Received: 26 January 2023 | Revised: 4 May 2023 | Accepted: 30 May 2023

DOI: 10.1111/mec.17047

FROM THE COVER

MOLECULAR ECOLOGY WILEY

Phenotypic and genomic dissection of colour pattern variation in a reef fish radiation

Floriane Coulmance[1,2,3] | Derya Akkaynak[4,5] | Yann Le Poul[6] | Marc P. Höppner[7] | W. Owen McMillan[3] | Oscar Puebla[1,2,3]

[1]Leibniz Center for Tropical Marine Research, Bremen, Germany

[2]Institute for Chemistry and Biology of the Marine Environment (ICBM), Oldenburg, Germany

[3]Smithsonian Tropical Research Institute (STRI), Panama, Republic of Panama

[4]Hatter Department of Marine Technologies, University of Haifa, Haifa, Israel

[5]Interuniversity Institute of Marine Sciences, Eilat, Israel

[6]Ludwig-Maximilians-Universität München, Munich, Germany

[7]Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany

Correspondence
Floriane Coulmance, Leibniz Center for Tropical Marine Research, Farenheitstraße 6, 28359 Bremen, Germany.
Email: floriane.coulmance-gayrard@leibniz-zmt.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: PU571/1-2; Smithsonian Institute for Biodiversity Genomics and Global Genome Initiative Grants Program; GEOMAR Helmholtz Centre for Ocean Research Kiel; The Future Ocean Cluster of Excellence

Handling Editor: Sean Rogers

Abstract

Coral reefs rank among the most diverse species assemblages on Earth. A particularly striking aspect of coral reef communities is the variety of colour patterns displayed by reef fishes. Colour pattern is known to play a central role in the ecology and evolution of reef fishes through, for example, signalling or camouflage. Nevertheless, colour pattern is a complex trait in reef fishes—actually a collection of traits—that is difficult to analyse in a quantitative and standardized way. This is the challenge that we address in this study using the hamlets (*Hypoplectrus* spp., Serranidae) as a model system. Our approach involves a custom underwater camera system to take orientation- and size-standardized photographs in situ, colour correction, alignment of the fish images with a combination of landmarks and Bézier curves, and principal component analysis on the colour value of each pixel of each aligned fish. This approach identifies the major colour pattern elements that contribute to phenotypic variation in the group. Furthermore, we complement the image analysis with whole-genome sequencing to run a multivariate genome-wide association study for colour pattern variation. This second layer of analysis reveals sharp association peaks along the hamlet genome for each colour pattern element and allows to characterize the phenotypic effect of the single nucleotide polymorphisms that are most strongly associated with colour pattern variation at each association peak. Our results suggest that the diversity of colour patterns displayed by the hamlets is generated by a modular genomic and phenotypic architecture.

KEYWORDS
colour pattern, genome-wide association, *Hypoplectrus*, modularity, radiation, reef fishes

1 | INTRODUCTION

The colour patterns displayed by reef fishes are among the most visually stunning traits in animals. Coral reefs are highly visual environments and colour pattern plays an important ecological and evolutionary role in reef fishes (Cott, 1940; Lorenz, 1962; Marshall, 1998; Marshall et al., 2019; Thayer, 1909). Colour pattern is, for example, involved in signalling (Cheney et al., 2009), camouflage (Phillips et al., 2017), mimicry (Randall, 2005) and mate choice (McMillan et al., 1999). Although the origin of coral reef fish families and functional groups goes back to the Palaeocene (66 million years ago), the majority of species arose within the last 5.3 million years, with sister species often differing primarily in terms of colour pattern (Bellwood et al., 2015, 2017). The analysis of reef fish

colour pattern provides therefore the opportunity to address a number of fundamental ecological and evolutionary processes (Salis et al., 2019) that include adaptation (Cortesi et al., 2015), speciation (Puebla et al., 2007) and adaptive radiation (Hench et al., 2022).

Nonetheless, colour pattern is a complex trait in reef fishes. It is actually a collection of traits rather than a single trait. Structurally, colour pattern involves all aspects of colour (e.g. hue, saturation, iridescence) as well as a diversity of patterns such as bars, stripes, bands, lines, gradients, speckles, blotches, spots and many more that can occur in many combinations. Developmentally, reef fishes present a diversity of pigment cell types (chromatophores) that goes well beyond the few types that are present in model species such as zebrafish (Parichy, 2021; Salis et al., 2019). Genetically, tens of patterning genes have been identified in fishes, but mostly in model fishes (Parichy, 2021; Salis et al., 2019), and we are only starting to unveil the genetic bases of colour pattern variation in reef fishes (Hench et al., 2019, 2022). These new insights have been facilitated by next-generation DNA sequencing technologies that now allow to routinely assemble and sequence entire fish genomes. Genotyping capacity has increased by six orders of magnitude, from about 10 loci with classic genetic markers to about 10 million loci with whole genomes, in less than two decades. As a consequence, progress starts to be hindered by phenotyping power, that is, the ability to describe phenotypic variation in detail, rather than genotyping power.

Reef fishes provide a good illustration of this situation. Studies of reef fish colour pattern have typically relied on manual scoring (e.g. Hemingson et al., 2020; Hench et al., 2022; Kelley et al., 2013), which is appropriate for discrete colour pattern elements (e.g. bars or eyespots). Nevertheless, manual scoring reduces colour pattern variation to a few predefined elements and is hardly applicable to complex and/or continuous variation. With recent advances in computer vision and deep learning, pre-trained convolutional neural networks provide an interesting avenue to analyse reef fish colour patterns. Nevertheless, variation in body shape can confound the analysis of colour pattern with this approach (Alfaro et al., 2019). An alternative avenue consists in aligning fish images and analysing the colour of each pixel of each aligned fish with principal component analysis (PCA). Hemingson et al. (2019) applied this approach to *Chaetodon* butterflyfishes reducing colour to four discrete categories ('yellow', 'black', 'white' and 'other'), and evidenced an effect of geographical range overlap and symmetry on colour pattern. Colour-standardized images would allow to analyse quantitatively the colour value of each pixel and thereby capture more subtle variation (e.g. gradients). Here, we develop an approach to do exactly that using the hamlets (*Hypoplectrus* spp., Serranidae) as a reef fish model system.

The hamlets are a genus of small predatory sea basses from the wider Caribbean. They are characterized by striking variation in colour pattern that delineates 18 recognized species (Lobel, 2011; Puebla et al., 2022). The hamlets are highly sympatric and very similar ecologically (Holt et al., 2008; Whiteman et al., 2007). Colour pattern is an important cue for mate choice and the hamlets show strong assortative mating both in the field and in experimental conditions (Barreto & McCartney, 2008; Domeier, 1994; Fischer, 1980; Puebla et al., 2012). Interspecific spawnings occur at low frequency in natural populations (<2% of observations). There are no apparent postzygotic barriers between species (Whiteman & Gage, 2007) and genetic data indicate that gene flow is ongoing among species (Hench et al., 2019, 2022). Colour pattern also likely plays an ecological role through aggressive mimicry and crypsis (Fischer, 1980; Puebla et al., 2007, 2018; Randall & Randall, 1960; Thresher, 1978). Colour pattern may therefore constitute a link between natural selection and reproductive isolation (Puebla et al., 2007), which would contribute to explain why the hamlets present the highest speciation rates documented in reef fishes (Siqueira et al., 2020). Nevertheless, colour pattern is a complex trait in the hamlets, involving all aspects of colour variation (e.g. hue, saturation) as well as a diversity of patterns (bars, dots, lines, spots, gradients, etc.). In addition, colour pattern is highly variable, not only among species but also within, both within and among locations (Aguilar-Perera, 2004; Robertson et al., 2019; Thresher, 1978), which complicates species delineation.

The hamlets are also very closely related genetically. They show low levels of genetic differentiation and divergence (Barreto & McCartney, 2008; Hench et al., 2022; Holt et al., 2011; Puebla et al., 2012, 2014) and do generally not sort into distinct mitochondrial haplogroups (Garcia-Machado et al., 2004; McCartney et al., 2003; Ramon et al., 2003). These low levels of genetic differentiation and divergence, combined with the diversity of colour patterns displayed by the group, provide a backdrop that is well suited for a genome-wide association study (GWAS) for colour pattern variation. Previous work has identified genetic associations with three colour pattern elements [vertical bars, saddle on the caudal peduncle and spot on the snout (Hench et al., 2022)]. Nevertheless, this analysis was limited to these three predefined traits and by the fact that they were scored manually.

The objective of this study is to characterize colour pattern variation in the *Hypoplectrus* radiation and its genomic bases. To this aim, we develop a protocol to take standardized photographs in situ, and generate a largely new dataset that includes both standardized photographs and whole genomes of the same individuals for 13 hamlet species. We analyse colour pattern variation quantitatively and with minimal human intervention, that is, without manual scoring of predefined colour pattern elements. We then use this quantified colour pattern variation as trait for a GWAS. We hypothesize that this approach will allow to identify the major colour pattern elements that contribute to phenotypic variation in the group, and to identify genetic associations with each colour pattern element. We expect to recover previously identified colour pattern elements such as vertical bars. We also expect to recover new colour pattern elements, colour elements in particular since colour varies continuously in the hamlets and could therefore not be scored manually in previous analyses.

