Proceedings of the National Academy of Sciences*, *100*(2), 593-598.
- Berner, D., & Salzburger, W.** (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, *31*(9), 491-499.
- Bielawski, J. P., & Yang, Z.** (2004). A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *Journal of molecular evolution*, *59*(1), 121-132.
- Blackledge, T. A., & Gillespie, R. G.** (2004). Convergent evolution of behavior in an adaptive radiation of Hawaiian web-building spiders. *Proceedings of the National Academy of Sciences*, *101*(46), 16228-16233
- Blackledge, T. A., & Gillespie, R. G.** (2004). Convergent evolution of behavior in an adaptive radiation of Hawaiian web-building spiders. *Proceedings of the National Academy of Sciences*, *101*(46), 16228-16233.
- Bono, L. M., Draghi, J. A., & Turner, P. E.** (2020). Evolvability costs of niche expansion. *Trends in Genetics*, *36*(1), 14-23.
- Bouckaert, R. R.** (2010). DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*, *26*(10), 1372-1373.
- Bowmaker, J. K.** (2008). Evolution of vertebrate visual pigments. *Vision research*, *48*(20), 2022-2041.
- Brakefield, P. M.** (2006). Evo-devo and constraints on selection. *Trends in Ecology & Evolution*, *21*(7), 362-368.
- Branco, S., Carpentier, F., de la Vega, R. C. R., Badouin, H., Snirc, A., Le Prieur, S., ... & Giraud,**

- T. (2018). Multiple convergent supergene evolution events in mating-type chromosomes. *Nature communications*, 9(1), 1-13.
- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., ... & Turner-Maier, J.** (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, 513(7518), 375-381.
- Bromham, L.** (2002). Molecular clocks in reptiles: life history influences rate of molecular evolution. *Molecular biology and evolution*, 19(3), 302-309.
- Bromham, L.** (2009). Why do species vary in their rate of molecular evolution?. *Biology letters*, 5(3), 401-404.
- Bromham, L., Cowman, P. F., & Lanfear, R.** (2013). Parasitic plants have increased rates of molecular evolution across all three genomes. *BMC evolutionary biology*, 13(1), 1-11.
- Bromham, L., Rambaut, A., & Harvey, P. H.** (1996). Determinants of rate variation in mammalian DNA sequence evolution. *Journal of molecular evolution*, 43(6), 610-621.
- Buston, P.** (2003). Size and growth modification in clownfish. *Nature*, 424(6945), 145-146.
- Buston, P. M., & García, M. B.** (2007). An extraordinary life span estimate for the clown anemonefish *Amphiprion percula*. *Journal of Fish Biology*, 70(6), 1710-1719.
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & 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. *Scientific reports*, 8(1), 1-12.
- Chausson, J., Srinivasan, M., & Jones, G. P.** (2018). Host anemone size as a determinant of social group size and structure in the orange clownfish (*Amphiprion percula*). *PeerJ*, 6, e5841.
- Chopelet, J., Waples, R. S., & Mariani, S.** (2009). Sex change and the genetic structure of marine fish populations. *Fish and Fisheries*, 10(3), 329-343.
- Chopra, K., Ishibashi, S., & Amaya, E.** (2019). Zebrafish *duox* mutations provide a model for human congenital hypothyroidism. *Biology Open*, 8(2).
- Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., ... & Kingsley, D. M.** (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *science*, 307(5717), 1928-1933.
- Cornetti, L., Valente, L. M., Dunning, L. T., Quan, X., Black, R. A., Hébert, O., & Savolainen, V.** (2015). The genome of the “great speciator” provides insights into bird diversification. *Genome Biology and Evolution*, 7(9), 2680-2691.
- Coscia, I., Chopelet, J., Waples, R. S., Mann, B. Q., & Mariani, S.** (2016). Sex change and effective population size: implications for population genetic studies in marine fish. *Heredity*, 117(4), 251-258.
- Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., ... & Postlethwait, J. H.** (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, 101(16), 6050-6055.
- Dabney, A., Storey, J. D., & Warnes, G. R.** (2010). qvalue: Q-value estimation for false discovery rate control. R package version 1.38.0, <https://github.com/StoreyLab/qvalue>
- Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., ... & Heliconius Genome Consortium.** (2012). Butterfly genome reveals promiscuous exchange of mimicry

- adaptations among species. *Nature*, 487(7405), 94.
- DiBattista, J. D., Howard Choat, J., Gaither, M. R., Hobbs, J. P. A., Lozano-Cortés, D. F., Myers, R. F., ... & Berumen, M. L.** (2016). On the origin of endemic species in the Red Sea. *Journal of Biogeography*, 43(1), 13-30.
- Diekmann, Y., & Pereira-Leal, J. B.** (2015). Gene tree affects inference of sites under selection by the branch-site test of positive selection. *Evolutionary Bioinformatics*, 11, EBO-S30902.
- Dixon, D. L., Jones, G. P., Munday, P. L., Planes, S., Pratchett, M. S., Srinivasan, M., ... & Thorrold, S. R.** (2008). Coral reef fish smell leaves to find island homes. *Proceedings of the Royal Society B: Biological Sciences*, 275(1653), 2831-2839.
- Draghi, J. A.** (2021). Asymmetric evolvability leads to specialization without trade-offs. *The American Naturalist*, 197(6), 644-657.
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M.** (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, 28(8), 2239-2252.
- Dusa, A.** (2021). venn: Draw Venn Diagrams. R package version 1.10. <https://cran.r-project.org/web/packages/venn/index.html>
- Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., ... & Mallet, J.** (2019). Genomic architecture and introgression shape a butterfly radiation. *Science*, 366(6465), 594-599.
- Elmer, K. R., & Meyer, A.** (2011). Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends in ecology & evolution*, 26(6), 298-306.
- Faber-Hammond, J. J., Bezault, E., Lunt, D. H., Joyce, D. A., & Renn, S. C.** (2019). The genomic substrate for adaptive radiation: copy number variation across 12 tribes of African cichlid species. *Genome biology and evolution*, 11(10), 2856-2874.
- Fan, S., & Meyer, A.** (2014). Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. *Frontiers in genetics*, 5, 163.
- Fautin, D. G., & Allen, G. R.** (1997). Anemone fishes and their host sea anemones. Western Australian Museum, Perth, 159 pp.
- Fedoroff, N. V.** (2012). Transposable elements, epigenetics, and genome evolution. *Science*, 338(6108), 758-767.
- Feiner, N.** (2016). Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards. *Proceedings of the Royal Society B: Biological Sciences*, 283(1840), 20161555.
- Fletcher, W., & Yang, Z.** (2010). The effect of insertions, deletions, and alignment errors on the branch-site test of positive selection. *Molecular biology and evolution*, 27(10), 2257-2267.
- Forsberg, R., & Christiansen, F. B.** (2003). A codon-based model of host-specific selection in parasites, with an application to the influenza A virus. *Molecular Biology and Evolution*, 20(8), 1252-1259.
- Frédérich, B., Sorenson, L., Santini, F., Slater, G. J., & Alfaro, M. E.** (2013). Iterative ecological radiation and convergence during the evolutionary history of damselfishes (Pomacentridae). *The American Naturalist*, 181(1), 94-113.
- Fricke, H. W.** (1979). Mating system, resource defence and sex change in the anemonefish Amphiprion akallopisos. *Zeitschrift für Tierpsychologie*, 50(3), 313-326.
- Fricke, H., & Fricke, S.** (1977). Monogamy and sex change by aggressive dominance in coral reef fish. *Nature*, 266(5605), 830-832.

- Froese, R., & Pauly, D.** (2000). *FishBase 2000: concepts designs and data sources* (Vol. 1594). WorldFish.
- Gainsford, A., Van Herwerden, L., & Jones, G. P.** (2015). Hierarchical behaviour, habitat use and species size differences shape evolutionary outcomes of hybridization in a coral reef fish. *Journal of Evolutionary Biology*, 28(1), 205-222.
- Gao, B., Shen, D., Xue, S., Chen, C., Cui, H., & Song, C.** (2016). The contribution of transposable elements to size variations between four teleost genomes. *Mobile DNA*, 7(1), 4.
- Gerlach, G., Atema, J., Kingsford, M. J., Black, K. P., & Miller-Sims, V.** (2007). Smelling home can prevent dispersal of reef fish larvae. *Proceedings of the national academy of sciences*, 104(3), 858-863.
- Givnish T & Sytsma K.** (1997). *Molecular Evolution and Adaptive Radiation*. Cambridge Univ.Press, New York.
- Givnish, T. J.** (2015). Adaptive radiation versus ‘radiation’ and ‘explosive diversification’: why conceptual distinctions are fundamental to understanding evolution. *New Phytologist*, 207(2), 297-303.
- Glor, R. E.** (2010). Phylogenetic insights on adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, 41, 251-270.
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Hansen, N. F.** (2010). A draft sequence of the Neandertal genome. *Science*, 328(5979), 710-722.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O.** (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 59(3), 307-321.
- 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 research*, 41(13), e129-e129.
- Hofmann, H. A., & Fernald, R. D.** (2000). Social status controls somatostatin neuron size and growth. *Journal of Neuroscience*, 20(12), 4740-4744.
- Iwata, E., Nagai, Y., & Sasaki, H.** (2010). Social rank modulates brain arginine vasotocin immunoreactivity in false clown anemonefish (*Amphiprion ocellaris*). *Fish physiology and biochemistry*, 36(3), 337-345.
- Iwata, E., Nagai, Y., Hyoudou, M., & Sasaki, H.** (2008). Social environment and sex differentiation in the false clown anemonefish, *Amphiprion ocellaris*. *Zoological Science*, 25(2), 123-128.
- Jombart, T., Kendall, M., Almagro-Garcia, J., & Colijn, C.** (2017). treespace: Statistical exploration of landscapes of phylogenetic trees. *Molecular ecology resources*, 17(6), 1385-1392.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... & Kingsley, D. M.** (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55-61.
- Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., ... & Jiggins, C. D.** (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, 477(7363), 203-206.
- Kapralov, M. V., Votintseva, A. A., & Filatov, D. A.** (2013). Molecular adaptation during a rapid adaptive radiation. *Molecular biology and evolution*, 30(5), 1051-1059.
- Katoh, K., & Standley, D. M.** (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.
- Katoh, K., & Standley, D. M.** (2016). A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics*, 32(13), 1933-1942.

- Kawecki, T. J.** (1994). Accumulation of deleterious mutations and the evolutionary cost of being a generalist. *The American Naturalist*, 144(5), 833-838.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Drummond, A.** (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Kelley, J. L., Fitzpatrick, J. L., & Merilaita, S.** (2013). Spots and stripes: ecology and colour pattern evolution in butterflyfishes. *Proceedings of the Royal Society B: Biological Sciences*, 280(1757), 20122730.
- Kimura, M.** (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of molecular evolution*, 16(2), 111-120.
- Kingman, G. A. R., Vyas, D. N., Jones, F. C., Brady, S. D., Chen, H. I., Reid, K., ... & Veeramah, K. R.** (2021). Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, 7(25), eabg5285.
- Kingsley, E. P., Manceau, M., Wiley, C. D., & Hoekstra, H. E.** (2009). Melanism in *Peromyscus* is caused by independent mutations in *Agouti*. *PloS one*, 4(7), e6435.
- Kocher, T. D., Conroy, J. A., McKaye, K. R., & Stauffer, J. R.** (1993). Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi are due to convergence. *Molecular phylogenetics and evolution*, 2(2), 158-165.
- Kondrashov, F. A.** (2012). Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proceedings of the Royal Society B: Biological Sciences*, 279(1749), 5048-5057.
- Kozak, K. M., Joron, M., McMillan, W. O., & Jiggins, C. D.** (2021). Rampant genome-wide admixture across the *Heliconius* radiation. *Genome Biology and Evolution*.
- Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., ... & Kronforst, M. R.** (2014). Doublesex is a mimicry supergene. *Nature*, 507(7491), 229-232.
- Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., ... & Burke, T.** (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature genetics*, 48(1), 79-83.
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., ... & Andersson, L.** (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371-375.
- Lavoie, C. A., Platt, R. N., Novick, P. A., Counterman, B. A., & Ray, D. A.** (2013). Transposable element evolution in *Heliconius* suggests genome diversity within Lepidoptera. *Mobile DNA*, 4(1), 1-10.
- Lecchini, D., Planes, S., & Galzin, R.** (2005a). Experimental assessment of sensory modalities of coral-reef fish larvae in the recognition of their settlement habitat. *Behavioral Ecology and Sociobiology*, 58(1), 18-26.
- Lecchini, D., Shima, J., Banaigs, B., & Galzin, R.** (2005b). Larval sensory abilities and mechanisms of habitat selection of a coral reef fish during settlement. *Oecologia*, 143(2), 326-334.
- Lehmann, R., Lightfoot, D. J., Schunter, C., Michell, C. T., 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*. *Molecular ecology resources*, 19(3), 570-585.
- Lewis, J. J., Geltman, R. C., Pollak, P. C., Rondem, K. E., Van Belleghem, S. M., Hubisz, M. J., ... & Reed, R. D.** (2019). Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry.

- Proceedings of the National Academy of Sciences*, 116(48), 24174-24183.
- Li, Q., Wen, H., Li, Y., Zhang, Z., Zhou, Y., & Qi, X.** (2019). Evidence for the direct effect of the NPFF peptide on the expression of feeding-related factors in spotted sea bass (*Lateolabrax maculatus*). *Frontiers in endocrinology*, 10, 545.
- Linnen, C. R., Poh, Y. P., Peterson, B. K., Barrett, R. D., Larson, J. G., Jensen, J. D., & Hoekstra, H. E.** (2013). Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science*, 339(6125), 1312-1316.
- Litsios, G., & Salamin, N.** (2014). Hybridisation and diversification in the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, 14(1), 245.
- Litsios, G., Pearman, P. B., Lanterbecq, D., Tolou, N., & Salamin, N.** (2014). The radiation of the clownfishes has two geographical replicates. *Journal of biogeography*, 41(11), 2140-2149.
- Litsios, G., Sims, C. A., Wüest, R. O., Pearman, P. B., Zimmermann, N. E., & Salamin, N.** (2012). Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC evolutionary biology*, 12(1), 212.
- Liu, Y., Cotton, J. A., Shen, B., Han, X., Rossiter, S. J., & Zhang, S.** (2010). Convergent sequence evolution between echolocating bats and dolphins. *Current Biology*, 20(2), R53-R54.
- Lobato, F. L., Barneche, D. R., Siqueira, A. C., Liedke, A. M., Lindner, A., Pie, M. R., ... & Floeter, S. R.** (2014). Diet and diversification in the evolution of coral reef fishes. *PLoS One*, 9(7), e102094.
- Lönnig, W. E., & Saedler, H.** (2002). Chromosome rearrangements and transposable elements. *Annual review of genetics*, 36(1), 389-410.
- Losos J. B.** (2009). Lizards in an evolutionary tree: ecology and adaptive radiation of anoles. Berkeley (CA): University of California Press.
- Losos, J. B.** (2011). Convergence, adaptation, and constraint. *Evolution: International Journal of Organic Evolution*, 65(7), 1827-1840.
- Lutzoni, F., & Pagel, M.** (1997). Accelerated evolution as a consequence of transitions to mutualism. *Proceedings of the National Academy of Sciences*, 94(21), 11422-11427.
- Lynch, M., & Conery, J. S.** (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.
- Machado, H. E., Jui, G., Joyce, D. A., Reilly, C. R., Lunt, D. H., & Renn, S. C.** (2014). Gene duplication in an African cichlid adaptive radiation. *BMC genomics*, 15(1), 1-13.
- Malinsky, M., Svardal, H., Tyers, A. M., Miska, E. A., Genner, M. J., Turner, G. F., & Durbin, R.** (2018). Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nature ecology & evolution*, 2(12), 1940-1955.
- Marcionetti, A., Rossier, V., Bertrand, J. A., Litsios, G., & Salamin, N.** (2018). First draft genome of an iconic clownfish species (*Amphiprion frenatus*). *Molecular ecology resources*, 18(5), 1092-1101.
- Marcionetti, A., Rossier, V., Roux, N., Salis, P., Laudet, V., & Salamin, N.** (2019). Insights into the genomics of clownfish adaptive radiation: Genetic basis of the mutualism with sea anemones. *Genome biology and evolution*, 11(3), 869-882.
- Marques, D. A., Meier, J. I., & Seehausen, O.** (2019). A combinatorial view on speciation and adaptive radiation.

Trends in ecology & evolution, 34(6), 531-544.

- Martin, C. H., & Richards, E. J.** (2019). The paradox behind the pattern of rapid adaptive radiation: how can the speciation process sustain itself through an early burst?. *Annual Review of Ecology, Evolution, and Systematics*, 50, 569-593.
- McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., ... & Seehausen, O.** (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, 586(7827), 75-79.
- Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O.** (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications*, 8(1), 1-11.
- Mirande, J. M.** (2017). Combined phylogeny of ray-finned fishes (Actinopterygii) and the use of morphological characters in large-scale analyses. *Cladistics*, 33(4), 333-350.
- Moretti, S., Laurency, B., Gharib, W. H., Castella, B., Kuzniar, A., Schabauer, H., ... & Robinson-Rechavi, M.** (2014). Selectome update: quality control and computational improvements to a database of positive selection. *Nucleic acids research*, 42(D1), D917-D921.
- Moyer, J. T., & Nakazono, A.** (1978). Protandrous hermaphroditism in six species of the anemonefish genus *Amphiprion* in Japan. *Japanese Journal of Ichthyology*, 25(2), 101-106.
- Moyer, J. T., & Nakazono, A.** (1978). Protandrous hermaphroditism in six species of the anemonefish genus *Amphiprion* in Japan. *Japanese Journal of Ichthyology*, 25(2), 101-106.
- Muschick, M., Indermaur, A., & Salzburger, W.** (2012). Convergent evolution within an adaptive radiation of cichlid fishes. *Current biology*, 22(24), 2362-2368.
- na Ayudhaya, P. T., Muangmai, N., Banjongsat, N., Singchat, W., Janekitkarn, S., Peyachoknagul, S., & Srikulnath, K.** (2017). Unveiling cryptic diversity of the anemonefish genera *Amphiprion* and *Premnas* (Perciformes: Pomacentridae) in Thailand with mitochondrial DNA barcodes. *Agriculture and Natural Resources*, 51(3), 198-205.
- Nabholz, B., Glémin, S., & Galtier, N.** (2008). Strong variations of mitochondrial mutation rate across mammals—the longevity hypothesis. *Molecular biology and evolution*, 25(1), 120-130.
- Natri, H. M., Merilä, J., & Shikano, T.** (2019). The evolution of sex determination associated with a chromosomal inversion. *Nature communications*, 10(1), 1-13.
- Nevado, B., Atchison, G. W., Hughes, C. E., & Filatov, D. A.** (2016). Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. *Nature Communications*, 7(1), 1-9.
- Ochi, H.** (1989). Mating behavior and sex change of the anemonefish, *Amphiprion clarkii*, in the temperate waters of southern Japan. *Environmental Biology of Fishes*, 26(4), 257-275.
- Ogden, T. H., & Rosenberg, M. S.** (2006). Multiple sequence alignment accuracy and phylogenetic inference. *Systematic biology*, 55(2), 314-328.
- Ohta, T.** (1993). An examination of the generation-time effect on molecular evolution. *Proceedings of the National Academy of Sciences*, 90(22), 10676-10680.
- Ollerton, J., McCollin, D., Fautin, D. G., & Allen, G. R.** (2007). Finding NEMO: nestedness engendered by mutualistic organization in anemonefish and their hosts. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609), 591-598.
- Pal, C., Maciá, M. D., Oliver, A., Schachar, I., & Buckling, A.** (2007). Coevolution with viruses drives the evolution of bacterial mutation rates. *Nature*, 450(7172), 1079-1081.

- Paradis, E., & Schliep, K.** (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526-528.
- Pardo-Diaz, C., Salazar, C., Baxter, S.W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W.O. & Jiggins, C.D.** (2012) Adaptive Introgression across Species Boundaries in Heliconius Butterflies. *PLOS Genetics* 8, e1002752.
- Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., ... & Brockhurst, M. A.** (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, 464(7286), 275-278.
- Planet, P. J.** (2006). Tree disagreement: measuring and testing incongruence in phylogenies. *Journal of biomedical informatics*, 39(1), 86-102.
- Projecto-Garcia, J., Natarajan, C., Moriyama, H., Weber, R. E., Fago, A., Cheviron, Z. A., ... & Storz, J. F.** (2013). Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proceedings of the National Academy of Sciences*, 110(51), 20669-20674.
- Protas, M. E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W. R., ... & Tabin, C. J.** (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature genetics*, 38(1), 107-111.
- Rebollo, R., Horard, B., Hubert, B., & Vieira, C.** (2010). Jumping genes and epigenetics: towards new species. *Gene*, 454(1-2), 1-7.
- Reddon, A. R., O'Connor, C. M., Marsh-Rollo, S. E., & Balshine, S.** (2012). Effects of isotocin on social responses in a cooperatively breeding fish. *Animal Behaviour*, 84(4), 753-760.
- Reddon, A. R., O'Connor, C. M., Marsh-Rollo, S. E., Balshine, S., Gozdowska, M., & Kulczykowska, E.** (2015). Brain nonapeptide levels are related to social status and affiliative behaviour in a cooperatively breeding cichlid fish. *Royal Society Open Science*, 2(2), 140072.
- Reddon, A. R., O'Connor, C. M., Nesjan, E., Cameron, J., Hellmann, J. K., Ligocki, I. Y., ... & Balshine, S.** (2017). Isotocin neuronal phenotypes differ among social systems in cichlid fishes. *Royal Society open science*, 4(5), 170350.
- Revell, L. J.** (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in ecology and evolution*, 3(2), 217-223.
- Roberts, M. B. V.** (1986). *Biology: a functional approach*. Nelson Thornes. p.274
- Rosenblum, E. B., Parent, C. E., & Brandt, E. E.** (2014). The molecular basis of phenotypic convergence. *Annual Review of Ecology, Evolution, and Systematics*, 45, 203-226.
- Rosenblum, E. B., Römler, H., Schöneberg, T., & Hoekstra, H. E.** (2010). Molecular and functional basis of phenotypic convergence in white lizards at White Sands. *Proceedings of the National Academy of Sciences*, 107(5), 2113-2117.
- Rubin, B. E., & Moreau, C. S.** (2016). Comparative genomics reveals convergent rates of evolution in ant-plant mutualisms. *Nature Communications*, 7(1), 1-11.
- Rundle, H. D., & Nosil, P.** (2005). Ecological speciation. *Ecology letters*, 8(3), 336-352.
- Rundle, H. D., Nagel, L., Boughman, J. W., & Schluter, D.** (2000). Natural selection and parallel speciation in sympatric sticklebacks. *Science*, 287(5451), 306-308.
- Sackton, T. B., & Clark, N.** (2019). Convergent evolution in the genomics era: new insights and directions. *Philos. Trans. R. Soc. B: Biol. Sci*, 374

- Salis, P., Roux, N., Huang, D., Marcionetti, A., Mouginot, P., Reynaud, M., ... & Laudet, V.** (2021). Thyroid hormones regulate the formation and environmental plasticity of white bars in clownfishes. *Proceedings of the National Academy of Sciences*, 118(23).
- Schliep, K. P.** (2011). phangorn: phylogenetic analysis in R. *Bioinformatics* 27, 592e593.
- Schluter, D.** (2000). *The ecology of adaptive radiation*. OUP Oxford.
- Schluter, D., & Conte, G. L.** (2009). Genetics and ecological speciation. *Proceedings of the National Academy of Sciences*, 106(Supplement 1), 9955-9962.
- Schluter, D., & Nagel, L. M.** (1995). Parallel speciation by natural selection. *The American Naturalist*, 146(2), 292-301.
- Schott, R. K., Refvik, S. P., Hauser, F. E., López-Fernández, H., & Chang, B. S.** (2014). Divergent positive selection in rhodopsin from lake and riverine cichlid fishes. *Molecular biology and evolution*, 31(5), 1149-1165.
- Seehausen, O.** (2004). Hybridization and adaptive radiation. *Trends in ecology & evolution*, 19(4), 198-207.
- Seehausen, O., Mayhew, P.J., Van Alphen, J.J.M.** (1999). Evolution of colour patterns in East African cichlid fish. *Journal of evolutionary biology*, 12(3), 514-534.
- Shao, F., Han, M., & Peng, Z.** (2019). Evolution and diversity of transposable elements in fish genomes. *Scientific reports*, 9(1), 1-8.
- 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.
- Sheridan, M. A., & Hagemester, A. L.** (2010). Somatostatin and somatostatin receptors in fish growth. *General and comparative endocrinology*, 167(3), 360-365.
- Shimodaira, H., & Hasegawa, M.** (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular biology and evolution*, 16(8), 1114-1114.
- Simpson, G. G.** (1953). *The major features of evolution*. Columbia University Press.
- Smit, A., Hubley, R., & Green, P.** (2013-2015). RepeatMasker Open-4.0. 2013. Institute of Systems Biology. <http://www.repeatmasker.org/>
- Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L.** (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular biology and evolution*, 32(5), 1342-1353.
- Soulebeau, A., Aubriot, X., Gaudeul, M., Rouhan, G., Hennequin, S., Haevermans, T., ... & Jabbour, F.** (2015). The hypothesis of adaptive radiation in evolutionary biology: hard facts about a hazy concept. *Organisms Diversity & Evolution*, 15(4), 747-761.
- Spady, T. C., Seehausen, O., Loew, E. R., Jordan, R. C., Kocher, T. D., & Carleton, K. L.** (2005). Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. *Molecular Biology and Evolution*, 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.
- Stern, D. L.** (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics*, 14(11), 751-764.
- Stroud, J. T., & Losos, J. B.** (2016). Ecological opportunity and adaptive radiation. *Annual Review of Ecology, Evolution, and Systematics*, 47.

