Received: 7 November 2023 | Revised: 12 May 2024 | Accepted: 17 May 2024

DOI: 10.1111/mec.17436

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Gene flow throughout the evolutionary history of a colour polymorphic and generalist clownfish

Sarah Schmid[1] | Marianne Bachmann Salvy[1] | Alberto Garcia Jimenez[1] | Joris A. M. Bertrand[1] | Fabio Cortesi[2,3] | Sara Heim[1] | Filip Huyghe[4] | Glenn Litsios[1] | Anna Marcionetti[1] | James L. O'Donnell[5] | Cynthia Riginos[3] | Valerio Tettamanti[2] | Nicolas Salamin[1]

[1]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland

[2]Queensland Brain Institute, the University of Queensland, Brisbane, Queensland, Australia

[3]School of the Environment, The University of Queensland, Brisbane, Queensland, Australia

[4]Marine Biology Laboratory, Department of Ecology and Biodiversity, Vrije Universiteit Brussel, Brussel, Belgium

[5]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, California, USA

Correspondence
Sarah Schmid, Department of Computational Biology, University of Lausanne, Lausanne, Switzerland.
Email: sarah.schmid@ik.me

Funding information
Schweizerischer Nationalfonds zur Förderung der Wissenschaftlichen Forschung, Grant/Award Number: 31003A-163428; Australian Research Council, Grant/Award Number: DE200100620 and DP180102363

Handling Editor: Nick Hamilton Barton

Abstract

Even seemingly homogeneous on the surface, the oceans display high environmental heterogeneity across space and time. Indeed, different soft barriers structure the marine environment, which offers an appealing opportunity to study various evolutionary processes such as population differentiation and speciation. Here, we focus on *Amphiprion clarkii* (Actinopterygii; Perciformes), the most widespread of clownfishes that exhibits the highest colour polymorphism. Clownfishes can only disperse during a short pelagic larval phase before their sedentary adult lifestyle, which might limit connectivity among populations, thus facilitating speciation events. Consequently, the taxonomic status of *A. clarkii* has been under debate. We used whole-genome resequencing data of 67 *A. clarkii* specimens spread across the Indian and Pacific Oceans to characterize the species' population structure, demographic history and colour polymorphism. We found that *A. clarkii* spread from the Indo-Pacific Ocean to the Pacific and Indian Oceans following a stepping-stone dispersal and that gene flow was pervasive throughout its demographic history. Interestingly, colour patterns differed noticeably among the Indonesian populations and the two populations at the extreme of the sampling distribution (i.e. Maldives and New Caledonia), which exhibited more comparable colour patterns despite their geographic and genetic distances. Our study emphasizes how whole-genome studies can uncover the intricate evolutionary past of wide-ranging species with diverse phenotypes, shedding light on the complex nature of the species concept paradigm.

KEYWORDS
fish, genomics/proteomics, population genetics-empirical, phylogeography

1 | INTRODUCTION

Marine ecosystems are usually thought to be highly connected (Conover et al., 2006; Shanks et al., 2003) due to the apparent lack of barriers to gene flow and the large population sizes of many marine species (Palumbi, 1992; Waples, 1998). Despite this seeming absence of boundaries, oceans and seas constitute structured ecosystems. Past geographical events, water salinity and temperature, as well as oceanic currents, are some of the multiple environmental factors that might form partial or "soft" barriers in the marine environment (Belanger et al., 2012; Choo et al., 2021; Marko, 2004; Meyer et al., 2005; Oleksiak, 2019; White et al., 2010). Studying population connectivity in this environment is appealing since it promises to provide critical insights into the impact of soft barriers on population structure and speciation.

The Indo-Pacific region is widely recognized as the centre of marine biodiversity, with the Coral Triangle being a significant biodiversity hotspot (Hughes et al., 2002; Lohman et al., 2011; Renema et al., 2008) not only for fishes but also for corals and other invertebrate species (Allen, 2008; Hoeksema, 2007). However, the factors that have led to this consistent biogeographic pattern still need to be fully understood (Pellissier et al., 2014). One approach to understanding the history of this region is to study intraspecific variation to make inferences about the factors that may have shaped the distribution of marine species. This can shed light on questions such as connectivity within a species, their extent of dispersal, their geographical origins and whether edge exhibit lower genetic diversity (Arroyo-Correa et al., 2023; Buckley et al., 2022). Nevertheless, obtaining sufficient population data in marine environments to infer the processes that lead to population divergence and structure remains challenging (Keyse et al., 2014). Consequently, the number of marine studies is limited compared to their terrestrial counterparts. Early investigations into the population structure of marine animals revealed that they tend to show limited genetic structure (Conover et al., 2006; Waples, 1998). However, most of these studies were based on a few genetic markers. More recently, the transition to whole-genome analyses enabled the detection of previously unidentified population differentiation (Allendorf et al., 2010; Ellegren, 2014). It also allowed researchers to go beyond the description of these patterns to test hypotheses about the processes leading to population structure (Hellberg, 2009). Such extensive genomic coverage further enhanced the estimation of population-genetic parameters, such as migration rates or effective population sizes. It enabled the formulation of neutral expectations with more confidence and improved inferences about population demography (Allendorf et al., 2010), considerably changing the perspective on marine ecosystems. Hence, in recent years there has been a switch from a vision of well-mixed marine populations to a view of populations with restricted gene flow displaying considerable levels of differentiation and intraspecific variability (Barabás & D'Andrea, 2016; Messer et al., 2016; Rudman et al., 2015; Van der Ven et al., 2021; Wood & Brodie, 2016). This change of paradigm challenges past assumptions about marine species dispersal and calls for an improved understanding of the factors influencing gene flow and their resulting evolutionary implications.

Clownfishes—also known as anemonefishes (Actinopterygii; Perciformes; Amphiprioninae)—are a fascinating group of reef fishes suitable for population genomic studies due to their unique life-history traits. In their natural habitat, clownfishes have formed mutualistic relationships with 10 different sea anemone species. Specialist species interact with only a subset of anemones, while generalist species can interact with the full range of hosts (Litsios et al., 2012). Moreover, clownfishes dispersal is restricted by their short pelagic larval duration (PLD; the period of development spent in the water column as plankton), which lasts from 10 to 15 days (Fautin & Allen, 1997; Roux et al., 2019, 2020), and exhibit high levels of self-recruitment (i.e. the proportion of larvae returning to and settling in their natal population; Beldade et al., 2012; Buston et al., 2012; Jones et al., 2005; Madduppa et al., 2014). This pelagic phase results in a general dispersal range of around five to one hundred kilometres (Salinas-de-León et al., 2012), which is much shorter compared to most other damselfish species (Thresher et al., 1989). It needs to be noted that long-distance dispersal has been recorded in some clownfish species, presumably due to strong oceanic currents (e.g. over 400 km in *Amphiprion omanensis*; Simpson et al., 2014). Besides a low dispersal ability and strong habitat specificity, clownfishes are protandrous—meaning that male reproductive organs mature before the female ones—and they form monogamous pairs with a social structure that is controlled by a strict hierarchy. These strong constraints put into question the demographic history and the microevolutionary processes involved in the range expansions observed in clownfishes (Huyghe & Kochzius, 2017; Litsios, Kostikova, & Salamin, 2014; Litsios, Pearman, et al., 2014), with some species displaying an unexpectedly broad geographic range.

Amphiprion clarkii (Bennett, 1830)—a highly colour-polymorphic and generalist clownfish species—exhibits the broadest geographic range among all clownfishes (Ollerton et al., 2007). Its range extends from the Micronesian islands of the Pacific Ocean to the Persian Gulf and stretches longitudinally from the Japanese coast to the Great Barrier Reef (Fautin & Allen, 1997). Nevertheless, the geographical origin of the species, the potential impact of sea surface currents and geographical barriers on the connectivity between populations, and the evolutionary processes behind its wide distribution remain unclear (Litsios, Kostikova, & Salamin, 2014; Litsios, Pearman, et al., 2014). Moreover, *A. clarkii* is considered an "extreme generalist" because of its ability to interact with all 10 sea anemone hosts, even if host preference can vary locally (Ollerton et al., 2007). This characteristic could either facilitate connectivity among populations due to the increased host anemone availability or—on the opposite—might reduce gene flow among neighbouring populations due to habitat-specific host selection. Finally, this species displays a high level of colour polymorphism, which has led to considerable confusion regarding their identification. Indeed, the background colour can range from light orange to entirely dark, and the number of emblematic white bars varies among individuals (Bell et al., 1982; Fautin & Allen, 1997). All

these characteristics started a several decade-long debate about the potential for *A. clarkii* to consist of a complex of multiple species. Phylogenetic reconstructions based on a few genetic markers showed contradictory results at the molecular level. *A. clarkii* specimens from the Indo-Pacific Ocean clustered by colouration rather than geographic origin (Litsios, Kostikova, & Salamin, 2014; Litsios, Pearman, et al., 2014; Rolland et al., 2018) while colour divergent *A. clarkii* morphotypes from the Indian Ocean did not differ significantly at the molecular level (Devi et al., 2021). Moreover, the evolutionary processes involved in the remarkable colour variation, geographical range and mutualistic interactions displayed by *A. clarkii* remain to be deciphered to provide certainty about the taxonomic status of the species.

In this study, we analysed the most extensive sampling of *A. clarkii* populations to date using whole-genome resequencing approaches to assess the taxonomic status of *A. clarkii* and thus evaluate the impact of soft barriers on the species. We inferred the demographic history and population structure and computed genome-wide phenotype associations to test for demographic expansions, restricted gene flow and the genetic association of colour polymorphism in *A. clarkii* populations. The broad geographical range of *A. clarkii* raises questions about the geographic origin of the species and the dynamic of the range expansion that followed. We tested two different hypotheses that propose either an origin in the Indo-Pacific Ocean—following the 'Centre of Origin' hypothesis (Briggs, 2003; Gaboriau et al., 2017)—or in the Papua New Guinea–Solomon Islands area, which could lead to a subsequent westward expansion through a stepping-stone process. We further hypothesised that the unique characteristics of *A. clarkii* compared to other clownfishes may indicate the presence of cryptic genetic differentiation, questioning its status as a single species. We thus estimated gene flow among populations to determine whether gene flow occurred throughout the evolutionary history of *A. clarkii*, was restricted to a brief period following population splits, occurred after secondary contact, or occurred recently. Finally, observed colour polymorphism could be the cause or consequence of reproductive isolation, leading to an association between potential genetic differences and colour morphs. We conducted genome-wide analyses to test if the genetic structure was associated with specific colour morphs and if genomic differentiation involved loci known to influence colour variation in fishes. Our study demonstrates the potential of whole-genome data to improve our understanding of the evolution of geographically wide-ranging and phenotypically diverse species.

2 | MATERIALS AND METHODS

2.1 | Sampling, library preparation and sequencing

Tissue samples consisting of pieces of dorsal fins of about 1 cm long (fin clips) from 67 *Amphiprion clarkii* specimens were collected by various collaborators between 2013 and 2018 from eight different populations across the Indo-Pacific (Schmid et al., 2023; Figure 1A; Table S1; Appendix S1 for sampling permits) and were stored in ethanol at −20°C. Pictures for morphometry and colour pattern analysis were taken for individuals coming from New Caledonia, Maldives and Indonesia. We extracted the genomic DNA of each fin clip following the DNeasy Blood and Tissue kit standard procedure and performed the final elution twice in 100 μL of AE buffer (QIAGEN, Hombrechtikon, Switzerland). We quantified the extracted DNA using Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA) and evaluated the integrity by electrophoresis. We followed the TruSeq Nano DNA library prep standard protocol to generate libraries with a 350 base pair insert size for whole-genome paired-end sequencing (Illumina, San Diego, USA). We validated the fragment length distribution of the libraries with a Bioanalyzer (Agilent Technologies, Santa Clara, USA). Sequencing was performed by the Genomic Technologies Facility at the University of Lausanne, Switzerland. Libraries generated in 2017 and 2018 were sequenced on five Illumina HiSeq 2500, 100 paired-end lanes, while libraries prepared in 2019 were sequenced on nine Illumina 4000 HiSeq, 150 paired-end lanes. We aimed for each sample to reach an approximate 10× coverage with both types of libraries.

2.2 | Sequenced data processing, mapping and genotyping

We trimmed the generated reads to remove adapter sequences using Cutadapt V2.3 (Martin, 2011). We removed reads shorter than 40 bp and with a Phred quality score below 40 with Sickle V1.33 (Joshi & Fass, 2011). We assessed read quality before and after processing with FastQC V0.11.7 (Andrews, 2010). We mapped the reads that were kept against the reference genome of *A. percula* (Lehmann et al., 2018) using BWA-MEM V0.7.17 (Li & Durbin, 2009) and subsequently sorted, indexed and filtered them according to mapping quality (>30) using SAMtools V1.8 (Li et al., 2009). Then, we assigned all the reads to read-groups using Picard Tools V2.20.7 (http://broadinstitute.github.io/picard/) and merged overlapping reads using ATLAS V0.9 (Link et al., 2017). We validated the mapping output using various statistics generated with BamTools V2.4.1 (Barnett et al., 2011). After mapping, we computed genotype likelihoods and estimated the major and minor alleles using the maximum likelihood estimation method with ATLAS V0.9 (Link et al., 2017). We filtered the resulting VCF file including 67 *A. clarkii* individuals with VCFtools V0.1.15 (Danecek et al., 2011) to keep only sites informative for all samples, with a minimum depth of 2 and a minimum Phred quality of 40.

