# The phenotypic and genetic basis of a reef fish radiation

Von der Fakultät für Mathematik und Naturwissenschaften der Carl von Ossietzky Universität Oldenburg zur Erlangung des Grades und Titels:

Doktor der Naturwissenschaften (Dr.rer.nat.)

angenommene Dissertation

von Frau Floriane Coulmance Gayrard

geboren am 19.07.1996 in Paris

Erster Gutachter:
Prof. Dr. Oscar Puebla

Zweite Gutachter:
Prof. Dr. Arne W. Nolte

Dritter Gutachter:
Dr. W. Owen McMillan

Tag der Disputation: 16.01.2025

# Contents

## Summary

The origins of the extraordinary diversity of life on Earth have long intrigued scientists. Evolutionary radiations, where species emerge from increased diversification rates, play an instrumental role in generating biodiversity. These events are often characterized and shaped by remarkable level of phenotypic diversity. At their early stages, phenotypic variation provides a substrate for selection and a means for further diversification. While the genomic underpinnings of speciation have been extensively investigated, phenotypic studies have been limited by analytical power, despite advances in data acquisition. The advent of computer vision now enables systematic, multivariate analyses of phenotypes, transforming our ability to examine their eco-evolutionary relevance. This thesis aims to investigate the evolution of color pattern phenotypes in the early stages of a reef fish radiation. I focus on the genus *Hypoplectrus* (hamlets), a group of 18+ species that exhibit striking color pattern differences and represent one of the most recent marine radiations. Color patterns in hamlets are implicated in visually based assortative mating, acting as prezygotic barriers, and may also play ecological roles through mimicry, thereby subjecting this trait to selection pressures. By integrating phenotypic and genomic approaches, I aim to characterize color pattern variation and explore its contribution to reproductive isolation at the onset of speciation. The genic view of speciation posits that at early stages a few loci can disproportionately contribute to reproductive isolation. By characterizing color pattern variation and its genomic basis in hamlets, I explore how this perspective aligns with patterns of speciation in this radiation. I first develop a quantitative pipeline for the analysis of standardized *in situ* photographs, providing the first multivariate, pixel-scale phenotypic dataset for reef fishes. Coupled with genomic association studies, this approach reveals three genomic regions of large effect that underlie color pattern variation. This modular genetic and phenotypic architecture suggests that hamlet diversity arises from different allele combinations at these loci. Expanding on this, I leverage a dataset of 571 photographs and genomic data from 327 individuals to examine interspecific and intraspecific variation. While phenotypic and genomic clusters of sympatric species are locally retrieved, these clusters become largely continuous at the scale of the whole radiation. This pattern, driven by intraspecific variation, suggests contributions from ancestral variation and hybridization, and underscores the role of local processes in sympatric speciation. Finally, I assess the genic view by integrating both whole-genome data and color pattern-associated genomic regions in a phylogenetic analysis. I find that reproductive isolation in hamlets does not translate into a phylogenetic signal, whether using whole-genome datasets or specific regions. As predicted under the genic view, these findings suggest that reproductive isolation can occur first without a detectable genomic signature. Taken together, my results support the view that speciation is a continuous and multidimensional process, where divergence accumulates along genetic and phenotypic axes, ultimately mediating reproductive isolation and diversification. These insights advance our comprehension of the complex interplay between genetics, phenotypes, and speciation, particularly in the early stages of evolutionary radiations.

## Zusammenfassung

Die Ursprünge der bemerkenswerten Vielfalt des Lebens auf der Erde sind seit langem ein zentrales wissenschaftliches Thema. Evolutionäre Radiationen, bei denen Arten durch zunehmende Diversifizierung hervorgehen, spielen eine entscheidende Rolle bei der Entstehung der biologischen Vielfalt. Außergewöhnliche phänotypische Vielfalt kennzeichnet und prägt oft solche Episoden, bietet in ihren frühen Stadien ein Substrat für die Selektion und ermöglicht eine weitere Diversifizierung. Während die genomischen Grundlagen der Artbildung zuletzt verstärkt Gegenstand der Forschung waren, blieben phänotypische Studien trotz der Fortschritte bei der Datenerfassung durch die analytischen Möglichkeiten begrenzt. Das Aufkommen der Computervision ermöglicht nun systematische, multivariate Analysen von Phänotypen, wodurch sich unsere Möglichkeiten zur Untersuchung ihrer ökoevolutionären Bedeutung verändern. Das Ziel dieser Arbeit ist es, die Evolution von Farbmuster-Phänotypen in den frühen Stadien einer Rifffisch-Radiation zu untersuchen. Ich konzentriere mich auf die Gattung *Hypoplectrus* (Hamletbarsche), eine Gruppe von mehr als 18 Arten, die auffällige Farbmusterunterschiede aufweisen und eine der jüngsten marinen Radiationen darstellen. Die Farbmuster der Hamletbarsche sind an der visuellen assortativen Paarung beteiligt, die als präzygotische Barriere wirkt, und können durch Mimikry eine ökologische Rolle spielen, wodurch dieses Merkmal neben sexuellem auch unter natürlichem Selektionsdruck steht. Mit phänotypischen und genomischen Ansätzen charakterisiere ich die Farbmuster und ihren Einfluss auf die reproduktive Isolation in frühen Stadien der Artbildung. Laut der “genic view of species” können wenige Gene dabei eine zentrale Rolle spielen. Ich untersuche, wie Farbmustervariationen und ihre genetische Grundlage die Artbildung bei Hamletbarschen beeinflussen. Ich entwickle eine Pipeline, die erstmals multivariate phänotypische Daten auf Pixelebene aus standardisierten *in situ*-Fotos von Rifffischen liefert. In Verbindung mit einer genomischen Assoziationsstudie enthüllt dieser Ansatz drei genomische Regionen mit großer Auswirkung auf die Variation der Farbmuster. Diese genetische und phänotypische Architektur lässt vermuten, dass die Vielfalt der Hamletbarsche aus verschiedenen Allelkombinationen an diesen Loci resultiert. Darauf aufbauend nutze ich einen Datensatz von 571 Fotos und genomische Daten von 327 Individuen, um inter- und intraspezifische Variation zu untersuchen. Sympatrische Arten zeigen lokal klare phänotypische und genetische Cluster, die auf der Ebene der gesamten Radiation jedoch zu einem Kontinuum verschwimmen. Dieses Muster, geprägt durch intraspezifische Variation, weist auf hohe anzestrale Variation und Hybridisierung hin und betont die Bedeutung lokaler Prozesse für die sympatrische Artbildung. Schließlich bewerte ich die “genic view” auf die Arten, indem ich sowohl Daten aus dem gesamten Genom, als auch Farbmuster-assoziierte genomische Regionen in eine phylogenetische Analyse integriere. Diese Ergebnisse deuten darauf hin, dass reproduktive Isolation zunächst ohne nachweisbare genomische Signatur entstehen kann, wie es die “genic view”. Meine Ergebnisse zeigen, dass Artbildung ein kontinuierlicher, mehrdimensionaler Prozess ist, bei dem genetische und phänotypische Divergenz zur reproduktiven Isolation und Diversifizierung führen.

## Resumen

Los orígenes de la increíble diversidad de la vida en la Tierra son una vieja incógnita científica. Las radiaciones evolutivas, en las que surgen varias especies a partir de un evento de diversificacion, juegan un papel central en la generación de biodiversidad. Estos episodios suelen caracterizarse y configurarse por una notable diversidad fenotípica. En sus primeras fases, la variación fenotípica proporciona una base en la cual selección actua y da paso a procesos evolutivos que incrementan la diversificación. Mientras que los bases geneticas de la especiación se han investigado ampliamente, los estudios fenotípicos se han visto limitados por la capacidad analítica, a pesar de los avances en la automatizacion y tecnologia usada en la adquisición de datos. La llegada de la visión por ordenador permite ahora realizar análisis sistemáticos y multivariados a caracteristicas fenotipicas, transformando nuestra capacidad para examinar su relevancia en procesos ecoevolutivos. El objetivo de esta tesis es investigar la evolución de los fenotipos de patrones de color en las primeras etapas de la radiación de un pez de arrecife. Me centro en el género *Hypoplectrus* (vacas), un grupo de más de 18+ especies que exhiben diferencias contrastantes en los patrones de color y representan una de las radiaciones marinas más recientes. Los patrones de color de los hamlets están asociados a el apareamiento selectivo, actuando como barreras precigóticas, y también desempeñan funciones ecológicas a través del mimetismo, sometiendo este rasgo a la presión de selección. Mediante la integración de metodos fenotípicos y genómicos, mi objetivo es caracterizar la variación en los patrones de color y explorar su contribución al aislamiento reproductivo en el inicio de la especiación. En primer lugar, desarrollo un proceso cuantitativo estandarizado para el análisis de fotografías in situ, proporcionando el primer conjunto de datos fenotípicos multivariados para peces de arrecife basado en una escala de píxeles. En conjunto con estudios de asociación genómica, identifico tres regiones genómicas de gran efecto asociadas con la variación del patrón de color. Esta arquitectura genética y fenotípica modular sugiere que la diversidad en coloracion de las vacas surge de diferentes combinaciones de alelos en estos loci. A partir de ahí, expando los analisis de este estudio usando un conjunto de 571 fotografías y datos genómicos de 327 individuos para examinar la variación inter e intraespecífica. Aunque se recuperan agrupaciones fenotípicas y genómicas de especies simpátricas a escalas locales, estas agrupaciones se vuelven en gran medida continuas a lo largo de toda la radiación. Este patrón, incentivado por la variación intraespecífica, sugiere la existencia de una variación ancestral y de procesos de hibridación, enfatizando a la vez el papel de los procesos locales en la especiación simpátrica. Por último evaluo genomas completo y las regiones genómicas asociadas a patrones de color en un análisis filogenético. Tal y como predice el “genic view of species”, encuentro que el aislamiento reproductivo puede ocurrir en primer lugar sin una señal genómica detectable. En conjunto, mis resultados apoyan la opinión de que la especiación es un proceso continuo y multidimensional, en el que la divergencia se acumula a lo largo de bases genéticas y fenotípicas, mediando en última instancia el aislamiento reproductivo y la diversificación.

## Résumé

Les origines de l'incroyable diversité de la vie sur Terre fascinent les chercheurs depuis longtemps. Les radiations évolutives sont essentielles à l'augmentation de cette biodiversité et sont souvent accompagnées d'une diversité phénotypique remarquable. À leurs débuts, cette variation offre un substrat pour la sélection et favorise une diversification accrue. Bien que les bases génomiques de la spéciation aient été largement explorées, les études sur les phénotypes ont longtemps été freinées par des limites analytiques, malgré les progrès technologiques d'acquisition de données. Aujourd'hui, grâce à la vision numérique, il est possible d'analyser les phénotypes de manière systématique et multivariée, ouvrant de nouvelles perspectives pour comprendre leur rôle éco-évolutif. Cette thèse a pour objectif d'étudier l'évolution des phénotypes liés aux motifs de coloration au cours des premières étapes de la radiation d'un poisson de récifs. Pour cela, j'évalue les hamlets (*Hypoplectrus*), un groupe de 18 espèces qui se caractérise par d'incroyables motifs de coloration et qui représente l'une des radiations marines les plus récentes. Les motifs de coloration chez les hamlets interviennent lors de l'accouplement assortatif visuel, jouant un rôle de barrière prézygotique, et pourraient également jouer un rôle écologique *via* le mimétisme, ce qui les place sous pression de sélection. En combinant des approches phénotypiques et génomiques, je cherche à examiner la variation des motifs de coloration et leur rôle dans l'isolement reproductif au début de la spéciation. Selon la théorie de la "genic view of species", quelques gènes clés peuvent fortement influencer l'isolement reproductif. En étudiant la variation des motifs de coloration et leurs bases génomiques, j'explore comment cette idée se reflète chez les hamlets. Je développe d'abord une méthode d'analyse quantitative de photographies prises *in situ*, ce qui constitue la première étude de pixels multivariés d'informations phénotypiques sur des poissons de coraux. En les utilisants dans une étude d'association génomique, je révèle trois loci ayant un effet majeur sur la variation des motifs de coloration. Cette architecture génétique et phénotypique modulaire suggère que la diversité des hamlets résulte de différentes combinaisons alléliques à ces loci. J'analyse ensuite 571 photographies et les génomes de 327 individus pour étudier la variation inter- et intraspécifique. Des clusters phénotypiques et génomiques sont visibles localement, mais deviennent continus à l'échelle de la radiation, reflétant l'influence de la variation ancestrale, de l'hybridation et des processus locaux dans la spéciation sympatrique. Enfin, j'évalue la "genic view" au niveau de génomes entiers et au régions sous-jacentes des motifs de coloration dans une perspective phylogénétique. Je constate que l'isolement reproductif chez les hamlets ne se traduit par un signal phylogénétique ni au niveau du génome entier, ni aux régions spécifiques. Ces résultats, conformes à la "genic view", suggèrent que l'isolement reproductif peut se manifester avant que ne se détecte une signature génomique. Mes résultats montrent que la spéciation est un processus continu et complexe, où la divergence génétique et phénotypique conduit à l'isolement reproductif et à la diversification. Cette thèse éclaire les liens entre génétique, phénotypes et spéciation, enrichissant notre compréhension des radiations évolutives.