- Sugawara, T., Imai, H., Nikaido, M., Imamoto, Y., & Okada, N.** (2009). Vertebrate rhodopsin adaptation to dim light via rapid meta-II intermediate formation. *Molecular biology and evolution*, 27(3), 506-519.
- Supple, M. A., Hines, H. M., Dasmahapatra, K. K., Lewis, J. J., Nielsen, D. M., Lavoie, C., ... & Counterman, B. A.** (2013). Genomic architecture of adaptive color pattern divergence and convergence in *Heliconius* butterflies. *Genome research*, 23(8), 1248-1257.
- Suyama, M., Torrents, D., & Bork, P.** (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic acids research*, 34(suppl_2), W609-W612.
- Svardal, H., Quah, F. X., Malinsky, M., Ngatungmaiera, B. P., Miska, E. A., Salzburger, W., ... & Durbin, R.** (2020). Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Molecular biology and evolution*, 37(4), 1100-1113.
- Syvanen, M.** (1984). The evolutionary implications of mobile genetic elements. *Annual review of genetics*, 18(1), 271-293.
- Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., & Dessimoz, C.** (2015). Current methods for automated filtering of multiple sequence alignments frequently worsen single-gene phylogenetic inference. *Systematic biology*, 64(5), 778-791.
- Tang, K. L., Stiassny, M. L., Mayden, R. L., & DeSalle, R.** (2021). Systematics of Damselfishes. *Ichthyology & Herpetology*, 109(1), 258-318.
- Taylor, J. S., & Raes, J.** (2004). Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.*, 38, 615-643.
- Thomas, J. A., Welch, J. J., Lanfear, R., & Bromham, L.** (2010). A generation time effect on the rate of molecular evolution in invertebrates. *Molecular biology and evolution*, 27(5), 1173-1180.
- Tian, X., Azpurua, J., Hine, C., Vaidya, A., Myakishev-Rempel, M., Ablueva, J., ... & Seluanov, A.** (2013). High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. *Nature*, 499(7458), 346-349.
- Trainor, B. C., & Hofmann, H. A.** (2006). Somatostatin regulates aggressive behavior in an African cichlid fish. *Endocrinology*, 147(11), 5119-5125.
- Trainor, B. C., & Hofmann, H. A.** (2007). Somatostatin and somatostatin receptor gene expression in dominant and subordinate males of an African cichlid fish. *Behavioural brain research*, 179(2), 314-320.
- UniProt Consortium.** (2018). UniProt: the universal protein knowledgebase. *Nucleic acids research*, 46(5), 2699.
- Van Valen, L.** (1973) A new evolutionary law. *Evol. Theory* 1, 1–30 .
- Vater, M., & Kössl, M.** (2004). The ears of whales and bats. *Echolocation in bats and dolphins*, 89-99.
- Verta, J. P., & Jones, F. C.** (2019). Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. *Elife*, 8, e43785.
- Very, N. M., & Sheridan, M. A.** (2002). The role of somatostatins in the regulation of growth in fish. *Fish Physiology and Biochemistry*, 27(3-4), 217-226
- Vizueta, J., Macías-Hernández, N., Arnedo, M. A., Rozas, J., & Sánchez-Gracia, A.** (2019). Chance and predictability in evolution: the genomic basis of convergent dietary specializations in an adaptive radiation. *Molecular ecology*, 28(17), 4028-4045.
- Vizueta, J., Macías-Hernández, N., Arnedo, M. A., Rozas, J., & Sánchez-Gracia, A.** (2019). Chance and predictability in evolution: the genomic basis of convergent dietary specializations in an adaptive radiation.

Molecular ecology, 28(17), 4028-4045.

- Waples, R. S., Mariani, S., & Benvenuto, C.** (2018). Consequences of sex change for effective population size. *Proceedings of the Royal Society B*, 285(1893), 20181702.
- Weadick, C. J., & Chang, B. S.** (2012). An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Molecular biology and evolution*, 29(5), 1297-1300.
- Welch, J. J., Bininda-Emonds, O. R., & Bromham, L.** (2008). Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC evolutionary biology*, 8(1), 1-12.
- Wickham, H.** (2016). *ggplot2: elegant graphics for data analysis*. Springer.
- Woolfit, M., & Bromham, L.** (2003). Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. *Molecular Biology and Evolution*, 20(9), 1545-1555.
- Woolfit, M., & Bromham, L.** (2005). Population size and molecular evolution on islands. *Proceedings of the Royal Society B: Biological Sciences*, 272(1578), 2277-2282.
- Xiong, P., Hulsey, C. D., Fruciano, C., Wong, W. Y., Nater, A., Kautt, A. F., ... & Franchini, P.** (2021). The comparative genomic landscape of adaptive radiation in crater lake cichlid fishes. *Molecular ecology*, 30(4), 955-972.
- Yang, Z.** (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, 24(8), 1586-1591.
- Yang, Z., & Dos Reis, M.** (2010). Statistical properties of the branch-site test of positive selection. *Molecular biology and evolution*, 28(3), 1217-1228.
- Yoder, J. B., Clancey, E., Des Roches, S., Eastman, J. M., Gentry, L., Godsoe, W., ... & Sarver, B. A. J.** (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of evolutionary biology*, 23(8), 1581-1596.
- Yoshizawa, K., & Johnson, K. P.** (2003). Phylogenetic position of Phthiraptera (Insecta: Paraneoptera) and elevated rate of evolution in mitochondrial 12S and 16S rDNA. *Molecular phylogenetics and evolution*, 29(1), 102-114.
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. Y.** (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, 8(1), 28-36.
- Zhen, Y., Aardema, M. L., Medina, E. M., Schumer, M., & Andolfatto, P.** (2012). Parallel molecular evolution in an herbivore community. *science*, 337(6102), 1634-1637.
- Zinzow-Kramer, W. M., Horton, B. M., McKee, C. D., Michaud, J. M., Tharp, G. K., Thomas, J. W., ... & Maney, D. L.** (2015). Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes, Brain and Behavior*, 14(8), 641-654.

Chapter 4

**Genomics of the diversification of the clownfish
skunk complex (*Amphiprion akallopisos*,
A. sandaracinos and *A. perideraion*)**

Genomics of the diversification of the clownfish skunk complex **(*Amphiprion akallopisos*, *A. sandaracinos* and *A. perideraion*)**

ABSTRACT

Clownfish are an iconic group of coral reef fish maintaining mutualism with sea anemones, which triggered the rapid diversification of the group. Within clownfishes, the skunk complex is particularly interesting as, besides ecological speciation, hybridization events were suggested to have frequently occurred throughout the divergence of the group. Hence, we aimed here at investigating the mechanisms underlying the diversification of the complex.

By taking advantage of their disjunct geographical distribution, we obtained whole-genome data of sympatric and allopatric populations of the three main species of the complex. We examined population structure, genomic divergence patterns, and introgression signals, and we performed demographic modeling to identify the more realistic diversification scenario. We excluded scenarios of strict isolation, of hybrid origin of *A. sandaracinos*, and ruled out the presence of extensive gene flow in sympatry. However, we discovered moderate gene flow from *A. perideraion* and the ancestor of *A. akallopisos*-*A. sandaracinos*, and weak gene flow throughout the group's diversification between the species in the Indo-Australian Archipelago. We identified introgressed regions in *A. sandaracinos*, and we detected two large regions of high divergence in *A. perideraion*, likely maintained by the disruption of recombination and suggesting additional hybridization events with species not considered here.

Altogether, these results show that hybridization events in the skunk complex are less pervasive than initially thought and suggest a role of sea anemone host repartition in maintaining the genetic identity of the species in sympatry. The study provides the first insights on the mechanisms underlying the group's diversification, bringing us a step closer to the understanding of clownfish adaptive radiation.

1. INTRODUCTION

How organisms adapt and diversify, how new species emerge, and how biodiversity is built up have been fundamental questions in evolutionary biology for over 150 years (Losos et al., 2013). Answering these questions is not straightforward as speciation is a complex process involving different and interplaying mechanisms, such as geographic isolation, genetic drift, natural selection, and sexual selection (Kirkpatrick & Ravigné, 2002; Gavrilets, 2003 & 2004; Coyne, 2004, van Doorn et al., 2009; Gavrilets & Losos, 2009; Schluter, 2009; Nosil et al., 2009a). One mechanism that has been widely explored in recent years is ecological speciation, defined as the evolution of reproductive isolation between populations as a result of ecologically-driven divergent natural selection (Doebeli & Dieckmann, 2003; Rundle & Nosil, 2005; Funk et al., 2006; Nosil et al., 2009a; Schluter, 2000, 2001 & 2009; Thibert-Plante & Hendry, 2011). Under this process, divergent natural selection can act in response to differences in environments (such as habitat structure, climate, and resources) or as a result of ecological interactions (such as predation or competition; Doebeli & Dieckmann, 2003; Schluter, 2000, 2001, 2009; Rundle & Nosil, 2005; Thibert-Plante & Hendry, 2011). The study of ecological speciation led to shifts of emphasis in the literature, away from stochastic changes in isolated populations and towards adaptive changes (Beheregaray et al., 2015).

While studies investigating ecological speciation were initially based on manipulative experiments (Rice & Hostert, 1993), evidence for ecological speciation in natural populations has also emerged (reviewed in Schluter, 2001), with widely cited examples including sticklebacks (Schluter, 1996), Darwin's finches (Grant, 1986) and *Heliconius* butterflies (Mallet et al., 1998). The advent of next-generation sequencing techniques further boosted the field of research by providing access to genomic data of non-model organisms, which extended the study of the role of ecology in adaptation and speciation across various evolutionary stages, environments, and taxa (e.g., Moritz et al., 2000; Shafer & Wolf, 2013; Ellegren & Sheldon, 2008; Seehausen et al., 2014).

The analysis of the genomes of closely related organisms identified loci involved in the adaptation to different ecological niches, which promoted population divergence, and eventually, speciation (e.g., Lawniczak et al., 2010; Michel et al., 2010; Jones et al., 2012a, 2012b; Nadeau et al., 2012; Andrew & Rieseberg, 2013; Gagnaire et al., 2013; Keller et al., 2013; Martin et al., 2013; Parchman et al., 2013; Soria-Carrasco et al., 2014; Malinsky et al., 2015). Similarly, hybridization - traditionally saw as a process limiting diversification (Mayr, 1963) - was also seen to promote ecological adaptation and speciation, further questioning the classical view of natural selection

acting on either new mutations or on standing genetic variation present within populations (Hermisson & Pennings, 2005; Barret & Schluter, 2008; Hedrick, 2013; Abbott et al., 2016; Lai et al., 2019). Introgressive hybridization of adaptive alleles was demonstrated in several adaptive radiations, such as in *Heliconius* butterflies (e.g., Dasmahapatra et al., 2012; Pardo-Diaz et al., 2012; Nadeau et al., 2012, 2013), Darwin's finches (Lamichhaney et al., 2015) and cichlid fish (Salzburger et al., 2002; Keller et al., 2013). Ancient hybridization has been shown to fuel the adaptive radiation of cichlid fishes (Meier et al., 2017; Svardal et al., 2020). Additionally, hybrid speciation, which involves the formation of novel genetic combinations and novel adaptations allowing the persistence of the hybrid lineage, was observed as a mechanism involved in species evolution (Gompert et al., 2006; Mallet, 2007; Abbott et al., 2010; Keller et al., 2013).

Genomic studies of multiple lineages along the speciation continuum (as defined in Seehausen et al., 2014) contributed considerably to our knowledge of the genetics of speciation (Nosil et al., 2009b; Strasburg et al., 2012; Seehausen et al., 2014). However, they revealed that besides ecological adaptation and gene flow, other processes such as incomplete lineage sorting, variation in recombination rate, linked selection, and demographic history also shaped the genomes of diverging lineages (Nachman, 2002; Noor & Bennet, 2009; Turner & Han, 2010; Cruickshank & Hahn, 2014; Pennisi, 2014; Cutter & Payseur, 2013; Burri, 2017; Rettelbach et al., 2019; Wang et al., 2019). As a result, further comparative genomic studies of lineages at different stages of divergence are necessary to disentangle these processes and to understand the interplay between genomic properties, ecology, geography, and demographic history in shaping species diversification.

Here, we aimed to study the mechanisms underlying the diversification of clownfishes (genera *Amphiprion* and *Premnas*, family Pomacentridae), an iconic group of coral reef fishes maintaining a mutualistic interaction with sea anemones, by focusing on the skunk complex. This group (also named as *akallopisos* group) is composed of three main species, namely *A. akallopisos*, *A. sandaracinos* and *A. perideraion* (Fautin & Allen., 1997), and a fourth one (*A. pacificus*) endemic to Fiji, Tonga, Samoa, and Wallis Island (Allen et al., 2010, not considered here). The complex diversified in the Indo-Australian Archipelago (IAA, or Coral Triangle region) ca. 1 to 5 MYA (Hubert et al., 2012; Cowman et al., 2011 & 2013; Litsios et al., 2012a,b; Frédérick et al., 2013; Litsios et al., 2014; Rabosky et al., 2013 & 2018). The members of this group differ from other clownfishes by their coloration, their more elongated shape and the presence of a white stripe along the dorsal ridge (Figure 1A; Fautin & Allen, 1997). Within the group, the species only show slight differences in color patterns. *A. akallopisos* and *A. perideraion* have a white caudal fin and

an orange pinkish body color, whereas the caudal fin and the body of *A. sandaracinos* are orange (Figure 1A; Fautin & Allen, 1997). *A. sandaracinos* has a slightly longer white stripe than the two other species, while *A. perideraion* has an additional white bar on the head (Figure 1A; Fautin & Allen, 1997). *A. akallopisos* and *A. perideraion* have similarly shaped teeth (incisiform), which differ in *A. sandaracinos* (conical), potentially indicating slightly different ecological adaptations (Fautin & Allen, 1997; Timm et al., 2008).

Three main characteristics make this group a particularly suitable system to study the process of diversification. First, hybridization events potentially shaped the diversification of the group (Litsios & Salamin, 2014). Indeed, cytonuclear inconsistency was detected in the complex, with *A. sandaracinos* being sister to *A. akallopisos* at the nuclear level but diverging from *A. perideraion* in the mitochondrial genome (Litsios & Salamin, 2014). This observation was further supported by studies carried out using different markers and showing contrasting species' phylogenetic relationships and divergence times (e.g., Hubert et al., 2012; Cowman et al., 2011 & 2013; Litsios et al., 2012a & 2012b; Frédérich et al., 2013; Litsios et al., 2014; Rabosky et al., 2013 & 2018). The cytonuclear incongruence suggests either a hybrid origin of *A. sandaracinos* or the presence of hybridization events during its diversification (Litsios & Salamin, 2014). Additionally, gene flow between all members of the complex may still be occurring in the IAA. Indeed, low interspecific barcode variation observed for the species in this region was suggested to result from hybridization within the complex rather than incomplete lineage (Steinke et al., 2009). This hypothesis was further supported by hybridization events being common in the aquarium trade (Pushparaj, 2010; He et al., 2019).

Second, the geographical distribution of the complex is vast, but the three species co-occur only in the IAA (Figure 1C; Allen, 1991), resulting in the presence of sympatric populations in Indonesia and allopatric populations, for instance, in the Western Indian Ocean for *A. akallopisos* and in the New Caledonian archipelago for *A. perideraion* (Allen, 1991, Figure 1C). The disjunct geographical distribution offers an ideal framework to test for differential hybridization events at different timescales.

Third, the species of the complex show divergence in the sea anemone host usage. *A. perideraion* is the most generalist species of the group and can interact with four sea anemone species (Figure 1A-B; Fautin & Allen, 1997; Litsios et al., 2012a), while *A. sandaracinos* and *A. akallopisos* can live in two different hosts (Figure 1A-B; Fautin & Allen, 1997; Litsios et al., 2012a). Each pair of the three species shares a common sea anemone host (Figure 1A-B).

Ecological divergence in host usage (in terms of the number of possible mutualistic sea anemone hosts) was suggested to be the driver of the ecological speciation and adaptive radiation in clownfishes, especially for the species with overlapping distributions centered in the IAA (Litsios et al., 2012a).

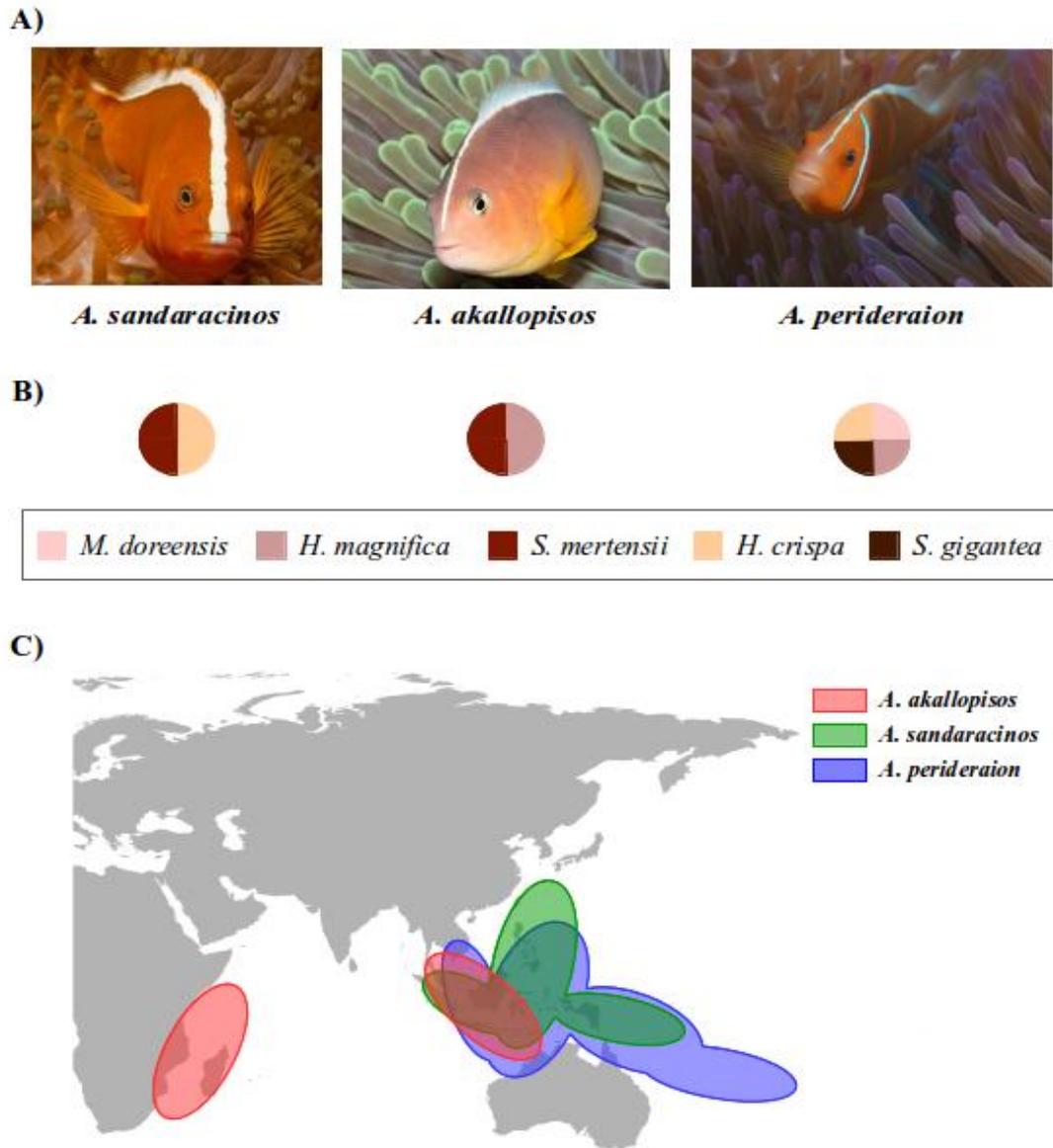


Figure 1. The three species of the skunk complex considered in this study (A), their host sea anemones (B) and their geographical distribution (C). B) The genera of the sea anemone species are *Macroactyla*, *Heteractis* and *Stichodactyla*. C) Geographical distributions were obtained from Allen (1991) and GBIF.org. See also Supplementary Figure S1.

Given the potential for ancient and recent hybridization events in the skunk complex and the ecological divergence in host usage potentially involved in their speciation, we investigated here the mechanisms underlying the diversification of this group of clownfishes. Their disjunct

geographical distribution offered a setup to differentiate between potential ancestral hybridization - which would be observed in sympatric and allopatric populations - and more recent gene flow - which would only be present in the IAA. We thus performed whole-genome sequencing of sympatric and allopatric populations of the three species. We examined the presence of ancestral gene flow between the species and examined the origin of *A. sandaracinos*. In the scenario of ancestral gene flow, the presence of divergent selection between habitat and host may have facilitated the speciation process. We also inspected the extent of recent gene flow in the IAA, as hybridization in sympatry is conceivable and may be facilitated by the three species sharing a common sea anemone host by pairs (Figure 1B). In this scenario of recent gene flow, hybridization would decrease the overall genomic divergence between the species in sympatry but not in allopatry, potentially revealing peaks of higher differentiation linked to the adaptation to different ecological niches (such as in Nadeau et al., 2012; Clarkson et al., 2014; Talla et al., 2017). Altogether, the results of this study will bring a better understanding of the role of hybridization (ancestral and/or recent gene flow) and ecological divergence in the diversification of the skunk complex, and more in general, in the radiation of clownfishes.

2. MATERIAL AND METHODS

2.1. Species selection and sampling

We obtained the overall geographic distribution of the three species selected for the study (*A. akallopisos*, *A. perideraion*, and *A. sandaracinos*) from Allen (1991) and GBIF.org (GBIF Occurrence Download <https://doi.org/10.15468/dl.rzpk14>; Fig. 1C, Supplementary Figure S1). The fourth species of the complex, the recently described *A. pacificus* (Allen et al., 2010), was not considered in the study as it is not co-occurring with the three other species. For each species, we sampled individuals from the IAA region, where the three species co-occur. For *A. perideraion* and *A. akallopisos*, we also included populations at the margin of the respective geographical distributions. This resulted in the sampling of one population of *A. sandaracinos* (Indonesia), two populations of *A. perideraion* (Indonesia and New Caledonia), and three populations of *A. akallopisos* (Indonesia, Mayotte, and Kenya), with five to six individuals per population (Supplementary Table S1). The protocol employed for all the sampling was similar, with fish caught in their host anemone using nets while SCUBA-diving. A piece of the caudal fin was taken and preserved in 96% ethanol immediately after the dive, in agreement with the permits delivered by the responsible institutions of the different countries.

As an outgroup was necessary for subsequent analyses, we included one *A. percula* individual from Lizard Island. A fin clip of the individual was provided by Fabio Cortesi (Queensland Brain Institute, University of Queensland, Australia).

2.2. DNA extraction, library preparation and DNA sequencing

Genomic DNA (gDNA) was extracted from the sampled fin tissue using DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. We prepared short-insert (350 bp) paired-end (PE) libraries from 100 ng of gDNA using TruSeq Nano DNA LT Library Preparation Kit (Illumina) following the manufacturer's instructions.

All the libraries of *A. akallopisos* and *A. perideraion* from New Caledonia were sequenced on Illumina HiSeq2000 at the Lausanne Genomic Technologies Facility, multiplexing five individuals per lane (read length of 100 bp). The libraries of *A. sandaracinos*, *A. perideraion* from Indonesia, and the *A. percula* individual were sequenced on Illumina HiSeq4000 at the Genomics Platform iGE3 (University of Geneva, Geneva), with a multiplex level of 8 individuals per lane (read length of 100 bp). The sequencing strategy led to a raw coverage of ca. 10X per individual.

2.3. Raw reads processing and mapping

We removed potential contamination by adapter from the raw reads with Cutadapt (v.1.13; Martin, 2011). We then trimmed the reads with Sickle (v.1.33; Joshi & Fass, 2011), with the following parameters: --qual-threshold 20, --length-threshold 40. We verified the quality of the reads with FASTQC (v.0.11.5; Andrews, 2010).

Two different reference genomes were available for the *Amphiprion* group: *A. percula* (Lehmann et al., 2018) and *A. frenatus* (Marcionetti et al., 2018). While the first has the advantage of a chromosome-level resolution, the second has a more recent divergence time with the skunk complex (4.96 for *A. frenatus* vs. 14.94 MYA for *A. percula*; from timetree.org; Kumar et al., 2017), which could affect the mapping quality. We mapped the reads using BWA on both references (v.0.7.15; Li & Durbin, 2009). We used the mapping statistics generated with bamtools (i.e., number of mapped reads, “proper pairs” and singletons; command *stats*, v.2.4.1; Barnett et al., 2011) and Picard tools (i.e., insert size distribution; command *CollectInsertSizeMetrics* v.2.4.1; <http://broadinstitute.github.io/picard>) to select the best reference genome for further analysis. The mapping statistics were similar between the two references (Supplementary Table S2), and we only considered the chromosome-level reference of *A. percula* for further analysis.

To decrease potential SNP calling errors, we filtered the mapped reads with samtools (command *view*, parameters: `-f 2 -F 256 -q 30`; v.1.3; Li et al., 2009) to keep only primary alignments and “proper pairs” (i.e., at the expected insert size and orientation) with a mapping quality higher than 30. We used *mergeReads* from ATLAS (v.0.9; Link et al., 2017) to remove potential redundant sequencing data originating from the overlap of paired reads. We verified the absence of soft-clipped positions with *assessSoftClipping* from ATLAS (v.0.9; Link et al., 2017).

2.4. SNP calling and filtering

For SNP calling, we employed the pipeline of ATLAS (v.0.9; Link et al., 2017) as it performs well at low-to-medium coverage, and it maintains a high accuracy in variant calling for moderately divergent species (Duchen & Salamin, 2020). For each sample (the 31 individuals of the skunk complex and one *A. percula* individual), we calculated the genotype likelihoods (GL) at each position of the reference genome using the GLF task with windows of 0.1 Mb. Window sizes of 0.1 Mb are sufficient to conduct GL inference with an average coverage of at least 4 (Kousathanas et al., 2016) while decreasing the computational effort. The GL files of all samples were used in the majorMinor step (MLE method) to infer major and minor alleles for each position and retrieve the genotype information of all individuals.

We used vcfTools (parameters: `--minQ 40 --minDP 2 --max-missing 0.9 --max-meanDP 40 --maxDP 40 --maf 0.02`; v.0.1.15; Danecek et al., 2011) to filter out invariant or low-confidence positions (variant quality score < 40), positions with coverage higher than 40, with a minor allele frequency (MAF) lower than 0.02, and with more than 10% of missing data. The MAF threshold removed positions where the minor allele occurred in a single individual (i.e., singletons), which can confound model-based inference of population structure (Linck & Battey, 2019).

We created a second dataset containing only samples of the skunk complex for analyses that did not require an outgroup. We removed the *A. percula* sample from the SNPs dataset, and we used vcfTools (v.0.1.15; Danecek et al., 2011) to filter out the resulting monomorphic positions and singletons. SNPs density was computed for different window sizes (100kb, 150kb, 200kb and 250kb) with vcfTools (v.0.1.15; Danecek et al., 2011). Results for different window sizes were consistent. Plotting was performed with the R package ggplot2 (v.3.0.0; Wickham, 2016).

Because of the presence of the outgroup in the variant calling process, SNPs specific to the skunk complex could have been lacking, for instance, in the case of triallelic positions. We

repeated the SNP calling as described above, but without the *A. percula* individual, to check that the number of potentially missed SNPs was low. The results were consistent, with only 326,456 additional SNPs detected (2% of the total number of SNPs).

2.5. Mitochondrial genome assembly

We used MITObim (v.1.9; Hahn et al., 2013) to reconstruct the entire mitochondrial genome of each sample of the skunk complex using two different reconstruction methods. The first one exploited previously published mitochondrial genomes as a reference, while the second used a barcode sequence to initiate the mitochondrial genome assembly. The sequences GB KJ833753 (complete mitochondrial genome, Li et al., 2015) and GB FJ582806 (COI gene, Steinke et al., 2009) were used for the mitochondrial genome reconstruction of *A. perideraion* samples. The sequences GB JF434730 (complete mitochondrial genome, Hubert et al., 2012) and JF434730 (COI gene, Hubert et al., 2012) were used for *A. akallopisos* and *A. perideraion* samples.

We confirmed the consistency of the two reconstruction methods with Geneious (v.10.2.2; Kearse et al., 2012), and we manually inferred the circularity of the sequence. The processed reads were mapped against the resulting mitochondrial genomes to check for the reconstruction success and assess coverage using Geneious (v.10.2.2; Kearse et al., 2012).