2.3 | Population structure and admixture

To explore population structure and admixture in *A. clarkii*, we performed a principal component analysis (PCA) based on the covariance between individuals and calculated admixture proportions using PCAngsd V1.21 (Meisner & Albrechtsen, 2018). This method

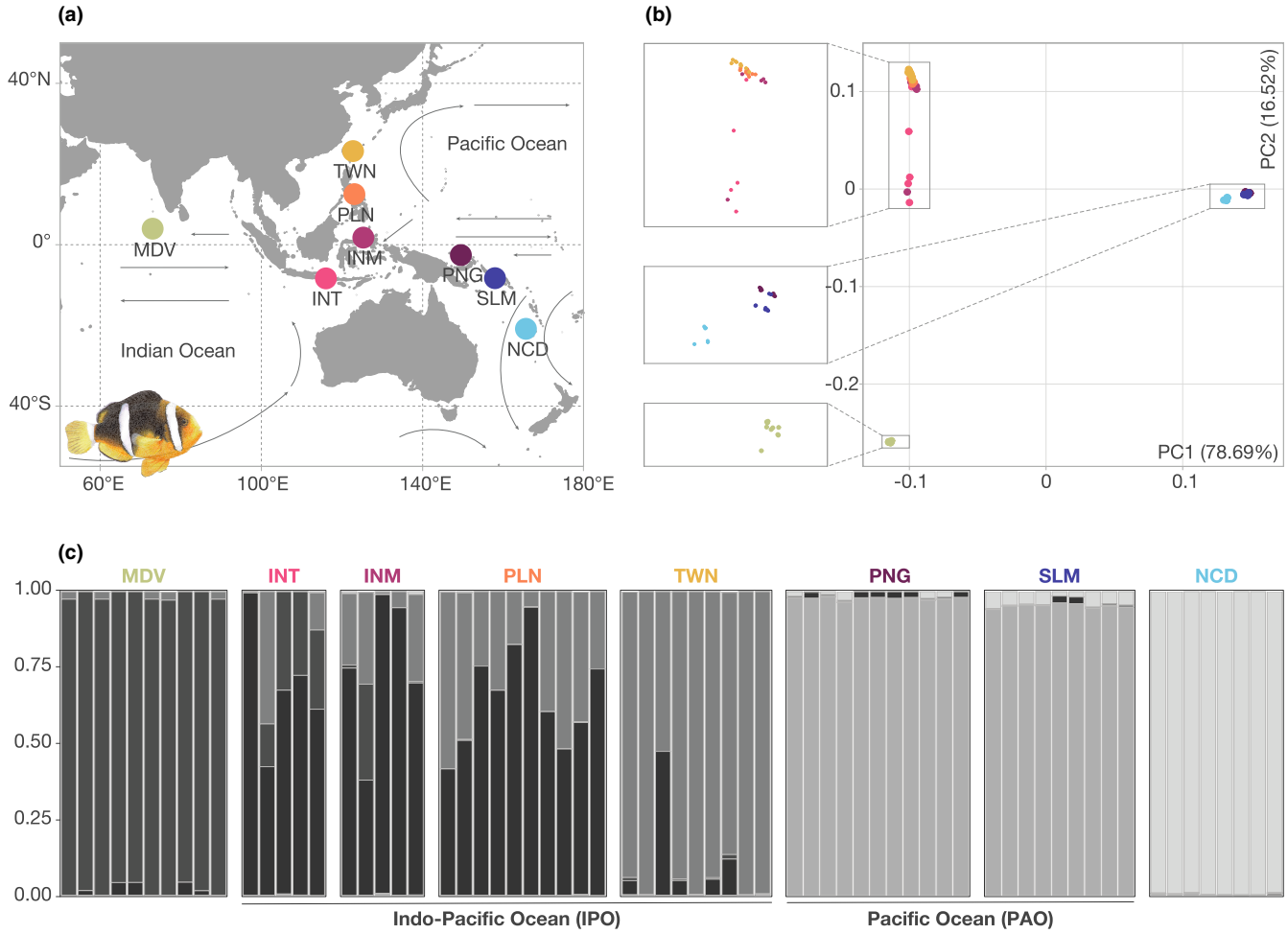

FIGURE 1 Sampling and population genetic structure of *Amphiprion clarkii*. (a) Coloured circles represent sampling sites for the study: MDV (Maldives), INT (Indonesia, Tulamben), INM (Indonesia, Manado), PLN (Philippines), TWN (Taiwan), PNG (Papua New Guinea), SLM (Solomon Islands) and NCD (New Caledonia). Arrows represent major ocean surface currents. (b) Principal component analysis showing a separation of the eastern Pacific populations (PNG, SLM, NCD) from the Indo-Pacific populations (MDV, INM, INT, PLN, TWN) on the first principal component axis (PC1), and further differentiation of the Maldives population (MDV) on the second axis (PC2). (c) Admixture bar plot based on PCAngsd *q*-values for $K = 5$. Shades of grey represent the five clusters membership for each individual.

works directly with genotype likelihoods and connects the PCA results with the admixture proportions by detecting the most likely number of ancestral populations *K* based on the PCA eigenvectors.

2.4 | Population genomics analysis in sliding windows and identification of outlier windows

To investigate the variation of genome-wide patterns of differentiation, divergence and diversity among *A. clarkii* populations, we estimated between-population differentiation (F_{ST}), between-population absolute divergence (d_{xy}) as well as population nucleotide diversity (π) in non-overlapping 5000 SNPs windows using popgenWindows.py (https://github.com/simonhmartin/genomics_general). To identify genomic regions with outlier values in all pairwise comparisons (F_{ST} and d_{xy}) or populations (π), we applied a *Z*-transformation to the window-based F_{ST}, d_{xy} and π estimates to normalize the values across all populations and pairwise-population comparisons. We considered as outliers those windows with mean normalized values outside the 1st and 99th percentiles.

2.5 | Demographic modelling

We modelled the demographic history of *A. clarkii* using *fastsimcoal2* V2.6.0.3, a coalescent approach to infer demographic parameters based on site frequency spectrum (SFS; Excoffier et al., 2013). To simplify the models and the computational power required, we used the four clusters identified with PCAngsd as populations. We grouped Indonesia, Philippines and Taiwan into a single "Indo-Pacific Ocean" (IPO) population and Papua New Guinea and Solomon Islands into a "Pacific Ocean" (PAO) population. We considered Maldives and New Caledonia as two separate populations. We estimated the folded 2D SFS of the minor allele using whole-genome SNPs for each population using the vcf2sfs.py script (https://github.com/marqueda/SFS-scripts). We used the resulting 2D SFS to compare a total of

eight distinct demographic scenarios (Figure 2). To first determine the origins of the four *A. clarkii* clusters (Maldives, IPO, PAO and New Caledonia), we built four scenarios based on previous biogeographical reconstruction (Litsios, Kostikova, & Salamin, 2014; Litsios, Pearman, et al., 2014). We modelled two scenarios with the PAO and IPO as the population of origin following a stepping stone process (models SS1 and SS2; Figure 2A,B), a scenario with the IPO population as the centre of origin (model CO; Figure 2C) and a scenario with a single ancestral population giving rise to both IPO and PAO populations, from which Maldives and New Caledonia respectively emerged (model SA; Figure 2D). We then used the best scenario explaining the origin of *A. clarkii* populations as the basis to build the four additional scenarios characterized by different timing of gene flow between populations: constant gene flow (Figure 2E), present-day exchanges between neighbouring populations (Figure 2F), gene flow after secondary contact (Figure 2G) and gene flow at the onset of the population splits (Figure 2H).

To estimate the demographic parameters (population split time, effective population size and migration rate), we ran *fastimcoal2* for each model using the observed multidimensional SFS. We specified an approximate mutation rate of 4E-8 (Delrieu-Trottin et al., 2017; Rolland et al., 2018), a 5-year generation time (Buston & García, 2007) and ignored monomorphic sites. We performed 200,000 coalescent simulations (*-n*) to approximate the expected SFS, ran 50 cycles of the expectation-conditional maximization (ECM; *-L*) to estimate the parameters and discarded entries with fewer than 10 SNPs per SFS count (-C). We ran each model 50 times to obtain the best combination of parameters estimates for our data. We selected the run with the highest likelihood and proceeded to 100 additional runs to obtain the likelihood distribution. We visualized the fit of the simulated to the observed SFS using the R script SFStools.r (https://github.com/marqueda/SFS-scripts/).

To select the best-fitting demographic model, we used the Akaike's information criterion (AIC), which considers the number of parameters when comparing the likelihood of the different models. Since the AIC can lead to an overestimation of the most likely model if SNPs are not independent (not LD-pruned), we also compared the likelihood distribution of each model. Two models with an overlap of 50% or more in their likelihood range were considered as not differing significantly from each other.

Finally, for the best-fitting demographic model, we calculated the confidence interval of the parameters by running 100 replicates of parametric bootstrap. For each replicate, we simulated SFS from the maximum composite likelihood estimates and re-estimated parameters based on 30 independent runs with 200,000 simulations and 20 ECM cycles (Excoffier et al., 2013; Ortego et al., 2018). We calculated the 95 percentile confidence intervals of every parameter based on the highest likelihood run of each bootstrap replicate.

2.6 | Gene ontology

We investigated whether specific gene ontology (GO) terms were enriched in the genomic windows displaying outlier values of F_{ST}, d_{xy}

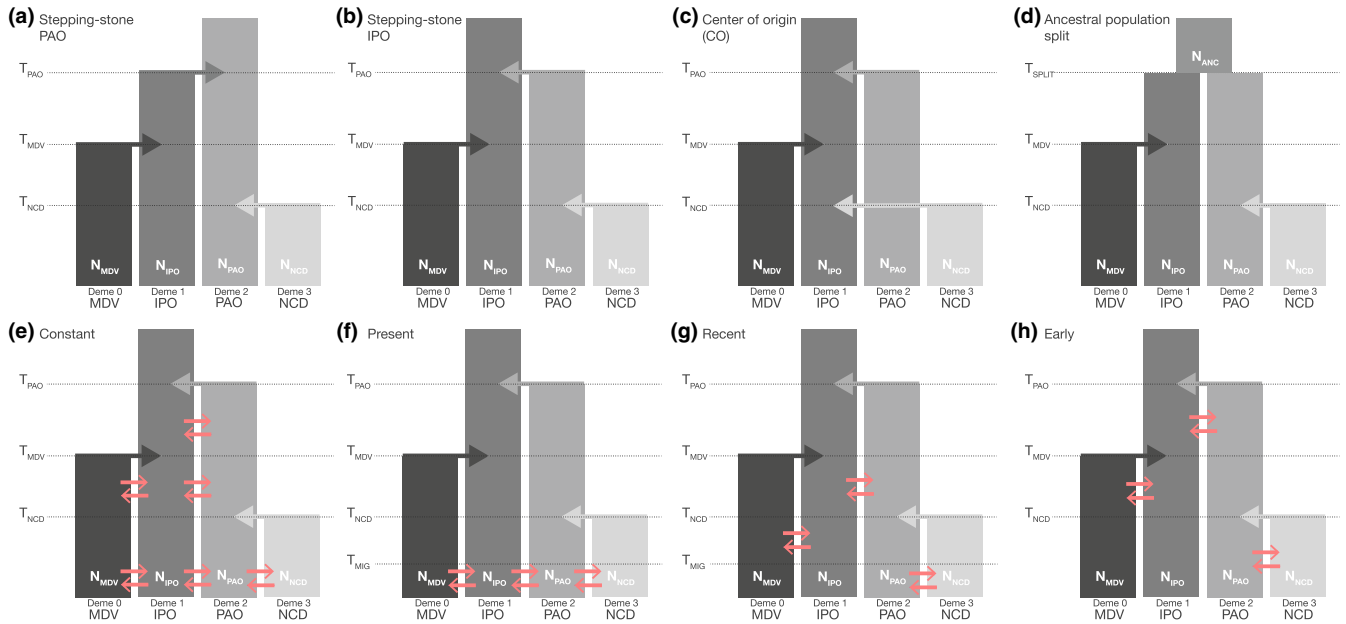

FIGURE 2 Demographic scenarios modelled with fastsimcoal26. Top scenarios depict the origins of *Amphiprion clarkii* populations. Stepping-stone models with either (a) the Pacific Ocean population as population of origin or (b) the Indo-Pacific Ocean population as origin; (c) Indo-Pacific Ocean population as centre of origin to all populations and (d) split from an ancestral population. Bottom models are characterized by different timing of gene flow. (e) Constant gene flow; (f) present gene flow; (g) gene flow after secondary-contact (recent); (h) gene flow at the onset of the population split (early). Colonization arrows (shades of grey) are depicted in coalescent time. The orange arrows represent migration (gene flow) between populations. IPO, Indo-Pacific Ocean (includes Indonesia, Philippines and Taiwan); MDV, Maldives; NC, New Caledonia; PAO, Pacific Ocean (includes Papua New Guinea and Solomon Islands).

and π, respectively. We used the 17,179 annotated genes inferred by Marcionetti et al. (2019), among which 14,002 were annotated with *biological process* GO terms. We retrieved gene annotations for all the genes within (or overlapping) the outlier genomic windows and performed GO enrichment analysis by comparing those annotations to the complete set of protein-coding annotated genes of *A. percula*. We used the R package topGO V2.44 (Alexa & Rahnenfuhrer, 2020) to perform the gene ontology enrichment analysis based on Fisher's exact test with the *weight01* algorithm. We defined a minimum node size of 10 and considered those GO terms with a raw *p*-value below .01 as significant, following the recommendations from the topGO manual.