# 01

# INTRODUCTION

"Again, we have many slight differences which may be called individual differences, [...]. No one supposes that all the individuals of the same species are cast in the very same mold. These individual differences are highly important for us, as they afford materials for natural selection to accumulate, [...]. That varieties of this doubtful nature are far from uncommon cannot be disputed."

*Charles Darwin, 1859*

## 1.1 Biodiversity and evolutionary radiations

Life on our planet and its diversity have fascinated biologists for centuries. Understanding how this diversity arises is valuable because biodiversity plays a central role in the functioning of ecosystems, which provide essential resources to the planet and valuable services to humans (Barbier et al., 2011). In coastal and marine ecosystems, the study of biodiversity enables better fisheries management, ensuring a sustainable food supply for people, while protecting the biological communities that live in these ecosystems. This biodiversity supports critical ecosystem functions such as water purification and climate regulation, which in turn benefit human well-being (Millennium ecosystem assessment, 2005). Loss of biodiversity negatively affects ecosystem health and services, and contributes to human health and social problems. These issues have been exacerbated by the accelerated loss of biodiversity over the last 50 years (Millennium ecosystem assessment, 2005). Understanding the origins of biodiversity and how its units (i.e. species) change over time is therefore valuable, as it can have a significant impact on human and planetary health.

Most of the biodiversity we observe on the planet has been attributed to processes of evolutionary radiation (Schluter, 2000), whereby an increase in diversity occurs within a given group of living organisms. Examples of evolutionary radiations can be found in different taxa ranging from plants to birds and fish. The radiation of flowering plants (angiosperms) during the Cretaceous period generated extraordinary biodiversity, accounting for ~90% of the 300,000 known land plant species (Hernández-Hernández and Wiens, 2020). Studying the origins and eco-evolutionary drivers of these radiations enhances our general understanding of the processes that shape present-day biodiversity.

Most studies of biodiversity have focused on adaptive radiations, where ecological opportunity and the evolution of novel key traits allow rapid diversification (Schluter, 2000; Stroud and Losos, 2020). Here, emerging species adapt to divergent ecological niches, generating variation for natural selection to act upon and thereby driving further differences between lineages (Schluter, 2000). Darwin's finches on the Galápagos Islands are a classic example of an adaptive radiation, where innovations in beak shape allowed them to exploit new feeding niches, leading to their diversification into 15 species

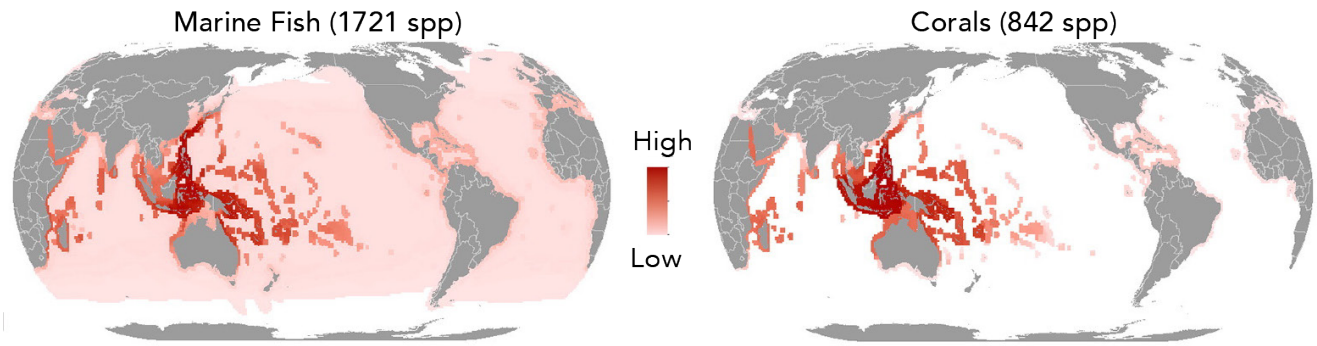



**Figure 1.1** — Global biodiversity distribution of marine fishes based on 1721 species (left) and corals based on 842 species (right). Modified from Jenkins and Van Houtan (2016), *Elsevier* license *5895520056135*.

(Grant and Grant, 2002). Sexual selection is another process that can generate biodiversity, operating independently of ecological divergence (natural selection), and instead driven by factors that are largely related to mate choice (Shuker and Kvarnemo, 2021). When both natural and sexual selection act on the same trait, it is called a magic trait (Wagner et al., 2012). Such traits are thought to play a significant role in fostering biodiversity, as reproductive isolation may develop more rapidly when a trait involved in mate choice is also subject to ecological selection (Thibert-Plante and Gavrilets, 2013). For example, in *Timema* stick insects, body color contributes to ecological adaptation through camouflage and to mating success through mate choice, bridging natural and sexual selection (Nosil et al., 2002; Comeault et al., 2016).

A bias in the study of evolutionary radiation is the longstanding underestimation of the emergence of new species in the ocean. This is surprising given that the ocean covers 70.8% of Earth's surface (Costello et al., 2013; Costello and Chaudhary, 2017), is thought to be where life itself originated (Martin et al., 2008; Lane and Xavier, 2024), and its vast seafloors represent the most extensive habitat on Earth (Snelgrove, 1999). As with terrestrial habitats, marine biodiversity hotspots are found in the tropics (Briggs, 2003; Bowen et al., 2013; Jenkins and Van Houtan, 2016). The Coral Triangle in the Indo-West Pacific is the major hotspot and is considered a marine center of diversity (Figure 1.1; Bowen et al., 2013). Other centers of biodiversity are located in the Caribbean, Antarctica and the northern Pacific, based on the high levels of endemism within these areas (Briggs, 2003). Because marine biodiversity is greater in these hotspots, they have a large influence on the global biodiversity (Briggs, 2003), highlighting the importance of understanding marine biodiversity to better comprehend global biodiversity.

Fragmented terrestrial habitats are thought to largely facilitate species diversification through reproductive isolation, however the ocean is a largely open area (Palumbi, 1994). Furthermore, in the complex life histories of marine organisms, there is often a pelagic larval stage, and this is believed to facilitate the oceans high connectivity (Kinlan and Gaines, 2003; Kokko and López-Sepulcre, 2006; Hernández et al., 2023). However, the paradigm of unlimited dispersal in the

ocean has been widely questioned (Barber et al., 2000) – we now know that ancient glacial events have repeatedly separated and merged marine habitats due to sea-level changes, meaning that fragmentation is also present in marine habitats (Palumbi, 1994). The closure of the Isthmus of Panama facilitated the Great American Interchange, while at the same time separating the eastern Pacific from the Caribbean, enabling the independent evolutionary trajectories of previously connected populations (Palumbi, 1994; Knowlton and Weigt, 1998). Because both biogeographic and reproductive barriers are not necessarily apparent in the ocean, many cryptic species have been mistakenly considered to be a single species (Victor, 2015). Consequently, the claim that one-third of the world's marine fish biodiversity is found on coral reefs is likely an underestimate, given the cryptic nature of many coral reef fish families, such as the *Gobiidae* (Brandl et al., 2018; Spalding and Grenfell, 1997; Robertson et al., 2022). Intrinsic biogeographic and habitat differences between terrestrial and marine environments mean that they differ in modes of reproductive isolation and evolution. Therefore, studying marine systems is crucial for a complete understanding of evolutionary radiation mechanisms.

**SUMMARY**

- *Biodiversity* is the variety of life forms on Earth
- In *adaptive radiations*, biodiversity is rapidly generated
- *Natural selection* acts on ecologically relevant traits
- *Sexual selection* acts on traits involved in mating success
- A *magic trait* is a phenotypic characteristic of a species that is under both natural and sexual selection
- ≧30% of the world's biodiversity is found in the ocean
- Intrinsic biogeographic and habitat differences between land and sea mean that they differ in modes of reproductive isolation and evolution

## 1.2 Species and speciation

Classifying biodiversity is a complex challenge, as it involves understanding diverse life forms and their genetic diversity in ever-evolving ecosystems. The diversity of individual organisms reveals differences in size, habitat choice, communication, reproduction, and other morphological, behavioral, and physiological traits. It is the relative similarities in these factors, especially morphology, that first led naturalists to group individuals together. Carl Linnaeus proposed a classification system with five ranks, from top to bottom: kingdom, class, order, genus and species (Linné, 1788). Later, Darwin proposed to tackle the species problem with an evolutionary model. In his view, species gradually accumulate small changes over time that may be favored by natural selection, meaning that populations eventually diverge to become two distinct species (Darwin's, 1859). His work inspired a great breadth of evolutionary studies in the XIX$^{th}$ and XX$^{th}$ centuries. Soon, concerns were raised about when to define different populations as different species, if the process, as Darwin described it, was gradual. Morphological classifications proved to be problematic, since sexual dissimilarities between females and males of the same species would lead

to them being classified as two separate species. This was the case for female and male cuttlefish, which differ drastically in size and body coloration, and were later reclassified as a single species by genetic approaches (Ponder et al., 2019). These early attempts to describe species and the dawn of the genetic era revealed the extent of the challenges in classifying biodiversity.

A multitude of species definitions and concepts currently exists, and no broad consensus has been reached. The Biological Species Concept (BSC) describes species as *"groups of interbreeding natural populations that are reproductively isolated from other such groups"* (Mayr, 1963). The Phylogenetic Species Concept, introduced by Cracraft (1983), relies on monophyly, where a species is defined as the smallest group of individuals that can be identified by a common pattern of ancestry and descent. The Evolutionary Species Concept defines species as *"a single lineage of ancestral descendant populations of organisms which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate"* (Simpson, 1961). The Ecological Species Concept defines a lineage as a species by its use of a distinct ecological niche, an adaptive zone in which the lineage evolves separately from others (Van Valen, 1976). In a recent review, Stankowski and Ravinet (2021b) show that different definitions are used depending on the field of research or the organisms studied. Among those who study genetics and genomics, the BSC prevails (Stankowski and Ravinet, 2021b). A shared aspect across all these definitions is that lineages diverge over the course of evolution, undergo distinct evolutionary processes, develop different ecological strategies, and exhibit morphological differences that eventually lead to complete reproductive isolation. This process of speciation, in which a single species splits into two distinct lineages, is particularly important for the study of evolutionary radiations and the origin of biodiversity units: species.

Seehausen et al. (2014) placed the BSC in a temporal context and described speciation as the *"origin of reproductive barriers among populations that permit the maintenance of genetic and phenotypic distinctiveness of these populations in geographical proximity"*. In these views, barriers that prevent the exchange of genetic material, so called gene flow, between two heterospecific individuals are central to achieve reproductive isolation. Barriers to gene flow can be caused by internal (intrinsic) or external (extrinsic) factors and can occur either before (prezygotic) or after (postzygotic) fertilization. Extrinsic prezygotic isolation mechanisms manifest as premating isolation, such as body color influencing mate choice in cichlids (Rometsch et al., 2020). Intrinsic prezygotic isolation may involve gametic incompatibility or isolation leading to removal of heterospecific sperm as observed in Trinidadian guppies, *Poecilia reticulata* (Ludlow and Magurran, 2006). Intrinsic postzygotic barriers manifest as hybrid inviability, reduced egg number or hybrid sterility (Coughlan and Matute, 2020; Haldane, 1922; Dobzhansky, 1936). Examples range from embryonic lethality in hybrid whitefish, *Coregonus clupeaformis* (Rogers and Bernatchez, 2006), to the disruptive sexual selection against F1 hybrids of *Heliconius melpomene* lineages (Naisbit et al., 2001). Extrinsic postzygotic

barriers result from environmental factors that affect the fitness of hybrids. These extrinsic barriers can interact with intrinsic postzygotic isolation, as shown by the accentuated negative effect of temperature on *Drosophila* hybrids carrying deleterious allelic combinations compared to pure strains (Coughlan and Matute, 2020; Miller and Matute, 2017).

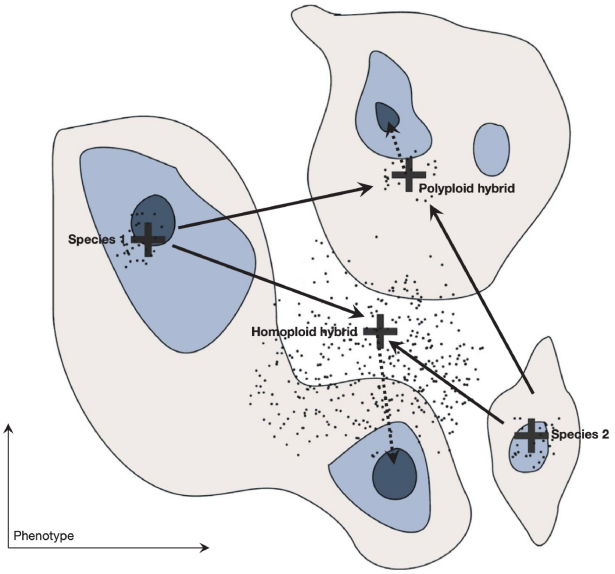



**Figure 1.2** — Hybrid speciation in an adaptive landscape. Optima are shown in blue, dots represent individuals. Hybrids that are far from phenotypic optima are represented by solid arrows, and hybrids that are close to an optima are represented by dotted arrows. Figure from Mallet (2007), *Springer Nature* license *5895520635167*.

