# The evolution of microendemism in a reef fish (*Hypoplectrus maya*)

Benjamin M. Moran[1,2,*], Kosmas Hench[1], Robin S. Waples[3], Marc P. Höppner[4], Carole C. Baldwin[5], W. Owen McMillan[6], and Oscar Puebla[1,6,7,*,8]

[1]GEOMAR Helmholtz Centre for Ocean Research Kiel, Marine Evolutionary Ecology, Düsternbrooker Weg 20, 24105 Kiel, Germany

[2]Department of Marine and Environmental Sciences, Northeastern University, 360 Huntington Ave., Boston, MA 02115, USA

[3]Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, Seattle, Washington, USA

[4]Institute of Clinical Molecular Biology, Kiel University, Rosalind-Franklin-Strasse 12, 24105 Kiel, Germany

[5]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, 10th and Constitution Ave, NW, Washington, DC 20560, USA

[6]Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Panamá, República de Panamá

[7]University of Kiel, Faculty of Mathematics and Natural Sciences, Christian-Albrechts-Platz 4, 24118 Kiel, Germany

[*]oscar.puebla@mail.mcgill.ca, moran.ben@husky.neu.edu

[8]Lead Contact

**Running Title: The evolution of marine microendemism**

## Abstract

Marine species tend to have extensive distributions, which are commonly attributed to the dispersal potential provided by planktonic larvae and the rarity of absolute barriers to dispersal in the ocean. Under this paradigm, the occurrence of marine microendemism without geographic isolation in species with planktonic larvae poses a dilemma. The recently described Maya hamlet (*Hypoplectrus maya*, Serranidae) is exactly such a case, being endemic to a 50-km segment of the Mesoamerican Barrier Reef System (MBRS). We use whole-genome analysis to infer the demographic history of the Maya hamlet and contrast it with the sympatric and pan-Caribbean black (*H. nigricans*), barred (*H. puella*) and butter (*H. unicolor*) hamlets, as well as the allopatric but phenotypically similar blue hamlet (*H. gemma*). We show that *H. maya* is indeed a distinct evolutionary lineage, with genomic signatures of inbreeding and a unique demographic history of continuous decrease in effective population size since it diverged from congeners just ~3000 generations ago. We suggest that this case of microendemism may be driven by the combination of a narrow ecological niche and restrictive oceanographic conditions in the southern MBRS, which is consistent with the occurrence of an unusually high number of marine microendemics in this region. The restricted distribution of the Maya hamlet, its decline in both census and effective population sizes, and the degradation of its habitat place it at risk of extinction. We conclude that the evolution of marine microendemism can be a fast and dynamic process, with extinction possibly occurring before speciation is complete.



**Keywords:** hamlets, *Hypoplectrus*, endemism, speciation, demographic inference.

# Introduction

Islands, in the sense of isolation, are much more rare in the ocean than on land. Many marine organisms are planktonic for a portion of their life cycle, allowing them to cross the pelagic expanses that limit terrestrial organisms (Palumbi, 1992). Planktonic dispersal leads to a greater rate of cosmopolitanism than in terrestrial communities and a corresponding rarity of microendemic species (that is, species that are endemic to unusally small areas), as colonists are less likely to exist in isolation long enough for reproductive isolation to evolve (Kay & Palumbi, 1987; Randall, 1998; L. Rocha & Bowen, 2008). Cases of marine microendemism that have been identified are generally in taxa with short or non-existent planktonic phases (Paulay & Meyer, 2002; Meyer, Geller & Paulay, 2005). This paradigm suggests planktonic larval duration (PLD) as a potential driver of range size in reef fishes; however, syntheses have shown that these factors are poorly correlated (Lester & Ruttenberg, 2005; Mora et al., 2012; Luiz et al., 2013). Instead, growing knowledge of marine dispersal suggests that reef fishes show lower average dispersal distances than expected based on PLD (G. Jones et al., 2009). This deviation is driven in part by local oceanographic processes, and by natal homing and habitat selectivity among larvae (Leis, 2006). As such, bio-physical coupling of oceanic currents and larval behavior is currently regarded as the primary determinant of dispersal patterns (G. Jones et al., 2009).

While mean dispersal in reef fishes is often limited to the scale of tens of kilometers, the potential for rare long-distance dispersal events remains (Simpson, Harrison, Claereboudt & Planes, 2014). Consequently, rates of microendemism are generally lower in marine fishes than in terrestrial animals (Kay & Palumbi, 1987; Paulay & Meyer, 2002). Marine fishes with exceptionally small ranges are generally found in the most isolated islands of habitat, or in zones where circulation patterns lead to a unidirectional transport of larvae towards inhospitable

habitat (Roberts et al., 2002). Nonetheless, given the importance of local-scale oceanography and behavior in determining reef-fish dispersal, marine microendemism might emerge without severe geographic isolation.

The Maya hamlet, *Hypoplectrus maya* (Serranidae), is a clear example of such microendemism without geographic isolation. This reef fish was identified by P.S. Lobel in the coastal lagoon of the Mesoamerican Barrier Reef System (MBRS) in Belize in 1993, and thereafter variously referred to as the "Belize" (Heemstra, Anderson Jr. & Lobel, 2002), "Belize Blue" (Ramon, Lobel & Sorenson, 2003) or "Mayan" (C. L. Smith et al., 2003) hamlet. Domeier (1994) identified the Maya hamlet as a population of the similarly-colored blue hamlet, *H. gemma*, which occurs in Florida, Cuba, and the northern Yucatan (Aguila-Perera & Tuz-Sulub, 2010; **Fig. 1**); however, based on the lack of melanized upper and lower margins on the caudal fin which are diagnostic of *H. gemma*, Lobel (2011) described the Maya hamlet (*H. maya*) as a distinct species. The Maya hamlet has been reported on the lagoon side of the MBRS between Wee Wee Cay and the Sapodilla Cays, corresponding to a range of approximately 50 linear kilometers of reef (though one vagrant individual was collected northward in 2010 on the seaward reef wall off Alligator Cay, Lobel, 2011; **Fig. 1**). This is an exceptionally small range considering that reef-fish distributions typically range between 2000 and 13,000 km in the Atlantic (Ruttenberg & Lester, 2015). As of 2003, the species was described as "common and abundant" in the Pelican Cays and the surrounding Rhomboidal Cays (C. L. Smith et al., 2003). In particular, Lobel (2011) noted the frequent occurrence of *H. maya* among mangrove roots, suggesting an ecological specificity to the complex array of shallow coral ridges and mangroves found within these cays.

The restricted range of the Maya hamlet contrasts sharply with the pan-Caribbean barred (*H. puella*), black (*H. nigricans*), and butter (*H. unicolor*) hamlets, which are also found in the coastal lagoon of the MBRS in Belize (B. Holt, Côté & Emerson, 2010; **Fig. 1**). These species

are sympatric throughout most of their range, with ongoing gene flow maintaining low levels of genetic differentiation despite strong assortative mating (Puebla, Bermingham, Guichard & Whiteman, 2007; Puebla, Bermingham & McMillan, 2014; Hench, Vargas, Höppner, McMillan & Puebla, 2019). More broadly, the genus includes a total of 19 species that vary widely in range size and abundance (B. Holt et al., 2010). It encompasses the entire continuum of genomic divergence, from species that are almost genetically identical (Barreto & McCartney, 2008; Puebla, Bermingham & Guichard, 2012; Puebla et al., 2014) to well-diverged species (Victor, 2012; Tavera & Acero, 2013). Hamlets are nonetheless reproductively isolated from a behavioral perspective through strong assortative mating (Fischer, 1980; Puebla et al., 2007; Barreto & McCartney, 2008), and described as valid species by ichthyologists (Lobel, 2011; Victor, 2012; Tavera & Acero, 2013; Victor & Marks, 2018). We herein follow this current, accepted nomenclature and consider them species, even though reproductive isolation is not always complete. In this regard we note that the biological species concept does not necessarily imply absolute isolation; many species that are considered good species do hybridize in nature, and hybridization also occurs above the species level (Mallet, 2005).

Hamlets are very similar from an ecological perspective (Whiteman & Gage, 2007; B. G. Holt, Emerson, Newton, Gage & Côté, 2008), yet the color patterns that characterize the different species appear to be ecologically relevant through crypsis (Thresher, 1978; Fischer, 1980) and mimicry (Randall & Randall, 1960; Thresher, 1978; Puebla et al., 2007; Puebla, Picq, Lesser & Moran, 2018). It has been suggested that speciation in the hamlets may be driven by a combination of natural (Thresher, 1978; Puebla et al., 2007) and sexual (Puebla, Bermingham & Guichard, 2012) selection, but it remains unclear whether the hamlets diverged in full sympatry or in allopatry followed by secondary contact as suggested by Domeier (1994). Regardless, the two-week planktonic larval stage of the hamlets (Domeier, 1994), the occurrence of hybrid spawnings in natural populations (Fischer, 1980; Puebla et al., 2007; Barreto

& McCartney, 2008; Puebla, Bermingham & Guichard, 2012), the apparent lack of post-zygotic barriers between species (Whiteman & Gage, 2007), the identification of hybrid and backcrossed individuals in the field (Hench et al., 2019), and the low levels of genetic differentiation among sympatric species (McCartney et al., 2003; Ramon et al., 2003; Puebla, Bermingham & Guichard, 2012) as well as allopatric populations within species (Puebla, Bermingham & Guichard, 2008, 2009; Picq, McMillan & Puebla, 2016) indicate that gene flow is pervasive among Caribbean hamlets, in contrast to the genetic isolation usually implied by microendemism. Given the biogeographic disparity observed in the hamlets, *H. maya* presents an ideal opportunity to understand the processes by which marine microendemism might arise or persist in the absence of geographic isolation.