2 | MATERIALS AND METHODS

2.1 | Fieldwork

Fieldwork was conducted at four locations in 2017 with the objective to collect standardized photographs and tissue samples that cover a large part of the hamlets colour pattern diversity and geographical range. This dataset includes 13 species and is largely new, with all images and 96 of the 113 genomes presented here for the first time. It notably includes five species (*H. floridae*, *H. gemma*, *H. guttavarius*, *H. chlorurus* and an undescribed species) as well as two locations (Florida and Puerto Rico) that were not considered in the previous manually scored GWAS (Hench et al., 2022). The four locations are the vicinity of the Keys Marine Laboratory in the Florida Keys (July 2017), of the Carrie Bow Cay field station in Belize (May 2017), of the Bocas del Toro field station in Panama (February 2017) and of the La Parguera field station in Puerto Rico (March 2017). All fieldwork was conducted under Smithsonian Tropical Research Institute (STRI) Institutional Animal Care and Use Committee (IACUC) protocol 2017-0101-2020-2 and Northeastern University IACUC protocol 17-0206R. Sampling was conducted under NOAA ONMS permit 2017-042 and FWCC permit SAL-17-1890A-SR (Florida), Fisheries Department permit 000026-17 (Belize), *Ministerio de Ambiente* permits SC/A-53-16 and SEX/A-35-17 and Access and Benefit-Sharing Clearing-House identifier ABSCH-IRCC-PA-241203-1 (Panama), and *Departamento de Recursos Naturales y Ambientales* research permit #2016-IC-127 (E) (Puerto Rico).

2.2 | Sampling procedure

Previous experience with the hamlets indicates that colour pattern is not well preserved in voucher specimens and often altered beyond the variation that is observed in the wild (Robertson et al., 2019) when hamlets are speared, manipulated, stressed for more than a few minutes, brought to the surface, taken out of the water, kept in captivity or euthanized (O. Puebla, personal observation). In this perspective, photographs of live individuals in their natural environment are ideal to capture colour pattern variation within and among species (e.g. Puebla et al., 2022; Robertson et al., 2019). Nevertheless, such photographs are not suited for quantitative analysis due to their lack of standardization in terms of size, colour and orientation. To address these limitations, we developed a protocol to take standardized photographs of live specimens in situ (Figure 1).

Briefly, the camera system consisted of a mirrorless Canon EOS M3 camera with an Ikelite underwater housing and two Ikelite DS51 strobes with white diffusers. A PVC frame held the camera at fixed distance and orientation from a neutral PVC background that included a colour checker (X-Rite Mini ColorChecker Classic) with a size standard. Hamlets were collected on scuba at depths ranging between 13 and 65 feet. This was done with hook-and-line, using small hooks whose barbels had been sanded to minimize injuries. Upon capture, the fish were transferred to a ziplock bag and held against the PVC background with an ID label while the other diver took the photographs.

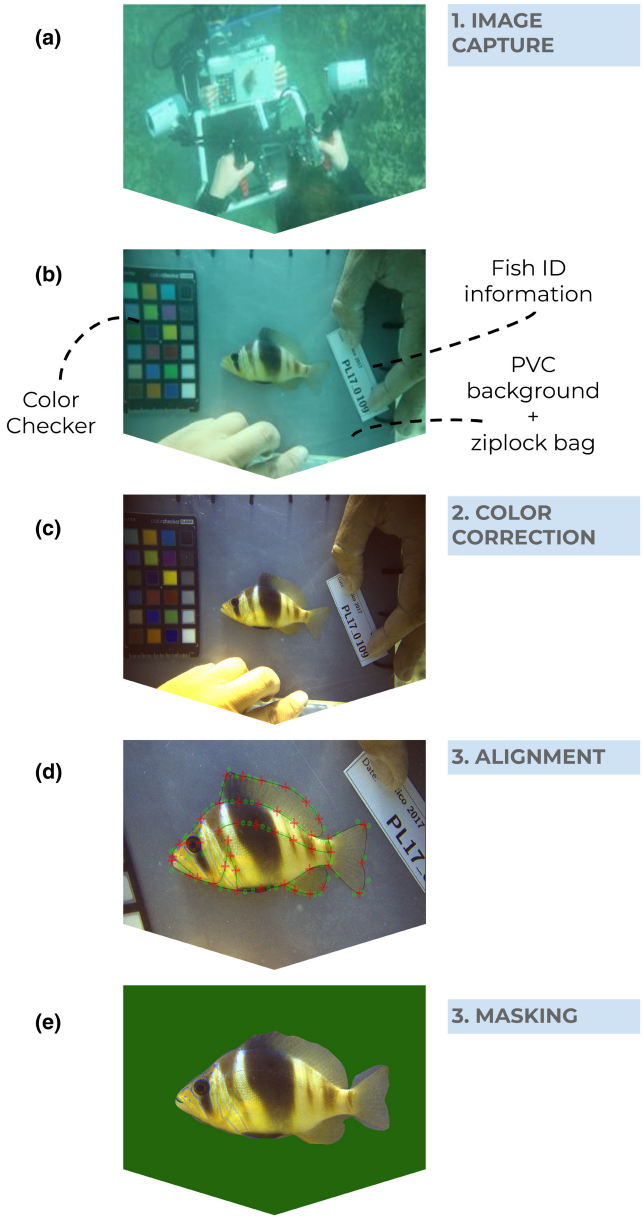

FIGURE 1 Overview of the image capture and standardization procedure. The fishes are captured in situ, photographed against a neutral background next to a colour checker and an ID label at a standardized distance and orientation from the camera lens. The fish images are then colour-standardized, aligned and masked.

The ziplock bag was then turned and the other side of the fish photographed. Photographs were initially taken with strobe lightning, and then both with and without strobe lightning as we started to suspect that strobe lightning may do more harm than good in terms of image analysis. All photographs were shot in *CR2* raw format since *jpg* format alters true colours irreversibly. A fin clip was taken and the fish was either released or collected as voucher. The entire operation took just a few (typically three) minutes per fish. Tissue samples were preserved in salt-saturated DMSO solution immediately after the dive. This method is efficient, highly selective and does not affect the reef (no chemicals or nets used). Photographs of a few selected individuals

taken before, during and after the sampling procedure indicate that colour pattern is not affected beyond the variation that is observed in the wild (Figure S1). Furthermore, observations of tagged individuals over 2 years indicate that this procedure does not noticeably alter their survival (O. Puebla, unpublished data).

2.3 | Image analysis

2.3.1 | Pre-processing

The image analysis was restricted to 113 samples for which we also sequenced full genomes (Figure 2, Table S1). The image analysis was also restricted to one photograph (the best one) of the left side of each fish, except for one sample (PL17-138uniflo) for which we only had the right side that we mirrored. The first step consisted in colour-correcting the images to standardize colours across the entire dataset. Raw *CR2* images were demosaiced with Adobe DNG converter and output as *dng* images for compatibility with MATLAB. Photographs were colour-corrected using the colour checker that was captured aside of the fish in each image and a MATLAB procedure derived from Akkaynak et al. (2014). The output of the colour correction was a new raster image (three-layer matrix) in *mat* format and a *png* image compressed in 8-bit. The second step consisted in creating a first rough mask to delimit the fish contour in all images. A MATLAB script was used to open the colour-corrected images and place outline landmarks around the fish body. This step resulted in a *mat* file containing the rough mask as a raster object for each fish.

2.3.2 | Alignment

Alignment of the fish images is necessary to ensure that each pixel position is as homologous as possible across all images. A first rough alignment was performed using the colour-corrected *png* images and the *mat* mask files to centre and orient the fish images. A more precise alignment was then conducted using a two-dimensional structure made of landmarks and Bézier curves following Le Poul et al. (2014). This structure was applied to each fish image, corrected manually when needed and used for the alignment of all fish images. Note that this procedure may involve slight deformations of the images. Aligned images were output in *png* format. To remove the background, an average image was created from the 113 aligned images. The outline of the fish body was drawn from the average image in Python. A *tif* mask was then created from the outline with the GNU Image Manipulation Program (GIMP). This resulted in a mask in *tiff* format in which the mask is white and the background black. The mask was applied to each fish image to remove the background. The images were then very slightly blurred with a 5-by-5 pixel convolution to remove fine-scale texture, and converted to the LAB format where the L channel contains luminosity information and the A and B channels colour information. Finally, each image was flattened into a long one-dimensional image vector to be used as input for PCA.

2.3.3 | PCA on colour pattern variation

PCA is an appropriate tool for the analysis of colour pattern because it allows to dissect colour pattern variation into a number of dimensions (principal components, PCs) that can be analysed either individually or altogether. PCA was conducted in python using the *sklearn* PCA function and limiting the number of PCs to 15. To visualize the different elements of colour pattern variation, heatmaps that display the contribution of each pixel to the variation of each PC were created using the PC eigenvectors.

2.4 | Genotyping

A total of 96 samples were sequenced anew for this study. In all, 17 additional samples were available from Moran et al. (2019), one from Hench et al. (2019) and two from Puebla et al. (2022). For the new samples, genomic DNA was extracted from tissue samples using Qiagen MagAttract High Molecular Weight kits. Libraries were prepared and sequenced at the Institute of Clinical Molecular Biology (Kiel University) on an Illumina NovaSeq 6000 at a mean raw sequencing depth of 27×. Four samples with a mean raw coverage <15× were removed from the analysis. After this filtering, our study consisted of 113 samples spanning 13 species and four locations. A list of all samples with metadata and accession numbers is provided in Table S1 and a summary of the sampling design is presented Figure 2.

All the samples considered in this study were genotyped jointly and anew. The variant calling procedure was adapted from the best practice recommendations for the GATK workflow (McKenna et al., 2010) provided by the Broad Institute (DePristo et al., 2011; Van der Auwera et al., 2013). GATK was used to transform the sequences from *fastq* to *uBAM* format, assign read groups and mark adapters. The sequences were then back-transformed to *fastq* format using GATK, mapped to the hamlet reference genome (Hench et al., 2019) using BWA (Li & Durbin, 2009) and merged with the *uBAM* files containing read groups information with GATK. Duplicated reads were removed, genotype likelihoods were called for each individual and merged for all samples. All individuals were then genotyped jointly based on the genotype likelihoods from all samples. The data set was filtered for a minor allele count $\geqslant 2$ and reduced to bialleic SNPs only using VCFtools (Danecek et al., 2011).