Reproductive isolation can also depend on biogeography (Butlin et al., 2008). Allopatric speciation occurs when a geographic barrier physically separates incipient species and prevents gene flow between them, as for anole lizards that evolved separately on different Caribbean islands (Stroud and Losos, 2020) or for shrimps that speciated after the Isthmus of Panama land bridge formed (Alves et al., 2024). This mode of speciation, although important and once thought to be prevalent (Coyne and Orr, 2004), is not necessarily the most common (Bolnick and Fitzpatrick, 2007). Sympatric speciation, where reproductive isolation occurs without such geographic constraints, is an alternative mode of evolution (Bolnick and Fitzpatrick, 2007). Gene flow and hybridization processes can contribute to sympatric speciation, thus discrediting the view that complete reproductive isolation is the best definition of what a species is. Rather, it places hybridization and gene flow as a potential mechanism for evolutionary radiation (Figure 1.2; Mallet, 2001a, 2007).

Hybrid speciation provides a means of avoiding long neutral processes, the accumulation of mutations and drift, by providing excess ancestral diversity for selection to act upon (Marques et al., 2019). Hybrid speciation boosts genetic variation by providing new genotypes with the potential to colonize new ecological niches, therefore generating new species of hybrid origin (Figure 1.2; Kronforst et al., 2013). Hybrid speciation thus challenges the BSC. Another aspect of hybridization that contradicts the BSC is introgression, where hybrids do not form a new species, but rather backcross to their parental line, causing gene flow to blur the defining differences between two populations (Figure 1.3; Harrison and Larson, 2014). At early stages of speciation, interspecific gene flow contributes to the Incomplete Lineage Sorting (ILS) of radiating lineages (Figure 1.3; Mallet et al., 2016). Because ancestral allele polymorphisms persist in descendant species due to gene flow at early stages of divergence, the gene tree will not be congruent with the actual species tree (Figure 1.3). The difficulty in distinguishing between ILS and introgression (Figure 1.3) poses further challenges to species classification and understanding the origins

of biodiversity (Guerrero and Hahn, 2017; Cruickshank and Hahn, 2014). Overall, hybrid speciation sheds light on the intricate mechanisms that can occur during speciation, confronting the BSC and highlighting the role of gene flow in gradual speciation processes.

It is now recognized that reproductive isolation is achieved gradually over time and need not to be absolute. Stankowski and Ravinet (2021a) formally defined the speciation continuum as a continuum of reproductive isolation. The view that there is a gradient of differentiation between sister lineages through time is supported and makes it difficult to pinpoint the exact separation between two lineages (Shaw and Mullen, 2014). For example, evidence for multiple introgression events between spotted whiptail lizard lineages of the *A. gularis* complex complicates the understanding of their evolutionary history (Barley et al., 2024). This concept of continuity can also be extended to genomes and phenotypes (Bolnick et al., 2023). Evidence for gene tranfer between closely related *Drosophila* species shows that reproductive isolation varies along genomes (Wang et al., 1997). Some loci within the same genome can become genealogically distinct, either because they contribute directly to selection against hybrids, or because they are trapped in parts of the genome that are subjected to such selection (Wu, 2001). These regions have been described as islands of speciation in the genic view of species formulated by Wu (2001). The recurrence of a similar polymorphism across a species radiation raises questions about the mechanisms that maintain it throughout speciation and blurs the distinctiveness of different species (Jamie and Meier, 2020). Examples range from the shared major histocompatibility complex (MHC) antigens across all 70,000 vertebrate species (Figueroa et al., 1988; Piertney and Oliver, 2006), to the blotch pattern on females of five out of 15 species of Lake Victoria cichlid *Neochromis* spp. (Dijkstra et al., 2008), and the body chiral dimorphism in 30 out of 35 species of Southeast Asian camaenid tree snails of the subgenus *Amphidromus* (Schilthuizen et al., 2007; Jamie and Meier, 2020). This multivariate gradient of speciation is now referred to as the speciation hypercube (Bolnick et al., 2023), highlighting the importance of combinatorial approaches to understanding evolutionary processes behind the origin of species (Mallet, 2001b). When examining evolutionary radiations, one can identify genotypic clusters across multiple loci or phenotypic clusters to determine which substrates selection is acting on. This approach also helps determine whether the radiation is in its early or late stages (Mallet, 2001b).

**SUMMARY**

- The *BSC*'s view of complete reproductive isolation has been challenged
- *Speciation* is the process by which two lineages become distinct over evolutionary time
- *Gene flow* is the exchange of genetic material between groups
- *Hybrid speciation* is an alternative to neutral processes for the rapid creation of diversity
- *Speciation continuum* views speciation as a gradual process rather than a discrete event
- The *genic view of species* suggests that few genomic regions underlie

species identity at early stages of divergence

## 1.3 Genomics and phenomics in evolution

Evolutionary processes involved in speciation leave detectable footprints in the genomes (Tajima, 1989). Evaluating such patterns at the DNA level through statistical inference can reveal which evolutionary or demographic forces are acting on populations (Gagnaire, 2020). Molecular markers such as mitochondrial DNA and microsatellites have made it possible to measure differentiation and dispersal during speciation (Salzburger et al., 2002; Natoli et al., 2004). However, these methods are limited to a few loci and provide less information about variation across the genome. In contrast, RAD-seq (restriction-site-associated DNA sequencing) and transcriptome sequencing have identified thousands of genomic loci and Single Nucleotide Polymorphisms (SNPs). The large amount of genome-wide data from these techniques has enabled the description of genome semipermeability, highlighting variation in genetic differentiation, divergence, and gene flow (Hohenlohe et al., 2010; Gagnaire et al., 2013). Both the advent of next-generation sequencing and the drop in sequencing costs marked the shift from genetics to genomics. The Human Genome Project (Lander, 2001) and the sequencing of medical model organisms, such as *Drosophila* (Adams et al., 2000) and mice (Waterston et al., 2002), paved the way for genome assembly and re-sequencing projects. These technologies were first applied to model systems in speciation research, such as Darwin's finches (Lamichhaney et al., 2015) and the African cichlids (Hulsey and Renn, 2009), and later became available to other study systems, including *Heliconius* butterflies (Martin et al., 2013), the three-spined stickleback *Gasterosteus* (Ravinet et al., 2018) and *Helianthus* sunflowers (Todesco et al., 2020). Studies of reproductive isolation and genome semipermeability in these taxa, along with the analysis of contiguous haplotype blocks, provided insights

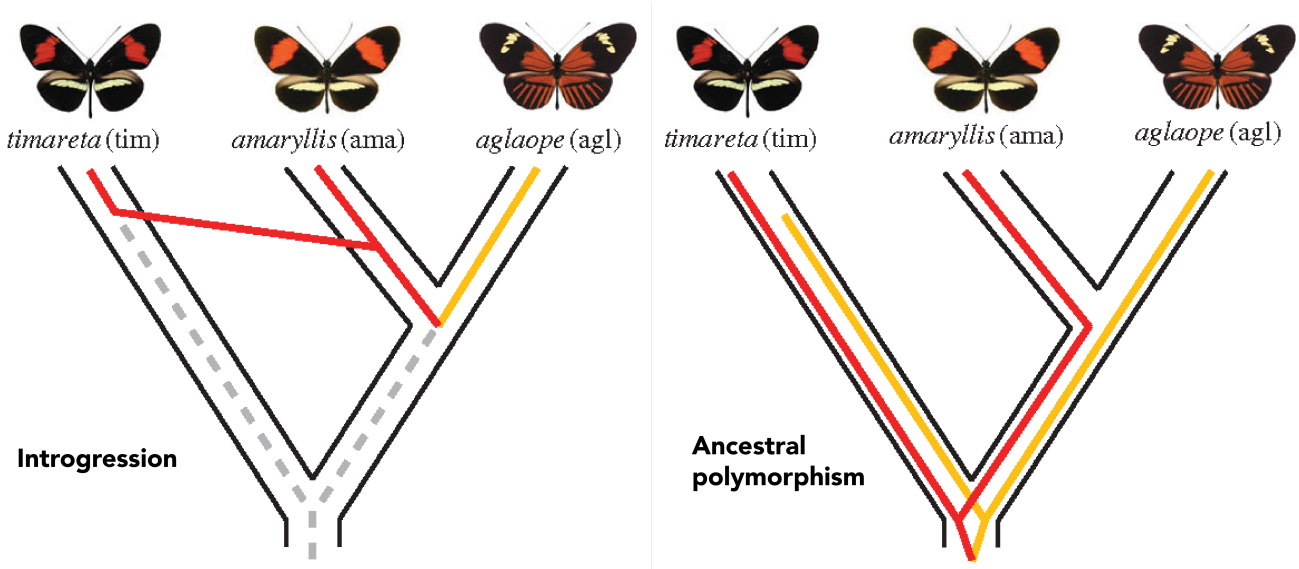



**Figure 1.3 —** Introgression (left) and ILS (right) in *Heliconius* butterflies. Modified from Smith and Kronforst (2013), *The Royal Society (UK.)* license *1539377-1*.

into introgression, demography, and selection (Duranton et al., 2018; Harris and Nielsen, 2013; Garud et al., 2021). Such genome-scale information provides valuable insights into the mechanisms that interact to facilitate evolutionary radiations and the origins of biodiversity.

A major challenge in studying speciation is that it occurs over long time frames, making it impossible to observe an entire speciation process from its onset to the achievement of complete reproductive isolation. Due to the difficulty of examining all stages along the speciation continuum, past evolutionary processes are often inferred from current biodiversity patterns and their phylogenetic relationships (Stroud and Losos, 2020). Typically, well-established radiations are studied and historical processes are inferred from DNA sequences of their extant species (Stroud and Losos, 2020). For example, Brawand et al. (2014) used present-day cichlid species to understand how neutral (non-adaptive) and adaptive processes contributed to their extraordinary diversity. They showed that neutral processes were crucial in amassing the genomic variation that selection subsequently sorted, and that the interaction of both processes was likely necessary to generate many and diverse new species in a very short time at the onset of the cichlid radiation (Brawand et al., 2014). An alternative approach to studying radiations is to examine pairs of species that represent different stages along the speciation continuum and reflect various levels of reproductive isolation (Kisel and Barraclough, 2010; Shaw and Gillespie, 2016). By assessing demographic and selective pressures or quantifying reproductive barriers across these species, it becomes possible to infer the evolutionary forces driving speciation. This approach focuses on more recent radiations that may still be undergoing differentiation, providing a window into the early processes where natural and sexual selection can be observed *"in action"*. Zhou et al. (2018) assembled a genome of finless porpoises (*Neophocaena* spp.) and re-sequenced 48 individuals. They showed that natural selection acting on osmotic stress responses to oceans and rivers reproductively isolated the Yangtze finless porpoise from other porpoise populations. Unique genetic adaptations define the Yangtze finless porpoise as a distinct species (Zhou et al., 2018). It is therefore necessary to look for DNA signatures at both macro- and microevolutionary scales to understand the drivers and processes that shape speciation and evolutionary radiations.