2.7 | Image processing, colour proportions and redundancy analyses (RDA)

For individuals with available pictures (28 individuals from Indonesia, Maldives and New Caledonia; see Table S1), we standardized the picture size (3000 by 4000 pixels) and resolution and corrected the white balance using Adobe Photoshop CC 2019. In addition, we set 28 morphological landmarks on each image using FIJI software (Schindelin et al., 2012). We used the landmarks to perform a Procrustes analysis using the R package MORPHO (Schlager, 2017) and aligned the images to a mean shape, ensuring that each pixel represented the same body part in all images. Then, we removed the background masking the fish outline considering a spline regression along the landmarks outer body landmarks. We used those processed pictures in RGB format in all subsequent image analyses.

2.7.1 | Colour proportions and colour patterns

We summarized the variation in colour proportion by estimating the proportion of black, white and orange across the body of each individual following Endler (2012). This method categorized the pixels of each image into predefined colour groups (black: R = 0, G = 0, B = 0; white: R = 255, G = 255, B = 255; and orange: R = 255, G = 165, B = 0) based on the minimum Euclidean distance from each pixel's RGB code to the three categories. The colour proportion is subsequently calculated by dividing the number of pixels in each colour category by the total number of pixels across all three categories (black, orange and white). We implemented this approach in a custom R script (see Data Accessibility). In addition, we implemented an approach similar to the R package patternize (Van Belleghem et al., 2018) to retrieve a more detailed representation of the colouration pattern variations in *A. clarkii*. Like patternize, our approach establishes homology between pattern positions across individuals through fixed landmarks. In summary, we vectorised our images and concatenated the RGB channels to create a 36,000,000 × 28 matrix (i.e. 3000 × 4000 pixels, times the three channels and 28 individuals). Then, we summarized the extracted colour patterns for each individual using a principal component analysis (PCA) with the *prcomp* function from the stats R package (version 3.6.2). To qualitatively assess the variation in colour patterns associated with each principal component (PC) axis, we reconstructed the minimum and maximum values of each PC axis while maintaining the remaining PC values at the centre of coordinates (0,0). This reconstruction was accomplished using the PC eigenvectors, which allowed us to reconstruct the three-layer structure of the image. Moreover, we scaled the output to values ranging from −1 to 1 and averaged the resulting three-layer raster stack to obtain the average importance of each pixel relative to the RGB image. This method was implemented in a custom R script (see Data Accessibility). Finally, we conducted hierarchical clustering analysis on the first six colour PCA axes utilizing the *hclust* function of the stats R package (V3.6.2) to summarize the information and group individual according to their colour pattern.

2.7.2 | Redundancy analysis (RDA)

We used a partial redundancy analysis (RDA; Legendre & Legendre, 1998) to identify potential SNPs that could be associated with differences in colour pattern among populations based on 28 individuals that had both SNP and colour data available. Following Capblancq & Forester, 2021 methodology (https://github.com/Capblancq/RDA-landscape-genomics), we used the first six colour PCA axes (hereafter referred as ColPCs) as the RDA explanatory variables, while constraining the axes with the population ID, the geographic coordinates of each population and the population genomic structure, which was quantified based on the loadings of the two first axes of a PCA on the genome-wide covariance among individuals. We removed monomorphic SNPs due to the populations subsampling and as well as SNPs that were less than 1 kb distant from one another. We performed RDAs on each chromosome separately and extracted the SNPs significantly correlated with at least one of the ColPCs. Using a chromosome-based approach allows us to be more conservative in the detection of outlier SNPs. We estimated the Malahanobis distance of each SNP in the multidimensional RDA space and identified significant SNPs using the probability that a given SNP is an outlier compared to the expected normal distribution of SNPs in the RDA space using a chi-squared test with an FDR correction (Benjamini & Hochberg, 1995). We randomly subsampled the same number of outlier SNPs across the genome and tested whether the geographic genomic structure was recovered using a limited number of SNPs. Finally, we looked for SNPs detected with the RDA analysis in the proximity (±10 kb) of pigmentation genes (Lorin et al., 2018) and summarized the SNPs information in a PCA, to compare the structure of the colour-related SNPs to the global population genomic structure.

3 | RESULTS

3.1 | Sequencing, mapping and SNPs calling statistics

We obtained an average of approximately 108 million raw paired reads across samples, with the number of raw reads ranging from ca. 57 million to ca. 127 million (Table S2). Batch effect due to the different sequencing platform used was negligible (Figure S1A,B). After trimming low-quality regions and removing low-quality reads, each sample contained between 54 to 119 million paired reads, which corresponds to an estimated raw coverage of 6.4× to 20.2× (Table S2).

We mapped the reads to the reference genome of *A. percula* (Lehmann et al., 2018) with between 83% and 89% of reads mapping accurately (i.e. pairs having the correct orientation and insert-size; Table S3). We further filtered the data by removing reads mapped with low-confidence and potentially redundant sequencing data arising from the overlap of paired reads to obtain a final coverage ranging between 4.5× and 14.4× (Table S3, Figure S1C). Following these steps, we computed genotype likelihoods, estimated the major and minor alleles, and filtered the resulting VCF file to obtain a final number of 21,932,247 SNPs.

3.2 | Population structure, differentiation and admixture

We used a PCA on the covariance matrix between individuals to get an insight into the population structure of *A. clarkii*. We found a clear separation among the Pacific Ocean populations (Papua New Guinea, Solomon Islands and New Caledonia), the Indo-Pacific populations (Taiwan, Philippines, Indonesia) and Indian Ocean (Maldives) populations along the first axis, which explained 78.69% of the variance. Along the second axis, which explained 16.52% of the variance, we observed a split between the Indo-Pacific populations and the Indian Ocean population (Figure 1B), which suggests that the genomic structure coincides with the spatial distribution of *A. clarkii*.

Then, to investigate potential ancestral introgression among populations, we calculated the admixture proportion of each individual. The optimal number of clusters inferred by PCAngsd was $K=5$, respectively splitting populations from (1) the Maldives, (2) Indonesia (Tulamben and Manado) and Philippines, (3) Taiwan, (4) Papua New Guinea and the Solomon Islands, and (5) New Caledonia (Figure 1C; Figure S2 for other K values). We observed substantially higher levels of admixture in the four Indo-Pacific populations compared to the Indian and Pacific Oceans populations, thus suggesting heterogeneous gene flow between *A. clarkii* populations across its geographic range.

3.2.1 | Genome-wide patterns of differentiation, divergence and diversity

We investigated whole-genome patterns of differentiation, divergence, and diversity across the eight *A. clarkii* populations by calculating F_{ST} and d_{xy} between each pair of population as well as within-population nucleotide diversity (π) in 5000 SNPs genomic windows (Tables S4–S6). We reported here the mean of these statistics across the genomic windows. We observed high disparities in mean F_{ST} for every 28 pairwise comparisons, with estimates ranging from 0.003 (Indonesia-Manado vs. Philippines) to 0.446 (Maldives vs. Papua New Guinea; Figure 3A,B; Figure S3). We found similar patterns for d_{xy}, with mean values ranging from 0.067 (Papua New Guinea vs. Solomon Islands) to 0.183 (Maldives vs Solomon Islands; Figure 3A–C; Figure S3), while mean within-population diversity (π) ranged from 0.065 (Papua New Guinea) to 0.134 (Indonesia-Tulamben; Figure 3A–D; Figure S3).

3.2.2 | Gene ontology of differentiation, divergence and diversity outlier genomic windows

To identify genomic regions of potential importance in the evolutionary history of *A. clarkii*, we ran a gene ontology (GO) analysis on genomic windows with mean *Z*-transformed F_{ST}, d_{xy} and π found outside the 1st and 99th percentiles of values. We detected seven outlier windows for F_{ST}, 29 for d_{xy} and 33 for π. Outlier windows were common to at least one-third of the pairwise comparisons or populations (Figure 3D; Table S7). We identified 92 genes in the F_{ST} outlier windows (74 were functionally annotated), 272 in d_{xy} (259 functionally annotated) and 488 in π (413 functionally annotated). We highlighted three, six and eight significantly enriched terms (p-value < .01; Table S8) in the F_{ST}, d_{xy} and π outlier regions, respectively. The significant GO terms were associated with defence response to bacterium (GO:0042742), linoleic acid metabolism (GO:0043651) and epoxygenase P450 pathway (GO:0019373) for both d_{xy} and π outlier regions, while they were associated with proteolysis during cellular protein catabolic process (GO:0051603) and urate metabolic process (GO:0046415) for F_{ST} outlier regions.

3.3 | Demographic history analysis

To determine the origin of *A. clarkii* populations, we modelled four demographic scenarios characterized by different populations of origin and colonization pathways (Figure 3A–D) using the coalescent modelling approach implemented in *fastsimcoal2*. The model with the highest likelihood was the stepping-stone model with an Indo-Pacific origin for the species (Figure 2B). This meant that the colonization of both Maldives and Pacific Ocean populations came from the Indo-Pacific Ocean population followed by a more recent colonization of New Caledonia from the Pacific Ocean population (see Table S9 for likelihood values and Figure S4 for SFS).

Based on this model, we built four additional scenarios defined by different timing of gene flow (Figure 3E–H) to test for differential levels of migration between neighbouring populations. We obtained the highest likelihood for the model of constant gene flow (Figure 2E), which consisted of heterogeneous migration levels among neighbouring populations but constantly occurring through time (Figure 4). The