2.6. Mitochondrial and nuclear phylogenetic reconstruction

We downloaded the mitochondrial genome of one *A. percula* individual from NCBI (GB NC_023966.1; Tao et al., 2016), and we aligned it with the reconstructed mitochondrial genomes of the skunk complex using MAFFT (default parameters; v.7.450; Katoh & Standley, 2013). We visually checked the alignment to remove poorly aligned regions, and we reconstructed the mitochondrial phylogenetic tree with RaxML (v.8.2.12; parameters: -n 100 -m GTRGAMMA -p 12345 -x 12345 -f a; Stamatakis, 2014), performing 100 bootstrap replicates.

We reconstructed the nuclear phylogenetic trees using the final set of SNPs containing the *A. percula* individual. We generated the SNPs alignments using the script parseVCF (https://github.com/simonhmartin/genomics_general). We inferred phylogenetic trees for each chromosome separately using RAxML (v.8.2.12; parameters: -n 100 -m GTRGAMMA -p 12345 -x 12345 -f a -asc-corr lewis; Stamatakis, 2014), with the ASC parameter to correct for the lack of invariant sites and performing 100 bootstrap replicates. We computed the final nuclear phylogenetic tree from the trees obtained for each chromosome using ASTRAL-III (v.5.6.3; default parameters; Zhang et al., 2018). We plotted the mitochondrial and nuclear phylogenetic

trees with the *cophylo* command of the R package phytools (v.0.6.44; Revell, 2012). We used the *A. percula* individual to root the two phylogenetic trees, and then we removed it from the plot.

2.7. Principal Component Analysis and Admixture analysis

To investigate the overall structure of populations and species, we converted the SNPs dataset to the BEAGLE format with VCFTToBeagle from ATLAS (v.0.9; Link et al., 2017), and we performed PCA and admixture analyses with PCAngsd (v.0.98; Meisner & Albrechtsen, 2018). The PCA analysis was performed on the whole dataset and on each chromosome separately to check for inconsistencies, but no difference was detected. In the admixture analysis, we estimated the individual ancestry coefficients considering from two to five ancestral populations (K). We did this by setting the parameters -e to the corresponding K-1 value, as recommended in the PCAngsd documentation. For each K, we ran the analysis multiple times with different seeds to ensure convergence, and we retained the solution with the highest likelihood. Automatic selection of the best number of ancestral populations based on the MAP test in PCAngsd resulted in K=3. We investigated potential inconsistencies between genotype likelihoods and genotype calls by performing PCA and admixture analyses with a second approach based on genotype calls, and similar results were obtained (see Supplementary Information S1 for details).

2.8. Estimation of population genomic metrics

We investigated the overall genomic differentiation between the populations and species by calculating the average F_{st} for each pair of populations, using vcftools (v.0.1.15; Danecek et al., 2011) on the SNPs dataset of the skunk complex (see Supplementary Information S2 for details).

We investigated the patterns of genomic diversity and divergence along the genome by calculating nucleotide diversity (π) as well as relative (F_{st}) and absolute (d_{xy}) genetic divergence across sliding windows. We used the SNPs dataset of the skunk complex and applied the popgenWindows.py script (https://github.com/simonhmartin/genomics_general) to calculate F_{st} as described in Hudson et al. (1992, equation 9), as well as π and d_{xy} as the average proportion of differences between all pairs of sequences, either within a population (π) or between two populations (d_{xy}). To account for the variability in SNPs density along the genome (Supplementary Figure S2), we selected the window sizes based on the number of variable sites (parameter: --windType sites). We performed the analysis with windows of 500, 1,000, 1,500, 2,000, 2,500, 3,000, 4,000 and 5,000 variable sites, approximately corresponding to an average window size of respectively 25, 50, 80, 100, 130, 150, 200 and 250 kb. For each window size, we calculated average π for each population and averages d_{xy} and F_{st} for each pairwise comparison. The results

for different windows were consistent (Supplementary Table S3), and we further considered only windows of 3,000 variable sites. We plotted the measures of π , F_{st} , and d_{xy} along the genome using the function *manhattan* from the qqman R package (v.0.1.4; Turner, 2014).

We defined regions of increased or decreased divergence (or nucleotide diversity) as the windows in the top or lower 1% of the F_{st} , d_{xy} , and π distributions.

2.9. Topology discordance in the nuclear genome

We investigated the presence of alternative topologies in the nuclear genome using *twisst* (Martin & Van Belleghem, 2017), which assesses the weighting - measured as the frequency of occurrence - of each possible topology describing the relationship between the different groups in a dataset (such as populations or species). We ran *twisst* on the SNPs dataset with the *A. percula* individual, which was used to root the topologies. Phylogenetic trees were reconstructed for non-overlapping sliding windows with PhyML (v.3.3.2; Guindon et al., 2010) using *phyml_sliding_windows.py* (https://github.com/simonhmartin/genomics_general). We used windows that were identical to those in the estimation of population genomic metrics, and we calculated the exact weightings for all possible subtrees (method “complete”) using 10,000 sampling iterations. The individuals were grouped by species (species-level analysis) or populations (population-level analysis; see Supplementary Information S3). Plots were produced in R with the script *plot_twisst.R* provided in *twisst*.

2.10. Tests for ancient admixture

The cytonuclear incongruence and the topological inconsistencies along the genome showed that *A. sandaracinos* is occasionally phylogenetically closer to *A. perideraion* than to *A. akallopisos*. This disparity might result from incomplete lineage sorting (ILS) or from past hybridization events between *A. perideraion* and *A. sandaracinos*. Thus, we tested for signals of introgression between these two species. For this, we applied ABBA-BABA tests (Green et al. 2010), computing the Patterson's *D* statistic for a genome-wide estimation of admixture (Green et al., 2010; Durand et al., 2011; see below), and the f_d statistic for the estimation of admixture proportions at specific loci (Martin et al., 2015; see below). We investigated the presence of additional signals of admixture (besides between *A. perideraion* and *A. sandaracinos*) with TreeMix, and with 3- and 4-populations tests (v.1.13.; Pickrell & Pritchard, 2012; see Supplementary Information S4 for details).

2.10.1. Genome-wide signal of admixture

For the genome-wide signal of admixture, we estimated allele frequencies at each site and for each population with the script *freq.py* (https://github.com/simonhmartin/genomics_general). We considered all samples of the same species a single population. We used *A. percula* as the outgroup species, and we set *A. sandaracinos* and *A. perideraion* as the potentially hybridizing populations (i.e., P2 and P3, respectively; see Supplementary Information S5 for details on the methods). We calculated Patterson's *D* statistic with the equation (2) reported in Durand et al. (2011; see Supplementary Information S5). We applied a block jackknife procedure to test the significance of the *D* statistic as implemented in the script *jackknife.R* (https://github.com/simonhmartin/genomics_general; Supplementary Information S5). To ensure independent blocks, we set a block size of 1Mb, resulting in 903 blocks.

We tested whether the introgression signal depended on the geographical origin of *A. perideraion* by setting P3 either as the Indonesian or the New Caledonian population, and we repeated the analyses described above. We also investigated whether all chromosomes showed consistent patterns of introgression by applying the procedure described above, but estimating the *D*-statistics, the standard error, and the Z-score for each chromosome independently.

We estimated the proportion of admixture with the *f* statistics (Durand et al., 2011), which compares the observed excess of ABBA over BABA sites to the expected excess under complete admixture. For this, we approximated the expected excess under complete admixture by setting the New Caledonian and Indonesian populations as P2 and P3, respectively. We then estimated ABBA and BABA proportions as described above. We calculated the *f* statistic as the ratio of the differences in ABBA and BABA frequencies between the original scenario (i.e., with *A. sandaracinos* as P2) and this "complete admixture" one. We obtained the standard error and 95% confidence interval of *f* by applying a block jackknife approach, as described above (see also Supplementary Information S5). To ensure that the geographic origin of *A. perideraion* populations did not bias the "complete admixture" scenario, we also estimated *f* by randomly distributing *A. perideraion* individuals in P2 and P3.

2.10.2. Detection of potentially introgressed regions

We identified candidate genomic regions of introgression (CRI) by estimating the *f_d* statistics (Martin et al., 2015) with *ABBABABAWindows.py* (available in https://github.com/simonhmartin/genomics_general), setting *A. sandaracinos* and *A. perideraion* as the potentially hybridizing populations (i.e., P2 and P3, respectively; Supplementary

Information S5). We applied the same window sizes described above (see *Section 2.8*) but removed windows containing less than 100 biallelic SNPs to avoid stochastic errors in the estimation of f_d (Martin et al., 2015). We defined CRI as the windows in the top 5% of the genome-wide f_d distribution. We used this threshold based on the estimate of the genome-wide admixture proportion between *A. sandaracinos* and *A. perideraion* (5.5% of admixed genome). We visually verified that the regions of high f_d also supported the alternative mitochondrial topology (see *Section 2.9*).

Introgressed regions generally show lower absolute genetic divergence (d_{xy} ; Smith & Kronforst, 2013; Martin et al., 2015). Thus, we used Welch two-sample t-tests (in R) to test for significant differences in d_{xy} between the CRI and the rest of the genome for each combination of species. We removed d_{xy} values for the regions on chromosome 18, which behave differently than the other chromosomes, potentially biasing the tests.

2.11. Gene content of the CRI and regions of increased/decreased divergence

We downloaded the structural gene annotation for 23,718 protein-coding genes of *A. percula* from the Ensembl database (release 99; <https://www.ensembl.org>). Functional annotation was available for 15,875 protein-coding genes, with 9,453 associated with *biological process* gene ontologies (GOs). We expanded the functional annotation of *A. percula* protein-coding genes based on the available annotation of *A. frenatus* (Marcionetti et al., 2018). The extended annotation resulted in 17,179 annotated genes, with 14,002 annotated with *biological process* GOs. For details on the methods, see Supplementary Information S6.

We retrieved the functional annotation of the genes located within (or partially overlapping with) the CRI (see *Section 2.10*) and the regions of increased/decreased genomic divergence between species (see *Section 2.8*). We performed GO enrichment analysis for the two types of regions separately by contrasting the annotations of these genes against those of all the annotated protein-coding genes of *A. percula*. We used the topGO package (v.2.26.0; Alexa & Rahnenfuhrer, 2016) available in Bioconductor (<http://www.bioconductor.org>), setting a minimum node size of 3. Fisher's exact tests with the weight01 algorithms were applied to examine the significance of enrichment, with p-values < 0.01 considered significant. We present here *raw p-values* instead of *p-values* corrected for multiple testing, following recommendations from the topGO manual.

2.12. Tests of hybrid speciation, gene flow and demographic reconstruction

To infer the demographic scenario that best fitted the genomic data, we performed model comparison under the likelihood framework developed in fastsimcoal2 (v.2.6; Excoffier et al., 2013). The goal was to confirm the presence of ancestral admixture between the species, better characterize it, and further investigate the presence of more recent gene flow between the species in the IAA.

We retrieved the SNPs dataset without the *A. percula* sample and further removed positions with missing data using vcftools (--max-missing 1; v.0.1.15; Danecek et al., 2011). This resulted in a total of 16,547,283 SNPs (100,713 removed positions). We considered the two populations of *A. akallopisos* from the Western Indian Ocean as a single population. We extracted the multidimensional folded site frequency spectra (SFS) for each population using easySFS (<https://github.com/isaacovercast/easySFS>). We selected the projection that maximized the number of segregating sites per population using the --preview option of easySFS. This corresponded to 10 samples per population.

The multidimensional SFS were used to compare a total of 16 distinct demographic models (Figure 2). We first compared a model of strict isolation (Figure 2A) with models where *A. sandaracinos* originated from the hybridization of *A. akallopisos* and *A. perideraion* (i.e., hybrid speciation) followed by strict isolation (Figure 2B) or by asymmetric gene flow between *A. akallopisos* and *A. sandaracinos* (Figure 2C). We further compared these models with scenarios of ancestral asymmetric gene flow between the *A. akallopisos*-*A. sandaracinos* ancestor and *A. perideraion* and/or between *A. sandaracinos* and *A. perideraion* (Figure 2D-F). We then built models to investigate whether more recent gene flow between the three species was likely, either alone (Figure 2G, 2H, 2I) or with ancestral gene flow (Figure 2J, 2K, 2L). In these scenarios, we modeled gene flow throughout the divergence of different pairs of Indonesian populations (*A. sandaracinos* - *A. perideraion*, Figure 2G, 2J; *A. sandaracinos* - *A. perideraion* and *A. akallopisos* - *A. sandaracinos*, Figure 2H, 2L; *A. sandaracinos* - *A. perideraion*, *A. akallopisos* - *A. sandaracinos* and *A. akallopisos* - *A. perideraion*, Figure 2I, 2L). Finally, we explored the timing of the recent gene flow by comparing models with gene flow only in populations from the IAA region (i.e., after the split of the allopatric populations of *A. akallopisos* and *A. perideraion* (Figure 2N, 2P), and models where gene flow was only possible before the split of the allopatric populations (Figure 2M, 2O).

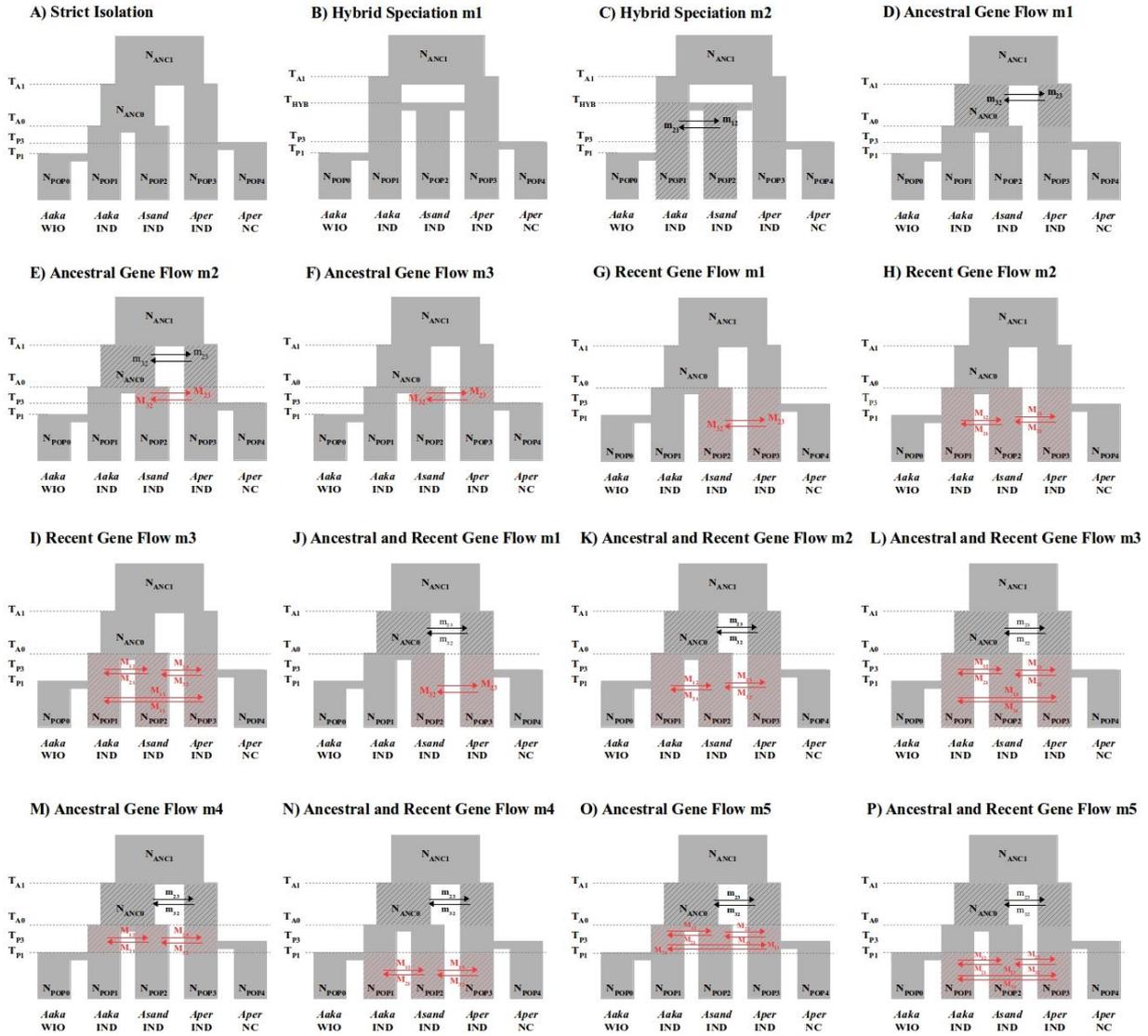


Figure 2: Summary of the 16 demographic models and their parameters evaluated using whole-genome SNPs. *Aaka*, *Asand*, and *Aper* correspond respectively to *A. akallopisos*, *A. sandaracinos*, and *A. perideraion*. WIO, IND, and NC correspond to the geographical origin of the population (respectively Western Indian Ocean, Indonesia, and New Caledonia). Gene flow between *A. perideraion* and the ancestor of *A. akallopisos* and *A. sandaracinos* is represented by black arrows. Red arrows correspond to more recent migrations (i.e., after the split between *A. akallopisos* and *A. sandaracinos*). We define “recent gene flow” when it is modeled to also occur in the present time. The exchanging populations were highlighted using hatching.

In all models, effective population sizes were estimated for each population and could vary at each splitting time. Parameters were estimated from the multidimensional SFS using fastsimcoal2 (v.2.6; Excoffier et al., 2013). We ran each model 50 times, performing 30 cycles of the expectation-conditional maximization (ECM) algorithm and considering 200,000 simulations to calculate the composite likelihood. We assumed a generation time of 5 years (as estimated for

A. percula, Buston & Garcia, 2007) and set a mutation rate of 4.0×10^{-8} (obtained from the average expected mutation rate per site per year for nuclear genomes of reef fishes and assuming a generation time of clownfishes of 5 years; Delrieu-Trottin et al., 2017). For each model, we retained the set of parameters with the highest final likelihood as the best point estimate. We compared the expected and observed SFS with the SFStools.R (available from <https://github.com/marqueda/SFS-scripts>).

We identified the best-fitting demographic model based on the rescaled Akaike's information criterion (AIC; Akaike, 1974). AIC values should, however, be interpreted with caution as linked sites are present in our data (Excoffier et al., 2013). We thus also examined the likelihood distributions obtained based on 100 expected SFS, each approximated using 1 million coalescent simulations under the parameters that maximized the likelihood for each model. An overlap of these distributions between models indicates no significant difference between the fit of alternative models.

For the best-supported model, we calculated confidence intervals of parameter estimates from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating parameters each time (Excoffier et al., 2013; Lanier et al., 2015; Ortego & Sork, 2018). For each bootstrap replicate, we performed 30 independent runs with 200,000 simulations and 20 ECM cycles. We used the parameter point estimates from the run with the highest likelihood of each bootstrapping replicate to compute the 95 percentile confidence intervals. We calculated points estimates of effective migration rates with $2 * N * m$ (with N the haploid population sizes and m the migration to the given population), obtaining the number of gene copies exchanged each generation (as in Bourgeois et al., 2020).

3. RESULTS

3.1. Whole Genome Sequencing, Mapping and SNP calling

The number of obtained raw paired-end reads (PEs) for each individual ranged from ca. 29 million pairs (sample MY085) to ca. 61 million pairs (sample FH4030), with a median of ca. 48 million PEs per sample (Supplementary Table S4). After trimming low-quality regions, the number of PEs per sample ranged from 27 to 57 million, corresponding to an estimated raw coverage between 5.8X and 12.0 X depending on the individual (Supplementary Table S4).

The mapping of the reads on the *A. percula* reference (Lehmann et al., 2018) resulted in between 84% and 88% of the reads mapping properly (with the pair at the expected insert size; Supplementary Table S2). We further filtered the reads to keep only those aligning with high confidence, and we removed potentially redundant sequencing data arising from the overlap of paired reads, obtaining a final coverage ranging between 4.6 X and 10.2 X (Supplementary Table S5).

The SNP calling strategy gave a total of 16,647,996 SNPs for the samples of the skunk complex, with an average SNPs density of 18.7 variants/kb. We observed regions of high and low SNPs density across all chromosomes (Supplementary Figure S2). When also considering the *A. percula* individual, the number of SNPs increased to 29,793,603 (see Supplementary Table S6 for additional information).

3.2. Cytonuclear discordance

We explored the phylogenetic relationship between the samples by reconstructing mitochondrial and nuclear phylogenetic trees. We observed that both trees showed a clear separation of the three species, followed by a separation of populations depending on their geographical origin (Figure 3). Only the samples of *A. akallopisos* from the two populations of the WIO were not differentiated at both the nuclear and mitochondrial levels.

Despite a clear separation of the species, the two genomes showed discordant topologies (Figure 3). At the nuclear level, *A. akallopisos* was sister to *A. sandaracinos*, while the latter was closer to *A. perideraion* at the mitochondrial level. This cytonuclear discordance was well supported, with bootstrap values higher than 0.95. We observed an additional inconsistency for one individual of *A. sandaracinos* (GB227), which grouped with its conspecific samples at the nuclear level but clustered within *A. perideraion* from Indonesia in the mitochondrial tree (bootstrap support > 0.95; Figure 3).

3.3. Population structure and admixture

The overall population structure of the skunk complex is shown in a PCA (Figure 4A). The first two axes explained 59.4% and 15.3% of the variance, respectively, and they split the three species into distinct clusters (Figure 4A, top). The first axis separated the *A. sandaracinos* and *A. akallopisos* individuals from *A. perideraion*, while the second divided *A. akallopisos* samples from *A. sandaracinos*. The third and fourth components explained 3.8% and 2.3% of the variance, respectively, and they separated the samples of the populations based on their geographical origin

(Figure 4A, bottom). The third axis split the Indonesian population of *A. perideraion* from the New Caledonian one, while the fourth axis divided the Indonesian population of *A. akallopisos* from the two populations in the Western Indian Ocean (WIO; Kenya and Mayotte). The admixture analysis also resulted in an overall separation of the samples by species and geography (Figure 4B). Indeed, depending on the number of ancestral populations *K* considered, individuals clustered with their respective species (for *K*=2 and *K*=3), followed by geographical populations (for *K*=4 and *K*=5; Figure 4B). The best number of ancestral populations was inferred to be *K*=3. We obtained consistent results with the method based on genotype-calls (Supplementary Information S1).

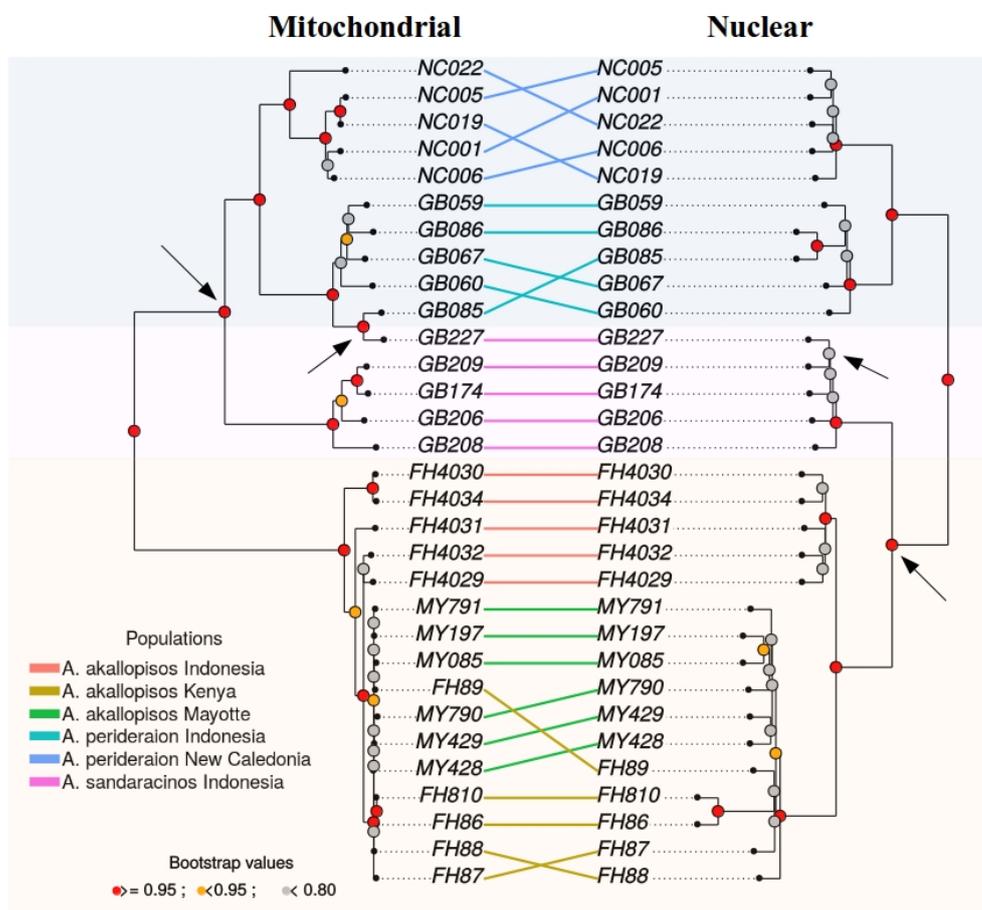


Figure 3: Mitochondrial and nuclear phylogenetic trees of *A. akallopisos*, *A. sandaracinos*, and *A. perideraion* samples. The mitochondrial phylogenetic tree was obtained with RAxML from the whole mitochondrial genome alignments (total of 16,747 bp). The nuclear phylogenetic tree was obtained with RAxML and ASTRAL-III from 29,793,603 SNPs. For both trees, 100 bootstrap replicates were performed. Red, orange and gray dots indicate bootstrap supports of respectively > 0.95 , between 0.8 and 0.95, and < 0.8 . Links between samples are drawn to highlight the topological differences. The different colors of the links correspond to the different populations. Arrows pinpoint the major cytonuclear inconsistencies. The phylogenetic trees were rooted with the *A. percula* outgroup, which was removed from the plot.

In the PCA, the populations of *A. akallopis* from Mayotte and Kenya formed a single group, while Indonesian individuals of different species were never clustered together in the PCA (Figure 4A), suggesting the absence of recent gene flow between the species. However, we observed some signals of shared ancestry between species and populations in the admixture plot (Figure 4B). For instance, with K=3 and K=4, samples of *A. akallopis* from Indonesia showed a low proportion (6%) of shared ancestry with *A. sandaracinos*. At K=2, while most of the ancestry of *A. sandaracinos* was shared with *A. akallopis*, a low proportion (13%) was in common with *A. perideraion* (Figure 5), suggesting potential ancestral gene flow between the two species.