Evolutionary radiations can also be characterized by remarkable phenotypic diversification, and its eco-evolutionary significance has been well documented (Kronforst and Papa, 2015). Scientists have always used images such as drawings, paintings, and photographs to record and quantify the shapes, patterns, and colors of life (Brumm et al., 2021; Meriam, 2017; Robertson and Van Tassell, 2019). With the advent of digital imaging, an incredible amount of phenotypic data has been accumulated, but challenges in the analysis capacity have limited the outputs (Lürig et al., 2021; Akkaynak et al., 2014). Manual image analysis is difficult to reproduce and time consuming, typically measures only a few traits at a time, and is subject to human bias (Lürig et al., 2021). Recent improvements in computer vision have accelerated the shift from phenetics to phenomics: the multivariate and comprehen-

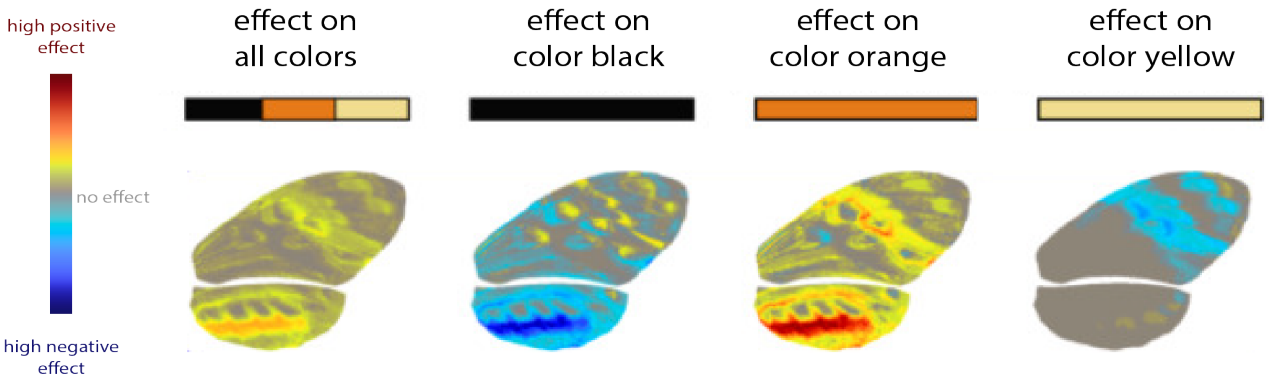



**Figure 1.4** — Effect of a genetic locus on wing color variation in *Heliconius* butterflies. Modified from Jay et al. (2022), *The Royal Society (UK.)* license *1539379-1*.

sive study of phenotypes (Houle et al., 2010). Computer vision refers to modern methods of both digital image acquisition and its processing (Weinstein, 2018). With computer vision, we can automatically extract meaningful information from images and thus collect phenotypic information on a very large scale (Houle et al., 2010; Lürig et al., 2021). Pixel data is high-dimensional, continuous, and captures the entire phenotypic variation, which is valuable for understanding the interplay between phenotypes, their genetic underpinnings, and the environment (Figure 1.4; Jiggins et al., 2017; Cuthill et al., 2017). Image analysis has been used to quantify external and internal phenotypes, behaviors, and biochemical properties at the organism level (Houle et al., 2010). Artificial Intelligence (AI), particularly computer vision, machine learning, and deep learning, has been successfully integrated with non-invasive imaging techniques to enable sensor-based monitoring and taxonomic identification (Høye et al., 2021; Guisande et al., 2010; Wäldchen and Mäder, 2018).

Phenotypes in evolution have been investigated for a wide range of organisms (plants, insects, fish, humans) and at both the macro- and microevolutionary scales (Houle et al., 2010). Navalón et al. (2020) showed that a tight beak-skull morphology is linked to the rapid diversification of Hawaiian honeycreepers and Darwin's finches, compared to other land birds. In more recent radiations, specific traits such as the jaw morphology, body coloration, and blotch patterning of cichlids have been studied (Darrin Hulsey, 2006; Urban et al., 2022; Gerwin et al., 2021; Streelman et al., 2003). A growing number of studies focus on color patterns because they are among the most diverse and variable traits in nature (Cuthill et al., 2017). In highly visual environments such as coral reefs, the importance of color patterns is amplified, as evidenced by the stunning variation in color patterns observed in reef fishes. This variation in color patterns appears to be central to the recent diversification of reef fishes, with the majority of species having emerged within the last 5.3 million years and closely related species differing primarily in color patterns (Bellwood et al., 2015, 2017). Color patterns are involved in camouflage and mimicry behaviors (Phillips et al., 2017; Randall, 2005), mate choice (McMillan et al., 1999), and vision (Gruber et al., 2016), which are likely to influence speciation dynamics by being under direct natural and sexual selection pressures (Cortesi et al., 2015; Salis et al., 2019; Puebla et al.,

2007; Hench et al., 2022). The analysis of reef fish color patterns provides a unique opportunity to address ecological and evolutionary questions, especially when considered in the context of evolutionary radiations.

**SUMMARY**

- *Genomics* developed with the advent of next-generation sequencing
- *Macroevolution* is the study of evolution above species level
- *Microevolution* studies evolution at the species or population level
- *Phenomics* is the multivariate and comprehensive study of phenotypes
- *Color pattern* is a multivariate biologically relevant phenotype, especially in reef fishes that inhabit highly visual coral reef environments

## 1.4 Hamlets

The genus *Hypoplectrus* provides a key case study of a recent reef fish radiation, characterized by remarkable phenotypic diversity in color patterns. This group of coral reef fishes is found in the wider Caribbean, with a distribution range extending from Tobago through the entire Caribbean and Gulf of Mexico to Florida, and further on to Bermuda (Figure 1.5; Robertson and Van Tassell, 2019). There are currently 18 described species in this genus, most of which have been recognized in the last 15 years (Figure 1.6; Flores et al., 2011; Lobel, 2011; Victor, 2012; Taveral and Acero, 2013; Puebla et al., 2022). Some species are endemic to specific regions, e.g. *H. liberte* is found solely in Haiti's northeastern coast (Victor and Marks, 2018), while *H. maya* is found only on the reef barrier of Belize (Lobel, 2011). Some species such as *H. atlahua*, *H. castroaguirrei* and *H. ecosur* are found only in the Gulf of Mexico. However, most hamlet species span the entire distribution range and up to nine species can be found co-occurring on the same reefs (Thresher, 1978; Puebla et al., 2012). Because of this wide sympatry and the fact that hamlets have small differences in ecology, morphology, and early genetic markers, their status as a single or separate species has long been debated (Randall and Randall, 1960; Thresher, 1978; Graves and Rosenblatt, 1980). The existence of their very diverse color variants has been acknowledged since the very first hamlet description:

> "...we find ourselves driven to the conclusion that all the common forms of *Hypoplectrus* probably constitute but a single species, subject to almost endless variations in color."
>
> Jordan (1890)

The different hamlets are easily distinguished by their distinctive color patterns (Puebla et al., 2022). The color patterns of hamlets are biologically relevant and are involved in both their predatory and mate choice behavior (Randall and Randall, 1960; Puebla et al., 2007, 2018). It has therefore been suggested that color pattern in hamlets is a magic trait subject to natural and sexual selection that drives the hamlet radiation (Puebla et al., 2007; Hench et al., 2019, 2022).

Hamlets are predatory and solitary coral reef fishes. They feed primarily on benthic invertebrates at shallow to mid-depths. Some non-significant differences in diet

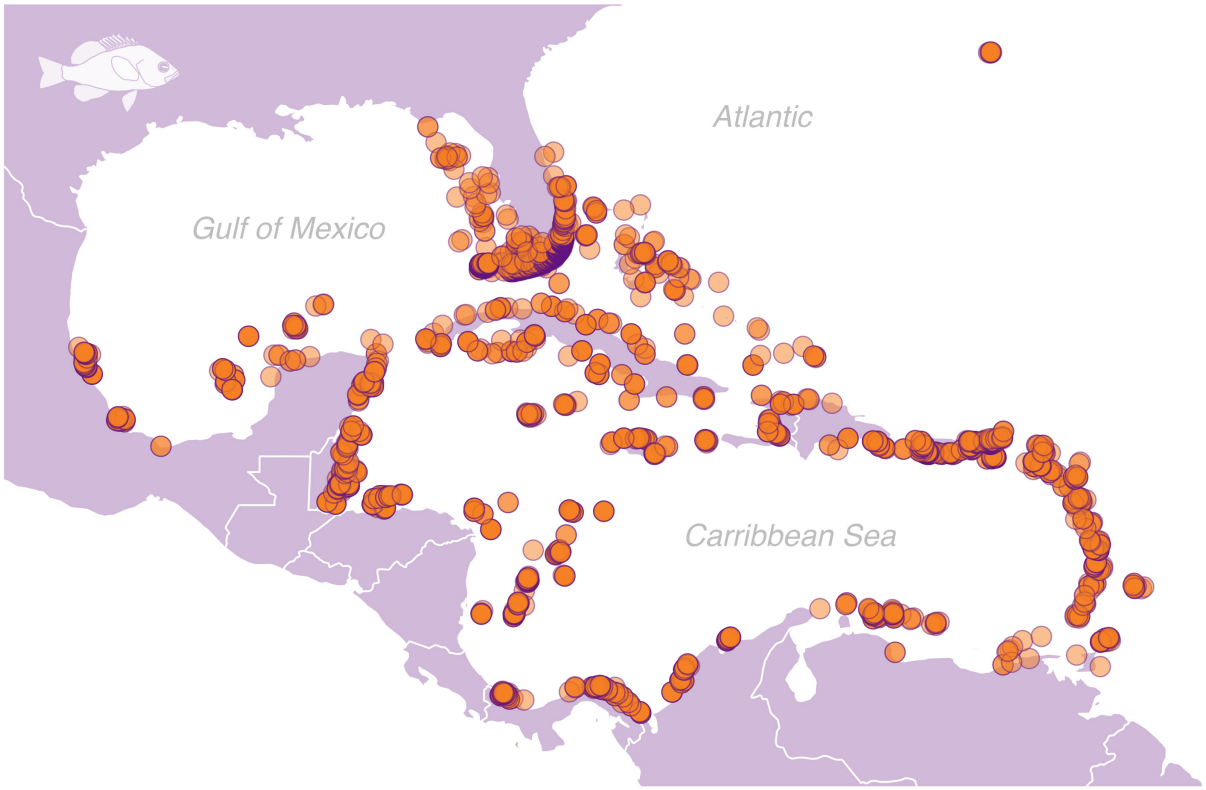



**Figure 1.5** — Distribution range of hamlets. Each point is an individual hamlet with a total of 6,424 records. Data from Robertson and Van Tassell (2019).

are observed, notably with *H. indigo*, which is the only hamlet that eats *Chromis cyanea* (Whiteman et al., 2007; Holt et al., 2008). Several hamlet species are considered to be aggressive mimics, increasing their predation success by imitating a more abundant and non-predatory fish species (Thresher, 1978; Randall and Randall, 1960; Puebla et al., 2007). Putative sympatric models have been suggested for the blue hamlet *H. gemma* and the butter hamlet *H. unicolor* (Randall and Randall, 1960; Puebla et al., 2007). Increased predation success was only demonstrated for *H. unicolor* (Puebla et al., 2007, 2018). The aggressive mimicry hypothesis involves both color patterns and behavior suggesting that these traits are under natural selection in hamlets. However, while this may be true for some hamlet species, this hypothesis cannot be extended to the entire radiation. Another axis for natural selection on hamlet color patterns would be camouflage, whereby disruptive vertical body bars would promote advantages in complex reef habitats (Phillips et al., 2017). This second hypothesis cannot be extended to the whole radiation either, since not all hamlets have bars. The complex reef habitat offers a multitude of ecological niches for fish to adapt to, and no single ecological driver appears to be solely responsible for the hamlet radiation.

While the role of natural selection in the hamlet radiation is uncertain, sexual selection on color pattern has been documented. Hamlets are simultaneously hermaphrodites and egg traders, meaning that mate choice must be reciprocal (Fischer, 1980). Each day at dawn, hamlets find and choose a reciprocal partner on the reef to mate with. This, combined with the widespread sympatry of hamlet species, leads to complex pairing dynamics (Puebla

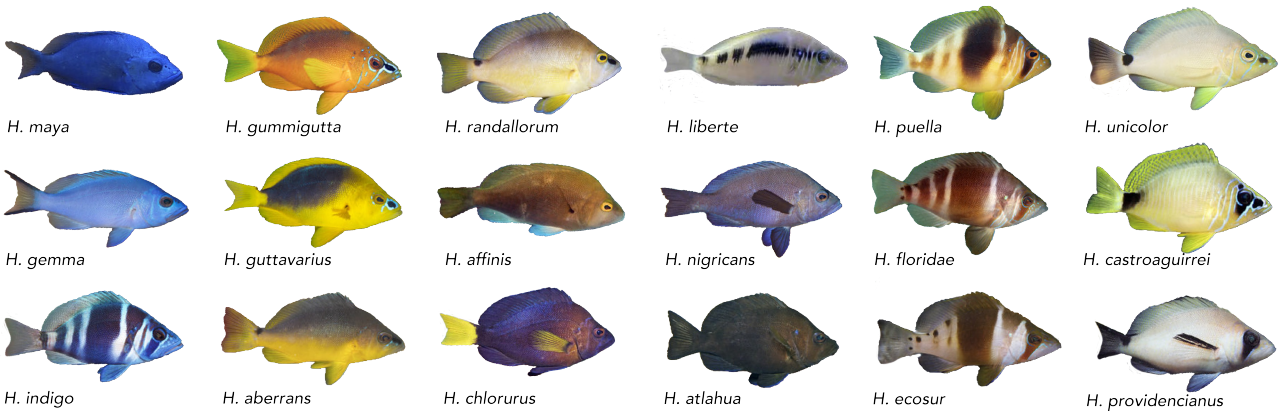



**Figure 1.6** — The 18 described hamlet species. Modified from Puebla et al. (2022) with authors' permission.

et al., 2012). Mate choice in hamlets is highly assortative, visually based, and may act as a prezygotic barrier to reproduction, limiting gene flow between species (Puebla et al., 2007). However, field surveys have shown that hybrids represent no more than 2.05% of observations, suggesting a low frequency of non-assortative matings in hamlets (Domeier, 1994). Hamlet hybrids have also been bred in captivity (Domeier, 1994). Puebla et al. (2012) showed that although mating with conspecifics is preferred, hybridization is possible when conspecifics are not available. Although sexual selection on color pattern limits gene flow through mate choice, it is likely that hamlets are not completely reproductively isolated.