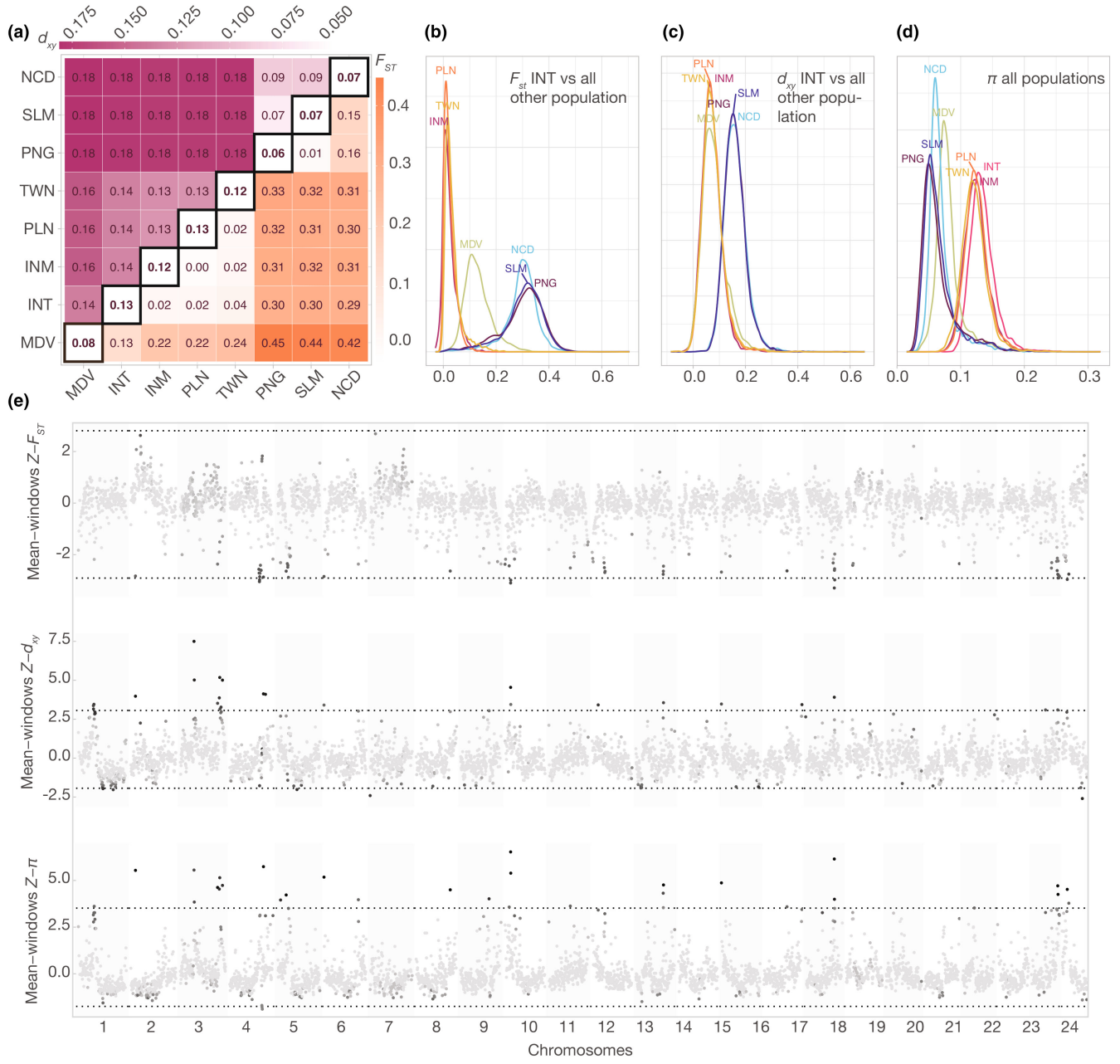

FIGURE 3 Genome-wide patterns of differentiation, divergence and diversity. (a) Mean population differentiation (F_{ST}; lower triangle), population divergence (d_{xy}; higher triangle) and diversity (π; diagonal) within and among *Amphiprion clarkii* populations calculated in 5000 SNPs genomic windows. (b) Levels of differentiation (F_{ST}), (c) divergence (d_{xy}) and (d) diversity (π) within and among *A. clarkii* populations calculated in 5000 SNPs genomic windows. For clarity, we illustrated only pairwise comparisons with the Indonesian (INT) population. See Figure S3 for pairwise comparisons of F_{ST} and d_{xy} across all populations. Each colour represents a different population for the pairwise comparison: MDV (Maldives), INT (Indonesia, Tulamben), INM (Indonesia, Manado), PLN (Philippines), TWN (Taiwan), PNG (Papua New Guinea), SLM (Solomon Islands) and NCD (New Caledonia). (e) Mean *Z*-transformed estimates across pairwise comparisons (F_{ST} and d_{xy}) and populations (π) for each genomic window. Points are coloured according to the number of pairwise comparisons or populations displaying an outlier value (not included within the 1st and 99th percentiles, which are illustrated by the dotted lines). Black points mean that all 28 pairwise comparisons or all eight populations show outlier values. Light grey points stand for no outlier value.

three other scenarios (Figure 2F–H) had lower likelihood distributions that were not overlapping with the likelihood distribution of the best model (Figure 4B; Table S9, Figure S5) and were therefore rejected.

We estimated the demographic parameters for the best model and highlighted that an initial split between Indo-Indian and Pacific populations (T_{PAO}) occurred 60,060 generations ago. Subsequent splits of the Maldives (T_{MDV}) and New Caledonia (T_{NC}) populations occurred, respectively, 11,591 and 2537 generations ago (Table 1). If we consider a generation time of five-years for clownfishes (Buston & García, 2007), our estimates corresponded to population divergence times of ca. 300 KYA (T_{PAO}), 58 KYA (T_{MDV}) and 12 KYA (T_{NC}), respectively. We obtained the highest estimated effective population

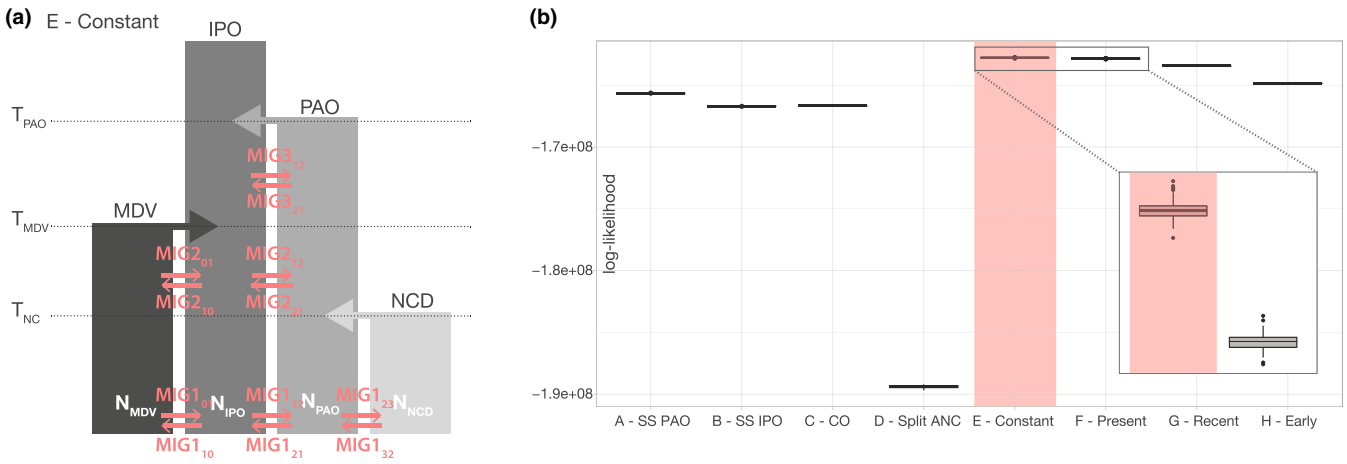

FIGURE 4 Demographic modelling results. (a) Scheme of the highest likelihood demographic scenario (E—Constant). Orange arrows represent gene flow between neighbouring populations. The parameter estimates for the best model are reported in Table 1. (b) Log-likelihood distributions for the eight tested demographic scenarios (see Figure 2). The best model is highlighted in orange, and a close-up of the two models with the highest likelihood distribution is displayed (E—constant, F—present). Likelihood distributions were obtained based on 100 expected SFS approximated using the parameters that maximize the likelihood for each model. Models with an overlap in their likelihood range indicate no significant difference between them. ANC, ancestral; CO, centre of origin; IPO, Indo-Pacific Ocean (includes Indonesia, Philippines and Taiwan); MDV, Maldives; NC, New Caledonia; PAO, Pacific Ocean (includes Papua New Guinea and Solomon Islands); SS, stepping-stone.

size for the Indo-Pacific Ocean population (N_{IPO}), followed by the Maldives population (N_{IPO}) and both Pacific Ocean (N_{PAO}) and New Caledonia populations (N_{NC}), the latter two having comparable sizes (Table 1). The estimated migration rates for the best model ranged between 9.23E-08 and 2.37E-04 (Table 1), with the highest migration rate corresponding to current gene flow between the New Caledonia and the Pacific Ocean populations ($MIG2_{32}$). We obtained the lowest migration rate for the current gene flow between Indo-Pacific and Pacific Ocean populations ($MIG2_{12}$; Figure 4A), although the confidence interval suggested a high level of uncertainty in the estimation of this parameter (Table 1).

3.4 | Colouration in *Amphiprion clarkii*

3.4.1 | Colour proportions and patterns

We first outlined the coarse colour phenotype of each *A. clarkii* individuals by calculating black, orange, and white proportions. We found a mean proportion for the black colour of 0.5, with values ranging from 0.19 (sample NC028) to 0.74 (sample GB030) between individuals. The orange proportion ranged from 0.04 (sample GB030) to 0.64 (sample NC028) with a mean of 0.31, and the white proportion displayed a mean of 0.19 with a minimum of 0.08 (sample MV186) and a maximum of 0.26 (MV033; Figure 5A,B, Table S11). The measurements confirmed the high variability in colouration patterns present in *A. clarkii* populations.

We further characterized the colour patterns of each individual based on the colour of each pixel and then presented the information in a dendrogram (Figure 5C) using the first six colour PCA axes, which accounted for over 60% of the variance across images (Figure S6). The first axis (colPC1) explained 20.7% of the variance and differentiated yellow versus black colour variation of the ventral region, pectoral, pelvic, anal and caudal fins as well as the extent and width of the white bands, while the second axis (colPC2) explained 16% of the variance and differentiated the orange versus black colouration of the ventral region, the intensity of yellow on the caudal fin and the shape of the middle white band (Figure S6A). The next four axis respectively explained 9.3%, 6%, 5.6% and 4.6% of the variance in colouration and differentiated the intensity of white on the middle band and the orange proportion of the ventral region (colPC3), the shape variation of the middle band (colPC4; Figure S6B), the colour variation on the dorsal and anal fins as well as the presence/absence of a peduncle white band (colPC5) and the dorso-ventral variation of the middle white band (colPC6; Figure S6C). The dendrogram summarizing this information highlight three main clusters: the first containing only one individual from New Caledonia, the second including all individuals from Indonesia, and the last encompassing the remaining individuals from New Caledonia and all those from Maldives (Figure 5C).

3.4.2 | RDA analysis

To identify genomic regions potentially involved in the difference in colour patterns among individuals and populations, we performed a partial RDA and removed the effect of neutral genomic structure and geographic distances. The final genomic dataset used for this analysis consisted in 650,670 SNPs. We detected 224 SNPs that had significantly larger Mahalanobis distances in the RDA space and could potentially be related to the colour pattern variation (Table S12; hereafter referred to as "significant SNPs"). The average number of

significant SNPs per chromosome was nine (corresponding to an average proportion of 9.7E-4), with some chromosomes displaying high density of significant SNPs compared to others (i.e. chromosomes 1, 2, 3 and 18; Figure S7). None of them were found in a genomic region related to pigmentation in other teleost fishes (Figure S8). As expected, the PCA performed on the 224 significant RDA SNPs shows a complete loss of population structure (Figure S9A). In contrast, a random sampling of 224 SNPs across the genome still recovered the expected genome-wide population structure (Figure S9B).

4 | DISCUSSION

The extreme variability in multiple phenotypic and ecological traits makes *A. clarkii* an intriguing case among clownfish species. It spans the broadest geographic range of all clownfishes, can establish a mutualistic relationship with any of the 10 host sea anemone species and displays a high level of colour polymorphism. We used population genomic, demographic and colour pattern analyses to assess the genetic cohesion between populations and to test different hypotheses about the colonization pathways that enabled this species to reach its broad geographic range. Population genetic analyses of eight clownfish populations revealed high level of differentiation between the two outermost populations, with decreasing levels of differentiation observed among populations that are closer geographically. We further show that *A. clarkii* originated in the Indo-Pacific Ocean and spread to its current range using a stepping-stone mode of dispersal. We identified substantial gene flow throughout the demographic history of the species, despite being heterogeneous over time and among populations. Finally, colour patterns differed noticeably among the Indonesian populations and the two populations at the extreme of the sampling distribution (i.e. Maldives and New Caledonia), which exhibited more comparable colour patterns despite being geographically distant from each other. Our results suggest that despite high differentiation between the Indian and Pacific Oceans populations, gene flow occurred throughout the evolutionary history of the species and is still ongoing. This highlights the intricate nature of the species paradigm.

4.1 | Population structure, genomic landscape of differentiation, divergence and nucleotide diversity

Our results based on whole-genome sequencing revealed substantial genetic structure among *A. clarkii* populations, with the highest level of differentiation between populations from the Western part of the geographic range (Maldives, Indonesia, Taiwan and Philippines) and populations from the Eastern part of the distribution (Papua New Guinea, Solomon Islands and New Caledonia).

Populations from Indonesia, Taiwan and the Philippines displayed the highest nucleotide diversity, consistent with their central distribution providing them with further opportunities for connections to other populations. The New Caledonia population, which is on the edge of the distribution, displayed in contrast higher nucleotide diversity across the genome than both Papua New Guinea and Solomon Islands core populations. A previous study on *A. clarkii* revealed that an edge population from Japan exhibited lower nucleotide diversity than the core populations (Clark et al., 2021), which was expected since populations from the edge of the geographic distribution usually face more substantial genetic drift and have a lower effective population size (N_e), two factors known to decrease genetic diversity (Glémin et al., 2003; Kawecki, 2008). The higher nucleotide diversity that we found in the New Caledonian population might be a consequence of the constant gene-flow from neighbouring populations, as suggested in other coral reef fishes (Bay & Caley, 2011).

Mean differentiation (F_{ST}) across windows was high among populations from different clusters compared to some other damselfish species at a similar spatial scale (Frédérich et al., 2012). This finding was expected because of the limited pelagic larval dispersal (Wellington & Victor, 1989; Ye et al., 2011), with larvae usually spreading within a perimeter of less than 27 km from the spawning site, even if long-distance migration has been observed previously (Pinsky et al., 2010). Furthermore, the social structure of clownfishes restricts larval settlement to not only a specific set of environmental variables but also to the presence of an appropriate sea anemone species. Even in the case of *A. clarkii* and its wide range of anemone hosts, this lifestyle further constrains the available settling sites and could reduce long-distance genetic exchanges between populations. Moreover, the observed population fragmentation might be a remnant of past sea-level fluctuations. The Pleistocene's last glacial maxima (ca. 20,000 years ago) witnessed the formation of the Indo-Pacific barrier, a nearly uninterrupted land barrier (Fleminger, 1986) that induced a substantial reduction of the connectivity between the Indian and Pacific Oceans basins (Voris, 2000). The resulting habitat fragmentation potentially led to vicariance and subsequent genetic drift in the formerly separated populations. This mechanism potentially left behind a genetic signature in current populations, as observed in many fish species (Fitzpatrick et al., 2011; Winters et al., 2010) and invertebrates (Barber et al., 2000; Lessios et al., 2003), with few exceptions to that (Gaither et al., 2010; Horne et al., 2008). Compared to the high inter-cluster genetic structure, differentiation among populations from the same clusters was relatively low and comparable with other clownfish species (Dohna et al., 2015; Timm et al., 2012) and damselfishes (Frédérich et al., 2012; Liu et al., 2014). The limited differentiation across the Indo-Pacific cluster can be explained by the action of the North Equatorial Current (NEC), which splits into the northward-flowing Kuroshio current and the southward Mindanao current (Toole et al., 1990) and likely has homogenized genetic variation across populations, like other coral reef fishes from the Coral Triangle (Liu et al., 2019).

