

# Recurrent gene flow events occurred during the diversification of clownfishes of the skunk complex

Anna Marcionetti<sup>1</sup>  | Joris A. M. Bertrand<sup>1,2</sup>  | Fabio Cortesi<sup>3</sup> | Giulia F. A. Donati<sup>4,5</sup> | Sara Heim<sup>1</sup> | Filip Huyghe<sup>6</sup>  | Marc Kochzius<sup>6</sup> | Loïc Pellissier<sup>5,7</sup>  | Nicolas Salamin<sup>1</sup>

<sup>1</sup>Department of Computational Biology, Génopode, University of Lausanne, Lausanne, Switzerland

<sup>2</sup>Laboratoire Génome et Développement Des Plantes (UMR 5096 UPVD/CNRS), University of Perpignan via Domitia, Perpignan, France

<sup>3</sup>School of the Environment and Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia

<sup>4</sup>EAWAG Swiss Federal Institute of Aquatic Science & Technology, Dübendorf, Switzerland

<sup>5</sup>Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland

<sup>6</sup>Marine Biology – Ecology, Evolution and Genetics, Vrije Universiteit Brussel (VUB), Pleinlaan 2, Brussels, Belgium

<sup>7</sup>Ecosystems and Landscape Evolution, Department of Environmental System Science, Institute of Terrestrial Ecosystems, ETH Zürich, Zurich, Switzerland

## Correspondence

Nicolas Salamin, Department of Computational Biology, University of Lausanne, 1015 Lausanne, Switzerland. Email: [nicolas.salamin@unil.ch](mailto:nicolas.salamin@unil.ch)

## Funding information

Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 31003A-163428; University of Lausanne

Handling Editor: Nick Hamilton Barton

## Abstract

Clownfish (subfamily Amphiprioninae) are an iconic group of coral reef fish that evolved a mutualistic interaction with sea anemones, which triggered the adaptive radiation of the clade. Within clownfishes, the “skunk complex” is particularly interesting. Besides ecological speciation, interspecific gene flow and hybrid speciation are thought to have shaped the evolution of the group. We investigated the mechanisms characterizing the diversification of this 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 (*Amphiprion akallopisos*, *A. perideraion* and *A. sandaracinos*). We examined population structure, genomic divergence and introgression signals and performed demographic modelling to identify the most realistic diversification scenario. We excluded scenarios of strict isolation or hybrid origin of *A. sandaracinos*. We discovered moderate gene flow from *A. perideraion* to the ancestor of *A. akallopisos* + *A. sandaracinos* and weak gene flow between the species in the Indo-Australian Archipelago throughout the diversification of the group. We identified introgressed regions in *A. sandaracinos* and detected in *A. perideraion* two large regions of high divergence from the two other species. While we found that gene flow has occurred throughout the species' diversification, we also observed that recent admixture was less pervasive than initially thought, suggesting a role of host repartition or behavioural barriers in maintaining the genetic identity of the species in sympatry.

## KEYWORDS

*Amphiprion*, comparative genomics, disjunct geographic distribution, hybridization, introgression, species diversification

Anna Marcionetti and Joris A. M. Bertrand are co-first authorship.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2024 The Authors. *Molecular Ecology* published by John Wiley & Sons Ltd.

## 1 | INTRODUCTION

The underlying causes and mechanisms of organisms' adaptation, diversification and the emergence of new species have been fundamental questions in evolutionary biology (Losos et al., 2013). Speciation is a complex process and involves different and interplaying mechanisms, such as geographic isolation, genetic drift, natural selection and sexual selection (Coyne & Orr, 2004; Gavrillets, 2004; Gavrillets & Losos, 2009; Nosil et al., 2009; Schluter, 2009). Since access to genomic data is increasing rapidly, the investigation of mechanisms underlying diversification and speciation has been extended to numerous taxa, environments, and evolutionary stages (Berner & Salzburger, 2015; Cerca et al., 2023; Seehausen et al., 2014). Studies have identified loci promoting ecological divergence and, eventually, speciation (e.g. Malinsky et al., 2015; Payseur & Rieseberg, 2016). Likewise, the prevalence and importance of hybridization in the evolution of species has become evident (Abbott et al., 2013; Payseur & Rieseberg, 2016; Taylor & Larson, 2019). While hybridization was traditionally seen as a process limiting diversification (Mayr, 1963), it is now recognized to promote ecological adaptation and speciation (Abbott et al., 2013; Taylor & Larson, 2019), and it is thus a pervasive key element of adaptive radiations (Berner & Salzburger, 2015; Cerca et al., 2023). Hybrid speciation (Olave et al., 2022) and ancient hybridization events that fuelled diversification (Meier et al., 2017, 2023; Svardal et al., 2020) have been observed in cichlid fishes, while introgressive hybridization linked to adaptive divergence has been found in *Heliconius* butterflies (Nadeau et al., 2013).

Genomic studies including those in cichlids and butterflies, have contributed considerably to our understanding of the molecular bases of diversification and speciation (Berner & Salzburger, 2015; Seehausen et al., 2014). However, more comparative studies considering different organisms at various stages of divergence are needed to fully grasp the interplay between genomic properties, ecology, geography and demographic history in shaping species diversification (Campbell et al., 2018; Seehausen et al., 2014; Wolf & Ellegren, 2017). The skunk complex (or *akallopisos* group) within the clownfish genus *Amphiprion* (family Pomacentridae) lends itself to this purpose. This complex is composed of three main species distributed across the Indo-Pacific Ocean, *Amphiprion akallopisos*, *A. sandaracinos* and *A. perideraion* (Figure 1a; Fautin & Allen, 1997), as well as a fourth species that is endemic to Fiji, Tonga, Samoa and Wallis Island (*A. pacificus*; Allen et al., 2010). Like all clownfishes, they maintain a mutualistic interaction with sea anemones (Fautin & Allen, 1997). However, the members of the skunk complex show divergence in host use (Table S1). *A. perideraion* is a generalist that can interact with four sea anemone species (*Heteractis magnifica*, *H. crispa*, *Macroactyla doreensis* and *Stichodactyla gigantea*; Fautin & Allen, 1997; Litsios et al., 2012), while *A. sandaracinos* and *A. akallopisos* only associate with two host species. The former can live in *S. gigantea* and *S. mertensii*, while the latter is found in *S. mertensii* and *H. magnifica* (Fautin & Allen, 1997; Litsios et al., 2012). The group diversified in the Indo-Australian Archipelago (IAA) c. 1–5 MYA (Cowman & Bellwood, 2013; Frédérick et al., 2013; Litsios

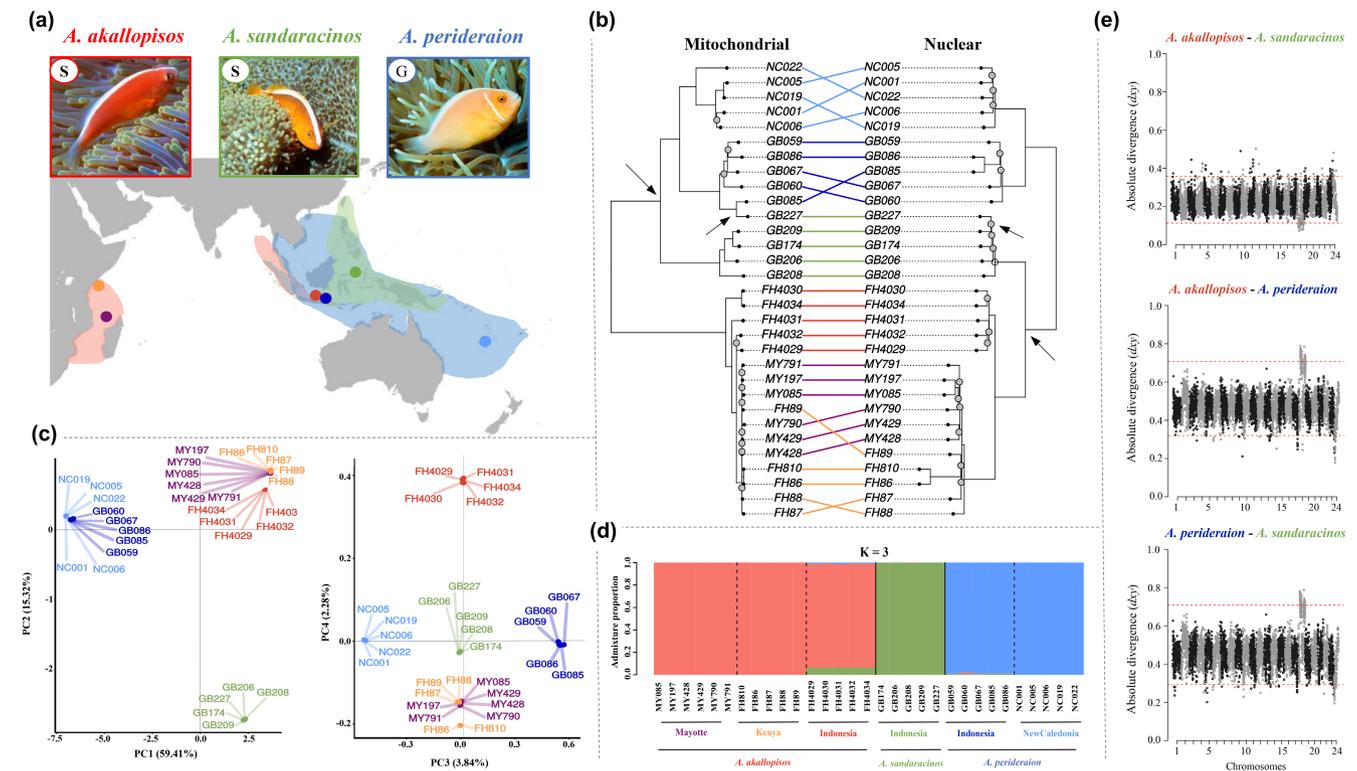
et al., 2012, 2014; Rabosky et al., 2018). Today, the skunk complex species have a vast and only partially overlapping geographical distribution (Figure 1a). Although the three species still co-occur in the IAA, *A. akallopisos* has a disjunct distribution and can also be found in the Western Indian Ocean (WIO), while *A. perideraion* is observed further east, reaching the New Caledonian (NC) archipelago (Allen, 1991; Figure 1a). Based on cytonuclear inconsistencies detected between mitochondrial and nuclear phylogenetic trees, ancestral hybridization events are thought to have shaped the diversification of the complex, and a hybrid origin has been suggested for *A. sandaracinos* (Litsios & Salamin, 2014). Furthermore, hybrids involving species of the skunk complex are commonly observed in the aquarium trade (Pedersen, 2014). Contemporary gene flow between the three species in the IAA is also likely, as suggested by the low interspecific barcode variation for the species in this region (Steinke et al., 2009). However, whole genome characterization is required to examine the actual extent of hybridization events in the evolution of the skunk complex and the potential role played by such gene flow in the origin of the species of the group.