2.5 | Population genetic analyses

2.5.1 | Genetic structure

Genetic differentiation (F_{ST}) between pairs of samples (species/populations) was computed with VCFtools following Weir & Cockerham (Weir & Cockerham, 1984), considering the genome-wide weighted mean. This analysis was restricted to the species/populations for

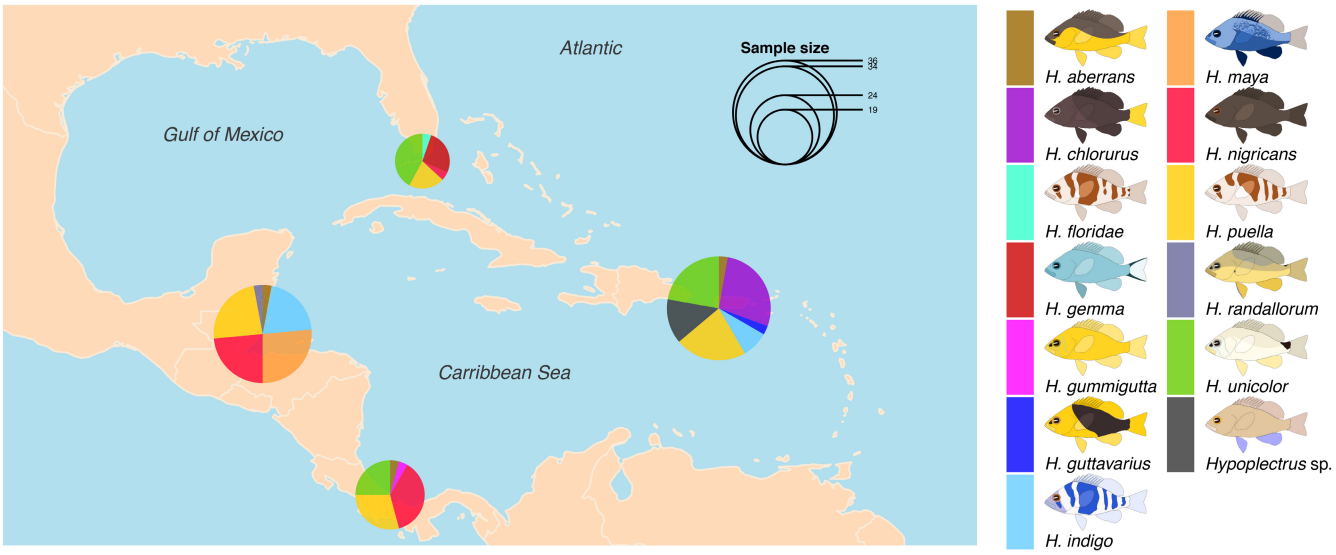

FIGURE 2 Sampling design. The dataset consisted of 113 samples spanning 13 species and 4 locations (Panama, Belize, Puerto Rico and Florida).

which we had at least seven samples. The same approach was used to estimate the joint F_{ST} among all species/populations for which we had at least seven samples.

A PCA was run on all samples using the R package SNPRelate (Zheng et al., 2012) to also explore genetic structure at the individual level. Linkage Disequilibrium (LD) among SNPs was limited by pruning SNPs with a LD threshold of 0.2 using a recursive sliding-window approach. The LD-pruned dataset consisted of 289,515 SNPs.

2.5.2 | Identification of hybrids and backcrosses

To identify potential hybrids and backcrosses in our dataset, we used the approach based on Mendelian inheritance of highly differentiated loci implemented in NewHybrids (Anderson & Thompson, 2002). This analysis was restricted to the species/populations for which we had at least six samples. For each pair of sympatric species, 800 of the most differentiated SNPs were selected and then filtered for a minimum physical distance of 5 kb with VCFtools to reduce physical linkage among them. In all, 80 SNPs were then randomly selected from the filtered set using bash scripting to ensure that all analyses are based on the same number of markers, and converted to the NewHybrids input format using PGDSpider (Lischer & Excoffier, 2012). The assignment to hybrid classes with NewHybrids was implemented in the R package parallelnewhybrid (Wringe et al., 2016), which was run with a burn-in of 10^6 iterations and 10^7 sweeps. Individuals that were assigned to one hybrid class (first-generation hybrid, second-generation hybrid or backcross) with a posterior probability >0.99 were considered high-probability hybrids or backcrosses.

2.6 | Genome-wide association study

To identify genomic regions associated with colour pattern variation, a multivariate GWAS was implemented using the PCs of the colour pattern PCA as traits. Specifically, the multivariate association test implemented in *PLINK* (Purcell et al., 2007) was used (Ferreira & Purcell, 2009) with the first five PCs of the colour pattern PCA as traits (considering more than five PCs did neither reveal additional association peaks nor noticeably change the multivariate GWAS results). This multivariate approach provides more statistical power than a univariate method that would consider each PC individually (Ferreira & Purcell, 2009). Nevertheless, it has the drawback of not accounting for population structure. Furthermore, it is also relevant to consider the associations of individuals PCs to single out the genetic associations for specific colour pattern elements. The GWAS was therefore repeated with (i) the univariate approach implemented in *PLINK* that does account for population structure, (ii) the univariate approach implemented in *GEMMA* (Zhou & Stephens, 2012) that also accounts for population structure and (iii) the multivariate approach used above but considering each PC individually.

Multivariate and univariate associations were first considered along the genome using the mean association −log (*p*-value) over 50-kb windows with 5-kb increments, which revealed a number of sharp association peaks. The strongest multivariate association peaks were identified using an arbitrary mean association −log (*p*-value) cut-off value of 2.5.

The strongest association peaks were then examined in more detail considering individual SNPs and the reference genome annotation. For each association peak, the multivariate phenotypic effect of the SNP showing the strongest association was characterized with a heatmap displaying the weighted effect of the five PCs on this SNP for each pixel using the PC eigenvectors.

3 | RESULTS

3.1 | Image analysis

The first five PCs of the colour pattern PCA explained 85.7% of the variation among the standardized images, and the PCs beyond PC5 accounted for less than 3% of variation each (Figure S2). The PCA showing PC1 versus PC5 is presented in Figure 3, the PCA showing PC2 versus PC4 is presented in Figure 4, and the PCA showing PC1 versus PC3 is presented in Figure S3.

PC1 accounted for 58.0% of the variation and broadly discriminated samples from the darkest coloured ones (*H. nigricans*, on the left of the PCA) to the lightest coloured ones (*H. unicolor*, on the right, Figure 3). The heatmap indicated that the pixels whose variation contributed most to PC1 were located on the ventral part and on the caudal and anal fins (Figure 3).

PC5 accounted for only 3.5% of the variation but broadly discriminated the samples that have vertical bars (*H. puella*, *H. indigo* and *H. floridae*, in the upper half of the PCA) from the samples that do not have vertical bars (in the lower half, Figure 3). In agreement with this pattern, the heatmap indicated that the pixels whose variation contributed most to PC5 were located on the vertical bars and caudal peduncle mark (Figure 3). Thus, the heatmaps for PC1 (Figure 3) and PC5 (Figure 3) were broadly the negative of each other.

PC4 accounted for 5.6% of the variation and broadly discriminated the samples that have a plain blue-coloured face and body (*H. maya* and *H. gemma*, in the upper half of the PCA) from the other samples (in the lower half, Figure 4). In agreement with this pattern, the heatmap indicated that the pixels whose variation contributed most to PC4 were uniformly distributed over the face and body (Figure 4).

Patterns were more diffuse for PC2 although it accounted for 10.6% of the variation. This PC broadly discriminated the samples that have a yellow colour component, in particular a yellow caudal fin (*H. guttavarius*, *H. chlorurus*, *H. aberrans* and *H. gummigutta*, on the right of the PCA) from the other samples (on the left, Figure 4). In agreement with this pattern, the heatmap indicated that the pixels whose variation contributed most to PC2 were located on the caudal fin (Figure 4).

PC3 accounted for 8.0% of the variation but did not appear to discriminate any colour pattern element in particular (Figure S3). Examination of the photographs indicated that the images that are in the upper part of the PCA in Figure S3 were characterized by over-exposure of the ventral anterior part of the body due to strobe lightning, which was used for 59 of the 113 images (52%). This is reflected in the heatmap for PC3 (Figure S3), which highlights this part of the body. PC3 appears therefore to capture variation that is in large part artefactual.

3.2 | Population genetic analyses

3.2.1 | Genetic structure

Genetic structure was low to moderate, with pairwise F_{ST} estimates ranging between 0.002 (*H. chlorurus* vs. *H. puella* in Puerto Rico) and 0.097 (*H. nigricans* from Panama vs. *H. indigo* from Belize, Table S2), and a joint F_{ST} estimate among all species/populations of 0.034. It is to be noted that these results are based on relatively small samples sizes (between 7 and 10 samples per species/population), which is expected to result in relatively low precision of F_{ST} estimates.

PC1 of the whole-genome PCA accounted for 1.5% of the genetic variation and discriminated the only *H. floridae* sample from the rest of the samples (Figure S4a,b). PC 2 accounted for 1.3% of the genetic variation and discriminated *H. indigo*, one *H. puella* sample from Florida, and the *H. floridae* sample. PC3 accounted for 1.1% of the genetic variation and discriminated the samples from Puerto Rico, the *H. nigricans* samples and the *H. gemma* samples.