Differentiation (F_{ST}) and divergence (d_{xy}) were highly heterogeneous across the genome. The heterogeneous distribution of

TABLE 1 Parameter estimates for the best model with constant gene flow between neighbouring populations using the coalescent modelling approach implemented in *fastsimcoal2*.

		Parameter	Abbreviation	Estimate	95% CI
Population size		Maldives	NMDV	5916	[5498; 8549]
		Indo-Pacific Ocean	NIPO	19,858	[19,082; 31,838]
		Pacific Ocean	NPAO	3634	[3478; 5282]
		New Caledonia	NNCD	3849	[3617; 6123]
Splitting time		New Caledonia	TNCD	2537	[2406; 4397]
		Maldives	TMDV	11,591	[11,591; 18,323]
		Pacific Ocean	TPAO	60,060	[43,062; 524,780]
Migration rate per generation	After NCD split	MDV → IPO	MIG1_01	5.92E-05	[3.98E-05; 6.40E-05]
		IPO → MDV	MIG1_10	3.73E-05	[2.37E-05; 4.51E-05]
		IPO → PAO	MIG1_12	2.51E-05	[1.27E-05; 2.51E-05]
		PAO → IPO	MIG1_21	9.31E-05	[6.50E-05; 1.02E-04]
		PAO → NCD	MIG1_23	1.57E-04	[1.15E-04; 1.79E-04]
		NCD → PAO	MIG1_32	2.37E-04	[1.53E-04; 2.80E-04]
	After MDV split	MDV → IPO	MIG2_01	3.85E-05	[1.68E-05; 5.14E-05]
		IPO → MDV	MIG2_10	4.78E-05	[2.78E-05; 5.88E-05]
		IPO → PAO	MIG2_12	9.23E-08	[3.92E-08; 1.69E-06]
		PAO → IPO	MIG2_21	7.70E-05	[4.51E-05; 7.70E-05]
	After POA split	IPO → PAO	MIG3_12	9.86E-05	[5.83E-05; 1.26E-04]
		PAO → IPO	MIG3_21	1.16E-05	[4.14E-08; 5.73E-05]

Note: Population sizes are haploid sizes. Splitting times are reported in generations. Migration rates are reported per generation. The model is illustrated in Figures 2E and 4A. 95% confidence intervals (CI) were obtained by parametric bootstrap.
Abbreviations: IPO, Indo-Pacific Ocean; MDV, Maldives; NCD, New Caledonia; PAO, Pacific Ocean.

functional elements and local recombination rate throughout the genome can influence the indirect selection strength and lead to highly heterogeneous patterns of F_{ST} and d_{xy} across the genome, as observed in various taxa (Burri et al., 2015; Pease & Hahn, 2013; Stankowski et al., 2019). Although no genomic window was considered as outlier across all 28 pairwise comparisons – suggesting that gene flow is homogenizing the variance among populations – the detection of highly differentiated parts of the genome comparable across some populations could point to regions that are under selection and might be adaptive, potentially explaining the widespread distribution of *A. clarkii*. Interestingly, no outlier regions for the high levels of differentiation and divergence across populations were enriched for biological functions known to be related to colouration or preference for host sea anemone—two highly variable traits among *A. clarkii* populations. Although we cannot rule out that the observed peaks of differentiation could act as barrier loci (i.e. "islands of speciation"; Turner et al., 2005), it is more likely that they reflect various factors influencing the level of nucleotide diversity. Indeed, processes such as recombination rate, genetic drift and local adaptation are known to affect the nucleotide diversity, which in turn influences the level of differentiation (Cruickshank & Hahn, 2014; Hartasánchez et al., 2014; Noor & Bennett, 2009; Wolf & Ellegren, 2017).

4.2 | Demographic reconstruction of *Amphiprion clarkii* colonization

4.2.1 | Origins of *Amphiprion clarkii* populations

The demographic modelling based on SNPs data supported the hypothesis of a centre of origin of all *A. clarkii* populations in the Indo-Pacific Ocean and was favoured over a centre of origin located in the Papua New Guinea-Solomon Islands area. According to this model, range expansion in *A. clarkii* was thus most likely achieved through a stepping stone process over multiple generations, like other clownfish species (Huyghe & Kochzius, 2018). Additionally, sea surface currents might have facilitated dispersal, as they are known to impact marine species connectivity (Kool et al., 2011). Expansion into new areas can thus be achieved even in species with characteristics linked to low dispersal ability. Indeed, a recent meta-analysis highlighted that life-history traits are not strong predictors of population structure and that traits such as pelagic larval duration were not good predictors of dispersal (Gandra et al., 2021). The migration pattern we uncovered for *A. clarkii*, is also consistent with the origin of the whole clownfish adaptive radiation, which started in the Indo-Pacific Ocean (Litsios, Kostikova, & Salamin, 2014; Litsios, Pearman, et al., 2014; Santini & Polacco, 2006). The findings further

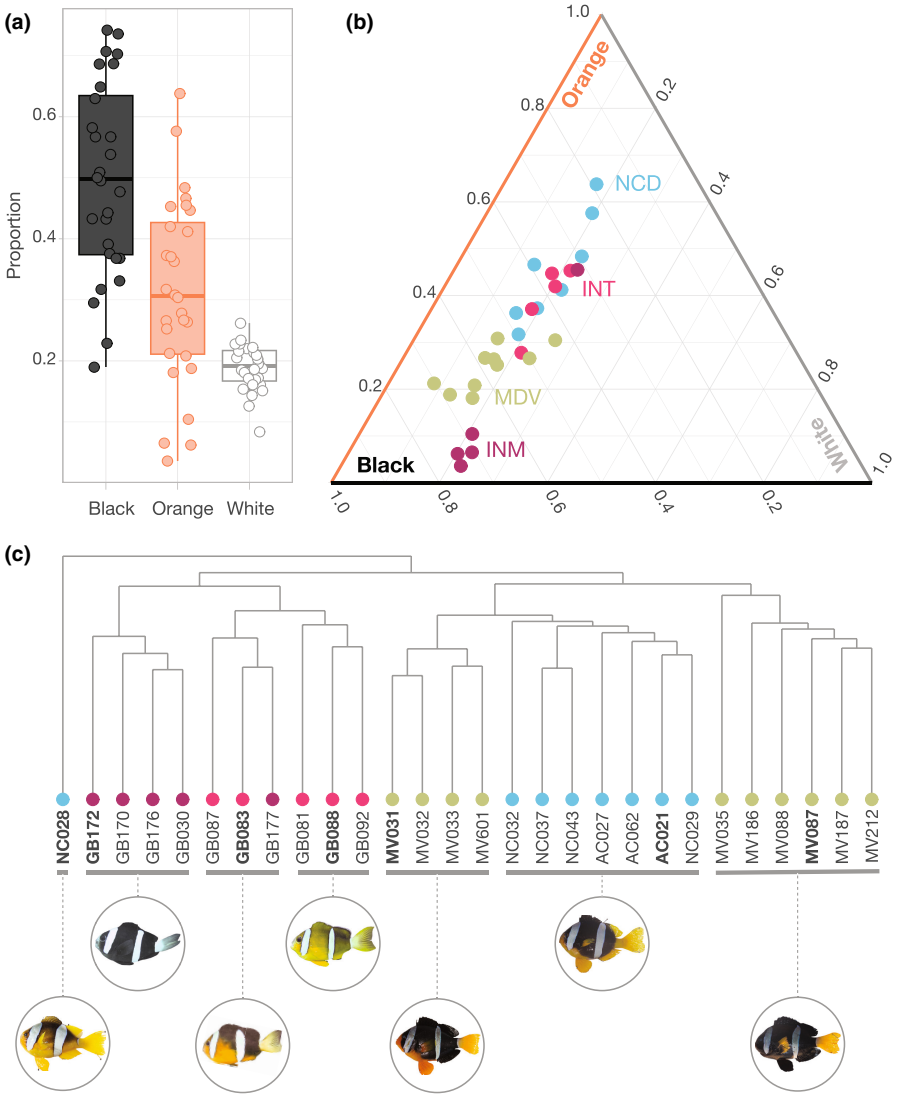

FIGURE 5 Colour proportions in *Amphiprion clarkii*. (a) Distribution of black, orange and white colour proportions for the 28 individuals with available pictures (see Table S10). (b) Ternary plot showing the proportion of each of black, orange and white for each individual. Coloured points represent the different populations. Blue: New Caledonia (NCD); green: Maldives (MDV); pink: Indonesia-Tulamben (INT); and purple: Indonesia-Manado (INM). (c) Dendrogram based on the hierarchical clustering of the 28 axis of the principal component analysis (PCA) summarizing the colour of each pixel for the 28 individuals. Pictures depict the typical individual for each cluster highlighted by a horizontal line. The individuals in bold correspond to the ones illustrating the colour pattern of the cluster. Coloured points represent the different populations.

support the possibility of range expansion and colonization for species with limited dispersal ability. More generally, this pattern of colonization coincides with the "Centre of Origin" hypothesis of the generation of the observed biodiversity (Briggs, 2003), which suggests that older lineages occur in the centre of the Coral Triangle and have longer evolutionary history compared to Pacific Ocean and Indian Ocean populations that were colonized more recently (Drew & Barber, 2009).

4.2.2 | Gene flow between *Amphiprion clarkii* populations

The best model supported the hypothesis of migration throughout *A. clarkii* demographic history. The low likelihood of the two models representing primary and secondary contacts of populations further supported that populations were constantly in contact with no interruption of gene flow between neighbouring populations. This result was somewhat unexpected since historical processes such as climatic oscillations across geological time are known to have significantly influenced the genetic structure of many marine species over time (Hewitt, 2000). Hence, we expected that gene flow in *A. clarkii* would have varied across geological time according to various environmental factors known to act as barriers in the marine environment, such as temperature, salinity or oceanic currents (Oleksiak, 2019). Gene flow might thus have occurred shortly after populations split (Malinsky et al., 2015; Schumer et al., 2018)—or following a period without exchange between populations (secondary contact; Canino et al., 2010; Liu et al., 2012; Nikolic et al., 2020).

Although the migration level was heterogeneous across populations and time, the extent of gene flow was considerably higher compared to what was previously found between sympatric species of clownfishes (ranging from 1.29E-8 to 7.08E-7; Marcionetti, 2021) and was more comparable to population-level gene flow in other marine species (ranging from 3E-6 to 4E-6; Nikolic et al., 2020). Additionally, *A. clarkii* populations diverged in the recent past, with at least 10 times lower divergence times among their populations than the divergence times found in sympatric clownfish species (about 45,697 generations; Marcionetti, 2021). The prevalence of gene flow throughout *A. clarkii*'s demographic history and the comparisons of

demographic parameters with other clownfish species suggest that populational processes—and not speciation processes—are driving *A. clarkii* diversification. Furthermore, the constant and substantial gene flow between *A. clarkii* populations builds on the pervasiveness of gene flow in the evolutionary history of clownfishes (Gainsford et al., 2020; Litsios & Salamin, 2014; Schmid et al., 2022). The presence of past gene flow among the *A. clarkii* populations might have enabled the species to reach a high level of genetic variation by reassembling old genetic variations into new combinations (Marques et al., 2019). The newly emerged variations might have enabled them to colonize various environmental niches and occupy all the 10 host sea anemone species, similar to the ancestral hybridization events between clownfish species that have potentially triggered their adaptive radiation (Schmid et al., 2022).

4.3 | Colouration in *Amphiprion clarkii*

We identified potential genomic regions associated with colouration patterns in *A. clarkii* but none of them were found to be linked to pigmentation in other teleost fishes (Lorin et al., 2018). The absence of known genomic regions related to pigmentation implies that the colouration patterns observed in *A. clarkii* may be specific to this species or that our analysis might not have fully captured the intricate genetic factors influencing colouration patterns. Intriguingly, the genomic regions potentially implicated in colouration in *A. clarkii* were more prevalent on certain chromosomes than others. Particularly, chromosome 18 exhibited a high density of SNPs possibly connected to colour patterns and has previously been highlighted in various studies concerning clownfishes (Marcionetti & Salamin, 2023; Marcionetti et al., 2024; Schmid et al., 2022). Indeed, chromosome 18 exhibits distinct genomic differentiation patterns in other species of clownfishes and also demonstrates unique phylogenetic relationships within the *Amphiprion* genus, possibly linked to a significant history of hybridization (Marcionetti & Salamin, 2023; Schmid et al., 2022).