The recent publication of a chromosome-resolution reference genome for the hamlets offers a new opportunity to understand the evolution of marine microendemism from a genomic perspective (Hench et al., 2019). Here, we test whether *H. maya* represents an evolutionarily distinct lineage from its three sympatric pan-Caribbean congeners (*H. puella*, *H. nigricans*, and *H. unicolor*) and from the allopatric but phenotypically similar species *H. gemma*. Considering the restricted range of *H. maya*, we also test the hypothesis that it may present genomic signatures of inbreeding (in terms of nucleotide diversity, heterozygosity, coefficient of inbreeding, relatedness and runs of homozygosity) relative to its more widely distributed congeners. Following the same line of thought, we then infer the demographic histories of the five species using Markovian Coalescent analyses of past effective population size ($N_e$). Finally, we estimate the recent effective population size of *H. maya* and discuss the potential causes and consequences of microendemism in marine species.

# Methods

## Sampling

We considered whole genomes of 12 individuals each of *H. puella*, *H. nigricans*, and *H. unicolor* from the Belize portion of the MBRS, available from a previous study (Hench et al., 2019). To this we added 10 *H. maya* samples collected in Belize in May 2017 under STRI IACUC protocol 2017-0101-2020-2, Northeastern University IACUC protocol 17-0206R, and Belize Fisheries Department permit 000026-17, as well as 5 *H. gemma* samples collected in the Florida Keys in July 2017 under the prior IACUC protocols, NOAA ONMS permit 2017-042, and Florida FWCC permit SAL-17-1890A-SR. Gill tissue for sequencing was preserved in salt-saturated DMSO buffer, and entire fish were preserved in 10% formalin until accessioned and stored as voucher specimens in 70-75% ethanol at the Smithsonian National Museum of National History (**Suppl. Tab. 2**).

## Field Surveys

In Belize, *H. maya* surveys were carried out opportunistically in May 2017 in the center of the species' known distribution (**Fig. 1**; **Suppl. Fig. 1**). Surveys targeted reef and mangrove habitat, including the specific cays where *H. maya* had been previously reported (Domeier, 1994; C. L. Smith et al., 2003; Lobel, 2011). In the latter case, snorkelers surveyed all mangrove habitat encircling the cay and those interior ponds which were accessible by boat. A combination of snorkeling (0-5 m) and SCUBA diving (5-15 m) was moreover used to haphazardly survey reef habitat on the MBRS exterior wall, fringing reefs around cays, and patch reefs within the lagoon.

Field surveys were also conducted in the Florida Keys in July 2017 using 4 x 100 m linear SCUBA transects to assess the densities of *H. gemma* and all other hamlets, and to evaluate

whether densities and relative abundances changed over the last 15 years (**Suppl. Fig. 2**). Average densities over all transects were compared to the yearly averages from stationary surveys (15 $m^2$ diameter) conducted throughout the Keys by the Florida Keys Reef Visual Census, which took place in all years between 2002 and 2016, except for 2013 and 2015 (S. G. Smith et al., 2011). To test for changes in community composition over time, years were divided into two periods with equal sampling effort, 2002–2008 and 2009–2017. The dissimilarity of community composition was tested using PERMANOVA (Anderson, 2001) with 999 permutations and the Bray-Curtis measure of ecological distance, as implemented in the vegan package in R (Oksanen et al., 2018).

## Genotyping

*Hypoplectrus gemma* and *H. maya* genomic DNA was extracted from gill tissue using a Qiagen MagAttract High Molecular Weight Kit and sequenced to a mean genome-wide coverage of $\sim$ 22X on an Illumina HiSeq 4000 (PE, 2x151) at the Institute of Clinical Molecular Biology (IKMB) in Kiel, Germany (**Suppl. Tab. 2**), following the same sequencing approach that was taken for the *H. puella*, *H. nigricans*, and *H. unicolor* samples (Hench et al., 2019). Raw reads were mapped to the *H. puella* reference genome (Hench et al., 2019) using BWA v0.7.12 (Li & Durbin, 2009), with an average mapping efficiency of 97.24% for *H. maya*, 98.62% for *H. gemma*, 99.16% for *H. unicolor*, 99.20% for *H. nigricans*, and 99.21% for *H. puella*. All samples considered in this study were genotyped together with a workflow adapted from GATK Best Practices, with hard filters for quality control (Van der Auwera et al., 2013). Specifically, reads were filtered to remove outliers in the ratio of Phred-scaled probability of the genotype to sequencing depth (QD $<$ 2.5), the Phred-scaled *p*-value from a Fisher's Exact Test for strand bias (FS $>$ 25.0), the Strand Odds Ratio (SOR $>$ 3.0), the root-mean-squared mapping quality across samples (MQ $<$ 58.0 or $>$ 62.0) and the *U*-values from rank-sum tests for differences in

mapping quality (|MQRankSum| > 2.5) and variant position within read (|ReadPosRankSum| > 2.5) in reference vs. alternate alleles. Additional filtering with respect to minor allele frequency and coverage was specific to each analysis and mentioned explicitly when applied. Two VCF data sets were generated: one including all (variant and invariant) callable sites (555 379 974 sites, 2.5% missing data), and another including only biallelic SNPs (11 419 868 sites, 0.4% missing data). A phased data set was generated from the biallelic VCF by using phase-informative reads and SHAPEIT2 (Delaneau, Howie, Cox, Zagury & Marchini, 2013).

## Population Genomics

To estimate the extent of physical linkage, $r^2$ was calculated using VCFtools between all pairs of SNPs with a minor allele frequency greater than 10% in 200 randomly placed windows of 30 kb each. Principal Component Analysis (PCA) was performed using the R package SNPRelate (Zheng et al., 2012). This analysis was conducted on all samples, repeated considering the Belize samples only (i.e. excluding *H. gemma*), and repeated again in Belize considering a minimum distance of 15 kb between SNPs to minimize the effect of linkage. Genome-wide differentiation ($F_{ST}$) was calculated for all pairs of species using the weighted mean approach implemented in VCFtools (Weir & Cockerham, 1984). $F_{ST}$ was also calculated in sliding windows (50 kb window with 5 kb increments) between *H. maya* and each other species in order to explore the distribution of differentiation across the genome. Heterozygosity and inbreeding coefficient ($F$) were calculated for each sample using VCFtools, and the data set including all callable sites was used to calculate nucleotide diversity ($\pi$) in non-overlapping 10 kb windows for each species using VCFtools (Danecek et al., 2011). The data set including all callable sites was also used to calculate absolute divergence ($d_{XY}$; Nei, 1987) for each species pair in non-overlapping 50 kb windows using popgenWindows.py (Martin, 2016). Genome-wide absolute divergence was then calculated by averaging over the windows using the number of

SNPs as weights. Relatedness was calculated between all pairs of individuals using the Maximum Likelihood Estimation (MLE) method implemented in SNPRelate as well as the unadjusted $A_{jk}$ statistic in VCFtools (Zheng et al., 2012; Yang et al., 2010). Runs of homozygosity (ROH) greater than 150 kb in length were identified using PLINK and located in the genome after filtering SNPs for a minor allele frequency > 5% across all individuals (Purcell et al., 2007). iHH12 (Torres, Szpiech & Hernandez, 2018) was also calculated over the entire genome (50 kb windows, 5 kb increments) using selscan (Szpiech & Hernandez, 2014) to look for signs of recent positive selection.

## Demographic Inference

Demographic history was inferred for each species using the Multiple Sequentially Markovian Coalescent (MSMC) v2.0.0 (Schiffels & Durbin, 2014). Data preparation was performed following `https://github.com/stschiff/msmc-tools`. Based on the recommendations of Nadachowska-Brzyska et al. (2016), variant sites were filtered for a minimum depth of 10X, and a maximum depth of twice the individual's mean depth. MSMC inference may be affected by deviations from neutrality caused by selection (Schrider, Shanku & Kern, 2016); as such, runs were repeated after excluding the regions above the 99.90$^{\text{th}}$ $F_{ST}$ percentile identified in Hench et al. (2019). Each MSMC run included 4 individuals, with the exception of 2 runs of 3 individuals in *H. maya* (**Suppl. Tab. 4**). Each individual was included in only one MSMC analysis; replicate runs are therefore independent sample-wise. To explore the history of divergence among species, we also considered the cross-population coalescence rate, scaled relative to the within-population coalescence rates (relative cross-coalescence rates, Schiffels & Durbin, 2014). Cross-coalescence rates were inferred for the maximum possible number of independent runs for each species pair, considering two individuals per species (four individuals per run). All MSMC runs were performed with a time segmentation pattern of $1 \cdot 2 + 25 \cdot 1 + 1 \cdot 2 + 1 \cdot 3$, and the

average of Watterson's estimator across input data sets, $\theta = 2.55 * 10^{-3}$. To explore whether the recent demographic trends observed in MSMC were an artifact of phasing switch errors, we also applied SMC++ v1.14.0, an extension of the Sequentially Markovian Coalescent that does not rely on phasing (Terhorst, Kamm & Song, 2017). A single SMC++ composite likelihood estimate for each species was created from the product of estimates across chromosomes, and across each possible "distinguished individual" in a species (see Terhorst et al., 2017). In both MSMC and SMC++, the mutation rate was set at $\mu = 3.7 * 10^{-8}$, based on the closest relative for which the value was known (Liu, Hansen & Jacobsen, 2016). Generation times (that is, the mean age of successfully reproducing individuals) for hamlets are uncertain, but likely fall between 1 and 3 years based on size, taxonomy, and habitat. Due to the resultant uncertainty, time is presented in terms of generations, with potential years on a secondary axis.

To complement estimates of past effective population size ($N_e$), we used a novel whole-genome implementation of recent $N_e$ estimation based on linkage disequilibrium (Hill, 1981; R. S. Waples, 2006). We first used GATK to subset the biallelic SNP set by species, then selected sites with no missing genotype calls and minor allele counts $> 2$ (i.e. minor allele frequency $> 0.1$ in *H. maya*). Each SNP set was then randomly subset into 100 non-overlapping sets, to which $N_e$ estimations were applied independently. We utilized a new feature of the LD method in NeEstimator v2.1 (released December 2017), calculating $N_e$ based on only interchromosomal comparisons (Do et al., 2014). Confidence intervals were obtained from the per-individual jackknife of A. Jones, Ovenden & Wang, 2016, as well as the distribution of $N_e$ across the 100 SNP subsets. This analysis was applied to all species except for *H. gemma* due to the low sample size for this species.

All scripts to reproduce our results from the raw data are available at `https://github.com/benmoran11/hamlets_endemism.git`.