3.2.2 | Identification of hybrids and backcrosses

One individual out of the 80 included in the NewHybrids analyses (1.25%) was identified as a high-probability (posterior probability >0.99) backcross. This proportion is broadly in line with the low occurrence (<2%) of interspecific spawnings reported in wild populations, including the populations from Florida, Belize and Bocas del Toro considered in this study (Barreto & McCartney, 2008; Fischer, 1980; Puebla et al., 2007, 2012). This sample was a *H. chlorurus* individual from Puerto Rico (PL_17_40), which had a typical *H. chlorurus* phenotype but came out as a *H. unicolor*/*H. chlorurus* backcross (Figure S5).

Two other samples provided notable results. The first one was a *H. puella* individual from Belize (PL_17_107), which had a typical *H. puella* colour pattern phenotype but came out as a hybrid (posterior probability 0.20) or backcross (posterior probability 0.80) between *H. puella* and *H. nigricans*. The second sample was a *H. puella* individual from Puerto Rico (PL17_17_53), which came out as a *H. chlorurus*/*H. puella* backcross with a posterior probability of 0.92. This individual was atypical, with vertical bars but a dark body and a bright yellow caudal fin, that is, a phenotypic intermediate between *H. chlorurus* and *H. puella*. These two samples are considered possible hybrids or backcrosses, although not high-probability ones since the assignment posterior probability to any category is <0.99%.

3.3 | Genome-wide association study

PCs 1 and 5 showed sharp univariate association peaks along the genome (Figure 3). PC5 (which is associated with vertical bars and caudal peduncle mark) presented a particularly clear pattern, with an outstanding association peak on Linkage Group (LG, putative chromosome) 12 (Figure 3). PC1 (which is associated with dark to clear ventral part, caudal fin and anal fin) showed more association peaks than PC5, distributed over several LGs (LG02, LG03, LG04, LG08, LG09, LG12, LG23), with a major peak on LG04 (Figure 3). Association peaks were lower for PC4 (which is associated with plain blue face and body) but a few peaks emerged, notably on LG04, LG08 and LG12 (Figure 4). PC2 (which is associated with

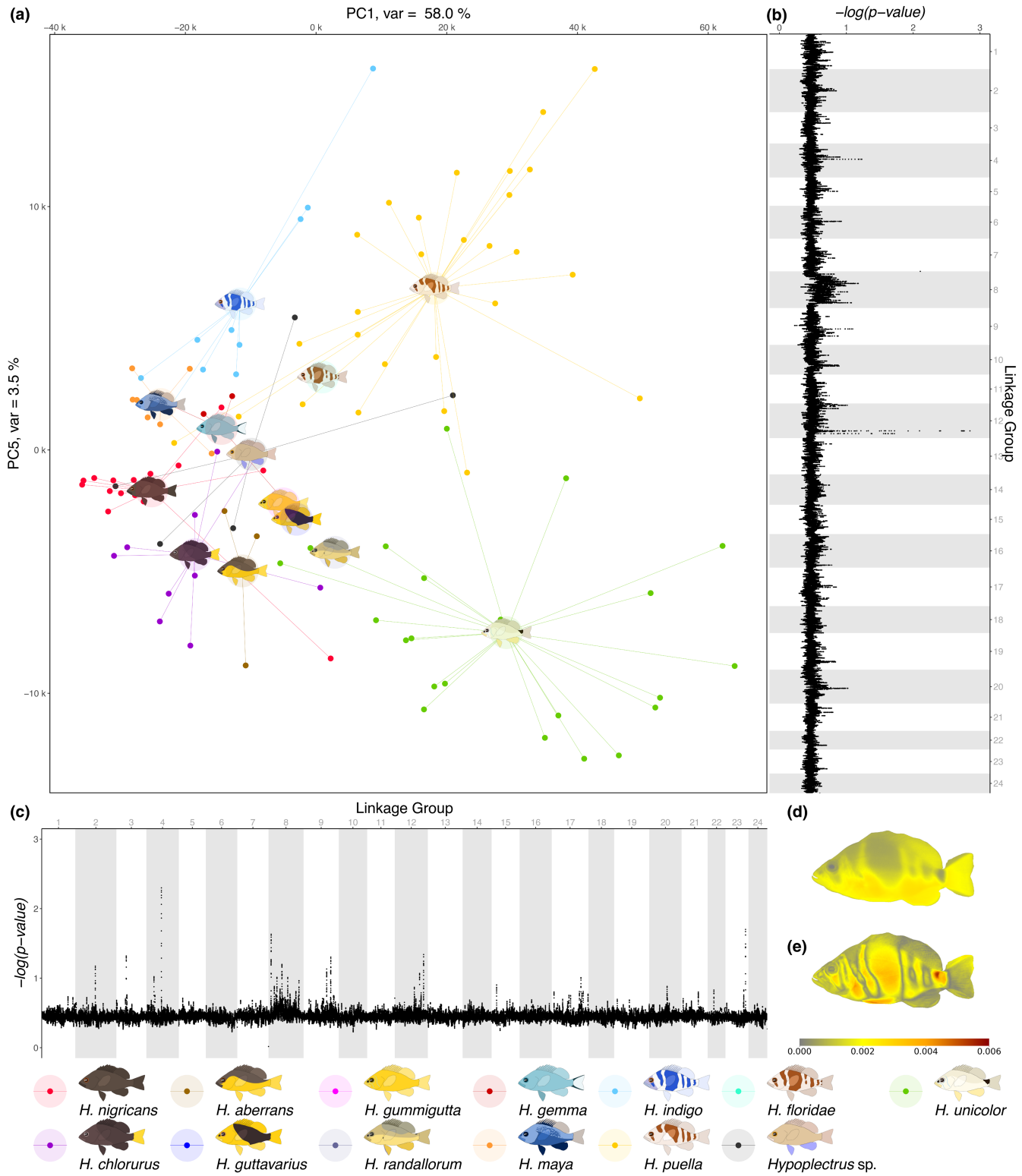

FIGURE 3 (a) Principal component (PC) 1 versus PC5 of the colour pattern PCA. (b) Association −log (*p*-value) of the univariate GWAS for PC5 along the hamlet genome. (c) Association −log (*p*-value) of the univariate GWAS for PC1 along the hamlet genome. (d) Heatmap showing the contribution of each pixel to PC1. (e) Heatmap showing the contribution of each pixel to PC5.

yellow caudal fin) showed a large number of relatively low association peaks distributed all over the genome (Figure 4). PC3 (which appears to capture artefactual variation) showed the fewest and lowest association peaks (Figure S3). The most notable one was in a region of LG06 where a small inversion had been previously identified (K. Hench & O. Puebla, unpublished). Higher levels of association were observed in LG08 for most PCs, particularly for PC2. This is likely due to a large (>20 Mb) low-recombining region—presumably

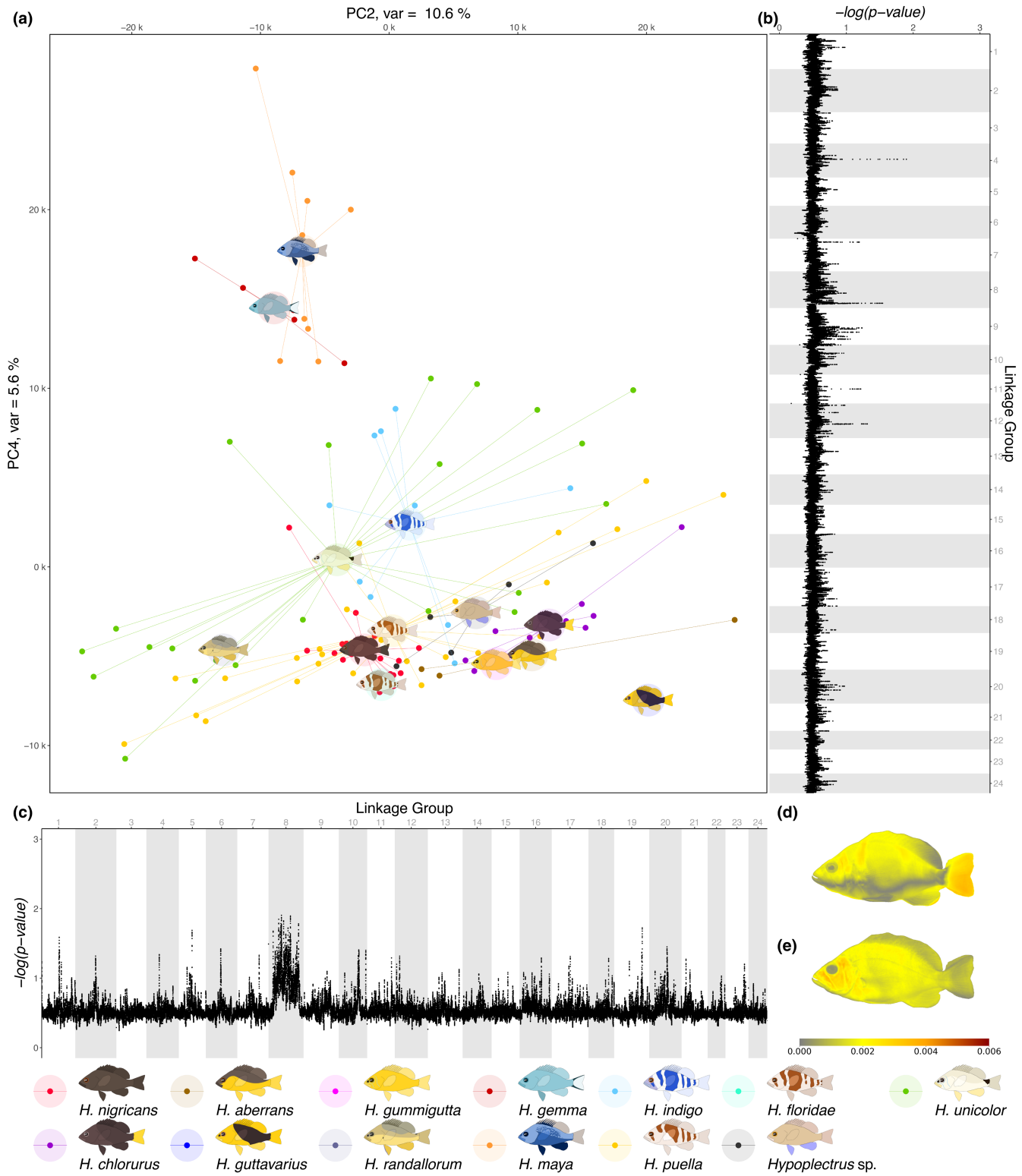

FIGURE 4 (a) Principal component (PC) 2 versus PC4 of the colour pattern PCA. (b) Association −log (*p*-value) of the univariate GWAS for PC4 along the hamlet genome. (c) association −log (*p*-value) of the univariate GWAS for PC2 along the hamlet genome. (d) Heatmap showing the contribution of each pixel to PC2. (e) Heatmap showing the contribution of each pixel to PC4.

an inversion—that has been previously identified in this LG (Hench et al., 2019; Theodosiou et al., 2016).