In this study, we investigated the processes characterizing the diversification of the skunk complex by sampling and sequencing sympatric and allopatric populations of the three species. Our goal was to test whether *A. sandaracinos* originated through hybrid speciation or whether its diversification was characterized by ancestral hybridization events, as suggested by the cytonuclear discordance. Under the hypotheses of ancestral gene flow, ecological divergence in host usage (i.e. specialists (S) vs. generalists (G), Figure 1a) may have contributed to the speciation process by contrasting the homogenization effect of gene flow. This type of ecological divergence (host specialists vs. host generalists) is thought to be the driver for the ecological speciation and adaptive radiation in clownfishes, especially for the species with overlapping distributions centred in the IAA (Litsios et al., 2012). Additionally, we examined the occurrence of recent interspecific gene flow in the IAA. The presence of gene flow between populations of different species could result in the homogenization of their genomic divergence, which, in turn, would allow the identification of genomic regions of increased differentiation potentially involved in the ecological divergence or speciation of the complex (e.g. Barrera-Guzmán et al., 2022; Bourgeois et al., 2020; Luzuriaga-Aveiga et al., 2021; Nadeau et al., 2013). We provide insights into the processes characterizing the evolution of the clownfish species of the skunk complex, and more broadly, advance our understanding of the radiation of clownfishes.

## 2 | MATERIALS AND METHODS

### 2.1 | Species selection and sampling

We obtained the overall geographic distribution of the species of the skunk complex from Allen (1991) and GBIF.org (GBIF Occurrence Download <https://doi.org/10.15468/dl.rzpk14>, accessed in January 2016; Figure 1a, Figure S1). We considered



**FIGURE 1** The three *Amphiprion* species considered in this study, their geographical distributions and the locations of the sampled populations (a). The classification based on the host sea anemone usage of the three species is reported, with S=Specialist and G=Generalist (Table S1). Cytonuclear discordance is observed (b), but the species and populations are overall well differentiated, as shown by the PCA (c) and the admixture plot (d). The  $d_{xy}$  along the 24 chromosomes for the different species pairs also show a clear genetic divergence of the populations, but two large outlier regions are observed on Chromosome 18 (e). *Amphiprion akallopisos*, *A. sandaracinos* and *A. perideraion* samples and populations are reported in shades of red, green and blue, respectively. (b) The mitochondrial phylogenetic tree was obtained with RAxML from the alignment of the whole mitochondrial genomes (16,747 bp), and the nuclear tree was obtained with RAxML and ASTRAL-III from 29,793,603 SNPs. Grey dots indicate bootstrap support lower than 0.8, and the arrows pinpoint the cytonuclear inconsistencies. The phylogenetic trees were rooted with *A. percula* as outgroup, which was removed from the plot. (c, d) The PCA and admixture results were obtained with PCAngsd on 29,793,603 SNPs. In the PCA, the first two components separate samples according to the species, whereas PC3 and PC4 separate the populations of *A. perideraion* and *A. akallopisos* according to geography. For the admixture analysis, multiple runs with different seeds for each  $K$  were performed to ensure convergence, and the highest likelihood solution was kept. The admixture plot represents the individual's ancestry proportions for the best number of ancestral populations  $K$  selected by PCAngsd ( $K=3$ ). Plots for different  $K$  are available in Figure S3. (e) The absolute genetic divergence ( $d_{xy}$ ) between the Indonesian populations of the three species.  $d_{xy}$  was computed along the 24 chromosomes, using sliding windows encompassing 3000 variable sites and corresponding to an average window size of 160 kb. Red dotted lines represent the upper and lower 1% of the  $d_{xy}$  distributions. Outlier regions of increased (*A. perideraion*-*A. akallopisos* and *A. perideraion*-*A. sandaracinos*) or decreased (*A. akallopisos*-*A. sandaracinos*) divergence are only observed on Chromosome 18. Similar between-species  $d_{xy}$  values and plots were obtained when using allopatric populations for the computation of  $d_{xy}$  (Figure S4). *Amphiprion* photos by Marc Kochzius.

three species: *Amphiprion akallopisos*, *A. perideraion* and *A. sandaracinos*. The three species show slight differences in colour patterns (Fautin & Allen, 1997). *A. akallopisos* and *A. perideraion* have a white caudal fin and an orange pinkish body colour, whereas the caudal fin and the body coloration of *A. sandaracinos* are orange (Figure 1a). *A. sandaracinos* has a slightly longer white stripe than the other two species, while *A. perideraion* has an additional white bar on the head (Figure 1a). *A. akallopisos* and *A. perideraion* have similarly shaped teeth (incisiform), which differ in *A. sandaracinos* (conical), potentially indicating slightly different ecological adaptations (Timm et al., 2008). We excluded the fourth species of the complex (*A. pacificus*) from the study. This species was more recently described (Allen et al., 2010) and little

information is available. It appears to be sister to *A. sandaracinos* (Allen et al., 2010), is only distributed in a limited number of islands in the South Pacific Ocean (Fiji, Tonga, Wallis and Samoa; Allen et al., 2010), and only found in sympatry at the margin of the distribution of *A. perideraion*. *A. pacificus* likely originated recently in the South Pacific Ocean by peripatric divergence from *A. sandaracinos* (Allen et al., 2010), and its evolution is not directly impacting the diversification history in the IAA of the three other species of the complex.

We collected individuals from the IAA region for each species (Table S2). *A. akallopisos* and *A. perideraion* populations were collected in Karimunjawa and Tulamben, where these two species and *A. sandaracinos* are found in sympatry (GBIF Occurrence Download

<https://doi.org/10.15468/dl.rzpk14>, accessed in January 2016; Hoey, Karimunjawa WCS fish data, accessed 2016). *A. sandaracinos* was collected in Manado, where it is sympatric to *A. perideraion* (GBIF Occurrence Download <https://doi.org/10.15468/dl.rzpk14>, accessed in January 2016). For *A. perideraion* and *A. akallopisos*, we also sampled populations at the margin of their geographic distribution, resulting in a total of three populations of *A. akallopisos* (Indonesia, Mayotte, and Kenya), two populations of *A. perideraion* (Indonesia and New Caledonia) and one population of *A. sandaracinos* (Indonesia). We sampled five to six individuals per population (Figure 1, Table S2). Fish were 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 official institutions of the different countries. A fin clip of one *A. percula* individual from Lizard Island, Northern Great Barrier Reef, Australia, was obtained from Fabio Cortesi (Queensland Brain Institute, University of Queensland, Australia) and used as an outgroup for subsequent analyses. We verified that the *A. percula* individual did not show signals of hybridization with members of the skunk complex, which could have biased the analyses (Appendix S1).

## 2.2 | DNA extraction, library preparation and DNA sequencing

Genomic DNA (gDNA) was extracted from the sampled tissue using a DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. We prepared short-insert (350bp) paired-end (PE) libraries from 100ng of gDNA using a TruSeq Nano DNA LT Library Preparation Kit (Illumina) following the manufacturer's instructions. The libraries corresponding to individuals of *A. sandaracinos*, *A. perideraion* from Indonesia, and the *A. percula* individual were sequenced on an Illumina HiSeq4000 at the Genomics Platform iGE3 (University of Geneva, Geneva), with a multiplex level of eight individuals per lane (read length: 100bp). The remaining libraries were sequenced on an Illumina HiSeq2000 at the Lausanne Genomic Technologies Facility, multiplexing five individuals per lane (read length: 100bp).

## 2.3 | Reads processing, mapping and variant calling

We removed adapter contamination from the raw reads with Cutadapt (v.1.13; Martin, 2011), trimmed the reads with Sickle (parameters: --qual-threshold 20, --length-threshold 40; v.1.33; Joshi & Fass, 2011) and verified their quality with FASTQC (v.0.11.5; Andrews, 2010). We mapped reads to the *A. percula* reference genome (INSDC Assembly [GCA\\_003047355.1](https://www.ncbi.nlm.nih.gov/assembly/GCA_003047355.1); Lehmann et al., 2018; Appendix S1) using BWA (v.0.7.15; Li & Durbin, 2009) and generated mapping statistics with bamtools (command *stats*, v.2.4.1; Barnett et al., 2011) and Picard tools (command *CollectInsertSizeMetrics* v.2.4.1; <http://broadinstitute.github.io/>

*picard*). 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 with a mapping quality higher than 30. We used *mergeReads* from ATLAS (v.0.9; Link et al., 2017) to remove redundant sequencing data originating from the overlap of paired reads and verified the absence of soft-clipped positions with *assessSoftClipping* from ATLAS (v.0.9; Link et al., 2017).

SNPs were called with the ATLAS pipeline (v.0.9; Link et al., 2017), which performs well at low-to-medium coverage and maintains a high accuracy in variant calling for moderately divergent species (Duchen & Salamin, 2021). Genotype likelihoods (GL) were computed with the *GLF* task (windows size of 0.1Mb), and SNPs were subsequently called with the *majorMinor* task (MLE method). Windows size of 0.1Mb allows reducing the computational effort while maintaining GL accuracy when the average coverage is at least 4 (Kousathanas et al., 2017). We filtered out invariant sites, low-quality SNPs and singletons using 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). With vcftools, we also generated a second SNPs dataset by removing the *A. percula* reads and subsequently filtering out resulting monomorphic positions and singletons. We repeated the SNP calling as described above but excluding the outgroup to assess its potential influence on SNPs specific to the skunk complex. Results were consistent, with a difference of only 326,456 SNPs (2% of the total number of SNPs). We computed SNPs densities for different window sizes (100, 150, 200 and 250kb) with vcftools (v.0.1.15; Danecek et al., 2011). Plotting was performed with the package ggplot2 (v.3.0.0; Wickham, 2016) in R (v.3.6.3; R core Team, 2020).