# Results

## Field Surveys

Only two Maya hamlets were sighted in the Pelican Cays and surrounding Rhomboidal Cays where *H. maya* was described as "common and abundant" by Smith et al. (2003). One individual was a juvenile found on the reef flat adjacent to Little Cat Cay at a depth of 1.5 m next to an *Orbicella* coral head. The other individual was found in proximity to a barrel sponge surrounded by *Acropora cervicornis* rubble at a depth of 3 m in "Tunicate Cove", a honeycomb of coral ridges adjacent to Cat Cay where Lobel (2011) collected the holotype and eight paratypes over seven years. In contrast, *H. maya* was the most abundant hamlet species in the shallow (1-5 m) *A. cervicornis* patch reefs near Laughing Bird Cay National Park, where the majority of samples were collected for this study. Two individuals were also sighted at a depth of 2 m on an *Orbicella*-dominated fringing reef in Bread and Butter Cay (**Suppl. Fig. 1**).

In Florida, 27 non-overlapping transects were conducted between Geiger Key and French Reef, encompassing the majority of the range surveyed by the FL Reef Visual Census (Key West to Key Largo; **Suppl. Fig. 2**). Total mean density of hamlets in 2017 was 4.8 $\pm$ 1.0 (SE) fish 1000 $m^{-2}$, while RVC estimates in 2002-2016 fell between 2.5 and 4.5 fish 1000 $m^{-2}$ (**Suppl. Fig. 3**). Hamlet community composition changed significantly between the first and second half of the temporal data set (**Suppl. Fig. 4**; PERMANOVA $P = 0.002$), with *H. unicolor* increasing in relative abundance at the expense of *H. gemma*. Between the two periods, mean *H. gemma* densities declined by more than 50%, from 0.41 $\pm$ 0.03 (SE) to 0.18 $\pm$ 0.03 (SE) fish 1000 $m^{-2}$.

## Population Genomics

Our linkage analysis indicates that physical linkage decays rapidly within 5 kb (**Suppl. Fig. 5**). Genome-wide PCA showed clear clustering of *H. maya*, *H. gemma*, and *H. nigricans*, with partial overlap between *H. puella* and *H. unicolor* (**Fig. 2**a). Similar patterns were obtained when considering only SNPs > 15 kb apart to minimize physical linkage (**Suppl. Fig. 6**). Genome-wide differentiation was greatest between *H. maya* and *H. gemma* ($F_{ST}$ = 0.060), lowest between *H. puella* and *H. unicolor*, ($F_{ST}$ = 0.004), and intermediate for the other species pairs ($F_{ST}$ = 0.014–0.040; **Tab. 1**). Sliding-window analysis revealed heterogeneous patterns of differentiation between *H. maya* and the four other species, with an accumulation of differentiation on linkage groups (LGs) 8 and 9 (likely due to large inversions, Hench et al., 2019) and a number of sharp peaks, some of which were repeated across different species comparisons (**Suppl. Fig. 7**a). The genomic regions above the 99.99$^{th}$ $F_{ST}$ percentile in comparisons involving *H. maya* are highlighted in **Suppl. Fig. 7** and the genes found within these regions are listed in **Suppl. Tab. 3**.

Heterozygosity was depressed in *H. maya* (median = 0.162) relative to other Belizean hamlets (median = 0.169 – 0.173) and to *H. gemma* (median = 0.181; **Fig. 2**d). Nucleotide diversity was also lowest in *H. maya* (median $\pi$ = 0.0047 versus 0.0049 – 0.0053 in other species), but the difference was small relative to the variation among 10 kb windows (**Suppl. Fig. 8**). For both maximum likelihood estimates and $A_{jk}$, mean relatedness was highest in *H. gemma* (mean MLE $r$ = 0.012, $A_{jk}$ = 0.054), followed by *H. maya* (mean MLE $r$ = 0.008, $A_{jk}$ = 0.032) and the other Belizean hamlet species (mean MLE $r$ = 0, $A_{jk}$ = -0.011 – 0.003 **Fig. 2**c, **Suppl. Fig. 9**). A positive outlier was observed between two *H. maya* individuals, suggesting inbreeding beyond background relatedness (MLE $r$ = 0.041, $A_{jk}$ = 0.093; **Fig. 2**c, **Suppl. Fig. 9**). Inbreeding in *H. maya* was also suggested by the higher inbreeding coefficients observed in this species (median $F$ = 0.068) relative to the other Belizean species (median $F$ =

0.008 – 0.030, Fig. 2e, note that this includes *H. unicolor* which is rare in Belize) as well as the markedly higher number of runs of homozygosity > 150 kb in *H. maya* relative to other species (Fig. 2b). The ROH were located all over the genome, indicating that the higher prevalence of ROH in *H. maya* is a genome-wide phenomenon. Nevertheless, ROH were disproportionately represented on LG2, LG9 and LG12, matching the $F_{ST}$ patterns (Suppl. Fig. 7b). This result was confirmed by the integrated haplotype homozygosity pooled (iHH12, Torres et al., 2018, Suppl. Fig. 7c), which is often used to detect signs of recent positive selection and thereby suggests that selection is also playing a role in these regions. The blue hamlet showed negative inbreeding coefficients (median = -0.040), yet this result should be interpreted with caution due to the low sample size for this species ($n = 5$).

## Demographic Inference

We used MSMC to identify demographic trends leading to current biogeographic patterns. The most ancient and two most recent time segments provided highly inconsistent $N_e$ estimates within species (Suppl. Fig. 10) and were therefore not considered, since this suggests unreliable inference (S. Schiffels, personal communication). All species presented very similar trends earlier than 3000 generations before present (gbp), suggesting that they diverged only recently (Fig. 3). Following an expansion until 2000 gbp, *H. maya* $N_e$ decreased continuously to a minimum of 12000 at 290 gbp (Fig. 3). The *H. gemma* and *H. nigricans* populations also decreased beginning 2000 gbp, but rebounded to a final $N_e$ of 50000 and 100000 $\pm$ 15000 (mean $\pm$ SE across *H. nigricans* runs), respectively. In contrast, *H. puella* and *H. unicolor* $N_e$ increased to final values of 120000 $\pm$ 18000 and 110000 $\pm$ 13000, respectively (Fig. 3). SMC++ analysis, which does not rely on phasing, confirmed that these general trends were not due to phasing switch errors (Suppl. Fig. 12). Though the heuristic calculation of time points limited SMC++ inference to $10^3$–$10^5$ gbp, we nonetheless observed a population expansion beginning

$10^4$ gbp in all species, a sharp decline in *H. maya*, and a limited decline in *H. gemma* (Suppl. Fig. 12). The most notable differences in the SMC++ results were large $N_e$ fluctuations between $10^5$ and $10^4$ gbp and a shift towards older times for the beginning of the declines in *H. maya* and *H. gemma* (Suppl. Fig. 12). For both analyses, results were qualitatively identical with and without the most diverged genomic regions that are likely under selection (Suppl. Fig. 10; Suppl. Fig. 11; Suppl. Fig. 12). The cross-coalescence results indicate that *H. gemma* diverged from the other species within $\sim$ 6000 gbp, followed by *H. nigricans* ($\sim$ 5000 gbp) and *H. maya* ($\sim$ 3000 gbp, Fig. 4). The barred and butter hamlets appear to have diverged even more recently ($\sim$ 2000 gbp), yet these results should be interpreted with caution due to ongoing gene flow between these two species in Belize (Hench et al., 2019, which may explain the observed cross-coalescence rates $>$ 1.0). Relative cross-coalescence was $>$ 0.01 in all comparisons until $<$ 500 gbp, and remained $>$ 0.05 throughout inference in two *H. puella–H. unicolor* runs (Fig. 4). As such, MSMC relative cross-coalescence supports other evidence of ongoing gene flow within the genus, especially between *H. puella* and *H. unicolor*.

For the estimation of recent $N_e$, quality filters left 3,296,967 suitable variant sites in *H. maya*, which were split into 100 non-overlapping data sets. Median estimated $N_e$ was 1584 individuals, with a minimum of 1002, and a maximum of 9478 (Fig. 5). Based on the Jones et al. (2016) jackknife variance method, NeEstimator estimated that the effective degrees of freedom associated with the 100 subsets ranged from 229647 to 532857; jackknife 95% confidence intervals had lower bounds between 277 and 528, and a consistent upper bound of infinity (Fig. 5). In contrast, the 100 replicates provided an empirical 95% CI of 1073 – 4426 effective individuals (Fig. 5). All 100 analyses for *H. puella*, *H. unicolor* and *H. nigricans* produced $N_e$ point estimates, as well as lower and upper confidence bounds, of infinity.

# Discussion

Our data confirm that *H. maya* represents a rare case of microendemism in reef fishes. From the moment of its scientific documentation, this species was confused with the phenotypically similar *H. gemma* of the northern Caribbean (Domeier, 1994). The diagnostic color pattern used to describe the new species and distinguish it from *H. gemma* (absence of black margins on the caudal fin, Lobel, 2011) is only found within the MBRS; however, such characteristics are strained as taxonomic identifiers in the hamlets, where intermediate phenotypes, polymorphism, and regional variants of described species are frequently observed. In particular, black margins on the caudal fin are polymorphic within other hamlet species and populations (O. Puebla, personal observation). As such, we sought first to establish the status of *H. maya* as a distinct evolutionary unit. Our analyses demonstrate that *H. maya* and *H. gemma* are distinct evolutionary lineages, despite their phenotypic similarity. In fact, whole-genome differentiation between these two species is markedly higher than any other allopatric or sympatric comparison within this study, and *H. maya* is also differentiated from the other three sympatric pan-Caribbean hamlets (**Tab. 1**; **Fig. 2**). The Maya hamlet can therefore be considered a separate species, so far as the biological species concept applies to the low differentiation and ongoing gene flow regime within *Hypoplectrus*.