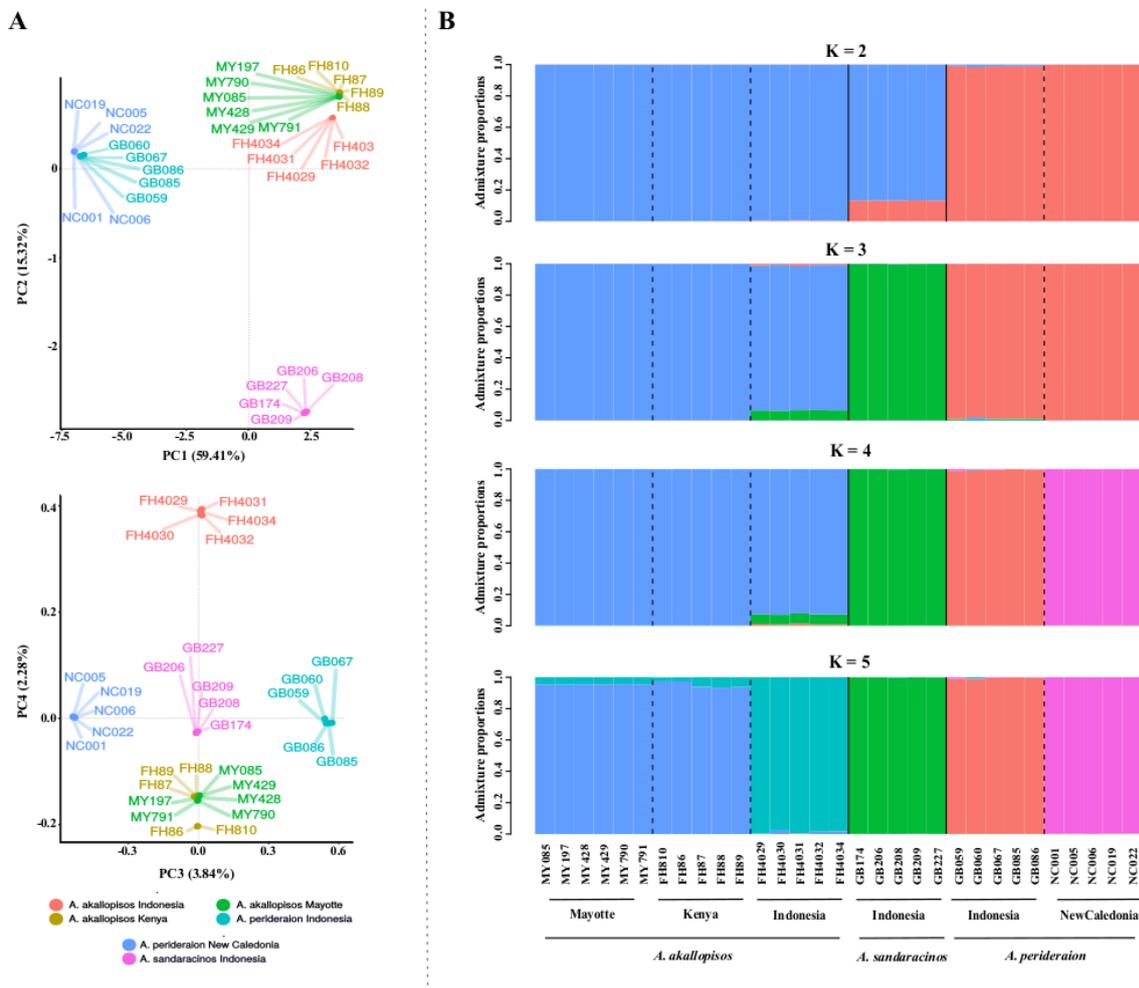


Figure 4: PCA (A) and admixture (B) results for the individuals of the skunk complex obtained from PCAngsd. In the PCA, the first two components (top) split samples according to the species (A), whereas PC3 and PC4 (bottom) separate the two populations of *A. perideraion* (Indonesia vs. New Caledonia) and *A. akallopis* (Indonesia vs. Western Indian Ocean). In the admixture plot, individuals' ancestry proportions for two to five ancestral populations K are reported. Multiple runs with different seeds for each K were performed to ensure convergence. The highest likelihood solutions for each K are reported. Automatic selection of the best K by PCAngsd resulted in K=3.

3.4. Species and populations genomic divergence

Overall, the average F_{st} calculated between all populations showed a clear species and population divergence, except for the two *A. akallopisos* populations of the WIO (Supplementary Information S2). However, when we examined the divergence across the genome, we identified high heterogeneity in the absolute (d_{xy}) and relative (F_{st}) genetic divergence (Figure 5). This result was stronger for the F_{st} calculations but remained valid for d_{xy} . The patterns were independent of the size of the sliding windows (Supplementary Table S3), or the populations considered within each species (Supplementary Figures S3 and S4).

We explored the distribution of windows of increased or decreased divergence (i.e., upper and lower 1% of the F_{st} and d_{xy} distributions) between the three species, and we observed that they were scattered across the chromosomes (Figure 5). However, an exception was observed for the comparisons between *A. akallopisos* – *A. perideraion* and *A. sandaracinos* – *A. perideraion*, where all the windows of increased absolute (d_{xy}) and relative (F_{st}) divergence were clustered in two regions of chromosome 18 (Figure 5, Supplementary Figure S5). These two regions also included all the windows of decreased absolute divergence (d_{xy}) between *A. akallopisos* and *A. sandaracinos* (Figure 5, Supplementary Figure S5), and they were characterized by a reduced nucleotide diversity π in all populations (Figure 6 and Supplementary Figure S6).

The two regions of differential divergence on chromosome 18 extended from ca. 2.9 Mb to 3.5 Mb and from ca. 7.2 Mb to 16.9 Mb. Within them, we identified 408 functionally annotated genes, and their GO enrichment analysis resulted in 13 enriched GOs (p -value < 0.01, Supplementary Table S7). Among them, we observed terms associated with the regulation of behavior (GO:0050795), the development of endoderm (GO:0007492), the morphogenesis of the epithelium (GO:1905332).

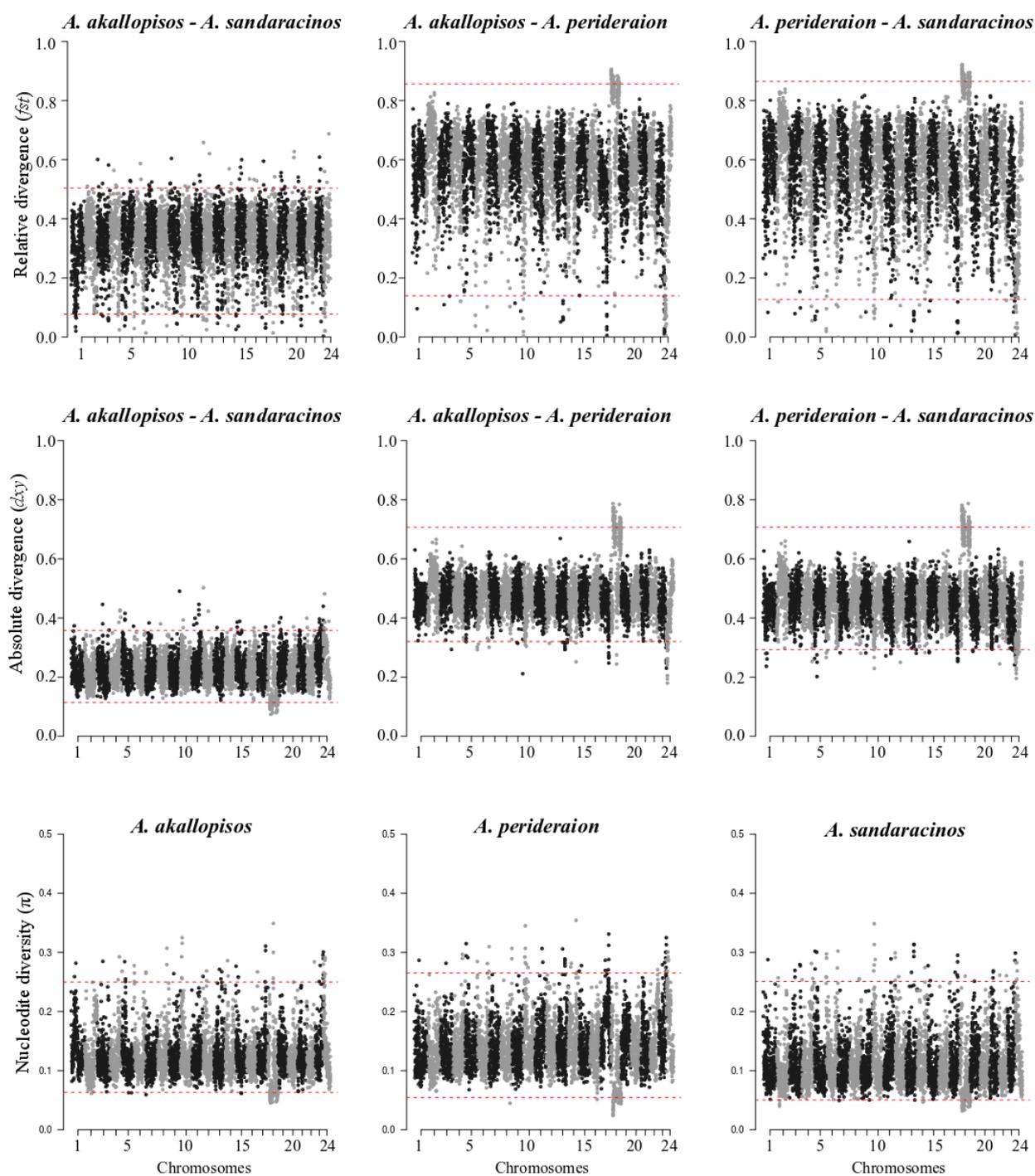


Figure 5: Sliding windows analysis of the relative (F_{st}) and absolute (d_{xy}) genetic divergence and nucleotide diversity (π) for the 24 for chromosomes. Results for sliding windows of 3,000 variable sites are reported. Red dotted lines represent the upper and lower 1% of each distribution. Plot for the Indonesian populations of each species are reported. Pairwise comparisons of F_{st} and d_{xy} for all the combinations of populations, and π for all populations are reported in Supplementary Figure S3-S5.

3.5. Species topological inconsistencies and signal of admixture

We explored the presence of topological inconsistencies across the nuclear genome of the skunk complex, which may reflect ancestral gene flow (i.e., before the population splits) between the species. Three different rooted topologies were possible (represented in Figure 6). Topology 1 represented the topology of the nuclear phylogenetic tree (Figure 3), topology 2 corresponded to the mitochondrial phylogenetic tree (Figure 3), and topology 3 was characterized by *A. akallopisos* and *A. perideraion* being sister species (Figure 6A). Topology 1 was the most frequent topology across the genome, with an average weighting higher than 80% (Figure 6A), and it was fully supported in 74.8% of the windows (Supplementary Table S8). Topology 2 was also represented in a relatively high proportion of genomic windows, with an average weighting of 13% (Figure 6A), and it was fully supported in 7.8% of windows. This proportion was consistent when considering different window sizes (Supplementary Table S8). The remaining topology 3 had an average weighting of 7% (Figure 6A) and was completely supported in only 1.8% of the windows (Supplementary Table S8).

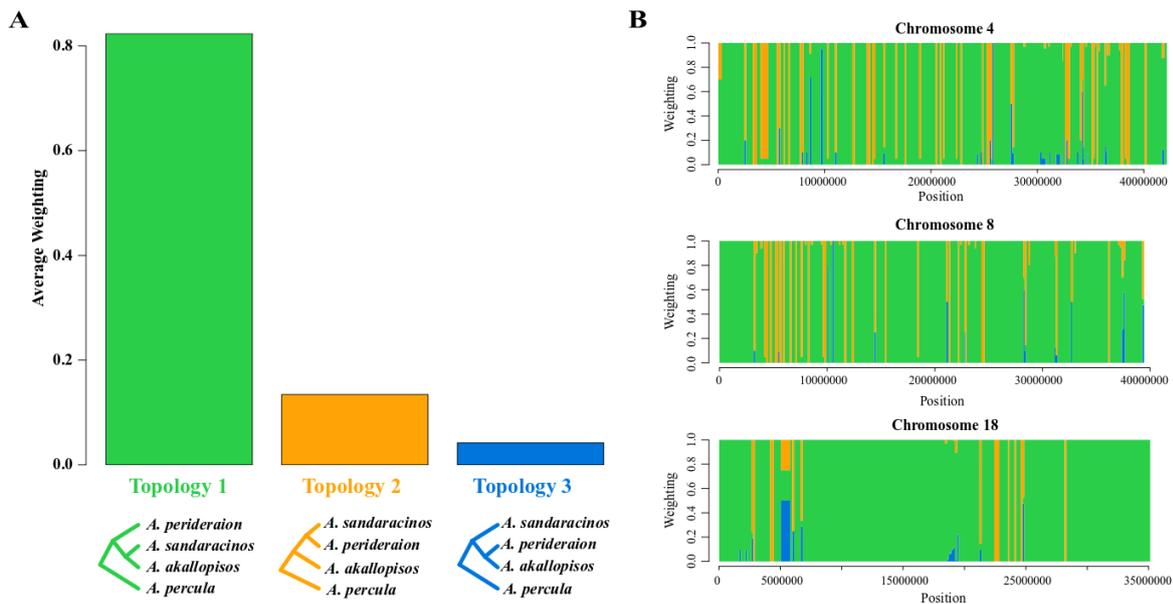


Figure 6: Whole-genome average weighting for each species-level topology (A), and examples of topological weighting across chromosomes 4, 8, and 18 (B). Results were obtained with *twisst*, for sliding windows of 3,000 variant sites. The weightings are determined by sampling in turn a single member of each species and identifying the topology matched by the resulting subtree. Weightings are calculated as the frequency of occurrence of each topology in each window. Average weightings (A) are calculated genome-wide. Topology 1 (green) represents the expected nuclear topology, and topology 2 (orange) corresponds to the mitochondrial topology. Chromosome 18 shows two large regions of full support for topology 1. Results obtained for different window sizes were consistent (Supplementary Table S8).

We looked at the distribution of the different topologies across the 24 chromosomes separately. We observed that most chromosomes had an average weighting and a proportion of fully-supported topologies consistent with the whole-genome results (Figure 6 and Supplementary Table S9). There were, however, three notable exceptions. First, chromosomes 23 and 24 showed an increased average weighting for the topology 3 compared to the other chromosomes and the whole genome (8.1% and 12% for chromosomes 23 and 24, respectively, vs. 4.2% for the whole genome; Supplementary Table S9). The prevalence of topology 3 was associated with a decreased proportion of windows fully supporting any of the three topologies (Supplementary Table S9 and Figure S7), which resulted from a highly heterogeneous SNP density profile in these regions (Supplementary Figure S2). Second, chromosome 18 showed a decreased support for both topologies 2 and 3 compared to the rest of the genome (4.6% and 1% for topology 2 and 3, respectively, vs. 13.4% and 4.2% for the whole genome; Figure 6B, Supplementary Table S9, and Figure S7). This decrease resulted from two large genomic regions that completely supported topology 1 (Figure 6B) and coincided with the two regions of increased genetic divergence between *A. akallopisos* – *A. perideraion* and *A. sandaracinos* – *A. perideraion* (Figure 5).

The presence of topology 2 across the genome (Figure 6), together with the cytonuclear discordance (Figure 3) and admixture analysis (Figure 4B), suggested the presence of ancestral gene flow between *A. sandaracinos* and *A. perideraion* but could also result from incomplete lineage sorting (ILS). Thus, we tested for evidence of hybridization with ABBA-BABA tests (Green et al., 2010; Durand et al., 2011). We found a genome-wide excess of ABBA sites that resulted in a significantly positive Patterson's *D* statistics (0.171 ± 0.0041 ; Z-score of 41.9), confirming a signal of hybridization between *A. sandaracinos* and *A. perideraion*. Results were consistent when considering the two populations of *A. perideraion* simultaneously or individually (Supplementary Table S10). The overall proportion of admixture estimated with the *f* statistic showed that about 5.5% of the genome is shared between *A. sandaracinos* and *A. perideraion* ($f = 0.056$, 95% CI [0.052-0.060]).

At the chromosome level, we found a significant Patterson's *D* statistic for all 24 chromosomes (Figure 7A, Supplementary Table S10), but, consistently with the overall genomic divergence results and *twisst* analysis, chromosome 18 showed the lowest *D* statistics (0.1 ± 0.03 , Z-score of 3.15; Fig. 7A, Supplementary Table S10). Similarly, chromosome 18 stands again out when looking at the *fd* statistics, with values lower than the other chromosomes (Fig. 7B) and with two large genomic regions showing extremely low *fd* ($fd < 0.01$; Fig. 7B).

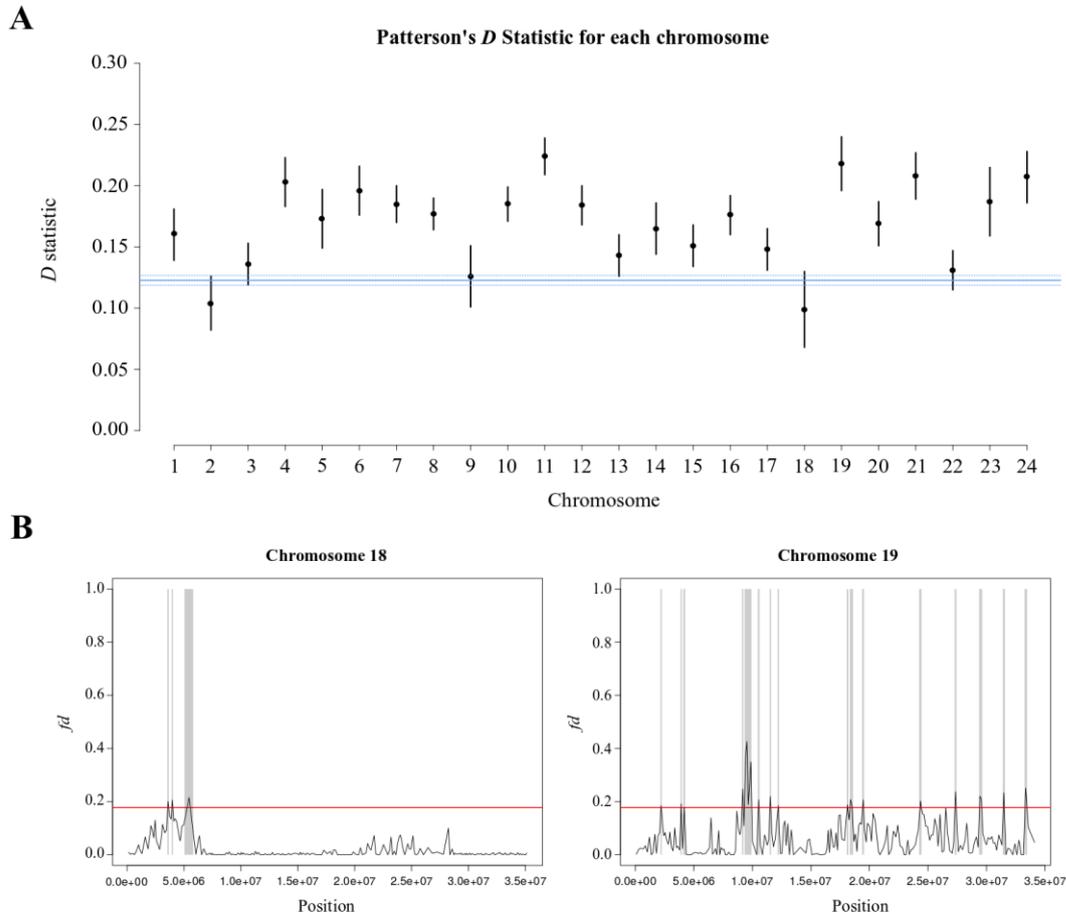


Figure 7: Patterson's D statistics of each chromosome (A) and examples of f_d distribution along chromosome 18 and 19 (B) for the test of introgression between *A. perideraion* and *A. sandaracinos*. A) The D statistics \pm s.e. are reported for each chromosome. In blue, the genome-wide estimate of $D \pm$ s.e. is reported. Standard errors were obtained with a block jackknife approach. All chromosomes show a D -statistic significantly deviating from zero, indicating introgression between *A. perideraion* and *A. sandaracinos*. B) Red lines represent the 95th percentile of the genome-wide f_d distribution, which was set as the cutoff for determining candidate regions of introgression CRI, (highlighted in gray). Chromosome 18 shows two regions of extremely low f_d ($f_d < 0.01$).

The candidate regions of introgression between *A. perideraion* and *A. sandaracinos* (CRI, i.e., upper 5% of the f_d distribution) were distributed across all 24 chromosomes (Fig. 7B and Supplementary Figure S8) and overlapped with the regions supporting the mitochondrial topology (i.e., *A. sandaracinos* and *A. perideraion* as sister species) detected by *twisst*. Because introgressed regions generally show lower absolute genetic divergence (d_{xy} ; Smith & Kronforst, 2013; Martin et al., 2015), we estimated d_{xy} between the different species in the CRI and in the rest of the genome. For the comparison between *A. perideraion* (Indonesia) and *A. sandaracinos*, we found that d_{xy} was significantly lower for the CRI (M=0.34, SD=0.045) compared to the rest of the genome (M=0.46, SD=0.061; $t(332)=-40.6$, $p < 0.001$; Figure 8). Consistent results were obtained

between *A. perideraion* from New Caledonia and *A. sandaracinos* (CRI: M=0.34, SD=0.046; non-CRI: M=0.46, SD=0.061; $t(330)=-40.3$, $p < 0.001$). In contrast, the CRI showed an increased d_{xy} between *A. sandaracinos* and *A. akallopisos* compared to the rest of the genome (CRI: M=0.32, SD=0.041; non-CRI: M=0.23, SD=0.043; $t(308)=36.4$, $p < 0.001$; Figure 8), while they showed again a decreased d_{xy} between *A. akallopisos* and *A. perideraion* (outliers regions: M=0.42, SD=0.060; non-outliers regions: M=0.47, SD=0.056; $t(332)=-40.64$, $p < 0.001$), despite this decrease was less pronounced (Figure 8).

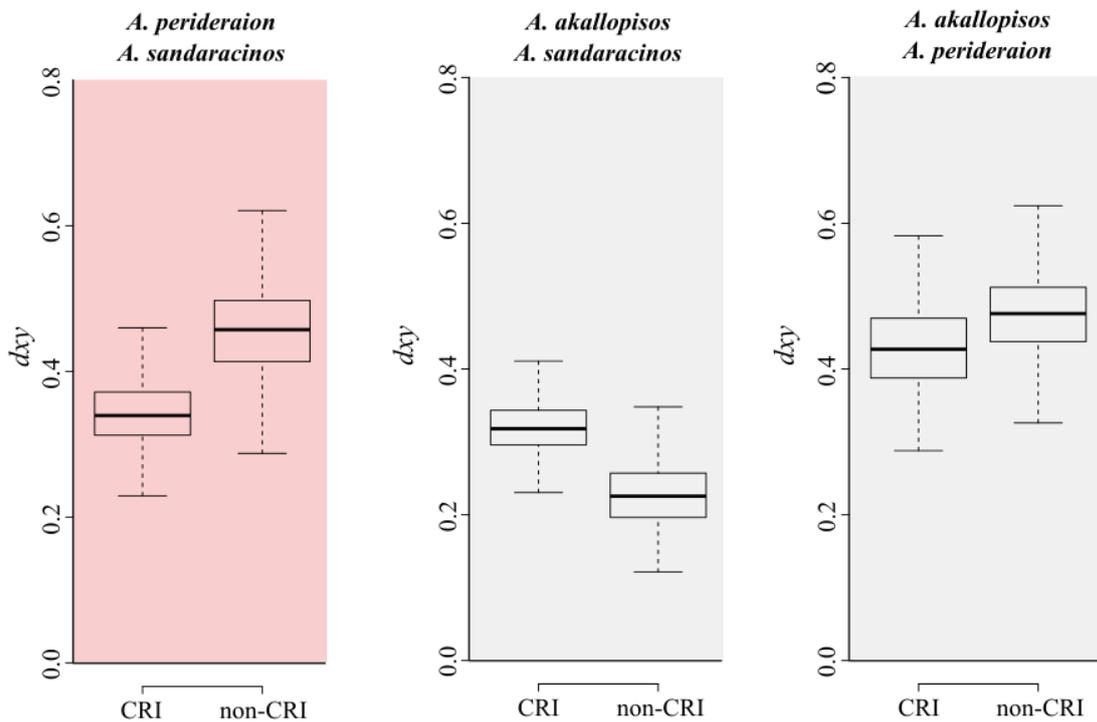


Figure 8: Comparisons of the absolute genetic divergence (d_{xy}) of the candidate regions of introgression (CRI, between *A. perideraion* and *A. sandaracinos*) and the rest of the genome (non-CRI) for each species pair. The CRI between *A. perideraion* and *A. sandaracinos* were defined as the regions in the top 5% of the f_d distribution. The comparison of *A. perideraion* and *A. sandaracinos* is highlighted in red, as it depicts the d_{xy} of the two introgressing species. Comparisons for Indonesian populations of the species are reported here, but comparable results were obtained for each population.

We identified 905 functionally annotated genes in the CRI between *A. perideraion* and *A. sandaracinos*. GO enrichment analysis on these genes resulted in 21 enriched GOs (p -value < 0.01 , Supplementary Table S11). Among them, we observed terms associated with the sensory perception of a light stimulus (GO:0050953), the adult feeding behavior (GO:0008343), the neuropeptide signaling pathway (GO:0007218), as well as terms linked with the immune system (Supplementary Table S11).

3.6. Populations topological inconsistencies and additional admixture events

While the Indonesian populations of *A. perideraion*, *A. sandaracinos*, and *A. akallopisos* were differentiated (Figure 4A) and showed relatively high genomic divergence (Supplementary Information S2 and Supplementary Figure S3), the admixture analysis suggested some level of shared ancestry between sympatric populations of *A. akallopisos* and *A. sandaracinos* (Figure 4B). We investigated possible gene flow events between the populations with TreeMix, which detected migration events between Indonesian populations of *A. akallopisos* and *A. sandaracinos*, besides the gene flow between *A. sandaracinos* and *A. perideraion* (see Supplementary Information S4).

We explored the presence of topological inconsistencies at the population level, with topologies branching the three sympatric populations together, which would suggest recent gene flow (i.e., after the population splits) in the IAA. By considering all the possible topologies branching the Indonesian populations of *A. akallopisos* and *A. sandaracinos* as sister species, we obtained an average weighting of only 3.35%. We observed even lower average weighting when considering topologies with the Indonesian populations of *A. perideraion* and *A. sandaracinos* as sister species (1.8% average weighting), and with those of *A. perideraion* and *A. akallopisos* branching together (1.0%). Additionally, we did not find any windows fully supporting any of these topologies. For more details, see the Supplementary Information S3.

3.7. Demographic reconstruction

A possible hybrid origin of *A. sandaracinos* has been suggested (Litsios & Salamin, 2014), and the results presented above indicate past gene flow between *A. perideraion* and *A. sandaracinos* (predating the split of *A. perideraion* populations). Additionally, although we did not find evidence of recent extensive gene flow in sympatry, some indications of admixture signatures specific to sympatric populations were observed (Figures 3 and 4B, Supplementary Information S4). Thus, we can take advantage of the disjunct geographical distribution of the species to better understand how the skunk complex diversified by further testing the presence of ancient hybridization between the species - which would influence both allopatric and sympatric populations -, and more recent introgression - which would affect only the populations in the IAA (Figure 2).

We contrasted different diversification scenarios without or with gene flow (Figure 2) that potentially result in the observed patterns of admixture between the species. We obtained the highest likelihood for model L, in which gene flow happened between *A. perideraion* and the ancestor of *A. akallopisos* - *A. sandaracinos*, and also between all species throughout

diversification of the group (Figures 1L and 9B). Models of strict isolation (Figure 1A) and hybrid speciation (Figures 1B and 11C) resulted in the lowest likelihoods (Figure 9A and Supplementary Table S12), and we can thus reject the hypothesis of a hybrid origin of *A. sandaracinos*, as well as the one of strict isolation in the diversification of the skunk complex. The other models also showed reduced likelihood compared to model L, with no likelihood distribution overlap (Figure 9A and Supplementary Table S12), and we can thus reject these scenarios.