In addition, the colouration-based clustering highlighted that populations at outer edges of the sampling distribution (i.e. New Caledonia and the Maldives) show more similarity to each other than to the Indonesian populations, which are located in the centre of the geographic distribution. The higher similarity in colour pattern between the two most geographically and genetically distant populations raises the possibility that there may be selective pressures or ecological conditions specific to the outer edges of the distribution range that are driving similar colouration patterns. Previous researches have described colour trait plasticity in *A. clarkii* (Devi et al., 2021) and how sea anemone hosts can influence clownfish species' colouration (Militz et al., 2016). Interestingly, other types of coral reef species that experienced a rapid increase in diversity also exhibit significant variation in colouration (*Hypoplectrus* spp.; Hench et al., 2022). This phenomenon has also been noted in other species such as the *Heliconius* butterflies, which have undergone adaptive radiation (Edelman et al., 2019). Colour pattern is thus a crucial ecological trait under strong selection in various cases of adaptive radiations (Campagna et al., 2017; Stryjewski & Sorenson, 2017; Belleghem et al., 2017), and various factors such as environmental heterogeneity and context-dependent selection can have a substantial impact on the appearance of colour variations (Sirkia & Qvarnstrom, 2021). Further investigations are needed to evaluate the impact of colour patterns on the fitness of *A. clarkii*, its underlying genetic causes and the potential role played by chromosome 18. This information will be crucial to understand the potential link between the colour polymorphism and the widespread geographic distribution of the species as well as its ability to interact with all the ten host sea anemone species.

4.4 | Can we consider *Amphiprion clarkii* as a single species?

Polymorphism's role in speciation has been well documented (Forsman et al., 2008; Gray & McKinnon, 2007) and could lead to an increase in speciation rates (Hugall & Stuart-Fox, 2012). The high colour polymorphism displayed by *A. clarkii*, as well as its broad geographic distribution and ability to interact with the 10 host sea anemone species, had been persistent arguments calling for further investigation of the phylogenetic status of the species (Litsios & Salamin, 2014). The contrast between the high differentiation (F_{ST}) of the two populations at the extreme of the sampling distribution and their similarity in colour patterns suggested that there is potential for ecological speciation. Indeed, similar colour patterns are not always linked to a common ancestry, as they could result from convergent evolution due to shared selection pressure (Meyer, 2006), as it was highlighted in some cichlid fishes (Duftner et al., 2007). However, the constant gene flow throughout the evolutionary history of the species and the absolute divergence among populations points to the other direction and suggest that *A. clarkii* should be considered as a single species entity. There are many ways to define what a species is (De Queiroz, 2007) and even more methodological approaches to delineate them (Petzold & Hassanin, 2020). A critical part of species delimitation is to consider morphological, behavioural, and ecological data in parallel to genomic data. We considered mainly genomic data in this study, and the incorporation of behavioural and ecological data is needed in the future. Indeed, field observations suggest that *A. clarkii* is an aggressive fish and is remarkably efficient at defending its host sea anemone against fishes feeding on its tentacles compared to other clownfishes. This bold behaviour might have enabled *A. clarkii* to colonize even smaller sea anemones neglected by other clownfish species (Santini & Polacco, 2006). Moreover, morphological data should be investigated more thoroughly. We did not consider more subtle variations in colour patterns, which might be a key feature that makes populations highly distinguishable since cryptic species often have divergent phenotypes that are hard to measure (Singhal et al., 2018). Another consideration for studies based on genomic data alone, is that in the past decade, various cases of speciation with gene flow

have been found (Capblancq et al., 2019; Wang et al., 2016). This shows that in some cases, only a limited part of the genome with restricted gene flow may be necessary to form new species. Thus, special care should be taken when investigating gene flow to consider genes potentially critical for adaptation, that is that are under selection (Cadena & Zapata, 2021; Vaux et al., 2021).

While the most intriguing questions may not revolve around whether *A. clarkii* is a single species, but rather on the complex nature of genetic and phenotypic diversity emerging over time, we still recommend that *A. clarkii* be regarded as a single species. This raises the question of the factors that enabled *A. clarkii* to reach such a broad distribution compared to all other clownfish species, despite their sedentary adult life and reduced larval dispersal time. This question is even more interesting considering previous suggestions of a trade-off between host specialization (i.e. number of host sea anemone species) and environmental specialization, with species interacting with a limited number of sea anemone species exhibiting a broader environmental niche (Litsios, Kostikova, & Salamin, 2014). However, the broad geographical range exhibited by *A. clarkii* might be a consequence of their ability to interact with the 10 host sea anemone species or their colour polymorphism. If one (or both) of those traits enhanced dispersal, the subsequent mating of the high-dispersing individuals might reinforce the phenotype's frequency and amplitude, leading to increased dispersal, a process known as spatial sorting (Comerford & Egan, 2022; Shine et al., 2011). This process can produce a shift in the phenotypes that improve dispersal against the potential negative impact on the organism's fitness in the newly colonized habitat. This study thus opens new paths for understanding how polymorphism emerged in *A. clarkii*, and why—contrary to other clownfish species—it did not diversify into multiple more specialized species. Future work should include a more detailed assessment of the species' ecology regarding environmental niche and interactions with the host sea anemone and other clownfish species.

AUTHOR CONTRIBUTIONS

MBS, NS and SS designed research. JB, FC, SH, FH, GL, AM, JO, CR and VT contributed to sample collection. MBS and SS performed research, lab work and analysed the data. AGJ carried out the colouration analyses. SS wrote the manuscript. All authors read, made corrections and approved the final version of the manuscript.

ACKNOWLEDGEMENTS

We thank the Lausanne Genomic Technologies Facility for sequencing the samples and the DCSR infrastructure of the University of Lausanne for the computational resources. 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. Finally, thanks to the two anonymous reviewers for their constructive comments, which contributed greatly to the final version of this manuscript. Funding was from the University of Lausanne funds, Swiss National Science Foundation, Grant Number: 31003A-163428. FC was supported by an Australian Research Council (ARC) Discovery Early Career Research Award DE200100620 and Discovery Project DP180102363. Open access funding provided by Universite de Lausanne.

CONFLICT OF INTEREST STATEMENT

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

DATA AVAILABILITY STATEMENT

Metadata were uploaded in GEOME (https://n2t.net/ark:/21547/R2565), and the raw Illumina reads were deposited in the SRA (BioProject ID: PRJNA1025355). The scripts and pictures used for the colour and RDA analysis as well as the final VCF file were deposited on Dryad (doi:10.5061/dryad.xwdbrv1nd).

BENEFIT-SHARING STATEMENT

Benefits from this research accrue from the sharing of our data and results on public databases (see above). We included all the collaborators providing tissue samples for this study as co-authors. All fieldwork was performed in agreement with local regulation and in collaboration with local entities. We thank the local authorities for permit to collect samples and help with field logistics (research permits number are listed in Appendix S1).

ORCID

Sarah Schmid https://orcid.org/0000-0002-9520-4086
Joris A. M. Bertrand https://orcid.org/0000-0002-3379-1019
Fabio Cortesi https://orcid.org/0000-0002-7518-6159
Anna Marcionetti https://orcid.org/0000-0002-2450-2879
Cynthia Riginos https://orcid.org/0000-0002-5485-4197
Valerio Tettamanti https://orcid.org/0000-0002-9078-734X

REFERENCES

Alexa, A., & Rahnenfuhrer, J. (2020). *topGO: Enrichment analysis for Gene Ontology. R package version 2.43.0*.

Allen, G. R. (2008). Conservation hotspots of biodiversity and endemism for Indo-Pacific coral reef fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *18*(5), 541–556. https://doi.org/10.1002/aqc.880

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697–709. https://doi.org/10.1038/nrg2844

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arroyo-Correa, B., Jordano, P., & Bartomeus, I. (2023). Intraspecific variation in species interactions promotes the feasibility of mutualistic assemblages. *Ecology Letters*, *26*(3), 448–459. https://doi.org/10.1111/ele.14163

Barabás, G., & D'Andrea, R. (2016). The effect of intraspecific variation and heritability on community pattern and robustness. *Ecology Letters*, *19*(8), 977–986. https://doi.org/10.1111/ele.12636

Barber, P. H., Palumbi, S. R., Erdmann, M. V., & Moosa, M. K. (2000). A marine Wallace's line? *Nature*, *406*(6797), 692–693. https://doi.org/10.1038/35021135

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. https://doi.org/10.1093/bioinformatics/btr174

Bay, L. K., & Caley, M. J. (2011). Greater genetic diversity in spatially restricted coral reef fishes suggests secondary contact among differentiated lineages. *Diversity*, *3*(3), 483–502. https://doi.org/10.3390/d3030483

Belanger, C. L., Jablonski, D., Roy, K., Berke, S. K., Krug, A. Z., & Valentine, J. W. (2012). Global environmental predictors of benthic marine biogeographic structure. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(35), 14046–14051. https://doi.org/10.1073/pnas.1212381109

Beldade, R., Holbrook, S. J., Schmitt, R. J., Planes, S., Malone, D., & Bernardi, G. (2012). Larger female fish contribute disproportionately more to self-replenishment. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1736), 2116–2121. https://doi.org/10.1098/rspb.2011.2433

Bell, L. J., Moyer, J. T., & Numachi, K. (1982). Morphological and genetic variation in Japanese populations of the anemonefish *Amphiprion clarkii*. *Marine Biology*, *72*(2), 99–108. https://doi.org/10.1007/BF00396909

Belleghem, S. M. V., Rastas, P., Papanicolaou, A., Martin, S. H., Arias, C. F., Supple, M. A., Hanly, J. J., Mallet, J., Lewis, J. J., Hines, H. M., Ruiz, M., Salazar, C., Linares, M., Moreira, G. R. P., Jiggins, C. D., Counterman, B. A., McMillan, W. O., & Papa, R. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. *Nature Ecology & Evolution*, *1*(3), 0052. https://doi.org/10.1038/s41559-016-0052

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Briggs, J. C. (2003). Marine centres of origin as evolutionary engines. *Journal of Biogeography*, *30*(1), 1–18. https://doi.org/10.1046/j.1365-2699.2003.00810.x

Buckley, S. J., Brauer, C. J., Unmack, P. J., Hammer, M. P., & Beheregaray, L. B. (2022). Variation in intraspecific demography drives localised concordance but species-wide discordance in response to past climatic change. *BMC Ecology and Evolution*, *22*(1), 35. https://doi.org/10.1186/s12862-022-01990-2

Burri, R., Nater, A., Kawakami, T., Mugal, C. F., Olason, P. I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L. Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S.-A., Sæthre, G.-P., Török, J., & Ellegren, H. (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. https://doi.org/10.1101/gr.196485.115

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. https://doi.org/10.1111/j.1095-8649.2007.01445.x

Buston, P. M., Jones, G. P., Planes, S., & Thorrold, S. R. (2012). Probability of successful larval dispersal declines fivefold over 1 km in a coral reef fish. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1735), 1883–1888. https://doi.org/10.1098/rspb.2011.2041

Cadena, C. D., & Zapata, F. (2021). The genomic revolution and species delimitation in birds (and other organisms): Why phenotypes should not be overlooked? *Ornithology*, *138*(2), ukaa069. https://doi.org/10.1093/ornithology/ukaa069

Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances*, *3*(5), e1602404. https://doi.org/10.1126/sciadv.1602404

Canino, M. F., Spies, I. B., Cunningham, K. M., Hauser, L., & Grant, W. S. (2010). Multiple ice-age refugia in Pacific cod, *Gadus macrocephalus*. *Molecular Ecology*, *19*(19), 4339–4351. https://doi.org/10.1111/j.1365-294X.2010.04815.x

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A swiss army knife for landscape genomics. *Methods in Ecology and Evolution*, *12*(12), 2298–2309. https://doi.org/10.1111/2041-210X.13722

Capblancq, T., Mavárez, J., Rioux, D., & Després, L. (2019). Speciation with gene flow: Evidence from a complex of alpine butterflies (*Coenonympha*, Satyridae). *Ecology and Evolution*, *9*(11), 6444–6457. https://doi.org/10.1002/ece3.5220

Choo, L. Q., Bal, T. M. P., Goetze, E., & Peijnenburg, K. T. C. A. (2021). Oceanic dispersal barriers in a holoplanktonic gastropod. *Journal of Evolutionary Biology*, *34*(1), 224–240. https://doi.org/10.1111/jeb.13735

Clark, R. D., Aardema, M. L., Andolfatto, P., Barber, P. H., Hattori, A., Hoey, J. A., Montes, H. R., & Pinsky, M. L. (2021). Genomic signatures of spatially divergent selection at clownfish range margins. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1952), 20210407. https://doi.org/10.1098/rspb.2021.0407