## 2.4 | Mitochondrial genome assembly

Mitochondrial genomes were reconstructed with MITObim (v.1.9; Hahn et al., 2013) by either using barcode sequences to initiate the assembly or previously published mitochondrial genomes as reference. The sequences GB KJ833753 (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 (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.

## 2.5 | Phylogenetic reconstruction

The mitochondrial genome of one *A. percula* individual was obtained from NCBI (GB NC\_023966.1; Tao et al., 2016). The mitochondrial genomes of all species were aligned using MAFFT (default

parameters; v.7.450; Katoh & Standley, 2013). We visually checked the alignment to remove poorly aligned regions and reconstructed the mitochondrial phylogenetic tree with RAxML (GTR+ $\Gamma$  model, 100 bootstrap; v.8.2.12; Stamatakis, 2014).

For the nuclear dataset, we generated SNPs alignments using the script parseVCF ([https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general)) and reconstructed phylogenetic trees for each chromosome separately using RAxML (GTR+ $\Gamma$  model, 100 bootstrap replicates, Lewis' ascertainment bias correction; v.8.2.12; Stamatakis, 2014). The nuclear phylogenetic tree was then inferred with ASTRAL-III (v.5.6.3; Zhang et al., 2018) from the chromosomal phylogenetic trees. The two resulting trees (mitochondrial and nuclear) were plotted with the *cophylo* command of the R package phytools (v.0.6.44; Revell, 2012). The *A. percula* individual was used to root the two phylogenetic trees and was removed for further analyses.

## 2.6 | Principal component analysis and admixture

We performed a principal component analysis (PCA) and admixture analysis with PCAnsd (v.0.98; Meisner & Albrechtsen, 2018). The PCA was performed on the whole dataset and on each chromosome separately to check for disparities. In the admixture analysis, we considered two to five ancestral populations ( $K$ ) and estimated individual ancestry coefficients by setting the parameter *-e* to the corresponding  $K-1$  value, as recommended in the PCAnsd guidelines. For each  $K$ , the analysis was run five times with different seeds to ensure convergence, and the outcome with the highest likelihood was retained. The best  $K$  was automatically selected by PCAnsd based on the MAP test.

## 2.7 | Estimation of population genomic metrics

We investigated the overall genomic differentiation between the populations and species by estimating the average  $F_{st}$  for each pair of populations using vcftools (v.0.1.15; Danecek et al., 2011; Appendix S2). Sliding windows measures of relative ( $F_{st}^r$ ) and absolute ( $d_{xy}$ ) genetic divergence and nucleotide diversity ( $\pi$ ) were calculated using *popgenWindows.py* ([https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general)). To account for the variability in SNPs density along the genome, window sizes were defined based on the number of variable sites (parameter: *--windType sites*). We performed analyses with windows encompassing 500–5000 variable sites, corresponding to an average window size of approximately 25–250 kb. The results for different window sizes were consistent (Table S3). In the main manuscript, we show the results for windows containing 3000 variable sites. We plotted the measures of  $\pi$ ,  $F_{st}^r$ , and  $d_{xy}$  along the genome using the *manhattan* function from the qqman R package (v.0.1.4; Turner, 2014). The regions of increased or decreased divergence (or nucleotide diversity) were defined as the windows in the upper or lower 1% of the  $F_{st}^r$ ,  $d_{xy}$  and  $\pi$  distributions.

## 2.8 | Topology discordance in the nuclear genome

We reconstructed phylogenetic trees for non-overlapping sliding windows along the genome with PhyML (v.3.3.2; Guindon et al., 2010) using *phyml\_sliding\_windows.py* ([https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general)) on the SNPs dataset including the *A. percula* individual, which was used to root the topologies. We used windows that were identical to those described above. We summarized the topologies by calculating the exact weightings of all possible subtrees using *twisst* (method “complete”, 10,000 sampling iterations; Martin & Van Belleghem, 2017). The individuals were grouped by species (species-level analysis) or populations (population-level analysis; see Appendix S3). Plots were produced in R with *plot\_twisst.R* provided in *twisst*.

## 2.9 | Tests for ancestral 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*. We investigated whether this disparity resulted from past hybridization events rather than from incomplete lineage sorting (ILS) by testing for signals of introgression between these two species using ABBA-BABA tests (Green et al., 2010). We computed the Patterson's  $D$  statistic for a genome-wide estimation of admixture (Durand et al., 2011; Green et al., 2010). Briefly, we estimated allele frequencies at each site for each population with *freq.py* ([https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general)). We considered *A. percula* as the outgroup and *A. perideraion* and *A. sandaracinos* as potentially hybridizing populations, and we calculated the  $D$  statistics with the equation (2) reported in Durand et al. (2011; Appendix S4). We applied a block jackknife procedure to test the significance of the  $D$  statistic using *jackknife.R* ([https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general)). To ensure independent blocks, we set a block size of 1 Mb, resulting in 903 blocks. We verified that the signal of introgression was consistent in all chromosomes and independent of the geographic origin of the populations (Appendix S4).

We estimated the proportion of admixture with the  $f$  statistics (Durand et al., 2011). We approximated the expected excess of ABBA over BABA sites under complete admixture by setting two *A. perideraion* populations (either geographic or random populations) as the hybridizing populations and estimating the site proportions as described above (Appendix S4). We obtained the standard error and 95% confidence interval of  $f$  by applying a block jackknife approach as described above.

We identified candidate genomic regions of introgression (CRI) along the genome between *A. perideraion* and *A. sandaracinos* by estimating the  $f_d$  statistics (Martin et al., 2015) with *ABBABABAwindows.py* ([https://github.com/simonhmartin/genomics\\_general](https://github.com/simonhmartin/genomics_general); Appendix S4). We used windows of 3000 variable sites but removed windows containing less than 100 biallelic SNPs to avoid stochastic errors in the  $f_d$  estimation (Martin et al., 2015).

We defined CRI as the windows in the top 5% of the genome-wide  $f_d$  distribution. This threshold was set based on the estimated genome-wide admixture proportion between *A. sandaracinos* and *A. perideraion* (i.e. 5.5% of admixed genomes). We visually verified that the regions of high  $f_d$  also supported the alternative mitochondrial topology. As introgressed regions generally show lower absolute genetic divergence ( $d_{xy}$ ; Martin et al., 2015; Smith & Kronforst, 2013), we tested for significant differences in  $d_{xy}$  between the CRI and the rest of the genome for each combination of species (Appendix S5). The presence of additional signals of admixture, besides that of *A. perideraion* and *A. sandaracinos*, was tested using TreeMix with 3- and 4-populations tests (v.1.13.; Pickrell & Pritchard, 2012; Appendix S6).

## 2.10 | 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 15,975 genes. We expanded this functional annotation based on the available information from the *A. frenatus* reference genome (Marcionetti et al., 2018). This resulted in 17,179 annotated genes, of which 14,002 had *biological process* gene ontologies annotations (GOs; Appendix S7). We performed GO enrichment analysis of the genes located within (or partially overlapping with) the CRI using the topGO package (parameters: Fisher's exact tests, weight01 algorithms, minimum node size of 3; v.2.26.0; Alexa & Rahnenfuhrer, 2016) and contrasting them against all the annotated protein-coding genes of *A. percula*. Results were considered significant at  $p$ -values  $<.01$ . No corrections for multiple testing were performed, following recommendations from the topGO manual. An identical strategy was employed for the GO enrichment analysis of the genes located in the regions of increased/decreased genomic divergence between species.