## The evolution of microendemism

Considering the restricted distribution of *H. maya* and its recent divergence, it provides a rare window into the evolution of marine microendemism. The heterogeneous landscape of genomic differentiation between *H. maya* and other *Hypoplectrus* species suggests that *H. maya* evolved under the effect of selection and may be locally adapted (**Suppl. Fig. 7**). Some of the highly differentiated regions evidenced here have been previously identified, and include genes involved in vision (*rorb*) and pigmentation (*sox10*) that may play a role in reproductive isolation

through visually-based assortative mating (Hench et al., 2019). We also note the presence of a sharp peak of differentiation on LG07 centered on the androgen receptor (AR) gene, which, although not above the 99.99$^{th}$ $F_{ST}$ percentile, is consistently and exclusively observed in comparisons involving *H. maya* (**Suppl. Fig. 7**a). A iHH12 signal was also observed at this locus (**Suppl. Fig. 7**c), suggesting that it is under positive selection. Androgens are involved in the development of sex-specific traits, including vision (Shao et al., 2014) and pigmentation (Lindsay, Webster & Schwabl, 2011). It remains to be shown whether this is the case in the hamlets, which have a very specific simultaneously hermaphroditic mating system whereby individuals reciprocally trade eggs for fertilization (Fischer, 1980).

All measures point to reduced genomic diversity and increased inbreeding in *H. maya* relative to pan-Caribbean congeners (**Fig. 2**). The Maya hamlet shows decreased heterozygosity, higher inbreeding coefficients, and more runs of homozygosity than sympatric congeners, as expected following a bottleneck or ongoing population decline (Nei, Maruyama & Chakraborty, 1975; Frankham, 1998). In contrast to the three pan-Caribbean species, background levels of relatedness are also $> 0$ in *H. maya*. Furthermore, we identified one pair of Maya hamlets that are much more related than background levels ($r = 0.041$, which corresponds to the level of relatedness that is expected between second cousins with a most recent common ancestor 3 generations ago; Wright, 1922; **Fig. 2**). These individuals were collected 34 km apart, at opposite ends of the sampling area, which is within the estimated dispersal potential of Belizean hamlets across three generations (Puebla, Bermingham & McMillan, 2012). Median nucleotide diversity was also 4-12% lower in *H. maya* than congeners (**Suppl. Fig. 8**). This difference may appear small, particularly in comparison to observed $\pi$ in other taxa: *H. maya* nucleotide diversity is $\sim 2$ times higher than that observed in *Ficedula* flycatchers, and $\sim 6$ times higher than that in humans (Primmer, Borge, Lindell & Sætre, 2002; International SNP Map Working Group, 2001). This high diversity is expected within the framework of high marine effective

population sizes, and is concordant with our inferred demographies: the hamlets experienced a pre-divergence bottleneck of $N_e \approx 30 \times 10^4$ (**Fig. 3**), as opposed to $20 \times 10^4$ and $1 \times 10^4$ in flycatchers and humans, respectively (Nadachowska-Brzyska et al., 2016; Li & Durbin, 2011).

We note that *H. gemma* presents striking population genomic patterns, with higher levels of heterozygosity and background relatedness, and lower (negative) inbreeding coefficients relative to the four other species (**Fig. 2**). We suggest that the high heterozygosity and apparent outbreeding observed in this species may be associated with the mixing of two lineages, from the Gulf of Mexico and Caribbean, in the Florida Keys (Ramon et al., 2003). As for the high levels of relatedness, they may be due to the ongoing decline of *H. gemma* populations in the Florida Keys documented by the transect data (**Suppl. Fig. 3**; **Suppl. Fig. 4**). We nevertheless reiterate caution with these hypotheses since they rely on only five *H. gemma*.

The analysis of present-day diversity and divergence is complemented by an understanding of the historical population dynamics in which they arose. Our approach allowed us to infer *Hypoplectrus* demographic histories up to $< 300$ generations before present, with a likely historical range of $\sim 300 - 900$ years ago (**Fig. 3**). Regardless of uncertainty in *Hypoplectrus* generation times, inference provided clear support for widely divergent demographic trends in *H. maya*, beginning near the last glacial maximum. While pan-Caribbean species began a growth trajectory ending with effective population sizes around 100000, *H. maya* began a monotonic decrease to $N_e \approx 12000$. In contrast, *H. gemma* $N_e$ declined to $\sim 30000$, and rebounded to $\sim 50000$. The divergent trajectories of these taxa provide further support for their evolutionary distinction. Cross-coalescence rates, too, support the developing picture of *Hypoplectrus* as an ongoing speciation event. Our analyses suggest four independent divergence windows, all falling during or after the last glacial maximum (**Fig. 4**). Extended gene flow is also suggested by this coalescent approach, with gene flow continuing into the current millennium in all lineages, and ongoing between *H. puella* and *H. unicolor*, the species pair between which high-probability

hybrid and back-crossed individuals have been previously identified (Hench et al., 2019). An explicit analysis of the history of gene flow—which may be complex—is beyond the scope of this study, and we note that the decrease in $N_e$ inferred in *H. maya* may also be interpreted in terms of a decrease in gene flow from other hamlet species and populations. Regardless, given the recent divergence of *H. maya*, it is likely that it arose within the MBRS and is thereby neoendemic to this area.

## Recent effective population size

The estimation of recent effective population size from linkage disequilibrium using whole-genome data has been limited by the computational scale of the necessary number of pairwise comparisons, as well as physical linkage, which decreases the effective degrees of freedom presented by each pair of loci (R. K. Waples, Larson & Waples, 2016). To eliminate bias due to physical linkage, we considered only interchromosomal comparisons. The remaining effects of non-independence among pairwise comparisons of loci were accounted for by the per-individual jackknife procedure of A. Jones et al., 2016, which calculates "effective degrees of freedom" and corresponding confidence intervals. In addition, we leveraged the scale of our data set to calculate $N_e$ estimates from 100 non-overlapping sets of markers, allowing an empirical evaluation of uncertainty in our estimate. These replicates display much less uncertainty than the jackknife confidence intervals would suggest; though no finite upper bounds could be placed on the jackknife CIs, 95% of our estimates fell between $\sim$ 1000 to 4500 (**Fig. 5**). Simulation-based analysis of pseudo-replication in genomic-scale LD data sets suggests that the Jones et al. (2016) jackknife confidence interval generally underestimates precision in LDNe, and that subsetting loci provides a more realistic assessment. On the other hand, genetic indices (like $r^2$ for unlinked loci) that reflect very recent demography are sensitive to the pedigree structure of the individuals in the sample. Replicating across many subsets of loci, all generated by the

same pedigree, will not capture uncertainty associated with differences between the pedigree structure of the sample and the pedigree structure of the population as a whole (King, Wakeley & Carmi, 2018). This argues for some caution in interpreting CIs for estimates of Ne for the hamlets, all of which are based on small samples of individuals. Nonetheless, given the order of magnitude of the $N_e$ estimates, this does not change our interpretation of the *H. maya* population as orders of magnitude smaller than its size at the beginning of speciation, including a tenfold reduction within the last few hundred generations.

Our recent $N_e$ estimate of $\sim$ 1600 contrasts with the rarity of *H. maya* in the field and its restricted distribution. Considering the dramatic decline of *H. maya* in the Pelican and surrounding Rhomboidal Cays within the last two decades documented here, this number may nevertheless be inflated by much higher effective population sizes just a few generations ago. It is also possible that *H. maya* $N_e$ is still affected by gene flow from pan-Caribbean hamlets, or that the population center of *H. maya* may not be in the Pelican and surrounding Rhomboidal Cays but around Laughing Bird Cay and further south, beyond the area surveyed here. Though effective population sizes as low as 500 were originally theorized as stable from a mutation-drift equilibrium perspective, the body of empirical evidence suggests that sizes of 1000-5000 are likely necessary to maintain fitness in perpetuity (Lande, 1995; Frankham, Bradshaw & Brook, 2014). As such, we suggest that the past and present effective population size of *H. maya* is by itself sufficient cause for concern regarding its long-term survival.

Our data also support the disparity in effective population size between *H. maya* and its congeners. $N_e$ estimation for *H. gemma* was not possible due to the low sample size for this species ($n$=5, which is below the validated range for LD-based estimation). Such a limitation is unfortunate given the recent decline in *H. gemma* census population reported here, and a renewed effort to estimate this species' $N_e$ is advised. In other species, though, infinite $N_e$ estimates were obtained with a larger sample size ($n$ = 12) than in *H. maya* ($n$ = 10), which

indicates that the pan-Caribbean species' $N_e$ can be reliably inferred as 'much larger' than that of *H. maya*. This is compatible with a previous Approximate Bayesian Computation estimate of $N_e$ of $\sim$ 15000 for *H. nigricans* on the BBR, an order of magnitude higher than our *H. maya* estimate (Puebla, Bermingham & McMillan, 2012).

## Microendemism in the MBRS

While the case of the Maya hamlet is remarkable, it is not unique. Twelve fish species are known to be endemic to the Belize section of the MBRS and the adjacent Honduran Bay Islands, representing over 20% of those endemic to the continental Caribbean (Floeter et al., 2008; Robertson & Van Tassell, 2015; **Suppl. Tab. 1**). Similar levels of microendemism are found among invertebrates (Rützler et al., 2000; Miloslavich et al., 2010). In the MBRS, the endemic fishes are distributed variably between the landward lagoon, the seaward barrier wall, and the associated atolls (Lobel, Rocha & Randall, 2009). This high level of microendemism may be due in part to the intense sampling and exploration of the southern MBRS (Miloslavich et al., 2010). Yet analogous cases of microendemism have been documented in the less intensively sampled Indo-Pacific (Allen, Erdmann & Hidayat, 2018; Allen, Erdmann & Cahyani, 2018), suggesting that such patterns may be more prevalent among reef-fish communities than previously recognized. Should broader sampling reveal similar concentrations of microendemics elsewhere in the Caribbean, in particular among small cryptobenthic fishes or in the mesophotic zone, the question of the underlying evolutionary processes will become even more pressing.