In addition to the direct hybrid observations, genetic analyses have identified hybrids and backcrosses in natural populations, confirming that hybridization and introgression are ongoing in hamlets (Hench et al., 2019, 2022). Hamlet species are highly genetically similar, congruent with the occurrence of gene flow. The very low levels of genetic differentiation and divergence among hamlet species fall within a range typically found among populations within species (Garcia-Machado et al., 2004; Puebla et al., 2007; Barreto and McCartney, 2008; Puebla et al., 2008, 2012, 2014). They also generally do not sort into distinct mitochondrial haplogroups (McCartney et al., 2003). Sharp peaks of differentiation stand out against a background of low differentiation, and comparisons between species pairs reveal varying degrees of divergence (Puebla et al., 2012; Hench et al., 2019, 2022). Both the widespread hybridization and the pattern of differentiation among hamlets are typical of early radiation stages (Figure 1.7; Wu, 2001).

While I will adopt the view that hamlets are separate species for this work, I recognize that hamlets may portray an early window of the speciation continuum (Figure 1.7; Puebla et al., 2012). Indeed, the hamlet radiation is estimated to have begun diverging very recently, $< 10,000$ generations (Hench et al., 2022), and remains largely unresolved phylogenetically (McCartney et al., 2003; Ramon et al., 2003; Garcia-Machado et al., 2004; Taveral and Acero, 2013; Hench et al., 2022). Hamlet fishes, with their diverse color patterns and their recent origin, provide a unique opportunity to study substrates for selection during the early stages of a radiation.

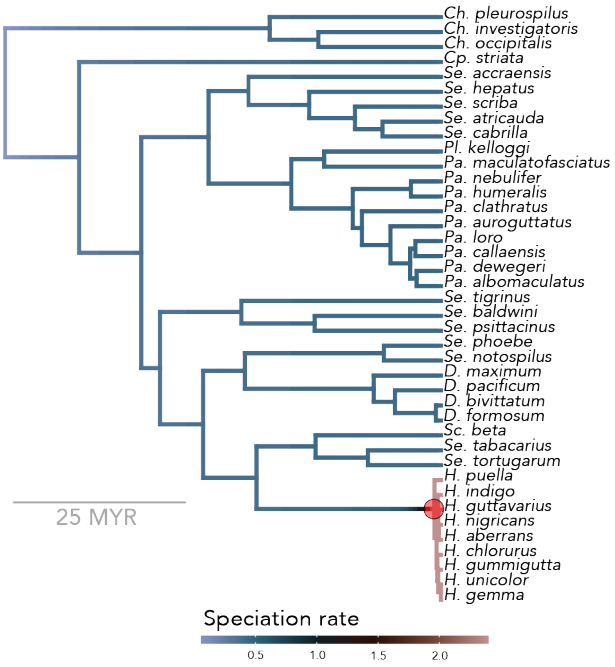



**Figure 1.7** — Serraninae subfamily tree reconstructed from the Fish Tree of Life (Rabosky et al., 2018). Short and red branches indicate that hamlets underwent a recent diversification burst not observed in related lineages. Modified from Hench et al. (2022), license *CC BY-NC-ND 4.0*.

**SUMMARY**

- *Hypolectrus* is a reef fish radiation
- Hamlets have striking variation in their color patterns
- Color pattern in hamlet is potentially a magic trait, strongly involved in sexual selection
- The hamlet radiation is recent and rather at the beginning of the speciation continuum

## 1.5 Thesis aims and outline

This thesis expands both genomic and phenomic hamlet data from the local scale to the scale of their entire geographic range. It also presents the first comprehensive and standardized dataset of hamlet photographs, marking a shift from human-biased qualitative phenotypic analysis to quantitative phenotypic analysis. It thus further supports hamlets as a model system to investigate speciation mechanisms at the microevolutionary scale. The main objective of my thesis was to characterize the phenotypic and genetic variation of an evolutionary radiation. The incredible color pattern diversity observed in the hamlet radiation, coupled with its recent origin, provides an ideal foundation for this investigation. In this context, I analyzed color pattern variation, uncovered its genomic basis, and evaluated its role in reproductive isolation between hamlet species. I first developed a quantitative pipeline for color pattern analysis from *in situ* photographs, quantified color pattern variation across hamlet species, and identified genes underlying such variation. I then assessed interspecific and intraspecific phenotypic and genomic variation to investigate the substrates of selection during the early stages of a radiation. Finally, I examined the hamlet radiation in a phylogenetic context, using both whole genomes and regions associated with phenotypic variation, according to the genic view of species.

The work within this thesis consists of the following chapters:

**Manuscript 1: Phenotypic and genomic dissection of color pattern variation in a reef fish radiation.**

**Floriane Coulmance**, Derya Akkaynak, Yann Le Poul, Marc P. Höppner, W. Owen McMillan, Oscar Puebla

The first manuscript quantifies color pattern variation in hamlets and identifies its genomic basis. I developed and implemented a novel quantitative pipeline for the analysis of *in situ* fish photographs. By combining this newly developed method

with a multivariate genome-wide association approach, I was able to identify three highly associated regions underlying hamlet phenotypic variation. This method also allows modeling the effect of specific loci on the hamlet body at a pixel resolution. The presented framework provides a quantitative support for the study of the eco-evolutionary significance of color pattern in reef fishes.

This work is published:

Coulmance, F., Akkaynak, D., Le Poul, Y., Höppner, M. P., McMillan, W. O. and Puebla, O. (2024). Phenotypic and genomic dissection of colour pattern variation in a reef fish radiation. Molecular Ecology *33*, e17047.

### Manuscript 2: Phenotypic variation within and between species during the early stages of radiation

**Floriane Coulmance**, Melanie J. Heckwolf, Jakob Gismann, Tane Kafle, Omar Domínguez-Domínguez, Karim Awhida, Martin Helmkampf, W. Owen McMillan, Oscar Puebla.

The second manuscript introduces the most extensive phenotypic dataset to date for hamlets, with 571 *in situ* photographs collected over seven fieldwork efforts, representing seven sites across their entire distribution range and 14 hamlet species. I leveraged the previously developed quantitative image analysis pipeline to investigate interspecific and intraspecific phenotypic variation. Combining this with genomic data from 327 hamlet individuals representing 18+ species, I showed that phenotypic and genomic clusters of local sympatric species show a substantial amount of overlap and become largely continuous at the scale of the whole radiation, highlighting the importance of local evolutionary processes in shaping radiations.

Work *in prep*, planned submission to:

Coulmance, F., Heckwolf, M. J., Gismann, J., Kafle, T., Domínguez-Domínguez, O., Awhida, K., Helmkampf, M., McMillan, W. O. and Puebla, O. *(in prep)*. Phenotypic variation within and between species during the early stages of radiation. Proceedings of the Royal Society B: Biological Sciences.

### Manuscript 3: Radiation with reproductive isolation in the near-absence of phylogenetic signal

**Floriane Coulmance***, Martin Helmkampf*, Melanie J. Heckwolf, Arturo Acero P., Alice Balard, Iliana Bista, Omar Domínguez-Domínguez, Paul B. Frandsen, Montserrat Torres-Oliva, Aintzane Santaquiteria, Jose Tavera, Benjamin C. Victor, D. Ross Robertson, Ricardo Betancur-R., W. Owen McMillan, Oscar Puebla

This manuscript explores the hamlet radiation from a phylogenetic perspective. Using a genomic dataset of 335 genomes from 18+ hamlet species, sampled across their entire geographic range, the first comprehensive phylogeny of the hamlet was constructed. I assessed the genic view of species by examining phylogenetic relationships at genomic regions underlying species divergence and color pattern variation. Phylogenies constructed using these regions or the whole genome failed to resolve the hamlet phylogeny at the species level, suggesting that reproductive isolation may unfold without a detectable genomic signature at early

stages.

Work *in review* in:

Coulmance, F.*, Helmkampf, M.*, Heckwolf, M. J., Acero, P. A., Balard, A., Bista, I., Domínguez-Domínguez, O., Frandsen, P. B., Torres-Oliva, M., Santaquiteria, A., Tavera, J., Victor, B. C., Robertson, D. R., Betancur-R., R., McMillan, W. O. and Puebla, O. *(in review)*. Radiation with reproductive isolation in the near-absence of phylogenetic signal. Science Advances.

* shared first authorship

## References


Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F. et al. (2000). The genome sequence of Drosophila melanogaster. Science *287*, 2185–2195.

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

Alves, D. F. R., Martinez, P. A., de Barros-Alves, S. d. P., de Almeida, A. C., Elias, M. A. M., Hurt, C. and Hultgren, K. M. (2024). Climatic niche evolution and speciation modes in the transisthmian Alpheus shrimps (Caridea: Alpheidae). Journal of Biogeography *51*, 1944–1957.

Barber, P. H., Palumbi, S. R., Erdmann, M. V. and Moosa, M. K. (2000). A marine Wallace's line? Nature *406*, 692–693.

Barbier, E. B., Hacker, S. D., Kennedy, C., Koch, E. W., Stier, A. C. and Silliman, B. R. (2011). The value of estuarine and coastal ecosystem services. Ecological monographs *81*, 169–193.

Barley, A. J., Nieto-Montes de Oca, A., Manríquez-Morán, N. L. and Thomson, R. C. (2024). Understanding Species Boundaries that Arise from Complex Histories: Gene Flow Across the Speciation Continuum in the Spotted Whiptail Lizards. Systematic Biology , syae040.

Barreto, F. S. and 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. and 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. and Mora, C. (2015). The evolution of fishes on coral reefs: fossils, phylogenies and functions. Ecology of fishes on coral reefs *6*, 55–63.

Bolnick, D. I. and Fitzpatrick, B. M. (2007). Sympatric speciation: models and empirical evidence. Annu. Rev. Ecol. Evol. Syst. *38*, 459–487.

Bolnick, D. I., Hund, A. K., Nosil, P., Peng, F., Ravinet, M., Stankowski, S., Subramanian, S., Wolf, J. B. and Yukilevich, R. (2023). A multivariate view of the speciation continuum. Evolution *77*, 318–328.

Bowen, B. W., Rocha, L. A., Toonen, R. J. and Karl, S. A. (2013). The origins of tropical marine biodiversity. Trends in ecology & evolution *28*, 359–366.

Brandl, S. J., Goatley, C. H., Bellwood, D. R. and Tornabene, L. (2018). The hidden half: ecology and evolution of cryptobenthic fishes on coral reefs. Biological Reviews *93*, 1846–1873.

Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A. Y., Lim, Z. W., Bezault, E. et al. (2014). The genomic substrate for adaptive radiation in African cichlid fish. Nature *513*, 375–381.

Briggs, J. C. (2003). Guest editorial: marine centres of origin as evolutionary engines. Journal of Biogeography *30*, 1–18.

Brumm, A., Oktaviana, A. A., Burhan, B., Hakim, B., Lebe, R., Zhao, J.-x., Sulistyarto, P. H., Ririmasse, M., Adhityatama, S., Sumantri, I. et al. (2021). Oldest cave art found in Sulawesi. Science Advances *7*, eabd4648.

Butlin, R. K., Galindo, J. and Grahame, J. W. (2008). Sympatric, parapatric or allopatric: the most important way to classify speciation? Philosophical Transactions of the Royal Society B: Biological Sciences *363*, 2997–3007.

Comeault, A. A., Carvalho, C. F., Dennis, S., Soria-Carrasco, V. and Nosil, P. (2016). Color phenotypes are under similar genetic control in two distantly related species of Timema stick insect. Evolution *70*, 1283–1296.

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. and Cheney, K. L. (2015). Phenotypic plasticity confers multiple fitness benefits to a mimic. Current Biology *25*, 949–954.

Costello, M. J., Bouchet, P., Boxshall, G., Fauchald, K., Gordon, D., Hoeksema, B. W., Poore, G. C., van Soest, R. W., Stöhr, S., Walter, T. C. et al. (2013). Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. PloS one *8*, e51629.

Costello, M. J. and Chaudhary, C. (2017). Marine biodiversity, biogeography, deep-sea gradients, and conservation. Current Biology *27*, R511–R527.

Coughlan, J. M. and Matute, D. R. (2020). The importance of intrinsic postzygotic barriers throughout the speciation process. Philosophical Transactions of the Royal Society B *375*, 20190533.

Coyne, J. and Orr, H. (2004). Speciation. 2004 Sunderland. MA: Sinauer Associates .

Cracraft, J. (1983). Species concepts and speciation analysis. Current ornithology *1*, 159–187.

Cruickshank, T. E. and Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular ecology *23*, 3133–3157.

Cuthill, I. C., Allen, W. L., Arbuckle, K., Caspers, B., Chaplin, G., Hauber, M. E., Hill, G. E., Jablonski, N. G., Jiggins, C. D., Kelber, A. et al. (2017). The biology of color. Science *357*, eaan0221.

Darrin Hulsey, C. (2006). Function of a key morphological innovation: fusion of the cichlid pharyngeal jaw. Proceedings of the Royal Society B: Biological Sciences *273*, 669–675.

Darwin's, C. (1859). On the origin of species. published on *24*, 1.

Dijkstra, P. D., Seehausen, O. and Groothuis, T. G. (2008). Intrasexual competition among females and the stabilization of a conspicuous colour polymorphism in a Lake Victoria cichlid fish. Proceedings of the Royal Society B: Biological Sciences *275*, 519–526.