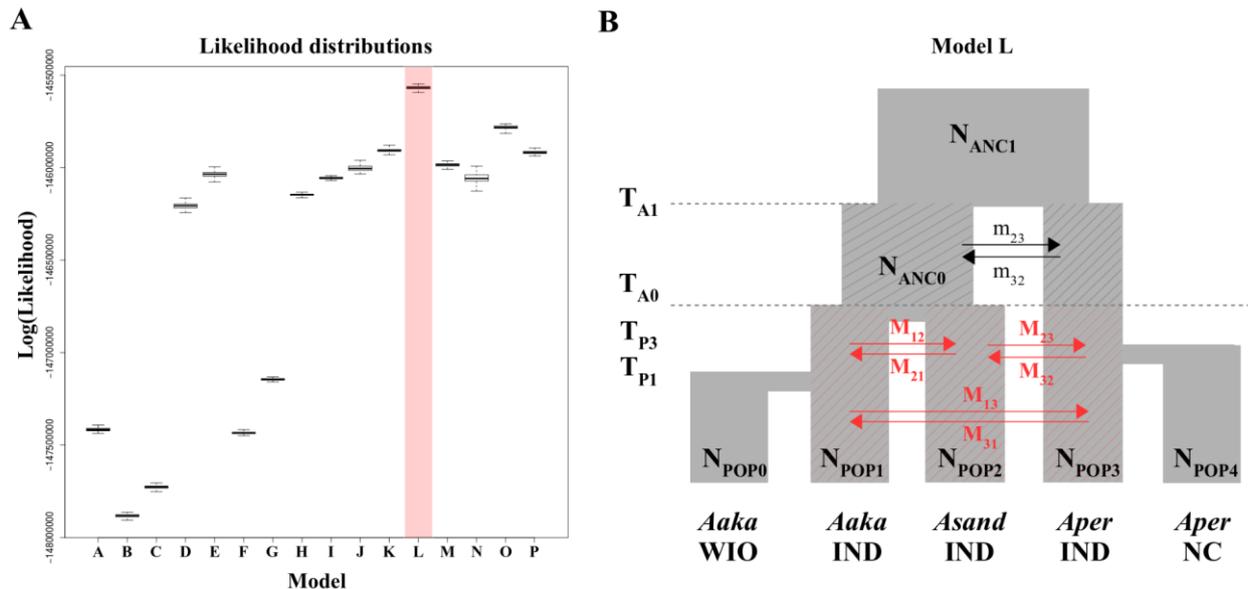


Figure 9: Likelihood distributions obtained for the 16 demographic scenarios (A) and the schematic representation of the resulting best model (B). A) The best resulting model (model L) is highlighted in red. All tested scenarios (A to P) are depicted in Figure 2. Likelihood distributions were obtained based on 100 expected SFS, each approximated using 1 million coalescent simulations under the parameters that maximized the likelihood for each model. An overlap of distributions between models indicates no significant difference between the fit. B) Gene flow between *A. perideraion* and the ancestor of *A. akallopisos* and *A. sandaracinos* is represented by black arrows. Red arrows correspond to more recent migrations (i.e., after the split between *A. akallopisos* and *A. sandaracinos*). We define “recent gene flow” when it is modeled to also occur in the present time. The exchanging populations were highlighted using hatching. Parameter estimates for the best model are reported in Table 2.

Parameter estimates obtained for the best model reported the split between the three species (T_{A1}) to be ca. 1.2 million generations and the one between *A. akallopisos* and *A. sandaracinos* (T_{A0}) to be ca. 0.3 million generations (Table 1). Considering a generation time of 5 years for clownfishes (Buston & Garcia, 2007), this corresponds to a divergence time of ca. 6 MYA and 1.5 MYA, respectively. The colonization from Indonesia to New Caledonia of *A. perideraion* (T_{P3}) was estimated to be ca. 509 KYA, while the split of *A. akallopisos* populations appears to be more recent (220 KYA; Table 1). The effective population sizes were estimated to be smaller for the *A.*

perideraion population of New Caledonia (N_{POP4}) and the *A. akallopisos* from the WIO (N_{POP0} ; Table 1). Within the IAA, *A. sandaracinos* showed the lowest effective population size (N_{POP2} , Table 1).

Estimates of the migration rates for the best model ranged between 1.3×10^{-8} and 5.14×10^{-6} , depending on the considered species and timing (Table 1). The highest rates corresponded to the ancestral migrations between *A. perideraion* and the ancestor of *A. akallopisos* - *A. sandaracinos* (migrations m_{32} and m_{23} , Table 1 and Figure 9B). Given the effective population sizes, the point estimates of effective migration rates ($2Nm$, with N the haploid population size) were 8.89 gene copies per generation from *A. perideraion* to the *A. akallopisos* – *A. sandaracinos* ancestor and 0.6 from the latter to *A. perideraion*. We obtained minor migration rates from *A. akallopisos* to *A. sandaracinos* ($M_{12} = 1 \times 10^{-6}$, Table 1 and Figure 9B) and from *A. perideraion* to *A. sandaracinos* ($M_{32} = 7.84 \times 10^{-7}$, Table 1 and Figure 9B), corresponding to effective migration rates of 0.49 and 0.38 gene copies per generation, respectively. The lowest migration rates were observed between *A. perideraion* and *A. akallopisos* ($M_{13} = 1.82 \times 10^{-8}$ and $M_{31} = 1.29 \times 10^{-8}$, Table 1 Figure 9B), corresponding to effective migration rates of 0.022 and 0.021 gene copies per generation, respectively.

Table 1. Parameter estimates for the best model with ancestral and recent gene flow between species (model L). Population sizes are haploid sizes. Divergence times are reported in generations. The model is depicted in Fig. 2L and 9B. 95% Confidence Intervals (CI) were obtained by parametric bootstrap.

Parameter	Estimate	95% CI	Parameter	Estimate	95% CI
N_{POP0}	161,285	[161,285; 161,285]	M_{13}	1.82×10^{-8}	[1.79×10^{-8} ; 1.86×10^{-8}]
N_{POP1}	832,856	[832,817.3; 832,894.6]	M_{31}	1.29×10^{-8}	[1.25×10^{-8} ; 1.33×10^{-8}]
N_{POP2}	244,917	[244,917; 244,917]	M_{21}	8.59×10^{-8}	[8.56×10^{-8} ; 8.63×10^{-8}]
N_{POP3}	620,804	[620,804; 620,804]	M_{12}	9.99×10^{-7}	[9.99×10^{-7} ; 9.99×10^{-7}]
N_{POP4}	148,393	[148,393; 148,393]	M_{23}	7.08×10^{-7}	[7.05×10^{-8} ; 7.11×10^{-8}]
N_{ANCO}	865,060	[865,060; 865,060]	M_{32}	7.84×10^{-7}	[7.84×10^{-7} ; 7.84×10^{-7}]
N_{ANCI}	686,668.3	[686,636.9; 686,699.7]	m_{32}	5.14×10^{-6}	[5.13×10^{-6} ; 5.143×10^{-6}]
T_{P1}	45,697.27	[45,628.75; 45,765.79]	m_{23}	4.82×10^{-7}	[4.79×10^{-7} ; 4.84×10^{-7}]
T_{P3}	101,809.9	[101,750.6; 101,869.2]			
T_{A0}	294,654.8	[294,519.2; 294,790.4]			
T_{A1}	1,188,136	[1,184,580; 1,191,692]			

4. DISCUSSION

The skunk complex offered a particularly interesting system to study the mechanisms underlying the diversification of clownfish. Indeed, in addition to ecological speciation, both ancestral and recent hybridization events were hypothesized to have shaped the divergence of the group (Steinke et al., 2009; Litsios et al., 2012a; Litsios & Salamin, 2014). A possible hybrid origin of *A. sandaracinos* has been suggested (Litsios & Salamin, 2014), and the hypothesis that the group has seen ancient hybridization events should be tested in detail. The disjunct geographical distribution of the three species, with both allopatric and sympatric populations, represented an excellent setting to test the presence of ancestral gene flow between the species - which would influence both allopatric and sympatric populations - and more recent introgression - which would affect only populations in the Indo-Australian Archipelago (IAA). Our hypothesis was that the presence of gene flow between populations of different species occurring in the sympatric region of the IAA could lead to the homogenization of the genomic divergence between the species, which, in turn, would allow the identification of genomic regions of increased differentiation potentially involved in the ecological divergence of the complex (e.g., Nadeau et al., 2012; Clarkson et al., 2014; Talla et al., 2017).

Here, we took advantage of this disjunct geographical distribution and performed comparative genomics analyses of sympatric and allopatric populations of the skunk complex. First, we found that *A. sandaracinos* does not have a hybrid origin but is sister to *A. akallopisos*. However, ancestral gene flow occurred, mainly from *A. perideraion* to the ancestor of *A. akallopisos* - *A. sandaracinos*, with *A. sandaracinos* maintaining introgressed genomic regions. This signal of introgression is absent - or at least substantially decreased - in *A. akallopisos*. This result suggests that hybridization events also occurred just before the speciation of *A. akallopisos* and *A. sandaracinos*. While we cannot exclude a role of selection, given the lower effective population size estimated for *A. sandaracinos*, genetic drift in this species was stronger and potentially fixed the introgressed regions more extensively.

Secondly, we detected evidence of low gene flow between the three species throughout their diversification. Nevertheless, this extent of gene flow was insufficient to impact the overall differentiation of the species in the IAA, as similar patterns of divergence were observed in the allopatric populations. Despite the species co-occurrence and the potential interaction with the same sea anemone species (Figure 1B), extensive gene flow is absent in the species, suggesting a role of host repartition in maintaining the genetic identity of the species in sympatry (see below).

Finally, we observed that two large regions on chromosome 18 show increased divergence between *A. perideraion* and the two other species. This pattern suggests the presence of divergent haplotypes in the ancestral population that predate speciation, or additional hybridization events of *A. perideraion* with a species outside the group, with these regions being likely maintained by the disruption of recombination. Thus, the pattern observed on this chromosome should be further investigated in the context of clownfish diversification.

4.1. Overall species divergence and ancestral hybridization shaping the diversification of the skunk complex

Genomic comparisons of the three species of the skunk complex demonstrated that the three species were, overall, well-differentiated. The PCA and admixture analyses resulted in a clear cluster of individuals by species, separating first *A. perideraion* from the two other species (first axis of the PCA, Figure 4A; K=2, Figure 4B), and secondly *A. sandaracinos* from *A. akallopisos* (second axis of the PCA, Figure 4A; K=3, Figure 4B). These results were also consistent with the nuclear phylogenetic tree (Figure 3), placing *A. sandaracinos* and *A. akallopisos* as sister species, and with the measures of overall genomic divergence (Supplementary Information S2), which were congruent with previous studies (Litsios et al., 2012a; 2014).

The demographic modeling performed on the SNPs data resulted in an estimated divergence time for *A. perideraion* and the ancestor of *A. akallopisos* - *A. sandaracinos* of 6 MYA, while the split between *A. akallopisos* and *A. sandaracinos* was estimated to be 1.5 MYA (Table 1). Previous phylogenetic studies based on nuclear and mitochondrial markers resulted in an estimated divergence time for *A. perideraion* – *A. sandaracinos* and *A. akallopisos* – *A. sandaracinos* of about 1.8 MYA and 2.5 MYA, respectively (median divergence time from timetree.org, Kumar et al., 2017, obtained from Cowman et al., 2011 & 2013; Litsios et al., 2012a & 2012b; Frédérick et al., 2013; Litsios et al., 2014; Rabosky et al., 2013 & 2018). However, in these studies, *A. sandaracinos* branched as sister species to *A. perideraion*, which is different from the results obtained in our study based on the SNPs dataset. The different placement of these species in these previous studies, which included informative mitochondrial markers, likely resulted from the cytonuclear inconsistency of the group (discussed below). It further affected the previously published age estimates and resulted in an underestimated divergence time for *A. sandaracinos*-*A. perideraion*, and, consequently, in an overestimated divergence time for *A. sandaracinos* and *A. akallopisos*. Regarding the estimated divergence time for *A. perideraion* and the *A. akallopisos*-*A. sandaracinos* ancestor, similar results were obtained in previous studies (6.1 MYA and 5.9 MYA

respectively in Rabosky et al., 2013 and 2018).

Despite a clear separation of the species, we observed cytonuclear discordance, confirming previous findings by Litsios and Salamin (2014). Indeed, similarly to what was obtained in previous phylogenetic studies mentioned above (e.g., Cowman et al., 2011 & 2013; Frédérick et al., 2013; Rabosky et al., 2013 & 2018), *A. sandaracinos* and *A. perideraion* branched as sister species in the mitochondrial phylogenetic tree (Figure 3). This topology was also observed throughout the nuclear genome (Figure 6), and ABBA-BABA tests revealed a significant signal of introgression between *A. perideraion* and *A. sandaracinos* (Figure 7), ruling out incomplete lineage sorting (ILS) as an alternative explanation for the topological inconsistencies (Martin et al., 2015; Martin & Van Belleghem, 2017). Because the signal of introgression was independent of the geographical origin of *A. perideraion*, the gene flow between the two species is likely to have occurred before the split of the New Caledonian population of *A. perideraion*. Additionally, the candidate regions of introgression not only showed the expected reduction of absolute genetic divergence (d_{xy}) between the two species (Figure 8; Smith & Kronforst, 2013; Martin et al., 2015) but also showed increased d_{xy} between *A. sandaracinos* and *A. akallopisos*, suggesting that the gene flow happened from *A. perideraion* to *A. sandaracinos* (Figure 8). These observations were confirmed by demographic modeling, which excluded the two scenarios of diversification in strict isolation and hybrid origin of *A. sandaracinos*. However, it confirmed the presence of moderate gene flow from *A. perideraion* to the ancestor of *A. akallopisos* and *A. sandaracinos* (effective migration rate of 8.89 gene copies per generation, Figure 9 and Table 1), plus a low migration rate between *A. perideraion* and *A. sandaracinos* throughout the diversification (effective migration rate of 0.38 gene copies per generation, Figure 9 and Table 1). Such levels of gene flow have been previously reported to generate patterns of introgression in other taxa (e.g., Hey, 2010; Wielgoss et al., 2014; Pereira et al., 2016; Godoy et al., 2018; Bourgeois et al., 2020).

We detected in our analyses the highest level of gene flow from *A. perideraion* to the *A. akallopisos*-*A. sandaracinos* ancestor, and we should thus expect a similar signal of introgression in *A. akallopisos* than in *A. sandaracinos*. However, the proportion of topological inconsistencies across the genome branching *A. akallopisos* as the sister species of *A. perideraion* was low. Further, the significance of the ABBA-BABA tests for *A. sandaracinos* implied a reduced frequency of BABA loci (i.e., shared loci between *A. akallopisos* and *A. perideraion*) compared to ABBA sites (i.e., shared loci between *A. sandaracinos* and *A. perideraion*), excluding not only ILS but also identical rates of introgression between *A. perideraion* and the two other species (Martin et al., 2015). This imbalance in the retention of the introgression signal between *A.*

akallopisos and *A. sandaracinos* may result from two processes that are not mutually exclusive. First, the introgressed regions that we identified may arise from the more recent but weaker migrations from *A. perideraion* to *A. sandaracinos*. However, this should have occurred before the split of the New Caledonian population of *A. perideraion* because no geographic signal was detected in any analyses. Second, the hybridization with the ancestor of *A. akallopisos*-*A. perideraion* may have occurred just before the speciation of *A. akallopisos* and *A. sandaracinos*, and an increased fixation of the introgressed regions by selection or drift could have occurred in *A. sandaracinos*.

By examining the gene content of the introgressed regions, we found genes with functions related to feeding behavior, perception of light, and immunity. These functions are broad and cannot be easily associated with shared ecological similarities in host usage and/or adaptive traits between *A. perideraion* to *A. sandaracinos*, which would suggest a role of natural selection in the fixation of the regions (such as in Grant et al., 2010; Whitney et al., 2010; Pardo-Diaz et al., 2012). While we cannot exclude a role of natural selection without further investigation of the functional traits of the species, the fixation (or increased fixation compared to *A. akallopisos*) of these regions in *A. sandaracinos* may have also occurred through neutral processes (Martin & Jiggins, 2017). Indeed, the estimated effective population size is lower in *A. sandaracinos* than in the two other IAA populations (Table 1), suggesting a stronger effect of genetic drift in this species (Martin & Jiggins, 2017; Sagonas et al. 2019).

The gene flow between *A. perideraion* and the ancestor of *A. akallopisos*-*A. sandaracinos* was present, was moderate, and occurred until the split of *A. akallopisos* and *A. sandaracinos*. While we cannot exclude a scenario of speciation with gene flow, where ecological divergence and adaptation were the main drivers of speciation, these results can also indicate a scenario of secondary contact, where gene flow happened following the speciation of *A. perideraion*. In that case, *A. perideraion* speciation may have occurred both through isolation or ecological divergence.

4.2. Overall populations divergence and low migration rates in the IAA

We observed that the allopatric populations of *A. perideraion* (IAA vs. New Caledonia) and *A. akallopisos* (IAA vs. WIO) were overall differentiated. Indeed, within the species, the PCA and admixture analyses showed a cluster of individuals by geography (third and fourth axis in the PCA, Figure 4A; K=4 and K=5, Figure 4B), and populations were monophyletic in the reconstructed mitochondrial and nuclear phylogenetic trees (Figure 3). Within the WIO, the population differentiation of *A. akallopisos* was reduced (Figure 3, Figure 4A-B).

These results are consistent with restricted gene flow across large geographic distances and with barriers to dispersal between the IAA and the WIO (Appleyard et al., 2002; Ridgway & Sampayo, 2005; Ragionieri et al., 2009; Huyghe & Kochzius, 2017 & 2018). As for most coral reef fishes, clownfishes have a biphasic life cycle, and dispersal occurs only at the larval stage (Leis & McCormick, 2002), whose duration is widely used as a proxy for the potential dispersal distance of a species (Shanks et al., 2003; Lester & Ruttenberg, 2005; Shanks, 2009; Hilário et al., 2015). In clownfish, the dispersal phase only lasts between 10 and 15 days (Fautin & Allen, 1997), typically corresponding to a dispersal range of 5-100 km (Planes et al., 2009; Salinas-de-León et al., 2012). Nevertheless, even with dispersal as long as 400 km documented for *A. omanensis* (Simpson et al., 2014), the exchange between the IAA and New Caledonia (ca. 6,000 km) or WIO (ca. 8,000 km) would be reduced, thus resulting in the population differentiation observed here. Within the WIO, the two populations of *A. akallopisos* (Kenya and Mayotte) are less differentiated due to the reduced geographical distance and the presence of the South Equatorial and the East African Coastal Currents (Figure 1 in Schouten et al., 2003), which could maintain the connectivity between populations.

The split of the WIO populations of *A. akallopisos* from the IAA was estimated to be 220 thousand years ago (kYA), while the separation of the New Caledonian population of *A. perideraion* happened 509 kYA (Table 1). However, we did not test for scenarios including gene flow between the allopatric populations (Figure 2), which could lead to older divergence estimates if gene flow is still occurring, even at a low rate, between these populations (as it is expected given their genetic structure, Supplementary Information S2). We also observed that estimates of effective population sizes were lower for WIO and New Caledonian populations than those of the IAA (Table 1), and that they also have slightly reduced nucleotide diversities (π ; Supplementary Table S3). These results are consistent with reduced effective population sizes at the margin of the geographical distributions (i.e., the central-marginal hypothesis, Eckert et al., 2008), likely resulting from founder events and range expansion (Peter & Slatkin, 2013; Schulte et al., 2013; Pierce et al., 2014; Braasch et al., 2019), and they further confirm the IAA origin of the skunk complex (Santini & Polacco, 2006; Litsios et al., 2014; Huyghe & Kochzius, 2017).

Within the IAA, the populations of the three species were clearly separated in the PCA and admixture analyses as well as in all phylogenetic reconstructions (Figures 3 and 4A-B). We further did not observe extensive population-level topological inconsistencies (Supplementary Information S3) nor differences in the genomic divergence heterogeneity of sympatric vs. allopatric comparisons (Figure 5 and Supplementary Figures S3-4, to note that the peaks on

chromosome 18 are present independently of the considered populations, as discussed below). Altogether, these results indicate the absence of extensive gene flow between the species in the IAA. Hybridization between members of the complex is frequent in the aquarium trade (Pushparaj, 2010; He et al., 2019) and was suggested to also happen in nature (Steinke et al., 2009). Indeed, members of the group can interact with the sea-anemone species (by pairs, Figure 1), and natural hybrids involving members of the complex exist (i.e., *A. leucokranos* and *A. thiellei*; Ollerton et al., 2007, Litsios & Salamin, 2014, Gainsford et al., 2015). Despite these observations, our findings show that gene flow in the IAA is not as common as previously thought. Hybridization events between the species may be avoided - or be substantially decreased - through competition-driven host repartition in sympatry, which would allow maintaining the genetic identity of the species as observed here. Preliminary results on clownfish-sea anemone host association frequency seem, indeed, to point in this direction (Titus et al., unpublished results, Supplementary Table S16). However, further studies are necessary before drawing firm conclusions.

Despite the absence of extensive hybridization, demographic modeling detected a low level of gene flow in the IAA between the three species and throughout their diversification (Figure 9, Table 1). The migration rate was low in all types of population exchanges (Table 1). As a comparison, extensive gene flow resulted in effective migration rates higher than 6 in the Reunion grey white-eye (Bourgeois et al., 2020), between 0.85 and 3.28 in *Bromelia hieronymi* (Godoy et al., 2018), and higher than 4.6 in the North Atlantic eels (Wielgoss et al., 2014). This weak gene flow may account for the small proportion of shared ancestry detected between *A. akallopisos* and *A. sandaracinos* in the IAA (Figure 4B, $K=3$ and $K=4$), and for the cytonuclear inconsistency observed for the *A. sandaracinos* sample GB227 (Figure 3, grouped with its conspecific at the nuclear level, but with *A. perideraion* of the IAA as the mitochondrial level). The retention of an allospecific mitochondrial genome in GB227 likely emerged from mitochondrial capture, in which hybridization events followed by selective backcrossing of hybrids with one of the parental species (here, *A. sandaracinos*) results in complete mitochondrial introgression (e.g., Toews & Brelsford; 2012; Perea et al., 2016; Bertrand et al., 2017). In such a case, evidence of introgression at the nuclear level is not expected (Toews & Brelsford; 2012; Perea et al., 2016), consistent with what was observed for sample GB227.

We did not detect extensive gene flow that would have homogenized the genomic divergence between species in the IAA, potentially revealing regions of increased differentiation important for the ecological divergence of the species. However, we still believe that the skunk complex represents an interesting system for studying the diversification mechanism in a marine

environment. Nevertheless, future studies should include the more recently described species *A. pacificus* (Allen et al., 2010) and the known natural hybrids originating from the species of the complex (*A. leucokranos* and *A. thiellei*; Ollerton et al., 2007, Litsios & Salamin, 2014; Gainsford et al., 2015). For instance, the integration of hybrids at different stages of backcrossing showing different phenotypes (Gainsford et al., 2015) could allow identifying regions of restricted introgression potentially involved in maintaining species identity (Payseur, 2010) and could permit to link the phenotypes to specific loci in the genome through genotype-phenotype association analyses (e.g., Brelsford et al., 2017; Aguillon et al., 2021).

Divergence on chromosome 18

Chromosome 18 showed particular patterns of differentiation in the skunk complex. Two large regions on the chromosomes (from ca. 2.9 Mb to 3.5 Mb and from ca. 7.2 Mb to 7.2 Mb to 16.9 Mb) showed an increased absolute (d_{xy}) and relative (F_{st}) genetic divergence between *A. perideraion* and the two other species but a reduced d_{xy} between *A. sandaracinos* and *A. akallopisos* (Figure 5), and this was independent of the populations considered (Supplementary Figures S3-4). While an increase in the F_{st} may emerge from the reduction in nucleotide diversity observed in these regions (Figure 5; Cruickshank & Hahn, 2014), measures of d_{xy} are less affected by the decrease of genetic variation (Cruickshank & Hahn, 2014; Burri et al., 2015; Wolf & Ellegren, 2017). Thus, these regions likely originated from the introgression in *A. perideraion* of species not considered in our study - which must have preceded the split of the New Caledonian population - or from the presence of diverged haplotypes in the ancestral populations that predate *A. perideraion* speciation (i.e., ancient haplotype, Cruickshank & Hahn, 2014). Previous phylogenomic studies including more clownfish species showed that, in these regions, *A. perideraion* displayed topological inconsistency and branched with the species *A. frenatus* and *A. melanopus*, suggesting past hybridization events with members of the *ephippium* clade (see Chapter 3).

Mechanisms maintaining the high divergence in these regions must be present. The most likely scenario consists of the two regions in *A. perideraion* carrying structural variations (such as inversions), which disrupt the recombination with the haplotypes of *A. akallopisos* and *A. sandaracinos* (Kirkpatrick, 2010; Stevison et al., 2011). This hypothesis is consistent with the absence of introgression from *A. perideraion* to *A. sandaracinos* observed in these two regions (Figures 6 and 7), with their low divergence between *A. akallopisos* and *A. sandaracinos* (Figures 5), and with the reduced nucleotide diversity observed in these windows (Figure 5 and

Supplementary Figure S6), which can result from a low recombination rate (Cruickshank & Hahn, 2014; Burri et al., 2015; Burri, 2017; Wolf & Ellegren, 2017). Chromosomal inversions fixed by selection and resulting in higher genetic divergence are, for instance, observed in the case of supergenes, i.e., genomic regions containing clusters of linked loci often controlling ecologically important traits (e.g., Joron et al., 2011; Kunte et al., 2014; Zinzow-Kramer et al., 2015; Küpper et al., 2016; Branco et al., 2018). The evolution of sex chromosomes may also result in particular chromosomal patterns (e.g., Natri et al., 2019). However, we believe that this is less relevant in clownfishes, as these species are sequential hermaphrodites with no sex chromosomes (Fricke, 1979; Moyer & Nakazono, 1978; Arai, 2011), and genes involved in the sex change are scattered throughout the genome (Casas et al., 2018).

Understanding how these two regions of higher divergence were acquired and maintained by *A. perideraion* will require future studies integrating additional clownfish species. The two regions on chromosome 18 contain genes with functions associated with the regulation of behavior, the development of the endoderm, and the epithelium's morphogenesis. Linking these functions to potential advantageous traits in *A. perideraion*, and thus assuming a potential role of selection, is not possible without further knowledge on functional traits of the species. Furthermore, demographic models only based on SNPs located in these regions - and their comparison with models based on the rest of the genome - will allow further insights into the evolutionary history of chromosome 18. Finally, establishing a recombination map and/or investigating structural variants across the genomes of clownfishes will be necessary to better understand the mechanisms underlying the observed patterns on chromosome 18.

5. CONCLUSIONS

Given the hypotheses of ecological speciation, ancestral hybridization, and recent gene flow within the group, the skunk complex represented an interesting system to study the mechanisms underlying the diversification process in clownfish. Additionally, the disjunct geographical distribution of the species offered an ideal framework to test for differential hybridization events at different timescales.