In accordance with the recognition of ocean currents as a limiting factor in marine dispersal (G. Jones et al., 2009), we suggest that local oceanography may be a primary cause of high microendemism in the MBRS. Drifters and numerical models have identified a system of temporally variable eddies that occur along the Belizean MBRS. Areas south of Glover's Reef ($\sim$ 16.75°N) experience slow, invariant transport to the south, while those found at or north

of this point experience variable transport dependent on the season: transport may be rapidly southward, or weakly northwestward (Ezer, Thattai, Kjerfve & Heyman, 2005; Tang, Sheng, Hatcher & Sale, 2006). Particles (e.g. planktonic larvae) which are transported southward either encounter the interior MBRS lagoon and the Honduran Bay Islands, or are carried into a gyre within the Gulf of Honduras (Richardson, 2005; Paris, Chérubin & Cowen, 2007). Of the 12 species endemic to the MBRS and southward Honduran islands, ten have northward boundaries at Carrie Bow Cay and Glover's Reef (Floeter et al., 2008; Robertson & Van Tassell, 2015). D'Aloia (2015) estimated the dispersal kernel of one of these endemics (*Elacatinus lori*) at the proposed oceanographic divide, and recovered an isotropic kernel of extremely small dispersal range. Such a pattern is consistent with an oceanographic limitation to range expansion, so long as these species originated in the southern MBRS under the current oceanographic regime. Multiple independent estimates of these species kernels across their entire range, extending the work of D'Aloia (2015), could shed further light on this hypothesis.

The case of the Maya hamlet is remarkable in that it is currently sympatric with congeners in terms of both distribution and microhabitat. This contrasts with other cases of microendemism in reef fishes, which show either allopatry or habitat divergence (Allen, Erdmann & Hidayat, 2018; Allen, Erdmann & Cahyani, 2018). Though *H. maya* overlaps in habitat with sympatric congeners, it may differ in its habitat specificity. Our qualitative observations indicate that *H. maya* is strongly associated with shallow (1–3 m) reef habitat. The Maya hamlet was nearly extirpated from the Pelican Cays as of 2017, coinciding with the degradation of shallow coral communities on the Cays' characteristics polygonal ridges (O. Puebla and B. Moran, personal observation). In contrast, *H. maya* was the dominant hamlet species on the shallow reefs west of Laughing Bird Cay, which harbored high coverage of *A. cervicornis* (O. Puebla and B. Moran, personal observation). Specialist adaptation to shallow *A. cervicornis* reefs would provide another explanation for the long-term $N_e$ decline of *H. maya* inferred by our MSMC

analyses, given the geological history of their range. The cays of the southern MBRS lagoon began as Pleistocene limestone surfaces, which were submerged by sea-level rise after the last glacial maximum (Macintyre, Precht & Aronson, 2000). *Acropora cervicornis* colonized this substrate, growing towards the surface at a rate of up to 8 m/1000 years (Westphall, 1986; Macintyre et al., 2000; Aronson, Macintyre, Precht, Murdoch & Wapnick, 2002). Where reef accretion outpaced sea level rise, the reef crest was colonized by the shallow-specialist coral *Porites divaricata*, and later red mangrove (*Rhizophora mangle*) trees (Neumann, 1985; Macintyre et al., 2000). The MBRS lagoon thus represents a non-equilibrium habitat in relative isolation, presenting the exceptional opportunity for reduced gene flow with outside populations, unfilled niches, and founder effects. If *H. maya* is indeed an *A. cervicornis* specialist that appeared in the mid-Holocene as inferred by our MSMC analyses, ecological succession after the last glacial maximum would have created a long-term natural decline in habitat availability throughout its existence. This, combined with a relatively short PLD of 14-22 days (Domeier, 1994) and the aforementioned oceanographic characteristics of the southern MBRS, may explain this case of micro-endemism and a long-term decline of *H. maya*. The generality of such forces could be tested in other cases of microendemism, both in geographically distinct cases within *Hypoplectrus* (Victor & Marks, 2018) and in phylogenetically distinct cases within the MBRS.

## Microendemism and extinction

Species with small ranges are particularly vulnerable to extinction, due to a combination of low total population size and increased threat presented by local extirpations (Gaston, 1998). This risk is further elevated in the case of ecological specialists, which exhibit a synergistic combination of lower population densities and lower tolerance to change (Munday, 2004). While *H. maya* population declines predated human influence, the reduction in habitat available to *H. maya* was likely accelerated in the last century by the drastic decline in Caribbean corals,

and acroporids in particular. This trend of reef degradation is largely attributable to coral disease outbreaks (Aronson & Precht, 2001), coastal development (Murray, 2007), decline of herbivorous fishes and invertebrates (Hughes, 1994), and ocean warming (Aronson, Precht, Macintyre & Murdoch, 2000). The MBRS lagoon, in particular, is currently threatened by clear-cutting of mangroves and dredging of shallow patch reefs to increase land values for real estate and touristic development (McKee & Vervaeke, 2009). Furthermore, the invasive lionfish constitutes a direct threat to the Maya hamlet and other Caribbean microendemic fishes (L. A. Rocha, Rocha, Baldwin, Weigt & McField, 2015). Such a combination of stressors provides a plausible explanation for the recent reduction of the *H. maya* population evidenced here by both genetic data and field surveys. Likewise, the recent decline in *H. gemma* evidenced by transect surveys (**Suppl. Fig. 3**) coincides with the loss of Florida Keys reef communities to disease and warming (Precht, Gintert, Robbart, Fura & Van Woesik, 2016). Collection by the aquarium trade may also play a role in the case of *H. gemma*, given the popularity of this species among public and private aquarists (O. Puebla and B. Moran, personal observation). Given the exceptionally small range of *H. maya*, its rarity, its long-term and recent decline in population size, its strong association with *A. cervicornis* and the ongoing degradation of its habitat, the persistence of this recently-diverged species is in jeopardy. The case of the Maya hamlet shows that the evolution of marine microendemism can be a fast and dynamic process, with extinction possibly occurring before speciation is complete.

## Acknowledgements

We thank Ryan Waples for feedback during the implementation of whole-genome LDNe, Stephan Schiffels for his guidance regarding MSMC, Derya Akkaynak for assistance in the field, as well as Katie Lotterhos and Phil Lobel for their input regarding population genomics and Belizean endemics. We are indebted to the staff of Carrie Bow Cay Field Station and

Keys Marine Laboratory for their assistance during sampling, to Allison and Carlos Estapé, and to Lisa Carne of the Belizean coral restoration non-profit Fragments of Hope. This work was conducted under National Geographic Society Young Explorers Grant #WW-037ER-17, a Summer Independent Research Fellowship from Northeastern University, and two scholarships (Undergraduate and Short-Term Research) from the German Academic Exchange Service to BM, a German Research Foundation and a Future Ocean Cluster of Excellence grant to OP, and a Global Genome Initiative and Smithsonian Institute for Biodiversity Genomics grant to CB, WOM, and OP.

## References

Aguila-Perera, A. & Tuz-Sulub, A. N. (2010). Scientific note Hypoplectrus gemma (Teleostei, Serranidae) is not endemic to southern Florida waters. *Pan-American Journal of Aquatic Sciences*, *5*(1), 143–146.

Allen, G., Erdmann, M. & Cahyani, N. D. (2018). Chrysiptera uswanasi, a new microendemic species of damselfish (Teleostei: Pomacentridae) from West Papua Province, Indonesia. *Journal of the Ocean Science Foundation*, *31*, 74–86.

Allen, G., Erdmann, M. & Hidayat, N. (2018). Pomacentrus bellipictus, a new microendemic species of damselfish (Pisces: Pomacentridae) from the Fakfak Peninsula, West Papua, Indonesia. *Journal of the Ocean Science Foundation*, *30*, 1–10.

Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46.

Aronson, R. B., Macintyre, I. G., Precht, W. F., Murdoch, T. J. & Wapnick, C. M. (2002). The expanding scale of species turnover events on coral reefs in Belize. *Ecological Monographs*, *72*(2), 233–249.

Aronson, R. B. & Precht, W. F. (2001). White-band disease and the changing face of Caribbean coral reefs. *Hydrobiologia*, *460*(1), 25–38.

Aronson, R. B., Precht, W. F., Macintyre, I. G. & Murdoch, T. J. (2000). Ecosystems: coral bleach-out in Belize. *Nature*, *405*(6782), 36.

Barreto, F. S. & McCartney, M. A. (2008). Extraordinary AFLP fingerprint similarity despite strong assortative mating between reef fish color morphospecies. *Evolution: International Journal of Organic Evolution*, *62*(1), 226–233.

D'Aloia, C. C., Bogdanowicz, S. M., Francis, R. K., Majoris, J. E., Harrison, R. G. & Buston, P. M. (2015). Patterns, causes, and consequences of marine larval dispersal. *Proceedings of the National Academy of Sciences*, *112*(45), 13940–13945.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group (2011). The variant call format and vcftools. *Bioinformatics*, *27*(15), 2156–2158.

Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F. & Marchini, J. (2013). Haplotype estimation using sequencing reads. *The American Journal of Human Genetics*, *93*(4), 687–696.

Do, C., Waples, R. S., Peel, D., Macbeth, G., Tillett, B. J. & Ovenden, J. R. (2014). Nees-timator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, *14*(1), 209–214.