Dobzhansky, T. (1936). Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. Genetics *21*, 113.

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

Duranton, M., Allal, F., Fraïsse, C., Bierne, N., Bonhomme, F. and Gagnaire, P.-A. (2018). The origin and remolding of genomic islands of differentiation in the European sea bass. Nature communications *9*, 2518.

Figueroa, F., Günther, E. and Klein, J. (1988). MHC polymorphism pre-dating speciation. Nature *335*, 265–267.

Fischer, E. A. (1980). The relationship between mating system and simultaneous hermaphroditism in the coral reef fish, Hypoplectrus nigricans (Serranidae). Animal behaviour *28*, 620–633.

Flores, L. F. D. M., Tello-Musi, J. L. and Martínez-Pérez, J. A. (2011). Descripción de una nueva especie del género Hypoplectrus (Actinopterigy: Serranidae) del Sistema Arrecifal Veracruzano, suroeste del Golfo de México. Revista de Zoología *22*, 1–10.

Gagnaire, P.-A. (2020). Comparative genomics approach to evolutionary process connectivity. Evolutionary Applications *13*, 1320–1334.

Gagnaire, P.-A., Pavey, S. A., Normandeau, E. and Bernatchez, L. (2013). The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution *67*, 2483–2497.

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

Garud, N. R., Messer, P. W. and Petrov, D. A. (2021). Detection of hard and soft selective sweeps from Drosophila melanogaster population genomic data. PLoS Genetics *17*, e1009373.

Gerwin, J., Urban, S., Meyer, A. and Kratochwil, C. F. (2021). Of bars and stripes: A Malawi cichlid hybrid cross provides insights into genetic modularity and evolution of modifier loci underlying colour pattern diversification. Molecular Ecology *30*, 4789–4803.

Grant, P. R. and Grant, B. R. (2002). Adaptive radiation of Darwin's finches: Recent data help explain how this famous group of Galapagos birds evolved, although gaps in our understanding remain. American scientist *90*, 130–139.

Graves, J. E. and Rosenblatt, R. H. (1980). Genetic relationships of the color morphs of the serranid fish Hypoplectrus unicolor. Evolution *34*, 240–245.

Gruber, D. F., Loew, E. R., Deheyn, D. D., Akkaynak, D., Gaffney, J. P., Smith, W. L., Davis, M. P., Stern, J. H., Pieribone, V. A. and Sparks, J. S. (2016). Biofluorescence in catsharks (Scyliorhinidae): fundamental description and relevance for elasmobranch visual ecology. Scientific Reports *6*, 24751.

Guerrero, R. F. and Hahn, M. W. (2017). Speciation as a sieve for ancestral polymorphism.

Guisande, C., Manjarrés-Hernández, A., Pelayo-Villamil, P., Granado-Lorencio, C., Riveiro, I., Acuña, A., Prieto-Piraquive, E., Janeiro, E., Matías, J., Patti, C. et al. (2010). IPez: an expert system for the taxonomic identification of fishes based on machine learning techniques. Fisheries Research *102*, 240–247.

Haldane, J. B. (1922). Sex ratio and unisexual sterility in hybrid animals. Journal of genetics *12*, 101–109.

Harris, K. and Nielsen, R. (2013). Inferring demographic history from a spectrum of shared haplotype lengths. PLoS genetics *9*, e1003521.

Harrison, R. G. and Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. Journal of Heredity *105*, 795–809.

Hench, K., Helmkampf, M., McMillan, W. O. and Puebla, O. (2022). Rapid radiation in a highly diverse marine environment. Proceedings of the National Academy of Sciences *119*, e2020457119.

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

Hernández, C. M., Paris, C. B., Vaz, A. C., Jones, B. T., Kellner, J. B., Richardson, D. E., Sponaugle, S., Cowen, R. K. and Llopiz, J. K. (2023). Diverse patterns of larval coral reef fish vertical distribution and consequences for dispersal and connectivity. Coral Reefs *42*, 453–465.

Hernández-Hernández, T. and Wiens, J. J. (2020). Why are there so many flowering plants? A multiscale analysis of plant diversification. The American Naturalist *195*, 948–963.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A. and Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS genetics *6*, e1000862.

Holt, B. G., Emerson, B. C., Newton, J., Gage, M. J. and 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.

Houle, D., Govindaraju, D. R. and Omholt, S. (2010). Phenomics: the next challenge. Nature reviews genetics *11*, 855–866.

Høye, T. T., Ärje, J., Bjerge, K., Hansen, O. L., Iosifidis, A., Leese, F., Mann, H. M., Meissner, K., Melvad, C. and Raitoharju, J. (2021). Deep learning and computer vision will transform entomology. Proceedings of the National Academy of Sciences *118*, e2002545117.

Hulsey, C. D. and Renn, S. C. (2009). Genomics and vertebrate adaptive radiation: a celebration of the first cichlid genome. Integrative and comparative biology *49*, 613–617.

Jamie, G. A. and Meier, J. I. (2020). The persistence of polymorphisms across species radiations. Trends in Ecology & Evolution *35*, 795–808.

Jay, P., Leroy, M., Le Poul, Y., Whibley, A., Arias, M., Chouteau, M. and 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.

Jenkins, C. N. and Van Houtan, K. S. (2016). Global and regional priorities for marine biodiversity protection. Biological Conservation *204*, 333–339.

Jiggins, C. D., Wallbank, R. W. and Hanly, J. J. (2017). Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns? Philosophical Transactions of the Royal Society B: Biological Sciences *372*, 20150485.

Jordan, D. S. (1890). A review of the genera and species of Serranidae found in the waters of America and Europe. US Government Printing Office.

Kinlan, B. P. and Gaines, S. D. (2003). Propagule dispersal in marine and terrestrial environments: a community perspective. Ecology *84*, 2007–2020.

Kisel, Y. and Barraclough, T. G. (2010). Speciation has a spatial scale that depends on levels of gene flow. The American Naturalist *175*, 316–334.

Knowlton, N. and Weigt, L. A. (1998). New dates and new rates for divergence across the Isthmus of Panama. Proceedings of the Royal Society of London. Series B: Biological Sciences *265*, 2257–2263.

Kokko, H. and López-Sepulcre, A. (2006). From individual dispersal to species ranges: perspectives for a changing world. Science *313*, 789–791.

Kronforst, M. R., Hansen, M. E., Crawford, N. G., Gallant, J. R., Zhang, W., Kulathinal, R. J., Kapan, D. D. and Mullen, S. P. (2013). Hybridization reveals the evolving genomic architecture of speciation. Cell reports *5*, 666–677.

Kronforst, M. R. and Papa, R. (2015). The functional basis of wing patterning in Heliconius butterflies: the molecules behind mimicry. Genetics *200*, 1–19.

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N. et al. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature *518*, 371–375.

Lander, E. S. (2001). Initial sequencing and analysis of the human germane. Nature *409*, 860–921.

Lane, N. and Xavier, J. C. (2024). To unravel the origin of life, treat findings as pieces of a bigger puzzle. Nature *626*, 948–951.

Linné, C. v. (1788). Systema naturae per regna tria naturae. Humboldt-Universität zu Berlin.

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

Ludlow, A. and Magurran, A. E. (2006). Gametic isolation in guppies (Poecilia reticulata). Proceedings of the Royal Society B: Biological Sciences *273*, 2477–2482.

Lürig, M. D., Donoughe, S., Svensson, E. I., Porto, A. and Tsuboi, M. (2021). Computer vision, machine learning, and the promise of phenomics in ecology and evolutionary biology. Frontiers in Ecology and Evolution *9*, 642774.

Mallet, J. (2001a). The speciation revolution. Journal of Evolutionary Biology *14*, 887–888.

Mallet, J. (2001b). Species, concepts of. Encyclopedia of biodiversity *5*, 427–440.

Mallet, J. (2007). Hybrid speciation. Nature *446*, 279–283.

Mallet, J., Besansky, N. and Hahn, M. W. (2016). How reticulated are species? BioEssays *38*, 140–149.

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

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., Blaxter, M., Manica, A., Mallet, J. and Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome research *23*, 1817–1828.

Martin, W., Baross, J., Kelley, D. and Russell, M. J. (2008). Hydrothermal vents and the origin of life. Nature Reviews Microbiology *6*, 805–814.

Mayr, E. (1963). Animal species and evolution. Harvard University Press.

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

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

Meriam, M. S. (2017). Metamorphosis insectorum Surinamensium [1705]. Ámsterdam: Gerardum Valk .

Millennium ecosystem assessment, M. (2005). Ecosystems and human well-being, vol. 5,. Island press Washington, DC.

Miller, C. J. and Matute, D. R. (2017). The effect of temperature on Drosophila hybrid fitness. G3: Genes, Genomes, Genetics *7*, 377–385.

Naisbit, R. E., Jiggins, C. D. and Mallet, J. (2001). Disruptive sexual selection against hybrids contributes to speciation between Heliconius cydno and Heliconius melpomene. Proceedings of the Royal Society of London. Series B: Biological Sciences *268*, 1849–1854.

Natoli, A., Peddemors, V. M. and Rus Hoelzel, A. (2004). Population structure and speciation in the genus Tursiops based on microsatellite and mitochondrial DNA analyses. Journal of evolutionary biology *17*, 363–375.

Navalón, G., Marugán-Lobón, J., Bright, J. A., Cooney, C. R. and Rayfield, E. J. (2020). The consequences of craniofacial integration for the adaptive radiations of Darwin's finches and Hawaiian honeycreepers. Nature Ecology & Evolution *4*, 270–278.

Nosil, P., Crespi, B. J. and Sandoval, C. P. (2002). Host-plant adaptation drives the parallel evolution of reproductive isolation. Nature *417*, 440–443.

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. Annual review of ecology and systematics *25*, 547–572.

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

Piertney, S. and Oliver, M. (2006). The evolutionary ecology of the major histocompatibility complex. Heredity *96*, 7–21.

Ponder, W. F., Lindberg, D. R. and Ponder, J. M. (2019). Biology and evolution of the mollusca, volume 1. CRC Press.

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

Puebla, O., Coulmance, F., Estape, C. J., Estape, A. M. and 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. and 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.

Rabosky, D. L., Chang, J., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., Alfaro, M. E. et al. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. Nature *559*, 392–395.

Ramon, M. L., Lobel, P. S. and 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-TAIPEI- *44*, 299.

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

Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A. and Kitano, J. (2018). The genomic landscape at a late stage of stickleback speciation: high genomic divergence interspersed by small localized regions of introgression. PLoS genetics *14*, e1007358.

Robertson, D. and Van Tassell, J. (2019). Shorefishes of the Greater Caribbean: online information system. Version 2.0. Smithsonian Tropical Research Institute, Balboa, Panamá .

Robertson, D. R., Estapé, C. J., Estapé, A. M., Richter, L., Peña, E. and Victor, B. (2022). An updated, illustrated inventory of the marine fishes of the US Virgin Islands. ZooKeys *1103*, 79.

Rogers, S. and Bernatchez, L. (2006). The genetic basis of intrinsic and extrinsic post-zygotic reproductive isolation jointly promoting speciation in the lake whitefish species complex (Coregonus clupeaformis). Journal of evolutionary biology *19*, 1979–1994.

Rometsch, S. J., Torres-Dowdall, J. and Meyer, A. (2020). Evolutionary dynamics of pre-and postzygotic reproductive isolation in cichlid fishes. Philosophical Transactions of the Royal Society B *375*, 20190535.

Salis, P., Lorin, T., Laudet, V. and Frédérich, B. (2019). Magic traits in magic fish: understanding color pattern evolution using reef fish. Trends in genetics *35*, 265–278.

Salzburger, W., Baric, S. and Sturmbauer, C. (2002). Speciation via introgressive hybridization in East African cichlids? Molecular Ecology *11*, 619–625.

Schilthuizen, M., Craze, P., Cabanban, A., Davison, A., Stone, J., Gittenberger, E. and Scott, B. (2007). Sexual selection maintains whole-body chiral dimorphism in snails. Journal of Evolutionary Biology *20*, 1941–1949.

Schluter, D. (2000). The ecology of adaptive radiation. OUP Oxford.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å. et al. (2014). Genomics and the origin of species. Nature Reviews Genetics *15*, 176–192.

Shaw, K. L. and Gillespie, R. G. (2016). Comparative phylogeography of oceanic archipelagos: Hotspots for inferences of evolutionary process. Proceedings of the National Academy of Sciences *113*, 7986–7993.

Shaw, K. L. and Mullen, S. P. (2014). Speciation continuum. Journal of Heredity *105*, 741–742.

Shuker, D. M. and Kvarnemo, C. (2021). The definition of sexual selection. Behavioral Ecology *32*, 781–794.

Simpson, G. G. (1961). Principles of animal taxonomy. Columbia University Press.

Smith, J. and Kronforst, M. R. (2013). Do Heliconius butterfly species exchange mimicry alleles? Biology letters *9*, 20130503.