Our results exclude the scenarios of strict isolation and hybrid origin of *A. sandaracinos* but support a scenario of ancestral gene flow between *A. perideraion* and the *A. akallopisos*-*A. sandaracinos* ancestor. However, contrary to what we expected, the three species only experienced low levels of gene flow throughout their diversification, and extensive gene flow was not observed

in the IAA. These results suggest a role of host repartition in maintaining the genetic identity of the species in sympatry. We also detected regions of introgression in *A. sandaracinos* and regions of increased divergence in *A. perideraion*. However, linking these regions to potential roles in the ecological divergence and adaptation is not reasonable without further studies on the functional traits of the species.

This study provides new cues for the further investigation of the mechanisms underlying clownfish diversification. For instance, in addition to the need for a better characterization of functional traits in the group, further studies on clownfish-sea anemone association frequency are necessary to advance our understanding of the level of species interaction in sympatry. Additionally, the integration of more clownfish species, such as the recently described *A. pacificus*, the known natural hybrids, and the species outside the skunk complex potentially hybridizing with *A. perideraion*, will further advance our knowledge of the diversification of this group. Nevertheless, our results provide the first insights into the diversification of the skunk complex, bringing us a step closer to the understanding of the mechanisms underlying clownfish adaptive radiation.

SUPPLEMENTARY MATERIAL

Given its considerable size, the supplementary material was not directly integrated into the manuscript, essentially to avoid excessive printing. The supplementary material is nevertheless easily accessible online. [Supplementary Information and Figures](#)¹ and [Supplementary Tables](#)² are available on OneDrive.

REFERENCES

- Abbott, R. J., Barton, N. H., & Good, J. M.** (2016). Genomics of hybridization and its evolutionary consequences. *Molecular ecology*.
- Abbott, R. J., Hegarty, M. J., Hiscock, S. J., & Brennan, A. C.** (2010). Homoploid hybrid speciation in action. *Taxon*, 59(5), 1375-1386.
- Aguillon, S. M., Walsh, J., & Lovette, I. J.** (2021). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proceedings of the Royal Society B*, 288(1943), 20201805.
- Akaike, H.** (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6), 716-723.
- Alexa, A. & Rahnenfuhrer, J.** (2016). topGO: Enrichment Analysis for Gene Ontology. R package version 2.26.0.

¹https://unils-my.sharepoint.com/:b:/g/personal/anna_marcionetti_unil_ch/ER1_4Pjfl5lJpohOQ16ySX8Bn7GFTKlzcMoqafwdHdBi_Q?e=0XbcrF

²https://unils-my.sharepoint.com/:x:/g/personal/anna_marcionetti_unil_ch/EU2YvfoIeRBJvb7Ym8H6KZgBIAFK0RY3LVfwc8NRrNVLxA?e=wNNbNI

- Allen, G. R., Drew, J., & Fenner, D.** (2010). *Amphiprion pacificus*, a new species of anemonefish (Pomacentridae) from Fiji, Tonga, Samoa, and Wallis Island. *Aquaculture*, 16, 129-138.
- Allen, G.R.** (1991). Damsel-fishes of the world. Mergus Publishers, Melle, Germany. 271 p
- Andrew, R. L., & Rieseberg, L. H.** (2013). Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution*, 67(9), 2468-2482.
- Andrews, S.** (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Appleyard, S. A., Ward, R. D., & Grewe, P. M.** (2002). Genetic stock structure of bigeye tuna in the Indian Ocean using mitochondrial DNA and microsatellites. *Journal of fish Biology*, 60(3), 767-770.
- Arai, R.** (2011). *Fish karyotypes: a check list*. Springer Science & Business Media.
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T.** (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, 27(12), 1691-1692.
- Barrett, R. D., & Schluter, D.** (2008). Adaptation from standing genetic variation. *Trends in ecology & evolution*, 23(1), 38-44.
- Beheregaray, L. B., Cooke, G. M., Chao, N. L., & Landguth, E. L.** (2015). Ecological speciation in the tropics: insights from comparative genetic studies in Amazonia. *Frontiers in genetics*, 5, 477.
- Bertrand, J. A., Borsa, P., & Chen, W. J.** (2017). Phylogeography of the sergeants *Abudefduf sexfasciatus* and *A. vaigiensis* reveals complex introgression patterns between two widespread and sympatric Indo-West Pacific reef fishes. *Molecular Ecology*, 26(9), 2527-2542.
- Bourgeois, Y. X., Bertrand, J. A., Delahaie, B., Holota, H., Thébaud, C., & Milá, B.** (2020). Differential divergence in autosomes and sex chromosomes is associated with intra-island diversification at a very small spatial scale in a songbird lineage. *Molecular Ecology*, 29(6), 1137-1153.
- Braasch, J., Barker, B. S., & Dlugosch, K. M.** (2019). Expansion history and environmental suitability shape effective population size in a plant invasion. *Molecular ecology*, 28(10), 2546-2558.
- Branco, S., Carpentier, F., de la Vega, R. C. R., Badouin, H., Snirc, A., Le Prieur, S., ... & Giraud, T.** (2018). Multiple convergent supergene evolution events in mating-type chromosomes. *Nature communications*, 9(1), 1-13.
- Brelsford, A., Toews, D. P., & Irwin, D. E.** (2017). Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. *Proceedings of the Royal Society B: Biological Sciences*, 284(1866), 20171106.
- Burri, R.** (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evolution Letters*, 1(3), 118-131.
- Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., ... & Hogner, S.** (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome research*, 25(11), 1656-1665.
- Buston, P. M., & García, M. B.** (2007). An extraordinary life span estimate for the clown anemonefish *Amphiprion percula*. *Journal of Fish Biology*, 70(6), 1710-1719.
- 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. *Scientific reports*, 8(1), 1-12.
- Clarkson, C. S., Weetman, D., Essandoh, J., Yawson, A. E., Maslen, G., Manske, M., ... & Donnelly, M. J.**

- (2014). Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nature communications*, 5(1), 1-10.
- Cowman, P. F., & Bellwood, D. R.** (2011). Coral reefs as drivers of cladogenesis: expanding coral reefs, cryptic extinction events, and the development of biodiversity hotspots. *Journal of evolutionary biology*, 24(12), 2543-2562.
- Cowman, P. F., & Bellwood, D. R.** (2013). The historical biogeography of coral reef fishes: global patterns of origination and dispersal. *Journal of biogeography*, 40(2), 209-224.
- Coyne, J.A., Orr, H.A.** (2004) Speciation. Sunderland, MA: Sinauer Assoc. 545 p.
- Cruickshank, T. E., & Hahn, M. W.** (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology*, 23(13), 3133-3157.
- Cutter, A. D., & Payseur, B. A.** (2013). Genomic signatures of selection at linked sites: unifying the disparity among species. *Nature Reviews Genetics*, 14(4), 262-274.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & McVean, G.** (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.
- Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., ... & Salazar, C.** (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, 487(7405), 94.
- Delrieu-Trottin, E., Mona, S., Maynard, J., Neglia, V., Veuille, M., & Planes, S.** (2017). Population expansions dominate demographic histories of endemic and widespread Pacific reef fishes. *Scientific reports*, 7(1), 1-13.
- Doebeli, M., & Dieckmann, U.** (2003). Speciation along environmental gradients. *Nature*, 421(6920), 259-264.
- Duchen, P., & Salamin, N.** (2020). A cautionary note on the use of haplotype callers in Phylogenomics. bioRxiv doi: 10.1101/2020.06.10.145011
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M.** (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, 28(8), 2239-2252.
- Eckert, C. G., Samis, K. E., & Loughheed, S. C.** (2008). Genetic variation across species' geographical ranges: the central–marginal hypothesis and beyond. *Molecular ecology*, 17(5), 1170-1188.
- Ellegren, H., & Sheldon, B. C.** (2008). Genetic basis of fitness differences in natural populations. *Nature*, 452(7184), 169-175.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., Foll, M.** (2013). Robust demographic inference from genomic and SNP data. *PLoS genetics*, 9(10).
- Fautin, D.G., Allen, G.R.** (1997). Anemonefishes and their host sea anemones. Perth: Western Australian Museum.
- Frédérich, B., Sorenson, L., Santini, F., Slater, G. J., & Alfaro, M. E.** (2013). Iterative ecological radiation and convergence during the evolutionary history of damselfishes (Pomacentridae). *The American Naturalist*, 181(1), 94-113.
- Fricke, H. W.** (1979). Mating system, resource defence and sex change in the anemonefish *Amphiprion akallopisos*. *Zeitschrift für Tierpsychologie*, 50(3), 313-326.
- Funk, D. J., Nosil, P., & Etges, W. J.** (2006). Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences*, 103(9), 3209-3213.
- Gagnaire, P. A., Pavey, S. A., Normandeau, E., & Bernatchez, L.** (2013). The genetic architecture of

reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. *Evolution*, 67(9), 2483-2497.

- Gainsford, A., Van Herwerden, L., & Jones, G. P.** (2015). Hierarchical behaviour, habitat use and species size differences shape evolutionary outcomes of hybridization in a coral reef fish. *Journal of evolutionary biology*, 28(1), 205-222.
- Gavrilets, S.** (2003). Perspective: models of speciation: what have we learned in 40 years?. *Evolution*, 57(10), 2197-2215.
- Gavrilets, S.** (2004). *Fitness landscapes and the origin of species* (MPB-41). Princeton University Press.
- Gavrilets, S., & Losos, J. B.** (2009). Adaptive radiation: contrasting theory with data. *Science*, 323(5915), 732-737.
- Godoy, F. M. D. R., Lenzi, M., Ferreira, B. H. D. S., Silva, L. V. D., Zanella, C. M., & Paggi, G. M.** (2018). High genetic diversity and moderate genetic structure in the self-incompatible, clonal *Bromelia hieronymi* (Bromeliaceae). *Botanical Journal of the Linnean Society*, 187(4), 672-688.
- Gompert, Z., Fordyce, J. A., Forister, M. L., Shapiro, A. M., & Nice, C. C.** (2006). Homoploid hybrid speciation in an extreme habitat. *Science*, 314(5807), 1923-1925.
- Grant, P. R.** (1986). *Ecology and evolution of Darwin's finches*. Princeton University Press.
- Grant, P. R., & Grant, B. R.** (2010). Conspecific versus heterospecific gene exchange between populations of Darwin's finches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 365(1543), 1065-1076.
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Hansen, N. F.** (2010). A draft sequence of the Neandertal genome. *Science*, 328(5979), 710-722.
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O.** (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic biology*, 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 research*, 41(13), e129-e129.
- He, S., Planes, S., Sinclair-Taylor, T. H., & Berumen, M. L.** (2019). Diagnostic nuclear markers for hybrid Nemos in Kimbe Bay, PNG-*Amphiprion chrysopterus* x *Amphiprion sandaracinos* hybrids. *Marine Biodiversity*, 49(3), 1261-1269.
- Hedrick, P. W.** (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular ecology*, 22(18), 4606-4618.
- Hermisson, J., & Pennings, P. S.** (2005). Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169(4), 2335-2352.
- Hey, J.** (2010). Isolation with migration models for more than two populations. *Molecular biology and evolution*, 27(4), 905-920.
- Hilário, A., Metaxas, A., Gaudron, S. M., Howell, K. L., Mercier, A., Mestre, N. C., ... & Young, C.** (2015). Estimating dispersal distance in the deep sea: challenges and applications to marine reserves. *Frontiers in Marine Science*, 2, 6.
- Hubert, N., Meyer, C. P., Bruggemann, H. J., Guerin, F., Komeno, R. J., Espiau, B., ... & Planes, S.** (2012). Cryptic diversity in Indo-Pacific coral-reef fishes revealed by DNA-barcoding provides new support to the

centre-of-overlap hypothesis. *PLoS one*, 7(3).

- Hudson, R. R., Boos, D. D., & Kaplan, N. L.** (1992). A statistical test for detecting geographic subdivision. *Molecular biology and evolution*, 9(1), 138-151.
- Huyghe, F., & Kochzius, M.** (2017). Highly restricted gene flow between disjunct populations of the skunk clownfish (*Amphiprion akallopisos*) in the Indian Ocean. *Marine Ecology*, 38(1), e12357.
- Huyghe, F., & Kochzius, M.** (2018). Sea surface currents and geographic isolation shape the genetic population structure of a coral reef fish in the Indian Ocean. *PloS one*, 13(3), e0193825.
- Jones, F. C., Chan, Y. F., Schmutz, J., Grimwood, J., Brady, S. D., Southwick, A. M., ... & Kingsley, D. M.** (2012b). A genome-wide SNP genotyping array reveals patterns of global and repeated species-pair divergence in sticklebacks. *Current biology*, 22(1), 83-90.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... & Kingsley, D. M.** (2012a). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55-61.
- Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., ... & Jiggins, C. D.** (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, 477(7363), 203-206.
- Joshi, N. A., & Fass, J. N.** (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software]. Available at <https://github.com/najoshi/sickle>.
- Katoh, K., & Standley, D. M.** (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4), 772-780.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Thierer, T.** (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
- Keller, I., Wagner, C. E., Greuter, L., Mwaiko, S., Selz, O. M., Sivasundar, A., ... & Seehausen, O.** (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular ecology*, 22(11), 2848-2863.
- Kirkpatrick, M.** (2010). How and why chromosome inversions evolve. *PLoS biology*, 8(9), e1000501.
- Kirkpatrick, M., & Ravigné, V.** (2002). Speciation by natural and sexual selection: models and experiments. *the american naturalist*, 159(S3), S22-S35.
- Kousathanas, A., Leuenberger, C., Link, V., Sell, C., Burger, J., & Wegmann, D.** (2017). Inferring heterozygosity from ancient and low coverage genomes. *Genetics*, 205(1), 317-332.
- Kumar, S., Stecher, G., Suleski, M., & Heddes, S. B.** (2017). TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular biology and evolution*, 34(7), 1812-1819
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics*, 25(14), 1754-1760.
- Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., ... & Kronforst, M. R.** (2014). Doublesex is a mimicry supergene. *Nature*, 507(7491), 229-232.
- Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., ... & Burke, T.** (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature genetics*, 48(1), 79-83.
- Lai, Y. T., Yeung, C. K., Omland, K. E., Pang, E. L., Hao, Y., Liao, B. Y., ... & Hung, H. Y.** (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences*, 116(6), 2152-2157.

- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., ... & Grant, B. R.** (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, *518*(7539), 371-375.
- Lanier, H. C., Massatti, R., He, Q., Olson, L. E., & Knowles, L. L.** (2015). Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology*, *24*(14), 3688-3705.
- Lawniczak, M. K. N., Emrich, S. J., Holloway, A. K., Regier, A. P., Olson, M., White, B., ... & Besansky, N. J.** (2010). Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science*, *330*(6003), 512-514.
- Lehmann, R., Lightfoot, D. J., Schunter, C., Michell, C. T., Ohyanagi, H., Mineta, K., ... & Gojobori, T.** (2018). Finding Nemo's Genes: A chromosome-scale reference assembly of the genome of the orange clownfish *Amphiprion percula*. *Molecular ecology resources*.
- Leis, J. M., & McCormick, M. I.** (2002). The biology, behavior, and ecology of the pelagic, larval stage of coral reef fishes. *Coral reef fishes: dynamics and diversity in a complex ecosystem*, 171-199.
- Lester, S. E., & Ruttenberg, B. I.** (2005). The relationship between pelagic larval duration and range size in tropical reef fishes: a synthetic analysis. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1563), 585-591.
- 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., ... & Durbin, R.** (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.
- Li, J., Chen, X., Kang, B., & Liu, M.** (2015). Mitochondrial DNA genomes organization and phylogenetic relationships analysis of eight anemonefishes (Pomacentridae: Amphiprioninae). *PLoS One*, *10*(4), e0123894.
- Linck, E., & Battey, C. J.** (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, *19*(3), 639-647.
- Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., & Wegmann, D.** (2017). ATLAS: analysis tools for low-depth and ancient samples. *BioRxiv*, 105346.
- Litsios, G., & Salamin, N.** (2014). Hybridization and diversification in the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, *14*(1), 245.
- Litsios, G., Pearman, P. B., Lanterbecq, D., Tolou, N., & Salamin, N.** (2014). The radiation of the clownfishes has two geographical replicates. *Journal of biogeography*, *41*(11), 2140-2149.
- Litsios, G., Pellissier, L., Forest, F., Lexer, C., Pearman, P. B., Zimmermann, N. E., & Salamin, N.** (2012b). Trophic specialization influences the rate of environmental niche evolution in damselfishes (Pomacentridae). *Proceedings of the Royal Society B: Biological Sciences*, *279*(1743), 3662-3669.
- Litsios, G., Sims, C. A., Wüest, R. O., Pearman, P. B., Zimmermann, N. E., & Salamin, N.** (2012a). Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC evolutionary biology*, *12*(1), 212.
- Losos, J. B., Arnold, S. J., Bejerano, G., Brodie III, E. D., Hibbett, D., Hoekstra, H. E., ... & Petrov, D. A.** (2013). Evolutionary biology for the 21st century. *PLoS Biology*, *11*(1).
- Mallet, J.** (2007). Hybrid speciation. *Nature*, *446*(7133), 279-283.
- Mallet, J., McMillan, W. O., & Jiggins, C. D.** (1998). Mimicry and warning color at the boundary between races

and species. *Endless forms: species and speciation*, 390-403.

- Marcionetti, A., Rossier, V., Bertrand, J. A., Litsios, G., & Salamin, N.** (2018). First draft genome of an iconic clownfish species (*Amphiprion frenatus*). *Molecular ecology resources*, 18(5), 1092-1101.
- Martin, M.** (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17(1), pp-10.
- Martin, S. H., & Van Belleghem, S. M.** (2017). Exploring evolutionary relationships across the genome using topology weighting. *Genetics*, 206(1), 429-438.
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... & Jiggins, C. D.** (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome research*, 23(11), 1817-1828.
- Martin, S. H., Davey, J. W., & Jiggins, C. D.** (2015). Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *Molecular biology and evolution*, 32(1), 244-257.
- Mayr, E.** (2013). *Animal species and evolution*. Harvard University Press.
- Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O.** (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications*, 8(1), 1-11.
- Meisner, J., & Albrechtsen, A.** (2018). Inferring population structure and admixture proportions in low depth NGS data. *BioRxiv*, 302463.
- Michel, A. P., Sim, S., Powell, T. H., Taylor, M. S., Nosil, P., & Feder, J. L.** (2010). Widespread genomic divergence during sympatric speciation. *Proceedings of the National Academy of Sciences*, 107(21), 9724-9729.
- Moritz, C., Patton, J. L., Schneider, C. J., & Smith, T. B.** (2000). Diversification of rainforest faunas: an integrated molecular approach. *Annual review of ecology and systematics*, 31(1), 533-563.
- Moyer, J. T., & Nakazono, A.** (1978). Protandrous hermaphroditism in six species of the anemonefish genus *Amphiprion* in Japan. *Japanese Journal of Ichthyology*, 25(2), 101-106.
- Nachman, M. W.** (2002). Variation in recombination rate across the genome: evidence and implications. *Current opinion in genetics & development*, 12(6), 657-663.
- Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., ... & Jiggins, C. D.** (2013). Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular ecology*, 22(3), 814-826.
- Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., ... & Mallet, J.** (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 343-353.
- Natri, H. M., Merilä, J., & Shikano, T.** (2019). The evolution of sex determination associated with a chromosomal inversion. *Nature communications*, 10(1), 1-13.
- Noor, M. A., & Bennett, S. M.** (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, 103(6), 439-444.
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D.** (2009). Divergent selection and heterogeneous genomic divergence. *Molecular ecology*, 18(3), 375-402.
- Nosil, P., Harmon, L. J., & Seehausen, O.** (2009a). Ecological explanations for (incomplete) speciation. *Trends in ecology & evolution*, 24(3), 145-156.

- Ollerton, J., McCollin, D., Fautin, D. G., & Allen, G. R.** (2007). Finding NEMO: nestedness engendered by mutualistic organization in anemonefish and their hosts. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609), 591-598.
- Ortego, J., Gugger, P. F., & Sork, V. L.** (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist*, 218(2), 804-818.
- Parchman, T. L., Gompert, Z., Braun, M. J., Brumfield, R. T., McDonald, D. B., Uy, J. A. C., ... & Buerkle, C. A.** (2013). The genomic consequences of adaptive divergence and reproductive isolation between species of manakins. *Molecular ecology*, 22(12), 3304-3317.
- Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., ... & Jiggins, C. D.** (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet*, 8(6), e1002752.
- Payseur, B. A.** (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular ecology resources*, 10(5), 806-820.
- Pennisi, E.** (2014). Disputed islands. *Science* 345: 611–613.
- Perea, S., Vukić, J., Šanda, R., & Doadrio, I.** (2016). Ancient mitochondrial capture as factor promoting mitonuclear discordance in freshwater fishes: a case study in the genus *Squalius* (Actinopterygii, Cyprinidae) in Greece. *PLoS One*, 11(12), e0166292.
- Pereira, R. J., Martínez-Solano, I., & Buckley, D.** (2016). Hybridization during altitudinal range shifts: nuclear introgression leads to extensive cyto-nuclear discordance in the fire salamander. *Molecular Ecology*, 25(7), 1551-1565.
- Peter, B. M., & Slatkin, M.** (2013). Detecting range expansions from genetic data. *Evolution*, 67(11), 3274-3289.
- Pickrell, J., & Pritchard, J.** (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*, 1-1.
- Pierce, A. A., Zalucki, M. P., Bangura, M., Udawatta, M., Kronforst, M. R., Altizer, S., ... & de Roode, J. C.** (2014). Serial founder effects and genetic differentiation during worldwide range expansion of monarch butterflies. *Proceedings of the Royal Society B: Biological Sciences*, 281(1797), 20142230.
- Planes, S., Jones, G. P., & Thorrold, S. R.** (2009). Larval dispersal connects fish populations in a network of marine protected areas. *Proceedings of the National Academy of Sciences*, 106(14), 5693-5697.
- Pushparaj, A.** (2010). Intra specific hybridization between *Amphiprion sebae* and *A. polymnus* under captive conditions. *Int J Biological Technology*, 1, 52-56.
- Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., ... & Alfaro, M. E.** (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559(7714), 392-395.
- Rabosky, D. L., Santini, F., Eastman, J., Smith, S. A., Sidlauskas, B., Chang, J., & Alfaro, M. E.** (2013). Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature communications*, 4(1), 1-8.
- Ragionieri, L., Fratini, S., Vannini, M., & Schubart, C. D.** (2009). Phylogenetic and morphometric differentiation reveal geographic radiation and pseudo-cryptic speciation in a mangrove crab from the Indo-West Pacific. *Molecular Phylogenetics and Evolution*, 52(3), 825-834.
- Rettelbach, A., Nater, A., & Ellegren, H.** (2019). How linked selection shapes the diversity landscape in *Ficedula* flycatchers. *Genetics*, 212(1), 277-285.

- Revell, L. J.** (2012). phytools: an R package for phylogenetic comparative biology (and other things). *Methods in ecology and evolution*, 3(2), 217-223.
- Rice, W. R., & Hostert, E. E.** (1993). Laboratory experiments on speciation: what have we learned in 40 years?. *Evolution*, 47(6), 1637-1653.
- Ridgway, T., & Sampayo, E. M.** (2005). Population genetic status of the western Indian Ocean: what do we know?. *Western Indian Ocean Journal of Marine Science*, 4(1), 1-10.
- Rundle, H. D., & Nosil, P.** (2005). Ecological speciation. *Ecology letters*, 8(3), 336-352.
- Sagonas, K., Runemark, A., Antoniou, A., Lymberakis, P., Pafilis, P., Valakos, E. D., ... & Hansson, B.** (2019). Selection, drift, and introgression shape MHC polymorphism in lizards. *Heredity*, 122(4), 468-484.
- Salinas-de-León, P., Jones, T., & Bell, J. J.** (2012). Successful determination of larval dispersal distances and subsequent settlement for long-lived pelagic larvae. *PLoS one*, 7(3), e32788.
- Salzburger, W., Baric, S., & Sturmbauer, C.** (2002). Speciation via introgressive hybridization in East African cichlids?. *Molecular Ecology*, 11(3), 619-625.
- Santini, S., & Polacco, G.** (2006). Finding Nemo: molecular phylogeny and evolution of the unusual life style of anemonefish. *Gene*, 385, 19-27.
- Schluter, D.** (1996). Ecological speciation in postglacial fishes. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 351(1341), 807-814.
- Schluter, D.** (2000) *The ecology of adaptive radiation*. Oxford University Press, Oxford, UK.
- Schluter, D.** (2001). Ecology and the origin of species. *Trends in ecology & evolution*, 16(7), 372-380.
- Schluter, D.** (2009). Evidence for ecological speciation and its alternative. *Science*, 323(5915), 737-741.
- Schouten, M. W., de Ruijter, W. P., Van Leeuwen, P. J., & Ridderinkhof, H.** (2003). Eddies and variability in the Mozambique Channel. *Deep Sea Research Part II: Topical Studies in Oceanography*, 50(12-13), 1987-2003.
- Schulte, U., Veith, M., Mingo, V., Modica, C., & Hochkirch, A.** (2013). Strong genetic differentiation due to multiple founder events during a recent range expansion of an introduced wall lizard population. *Biological Invasions*, 15(12), 2639-2649.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... & Brelsford, A.** (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3), 176-192.
- Shafer, A. B., & Wolf, J. B.** (2013). Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, 16(7), 940-950.
- Shanks, A. L.** (2009). Pelagic larval duration and dispersal distance revisited. *The biological bulletin*, 216(3), 373-385.
- Shanks, A. L., Grantham, B. A., & Carr, M. H.** (2003). Propagule dispersal distance and the size and spacing of marine reserves. *Ecological applications*, 13(sp1), 159-169.
- Simpson, S. D., Harrison, H. B., Claereboudt, M. R., & Planes, S.** (2014). Long-distance dispersal via ocean currents connects Omani clownfish populations throughout entire species range. *PLoS One*, 9(9), e107610.
- Smith, J., & Kronforst, M. R.** (2013). Do Heliconius butterfly species exchange mimicry alleles?. *Biology letters*, 9(4), 20130503.
- Soria-Carrasco, V., Gompert, Z., Comeault, A. A., Farkas, T. E., Parchman, T. L., Johnston, J. S., ... & Egan, S. P.** (2014). Stick insect genomes reveal natural selection's role in parallel speciation. *Science*, 344(6185), 738-742.