Comerford, M., & Egan, S. P. (2022). The potential role of spatial sorting in speciation and adaptive radiations. *Population Ecology*, *2022*, 1–8. https://doi.org/10.1002/1438-390X.12108

Conover, D. O., Clarke, L. M., Munch, S. B., & Wagner, G. N. (2006). Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology*, *69*(sc), 21–47. https://doi.org/10.1111/j.1095-8649.2006.01274.x

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. https://doi.org/10.1111/mec.12796

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. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879–886. https://doi.org/10.1080/10635150701701083

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), 40,519. https://doi.org/10.1038/srep40519

Devi, V. S., Salitha, S., Jose, S., Paramasivam, P., Anand, J., Ajith, K. T. T., & Kuldeep, K. L. (2021). Molecular characterization on the morphotypes of *Amphiprion clarkii* (Bennett, 1830) from Indian waters with biogeographic notes. *Thalassas: An International Journal of Marine Sciences*, *37*, 477–485. https://doi.org/10.1007/s41208-021-00330-3

Dohna, T. A., Timm, J., Hamid, L., & Kochzius, M. (2015). Limited connectivity and a phylogeographic break characterize populations of the pink anemonefish, *Amphiprion perideraion*, in the Indo-Malay archipelago: Inferences from a mitochondrial and microsatellite loci. *Ecology and Evolution*, *5*(8), 1717–1733. https://doi.org/10.1002/ece3.1455

Drew, J., & Barber, P. H. (2009). Sequential cladogenesis of the reef fish *Pomacentrus moluccensis* (Pomacentridae) supports the peripheral origin of marine biodiversity in the Indo-Australian archipelago. *Molecular Phylogenetics and Evolution*, *53*(1), 335–339. https://doi.org/10.1016/j.ympev.2009.04.014

Duftner, N., Sefc, K. M., Koblmüller, S., Salzburger, W., Taborsky, M., & Sturmbauer, C. (2007). Parallel evolution of facial stripe patterns in the *Neolamprologus brichardi/pulcher* species complex endemic to Lake Tanganyika. *Molecular Phylogenetics and Evolution*, *45*(2), 706–715. https://doi.org/10.1016/j.ympev.2007.08.001

Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., García-Accinelli, G., Belleghem, S. M. V., Patterson, N., Neafsey, D. E., Challis, R., Kumar, S., Moreira, G. R. P., Salazar, C., Chouteau,

M., Counterman, B. A., Papa, R., Blaxter, M., Reed, R. D., … Mallet, J. (2019). Genomic architecture and introgression shape a butterfly radiation. *Science*, *366*(6465), 594–599. https://doi.org/10.1126/science.aaw2090

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, *29*(1), 51–63. https://doi.org/10.1016/j.tree.2013.09.008

Endler, J. A. (2012). A framework for analysing colour pattern geometry: Adjacent colours. *Biological Journal of the Linnean Society*, *107*(2), 233–253. https://doi.org/10.1111/j.1095-8312.2012.01937.x

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. https://doi.org/10.1371/journal.pgen.1003905

Fautin, D. G., & Allen, G. R. (1997). *Field guide to anemonefishes and their host sea anemones* (2nd revised ed.) ed. Western Australian Museum.

Fitzpatrick, J. M., Carlon, D. B., Lippe, C., & Robertson, D. R. (2011). The West Pacific diversity hotspot as a source or sink for new species? Population genetic insights from the Indo-Pacific parrotfish *Scarus rubroviolaceus*. *Molecular Ecology*, *20*(2), 219–234. https://doi.org/10.1111/j.1365-294X.2010.04942.x

Fleminger, A. (1986). The Pleistocene equatorial barrier between the Indian and Pacific oceans and a likely cause for Wallace's line. *UNESCO Technical Papers in Marine Science*, *49*, 84–97.

Forsman, A., Ahnesjö, J., Caesar, S., & Karlsson, M. (2008). A model of ecological and evolutionary consequences of color polymorphism. *Ecology*, *89*(1), 34–40. https://doi.org/10.1890/07-0572.1

Frédérich, B., Liu, S.-Y. V., & Dai, C.-F. (2012). Morphological and genetic divergences in a coral reef damselfish, *Pomacentrus coelestis*. *Evolutionary Biology*, *39*(3), 359–370. https://doi.org/10.1007/s11692-011-9158-z

Gaboriau, T., Leprieur, F., Mouillot, D., & Hubert, N. (2017). Influence of the geography of speciation on current patterns of coral reef fish biodiversity across the Indo-Pacific. *Ecography*, *41*(8), 1295–1306. https://doi.org/10.1111/ecog.02589

Gainsford, A., Jones, G. P., Hobbs, J. A., Heindler, F. M., & Herwerden, L. (2020). Species integrity, introgression, and genetic variation across a coral reef fish hybrid zone. *Ecology and Evolution*, *10*(21), 11998–12014. https://doi.org/10.1002/ece3.6769

Gaither, M. R., Toonen, R. J., Robertson, D. R., Planes, S., & Bowen, B. W. (2010). Genetic evaluation of marine biogeographical barriers: Perspectives from two widespread Indo-Pacific snappers (*Lutjanus kasmira* and *Lutjanus fulvus*). *Journal of Biogeography*, *37*(1), 133–147. https://doi.org/10.1111/j.1365-2699.2009.02188.x

Gandra, M., Assis, J., Martins, M. R., & Abecasis, D. (2021). Reduced global genetic differentiation of exploited marine fish species. *Molecular Biology and Evolution*, *38*(4), 1402–1412. https://doi.org/10.1093/molbev/msaa299

Glémin, S., Ronfort, J., & Bataillon, T. (2003). Patterns of inbreeding depression and architecture of the load in subdivided populations. *Genetics*, *165*(4), 2193–2212. https://doi.org/10.1093/genetics/165.4.2193

Gray, S. M., & McKinnon, J. S. (2007). Linking color polymorphism maintenance and speciation. *Trends in Ecology & Evolution*, *22*(2), 71–79. https://doi.org/10.1016/j.tree.2006.10.005

Hartasánchez, D. A., Vallès-Codina, O., Brasó-Vives, M., & Navarro, A. (2014). Interplay of interlocus gene conversion and crossover in segmental duplications under a neutral scenario. *G3: Genes, Genomes, Genetics*, *4*(8), 1479–1489. https://doi.org/10.1534/g3.114.012435

Hellberg, M. E. (2009). Gene flow and isolation among populations of marine animals. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 291–310. https://doi.org/10.1146/annurev.ecolsys.110308.120223

Hench, K., Helmkampf, M., McMillan, W. O., & Puebla, O. (2022). Rapid radiation in a highly diverse marine environment. *Proceedings of the National Academy of Sciences*, *119*(4), e2020457119. https://doi.org/10.1073/pnas.2020457119

Hewitt, G. (2000). The genetic legacy of the quaternary ice ages. *Nature*, *405*(6789), 907–913. https://doi.org/10.1038/35016000

Hoeksema, B. W. (2007). Delineation of the Indo-Malayan Centre of maximum marine biodiversity: The coral triangle. In W. Renema (Ed.), *Biogeography, time, and place: Distributions, barriers, and islands* (pp. 117–178). Springer Netherlands. https://doi.org/10.1007/978-1-4020-6374-9_5

Horne, J. B., van Herwerden, L., Choat, J. H., & Robertson, D. R. (2008). High population connectivity across the Indo-Pacific: Congruent lack of phylogeographic structure in three reef fish congeners. *Molecular Phylogenetics and Evolution*, *49*(2), 629–638. https://doi.org/10.1016/j.ympev.2008.08.023

Hugall, A. F., & Stuart-Fox, D. (2012). Accelerated speciation in colour-polymorphic birds. *Nature*, *485*(7400), 631–634. https://doi.org/10.1038/nature11050

Hughes, T. P., Bellwood, D. R., & Connolly, S. R. (2002). Biodiversity hotspots, centres of endemicity, and the conservation of coral reefs. *Ecology Letters*, *5*(6), 775–784. https://doi.org/10.1046/j.1461-0248.2002.00383.x

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. https://doi.org/10.1111/maec.12357

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. https://doi.org/10.1371/journal.pone.0193825

Jones, G. P., Planes, S., & Thorrold, S. R. (2005). Coral reef fish larvae settle close to home. *Current Biology*, *15*(14), 1314–1318. https://doi.org/10.1016/j.cub.2005.06.061

Joshi, N., & Fass, J. (2011). *Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files*. https://github.com/najoshi/sickle

Kawecki, T. J. (2008). Adaptation to marginal habitats. *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 321–342. https://doi.org/10.1146/annurev.ecolsys.38.091206.095622

Keyse, J., Crandall, E. D., Toonen, R. J., Meyer, C. P., Treml, E. A., & Riginos, C. (2014). The scope of published population genetic data for Indo-Pacific marine fauna and future research opportunities in the region. *Bulletin of Marine Science*, *90*, 47–78. https://doi.org/10.5343/bms.2012.1107

Kool, J. T., Paris, C. B., Barber, P. H., & Cowen, R. K. (2011). Connectivity and the development of population genetic structure in Indo-West Pacific coral reef communities. *Global Ecology and Biogeography*, *20*(5), 695–706. https://doi.org/10.1111/j.1466-8238.2010.00637.x

Legendre, P., & Legendre, L. (1998). *Numerical ecology*. Elsevier.

Lehmann, R., Lightfoot, D. J., Schunter, C., Michell, C. T., Ohyanagi, H., Mineta, K., Foret, S., Berumen, M., 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. https://doi.org/10.1101/278267

Lessios, H. A., Kane, J., & Robertson, D. R. (2003). Phylogeography of the pantropical sea urchin *Tripneustes*: Contrasting patterns of population structure between oceans. *Evolution*, *57*(9), 2026–2036. https://doi.org/10.1111/j.0014-3820.2003.tb00382.x

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The sequence alignment/map format

and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Link, V., Kousathanas, A., Veeramah, K., Sell, C., Scheu, A., & Wegmann, D. (2017). ATLAS: Analysis tools for low-depth and ancient Samples. *BioRxiv*. https://doi.org/10.1101/105346

Litsios, G., Kostikova, A., & Salamin, N. (2014). Host specialist clownfishes are environmental niche generalists. *Proceedings of the Royal Society of London B: Biological Sciences*, *281*(1795), 20133220. https://doi.org/10.1098/rspb.2013.3220

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. https://doi.org/10.1111/jbi.12370

Litsios, G., & Salamin, N. (2014). Hybridisation and diversification in the adaptive radiation of clownfishes. *BMC Evolutionary Biology*, *14*(1), 245. https://doi.org/10.1186/s12862-014-0245-5

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. https://doi.org/10.1186/1471-2148-12-212

Liu, M., Lin, L., Gao, T., Yanagimoto, T., Sakurai, Y., & Grant, W. S. (2012). What maintains the central North Pacific genetic discontinuity in Pacific herring? *PLoS One*, *7*(12), e50340. https://doi.org/10.1371/journal.pone.0050340

Liu, S.-Y. V., Chang, F.-T., Borsa, P., Chen, W.-J., & Dai, C.-F. (2014). Phylogeography of the humbug damselfish, *Dascyllus aruanus* (Linnaeus, 1758): Evidence of Indo-Pacific vicariance and genetic differentiation of peripheral populations. *Biological Journal of the Linnean Society*, *113*(4), 931–942. https://doi.org/10.1111/bij.12378

Liu, S. Y. V., Tuanmu, M.-N., Rachmawati, R., Mahardika, G. N., & Barber, P. H. (2019). Integrating phylogeographic and ecological niche approaches to delimitating cryptic lineages in the blue–green damselfish (*Chromis viridis*). *PeerJ*, *7*, e7384. https://doi.org/10.7717/peerj.7384

Lohman, D. J., De Bruyn, M., Page, T., Von Rintelen, K., Hall, R., Ng, P. K. L., Shih, H.-T., Carvalho, G. R., & Von Rintelen, T. (2011). Biogeography of the Indo-Australian archipelago. *Annual Review of Ecology, Evolution, and Systematics*, *42*(1), 205–226. https://doi.org/10.1146/annurev-ecolsys-102,710-145,001

Lorin, T., Brunet, F. G., Laudet, V., & Volff, J.-N. (2018). Teleost fish-specific preferential retention of pigmentation gene-containing families after whole genome duplications in vertebrates. *G3: Genes, Genomes, Genetics*, *8*(5), 1795–1806. https://doi.org/10.1534/g3.118.200201

Madduppa, H. H., Timm, J., & Kochzius, M. (2014). Interspecific, spatial and temporal variability of self-recruitment in anemonefishes. *PLoS One*, *9*(2), e90648. https://doi.org/10.1371/journal.pone.0090648

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. https://doi.org/10.1126/science.aac9927

Marcionetti, A. (2021). *Genomics of clownfish adaptive radiation*. University of Lausanne.