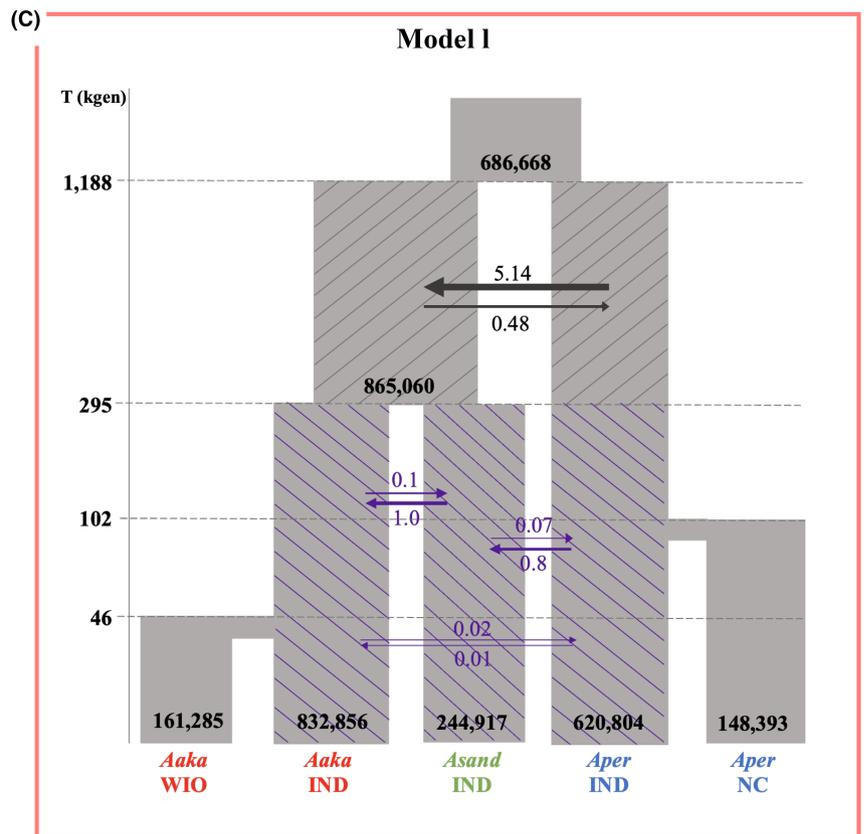
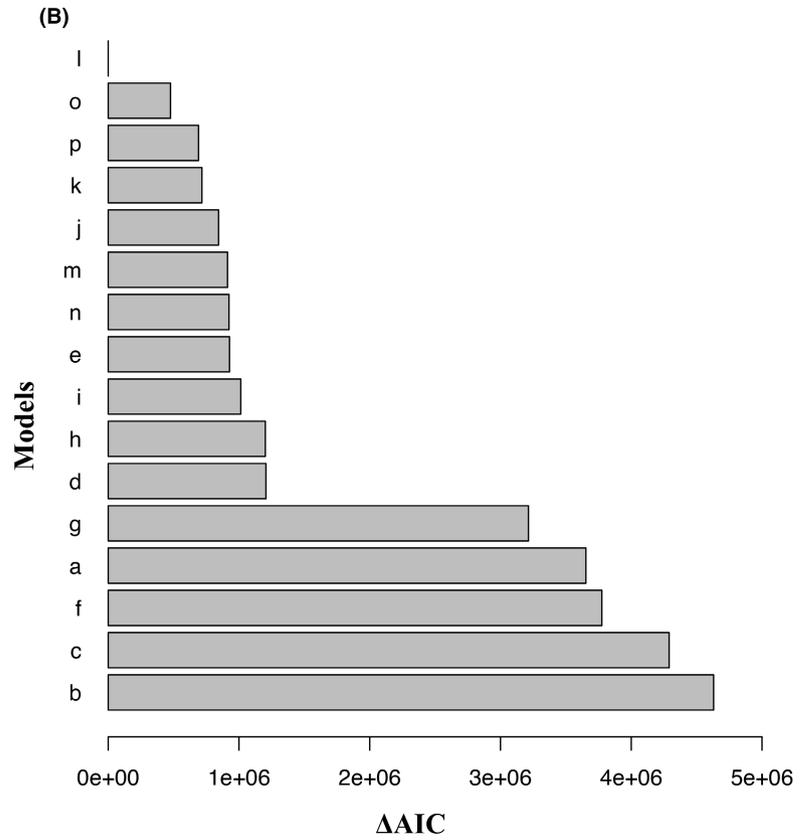
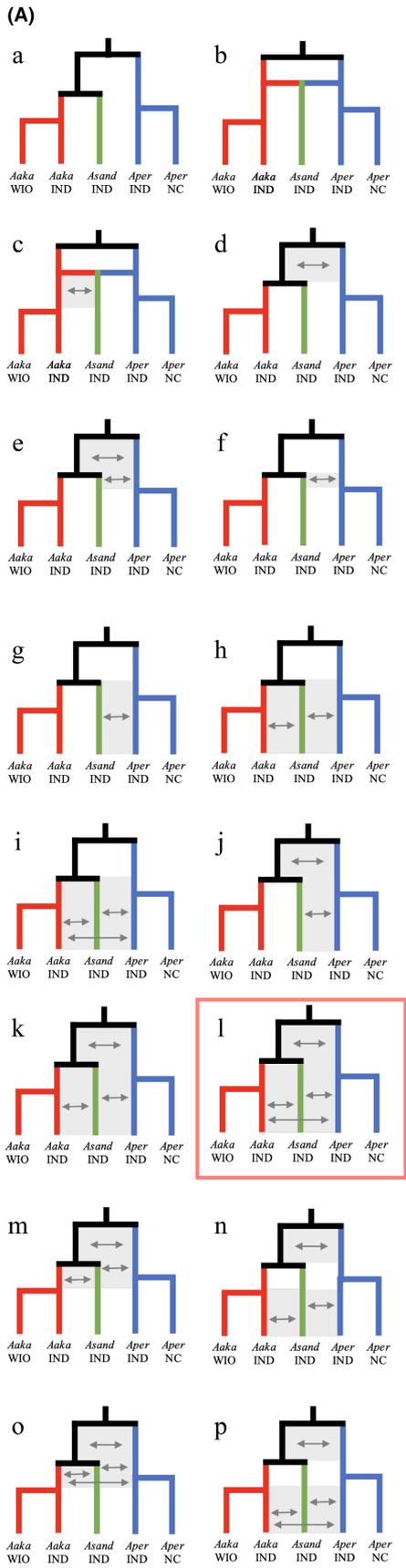
## 2.11 | Tests of hybrid speciation, gene flow and demographic reconstruction

To infer the demographic scenario that best fitted the genomic data and test the hypotheses for the evolution of the skunk complex,

we performed model comparison under the coalescent framework developed in fastsimcoal2 (v.2.6; Excoffier et al., 2013). We removed SNPs with missing data using vcfTools (--max-missing 1; total SNPs kept: 16,547,283; v.0.1.15; Danecek et al., 2011), and we complemented this dataset with  $6.1 \times 10^8$  monomorphic sites (Appendix S8). We treated the two populations of *A. akallopisos* from the WIO (Kenya, Mayotte) as a single population. We extracted the multidimensional folded site frequency spectra (SFS) for each population using easySFS (<https://github.com/isaacovercast/easySFS>), considering 10 samples per population (Appendix S8). The multidimensional SFS were used to compare 16 distinct demographic models (Figure 2A, Appendix S8). We first compared a model of strict isolation with models of hybrid origin of *A. sandaracinos* followed by strict isolation or by asymmetric gene flow with *A. akallopisos* (Figure 2A,a-c). We further compared these models with scenarios of ancestral asymmetric gene flow between *A. perideraion* and either the *A. akallopisos* + *A. sandaracinos* ancestor or *A. sandaracinos* (Figure 2A,d-f). We then built models to investigate whether more recent gene flow (i.e. throughout the divergence of the species) on its own (Figure 2A,g-i) or with ancestral gene flow (Figure 2A,j-l), was likely. Finally, we explored the timing of the recent gene flow by comparing models with gene flow occurring only in the populations of the IAA (i.e. after the split of the allopatric populations; Figure 2A,n,p), and models where gene flow was only possible before the split of the allopatric populations (Figure 2A,m,o). We did not include intraspecific gene flow in our demographic scenarios because it would have required the evaluation of additional and even more complex demographic models, which was out of the scope of this study. This choice could lead to an underestimation of divergence times if intraspecific gene flow is occurring between the populations (Barley et al., 2018; Leaché et al., 2014). It could also affect the estimation of population sizes, but we are only interested here in relative comparisons of  $N_e$ , which have been shown to be valid despite the underestimation of gene flow (Leaché et al., 2014).

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 & García, 2007) and set a mutation rate of  $4.0 \times 10^{-8}$  (derived from Delrieu-Trottin et al., 2017, see Appendix S8). We identified

**FIGURE 2** Schematic summaries of the 16 demographic models representing the possible diversification scenarios of the skunk complex (A) and AIC scores obtained for them (B). Model I with the best-fit is highlighted in red in (A), and a magnified version with the estimated parameters is reported in (C). The divergence times are reported in thousands of generations (kgen), and effective population sizes ( $N_e$ , reported in black in the plot) are haploid sizes. Arrows represent gene flow between the populations (represented here forward in time), with numbers  $\times 10^{-6}$  corresponding to the estimated migration rates ( $m$ ). Effective migration rates are calculated as  $2N_e m$ . The exchanging populations are highlighted using hatching, with different colours (grey and purple) corresponding to the timing of gene flow. *Aaka*, *Asand* and *Aper* correspond to *A. akallopisos*, *A. sandaracinos* and *A. perideraion*, respectively. WIO, IND and NC correspond to the geographical origin of the populations (Western Indian Ocean, Indonesia and New Caledonia, respectively). (A) Arrows in the schematic models represent gene flow between the populations. Models of strict isolation (a), hybrid origin of *A. sandaracinos* (b, c) and divergence with different levels and timing of gene flow (d-p) were tested. (B) For each model, the AIC scores are represented as  $\Delta AIC$ , which is the difference between the AIC scores of each model and the best model. Additional information on the analyses and results is provided in Appendix S8.



the best-fitting demographic model based on both the AIC and the likelihood distributions obtained for each model (Appendix S8).

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 et al., 2018). For each bootstrap replicate, we performed 30 independent runs with 200,000 simulations and 20 ECM cycles. We calculated point estimates of effective migration rates (i.e. “gene flow”) with  $2 N_e m$  (with  $N_e$  the haploid effective population sizes and  $m$  the estimated migration rate), obtaining the number of gene copies exchanged each generation (Bourgeois et al., 2020). We defined limited ( $N_e m = 0.01-0.1$ ), weak ( $N_e m = 0.1-1$ ) and moderate ( $N_e m = 1-10$ ) gene flow as in Samuk and Noor (2022).

### 3 | RESULTS

#### 3.1 | Whole genome sequencing, mapping and SNP calling

The number of raw paired-end reads (PEs) for each individual ranged from c. 29 million pairs to c. 61 million pairs, with a median of c. 48 million PEs per sample (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 12.0x depending on the individual (Table S4). Mapping to the *A. percula* reference genome resulted in the mapping of 84%–88% of reads, with a final average coverage ranging between 4.6x and 10.2x (Table S4).

The SNP calling strategy resulted in 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 (Figure S2). When the *A. percula* individual was included, the number of SNPs increased to 29,793,603 (Table S5).

#### 3.2 | Cytonuclear discordances in the skunk complex

We explored the phylogenetic relationship between the samples by reconstructing mitochondrial and nuclear phylogenetic trees. We found that both trees showed a clear separation of the three species, followed by a separation of populations depending on their geographical origin (Figure 1b). Only the samples of *A. akallopisos* from the two populations of the WIO were not differentiated at both the nuclear and mitochondrial data.

Despite a clear separation of the species, the two genomes showed discordant topologies (Figure 1b). 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 conspecifics at the nuclear level, but clustered within the *A. perideraion* group from Indonesia in the mitochondrial tree (bootstrap support >0.95; Figure 1b).

#### 3.3 | Overall clear divergence of the species and populations

We investigated the overall population structure by performing a PCA on the nuclear SNPs. We found that the first two axes split the three species into distinct clusters and explained 59.4% and 15.3% of the variance, respectively (Figure 1c). The first axis separated the *A. sandaracinos* and *A. akallopisos* individuals from *A. perideraion*, while the second separated *A. akallopisos* from *A. sandaracinos*. The third and fourth components explained 3.8% and 2.3% of the variance, respectively. They separated the populations based on their geographical origin (Figure 1c). 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 individuals by species and geography. Indeed, the best number of ancestral populations was inferred to be  $K=3$  and clustered the samples by species (Figure 1d), but populations were separated with increasing  $K$  ( $K=4$  and  $K=5$ ; Figure S3).