The multivariate GWAS considering the first five PCs altogether retrieved the same association peaks that were identified by the univariate analyses, and the same higher levels of association in the low-recombining region of LG08 (Figure 5a). When considering each PC individually, the multivariate approach provided similar association patterns as the univariate analyses (Figure S6).

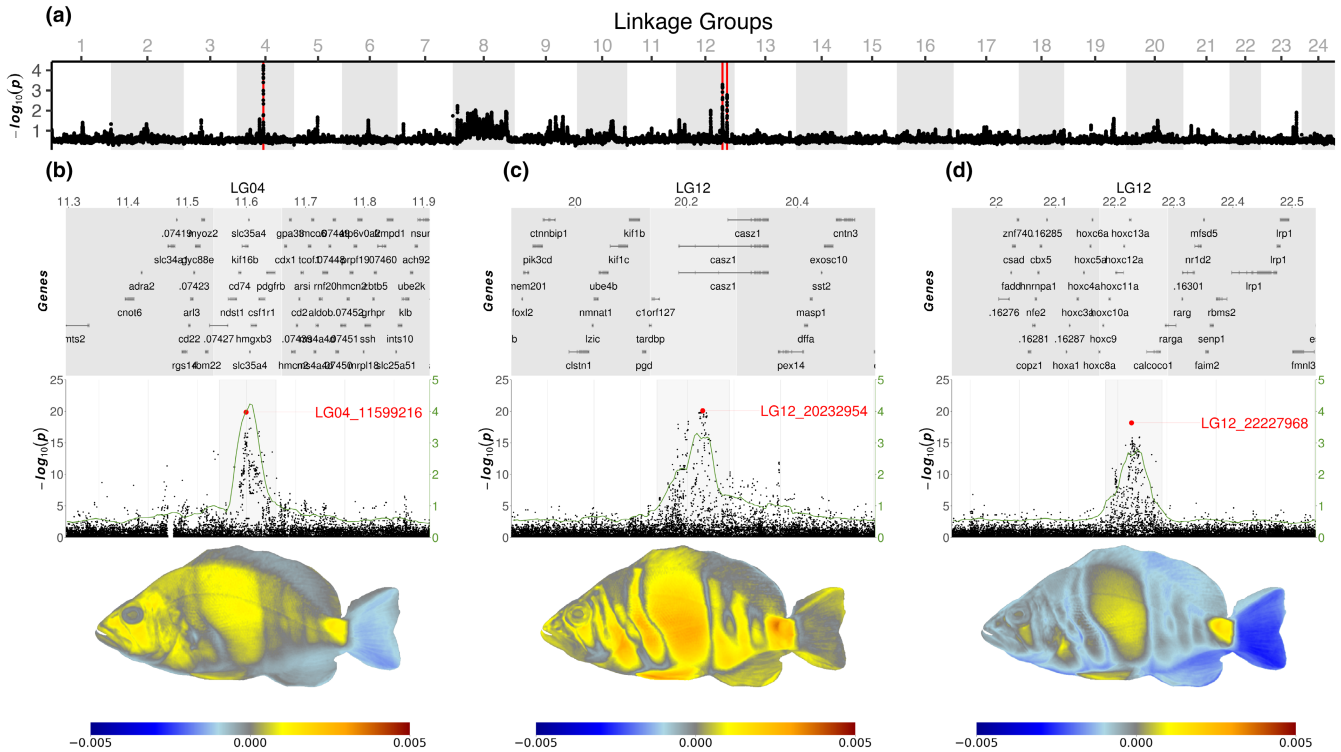

FIGURE 5 Results of the multivariate GWAS on the first five PCs of the PCA on colour pattern variation. (a) Manhattan plot showing the association −log (*p*-value) along the hamlet genome. The three major association peaks are highlighted in red. (b–d) Close-ups on the three major association peaks (in the same order from left to right as in the Manhattan plot). Each panel shows, from top to bottom, the position along the genome (in Mb), the gene annotation, the association −log (*p*-value) on a SNP-basis (most strongly associated SNP highlighted and annotated in red, scale on the left axis) and smoothed over 50-kb windows (green line, scale on the right axis), and a fish heatmap showing the phenotypic effect of the most strongly associated SNP. Note that the direction of the association (positive or negative) is arbitrary for each plot (it depends on which allele was considered the reference allele).

Three major association peaks stood out in the multivariate GWAS. A close-up on these peaks allows to explore association patterns in relation to the gene annotation, identify the SNP that is most strongly associated with colour pattern variation for each peak and characterize the multivariate phenotypic effect of this SNP (Figure 5b–d).

The association peak on LG04 is located in a gene-dense region. The most strongly associated SNP is upstream of the Solute Carrier Family 35 Member A4 (*slc35a4*) gene. It shows a positive association with most of the body, notably the head, the central bar and the saddle on the caudal peduncle, and a negative association with the caudal fin (Figure 5b). It is to be noted that the direction of the association for each SNP—positive or negative—is arbitrary (it depends on which allele was considered the reference allele).

The first association peak on LG12 is located in an intron of *casz1*, a castor zinc finger transcription factor. The most strongly associated SNP is associated with the seven vertical bars and, to a lesser extent, with the anterior part of the dorsal fin, the anal fin and the margins of the caudal fin (Figure 5c).

The second association peak on LG12 is located on the *hoxca* gene cluster and the most strongly associated SNP is in an exon of the *hoxc13a* gene. It is positively associated with the saddle on the caudal peduncle, the central bar and the eye bar, and negatively associated with a large part of the rest of the body, the caudal fin in particular (Figure 5d).

Smaller association peaks identify a number of additional genes (*atp13a3*, *znf711*, *glp1r*, *elovl5*, *sox10*, *smox*, *pard6b*, *fmod*, *tango6*, *tmem245*, *nxpe3*, *sers1*, *vit*, *matk*, *sin3b*, *naaladl2*, *tgm1*, *kit*), some of which are known to play a role in pigmentation (e.g. *sox10*, *smox*). These are not analysed in detail due to the limited power provided by our sample size to characterize small-effect loci.

4 | DISCUSSION

PCA of the standardized images dissects colour pattern variation into broadly interpretable colour patterns elements: dark to light ventral part, caudal and anal fin on PC1, yellow caudal fin on PC2, blue face and body on PC4, and vertical bars and caudal peduncle mark on PC5. As expected, we recover previously identified pattern elements (vertical bars and caudal peduncle mark), but also new specific colour elements that could not be scored manually in previous analyses. The univariate GWAS allows then to dissect the genetic associations with each colour pattern element independently. The colour pattern elements identified differ in terms of what aspect of colour pattern variation they capture, the proportion of colour

pattern variation they explain, and also in terms of genetic associations. For example, PC5 explains only 3.5% of colour pattern variation but is associated with discrete and specific colour pattern elements (vertical bars and caudal peduncle mark), and presents one outstanding association peak on LG12. On the other hand, PC1 explains 58% of colour pattern variation, is associated with a more continuous colour pattern element (dark to light ventral part, caudal fin and anal fin), and shows several associations peaks on different LGs.

The colour pattern elements identified are independent from each other from an image analysis perspective, but this does not necessarily imply that they are independent from a genetic perspective. On the contrary, the observation that the same association peaks emerge for different PCs (e.g. on LG04 for PC1 and PC4) suggests that some genomic regions may play a role in several colour pattern elements. This is where the multivariate GWAS is relevant because it allows to explore genetic associations—and characterize the phenotypic effect of the most strongly associated SNPs—across all PCs.

The multivariate GWAS identifies three major association peaks on LG04 and LG12. It is to be noted that although the two major peaks on LG12 appear close to each other on a whole-genome scale, they are separated by two megabases, which is well beyond physical linkage distance in the hamlets (Hench et al., 2019; Moran et al., 2019). Thus, these two association peaks are not physically linked. The most strongly associated SNP in each genomic region show strong associations (in the order of 1e-20) and distinct phenotypic effects.