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*(11), e17347. https://doi.org/10.1111/mec.17347

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. https://doi.org/10.1093/gbe/evz042

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. https://doi.org/10.1093/gbe/evad088

Marko, P. B. (2004). "What's larvae got to do with it?" Disparate patterns of post-glacial population structure in two benthic marine gastropods with identical dispersal potential. *Molecular Ecology*, *13*(3), 597–611. https://doi.org/10.1046/j.1365-294x.2004.02096.x

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. https://doi.org/10.1016/j.tree.2019.02.008

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, *210*(2), 719–731. https://doi.org/10.1534/genetics.118.301336

Messer, P. W., Ellner, S. P., & Hairston, N. G. (2016). Can population genetics adapt to rapid evolution? *Trends in Genetics*, *32*(7), 408–418. https://doi.org/10.1016/j.tig.2016.04.005

Meyer, A. (2006). Repeating patterns of mimicry. *PLoS Biology*, *4*(10), e341. https://doi.org/10.1371/journal.pbio.0040341

Meyer, C. P., Geller, J. B., & Paulay, G. (2005). Fine scale endemism on coral reefs: Archipelagic differentiation in turbinid gastropods. *Evolution*, *59*(1), 113–125. https://doi.org/10.1111/j.0014-3820.2005.tb00899.x

Militz, T. A., McCormick, M. I., Schoeman, D. S., Kinch, J., & Southgate, P. C. (2016). Frequency and distribution of melanistic morphs in coexisting population of nine clownfish species in Papua New Guinea. *Marine Biology*, *163*(10), 200. https://doi.org/10.1007/s00227-016-2972-1

Nikolic, N., Liu, S., Jacobsen, M. W., Jónsson, B., Bernatchez, L., Gagnaire, P.-A., & Hansen, M. M. (2020). Speciation history of European (*Anguilla Anguilla*) and American eel (*A. rostrata*), analysed using genomic data. *Molecular Ecology*, *29*(3), 565–577. https://doi.org/10.1111/mec.15342

Noor, M. A. F., & 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. https://doi.org/10.1038/hdy.2009.151

Oleksiak, M. F. (2019). Adaptation without boundaries: Population genomics in marine systems. In O. P. Rajora (Ed.), *Population genomics: Concepts, approaches and applications* (pp. 587–612). Springer International Publishing. https://doi.org/10.1007/13836_2018_32

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. https://doi.org/10.1098/rspb.2006.3758

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. https://doi.org/10.1111/nph.14951

Palumbi, S. R. (1992). Marine speciation on a small planet. *Trends in Ecology & Evolution*, *7*(4), 114–118. https://doi.org/10.1016/0169-5347(92)90144-Z

Pease, J. B., & Hahn, M. W. (2013). More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution; International Journal of Organic Evolution*, *67*(8), 2376–2384. https://doi.org/10.1111/evo.12118

Pellissier, L., Leprieur, F., Parravicini, V., Cowman, P. F., Kulbicki, M., Litsios, G., Olsen, S. M., Wisz, M. S., Bellwood, D. R., & Mouillot, D. (2014). Quaternary coral reef refugia preserved fish diversity. *Science*, *344*(6187), 1016–1019. https://doi.org/10.1126/science.1249853

Petzold, A., & Hassanin, A. (2020). A comparative approach for species delimitation based on multiple methods of multi-locus DNA sequence analysis: A case study of the genus *Giraffa* (Mammalia,

Cetartiodactyla). *PLoS One*, *15*(2), e0217956. https://doi.org/10.1371/journal.pone.0217956

Pinsky, M. L., Montes, H. R., & Palumbi, S. R. (2010). Using isolation by distance and effective density to estimate dispersal scales in anemonefish. *Evolution; International Journal of Organic Evolution*, *64*(9), 2688–2700. https://doi.org/10.1111/j.1558-5646.2010.01003.x

Renema, W., Bellwood, D. R., Braga, J. C., Bromfield, K., Hall, R., Johnson, K. G., Lunt, P., Meyer, C. P., McMonagle, L. B., Morley, R. J., O'Dea, A., Todd, J. A., Wesselingh, F. P., Wilson, M. E. J., & Pandolfi, J. M. (2008). Hopping hotspots: Global shifts in marine biodiversity. *Science*, *321*(5889), 654–657. https://doi.org/10.1126/science.1155674

Rolland, J., Silvestro, D., Litsios, G., Faye, L., & Salamin, N. (2018). Clownfishes evolution below and above the species level. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1873), 20171796. https://doi.org/10.1098/rspb.2017.1796

Roux, N., Salis, P., Lambert, A., Logeux, V., Soulat, O., Romans, P., Frédérich, B., Lecchini, D., & Laudet, V. (2019). Staging and normal table of postembryonic development of the clownfish (*Amphiprion ocellaris*). *Developmental Dynamics: An Official Publication of the American Association of the Anatomists*, *248*(7), 545–568. https://doi.org/10.1002/dvdy.46

Roux, N., Salis, P., Lee, S.-H., Besseau, L., & Laudet, V. (2020). Anemonefish, a model for eco-Evo-devo. *EvoDevo*, *11*, 20. https://doi.org/10.1186/s13227-020-00166-7

Rudman, S. M., Rodriguez-Cabal, M. A., Stier, A., Sato, T., Heavyside, J., El-Sabaawi, R. W., & Crutsinger, G. M. (2015). Adaptive genetic variation mediates bottom-up and top-down control in an aquatic ecosystem. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1812), 20151234. https://doi.org/10.1098/rspb.2015.1234

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. https://doi.org/10.1371/journal.pone.0032788

Santini, S., & Polacco, G. (2006). Finding Nemo: Molecular phylogeny and evolution of the unusual life style of anemonefish. *Gene*, *385*, 19–27. https://doi.org/10.1016/j.gene.2006.03.028

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., & Cardona, A. (2012). Fiji: An open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. https://doi.org/10.1038/nmeth.2019

Schlager, S. (2017). Chapter 9 – Morpho and Rvcg – Shape analysis in R: R-packages for geometric morphometrics, shape analysis and surface manipulations. In G. Zheng, S. Li, & G. Székely (Eds.), *Statistical shape and deformation analysis* (pp. 217–256). Academic Press. https://doi.org/10.1016/B978-0-12-810,493-4.00011-0

Schmid, S., Bachmann Salvy, M., Garcia Jimenez, A., Bertrand, J. A. M., Cortesi, F., Heim, S., Huyghe, F., Litsios, G., Marcionetti, A., O'Donnell, J., Riginos, C., Tettamanti, V., & Salamin, N. (2023). *Amphiprion clarkii* population genomics. Genomic Observatories MetaDatabase (GEOME). https://n2t.net/ark:/21547/FbA2

Schmid, S., Micheli, B., Cortesi, F., Donati, G. F. A., & Salamin, N. (2022). Extensive hybridisation throughout clownfishes evolutionary history. *bioRxiv*. https://doi.org/10.1101/2022.07.08.499304

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., Blazier, J. C., Sankararaman, S., Andolfatto, P., Rosenthal, G. G., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 656–660. https://doi.org/10.1126/science.aar3684

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. https://doi.org/10.1890/1051-0761(2003)013[0159:PDDATS]2.0.CO;2

Shine, R., Brown, G. P., & Phillips, B. L. (2011). An evolutionary process that assembles phenotypes through space rather than through time. *Proceedings of the National Academy of Sciences*, *108*(14), 5708–5711. https://doi.org/10.1073/pnas.1018989108

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. https://doi.org/10.1371/journal.pone.0107610

Singhal, S., Hoskin, C. J., Couper, P., Potter, S., & Moritz, C. (2018). A framework for resolving cryptic species: A case study from the lizards of the Australian wet tropics. *Systematic Biology*, *67*(6), 1061–1075. https://doi.org/10.1093/sysbio/syy026

Sirkia, P. M., & Qvarnstrom, A. (2021). Adaptive coloration in pied flycatchers (*Ficedula hypoleuca*)-the devil is in the detail. *Ecology and Evolution*, *11*(4), 1501–1525. https://doi.org/10.1002/ece3.7048

Stankowski, S., Chase, M. A., Fuiten, A. M., Rodrigues, M. F., Ralph, P. L., & Streisfeld, M. A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of monkeyflowers. *PLoS Biology*, *17*(7), e3000391. https://doi.org/10.1371/journal.pbio.3000391

Stryjewski, K. F., & Sorenson, M. D. (2017). Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology & Evolution*, *1*(12), 1912. https://doi.org/10.1038/s41559-017-0364-7

Thresher, R. E., Colin, P. L., & Bell, L. J. (1989). Planktonic duration, distribution and population structure of western and central Pacific damselfishes (Pomacentridae). *Copeia*, *1989*(2), 420–434. https://doi.org/10.2307/1445439

Timm, J., Planes, S., & Kochzius, M. (2012). High similarity of genetic population structure in the false clown anemonefish (*Amphiprion ocellaris*) found in microsatellite and mitochondrial control region analysis. *Conservation Genetics*, *13*(3), 693–706. https://doi.org/10.1007/s10592-012-0318-1

Toole, J. M., Millard, R. C., Wang, Z., & Pu, S. (1990). Observations of the Pacific north equatorial current bifurcation at the Philippine coast. *Journal of Physical Oceanography*, *20*(2), 307–318. https://doi.org/10.1175/1520-0485(1990)020<0307:OOTPNE>2.0.CO;2

Turner, T. L., Hahn, M. W., & Nuzhdin, S. V. (2005). Genomic islands of speciation in *Anopheles gambiae*. *PLoS Biology*, *3*(9), e285. https://doi.org/10.1371/journal.pbio.0030285

Van Belleghem, S. M., Papa, R., Ortiz-Zuazaga, H., Hendrickx, F., Jiggins, C. D., Owen McMillan, W., & Counterman, B. A. (2018). Patternize: An R package for quantifying colour pattern variation. *Methods in Ecology and Evolution*, *9*(2), 390–398. https://doi.org/10.1111/2041-210X.12853

Van der Ven, R. M., Flot, J.-F., Buitrago-López, C., & Kochzius, M. (2021). Population genetics of the brooding coral *Seriatopora hystrix* reveals patterns of strong genetic differentiation in the Western Indian Ocean. *Heredity*, *126*(2), 351–365. https://doi.org/10.1038/s41437-020-00379-5

Vaux, F., Bohn, S., Hyde, J. R., & O'Malley, K. G. (2021). Adaptive markers distinguish north and South Pacific albacore amid low population differentiation. *Evolutionary Applications*, *14*(5), 1343–1364. https://doi.org/10.1111/eva.13202

Voris, H. K. (2000). Maps of Pleistocene sea levels in Southeast Asia: Shorelines, river systems and time durations. *Journal of Biogeography*, *27*(5), 1153–1167. https://doi.org/10.1046/j.1365-2699.2000.00489.x

Wang, N., Liang, B., Wang, J., Yeh, C.-F., Liu, Y., Liu, Y., Liang, W., Yao, C.-T., & Li, S.-H. (2016). Incipient speciation with gene flow on a continental Island: Species delimitation of the Hainan hwamei (*Leucodioptron canorum owstoni*, Passeriformes, Aves). *Molecular Phylogenetics and Evolution*, *102*, 62–73. https://doi.org/10.1016/j.ympev.2016.05.022

Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, *89*(5), 438–450. https://doi.org/10.1093/jhered/89.5.438

Wellington, G. M., & Victor, B. C. (1989). Planktonic larval duration of one hundred species of Pacific and Atlantic damselfishes (Pomacentridae). *Marine Biology*, *101*(4), 557–567. https://doi.org/10.1007/BF00541659

White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., & Toonen, R. J. (2010). Ocean currents help explain population genetic structure. *Proceedings of the Royal Society B: Biological Sciences*, *277*(1688), 1685–1694. https://doi.org/10.1098/rspb.2009.2214

Winters, K. L., van Herwerden, L., Choat, J. H., & Robertson, D. R. (2010). Phylogeography of the Indo-Pacific parrotfish *Scarus psittacus*: Isolation generates distinctive peripheral populations in two oceans. *Marine Biology*, *157*(8), 1679–1691. https://doi.org/10.1007/s00227-010-1442-4

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), 87. https://doi.org/10.1038/nrg.2016.133

Wood, C. W., & Brodie, E. D. (2016). Evolutionary response when selection and genetic variation covary across environments. *Ecology Letters*, *19*(10), 1189–1200. https://doi.org/10.1111/ele.12662

Ye, L., Yang, S.-Y., Zhu, X.-M., Liu, M., Lin, J.-Y., & Wu, K.-C. (2011). Effects of temperature on survival, development, growth and feeding of larvae of yellowtail clownfish *Amphiprion clarkii* (Pisces: Perciformes). *Acta Ecologica Sinica*, *31*(5), 241–245. https://doi.org/10.1016/j.chnaes.2011.06.003

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: Schmid, S., Bachmann Salvy, M., Garcia Jimenez, A., Bertrand, J. A. M., Cortesi, F., Heim, S., Huyghe, F., Litsios, G., Marcionetti, A., O'Donnell, J. L., Riginos, C., Tettamanti, V., & Salamin, N. (2024). Gene flow throughout the evolutionary history of a colour polymorphic and generalist clownfish. *Molecular Ecology*, *33*, e17436. https://doi.org/10.1111/mec.17436