In both the PCA and admixture results, individuals from the sympatric populations of the IAA never clustered together (Figure 1c,d). Some signals of shared ancestry among species and populations were nevertheless observed in the admixture plots. For instance, with  $K=3$  and  $K=4$ , samples of *A. akallopisos* from Indonesia showed a low proportion (6%) of shared ancestry with *A. sandaracinos*, while at  $K=2$ , a low proportion (13%) of *A. sandaracinos* ancestry was in common with *A. perideraion* (Figure 1d, Figure S3).

Overall, measures of genomic divergence mirrored the PCA and admixture results. The averages  $F_{st}$  calculated between all populations showed a clear species ( $F_{st}$  from 0.27 to 0.55) and population ( $F_{st}$  around 0.1) divergence, except for the two *A. akallopisos* populations of the WIO ( $F_{st}$  of 0.01; Appendix S2). Nevertheless, we identified high heterogeneity in the absolute ( $d_{xy}$ ) and relative ( $F_{st}$ ) genetic divergence across the genome (Figure 1e, Figures S4 and S5). 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 (Table S3) or the considered populations within each species (Figures S4 and S5). We looked for outlier regions of divergence (i.e. upper/lower 1% of the  $F_{st}$  and  $d_{xy}$  distributions) and found that all the windows of increased  $d_{xy}$  and  $F_{st}$  clustered in two regions of chromosome 18 for the comparisons *A. akallopisos*–*A. perideraion* and *A. sandaracinos*–*A. perideraion* (Figure 1e, Figure S5). These two regions also showed the lowest  $d_{xy}$  between *A. akallopisos* and *A. sandaracinos*, and they were characterized by a reduced

nucleotide diversity ( $\pi$ ) in all populations (Figure 1e, Figures S4–S6). In these two regions, extending from c. 2.9 to 3.5 Mb and from c. 7.2 to 16.9 Mb, we identified 408 functionally annotated genes, whose GO enrichment analysis resulted in 13 enriched GOs ( $p$ -value  $< .01$ , Table S6). Among them, we observed terms associated with the regulation of behaviour (GO:0050795), the development of endoderm (GO:0007492) and the morphogenesis of the epithelium (GO:1905332).

### 3.4 | Evidence from ancestral hybridization events and introgression in *A. sandaracinos*

We investigated the presence of topological inconsistencies along the nuclear genome of the skunk complex that may reflect ancestral gene flow before the split of populations within each species. Three different rooted topologies are possible linking the three species (Figure 3a). The most frequent topology was the nuclear phylogenetic tree (“nuclear” topology in Figure 3a), which had an average weighting higher than 80.0% and was fully supported in 74.8% of the windows (Figure 3a, Table S7). The topology of the mitochondrial phylogenetic tree (“mitochondrial” topology in Figure 3a) was also represented in a relatively high proportion of the genome, being fully supported in 7.8% of the windows and having an average weighting of 13.0% (Figure 3a, Table S7). The last possible topology (“alternative” topology in Figure 3a), characterized by *A. akallopisos* and *A. perideraion* being sister species, was fully supported in only 1.8% of the windows and had an average weighting of 7.0% (Figure 3a, Table S7). Overall, the three topologies were homogeneously distributed across the 24 chromosomes (Figure S7, Table S8). However, the weightings for the “mitochondrial” and “alternative” topologies were reduced on chromosome 18 (weightings of 4.6% and 1%, respectively) due to the presence of two large regions with full support for the “nuclear” topology (Figure 3b, Figure S7, Table S8). These two regions coincided with two outlier regions of divergence (Figure 1e).

The frequency of the “mitochondrial” topology along the nuclear genome (Figure 3a), together with the cytonuclear discordance (Figure 1b) and the admixture result for  $K=2$  (Figure S3), suggested the presence of ancestral gene flow between *A. sandaracinos* and *A. perideraion*. This hypothesis was confirmed by significant and positive Patterson's  $D$  statistics ( $0.171 \pm 0.0041$ ; Z-score of 41.9;  $p$ -value  $< 1E-045$ ; Figure 3c, Table S9), with an overall proportion of admixture between the two species estimated to be 5.6% ( $f$  statistics = 0.056, 95% CI [0.052–0.060]). The signal of admixture was detected in all 24 chromosomes, all of which showed significant  $D$  statistics (Figure 3c, Table S9) and the presence of candidate regions of introgression between the two species (CRI; upper 5% of the  $f_d$  distribution; Figure 3d, Figure S8). Chromosome 18 differed from the remaining chromosomes, showing the lowest values for the  $D$  statistics ( $0.1 \pm 0.03$ , Z-score of 3.15; Figure 3c, Table S9) and containing two large regions with an extremely low  $fd$  ( $< 0.01$ ) and the absence of CRI (Figure 3d).

We investigated the gene content of the CRI and identified 905 functionally annotated genes and 21 enriched GOs ( $p$ -value  $< .01$ , Table S10). We observed terms associated with the sensory perception of a light stimulus (GO:0050953), the adult feeding behaviour (GO:0008343), the neuropeptide signalling pathway (GO:0007218), as well as terms linked with the immune system (Table S10).

### 3.5 | Weak evidence for ongoing extensive hybridization in the IAA

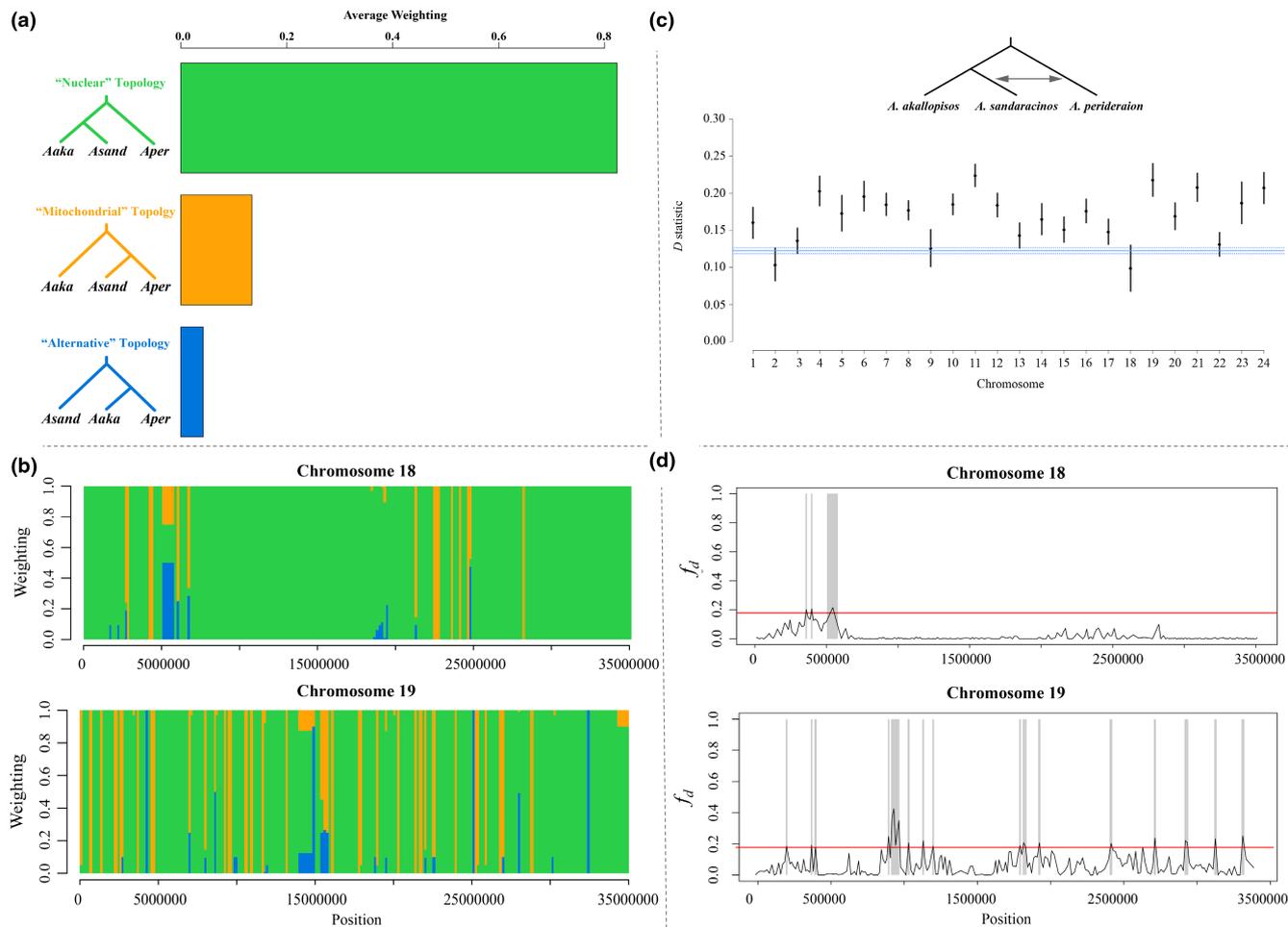
While the Indonesian populations of *A. perideraion*, *A. sandaracinos* and *A. akallopisos* were differentiated from each other (Figure 1c) and showed a relatively high genomic divergence (Appendix S2, Figure S4), the admixture analysis suggested some level of shared ancestry between the Indonesian populations of *A. akallopisos* and *A. sandaracinos* (Figure S3). We investigated possible gene flow events in the IAA between the populations with TreeMix, which detected migration events between the Indonesian populations of *A. akallopisos* and *A. sandaracinos*, besides gene flow between *A. sandaracinos* and *A. perideraion* (see Appendix S6).

We explored the presence of topological inconsistencies at the population level, with topologies grouping the three Indonesian populations together, which would suggest recent (i.e. after the population splits) gene flow in the IAA. We found an average topology weighting of only 3.4% for the topologies grouping the Indonesian populations of *A. akallopisos* and *A. sandaracinos*. 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%). In addition, we did not find any windows fully supporting any of these topologies. For more details, see Appendix S3.