- Stamatakis, A.** (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312-1313.
- Steinke, D., Zemplak, T. S., & Hebert, P. D.** (2009). Barcoding Nemo: DNA-based identifications for the ornamental fish trade. *PLoS one*, 4(7).
- Stevison, L. S., Hoehn, K. B., & Noor, M. A.** (2011). Effects of inversions on within-and between-species recombination and divergence. *Genome biology and evolution*, 3, 830-841.
- Strasburg, J. L., Sherman, N. A., Wright, K. M., Moyle, L. C., Willis, J. H., & Rieseberg, L. H.** (2012). What can patterns of differentiation across plant genomes tell us about adaptation and speciation?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 364-373.
- Svardal, H., Quah, F. X., Malinsky, M., Ngatunga, B. P., Miska, E. A., Salzburger, W., ... & Durbin, R.** (2020). Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Molecular biology and evolution*, 37(4), 1100-1113.
- Talla, V., Kalsoom, F., Shipilina, D., Marova, I., & Backström, N.** (2017). Heterogeneous patterns of genetic diversity and differentiation in European and Siberian chiffchaff (*Phylloscopus collybita abietinus*/*P. tristis*). *G3: Genes, Genomes, Genetics*, 7(12), 3983-3998.
- Tao, Y., Li, J. L., Liu, M., & Hu, X. Y.** (2016). Complete mitochondrial genome of the orange clownfish *Amphiprion percula* (Pisces: Perciformes, Pomacentridae). *Mitochondrial DNA Part A*, 27(1), 324-325.
- Thibert-Plante, X., & Hendry, A. P.** (2011). Factors influencing progress toward sympatric speciation. *Journal of evolutionary biology*, 24(10), 2186-2196.
- Timm, J., Figiel, M., & Kochzius, M.** (2008). Contrasting patterns in species boundaries and evolution of anemonefishes (Amphiprioninae, Pomacentridae) in the centre of marine biodiversity. *Molecular Phylogenetics and Evolution*, 49(1), 268-276.
- Toews, D. P., & Brelsford, A.** (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907-3930.
- Turner, S. D.** (2014). qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*, 005165.
- Turner, T. L., & Hahn, M. W.** (2010). Genomic islands of speciation or genomic islands and speciation?. *Molecular ecology*, 19(5), 848-850.
- van Doorn, G. S., Edelaar, P., & Weissing, F. J.** (2009). On the origin of species by natural and sexual selection. *Science*, 326(5960), 1704-1707.
- Wang, X., Maher, K. H., Zhang, N., Que, P., Zheng, C., Liu, S., ... & Liu, Y.** (2019). Demographic histories and genome-wide patterns of divergence in incipient species of shorebirds. *Frontiers in genetics*, 10, 919.
- Whitney, K. D., Randell, R. A., & Rieseberg, L. H.** (2010). Adaptive introgression of abiotic tolerance traits in the sunflower *Helianthus annuus*. *New Phytologist*, 187(1), 230-239.
- Wickham, H.** (2016). *ggplot2: elegant graphics for data analysis*. Springer.
- Wielgoss, S., Gilabert, A., Meyer, A., & Wirth, T.** (2014). Introgressive hybridization and latitudinal admixture clines in North Atlantic eels. *BMC evolutionary biology*, 14(1), 61.
- Wielgoss, S., Gilabert, A., Meyer, A., & Wirth, T.** (2014). Introgressive hybridization and latitudinal admixture clines in North Atlantic eels. *BMC evolutionary biology*, 14(1), 61.
- Wolf, J. B., & Ellegren, H.** (2017). Making sense of genomic islands of differentiation in light of speciation.

Nature Reviews Genetics, 18(2), 87.

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics*, 19(6), 153.

Zinzow-Kramer, W. M., Horton, B. M., McKee, C. D., Michaud, J. M., Tharp, G. K., Thomas, J. W., ... & Maney, D. L. (2015). Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes, Brain and Behavior*, 14(8), 641-654.

DISCUSSION AND PERSPECTIVES

With the advent of competitive sequencing technologies, the central question of how species diversify has increasingly been approached from a genomic perspective (Seehausen et al., 2014; Campbell et al., 2018; Marques et al., 2019). Nevertheless, we are still far from comprehending the interplay between intrinsic genomic architecture and extrinsic biological and ecological factors in determining the diversification potential of organisms. Consequently, further studies of lineages across the full breadth of the Tree of life are necessary to achieve a comprehensive understanding of how Earth's biodiversity is created (Seehausen et al., 2014; Campbell et al., 2018). In this sense, clownfishes represent a fascinating system. With their mutualism with sea anemones viewed as the trigger of their rapid diversification, this group represents a rare example of potential adaptive radiation in the marine environment (Litsios et al., 2012; Bowen et al., 2020). In my thesis, I created the genomic resources (Chapters 1 and 2) to study the genetic mechanisms behind this mutualistic interaction (Chapter 2). I then combined comparative (Chapter 3) and population genomic (Chapter 4) approaches to provide the first insights on the genomic architecture underlying this diversification.

Building up the genomic resources for clownfish

In chapter 1, I created the first genomic resource of clownfish. Using low-coverage PacBio long reads and high-coverage Illumina short reads, I assembled and annotated the genome of the tomato clownfish (*Amphiprion frenatus*). I obtained a reference genome of good quality comparable to the genomic resources obtained at the time. **In chapter 2**, I extended this resource with the assembly and annotation of nine additional closely related but ecologically divergent clownfish species and the damselfish *P. moluccensis*. I assessed the possibility of decreasing the economic and computational efforts required for *de novo* assembly strategies by developing and evaluating a reference-guided approach on *A. ocellaris*. This method resulted in a genome assembly equivalent to the one obtained with the *de novo* strategy, and therefore, I applied it to the additional species and obtained genome assemblies and annotations of overall good quality.

These genomic resources allowed me to study the genetic mechanisms behind the clownfish-sea anemone mutualism (Chapter 2) and the genomic architecture of clownfish radiation (Chapter 3). Initially, I also employed the *A. frenatus* reference genome to investigate the diversification of the skunk complex (Chapter 4). However, I quickly realized that analyses such as genome scans on fragmented assemblies are challenging, both in the implementation

and interpretation of the results. The concurrent publication and availability of a chromosome-scale assembly for *A. percula* (Lehmann et al., 2019) allowed me to refine my analyses on the skunk complex, helping me realize the importance of sharing resources and data in the scientific community to boost research.

While the access to one chromosome-level reference is valuable, it is not sufficient for comparative analyses investigating, for instance, structural variations, which widely contribute to adaptation and speciation (e.g., Wellenreuther et al., 2019; Mérot et al., 2020; Zhang et al., 2021). The results I obtained in my thesis suggest genomic inversions in clownfishes (on chromosome 18, Chapters 3 and 4), and future studies should advance towards the characterization of these regions and the evaluation of their role in clownfish diversification (further discussed below). For this, the acquisition of genomic data spanning the entire length of the inversions is necessary. Since we have candidate inverted regions, an approach could consist of baits' design followed by targeted long-read sequencing (e.g., Miller et al., 2020). However, at the moment, these methods are principally developed for medium-sized targets (typically, up to 20 kb, Wenger et al., 2019), and their application for the megabases-long candidate regions in clownfishes could be challenging. As third-generation sequencing keeps progressing in throughput, accuracy, and cost reduction (Amarasinghe et al., 2020), it is increasingly conceivable to elude targeted sequencing and directly perform long-read sequencing of whole genomes to obtain high-quality chromosomal-level assemblies of multiple clownfish species.

Multiple species' chromosomal-scale assemblies not only represent a valuable resource to study genomic inversions, but they also permit to better evaluate the extent of additional intra- and inter-specific structural variation, such as chromosomal rearrangements (e.g., de Sousa et al., 2021; Jackson et al., 2021), copy number variants (e.g., Faber-Hammond et al., 2019; North et al., 2020), and indels (McGee et al., 2020). These variations are often involved in the adaptation and speciation processes (Wellenreuther et al., 2019; Zhang et al., 2021), but until recently, they were difficult to identify given the fragmented nature of assemblies and the reference-based approaches. I am confident that in the near future, chromosomal-level, unbiased and accurately annotated genomes of a multitude of organisms will be easily accessible (see Rhie et al., 2021), further boosting the study of the genomic basis of the diversification process.

The mechanisms underlying the interaction with sea anemones

In chapter 2, I studied the genetic mechanisms underlying the clownfish mutualistic interaction with sea anemones. I principally investigated patterns of positive selection on coding sequences, and I identified 17 genes that presented signals of positive selection at the beginning of clownfish diversification. Among them, I identified genes with functions associated with *N*-acetylated sugars, which are likely involved in triggering the release of toxins by sea anemones. I verified the absence of false-positive results and demonstrated the expression of these genes in the skin of clownfishes, corresponding to the surface of interaction with the sea anemones. Overall, in this chapter, I discovered candidate genes likely to have played a role in the evolution of mutualism with sea anemones.

The mechanisms behind clownfishes' ability to live within toxic sea anemones are still far from being entirely elucidated (Mebs et al., 2009, da Silva & Nedosyko, 2016). The most accredited view is that clownfishes secrete a mucus acting as molecular mimicry and preventing sea anemones from recognizing them, thus avoiding the subsequent release of toxins (Mebs et al., 2009; da Silva & Nedosyko, 2016). The results obtained in this chapter and the findings that clownfish mucus composition differs from other teleosts by its glycoproteins content (e.g., Balamurugan et al., 2014, Abdullah & Saad, 2015) reinforce this hypothesis. The need for some species of an acclimation period - which was also hypothesized to corresponds to the time necessary for clownfishes to cover themselves with sea anemones mucus as camouflage (see Mebs et al., 2009) - then reflects the necessity of clownfishes to remodel their mucus content and to increase its production following the first interaction with sea anemones (see also Balamurugan et al., 2014). This hypothesis does not exclude the reinforcement of clownfish protection through their covering of sea anemones mucus but further supports the innate nature of clownfishes' ability to live within sea anemones.

In this sense, during my thesis, I tried to explore changes in gene expression in clownfish skin during the acclimation to sea anemones, and thus likely associated with mutualism. Collaborators at the OOB (Observatoire Océanologique de Banyuls-sur-Mer) set up the experimental conditions and extracted RNA from the skin of naïve (i.e., never being in contact with sea anemones) *A. ocellaris* individuals before, during, and after the contact with the sea anemones. I performed transcriptomic analysis on these samples, but the high variability in gene expression within each condition (time points) prevented us from finding significant differentially expressed genes associated with the interaction with sea anemones.

Considering the presence of mRNA in fish mucus (Ren et al., 2015; Brinchmann et al., 2016; Parida et al., 2018) and the little understanding that we have of its role (Brinchmann et al., 2016), I evaluated the possibility of analyzing it in the context of acclimation. Compared to skin tissues, the approach on mucus mRNA has the advantage of being the direct surface of interaction with sea anemones. Additionally, it avoids fish euthanization, allowing the analysis of individuals at different time points and reducing variability. I set up a pilot experiment to confirm the presence of mRNA in clownfish mucus and evaluate the possibility of characterizing it through sequencing (see Annex 5). The preliminary results - obtained with the aid of a Master's student - demonstrated the feasibility of transcriptomic analysis on mucus mRNA, which appears enriched with genes involved in immunity responses. We also noticed that phylogenetically distant specialist species presented a more similar transcriptome than their closely related generalists.

In order to advance our understanding of the molecular basis of this mutualism, future studies should continue in this direction, investigating gene expression (through transcriptomics), protein content (through proteomics, e.g., Fæste et al., 2020), and metabolites composition (through metabolomics, e.g., Reverter et al., 2017) in clownfish mucus, and comparing it with the composition of closely related damselfish species not interacting with the sea anemones. Furthermore, changes in mucus molecular content during clownfish acclimation to sea anemones should be investigated in an experimental setup, as we initially performed for the skin transcriptomic analysis. The differences in proteins and metabolites before and after the interaction could provide candidate molecules likely remodeling the molecular mimicry of clownfish mucus. Similarly, changes in gene expression could allow a link with clownfish genomics, providing a direction for further investigations of genomic changes between clownfishes and other non-mutualistic teleosts associated with the evolution of mutualism.

This approach should also be applied to multiple clownfish species presenting contrasting preferences in host usage (such as specialist and generalist species). Hence, the interspecific comparison of the mucus composition could allow the characterization of the different responses to distinct sea anemones interactions, potentially identifying the disparities in mucus molecular traits associated with ecological divergence (i.e., clownfish functional traits for sea anemones interaction, see also below). It is worth mentioning that experimental setups of the acclimation of naïve clownfish have already been adopted to analyze the transformation of skin microbiomes (Roux et al., 2019; Émie et al., 2021) or the differences in mucus specific

glycoproteins (Balamurugan et al., 2014). However, they have never been employed on multiple species or for a comprehensive analysis aiming to characterize the full extent of changes in gene expression, protein content, and metabolite composition occurring during acclimation.

Coupling these experiments with the concurrent acquisition of sea anemones samples before and after their interaction with clownfish will also be helpful to evaluate to which extent sea anemones respond to their residents (as observed for the microbiome, Titus et al., 2020) and thus understand whether this mutualism requires adaptations from both partners. Finally, analyzing the mucus composition of other fishes that usually avoid sea anemone tentacles but occasionally make contact with them (Randall & Fautin, 2002) could provide insights on whether similar mechanisms of protection have evolved multiple independent times in fish. When our knowledge on the mechanism underlying this interaction will be significantly extended, functional testing of candidate genes (such as the positively selected genes I identified in this chapter) through CRISPR/Cas9 gene-editing techniques, which have been developed for clownfish (Mitchell et al., 2020), could then be conceived.

The genomics substrate of clownfish diversification

In chapter 3, I studied the genomic architecture of clownfish adaptive radiation. Are clownfish characterized by a higher genomic diversity than other non-radiating lineages that may have facilitated their rapid diversification? I identified bursts of transposable elements and detected a general acceleration in coding sequence evolution in clownfishes. I also detected topological inconsistencies likely deriving from hybridization events that occurred during the group diversification, with a particular signal on chromosome 18 likely maintained by genomic inversions. I also discovered signatures of positive selection associated with the diversification in 5.4% of the analyzed genes. Among them, I identified genes with functions affecting social behavior that are likely to have participated in the evolution of clownfish's size-based dominance hierarchy. Altogether, these results demonstrate that clownfishes share some genomic characteristics with other adaptive radiations (Dasmahapatra et al., 2012; Jones et al., 2012; Brawand et al., 2014; Lamichhaney et al., 2015; Berner & Salzburger, 2015) - that potentially facilitated their diversification - and add up to the understanding of the link between genomic content and the extent of diversification.

The overall acceleration in coding sequence evolution observed in the group is similar to what was previously observed in ant-plant mutualism (Rubin & Moreau, 2016). This opens the

question of whether the evolution of mutualism increases the rate of evolution in a similar way as what observed for antagonistically coevolving species (the Red Queen Hypothesis, Van Valen, 1971; Rubin & Moreau, 2016) rather than reducing it as expected in the Red King Effect theory (Bergstrom & Lachmann, 2003). Before drawing robust conclusions, it is necessary to assess the effective population sizes in clownfishes (such as with G-PhoCS; Gronau et al., 2011), as they influence the rate of coding sequence evolution (Lynch & Walsh, 2007). Nevertheless, these results suggest that the mutualistic interaction itself may increase the rate of evolution in clownfishes and show the importance of further studies of other mutualistic relationships to understand the link between mutualism and the rate of coding sequence evolution.

Other genomic features often observed in adaptive radiations are structural variants, changes in regulatory sequences (Dasmahapatra et al., 2012; Jones et al., 2012; Brawand et al., 2014; Lamichhaney et al., 2015; Berner & Salzburger, 2015), and, more recently, high levels of heterozygosity (Ronco et al., 2021). Future studies should characterize and evaluate the role of these elements in clownfish diversification. As mentioned above, chromosomal-level assemblies could facilitate the analysis of structural variants, and I will further discuss the importance of continuing the study of the inconsistencies observed on clownfish chromosome 18 in subsequent subsections. Similarly, changes in regulatory sequences - which modify gene expression - play a key role in the evolution of phenotypes such as morphology, coloration, and behavior, especially in closely related taxa (reviewed in Wray, 2007; Stern & Orgogozo, 2008). For these reasons, during my thesis, I created a Master project (Annex 4) aiming to identify clownfish conserved non-coding regions - likely containing regulatory sequences - and evaluate their evolution (pipeline as in Brawand et al., 2014). However, given the overall fragmentation of the used assemblies and the generally challenging identification of regulatory elements *in silico* (Worsley-Hunt et al., 2011), the obtained results were not entirely meaningful. Since then, new methods based on machine learning have been developed and could facilitate the identification of *cis*-regulatory elements *in silico*. For instance, the DeepArk model (Cofer et al., 2020) is a promising tool, as it managed to accurately identify ChIP-Seq and ATAC-seq peaks in the *Oryzias latipes* (Medaka) genome, training the model on data from *Danio rerio*. This method should be tested on clownfishes to identify and evaluate the extent of *cis*-regulatory changes (through subsequent divergence analysis, as in Brawand et al., 2014) observed during their diversification, without investing in ChIP-seq experiments, whole-genome expression arrays or extensive transcriptomics approaches.

While it is important to characterize the genomic features described above, it is also essential to evaluate the extent of hybridization in clownfish evolution. Indeed, it has been proposed that the key genomic characteristic that drives rapid diversification is likely to be the access to ancient genetic variation through gene flow (Berner & Salzburger, 2015). Since the first genomic studies on adaptive radiation, evidence for this hypothesis has been accumulating (e.g., Han et al., 2017; Bassham et al., 2018; McGee et al., 2020; Kozak et al., 2021). It is also worth noticing that a recent study on the genomic features of Lake Tanganyika's cichlids showed that the only genomic characteristic correlating with species richness was heterozygosity, which was in turn explained by the levels of gene flow within the analyzed tribes coupled with a reduced gene flow among them (Ronco et al., 2021).

In clownfishes, I detected topological inconsistencies across the genome. While I only considered ten clownfish species, the level of inconsistency was already rather important. Indeed, I observed disparities both in the branching of species clades and in the position of *A. perideraion*, which branched outside its clade and with the *ephippium* group on chromosome 18 (see also discussion below). These topological inconsistencies suggest that ancestral hybridization events happened during the diversification of the group. However, incomplete lineage sorting cannot yet be excluded, and future studies should investigate the origin of these disparities. A possible approach consists of applying ABBA-BABA tests and the related statistics (Green et al., 2010; Durand et al., 2011; Reich et al., 2009). For instance, with *P. moluccensis* as the outgroup, it would be possible to alternatively consider the clownfish clades as the potential donor or recipient populations, according to the topologies observed (as in Meier et al., 2017; Svardal et al., 2020). In the case of a significant hybridization signal, candidate regions of introgression between clades could be investigated with f_d statistics (Martin et al., 2015). Using the reference genome of *A. percula* could considerably facilitate this approach by allowing contiguous sliding windows across the chromosomes. Furthermore, the integration of additional clownfish species in the study could permit to re-evaluate the extent of topological inconsistencies and provide additional replicates per clade in the introgression analyses. If these analyses confirm a significant gene flow between clades, we could conclude that ancestral hybridization shaped clownfish diversification. The evaluation of the content of the introgressed regions, together with a more precise definition of functional traits in clownfishes (see below), could then allow assessing whether such hybridization events are linked with ecological divergence and adaptation, fueling clownfish adaptive radiation, as observed for instance in cichlids (Meier et al., 2017; Svardal et al., 2020).

The widespread parallel evolution observed in adaptively radiating lineages is also resulting from the exchange of ancient genetic variation among populations (e.g., Cresko et al., 2004; Colosimo et al., 2005; Dasmahapatra et al., 2012; Lawson & Petren, 2017; Lewis et al., 2019). Indeed, divergent selection can act on these genetic variants, driving populations' adaptation to different ecological niches. This process can be repeated on replicated systems, resulting in the phenotypic convergence of ecologically similar species achieved by adaptation through identical genetic mechanisms that were spread by hybridizations (e.g., Cresko et al., 2004; Colosimo et al., 2005; Dasmahapatra et al., 2012; Lawson & Petren, 2017; Lewis et al., 2019). In this sense, in **chapter 3**, I also evaluated the extent of parallel evolution present in clownfishes. Did ecologically-convergent clownfishes acquire their convergent phenotypes through the same genetic mechanisms? To address this question, I used the genomic data of the five pairs of closely related but ecologically divergent species to examine the presence of hybridization events or genes with differences in evolutionary rates and selective pressures associated with ecological differentiation. While I did not find evidence of hybridization events, I detected genes with patterns of differential selective pressure associated with ecological divergence, potentially involved in the parallel adaptation to similar ecological niches.

Further studies are nevertheless needed to validate these results and provide insights into the potential adaptive nature of the genes I identified here. First, it is crucial to estimate the number of genes with parallel patterns of selection that are expected in the case of diversification without ecological divergence. Indeed, a recent study reported shared positively selected genes in 10 to 15 pairs of allopatric species of angiosperms with no apparent ecological differentiation (Dong et al., 2019). A possible approach to obtain this estimation is to acquire genomic data of five pairs of similarly closely related species but with no ecological divergence (i.e., pairs of specialists or generalists) and apply identical analyses (i.e., as performed on the ecologically divergent pairs of species). In the case of the absence (or substantial decrease) of genes detected in this scenario, an association between the parallel differences in selective pressures and ecological divergence and adaptation could be confirmed. Secondly, there is an essential need for an accurate description of the clownfish traits that are functionally related to their ecological niche (see below). Only then could linking the function of the genes, the ecological niche, and the observed phenotypes be conceivable.

Within the diversification of the skunk complex

In chapter 4, I took advantage of the availability of allopatric and sympatric populations in the skunk complex to investigate the mechanisms underlying their diversification through population genomic approaches. I found that the group's diversification did not occur in strict isolation and rejected a hybrid origin of *A. sandaracinos*. I discovered that the species experienced moderate ancestral gene flow, which lessened but persisted in sympatry. I also identified introgressed regions from *A. perideraion* in *A. sandaracinos* and two highly divergent regions on chromosome 18. Altogether, these results provide the first insights into the diversification of the skunk complex and show that hybridization events in the IAA are less pervasive than initially thought.

In order to better characterize the diversification of the skunk complex, further studies should also consider the recently described species *A. pacificus* (Allen et al., 2010). This species is basal to the *A. akallopisos*-*A. sandaracinos* pair (see Introduction, Figure 1). Integrating it in further studies, such as in the demographic modeling approach, could allow more accurate characterization of the timing of gene flow between *A. perideraion* and the ancestor of *A. akallopisos*-*A. sandaracinos*, potentially discerning between speciation with gene flow (i.e., if gene flow also occurred before the split with *A. pacificus*) or secondary contact.

The low level of hybridization between the three species in the IAA suggests some level of host repartition in sympatry. Despite having the ability to interact with the same sea anemones hosts, the species may vary their host selection in sympatry, likely due to interspecific competition. In clownfish, some species are known to be competitively superior to others (e.g., Fautin, 1986; Srinivasan, 1999; Huebner et al., 2012), with body size as a major contributor to competitiveness (Srinivasan, 1999; Hattori, 2002). However, species of the skunk complex have never been tested in this context. Because *A. perideraion* is the smallest species in the group (approximate total length of 10 cm, 11 cm, and 14 cm for respectively *A. perideraion*, *A. akallopisos*, and *A. sandaracinos*; Allen, 1991), competition could compel it to associate with sea anemone hosts not exploited by other species, reducing encountering frequencies and hybridization probabilities. This hypothesis should be tested in an experimental framework, such as in Srinivasan et al. (1999), but is supported by preliminary results on host frequency association (Titus et al., unpublished results).

Although gene flow in sympatry is low, hybridization between *A. perideraion* and *A. sandaracinos* occurred, with the latter having fixed introgressed regions. Here again, to link these regions to a potential adaptive role in the evolution of this species, further studies on clownfishes' functional traits are necessary (see below). In addition, patterns observed on chromosome 18 suggest the presence of additional hybridization events between *A. perideraion* and species not considered in the study, probably belonging to the *ephippium* clade (see chapter 3). However, differential sorting among incipient species of diverged haplotypes that predate speciation cannot be excluded (Cruickshank & Hahn, 2014). As previously mentioned, future studies should further investigate these regions of chromosome 18, and the availability of chromosome-level assemblies of multiple species should facilitate their characterization. Nevertheless, it is also essential to understand the evolutionary origin of these regions to be able to assess their potential role in the diversification of the skunk complex, and more in general, in clownfishes. A possible approach could consist in integrating species of the *ephippium* clade and comparing demographic models of evolution based on whole-genome SNPs (excluding chromosome 18) or SNPs in these regions, with methods based on coalescent simulations, such as *fastsimcoal2* (Excoffier et al., 2013), or in an Approximate Bayesian Computation framework, such as implemented in DIYABC (Cornuet et al., 2014). While similar methods have been used to investigate the evolution of a supergene regulating trans-species social polymorphism in ants (Yan et al., 2020), the feasibility of these approaches should nevertheless be tested before extensive analysis, as those methods usually are not designed for such divergence between species. The link between these regions and a potential adaptive role would then require additional information on clownfish functional traits (see below).

While no extensive hybridization was observed in the skunk complex in sympatry, I still believe that we could obtain insights into the diversification of clownfish through the analyses of hybridization events, where recent hybridization potentially reveals ecologically important regions (such as in Nadeau et al., 2012; Clarkson et al., 2014; Talla et al., 2017). In this sense, future genomic studies should focus on natural clownfish hybrids and their parental species (for instance, *A. chrysopterus*, *A. sandaracinos*, and *A. leucokranos*). A genetic study on these species was recently performed (Gainsford, 2020) and indicated that parental species remain genetically distinct, whereas hybrids are genetically similar to *A. sandaracinos* because of persistent backcrossing. Nevertheless, this study was performed only on a few mitochondrial and nuclear markers. Through population genomic approaches, further studies should

investigate the genomic heterogeneity in these parental and hybrid species to assess the presence of regions of restricted introgression that could be important in maintaining species identities (Payseur, 2010). Additionally, the presence of hybrids at different stages of backcrossing showing different phenotypes could allow linking the phenotypes to specific loci in the genome through genotype–phenotype associations analyses (e.g., Brelsford et al., 2017; Aguillon et al., 2021).

Stepping away from the genomics and back to ecology and functional traits

In 2012, Litsios and coauthors observed that, following the acquisition of their mutualism with sea anemones, clownfishes diversified into several ecological niches linked with host and habitat use and developed rapid and convergent phenotypes correlated to the host-associated ecological niches, suggesting that the group likely experienced adaptive radiation through ecological speciation. Clownfish ecological niches were described based on the reported interaction with sea anemones species, and the major axis of variation corresponded to the number of possible hosts (i.e., specialist vs. generalist species). The measured phenotypes mostly corresponded to clownfish morphological traits and the number of gills rakers. Although a correlation between phenotypes and environments was detected, linking these clownfish traits with the adaptive advantage in the described ecological niches is difficult. Consequently, it is even more challenging to link the observed genomic patterns to relevant functions in adaptation and diversification. It is thus essential for future studies to first confirm the described ecological niche in the light of the new knowledge on clownfish-sea anemone interactions, and secondly, to investigate clownfish traits that are functionally related to these ecological niches.

Classical clownfish-sea anemones host associations are derived from the overall observation of the interactions in nature (Dunn, 1981; Fautin & Allen, 1997; Litsios et al., 2012; da Silva & Nedosyko, 2016). However, differences in host frequency association (or host preference) are not reported, and some observations of interaction may only derive from sporadic associations driven by interspecific competition (Den Hartog, 1997). Titus et al. (unpublished results) started to investigate the clownfish-sea anemones frequency association as a proxy of host preference. For this, they performed literature searches for independent confirmation of host use and took advantage of the large number of publicly available pictures from SCUBA divers to assess the overall clownfish-sea anemone association frequency. Preliminary results suggest differences in the overall host usage compared to what has been previously reported and exploited (i.e., Fautin & Allen, 1997; Litsios et al., 2012; da Silva &

Nedosyko, 2016). While further investigation is needed, these results need to be exploited to re-evaluate the description of clownfish ecological niches.