Snelgrove, P. V. (1999). Getting to the bottom of marine biodiversity: sedimentary habitats: ocean bottoms are the most widespread habitat on earth and support high biodiversity and key ecosystem services. BioScience *49*, 129–138.

Spalding, M. and Grenfell, A. (1997). New estimates of global and regional coral reef areas. Coral reefs *16*, 225–230.

Stankowski, S. and Ravinet, M. (2021a). Defining the speciation continuum. Evolution *75*, 1256–1273.

Stankowski, S. and Ravinet, M. (2021b). Quantifying the use of species concepts. Current Biology *31*, R428–R429.

Streelman, J. T., Albertson, R. C. and Kocher, T. D. (2003). Genome mapping of the orange blotch colour pattern in cichlid fishes. Molecular Ecology *12*, 2465–2471.

Stroud, J. T. and Losos, J. B. (2020). Bridging the process-pattern divide to understand the origins and early stages of adaptive radiation: a review of approaches with insights from studies of Anolis lizards. Journal of Heredity *111*, 33–42.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics *123*, 585–595.

Taveral, J. and 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*, 29–39.

Thibert-Plante, X. and Gavrilets, S. (2013). Evolution of mate choice and the so-called magic traits in ecological speciation. Ecology letters *16*, 1004–1013.

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

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B., Imerovski, I. et al. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. Nature *584*, 602–607.

Urban, S., Gerwin, J., Hulsey, C. D., Meyer, A. and Kratochwil, C. F. (2022). The repeated evolution of stripe patterns is correlated with body morphology in the adaptive radiations of East African cichlid fishes. Ecology and Evolution *12*, e8568.

Van Valen, L. (1976). Ecological species, multispecies, and oaks. Taxon *25*, 233–239.

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. (2015). How many coral reef fish species are there? Cryptic diversity and the new molecular taxonomy. Cambridge University Press, Cambridge, United Kingdom.

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

Wagner, C. E., Harmon, L. J. and Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. Nature *487*, 366–369.

Wäldchen, J. and Mäder, P. (2018). Machine learning for image based species identification. Methods in Ecology and Evolution *9*, 2216–2225.

Wang, R. L., Wakeley, J. and Hey, J. (1997). Gene flow and natural selection in the origin of Drosophila pseudoobscura and close relatives. Genetics *147*, 1091–1106.

Waterston, R., Chinwalla, A., Cook, L., Delehaunty, K., Fewell, G., Fulton, L., Fulton, R., Graves, T., Hillier, L., Lindblad-Toh, K. et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Weinstein, B. G. (2018). A computer vision for animal ecology. Journal of Animal Ecology *87*, 533–545.

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

Wu, C.-I. (2001). The genic view of the process of speciation. Journal of evolutionary biology *14*, 851–865.

Zhou, X., Guang, X., Sun, D., Xu, S., Li, M., Seim, I., Jie, W., Yang, L., Zhu, Q., Xu, J. et al. (2018). Population genomics of finless porpoises reveal an incipient cetacean species adapted to freshwater. Nature communications *9*, 1276.

# 02

# PHENOTYPIC AND GENOMIC DISSECTION OF COLOR PATTERN VARIATION IN A REEF FISH RADIATION

**Floriane Coulmance,** Derya Akkaynak, Yann Le Poul, Marc P. Höppner, W. Owen McMillan, Oscar Puebla

The re-print within this thesis is in agreement with *John Wiley and Sons* under license number *5895510926740*.
**Coulmance, F.**, Akkaynak, D., Le Poul, Y., Höppner, M. P., McMillan, W. O. and Puebla, O. (2024). Phenotypic and genomic dissection of colour pattern variation in a reef fish radiation. Molecular Ecology 33, e17047.

## Abstract

Coral reefs rank among the most diverse species assemblages on Earth. A particularly striking aspect of coral reef communities is the variety of color patterns displayed by reef fishes. Color pattern is known to play a central role in the ecology and evolution of reef fishes through, for example, signaling or camouflage. Nevertheless, color pattern is a complex trait in reef fishes—actually a collection of traits—that is difficult to analyze 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*, color correction, alignment of the fish images with a combination of landmarks and Bézier curves, and principal component analysis on the color value of each pixel of each aligned fish. This approach identifies the major color 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 color pattern variation. This second layer of analysis reveals sharp association peaks along the hamlet genome for each color pattern element and allows to characterize the phenotypic effect of the single nucleotide polymorphisms that are most strongly associated with color pattern variation at each association peak. Our results suggest that the diversity of color patterns displayed by the hamlets is generated by a modular genomic and phenotypic architecture.

## Keywords

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

## Author's contribution

**FC** contributed to the development of the project, conducted all the analyses and wrote the first draft of the manuscript. DA contributed to the development of the camera system, part of the fieldwork and provided the scripts for color correction. YLP contributed to the image analysis and provided the scripts for the image alignment. MPH contributed to the sequencing. WOM contributed to development of the project and part of the fieldwork. OP 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.

## 2.1 Introduction

The color patterns displayed by reef fishes are among the most visually stunning traits in animals. Coral reefs are highly visual environments and color pattern plays an important ecological and evolutionary role in reef fishes (Thayer, 1909; Cott, 1940; Lorenz, 1962; Marshall, 1998; Marshall et al., 2019). Color pattern is for example involved in signaling (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 Paleocene (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 color pattern (Bellwood et al. 2015, 2017). The analysis of reef fish color 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, color pattern is a complex trait in reef fishes. It is actually a collection of traits rather than a single trait. Structurally, color pattern involves all aspects of color (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 (Salis et al. 2019; Parichy 2021). Genetically, tens of patterning genes have been identified in fishes, but mostly in model fishes (Salis et al., 2019; Parichy, 2021), and we are only starting to unveil the genetic bases of color 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 ten loci with classic genetic markers to about ten million loci with whole genomes, in less than two decades. As a consequence, progress starts to be hindered by phenotyping power, i.e. the ability to describe phenotypic variation in detail, rather than genotpying power.

Reef fishes provide a good illustration of this situation. Studies of reef fish color pattern have typically relied on manual scoring (e.g. Kelley et al., 2013; Hemingson et al., 2020; Hench et al., 2022), which is appropriate for discrete color pattern elements (e.g. bars or eyespots). Nevertheless, manual scoring reduces color 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 analyze reef fish color patterns. Nevertheless, variation in body shape can confound the analysis of color pattern with this approach (Alfaro et al., 2019). An alternative avenue consists in aligning fish images and analyzing the color of each pixel of each aligned fish with Principal Component Analysis (PCA). Hemingson et al. (2019) applied this approach to *Chaetodon* butterflyfishes reducing color to four discrete categories, (‘yellow’, ‘black’, ‘white’

and 'other'), and evidenced an effect of geographic range overlap and symmetry on color pattern. Color-standardized images would allow to analyze quantitatively the color 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 color pattern that delineates 18 recognized species (Lobel 2011; Puebla et al. 2022). The hamlets are highly sympatric and very similar ecologically (Whiteman et al., 2007; Holt et al., 2008). Color pattern is an important cue for mate choice and the hamlets show strong assortative mating both in the field and in experimental conditions (Fischer, 1980; Domeier, 1994; Barreto and McCartney, 2008; Puebla et al., 2012). Inter-specific spawning occur at low frequency in natural populations (<2% of observations). There are no apparent postzygotic barriers between species (Whiteman and Gage, 2007) and genetic data indicate that gene flow is ongoing among species (Hench et al., 2019, 2022). Color pattern also likely plays an ecological role through aggressive mimicry and crypsis (Randall and Randall, 1960; Thresher, 1978; Fischer, 1980; Puebla et al., 2007, 2018). Color 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, color pattern is a complex trait in the hamlets, involving all aspects of color variation (e.g. hue, saturation) as well as a diversity of patterns (bars, dots, lines, spots, gradients, ...). In addition, color pattern is highly variable, not only among species but also within, both within and among locations (Thresher, 1978; Aguilar-Perera, 2004; Robertson et al., 2019), which complicates species delineation.

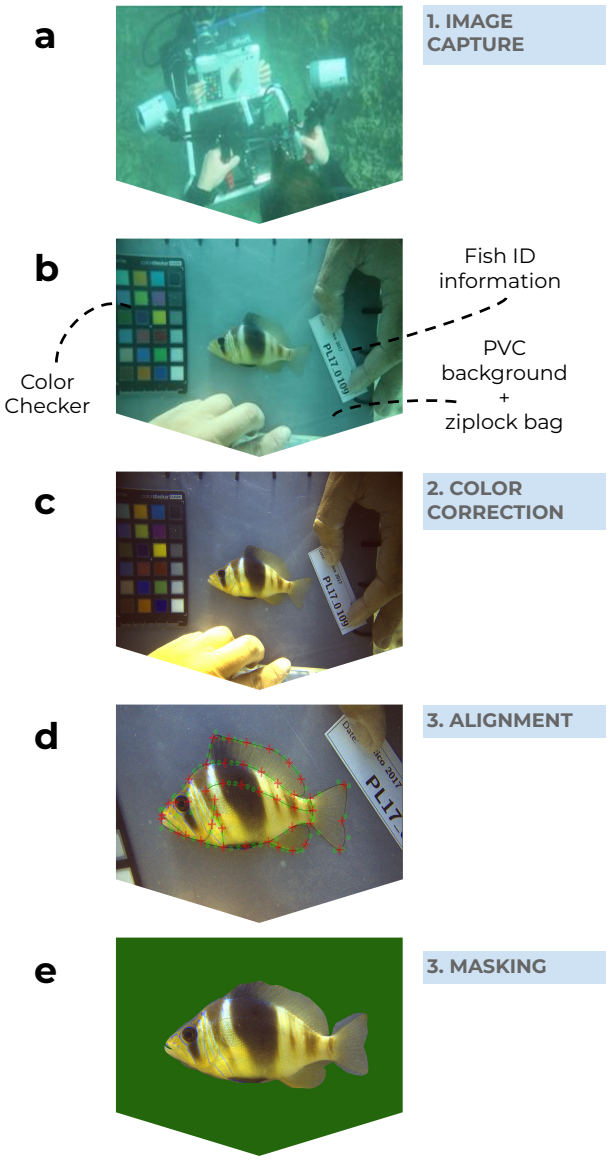



**Figure 2.1** — Overview of the image capture and standardization procedure. The fishes are captured *in situ*, photographed against a neutral background next to a color checker and an ID label at a standardized distance and orientation from the camera lens. The fish images are then color-standardized, aligned and masked.

The hamlets are also very closely related genetically. They show low levels of genetic differentiation and divergence (Holt et al., 2011; Barreto and McCartney, 2008; Puebla et al., 2012, 2014; Hench et al., 2022) and do generally not sort into distinct mitochondrial haplogroups (McCartney et al., 2003; Ramon et al., 2003; Garcia-Machado et al., 2004). These low levels of genetic differentiation and divergence, combined with the diversity of color patterns displayed by the group, provide a backdrop that is well suited for a Genome-Wide Association Study (GWAS) for color pattern variation. Previous work has identified genetic associations with three color 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 color 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 analyze color pattern variation quantitatively and with minimal human intervention, i.e. without manual scoring of predefined color pattern elements. We then use this quantified color pattern variation as trait for a GWAS. We hypothesize that this approach will allow to identify the major color pattern elements that contribute to phenotypic variation in the group, and to identify genetic associations with each color pattern element. We expect to recover previously identified color pattern elements such as vertical bars. We also expect to recover new color pattern elements, color elements in particular since color varies continuously in the hamlets and could therefore not be scored manually in previous analyses.

## 2.2 Materials and Methods

### 2.2.1 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 color pattern diversity and geographic range. This dataset includes 13 species and is largely new, with all images and 101 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).

### 2.2.2 Sampling procedure

Previous experience with the hamlets indicates that color 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 color pattern variation within and among species [e.g. Robertson et al., 2019; Puebla et al., 2022]. Nevertheless, such photographs are not suited for quantitative analysis due to their lack of standardization in terms of size, color and orientation. In order to address these limitations, we developed a protocol to take standardized photographs of live specimens *in situ* (Figure 2.1).

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 color 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. 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 colors 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 color pattern is not affected beyond the variation that is observed in the wild (Figure S2.1). Furthermore, observations of tagged individuals over two years indicate that this procedure does not noticeably alter their survival (O. Puebla, unpublished data).