### 3.6 | Demographic reconstruction

We contrasted different diversification scenarios with or without gene flow (Figure 2), potentially resulting in the observed patterns of admixture between the species. We obtained the highest likelihood and lowest AIC for model I, in which gene flow happened between *A. perideraion* and the ancestor of *A. akallopisos* + *A. sandaracinos* and between all species throughout the diversification of the group (Figure 2, Appendix S8). Models of strict isolation and hybrid speciation resulted in the lowest likelihoods and highest AIC, and the additional models tested also showed reduced support compared to model I, with no overlap in the likelihood distributions (Figure 2, Appendix S8). Thus, we rejected the hypotheses of a hybrid origin of *A. sandaracinos*, of strict isolation during the group diversification, and the scenarios represented by the additional models.

Parameter estimates obtained for the best model reported the split between the three species to have occurred c. 1.2 million generations ago and the one between *A. akallopisos* and *A. sandaracinos*



**FIGURE 3** Whole-genome average weighting for each species-level topology (a) and examples of topological weighting across chromosomes 18 and 19 (b). (c) The “mitochondrial” topology along the nuclear genome suggested ancestral gene flow between *A. sandaracinos* and *A. perideraion*, which was confirmed by the significantly positive values of Patterson's  $D$  statistics. (d) Examples of the  $f_d$  distribution along Chromosomes 18 and 19. Candidate regions of introgression (CRI) between *A. sandaracinos* and *A. perideraion* are reported in grey (d). (a, b) Results were obtained with *twisst*, for sliding windows of 3000 variant sites. The weightings are determined by successively sampling 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, and the average weightings are calculated genome-wide. Patterns observed on Chromosome 19 are similar to those of the other chromosomes, while Chromosome 18 shows two large regions of full support for the “nuclear” topology (green). (c) The  $D$  statistics  $\pm$  SE are reported for each chromosome. Plotted in blue is the genome-wide estimate of  $D \pm$  SE. Standard errors were obtained with a block jack-knife approach. All chromosomes show a  $D$ -statistic significantly deviating from zero, indicating introgression between *A. perideraion* and *A. sandaracinos*. (d) Red lines represent the 95th percentile of the genome-wide  $f_d$  distribution, which was set as the cut-off for determining the CRI. Here also, patterns observed on Chromosome 19 are similar to those of the other chromosomes, while Chromosome 18 shows two regions of extremely low ( $<0.01$ )  $f_d$  and a reduced number of CRI.

c. 0.3 million generations ago (Figure 2, Appendix S8). Considering a generation time of 5 years for clownfishes (Buston & García, 2007), this corresponds to a divergence time of c. 6 MYA and 1.5 MYA, respectively. The colonization of New Caledonia by *A. perideraion* was estimated to be c. 509 kYA (101,809 generations), while the split of *A. akallopisos* populations appears to be more recent (ca 45,700 generations 228 kYA; Figure 2c, Appendix S8). The lowest population size estimates were observed for the New Caledonian population of *A. perideraion* and the *A. akallopisos* population from the WIO, while *A. sandaracinos* showed the smallest effective population size in the IAA (Figure 2c, Appendix S8). Estimates of the migration rates for the best model ranged between  $1.3 \times 10^{-8}$  and  $5.14 \times 10^{-6}$ , depending

on the timing and species considered (Figure 2c, Appendix S8). The highest rate ( $5.14 \times 10^{-6}$ ) was observed from *A. perideraion* to the ancestor of *A. akallopisos* + *A. sandaracinos* and corresponded to an effective migration rate ( $2 N_e m$ ) of 6.38 gene copies per generation. We obtained lower migration rates from *A. sandaracinos* to *A. akallopisos* ( $1 \times 10^{-6}$ ) and from *A. perideraion* to *A. sandaracinos* ( $7.84 \times 10^{-7}$ ), corresponding to effective migration rates ( $2 N_e m$ ) of 0.49 and 0.97 gene copies per generation, respectively (Figure 2c, Appendix S8). The lowest migration rates were observed from *A. akallopisos* to *A. perideraion* ( $1.82 \times 10^{-8}$ ) and vice versa ( $1.29 \times 10^{-8}$ ) and corresponded to effective migration rates of 0.03 and 0.02 gene copies per generation, respectively (Figure 2c, Appendix S8).

## 4 | DISCUSSION

In addition to ecological speciation, gene flow between species and hybrid speciation were previously hypothesized to be implicated with the divergence of the skunk group (Litsios et al., 2012; Litsios & Salamin, 2014). Here, we took advantage of the disjunct geographical distribution of the group to explore these hypotheses and investigate the processes characterizing the diversification of the complex. We rejected the scenario of hybrid origin of *A. sandaracinos* and demonstrated that weak to moderate gene flow occurred throughout the diversification of the group. We identified introgressed regions in *A. sandaracinos* and detected in *A. perideraion* two large regions of high divergence from the two other species. While a clear link with the species' ecology remains challenging to draw, these results provide the first insights into the evolution of this complex.

Genomic comparisons demonstrated that the three species of the complex were, overall, well-differentiated, with first the separation of *A. perideraion* from *A. akallopisos* + *A. sandaracinos* and then the split of the latter two. Our results further indicated the presence of cytonuclear discordance in the group as previously reported (Litsios & Salamin, 2014), and similar topological inconsistencies were also observed along the nuclear genome. Furthermore, an individual of *A. sandaracinos* (GB227) harboured a mitochondrial genome clustering with the Indonesian population of *A. perideraion*. While such inconsistencies could purely arise from ancestral polymorphism and incomplete lineage sorting (e.g. Lee-Yaw et al., 2019; Toews & Brelsford, 2012), we found evidence of distinct hybridization events that occurred at different stages of the diversification of the skunk complex. Demographic modelling allowed us to rule out a hybrid origin of *A. sandaracinos* and identify a complex evolutionary scenario involving moderate gene flow (effective migration rate of 6.38 gene copies per generation) from *A. perideraion* to the *A. akallopisos* + *A. sandaracinos* ancestor and weak gene flow between the three species throughout their evolution. Within these weak exchanges between the analysed IAA populations, the highest effective migration rate (0.97) was observed from *A. perideraion* to *A. sandaracinos*. These findings were further supported by the results of ABBA-BABA tests, which indicated past hybridization events between *A. perideraion* and *A. sandaracinos*.

Candidate regions that were introgressed in *A. sandaracinos* genomes following hybridization with *A. perideraion* (candidate regions of introgression, CRI) could be identified. However, the mechanisms underlying the fixation of these CRI (genetic drift or selection) remain unclear. CRI contains genes with functions related to feeding behaviour, 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* and *A. sandaracinos*, which would suggest a role of natural or even sexual selection in the fixation of the regions (i.e. adaptive introgression; Hedrick, 2013). These past hybridization events between the two species could also account for the observed cytonuclear discordance. Hybridization could have led to the introgression and replacement

of *A. sandaracinos* mitochondrial DNA by that of *A. perideraion* (i.e. mitochondrial capture, Perea et al., 2016; Toews & Brelsford, 2012), a phenomenon already documented in other Pomacentridae in the same geographic context (e.g. Bertrand et al., 2017). The processes of hybridization and introgression likely occurred more than once, as suggested by the additional inconsistency observed for the individual GB227 (*A. sandaracinos* with mitochondrial genome of Indonesian *A. perideraion*).

The signal of hybridization that we found with the ABBA-BABA tests was independent of the population of *A. perideraion* considered (Indonesian or New Caledonian). The detected introgression likely predated the divergence between the two *A. perideraion* populations, but the sample GB227 further indicates that gene flow in the IAA likely occurred also after the split of the two *A. perideraion* populations. Since the individual GB227 did not show additional *A. perideraion* ancestry at the nuclear level, selective backcrossing might have erased the extra introgression signal from the nuclear DNA but still resulted in the cytonuclear discordance observed (Bertrand et al., 2017; Perea et al., 2016; Toews & Brelsford, 2012). A lack of power in the analysis could also be conceivable as only one distinctive individual was present (GB227), and the extent of additional introgression was likely small (no evident pattern in the admixture plot or the *twisst* analyses).

Interspecific gene flow was detected not only between *A. perideraion* and *A. sandaracinos* but generally between the three analysed populations in the IAA. The levels of gene flow were limited (effective migration rates from 0.01 to 0.97), with the lowest migration rates observed between the IAA populations of *A. akallopisos* and *A. perideraion*. Hybridization events involving members of the skunk complex are prevalent in the aquarium trade and were expected to be frequent also in nature (Steinke et al., 2009), likely facilitated by the possible cohabitation of different clownfish species within the same sea anemone hosts (Table S1; Gainsford et al., 2015). On the one hand, the level of gene flow from *A. akallopisos* to *A. sandaracinos* (effective migration rates of 0.14) could be underestimated and only partially reflect what is occurring in fully sympatric regions, because the *A. sandaracinos* population came from Manado, where *A. akallopisos* is not present. Thus, higher gene flow from *A. akallopisos* to *A. sandaracinos* could be present in sympatry, with the admixed populations then dispersing and gradually diluting the introgression signal as the distance from the sympatric region increases. To confirm that, sequencing of additional *A. sandaracinos* populations from sympatric locations such as Karimunjawa and/or Tulamben would be necessary. On the other hand, despite being from different locations, Indonesian populations of *A. akallopisos* and *A. perideraion* were collected at sites where the three species are sympatric. While we cannot exclude that populations experiencing stronger interspecific gene flow exist but were not sampled, our results suggest that, at least within these sympatric populations, gene flow is less pervasive than initially thought.

Interspecific competition drives host repartition and specialization in regions of co-occurrence, pushing the different

clownfish species to settle in distinct sea anemone host species (Garcia et al., 2023). Thus, micro-habitat specialization might further reduce cohabitation and interactions among different species in the IAA. This could reduce (or even prevent) hybridization events, thereby preserving clear genetic boundaries between species, as observed in this study. Species-specific behaviour that contributes to maintaining reproductive isolation could also limit interspecific gene flow. In particular, the social structure consisting of a dominant breeding pair and subordinate non-breeding individuals is likely to reduce the opportunities for hybridization by maintaining a stable reproductive couple over longer periods (several years in some cases; Gainsford et al., 2015). Opportunities for introgression could also be reduced by strong mate selection, with hybrid individuals struggling to find suitable mates, as they may not fit the preferences of individuals from either parental species.