Domeier, M. L. (1994). Speciation in the serranid fish Hypoplectrus. *Bulletin of Marine Science*, *54*(1), 103–141.
Ezer, T., Thattai, D. V., Kjerfve, B. & Heyman, W. D. (2005). On the variability of the flow along the Meso-American Barrier Reef system: a numerical model study of the influence of the Caribbean current and eddies. *Ocean Dynamics*, *55*(5-6), 458–475.
Fischer, E. A. (1980). The relationship between mating system and simultaneous hermaphroditism in the coral reef fish, Hypoplectrus nigricans (Serranidae). *Animal Behaviour*, *28*(2), 620–633.
Floeter, S. R., Rocha, L. A., Robertson, D. R., Joyeux, J., Smith-Vaniz, W. F., Wirtz, P., ... Bernardi, G. (2008). Atlantic reef fish biogeography and evolution. *Journal of Biogeography*, *35*(1), 22–47.
Frankham, R. (1998). Inbreeding and extinction: island populations. *Conservation Biology*, *12*(3), 665–675.
Frankham, R., Bradshaw, C. J. & Brook, B. W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, *170*, 56–63.
Gaston, K. J. (1998). Ecology: rarity as double jeopardy. *Nature*, *394*(6690), 229.
Heemstra, P. C., Anderson Jr., W. D. & Lobel, P. S. (2002). Serranidae: Groupers (seabasses, creolefish, coney, hamlets, anthiines, and soapfishes). In K. E. Carpenter (Ed.), *The living marine resources of the Western Central Atlantic, Volume 2: Bony fishes, part 1 (Acipenseridae to Grammatidae)* (p. 1308-1369). FAO Species Identification Guide for Fishery Purposes and American Society of Ichthyologists and Herpetologists Special Publication No. 5, FAO, Rome, Italy, 601–1374.
Hench, K., Vargas, M., Höppner, M. P., McMillan, W. O. & Puebla, O. (2019). Interchromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology & Evolution*. doi: 10.1038/s41559-019-0814-5
Hill, W. G. (1981). Estimation of effective population size from data on linkage disequilibrium. *Genetics Research*, *38*(3), 209–216.
Holt, B., Côté, I. & Emerson, B. (2010). Signatures of speciation? distribution and diversity of Hypoplectrus (Teleostei: Serranidae) colour morphotypes. *Global Ecology and Biogeography*, *19*(4), 432–441.
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.
Hughes, T. P. (1994). Catastrophes, phase shifts, and large-scale degradation of a Caribbean coral reef. *Science*, *265*(5178), 1547–1551.
International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, *409*(6822), 928.
Jones, A., Ovenden, J. & Wang, Y. (2016). Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. *Heredity*, *117*(4), 217.
Jones, G., Almany, G., Russ, G., Sale, P., Steneck, R., Van Oppen, M. & Willis, B. (2009). Larval retention and connectivity among populations of corals and reef fishes: history, advances and challenges. *Coral Reefs*, *28*(2), 307–325.
Kay, E. A. & Palumbi, S. R. (1987). Endemism and evolution in Hawaiian marine invertebrates. *Trends in Ecology & Evolution*, *2*(7), 183–186.
King, L., Wakeley, J. & Carmi, S. (2018). A non-zero variance of tajima's estimator for two sequences even for infinitely many unlinked loci. *Theoretical Population Biology*, *122*, 22–29.
Lande, R. (1995). Mutation and conservation. *Conservation Biology*, *9*(4), 782–791.
Leis, J. M. (2006). Are larvae of demersal fishes plankton or nekton? *Advances in Marine Biology*, *51*, 57–141.

Lester, S. E. & Ruttenberg, B. I. (2005). The relationship between pelagic larval duration and range size in tropical reef fishes: a synthetic analysis. *Proceedings of the Royal Society of London B: Biological Sciences*, *272*(1563), 585–591.

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

Li, H. & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493.

Lindsay, W. R., Webster, M. S. & Schwabl, H. (2011). Sexually selected male plumage color is testosterone dependent in a tropical passerine bird, the red-backed fairy-wren (malurus melanocephalus). *PLoS One*, *6*(10), e26067.

Liu, S., Hansen, M. M. & Jacobsen, M. W. (2016). Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. *Molecular Ecology*, *25*(20), 5187–5202.

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

Lobel, P. S., Rocha, L. A. & Randall, J. E. (2009). Color phases and distribution of the western Atlantic labrid fish, Halichoeres socialis. *Copeia*(1), 171–174.

Luiz, O. J., Allen, A. P., Robertson, D. R., Floeter, S. R., Kulbicki, M., Vigliola, L., ... Madin, J. S. (2013). Adult and larval traits as determinants of geographic range size among tropical reef fishes. *Proceedings of the National Academy of Sciences*, 201304074.

Macintyre, I. G., Precht, W. F. & Aronson, R. B. (2000). Origin of the Pelican Cays ponds, Belize. *Atoll Research Bulletin*(466).

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229–237.

Martin, S. H. (2016). *genomics_general: General tools for genomic analyses; a github repository.* Retrieved from https://github.com/simonhmartin/genomics_general.git

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*(11), 2963–2973.

McKee, K. L. & Vervaeke, W. C. (2009). Impacts of human disturbance on soil erosion potential and habitat stability of mangrove-dominated islands in the Pelican Cays and Twin Cays Ranges, Belize. *Smithsonian Contributions to the Marine Sciences*(38).

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.

Miloslavich, P., Díaz, J. M., Klein, E., Alvarado, J. J., Díaz, C., Gobin, J., ... Ortiz, M. (2010). Marine biodiversity in the Caribbean: regional estimates and distribution patterns. *PloS One*, *5*(8), e11916.

Mora, C., Treml, E. A., Roberts, J., Crosby, K., Roy, D. & Tittensor, D. P. (2012). High connectivity among habitats precludes the relationship between dispersal and range size in tropical reef fishes. *Ecography*, *35*(1), 89–96.

Munday, P. L. (2004). Habitat loss, resource specialization, and extinction on coral reefs. *Global Change Biology*, *10*(10), 1642–1647.

Murray, G. (2007). Constructing paradise: the impacts of big tourism in the Mexican coastal zone. *Coastal Management*, *35*(2-3), 339–355.

Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology*, *25*(5), 1058–1072.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Nei, M., Maruyama, T. & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, *29*(1), 1–10.

Neumann, A. (1985). Reef response to sea-level rise: keep-up, catch-up, or give-up. In *Proceedings of the fifth international coral reef congress, Tahiti* (Vol. 3, pp. 105–110).

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2018). vegan: Community ecology package [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=vegan (R package version 2.5-2)

Palumbi, S. R. (1992). Marine speciation on a small planet. *Trends in Ecology & Evolution*, *7*(4), 114–118.

Paris, C. B., Chérubin, L. M. & Cowen, R. K. (2007). Surfing, spinning, or diving from reef to reef: effects on population connectivity. *Marine Ecology Progress Series*, *347*, 285–300.

Paulay, G. & Meyer, C. (2002). Diversification in the tropical Pacific: comparisons between marine and terrestrial systems and the importance of founder speciation. *Integrative and Comparative Biology*, *42*(5), 922–934.

Picq, S., McMillan, W. O. & Puebla, O. (2016). Population genomics of local adaptation versus speciation in coral reef fishes (Hypoplectrus spp, Serranidae). *Ecology and Evolution*, *6*(7), 2109–2124.

Precht, W. F., Gintert, B. E., Robbart, M. L., Fura, R. & Van Woesik, R. (2016). Unprecedented disease-related coral mortality in Southeastern Florida. *Scientific Reports*, *6*, 31374.

Primmer, C., Borge, T., Lindell, J. & Sætre, G.-P. (2002). Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology*, *11*(3), 603–612.

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*(6), 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*(11), 3087–3098.

Puebla, O., Bermingham, E. & Guichard, F. (2012). Pairing dynamics and the origin of species. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1731), 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 of London B: Biological Sciences*, *274*(1615), 1265–1271.

Puebla, O., Bermingham, E. & McMillan, W. O. (2012). On the spatial scale of dispersal in coral reef fishes. *Molecular Ecology*, *21*(23), 5675–5688.

Puebla, O., Bermingham, E. & McMillan, W. O. (2014). Genomic atolls of differentiation in coral reef fishes (Hypoplectrus spp., Serranidae). *Molecular Ecology*, *23*(21), 5291–5303.

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*(4), 1127–1137.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., ... Sham, P. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 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*(11), 2975–2980.

Randall, J. E. (1998). Zoogeography of shore fishes of the Indo-Pacific region. *Zoological Studies*, *37*(4), 227–268.

Randall, J. E. & Randall, H. A. (1960). Examples of mimicry and protective resemblance in tropical marine fishes. *Bulletin of Marine Science*, *10*(4), 444–480.

Richardson, P. L. (2005). Caribbean current and eddies as observed by surface drifters. *Deep Sea Research Part II: Topical Studies in Oceanography*, *52*(3-4), 429–463.

Roberts, C. M., McClean, C. J., Veron, J. E., Hawkins, J. P., Allen, G. R., McAllister, D. E., ... Werner, T. B. (2002). Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science*, *295*(5558), 1280–1284.

Robertson, D. R. & Van Tassell, J. (2015). *Shorefishes of the Greater Caribbean: online information system.* Smithsonian Tropical Research Institute, Balboa, Panamá.

Rocha, L. & Bowen, B. (2008). Speciation in coral-reef fishes. *Journal of Fish Biology*, *72*(5), 1101–1121.

Rocha, L. A., Rocha, C. R., Baldwin, C. C., Weigt, L. A. & McField, M. (2015). Invasive lionfish preying on critically endangered reef fish. *Coral Reefs*, *34*(3), 803–806.

Ruttenberg, B. I. & Lester, S. E. (2015). Patterns and processes in geographic range size in coral reef fishes. In C. Mora (Ed.), *Ecology of fishes on coral reefs* (p. 97). Cambridge University Press.

Rützler, K., Díaz, M. C., van Soest, R. W. M., Zea, S., Smith, K. P., Alvarez, B. & Wulff, J. (2000). Diversity of sponge fauna in mangrove ponds, Pelican Cays, Belize. *Atoll Research Bulletin*(476).

Schiffels, S. & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919.

Schrider, D. R., Shanku, A. G. & Kern, A. D. (2016). Effects of linked selective sweeps on demographic inference and model selection. *Genetics*, genetics–116.

Shao, Y. T., Wang, F.-Y., Fu, W.-C., Yan, H. Y., Anraku, K., Chen, I.-S. & Borg, B. (2014). Androgens increase lws opsin expression and red sensitivity in male three-spined sticklebacks. *PLoS One*, *9*(6), e100330.

Simpson, S. D., Harrison, H. B., Claereboudt, M. R. & Planes, S. (2014). Long-distance dispersal via ocean currents connects Omani clownfish populations throughout entire species range. *PLoS One*, *9*(9), e107610.

Smith, C. L., Tyler, J. C., Davis, W. P., Jones, R. S., Smith, D. G. & Baldwin, C. C. (2003). Fishes of the Pelican Cays, Belize. *Atoll Research Bulletin*, *497*.

Smith, S. G., Ault, J. S., Bohnsack, J. A., Harper, D. E., Luo, J. & McClellan, D. B. (2011). Multispecies survey design for assessing reef-fish stocks, spatially explicit management performance, and ecosystem condition. *Fisheries Research*, *109*(1), 25–41.