The first association peak on LG12 is particularly striking as it identifies the seven vertical bars that are characteristic of the barred hamlet (*H. puella*), the Florida hamlet (*H. floridae*) and the indigo hamlet (*H. indigo*) in our dataset. This result is consistent with the univariate analysis for PC5, which captures the vertical bars and reveals the same outstanding association peak on LG12. This result is also consistent with our previous GWAS (based on a different dataset) where the presence/absence of vertical bars was scored manually, revealing here again the same outstanding association peak on LG12 (Hench et al., 2022). This validates both this specific result for the vertical bars and our new GWAS approach more generally. The fundamental difference is that the vertical bars now emerge from the analysis without having been identified a priori and scored manually, and represent just one among several colour pattern elements identified. Indeed, although the vertical bars are a conspicuous element of colour pattern variation in the hamlets and show an outstanding association peak, they represent only 3.5% of variation in the colour pattern PCA. In terms of gene annotation, this association peak points to a castor zinc finger transcription factor (*casz1*). Interestingly, this gene has been shown to be expressed in the photoreceptors of mice (Mattar et al., 2015, 2018) and zebrafish (Ogawa & Corbo, 2021), and is also consistently expressed in the hamlet retina (Hench et al., 2019). This suggests a possible pleiotropic effect of this locus on vision and pigmentation, or tight physical coupling between vision and pigmentation loci as in *Heliconius* butterflies (Rossi et al., 2020). Either of these situations would contribute to explain the explosive radiation of the hamlets.

The second association peak on LG12 identifies, among others, the saddle on the caudal peduncle that is characteristic of the butter hamlet (*H. unicolor*). This result is consistent with the univariate analyses for PC1 and PC5, which discriminate the butter hamlet from the other species and reveal the same association peak on LG12. This result is also consistent with our previous GWAS analysis where we had scored the presence/absence of the saddle on the caudal peduncle manually, revealing here again the same association peak on LG12 (Hench et al., 2022). Nevertheless, the results are less clear-cut for the saddle on the caudal peduncle than for vertical bars because other colour pattern elements are also identified by this association peak (eye bar, central bar, caudal fin), and because other association peaks also emerged when the saddle on the caudal peduncle was scored manually (Hench et al., 2022). Furthermore, the saddle on the caudal peduncle is characteristic of the butter hamlet, and although it is polymorphic within this species (O. Puebla, personal observation), this variation was not captured in our samples. All butter hamlets in our dataset have a saddle on the caudal peduncle and none of the other samples have one (note that the saddle on the caudal peduncle differs from the most posterior vertical bar in terms of shape and its melanic nature). We therefore have limited power to disentangle the genetic associations with this specific colour pattern element from other colour pattern elements that are characteristic of the butter hamlet. In terms of gene annotation, the second association peak on LG12 points to *hoxc13a*, another transcription factor. Hox genes are well known for their role in the patterning of tissues along the body axis and their 3′–5′ organization in tight genomic clusters that reflects their anterior to posterior expression during development (Carroll et al., 2005). They have also been shown to be involved in pigmentation, notably in *Drosophila* (Jeong et al., 2006), Nymphalidae butterflies (Saenko et al., 2011), bumble bees (Tian et al., 2019) and crows (Poelstra et al., 2015). *hoxc13a* in particular is the most 5′ gene of the hoxca gene cluster and has been shown to be expressed at the caudal bud and pigment appearance stages in fishes (Jakovlić & Wang, 2016; Thummel et al., 2004). This is consistent with the possibility that *hoxc13a* may be involved in the development of the saddle on the caudal peduncle in the hamlets. Functional analyses are needed to test this hypothesis.

The association peak on LG04 is the strongest multivariate association peak. It is also the most pervasive association peak as it emerges in the univariate analyses for PC1, PC2, PC4 and PC5. This genomic region appears therefore to play a prominent role in colour pattern variation. In the univariate analyses, this association peak is strongest for PC1, which is associated with light to dark colour. It is also the strongest association peak for PC4 that discriminates the Maya hamlet (*H. maya*) and the blue hamlet (*H. gemma*), the two species with a plain blue face and body, from the other samples. This genomic region, which is gene dense, appears therefore to be involved in body coloration. Accordingly, the phenotypic effect of the most strongly associated SNP recovers most of the body. Yet, it also recovers specific colour pattern elements (saddle on the caudal peduncle, central bar, spot on the

snout), suggesting that this genomic region may also be involved in patterning.

The fact that the low-recombining region on LG08 presents noticeably higher levels of association for most PCs suggests that this region may also be involved in colour patterning. This is particularly true for PC2, which shows the highest associations on LG08. Chromosomal inversions have been shown to be associated with colour pattern variation in a variety of taxa including *Drosophila* (Hatadani et al., 2004), horses (Brooks et al., 2007), stick insects (Lindtke et al., 2017), redpoll finches (Funk et al., 2021) and deer mice (Hager et al., 2022). Furthermore, the low-recombining region on LG08 is extensive (about 20 Mb, 75% of the LG) and distinct association peaks can be identified within this region. This suggests that it may act as a supergene that captures specific combinations of alleles at several loci associated with different colour pattern elements, as reported in *Heliconius* butterflies (Jay et al., 2022). Nevertheless, larger sample sizes will be needed to identify the specific colour pattern elements associated with the different peaks on LG08 since the signal is weaker than the three major peaks in the multivariate GWAS. Furthermore, extensive analyses on both previous and this dataset failed to identify distinct groups corresponding to the non-inverted homozygotes, non-inverted/inverted heterozygotes and inverted homozygotes at the putative inversion on LG08 (K. Hench, F. Coulmance, unpublished). This suggests that if the low-recombining region is indeed an inversion (or a cluster of several inversions), then it is a recent one.

4.1 | Phenotypic and genetic modularity

A striking aspect of the colour pattern elements identified by the PCA (colour of the face, body, ventral part, caudal fin and anal fin, vertical bars and saddle on the caudal peduncle) is that they constitute the basic 'building blocks' of colour pattern variation in the whole radiation. Indeed, the colour patterns displayed by the different species are broadly different combinations of these basic elements. Furthermore, the GWAS indicates that each colour pattern element is associated with a relatively small number of large-effect loci. This suggests a modular (or combinatorial; Marques et al., 2019) genetic and phenotypic architecture, whereby the variety of colour patterns displayed by the hamlets is generated by different combination of alleles at these loci. In addition, the high levels of sympatry of the hamlets and the occurrence of hybridization and introgression in the group (Barreto & McCartney, 2008; Domeier, 1994; Fischer, 1980; Hench et al., 2019, 2022; Puebla et al., 2012; this study) can contribute to generate new phenotypic diversity rapidly. This would provide a mechanism for the exceptionally rapid radiation of the hamlets, as suggested for other radiations in tropical butterflies and finches where colour pattern also plays a prominent role (Campagna et al., 2017; Stryjewski & Sorenson, 2017; Van Belleghem et al., 2017). Such a modular genetic architecture appears to also involve pleiotropic effects in the hamlets since, as also documented in *Heliconius* butterflies (Morris et al., 2019), some genomic regions are associated with several colour pattern elements.

4.2 | Advantages and limitations

The major advantage of our approach is that it considers colour pattern variation in its entirety, without a priori identification of colour pattern elements, manual scoring or clustering of pixel colour values into discrete colour categories. Furthermore, it generates results that are interpretable from both a colour pattern and a genetic perspective. The PCA dissects colour pattern variation into its major elements, and the univariate GWAS identifies associations with each element independently. The multivariate GWAS then identifies associations across all these elements and allows to characterize the phenotypic effect of the most strongly associated SNPs.

Nevertheless, our study remains limited by several factors. While our sample size of 113 allows to identify major colour pattern elements and large-effect loci, it is too low to characterize more subtle elements of colour pattern variation and small-effect loci. Such small-effect loci appear to also contribute to colour pattern variation, as suggested by a number of smaller association peaks throughout the genome and the signal at the low-recombining region of LG08. A larger sample size is warranted to achieve higher resolution.

Furthermore, the multivariate GWAS implemented here does not account for population structure. Yet, the consistency of the results with the univariate GWAS, which does account for population structure, suggests that our results are not strongly affected by population structure. This may result from (i) the relatively low levels of genetic structure within hamlets species (Puebla et al., 2008, 2009, this study), (ii) the exceptionally low levels of genetic differentiation among hamlet species (Barreto & McCartney, 2008; Hench et al., 2022; Holt et al., 2011; Puebla et al., 2012, 2014; this study) and (iii) the fact that most major colour pattern elements (dark to light ventral part, caudal fin and anal fin, yellow caudal fin, blue face and body, vertical bars) were captured in different genomic backgrounds, that is, in different species and/or locations, by our sampling design. In this regard, we note that the major reason why we could not correct for genetic structure in the multivariate GWAS is sample size. Thus, a higher sample size would not only allow to identify more subtle colour pattern elements and small-effect loci, but also correct for genetic structure in the multivariate GWAS.

Another limitation of our study is the artefactual variation introduced by strobe lighting, which appears to generate about 8% of the variation in the image analysis. Strobe lighting was meant to standardize images by providing homogeneous and constant illumination, but it ultimately provided inconsistent lighting, even among repeated photographs of the same fish taken seconds apart. The strobes also generated under-exposure, over-exposure and reflection artefacts. Considering that light was not strongly limiting at the depths at which the pictures were taken and that the colour standard allows to correct for differences in light conditions between

photographs, we advise to not use strobes. Exploration of our data with strobed versus non-strobed images suggests that species should cluster better in the colour pattern PCA when non-strobed images only are used.

Our approach is also limited to the visible spectrum, while it is known that patterns outside of the visible spectrum, for example, in the UV (Mitchell et al., 2022; Siebeck et al., 2010), can be important in some reef fishes. This limitation may be addressed using a hyperspectral camera instead of a commercial off-the-shelf camera and underwater housing.

Finally, it is important to note that our approach is meant to describe colour pattern variation per se, as opposed to how colour patterns are perceived by, for example, conspecifics, congeners, predators or preys, which requires visual modelling (e.g. Pierotti et al., 2020). In this regard, we note that the GWAS did not identify associations with short- and long-wave sensitive opsin genes (*sws2aα*, *sws2aβ*, *sws2b* and *lws*), which are characterized by a sharp peak of differentiation between species on LG17 (Hench et al., 2019). This confirms that our approach singles out the effect of colour pattern from other traits that may also differ among species.