The level of gene flow observed between the analysed populations in the IAA was, thus, weaker than initially hypothesized. This limited level of gene flow prevented the genomic homogenization that would be essential for uncovering peaks of high differentiation potentially linked with ecologically important traits or the speciation process. In comparison, effective migration rates higher than nine gene copies/generation were detected between highland and lowland bird forms of the Réunion grey white-eye, *Zosterops borbonicus*. Such gene flow resulted in the homogenization of the genome and the detection of loci involved in ecological adaptation (Bourgeois et al., 2020). Similarly, genome-wide homogenization of two *Ramphocelus* tanager species was obtained thanks to even higher (>20) migration rates, allowing the identification of loci involved in plumage colour divergence (Luzuriaga-Aveiga et al., 2021).

Peaks of high differentiation resulting from genomic homogenization were not observed in the analysed populations of the IAA. Singular patterns of genomic differentiation were found on chromosome 18, with two large regions (from c. 2.9 to 3.5 Mb and from c. 7.2 to 16.9 Mb) of reduced genetic divergence between *A. sandaracinos* and *A. akallopisos* but increased divergence with *A. perideraion* (Figure 1e, Figures S4 and S5). Previous phylogenomic and comparative studies on clownfish showed that in these same regions of chromosome 18, the relationship between clownfish species was distinct compared to the rest of the genome (Marcionetti & Salamin, 2023; Schmid et al., 2022). For instance, *A. perideraion* was not sister to the other species of skunk complex but to the *ephippium-clarkii* clade (Marcionetti & Salamin, 2023; Schmid et al., 2022). These regions also showed increased signals of introgression between different clownfish clades, such as between *A. perideraion* and the members of the *ephippium-clarkii* clade (Marcionetti & Salamin, 2023; Schmid et al., 2022). The authors suggested that the hybridization signals on chromosome 18 were potentially maintained across the whole clownfish group through genomic inversions disrupting the recombination (Marcionetti & Salamin, 2023; Schmid et al., 2022). This was recently confirmed with long-read sequencing, which highlighted a large inversion on *A. perideraion* chromosome 18 that was absent

in *A. akallopisos* (Marcionetti A, Salamin N, unpublished data). *A. sandaracinos* individuals likely carry the non-inverted haplotype, as suggested by the lower divergence observed with *A. akallopisos* on chromosome 18. Additionally, the two haplotypes (inverted or not) seem to be fixed in the species, as the genomic divergence results obtained with the different populations were consistent (Figures S4 and S5). This hypothesis must be confirmed by assessing additional samples and populations per species. Genomic inversions are often responsible for the persistence of supergenes (e.g. Branco et al., 2018; Küpper et al., 2016; Zinzow-Kramer et al., 2015). In these regions of chromosome 18, we found genes potentially involved in the ecological preference of the species, with, for instance, enrichment for genes associated with the regulation of behaviour. The role played by these regions—if any—in species divergence cannot be readily determined without further investigations.

The diversification of the skunk complex was characterized by recurrent gene flow events between the three species. We do not exclude that additional factors influencing population dynamics were at play during the evolution of the group. For instance, Pleistocene glacial cycles resulted in sea-level fluctuations that seriously impacted coastal marine life, resulting in changes in range and population sizes (Ludt & Rocha, 2015). Such events likely affected each of the three species analysed here but studying population dynamics would require many additional populations to trace the evolution of each species separately, a different aim than the present study. Because the population dynamics within each species—such as bottlenecks—were not explicitly modelled in the study, estimates such as divergent times and population sizes obtained here should be taken cautiously (Barley et al., 2018; Leaché et al., 2014; Momigliano et al., 2021). The absence of bottleneck events in the models does not affect the conclusion about gene flow between species during their diversification. The issue arising from not modelling bottleneck events when they, in fact, occurred would be a lack of power to detect gene flow rather than an overestimation of it (e.g. Momigliano et al., 2021).

We observed that, within each species, populations were distinctly separated by geography, especially when large distances separated the populations. The divergence between the IAA populations of *A. perideraion* and *A. akallopisos* and the peripheral populations (i.e. NC and WIO, respectively) was relatively high, while within the WIO, the differentiation between *A. akallopisos* populations was lower (Figure S4). The results obtained for the *A. akallopisos* populations were consistent with the patterns of divergence previously observed for the species (Huyghe & Kochzius, 2017, 2018). Precisely, in *A. akallopisos*, the connectivity between the IAA and the WIO (c. 8000 km) was reduced because of the low dispersal ability of the species, and intraspecific gene flow was likely possible only through a stepping-stone model over several generations (Huyghe & Kochzius, 2017). In the WIO, the reduced geographical distances and the presence of the South Equatorial and the East African Coastal Currents maintained higher connectivity between populations (Huyghe & Kochzius, 2018). Because

of the limited number of populations per species available here, the generalization of the findings to explain the evolutionary history of each species separately needs caution. The considerable geographic distance between the IAA and New Caledonia (c. 6000 km) and the low potential for long-range dispersal in clownfishes (<100 km; Almany et al., 2017; Planes et al., 2009; but see Simpson et al., 2014 for dispersal of 400 km for *A. omanensis*) suggest that a scenario of gene flow through a stepping-stone process could also be expected for *A. perideraion*. Additional population genetic studies focusing on this species alone would be needed to verify this hypothesis.

We considered here the three main species of the skunk complex, but did not consider the fourth species, *A. pacificus*. *A. pacificus* was recently described, and little information is available on the species (Allen et al., 2010). It is thought to have originated in the South Pacific Ocean by peripatric divergence from *A. sandaracinos*, and it is present exclusively on a few islands there (Fiji, Tonga, Wallis and Samoa; Allen et al., 2010). Samples for this species were not available, but including it in the study would not have altered the conclusions on the processes characterizing the diversification of the skunk complex in the IAA. The origin of *A. pacificus* is intriguing, and further studies should be conducted to investigate its evolutionary history.

## 5 | CONCLUSION

Our study on the skunk group's diversification provides the first insights into the multiple mechanisms marking the evolution of this complex. We were able to reject the hypothesis of a hybrid origin for *A. sandaracinos* but unveil the occurrence of moderate ancestral gene flow from *A. perideraion* to the *A. akallopisos*+*A. sandaracinos* ancestor. A low level of gene flow was also detected between the species in sympatry. However, such a level was insufficient to impact the overall genetic divergence significantly, suggesting potential host repartition and behaviour-mediated reproductive barriers as factors maintaining genetic identity in sympatry. Furthermore, our study detected introgressed regions from *A. perideraion* to *A. sandaracinos* and also identified two large regions of increased divergence on Chromosome 18, indicative of potential additional hybridization events with species outside the skunk complex. The findings provide valuable insights into the mechanisms having influenced the divergence of this intriguing species complex. However, further research is necessary to better understand the specific demographic, ecological and behavioural factors responsible for maintaining genetic isolation among sympatric species within the skunk group, and, at a broader scale, advance our understanding of speciation processes in diverse ecosystems. Investigating how hybridization and gene flow influence the evolutionary trajectory of species can provide valuable insights for conservation efforts, especially in the context of biodiversity management and the preservation of unique and complex ecosystems.

## AUTHOR CONTRIBUTIONS

JAMB, AM and NS designed the study. AM and JAMB performed the research and analyses. FC, MK, LP, GFAD, SH, AM and JAMB contributed to the sample collection. AM wrote the manuscript. All authors read, made corrections and approved the final version of the manuscript.

## ACKNOWLEDGEMENTS

We thank the local authorities of Indonesia, Kenya, Mayotte and New Caledonia for the permits to collect samples and their help with field logistics. A special thanks goes to the Quesnel family for the help and logistical support they provided during fieldwork in New Caledonia. We also thank the staff at the Lizard Island Research Station for fieldwork support, and acknowledge the Dingaal, Ngurrumungu and Thanhil peoples as traditional owners of the lands and waters of the Lizard Island region. We are appreciative of the Lausanne Genomic Technology Facility for the sequencing and the DCSR computing infrastructure of the University of Lausanne for computational support/resources. We thank Diego A. Hartasánchez and two anonymous reviewers, whose constructive comments helped to improve the manuscript. Open access funding provided by Universite de Lausanne.

## FUNDING INFORMATION

The work was supported by funds from the University of Lausanne, the Swiss National Science Foundation, Grand Number: 31003A-163428. FC was supported by an Australian Research Council (ARC) Discovery Early Career Research Award DE200100620 and Discovery Project DP180102363.

## CONFLICT OF INTEREST STATEMENT

The authors declare that there are no competing interests in the publication of this work.

## DATA AVAILABILITY STATEMENT

Raw Illumina reads are deposited in the SRA (BioProject PRJNA1022585) and mitochondrial assemblies are available in the Zenodo repository (<https://doi.org/10.5281/zenodo.10775242>). Metadata are also stored in the SRA (BioProject PRJNA1022585) using the MIGS eukaryote version 6.0 package.

## BENEFIT-SHARING STATEMENT

Benefits Generated: Benefits from this research accrue from the sharing of our data and results on public databases as described above. All fieldwork was performed in agreement with local regulations and in collaboration with local entities (University of Mayotte; University of New Caledonia; Kenya Marine and Fisheries Research Institute; Universitas Hasanuddin, Indonesia) following the "Access and Benefit Sharing" (ABS) principles of the Nagoya Protocol established by the Convention of Biological Diversity (CBD). We thank the local authorities for permits to collect samples and help with field logistics in Indonesia, Mayotte, Kenya, New Caledonia and Australia. Research permits number are the

following: Mayotte: 06/UTM/2016; New Caledonia: N°60912-895-2017/JJC; Kenya: NCST/RRI/12/1/BS/250; Indonesia: 55/PSTK/UH/XI/04; Australia: Great Barrier Marine Park (GBRMPA): G17/38160.1.