Szpiech, Z. A. & Hernandez, R. D. (2014, 10). selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular biology and evolution*, *31*(10), 2824–2827. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/25015648https://www.ncbi.nlm.nih.gov/pmc/PMC4166924/ doi: 10.1093/molbev/msu211

Tang, L., Sheng, J., Hatcher, B. G. & Sale, P. F. (2006). Numerical study of circulation, dispersion, and hydrodynamic connectivity of surface waters on the Belize shelf. *Journal of Geophysical Research: Oceans*, *111*(C1).

Tavera, J. & Acero, A. P. (2013). Description of a new species of Hypoplectrus (Perciformes: Serranidae) from the Southern Gulf of Mexico. *Aqua: International Journal of Ichthyology*, *19*(1), 29–39.

Terhorst, J., Kamm, J. A. & Song, Y. S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, *49*(2), 303.

Thresher, R. E. (1978). Polymorphism, mimicry, and the evolution of the hamlets (Hypoplectrus, Serranidae). *Bulletin of Marine Science*, *28*(2), 345–353.

Torres, R., Szpiech, Z. A. & Hernandez, R. D. (2018). Human demographic history has amplified the effects of background selection across the genome. *PLoS Genetics*, *14*(6), 1–27. Retrieved from https://doi.org/10.1371/journal.pgen.1007387 doi: 10.1371/journal.pgen.1007387

UNEP-WCMC, WorldFish Centre, WRI & TNC. (2010). Global distribution of coral reefs, compiled from multiple sources including the Millennium Coral Reef Mapping Project. Version 2.0, updated by UNEP-WCMC. Includes contributions from IMaRS-USF and IRD (2005), IMaRS-USF (2005) and Spalding et al. (2001). Cambridge (UK): UNEP World Conservation Monitoring Centre. (URL: http://data.unepwcmc.org/datasets/1)

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., ... DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*(SUPL.43). doi: 10.1002/0471250953.bi1110s43

Victor, B. C. (2012). Hypoplectrus floridae n. sp. and Hypoplectrus ecosur n. sp., two new barred hamlets from the Gulf of Mexico (Pisces: Serranidae): more than 3% different in COI mtDNA sequence from the Caribbean Hypoplectrus species flock. *Journal of the Ocean Science Foundation*, *5*, 1–19.

Victor, B. C. & Marks, K. (2018). Hypoplectrus liberte, a new and endangered microendemic hamlet from Haiti (Teleostei: Serranidae). *Journal of the Ocean Science Foundation*, *31*, 8–17.

Waples, R. K., Larson, W. A. & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*, *117*(4), 233–240.

Waples, R. S. (2006). A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics*, *7*(2), 167–184.

Weir, B. S. & Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Westphall, M. (1986). *Anatomy and history of a ringed-reef complex, Belize, Central America*. Unpublished master's thesis, University of Miami, Coral Gables, Florida.

Whiteman, E. & Gage, M. (2007). No barriers to fertilization between sympatric colour morphs in the marine species flock Hypoplectrus (Serranidae). *Journal of Zoology*, *272*(3), 305–310.

Wright, S. (1922). Coefficients of inbreeding and relationship. *The American Naturalist*, *56*(645), 330–338.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., ... Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, *42*(7), 565–569.

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*(24), 3326–3328.

## Data Accessibility

All new raw sequences (*H. maya* and *H. gemma*) have been deposited at the European Nucleotide Archive (ENA) under project accession number PRJEB29705; the individual accession numbers for these samples are provided in **Suppl. Tab. 2**. Previously sequenced Belizean samples (*H. nigricans*, *H. puella*, and *H. unicolor*) are available under ENA project PRJEB27858. The biallelic SNP genotypes in VCF format for all samples are also available in Dryad (doi:10.5061/dryad.hp388dm).

## Author contributions

BM conceived of the study, conducted field work and data analyses, and wrote the manuscript. OP conceived of the study, conducted field work, and contributed to the manuscript. RWS contributed to data analyses and the manuscript. KH contributed to the data analyses. WOM and MH contributed to genome sequencing. CB contributed to the curation of specimens. All co-authors provided feedback on the manuscript.

# Tables

**Table 1** Estimates of genome-wide differentiation and divergence among the *Hypoplectrus* species considered in this study; $F_{ST}$ above the diagonal, and $d_{XY}$ below.

| **Species** | *H. gemma* | *H. maya* | *H. nigricans* | *H. puella* | *H. unicolor* |
|---|---|---|---|---|---|
| *H. gemma* | — | 0.060 | 0.040 | 0.037 | 0.033 |
| *H. maya* | 0.00379 | — | 0.039 | 0.026 | 0.028 |
| *H. nigricans* | 0.00378 | 0.00360 | — | 0.016 | 0.014 |
| *H. puella* | 0.00379 | 0.00357 | 0.00360 | — | 0.004 |
| *H. unicolor* | 0.00379 | 0.00360 | 0.00361 | 0.00359 | — |

# Figures

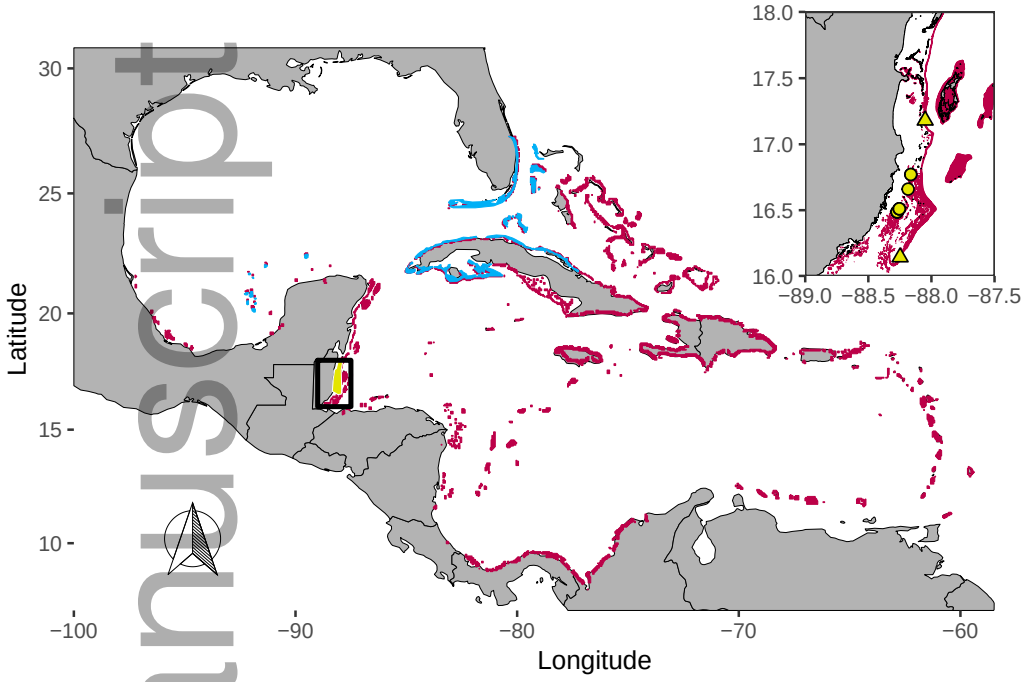


**Figure 1** Ranges of the *Hypoplectrus* species considered in this study. *H. puella*, *H. nigricans*, and *H. unicolor* occur throughout the Greater Caribbean (purple). *H. gemma* (blue) is restricted to the Northern Caribbean, and *H. maya* (yellow) to a section of the Mesoamerican Barrier Reef System (MBRS) in Belize. Inset: reports of *H. maya* from the literature (triangles) and the current study (circles, note that some of these locations had been reported before). Distribution of pan-Caribbean hamlets extrapolated from the WCMC-008 Global Distribution of Coral Reefs (UNEP-WCMC et al., 2010).