4.3 | Perspectives

This study constitutes a proof-of-concept for the quantitative analysis of colour pattern variation in reef fishes. The results illustrate the potential of our approach to dissect colour pattern variation into interpretable colour pattern elements and detect genetic associations with these colour pattern elements, both independently with a univariate GWAS and altogether with a multivariate GWAS. The method may by further refined using larger sample sizes, non-strobed images only, a multivariate GWAS that accounts for population genetic structure or a hyperspectral camera. These improvements provide potential to identify and characterize subtle colour pattern variation, not only among species but also within, providing a quantitative basis for study of the eco-evolutionary significance of colour pattern in reef fishes.

AUTHOR CONTRIBUTIONS

Floriane Coulmance contributed to the development of the project, conducted all the analyses and wrote the first draft of the manuscript. Derya Akkaynak contributed to the development of the camera system, part of the fieldwork and provided the scripts for colour correction. Yann Le Poul contributed to the image analysis and provided the scripts for the image alignment. Marc P. Höppner contributed to the sequencing. W. Owen McMillan contributed to development of the project and part of the fieldwork. Oscar Puebla developed the project, contributed to the development of the camera system and to all the fieldwork, provided guidance and feedback throughout the analyses, and contributed to the interpretation of the results and the writing. All authors provided feedback on the manuscript.

ACKNOWLEDGEMENTS

This study was funded by an individual German Research Foundation grant to O.P. (PU571/1-2), a Smithsonian Institute for Biodiversity Genomics and Global Genome Initiative Grants Program grant to W.O.M., O.P. and Carol Baldwin, and support to O.P. from the GEOMAR Helmholtz Centre for Ocean Research Kiel and The Future Ocean Cluster of Excellence. We thank the authorities of Panama, Puerto Rico, Belize and Florida for their support as well as the Keys Marine Laboratory, the Carrie Bow Cay, the Bocas del Toro and the La Parguera field stations. We are grateful to Allison and Carlos Estapé, Kosmas Hench, Benjamin M. Moran and Sophie Picq for their contribution to the fieldwork, to Ricardo Betancur for support in Puerto Rico and to Melanie Heckwolf, Martin Helmkampf, Paul Jay and D. Ross Robertson for their input. Open Access funding enabled and organized by Projekt DEAL.

FUNDING INFORMATION

German Research Foundation, individual grant to O.P. number PU571/1-2; Smithsonian Institute for Biodiversity Genomics and Global Genome Initiative Grants Program grant to W.O.M., O.P. and Carol Baldwin; GEOMAR Helmholtz Centre for Ocean Research Kiel and The Future Ocean Cluster of Excellence support to O.P.

CONFLICT OF INTEREST STATEMENT

The authors declare they have no conflict of interest.

DATA AVAILABILITY STATEMENT

The raw sequencing data were deposited in the European Nucleotide Archive (project accession number PRJEB61240, individual sample accession numbers listed in Table S1). All scripts needed to replicate our analyses from raw data to figures were deposited in GitHub https://github.com/florianecoulmance/hamlet_color. Access and Benefit-Sharing Clearing-House identifier ABSCH-IRCC-PA-241203-1.

ORCID

Oscar Puebla https://orcid.org/0000-0001-9700-5841

REFERENCES

Aguilar-Perera, A. (2004). Variations in morphology and coloration in the black hamlet, Hypoplectrus nigricans (Teleostei: Serranidae). *Caribbean Journal of Science*, *40*, 150–154.

Akkaynak, D., Treibitz, T., Xiao, B., Gürkan, U. A., Allen, J. J., Demirci, U., & Hanlon, R. T. (2014). Use of commercial off-the-shelf digital cameras for scientific data acquisition and scene-specific color calibration. *Journal of the Optical Society of America. A*, *31*, 312–321.

Alfaro, M. E., Karan, E. A., Schwartz, S. T., & Shultz, A. J. (2019). The evolution of color pattern in butterflyfishes (Chaetodontidae). *Integrative and Comparative Biology*, *59*, 604–615.

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*, 1217–1229.

Barreto, F. S., & McCartney, M. A. (2008). Extraordinary AFLP fingerprint similarity despite strong assortative mating between reef fish color morphospecies. *Evolution*, *62*, 226–233.

Bellwood, D. R., Goatley, C. H., & Bellwood, O. (2017). The evolution of fishes and corals on reefs: Form, function and interdependence. *Biological Reviews*, *92*, 878–901.

Bellwood, D. R., Goatley, C. H., Cowman, P. F., & Bellwood, O. (2015). The evolution of fishes on coral reefs: Fossils, phylogenies and functions. *Ecology of Fishes on Coral Reefs*, *6*, 55–63.

Brooks, S., Lear, T., Adelson, D., & Bailey, E. (2007). A chromosome inversion near the KIT gene and the Tobiano spotting pattern in horses. *Cytogenetic and Genome Research*, *119*, 225–230.

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*, e1602404.

Carroll, S. B., Grenier, J. K., & Weatherbee, S. D. (2005). *From DNA to diversity, molecular genetics and the evolution of animal design* (2nd ed.). Blackwell Publishing Ltd.

Cheney, K. L., Grutter, A. S., Blomberg, S. P., & Marshall, N. J. (2009). Blue and yellow signal cleaning behavior in coral reef fishes. *Current Biology*, *19*, 1283–1287.

Cortesi, F., Feeney, W. E., Ferrari, M. C., Waldie, P. A., Phillips, G. A., McClure, E. C., Sköld, H. N., Salzburger, W., Marshall, N. J., & Cheney, K. L. (2015). Phenotypic plasticity confers multiple fitness benefits to a mimic. *Current Biology*, *25*, 949–954.

Cott, H. B. (1940). *Adaptive coloration in animals*. Methuen & Co., Ltd.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., Del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*, 491–498.

Domeier, M. L. (1994). Speciation in the serranid fish *Hypoplectrus*. *Bulletin of Marine Science*, *54*, 103–141.

Ferreira, M. A., & Purcell, S. M. (2009). A multivariate test of association. *Bioinformatics*, *25*, 132–133.

Fischer, E. A. (1980). Speciation in the hamlets (*Hypoplectrus*: Serranidae): A continuing enigma. *Copeia*, *1980*, 649–659.

Funk, E. R., Mason, N. A., Pálsson, S., Albrecht, T., Johnson, J. A., & Taylor, S. A. (2021). A supergene underlies linked variation in color and morphology in a Holarctic songbird. *Nature Communications*, *12*, 6833.

Garcia-Machado, E., Monteagudo, P. C., & Solignac, M. (2004). Lack of mtDNA differentiation among hamlets (*Hypoplectrus*, Serranidae). *Marine Biology*, *144*, 147–152.

Hager, E. R., Harringmeyer, O. S., Wooldridge, T. B., Theingi, S., Gable, J. T., McFadden, S., Neugeboren, B., Turner, K. M., Jensen, J. D., & Hoekstra, H. E. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science*, *377*, 399–405.

Hatadani, L., Baptista, J., Souza, W., & Klaczko, L. (2004). Colour polymorphism in Drosophila mediopunctata: Genetic (chromosomal) analysis and nonrandom association with chromosome inversions. *Heredity*, *93*, 525–534.

Hemingson, C. R., Cowman, P. F., & Bellwood, D. R. (2020). Body size determines eyespot size and presence in coral reef fishes. *Ecology and Evolution*, *10*, 8144–8152.

Hemingson, C. R., Cowman, P. F., Hodge, J. R., & Bellwood, D. R. (2019). Colour pattern divergence in reef fish species is rapid and driven by both range overlap and symmetry. *Ecology Letters*, *22*, 190–199.

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 of the United States of America*, *119*, e2020457119.

Hench, K., Vargas, M., Höppner, M. P., McMillan, W. O., & Puebla, O. (2019). Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology & Evolution*, *3*, 657–667.

Holt, B. G., Côté, I. M., & Emerson, B. C. (2011). Searching for speciation genes: Molecular evidence for selection associated with colour morphotypes in the Caribbean reef fish genus *Hypoplectrus*. *PLoS One*, *6*, e20394.

Holt, B. G., Emerson, B. C., Newton, J., Gage, M. J., & Côté, I. M. (2008). Stable isotope analysis of the *Hypoplectrus* species complex reveals no evidence for dietary niche divergence. *Marine Ecology Progress Series*, *357*, 283–289.

Jakovlić, I., & Wang, W.-M. (2016). Expression of *hox* paralog group 13 genes in adult and developing *Megalobrama amblycephala*. *Gene Expression Patterns*, *21*, 63–68.

Jay, P., Leroy, M., Le Poul, Y., Whibley, A., Arias, M., Chouteau, M., & Joron, M. (2022). Association mapping of colour variation in a butterfly provides evidence that a supergene locks together a cluster of adaptive loci. *Philosophical Transactions of the Royal Society B*, *377*, 20210193.

Jeong, S., Rokas, A., & Carroll, S. B. (2006). Regulation of body pigmentation by the abdominal-B Hox protein and its gain and loss in drosophila evolution. *Cell*, *125*, 1387–1399.

Kelley, J. L., Fitzpatrick, J. L., & Merilaita, S. (2013). Spots and stripes: Ecology and colour pattern evolution in butterflyfishes. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 20122730.

Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V., & Joron, M. (2014). Evolution of dominance mechanisms at a butterfly mimicry supergene. *Nature Communications*, *5*, 1–8.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Lindtke, D., Lucek, K., Soria-Carrasco, V., Villoutreix, R., Farkas, T. E., Riesch, R., Dennis, S. R., Gompert, Z., & Nosil, P. (2017). Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology*, *26*, 6189–6205.