In order to investigate clownfish traits that are functionally related to ecological niches, future studies should involve a broad characterization of clownfish phenotypes. This should include morphologies (such as with geometric morphometric approaches, e.g., Clabaut et al., 2007), coloration, body stripes (which may be ecologically important, see Salis et al., 2018, 2021), and chemical biology of clownfish-sea anemone interaction, which is still poorly understood (Mebs et al., 2009) but could be advanced by comprehensive studies on clownfish mucus (see above). These traits should then be integrated into comparative analyses of ecology, also using the phylogenetic context, to investigate further the role of ecology in driving clownfish diversification. Additionally, after identifying the key phenotypes linked with ecological divergence, testing their fitness consequences (i.e., trait utility) in a strict experimental setting would be necessary to confirm the adaptive radiation of the group.

These analyses are essential to achieve a description of different clownfish ecotypes, which will dramatically help in the interpretation of the genomic patterns observed in the study. Only at that moment will it be possible to understand further the potential role of the genes identified in this thesis (e.g., the candidate introgressed genes, the genes observed on chromosome 18, and the ones showing patterns of parallel evolution. And only then will it be conceivable to validate the functions of the genes using CRISPR/Cas9 genome editing approaches (Bono et al., 2015; Mitchell et al., 2021).

In conclusion, the research project I developed in my thesis provides the first genomic insights into the acquisition of clownfish mutualism with sea anemones and the mechanisms underlying the diversification of the group. The results I obtained integrate the constantly growing collection of studies investigating the mechanisms governing species diversification, which brings us a step closer to the understanding of how biodiversity on Earth is created.

REFERENCES

- Abdullah, N. S., & Saad, S.** (2015). Rapid detection of N-acetylneuraminic acid from false clownfish using HPLC-FLD for symbiosis to host sea anemone. *Asian Journal of Applied Sciences*, 3(5).
- Aguillon, S. M., Walsh, J., & Lovette, I. J.** (2021). Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. *Proceedings of the Royal Society B*, 288(1943), 20201805.
- Allen, G. R., Drew, J., & Fenner, D.** (2010). *Amphiprion pacificus*, a new species of anemonefish
- Allen, G.R.**, (1991). *Damselfishes of the world*. Mergus Publishers, Melle, Germany. 271 p

- Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., & Gouil, Q.** (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome biology*, *21*(1), 1-16.
- Balamurugan, J., Kumar, T. A., Kannan, R., & Pradeep, H. D.** (2014). Acclimation behaviour and biochemical changes during anemonefish (*Amphiprion sebae*) and sea anemone (*Stichodactyla haddoni*) symbiosis. *Symbiosis*, *64*(3), 127-138.
- Bassham, S., Catchen, J., Lescak, E., von Hippel, F. A., & Cresko, W. A.** (2018). Repeated selection of alternatively adapted haplotypes creates sweeping genomic remodeling in stickleback. *Genetics*, *209*(3), 921-939.
- Bergstrom, C. T., & Lachmann, M.** (2003). The Red King effect: when the slowest runner wins the coevolutionary race. *Proceedings of the National Academy of Sciences*, *100*(2), 593-598.
- Berner, D., & Salzburger, W.** (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, *31*(9), 491-499.
- Bono, J. M., Olesnick, E. C., & Matzkin, L. M.** (2015). Connecting genotypes, phenotypes and fitness: harnessing the power of CRISPR/Cas9 genome editing. *Molecular ecology*, *24*(15), 3810-3822.
- Bowen, B. W., Forsman, Z. H., Whitney, J. L., Fauci, A., Hoban, M., Canfield, S. J., ... & Toonen, R. J.** (2020). Species radiations in the sea: What the flock?. *Journal of Heredity*, *111*(1), 70-83.
- Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., ... & Turner-Maier, J.** (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, *513*(7518), 375-381.
- Brelsford, A., Toews, D. P., & Irwin, D. E.** (2017). Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1866), 20171106.
- Brinchmann, M. F.** (2016). Immune relevant molecules identified in the skin mucus of fish using-omics technologies. *Molecular BioSystems*, *12*(7), 2056-2063.
- Campbell, C. R., Poelstra, J. W., & Yoder, A. D.** (2018). What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, *124*(4), 561-583.
- Clabaut, C., Bunje, P. M., Salzburger, W., & Meyer, A.** (2007). Geometric morphometric analyses provide evidence for the adaptive character of the Tanganyikan cichlid fish radiations. *Evolution*, *61*(3), 560-578.
- Clarkson, C. S., Weetman, D., Essandoh, J., Yawson, A. E., Maslen, G., Manske, M., ... & Donnelly, M. J.** (2014). Adaptive introgression between *Anopheles* sibling species eliminates a major genomic island but not reproductive isolation. *Nature communications*, *5*(1), 1-10.
- Cofer, E. M., Raimundo, J., Tadych, A., Yamazaki, Y., Wong, A. K., Theesfeld, C. L., ... & Troyanskaya, O. G.** (2020). DeepArk: modeling cis-regulatory codes of model species with deep learning. *BioRxiv*.
- Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., ... & Kingsley, D. M.** (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *science*, *307*(5717), 1928-1933.
- Cornuet, J. M., Pudlo, P., Veyssier, J., Dehne-Garcia, A., Gautier, M., Leblois, R., ... & Estoup, A.** (2014). DIYABC v2. 0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, *30*(8), 1187-1189.

- Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., ... & Postlethwait, J. H.** (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, *101*(16), 6050-6055.
- Cruickshank, T. E., & Hahn, M. W.** (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular ecology*, *23*(13), 3133-3157.
- da Silva, K. B. & Nedosyko, A.** (2016) Sea Anemones and Anemonefish: A Match Made in Heaven. In Goffredo, S. & Dubinsky, Z. (eds.) *The Cnidaria, Past, Present and Future*, 425–438 (Springer International Publishing, 2016)
- Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., ... & Heliconius Genome Consortium.** (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94.
- de Sousa, R. P. C., Silva-Oliveira, G. C., Furo, I. O., de Oliveira-Filho, A. B., de Brito, C. D. B., Rabelo, L., ... & Vallinoto, M.** (2021). The role of the chromosomal rearrangements in the evolution and speciation of Elopiformes fishes (Teleostei; Elopomorpha). *Zoologischer Anzeiger*, *290*, 40-48.
- Den Hartog, J. C.** (1997). Notes on the genus *Amphiprion* Bloch & Schneider, 1801 (Teleostei: Pomacentridae) and its host sea anemones in the Seychelles. *Zoologische Mededelingen*, *71*(17), 181-188.
- Dong, Y., Chen, S., Cheng, S., Zhou, W., Ma, Q., Chen, Z., ... & Xiang, Q. Y. J.** (2019). Natural selection and repeated patterns of molecular evolution following allopatric divergence. *Elife*, *8*, e45199.
- Dunn, D. F.** (1981). The clownfish sea anemones: Stichodactylidae (Coelenterata: Actiniaria) and other sea anemones symbiotic with pomacentrid fishes. *Transactions of the American Philosophical Society*, *71*(1), 3-115.
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M.** (2011). Testing for ancient admixture between closely related populations. *Molecular biology and evolution*, *28*(8), 2239-2252.
- Émie, A. G., François-Étienne, S., Sidki, B., & Nicolas, D.** (2021). Microbiomes of clownfish and their symbiotic host anemone converge before their first physical contact. *Microbiome*, *9*(1), 1-15.
- Excoffier, L., Dupanloup, I., Huerta-Sánchez, E., Sousa, V. C., & Foll, M.** (2013). Robust demographic inference from genomic and SNP data. *PLoS genetics*, *9*(10), e1003905.
- Faber-Hammond, J. J., Bezault, E., Lunt, D. H., Joyce, D. A., & Renn, S. C.** (2019). The genomic substrate for adaptive radiation: copy number variation across 12 tribes of African cichlid species. *Genome biology and evolution*, *11*(10), 2856-2874.
- Fæste, C. K., Tartor, H., Moen, A., Kristoffersen, A. B., Dhanasiri, A. K. S., Anonsen, J. H., ... & Grove, S.** (2020). Proteomic profiling of salmon skin mucus for the comparison of sampling methods. *Journal of Chromatography B*, *1138*, 121965.
- Fautin, D. G.** (1986). Why do anemonefishes inhabit only some host actinians?. *Environmental Biology of Fishes*, *15*(3), 171-180.
- Fautin, D.G., Allen, G.R.** (1997). Anemonefishes and their host sea anemones. Perth: Western Australian Museum.
- Gainsford, A., Jones, G. P., Hobbs, J. P. A., Heindler, F. M., & van Herwerden, L.** (2020). Species integrity, introgression, and genetic variation across a coral reef fish hybrid zone. *Ecology and evolution*, *10*(21), 11998-12014.

- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., ... & Hansen, N. F.** (2010). A draft sequence of the Neandertal genome. *Science*, 328(5979), 710-722.
- Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G., & Siepel, A.** (2011). Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics*, 43(10), 1031-1034.
- Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T.** (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome research*, 27(6), 1004-1015.
- Hattori, A.** (2002). Small and large anemonefishes can coexist using the same patchy resources on a coral reef, before habitat destruction. *Journal of Animal Ecology*, 71(5), 824-831.
- Huebner, L. K., Dailey, B., Titus, B. M., Khalaf, M., & Chadwick, N. E.** (2012). Host preference and habitat segregation among Red Sea anemonefish: effects of sea anemone traits and fish life stages. *Marine Ecology Progress Series*, 464, 1-15
- Jackson, C. E., Xu, S., Ye, Z., Pfrender, M. E., Lynch, M., Colbourne, J. K., & Shaw, J. R.** (2021). Chromosomal rearrangements preserve adaptive divergence in ecological speciation. *BioRxiv*.
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., ... & Kingsley, D. M.** (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55-61.
- Kozak, K. M., Joron, M., McMillan, W. O., & Jiggins, C. D.** (2021). Rampant genome-wide admixture across the *Heliconius* radiation. *Genome Biology and Evolution*.
- Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., ... & Andersson, L.** (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518(7539), 371-375.
- Lawson, L. P., & Petren, K.** (2017). The adaptive genomic landscape of beak morphology in Darwin's finches. *Molecular ecology*, 26(19), 4978-4989.
- Lehmann, R., Lightfoot, D. J., Schunter, C., Michell, C. T., Ohyanagi, H., Mineta, K., ... & Ravasi, T.** (2019). Finding Nemo's Genes: A chromosome-scale reference assembly of the genome of the orange clownfish *Amphiprion percula*. *Molecular ecology resources*, 19(3), 570-585.
- Lewis, J. J., Geltman, R. C., Pollak, P. C., Rondem, K. E., Van Belleghem, S. M., Hubisz, M. J., ... & Reed, R. D.** (2019). Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proceedings of the National Academy of Sciences*, 116(48), 24174-24183.
- Litsios, G., Sims, C. A., Wüest, R. O., Pearman, P. B., Zimmermann, N. E., & Salamin, N.** (2012). Mutualism with sea anemones triggered the adaptive radiation of clownfishes. *BMC evolutionary biology*, 12(1), 212.
- Lynch, M., & Walsh, B.** (2007). *The origins of genome architecture* (Vol. 98). Sunderland, MA: Sinauer associates.
- Marques, D. A., Meier, J. I., & Seehausen, O.** (2019). A combinatorial view on speciation and adaptive radiation. *Trends in ecology & evolution*, 34(6), 531-544.
- Martin, S. H., Davey, J. W., & Jiggins, C. D.** (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular biology and evolution*, 32(1), 244-257.
- McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., ... & Seehausen, O.** (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, 586(7827), 75-79.
- Mebs, D.** (2009). Chemical biology of the mutualistic relationships of sea anemones with fish and crustaceans.

Toxicon, 54(8), 1071-1074.

- Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O.** (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications*, 8(1), 1-11.
- Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M.** (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, 35(7), 561-572.
- Miller, D. E., Sulovari, A., Wang, T., Loucks, H., Hoekzema, K., Munson, K. M., ... & University of Washington Center for Mendelian Genomics.** (2020). Targeted long-read sequencing resolves complex structural variants and identifies missing disease-causing variants. *BioRxiv*.
- Mitchell, L. J., Tettamanti, V., Marshall, J. N., Cheney, K., & Cortesi, F.** (2020). CRISPR/Cas9-mediated generation of biallelic G0 anemonefish (*Amphiprion ocellaris*) mutants. *BioRxiv*.
- Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., ... & Jiggins, C. D.** (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 343-353.
- North, H. L., Caminade, P., Severac, D., Belkhir, K., & Smadja, C. M.** (2020). The role of copy-number variation in the reinforcement of sexual isolation between the two European subspecies of the house mouse. *Philosophical Transactions of the Royal Society B*, 375(1806), 20190540.
- Parida, S., Mohapatra, A., Kar, B., Mohanty, J., & Sahoo, P. K.** (2018). Transcriptional analysis of immune-relevant genes in the mucus of *Labeo rohita*, experimentally infected with *Argulus siamensis*. *Acta parasitologica*, 63(1), 125-133.
- Payseur, B. A.** (2010). Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular ecology resources*, 10(5), 806-820.
- Randall, J., & Fautin, D.** (2002). Fishes other than anemonefishes that associate with sea anemones. *Coral Reefs*, 21(2), 188-190.
- Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L.** (2009). Reconstructing Indian population history. *Nature*, 461(7263), 489-494.
- Ren, Y., Zhao, H., Su, B., Peatman, E., & Li, C.** (2015). Expression profiling analysis of immune-related genes in channel catfish (*Ictalurus punctatus*) skin mucus following *Flavobacterium columnare* challenge. *Fish & shellfish immunology*, 46(2), 537-542.
- Reverter, M., Sasal, P., Banaigs, B., Lecchini, D., Lecellier, G., & Tapissier-Bontemps, N.** (2017). Fish mucus metabolome reveals fish life-history traits. *Coral Reefs*, 36(2), 463-475.
- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., ... & Jarvis, E. D.** (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), 737-746.
- Ronco, F., Matschiner, M., Böhne, A., Boila, A., Büscher, H. H., El Taher, A., ... & Salzburger, W.** (2021). Drivers and dynamics of a massive adaptive radiation in cichlid fishes. *Nature*, 589(7840), 76-81.
- Roux, N., Lami, R., Salis, P., Magré, K., Romans, P., Masanet, P., ... & Laudet, V.** (2019). Sea anemone and clownfish microbiota diversity and variation during the initial steps of symbiosis. *Scientific reports*, 9(1), 1-13.
- Rubin, B. E., & Moreau, C. S.** (2016). Comparative genomics reveals convergent rates of evolution in ant-plant mutualisms. *Nature Communications*, 7(1), 1-11.

- Salis, P., Roux, N., Huang, D., Marcionetti, A., Mouginot, P., Reynaud, M., ... & Laudet, V.** (2021). Thyroid hormones regulate the formation and environmental plasticity of white bars in clownfishes. *Proceedings of the National Academy of Sciences*, 118(23).
- Salis, P., Roux, N., Soulat, O., Lecchini, D., Laudet, V., & Frédérick, B.** (2018). Ontogenetic and phylogenetic simplification during white stripe evolution in clownfishes. *BMC biology*, 16(1), 1-13.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... & Brelsford, A.** (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3), 176-192.
- Srinivasan, M., Jones, G. P., & Caley, M. J.** (1999). Experimental evaluation of the roles of habitat selection and interspecific competition in determining patterns of host use by two anemonefishes. *Marine Ecology Progress Series*, 186, 283-292.
- Stern, D. L., & Orgogozo, V.** (2008). The loci of evolution: how predictable is genetic evolution?. *Evolution: International Journal of Organic Evolution*, 62(9), 2155-2177.
- Svardal, H., Quah, F. X., Malinsky, M., Ngatunga, B. P., Miska, E. A., Salzburger, W., ... & Durbin, R.** (2020). Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Molecular biology and evolution*, 37(4), 1100-1113.
- Talla, V., Kalsoom, F., Shipilina, D., Marova, I., & Backström, N.** (2017). Heterogeneous patterns of genetic diversity and differentiation in European and Siberian chiffchaff (*Phylloscopus collybita abietinus*/P. tristis). *G3: Genes, Genomes, Genetics*, 7(12), 3983-3998.
- Titus, B. M., Laroche, R., Rodríguez, E., Wirshing, H., & Meyer, C. P.** (2020). Host identity and symbiotic association affects the taxonomic and functional diversity of the clownfish-hosting sea anemone microbiome. *Biology letters*, 16(2), 20190738.
- Van Valen, L.** (1973) A new evolutionary law. *Evol. Theory* 1, 1–30 .
- Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L.** (2019). Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular ecology*, 28(6), 1203-1209.
- Wenger, A. M., Peluso, P., Rowell, W. J., Chang, P. C., Hall, R. J., Concepcion, G. T., ... & Hunkapiller, M. W.** (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature biotechnology*, 37(10), 1155-1162.
- Worsley-Hunt, R., Bernard, V., & Wasserman, W. W.** (2011). Identification of cis-regulatory sequence variations in individual genome sequences. *Genome medicine*, 3(10), 1-14.
- Wray, G. A.** (2007). The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8(3), 206-216.
- Yan, Z., Martin, S. H., Gotzek, D., Arsenaault, S. V., Duchon, P., Helleu, Q., ... & Keller, L.** (2020). Evolution of a supergene that regulates a trans-species social polymorphism. *Nature ecology & evolution*, 4(2), 240-249.
- Zhang, L., Reifová, R., Halenková, Z., & Gompert, Z.** (2021). How Important Are Structural Variants for Speciation?. *Genes*, 12(7), 1084.

Appendices

Annex 1: Salis et al. (2019). Published in *Pigment Cell & Melanoma Research*

The complete article is available on *Pigment Cell & Melanoma Research* [online](#) and on [OneDrive](#)¹ for easy access. To avoid excessive printing, I report here only the abstract.

Developmental and comparative transcriptomic identification of iridophore contribution to white barring in clownfish

Salis, P., Lorin, T., Lewis, V., Rey, C., Marcionetti, A., Escande, M.L., Roux, N., Besseau, L., Salamin, N., Sémon, M., Parichy, D., Volf, J.N., Laudet, V.

ABSTRACT

Actinopterygian fishes harbor at least eight distinct pigment cell types, leading to a fascinating diversity of colors. Among this diversity, the cellular origin of the white color appears to be linked to several pigment cell types such as iridophores or leucophores. We used the clownfish *Amphiprion ocellaris*, which has a color pattern consisting of white bars over a darker body, to characterize the pigment cells that underlie the white hue. We observe by electron microscopy that cells in white bars are similar to iridophores. In addition, the transcriptomic signature of clownfish white bars exhibits similarities with that of zebrafish iridophores. We further show by pharmacological treatments that these cells are necessary for the white color. Among the top differentially expressed genes in white skin, we identified several genes (*fhl2a*, *fhl2b*, *saiyan*, *gpnmb*, and *apoD1a*) and show that three of them are expressed in iridophores. Finally, we show by CRISPR/Cas9 mutagenesis that these genes are critical for iridophore development in zebrafish. Our analyses provide clues to the genomic underpinning of color diversity and allow identification of new iridophore genes in fish.

KEYWORDS

Amphiprion, clownfish, coral reef fish, iridophore, leucophore, transcriptomic

¹https://unils-my.sharepoint.com/:b:/g/personal/anna_marcionetti_unil_ch/EaWL35GuSGNLnmXgynDEJOQBbsIvRuUTjRpsvJizp33JOW?e=nLeR8y

Annex 2: Salis et al. (2021). Published in *PNAS*

The complete article is available on *PNAS* [online](#) and on [OneDrive](#)² for easy access. To avoid excessive printing, I report here only the abstract.

Thyroid hormones regulate the formation and environmental plasticity of white bars in clownfishes

Salis, P., Roux, N., Huang, D., Marcionetti, A., Mouginot, P., Reynaud, M., Salles, O., Salamin, N., Pujol, B., Parichy, D.M., Planes, S., Laudet, V.

ABSTRACT

Determining how plasticity of developmental traits responds to environmental conditions is a challenge that must combine evolutionary sciences, ecology, and developmental biology. During metamorphosis, fish alter their morphology and color pattern according to environmental cues. We observed that juvenile clownfish (*Amphiprion percula*) modulate the developmental timing of their adult white bar formation during metamorphosis depending on the sea anemone species in which they are recruited. We observed an earlier formation of white bars when clownfish developed with *Stichodactyla gigantea* (*Sg*) than with *Heteractis magnifica* (*Hm*). As these bars, composed of iridophores, form during metamorphosis, we hypothesized that timing of their development may be thyroid hormone (TH) dependent. We treated clownfish larvae with TH and found that white bars developed earlier than in control fish. We further observed higher TH levels, associated with rapid white bar formation, in juveniles recruited in *Sg* than in *Hm*, explaining the faster white bar formation. Transcriptomic analysis of *Sg* recruits revealed higher expression of *duox*, a dual oxidase implicated in TH production as compared to *Hm* recruits. Finally, we showed that *duox* is an essential regulator of iridophore pattern timing in zebrafish. Taken together, our results suggest that TH controls the timing of adult color pattern formation and that shifts in *duox* expression and TH levels are associated with ecological differences resulting in divergent ontogenetic trajectories in color pattern development.

KEYWORDS

Pigmentation, developmental plasticity, clownfishes, thyroid, hormones, metamorphosis

²https://unils-my.sharepoint.com/:b:/g/personal/anna_marcionetti_unil_ch/ESThCEX6_kFNoz1ttNAd8cwBiQk3areL_6xsssNK41teSg?e=wMzX06

Annex 3: Serrano-Serrano et al. *in prep*

The complete article is available on [OneDrive](#)³ for easy access. To avoid excessive printing, I report here only the abstract. [Supplementary Information and Figures](#) and [Supplementary Tables](#) are also available online.

Convergent changes in gene expression associated with repeated transitions between hummingbird and bee pollinated flowers

Serrano-Serrano[§], M.L., Marcionetti[§], A., Perret, M., Salamin N.

§ these two authors contributed equally to the work and should be considered co-first authors

ABSTRACT

Background: The repeated evolution of convergent floral shapes and colors in angiosperms has been largely interpreted as the response to pollinator-mediated selection to maximize the attraction and efficiency of specific groups of pollinators. The genetic mechanisms contributing to certain flower traits have been studied in detail for model species. Still, the extent to which flowers are free to vary and how predictable are the genetic changes underlying flower adaptation to pollinator shifts remain largely unknown. Here, we aimed at detecting the genetic basis of the repeated evolution of flower phenotypes associated with pollinator shifts.

Results: We assembled and compared de novo transcriptomes of three phylogenetically independent pairs of Gesneriaceae species, each with contrasting flower phenotype adapted to either bee or hummingbird pollination. We assembled and analyzed a total of 14,059 genes, and we showed that changes in expression in 550 of them were associated with the pollination syndromes. Among those, we observed genes with function associated with floral color, scent, development and nectar secretion. These genes represent candidate genes involved in the build-up of the convergent floral phenotypes.

Conclusions: This study provides insights into the molecular mechanisms underlying the repeated evolution of pollination syndromes. Although additional lineage-specific responses are observed, these results suggest that convergent evolution of gene expression is involved in the build-up of the pollination syndromes. Our results show that Gesneriaceae represent a promising model clade for the genomic analysis of floral evolution in tropical plants.

KEYWORDS

Gesneriaceae transcriptomes, Neotropics, Flower development, Parallel evolution, Pollinator-mediated selection

³https://unils-my.sharepoint.com/:b/g/personal/anna_marcionetti_unil_ch/EV7cMwqSaD9FqF-XuDABLw8BigU8Nwndowv8EtL_8Y2PFw?e=gDdFut

Annex 4: Master Thesis of Ms. Virginie Ricci

During my thesis, I set up a Master Project aiming at identifying and analyzing conserved non-coding regions in clownfish genomes. I worked alongside and supervised Ms. Virginie Ricci for the whole duration of the project. The complete Master Thesis of Ms. Virginie Ricci is available on [OneDrive](#)⁴.

Identification of putative CNE potentially associated with clownfish adaptive radiation

Virginie RICCI

Master Thesis of Science In Molecular Life Sciences, University of Lausanne

January 2018

Supervisor: **Anna Marcionetti**

Director: **Nicolas Salamin**

ABSTRACT

The adaptive radiation of clownfishes is mostly explained by their mutualistic interaction with sea anemones. Clownfishes are considered either as generalist or specialist species depending on the number of mutualistic hosts they interact with. Their phylogeny suggests that the adaptation to one or several sea anemones happens independently several times. Modeling approaches hypothesize that the adaptive radiation is more likely to occur if the number of underlying loci is small. The emergence of adaptive phenotypic traits may be therefore promoted by few alterations in both coding and non-coding DNA sequences. In comparison with mutations in coding DNAs which broadly affect either gene structure or function, alterations in specific non-coding DNAs may dynamically influence gene expression through time and cell location. In particular, changes in conserved non-coding elements (CNEs) potentially binding transcription factors (Transcription Factor Binding Sites TFBSs) may be involved in the regulation of adaptive genes. Among the genomic changes underlying the clownfish adaptive radiation, we hypothesized that convergent genomic alterations may reflect convergence in adaptive phenotypic modifications potentially associated with host specialization. In this project, we developed an in-silico pipeline to identify CNEs that are putative TFBSs responsible for clownfish diversification and host specialization. Although further experimental validations are needed, our results provide the first candidate regions of non-coding DNA that may be associated with the clownfish adaptive radiation.

⁴https://unils-my.sharepoint.com/:b:/g/personal/anna_marcionetti_unil_ch/Eem6PdQybtDgv6WpKGvgOMBxIHwCCGkcbAvr4fs488fdUg?e=BAJn17

Annex 5: Master Thesis of Ms. Sagane Dind

During my thesis, I set up a Master Project investigating the presence of mRNA in clownfish mucus and characterizing it through sequencing. I worked alongside and supervised Ms. Sagane Dind for the whole duration of the project. The complete Master Thesis of Ms. Sagane Dind is available on [OneDrive](#)⁵.

Genetic basis of the symbiosis between clownfish and sea anemones: transcriptomic profile of the epidermal mucus of three clownfish species

Sagane DIND

Master Thesis of Science In Behaviour, Evolution and Conservation, specialisation

“Computational Ecology and Evolution”, University of Lausanne

January 2019

Supervisor: **Anna Marcionetti**

Director: **Nicolas Salamin**

ABSTRACT

Clownfish (genera *Premnas* and *Amphiprion*) are well known for their mutualistic relationship with sea anemones, which are highly toxic for most other fish. Although it is believed that clownfish mucus plays a key role in the immunity from sea anemones toxicity and thus allows for this interaction, the mechanisms underlying this protection are still not well understood. While studies focusing on gene expression, protein characterization and enzyme activities of fish mucus exist, no studies have been performed to understand the composition of clownfish mucus. Here, we first compared *A. ocellaris* mucus and skin transcriptomic profiles. We found that the gene expression pattern in the mucus is very different from the skin, with an enrichment of genes involved in the immunity response found overexpressed in the mucus. We then compared the mucus transcriptome of three clownfish species with similar (*P. biaculeatus* and *A. frenatus*) or divergent (*A. ocellaris*) habitat and host usage. We found that similarities in mucus expression patterns were primarily associated with the species ecology rather than phylogenetic relationship, suggesting a role in the protection from different sea anemones hosts. This study brings new insights on the genetic basis and the mechanisms of clownfish protection from sea anemones toxicity.

⁵https://unils-my.sharepoint.com/:b:/g/personal/anna_marcionetti_unil_ch/EXUQgbmDBVIPiAW5TPOtuvMB558hAEr6L5F-RQHBHhEWfA?e=qz81dJ