## ORCID

Anna Marcionetti  <https://orcid.org/0000-0002-2450-2879>

Joris A. M. Bertrand  <https://orcid.org/0000-0002-3379-1019>

Filip Huyghe  <https://orcid.org/0000-0002-9352-9098>

Loïc Pellissier  <https://orcid.org/0000-0002-2289-8259>

## REFERENCES

- Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., ... Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, 26(2), 229–246.
- Alexa, A., & Rahnenfuhrer, J. (2016). *topGO: Enrichment analysis for gene ontology*. R package version 2.26.0.
- Allen, G. R. (1991). *Damselfishes of the world*. Mergus Publishers. 271 p.
- 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.
- Almany, G. R., Planes, S., Thorrold, S. R., Berumen, M. L., Bode, M., Saenz-Agudelo, P., Bonin, M. C., Frisch, A. J., Harrison, H. B., Messmer, V., Nanninga, G. B., Priest, M. A., Srinivasan, M., Sinclair-Taylor, T., Williamson, D. H., & Jones, G. P. (2017). Larval fish dispersal in a coral-reef seascape. *Nature Ecology & Evolution*, 1(6), 0148.
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Barley, A. J., Brown, J. M., & Thomson, R. C. (2018). Impact of model violations on the inference of species boundaries under the multi-species coalescent. *Systematic Biology*, 67(2), 269–284.
- 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.
- Barrera-Guzmán, A. O., Aleixo, A., Faccio, M., Dantas, S. D. M., & Weir, J. T. (2022). Gene flow, genomic homogenization and the timeline to speciation in Amazonian manakins. *Molecular Ecology*, 31(15), 4050–4066.
- Berner, D., & Salzburger, W. (2015). The genomics of organismal diversification illuminated by adaptive radiations. *Trends in Genetics*, 31(9), 491–499.
- 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.
- Branco, S., Carpentier, F., de la Vega, R. C. R., Badouin, H., Snirc, A., Le Prieur, S., Coelho, M. A., de Vienne, D. M., Hartmann, F. E., Begerow, D., Hood, M. E., & Giraud, T. (2018). Multiple convergent supergene evolution events in mating-type chromosomes. *Nature Communications*, 9(1), 1–13.
- 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.
- 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.
- Cerca, J., Cotoras, D. D., Bieker, V. C., De-Kayne, R., Vargas, P., Fernández-Mazuecos, M., López-Delgado, J., White, O., Stervander, M., Geneva, A. J., Guevara Andino, J. E., Meier, J. I., Roebler, L., Brée, B., Patiño, J., Guayasamin, J. M., Torres, M. L., Valdebenito, H., Castañeda, M. D. R., ... Martin, M. D. (2023). Evolutionary genomics of oceanic island radiations. *Trends in Ecology & Evolution*, 38, 631–642.
- 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*. Sinauer Associates. 545 p.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158.
- 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.
- Duchen, P., & Salamin, N. (2021). A cautionary note on the use of haplotype callers in phylogenomics. *Systematic Biology*, 70(4), 844–854.
- 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.
- 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.
- Fautin, D. G., & Allen, G. R. (1997). *Anemonefishes and their host sea anemones*. 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.
- 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.
- García, A., Broennimann, O., Guisan, A., Gaboriau, T., & Salamin, N. (2023). Implementation of biotic interactions in niche analyses unravels the patterns underneath community composition in clownfishes. *bioRxiv*, 2023-03.
- 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.
- Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M. H., Hansen, N. F., Durand, E. Y., Malaspina, A. S., Jensen, J. D., Marques-Bonet, T., Alkan, C., Prüfer, K., Meyer, M., Burbano, H. A., ... Pääbo, S. (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.
- 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.

- Hubert, N., Meyer, C. P., Bruggemann, H. J., Guerin, F., Komono, R. J., Espiau, B., Causse, R., Williams, J. T., & 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), e28987.
- 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.
- Joshi, N. A., & Fass, J. N. (2011). *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files* (Version 1.33) [Software]. <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., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & 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.
- 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.
- Küpper, C., Stocks, M., Risse, J. E., Dos Remedios, N., Farrell, L. L., McRae, S. B., Morgan, T. C., Karlionova, N., Pinchuk, P., Verkuil, Y. I., Kitaysky, A. S., Wingfield, J. C., Piersma, T., Zeng, K., Slate, J., Blaxter, M., Lank, D. B., & Burke, T. (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics*, 48(1), 79–83.
- 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.
- Leaché, A. D., Harris, R. B., Rannala, B., & Yang, Z. (2014). The influence of gene flow on species tree estimation: A simulation study. *Systematic Biology*, 63(1), 17–30.
- Lee-Yaw, J. A., Grassa, C. J., Joly, S., Andrew, R. L., & Rieseberg, L. H. (2019). An evaluation of alternative explanations for widespread cytonuclear discordance in annual sunflowers (*Helianthus*). *New Phytologist*, 221(1), 515–526.
- Lehmann, R., Lightfoot, D. J., Schunter, C., Michell, C. T., Ohyanagi, H., Mineta, K., Foret, S., Berumen, M. L., Miller, D. J., Aranda, M., Gojobori, T., Munday, P. L., & Ravasi, T. (2018). 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.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079.
- 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.
- 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., 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). Hybridization and diversification in the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, 14(1), 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(1), 212.
- Losos, J. B., Arnold, S. J., Bejerano, G., Brodie, E. D., III, Hibbett, D., Hoekstra, H. E., Mindell, D. P., Monteiro, A., Moritz, C., Orr, H. A., Petrov, D. A., Renner, S. S., Ricklefs, R. E., Soltis, P. S., & Turner, T. L. (2013). Evolutionary biology for the 21st century. *PLoS Biology*, 11(1), e1001466.
- Ludt, W. B., & Rocha, L. A. (2015). Shifting seas: The impacts of Pleistocene sea-level fluctuations on the evolution of tropical marine taxa. *Journal of Biogeography*, 42(1), 25–38.
- Luzuriaga-Aveiga, V. E., Ugarte, M., & Weir, J. T. (2021). Distinguishing genomic homogenization from parapatric speciation in an elevationally replacing pair of *Ramphocelus* tanagers. *Molecular Ecology*, 30(21), 5517–5529.
- Malinsky, M., Challis, R. J., Tyers, A. M., Schiffels, S., Terai, Y., Ngatunga, B. P., Miska, E. A., Durbin, R., Genner, M. J., & Turner, G. F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, 350(6267), 1493–1498.
- 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., & Salamin, N. (2023). Insights into the genomics of clownfish adaptive radiation: The genomic substrate of the diversification. *Genome Biology and Evolution*, 15(7), evad088.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, 17(1), 10.
- 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.
- Martin, S. H., & Van Belleghem, S. M. (2017). Exploring evolutionary relationships across the genome using topology weighting. *Genetics*, 206(1), 429–438.
- Mayr, E. (1963). *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.
- Meier, J. I., McGee, M. D., Marques, D. A., Mwaiko, S., Kishe, M., Wandera, S., Neumann, D., Mrosso, H., Chapman, L. J., Chapman, C. A., Kaufman, L., Taabu-Munyaho, A., Wagner, C. E., Bruggmann, R., Excoffier, L., & Seehausen, O. (2023). Cycles of fusion and fission enabled rapid parallel adaptive radiations in African cichlids. *Science*, 381(6665), eade2833.
- Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, 210(2), 719–731.
- Momigliano, P., Florin, A. B., & Merilä, J. (2021). Biases in demographic modeling affect our understanding of recent divergence. *Molecular Biology and Evolution*, 38(7), 2967–2985.
- Nadeau, N. J., Martin, S. H., Kozak, K. M., Salazar, C., Dasmahapatra, K. K., Davey, J. W., Baxter, S. W., Blaxter, M. L., Mallet, J., & Jiggins, C. D. (2013). Genome-wide patterns of divergence and gene flow across a butterfly radiation. *Molecular Ecology*, 22(3), 814–826.
- Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, 24(3), 145–156.
- Olave, M., Nater, A., Kautt, A. F., & Meyer, A. (2022). Early stages of sympatric homoploid hybrid speciation in crater lake cichlid fishes. *Nature Communications*, 13(1), 5893.
- 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.

- Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, 25(11), 2337–2360.
- Pedersen, M. (2014, October). Playing with matches: Hybrid Clownfishes – They can fuel the fires of conservation or burn everything to the ground. *Coral Magazine*, 11(5), 62–72.
- 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.
- Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, 8(11), e1002967.
- 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 of the United States of America*, 106(14), 5693–5697.
- R core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., & Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*, 559(7714), 392–395.
- Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223.
- Samuk, K., & Noor, M. A. (2022). Gene flow biases population genetic inference of recombination rate. *G3*, 12(11), jkac236.
- Schluter, D. (2009). Evidence for ecological speciation and its alternative. *Science*, 323(5915), 737–741.
- Schmid, S., Micheli, B., Cortesi, F., Donati, G., & Salamin, N. (2022). Extensive hybridisation throughout clownfishes evolutionary history. *bioRxiv*, 2022-07.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G. P., Bank, C., Brännström, A., Brelsford, A., Clarkson, C. S., Eroukhanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews Genetics*, 15(3), 176–192.
- 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.
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.
- Steinke, D., Zemlak, T. S., & Hebert, P. D. (2009). Barcoding nemo: DNA-based identifications for the ornamental fish trade. *PLoS One*, 4(7), e6300.
- Svardal, H., Quah, F. X., Malinsky, M., Ngatunga, B. P., Miska, E. A., Salzburger, W., Genner, M. J., Turner, G. F., & Durbin, R. (2020). Ancestral hybridization facilitated species diversification in the Lake Malawi cichlid fish adaptive radiation. *Molecular Biology and Evolution*, 37(4), 1100–1113.
- 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 DNA Mapping, Sequencing, and Analysis*, 27(1), 324–325.
- Taylor, S. A., & Larson, E. L. (2019). Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. *Nature Ecology & Evolution*, 3(2), 170–177.
- 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.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.
- Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, 18(2), 87–100.
- 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., Tuttle, E. M., Yi, S., & 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.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Marcionetti, A., Bertrand, J. A. M., Cortesi, F., Donati, G. F. A., Heim, S., Huyghe, F., Kochzius, M., Pellissier, L., & Salamin, N. (2024). Recurrent gene flow events occurred during the diversification of clownfishes of the skunk complex. *Molecular Ecology*, 33, e17347. <https://doi.org/10.1111/mec.17347>