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*, 298–299.

Lobel, P. S. (2011). A review of the Caribbean hamlets (Serranidae, Hypoplectrus) with description of two new species. *Zootaxa*, *3096*, 1–17.

Lorenz, K. (1962). The function of color in coral reef fishes. *Proceedings of the Royal Institution of Great Britain*, *39*, 382–296.

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A combinatorial view on speciation and adaptive radiation. *Trends in Ecology & Evolution*, *34*, 531–544.

Marshall, J. (1998). Why are reef fish so colorful? *Scientific American*, *9*, 54–57.

Marshall, N. J., Cortesi, F., de Busserolles, F., Siebeck, U. E., & Cheney, K. L. (2019). Colours and colour vision in reef fishes: Past, present and future research directions. *Journal of Fish Biology*, *95*, 5–38.

Mattar, P., Ericson, J., Blackshaw, S., & Cayouette, M. (2015). A conserved regulatory logic controls temporal identity in mouse neural progenitors. *Neuron*, *85*, 497–504.

Mattar, P., Stevanovic, M., Nad, I., & Cayouette, M. (2018). Casz1 controls higher-order nuclear organization in rod photoreceptors. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E7987–E7996.

McCartney, M. A., Acevedo, J., Heredia, C., Rico, C., Quenoville, B., Bermingham, E., & Mcmillan, W. O. (2003). Genetic mosaic in a marine species flock. *Molecular Ecology*, *12*, 2963–2973.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303.

McMillan, W. O., Weigt, L. A., & Palumbi, S. R. (1999). Color pattern evolution, assortative mating, and genetic differentiation in brightly colored butterflyfishes (Chaetodontidae). *Evolution*, *53*, 247–260.

Mitchell, L. J., Cortesi, F., Marshall, N. J., & Cheney, K. L. (2022). Higher ultraviolet skin reflectance signals submissiveness in the anemonefish, *Amphiprion akindynos*. *Behavioral Ecology*, *34*, 19–32.

Moran, B. M., Hench, K., Waples, R. S., Höppner, M. P., Baldwin, C. C., McMillan, W. O., & Puebla, O. (2019). The evolution of microendemism in a reef fish (*Hypoplectrus maya*). *Molecular Ecology*, *28*, 2872–2885.

Morris, J., Navarro, N., Rastas, P., Rawlins, L. D., Sammy, J., Mallet, J., & Dasmahapatra, K. K. (2019). The genetic architecture of adaptation: Convergence and pleiotropy in Heliconius wing pattern evolution. *Heredity*, *123*, 138–152.

Ogawa, Y., & Corbo, J. C. (2021). Partitioning of gene expression among zebrafish photoreceptor subtypes. *Scientific Reports*, *11*, 1–13.

Parichy, D. M. (2021). Evolution of pigment cells and patterns: Recent insights from teleost fishes. *Current Opinion in Genetics & Development*, *69*, 88–96.

Phillips, G. A., How, M. J., Lange, J. E., Marshall, N. J., & Cheney, K. L. (2017). Disruptive colouration in reef fish: Does matching the background reduce predation risk? *Journal of Experimental Biology*, *220*, 1962–1974.

Pierotti, M. E., Wandycz, A., Wandycz, P., Rebelein, A., Corredor, V. H., Tashiro, J. H., Castillo, A., Wcislo, W. T., McMillan, W. O., & Loew, E. R. (2020). Aggressive mimicry in a coral reef fish: The prey's view. *Ecology and Evolution*, *10*, 12990–13010.

Poelstra, J. W., Vijay, N., Hoeppner, M. P., & Wolf, J. B. W. (2015). Transcriptomics of colour patterning and coloration shifts in crows. *Molecular Ecology*, *24*, 4617–4628.

Puebla, O., Bermingham, E., & Guichard, F. (2008). Population genetic analyses of Hypoplectrus coral reef fishes provide evidence that local processes are operating during the early stages of marine adaptive radiations. *Molecular Ecology*, *17*, 1405–1415.

Puebla, O., Bermingham, E., & Guichard, F. (2009). Estimating dispersal from genetic isolation by distance in a coral reef fish (*Hypoplectrus puella*). *Ecology*, *90*, 3087–3098.

Puebla, O., Bermingham, E., & Guichard, F. (2012). Pairing dynamics and the origin of species. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1085–1092.

Puebla, O., Bermingham, E., Guichard, F., & Whiteman, E. (2007). Colour pattern as a single trait driving speciation in Hypoplectrus coral reef fishes? *Proceedings of the Royal Society B: Biological Sciences*, *274*, 1265–1271.

Puebla, O., Bermingham, E., & McMillan, W. O. (2014). Genomic atolls of differentiation in coral reef fishes (*Hypoplectrus* spp., Serranidae). *Molecular Ecology*, *23*, 5291–5303.

Puebla, O., Coulmance, F., Estapé, C. J., Estapé, A. M., & Robertson, D. R. (2022). A review of 263 years of taxonomic research on Hypoplectrus (Perciformes: Serranidae), with a redescription of Hypoplectrus affinis (Poey, 1861). *Zootaxa*, *5093*, 101–141.

Puebla, O., Picq, S., Lesser, J. S., & Moran, B. (2018). Social-trap or mimicry? An empirical evaluation of the *Hypoplectrus unicolor* – *Chaetodon capistratus* association in Bocas del Toro, Panama. *Coral Reefs*, *37*, 1127–1137.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*, 559–575.

Ramon, M. L., Lobel, P. S., & Sorenson, M. D. (2003). Lack of mitochondrial genetic structure in hamlets (*Hypoplectrus* spp.): Recent speciation or ongoing hybridization? *Molecular Ecology*, *12*, 2975–2980.

Randall, J. E. (2005). A review of mimicry in marine fishes. *Zoological Studies*, *44*, 299.

Randall, J. E., & Randall, H. A. (1960). Examples of mimicry and protective resemblance in tropical marine fishes. *Bulletin of Marine Science*, *10*, 444–480.

Robertson, D. R., Pérez-España, H., Domnguez-Domnguez, O., Estapé, C. J., & Estapé, A. M. (2019). An update to the inventory of shorefishes from the Parque Nacional Sistema Arrecifal Veracruzano, Veracruz, México. *ZooKeys*, *882*, 127–157.

Rossi, M., Hausmann, A. E., Thurman, T. J., Montgomery, S. H., Papa, R., Jiggins, C. D., McMillan, W. O., & Merrill, R. M. (2020). Visual mate preference evolution during butterfly speciation is linked to neural processing genes. *Nature Communications*, *11*, 1–10.

Saenko, S. V., Marialva, M. S., & Beldade, P. (2011). Involvement of the conserved Hox gene Antennapedia in the development and evolution of a novel trait. *EvoDevo*, *2*, 9.

Salis, P., Lorin, T., Laudet, V., & Frederich, B. (2019). Magic traits in magic fish: Understanding color pattern evolution using reef fish. *Trends in Genetics*, *35*, 265–278.

Siebeck, U. E., Parker, A. N., Sprenger, D., Mäthger, L. M., & Wallis, G. (2010). A species of reef fish that uses ultraviolet patterns for covert face recognition. *Current Biology*, *20*, 407–410.

Siqueira, A. C., Morais, R. A., Bellwood, D. R., & Cowman, P. F. (2020). Trophic innovations fuel reef fish diversification. *Nature Communications*, *11*, 1–11.

Stryjewski, K. F., & Sorenson, M. D. (2017). Mosaic genome evolution in a recent and rapid avian radiation. *Nature Ecology & Evolution*, *1*, 1912–1922.

Thayer, G. H. (1909). *Concealing-coloration in the animal kingdom: An exposition of the laws of disguise through color and pattern: Being a summary of Abbott H*. Thayer's discoveries.

Theodosiou, L., McMillan, W., & Puebla, O. (2016). Recombination in the eggs and sperm in a simultaneously hermaphroditic vertebrate. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20161821.

Thresher, R. E. (1978). Polymorphism, mimicry, and the evolution of the hamlets (Hypoplectrus, Serranidae). *Bulletin of Marine Science*, *28*, 345–353.

Thummel, R., Li, L., Tanase, C., Sarras, M. P., & Godwin, A. R. (2004). Differences in expression pattern and function between zebrafish *hoxc13* orthologs: Recruitment of *hoxc13b* into an early embryonic role. *Developmental Biology*, *274*, 318–333.

Tian, L., Rahman, S. R., Ezray, B. D., Franzini, L., Strange, J. P., Lhomme, P., & Hines, H. M. (2019). A homeotic shift late in development drives mimetic color variation in a bumble bee. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 11857–11865.

Van Belleghem, S. M., 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*, 1–12.

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*, 11–10.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358–1370.

Whiteman, E., Côté, I., & Reynolds, J. (2007). Ecological differences between hamlet (*Hypoplectrus*: Serranidae) colour morphs: Between-morph variation in diet. *Journal of Fish Biology*, *71*, 235–244.

Whiteman, E., & Gage, M. (2007). No barriers to fertilization between sympatric colour morphs in the marine species flock *Hypoplectrus* (Serranidae). *Journal of Zoology*, *272*, 305–310.

Wringe, B., Stanley, R., Jeffery, N., Anderson, E., & Bradbury, I. (2016). Parallelnewhybrid: An R package for the parallelization of hybrid detection using NEWHYBRIDS.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*, 3326–3328.

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*, 821–824.

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: Coulmance, F., Akkaynak, D., Le Poul, Y., Höppner, M. P., McMillan, W. O., & Puebla, O. (2023). Phenotypic and genomic dissection of colour pattern variation in a reef fish radiation. *Molecular Ecology*, *00*, 1–15. https://doi.org/10.1111/mec.17047