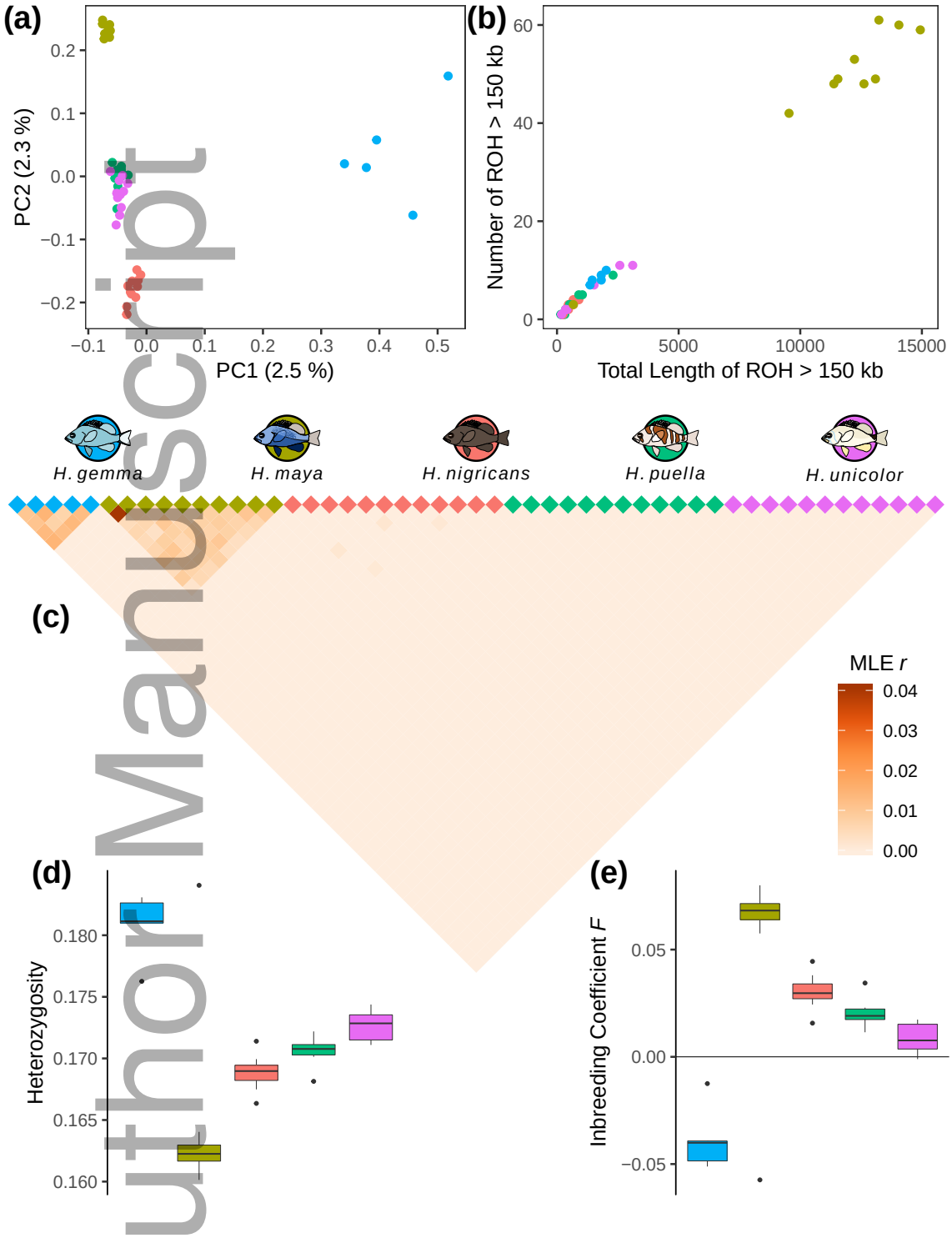


**Figure 2** Population genomics of five *Hypoplectrus* species. **a)** Principal Component Analysis (PCA) based on whole genome data from all individuals in this study. Proportion of explained variance for the first two PCs listed on axes. **b)** Runs of Homozygosity (ROH) > 150 kb in each individual. The total number of ROH with length > 150 kb is plotted against the summed length of those ROH. **c)** Genome-wide Maximum Likelihood Estimation (MLE) of relatedness between all pairs of samples. **d)** Heterozygosity, calculated genome-wide for each individual. **e)** Inbreeding coefficient $F$, calculated genome-wide for each individual. Central bars represent median values, and boxes $25^{th}$ – $75^{th}$ percentile intervals. Whiskers show data within 1.5 × interquartile range, and dots are outliers beyond this range.

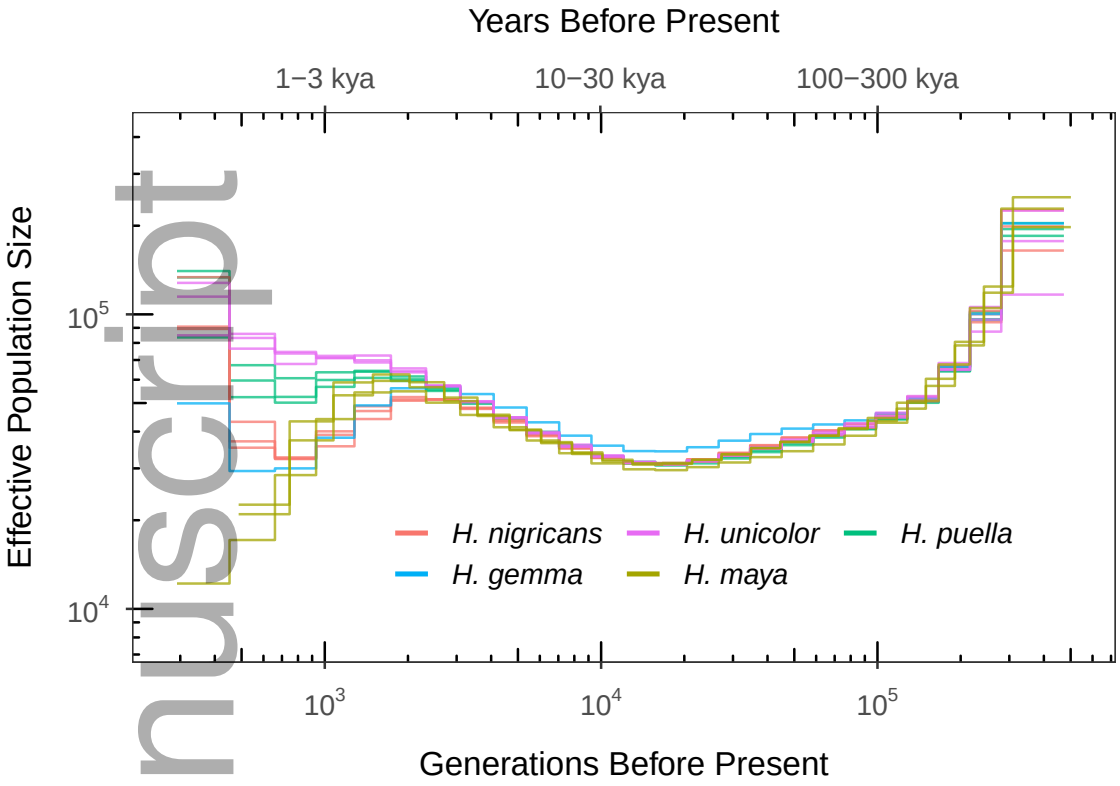


**Figure 3** MSMC inference of effective population size over time in the five species. Each analysis is based on 3-4 genomes and each genome is used in only one analysis. All estimates are scaled with a per-site mutation rate $\mu = 3.7 * 10^{-8}$. The most ancient and two most recent time segments are omitted due to unreliable inference (see text).

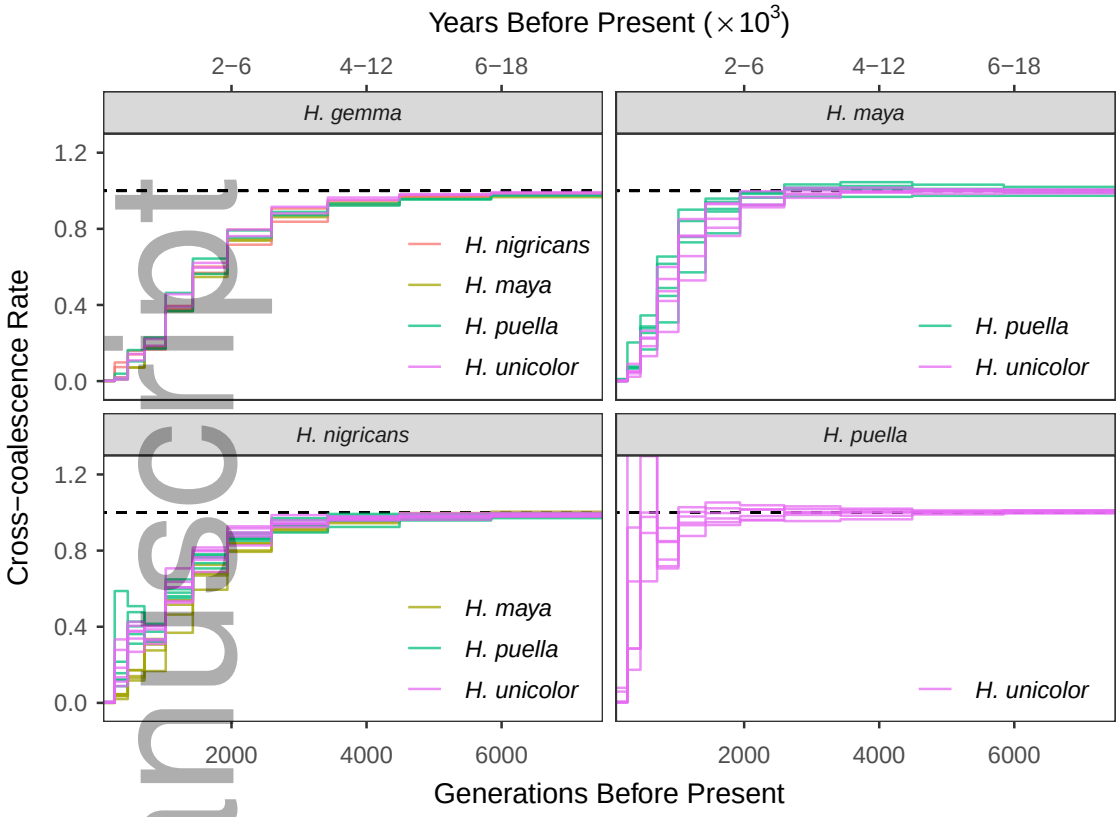


**Figure 4** MSMC cross-coalescence inference of divergence times between all pairs of species. Each line represents an independent run including 2 individuals from each species. Panel headers identify the first species in the comparison, and colors the second. All estimates are scaled with a per-site mutation rate $\mu = 3.7 * 10^{-8}$. In a given time interval, a relative cross-coalescence rate of 1 (dashed line) indicates totally shared ancestry, and a rate of 0 indicates no shared ancestry.

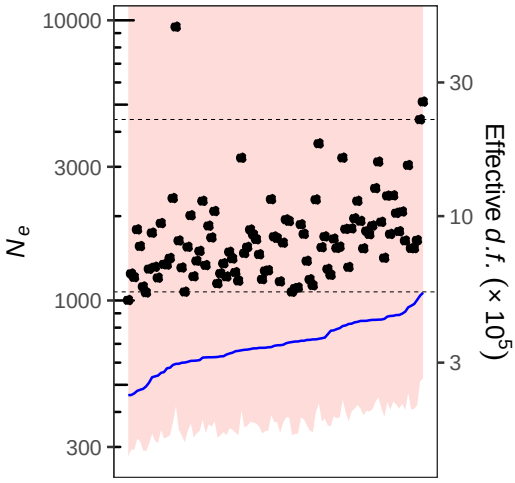


**Figure 5** Estimates of *H. maya* recent effective population size from inter-chromosomal LD among 100 non-overlapping SNP subsets. $N_e$ point estimates (black points) are ordered by effective degrees of freedom (blue line) inferred from the individual-wise jackknife procedure of Jones et al. (2016). Corresponding $N_e$ 95% CIs (red shading) extend to positive infinity in all estimates. Empirical 95% CI is denoted by dashed horizontal lines. Both vertical axes are log-scaled.