### 2.2.3 Image analysis

#### Pre-processing

The image analysis was restricted to 113 samples for which we also sequenced full genomes (Figure 2.2, Table S2.1). 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 color-correcting the im-

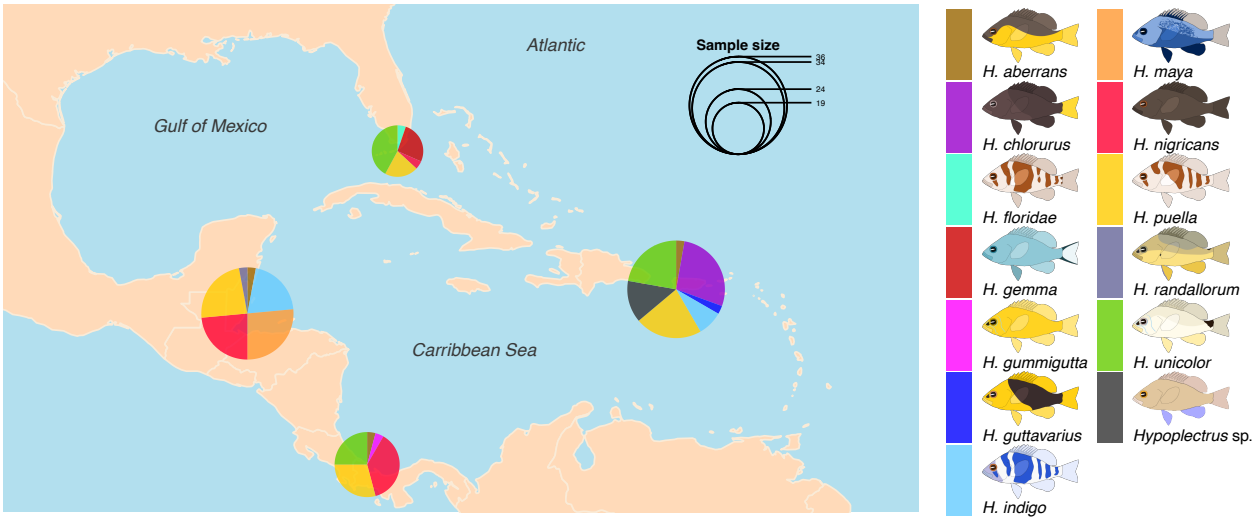



**Figure 2.2** — Sampling design. The dataset consisted of 113 samples spanning 13 species and 4 locations (Panama, Belize, Puerto Rico and Florida).

ages to standardize colors across the entire dataset. Raw *CR2* images were demosaicked with Adobe DNG converter and output as *dng* images for compatibility with MATLAB. Photographs were color-corrected using the color 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 color 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 color-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.

### 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 color-corrected *png* images and the *mat* mask files to center 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. In order 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 color informa-

tion. Finally, each image was flattened into a long one-dimensional image vector to be used as input for Principal Component Analysis (PCA).

#### PCA on color pattern variation

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

### 2.2.4 Genotyping

A total of 101 samples were sequenced anew for this study. Fifteen additional samples were available from Moran et al. (2019) and one from Hench et al. (2019). 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 S2.1 and a summary of the sampling design is presented Figure 2.2.

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 and 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 dataset was filtered for a minor allele count $\geqslant 2$ and reduced to biallelic SNPs only using VCFtools (Danecek et al., 2011).

### 2.2.5 Population genetic analyses

#### Genetic structure

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

#### Identification of hybrids and backcrosses

In order 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 and 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. 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 and 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.

### 2.2.6 Genome-Wide Association Study

In order to identify genomic regions associated with color pattern variation, a multivariate GWAS was implemented using the PCs of the color pattern PCA as traits. Specifically, the multivariate association test implemented in *PLINK* (Purcell et al., 2007) was used (Ferreira and Purcell, 2009) with the first five PCs of the color 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 and 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 color 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 and 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)* cutoff 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 display-

ing the weighted effect of the five PCs on this SNP for each pixel using the PC eigenvectors.

## 2.3 Results

### 2.3.1 Image analysis

The first five PCs of the color 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 S2.2). The PCA showing PC1 versus PC5 is presented in Figure 2.3a, the PCA showing PC2 versus PC4 is presented in Figure 2.4a, and the PCA showing PC1 versus PC3 is presented in Figure S2.3a.

PC1 accounted for 58.0% of the variation and broadly discriminated samples from the darkest colored ones (*H. nigricans*, on the left of the PCA) to the lightest colored ones (*H. unicolor*, on the right, Figure 2.3a). 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 2.3d).

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 2.3a). 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 2.3e). Thus, the heatmaps for PC1 (Figure 2.3d) and PC5 (Figure 2.3e) 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-colored face and body (*H. maya* and *H. gemma*, in the upper half of the PCA) from the other samples (in the lower half, Figure 2.4a). 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 2.4e).

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 color 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 2.4a). In agreement with this pattern, the heatmap indicated that the pixels whose variation contributed most to PC2 were located on the caudal fin (Figure 2.4d).

PC3 accounted for 8.0% of the variation but did not appear to discriminate any color pattern element in particular (Figure S2.3a). Examination of the photographs indicated that the images that are in the upper part of the PCA in Figure S2.3a 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 S2.3e), which highlights this part of the body. PC3 appears therefore to capture variation that is in large part artefactual.

### 2.3.2 Population genetic analyses

#### Genetic structure

Genetic structure was low to moderate, with pairwise $F_{ST}$ estimates ranging between 0.002 (*H. chlorurus* versus *H. puella*

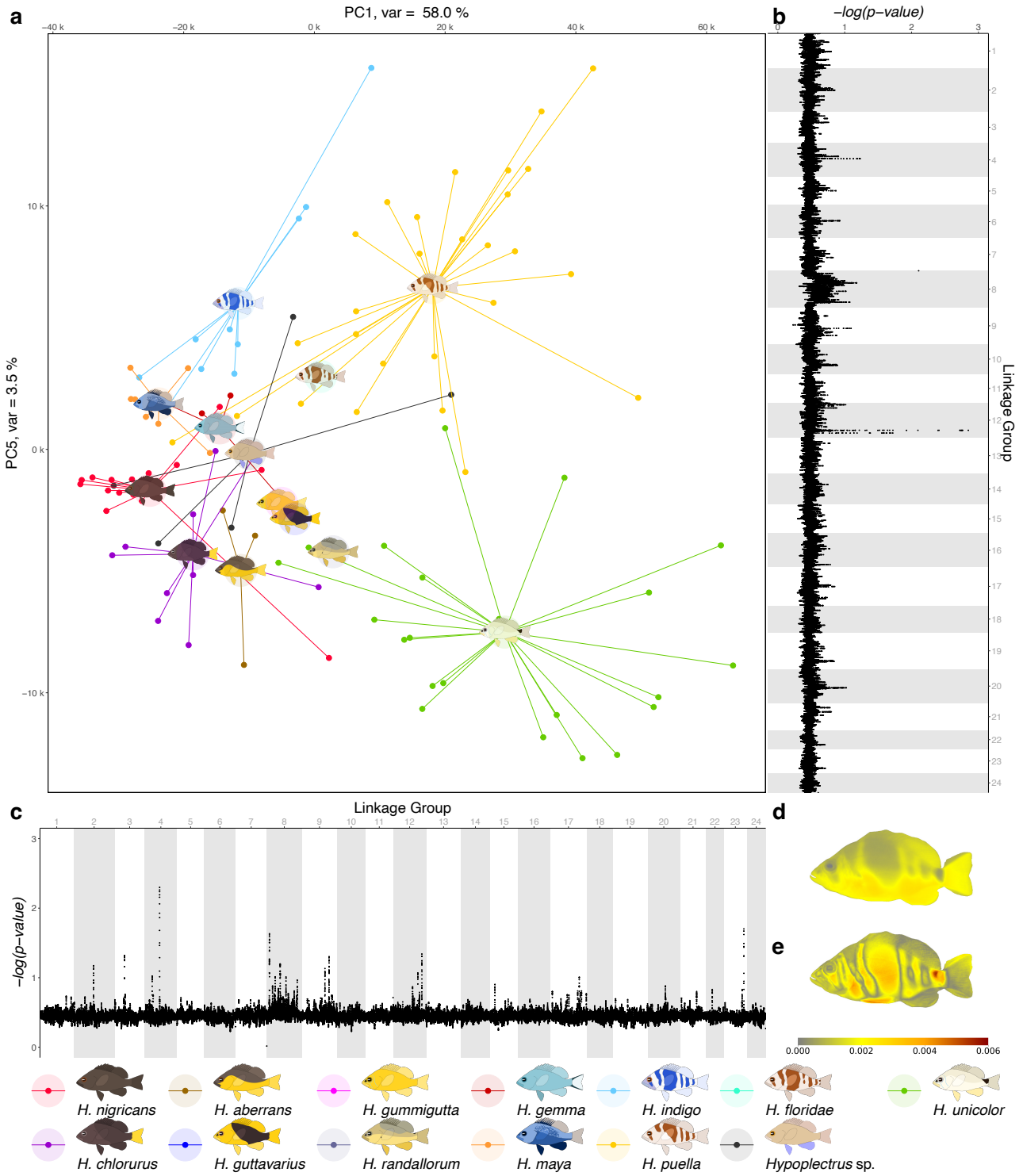



**Figure 2.3** — **a.** Principal Component (PC) 1 *versus* PC5 of the color pattern PCA. **b.** association *-log(p-value)* of the univariate GWAS for PC5 along the hamlet genome. **c.** association *-log(p-value)* of the univariate GWAS for PC1 along the hamlet genome. **d.** heatmap showing the contribution of each pixel to PC1. **e.** heatmap showing the contribution of each pixel to PC5.

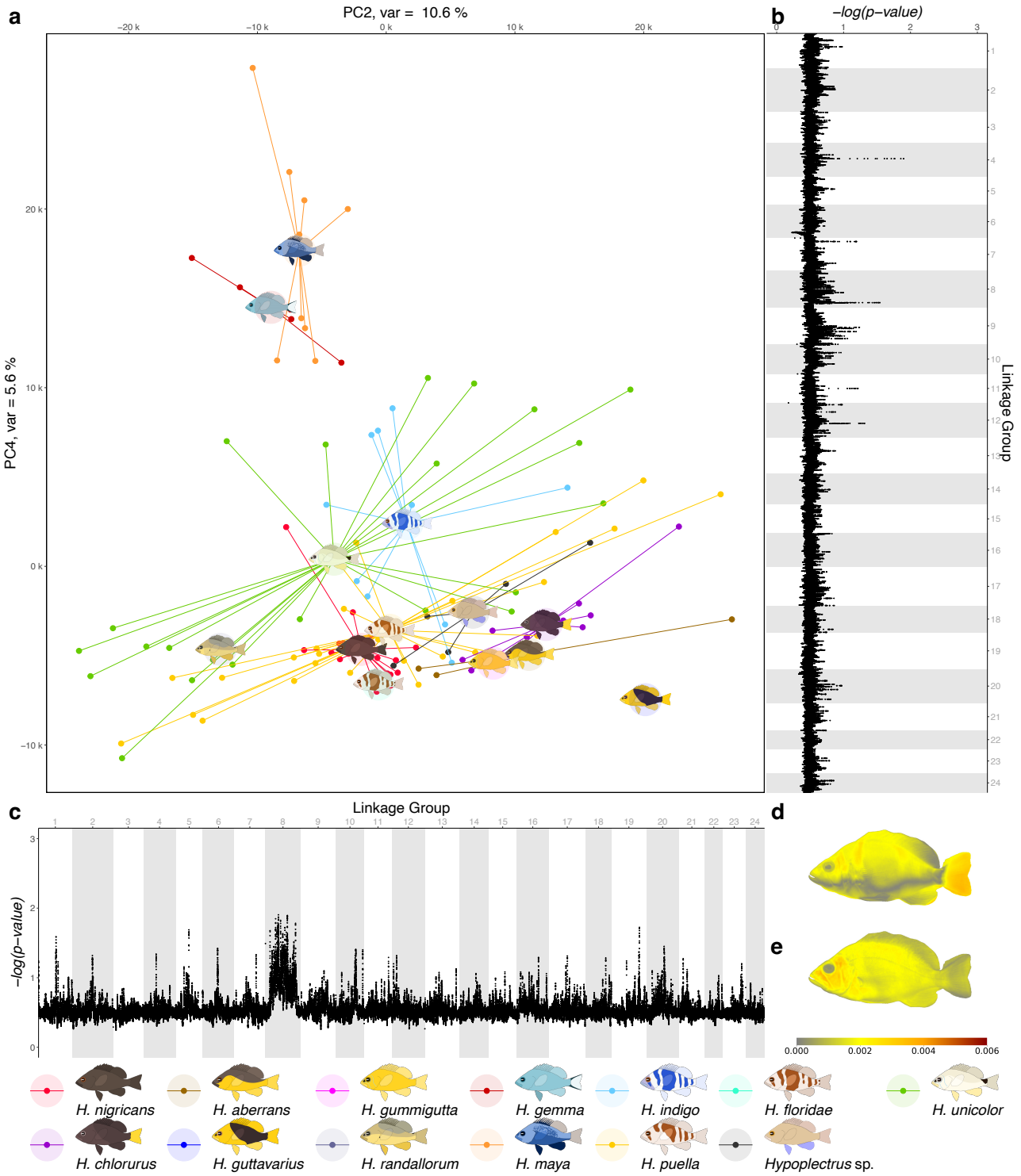



**Figure 2.4** — **a.** Principal Component (PC) 2 *versus* PC4 of the color pattern PCA. **b.** association -*log(p-value)* of the univariate GWAS for PC4 along the hamlet genome. **c.** association -*log(p-value)* of the univariate GWAS for PC2 along the hamlet genome. **d.** heatmap showing the contribution of each pixel to PC2. **e.** heatmap showing the contribution of each pixel to PC4.

in Puerto Rico) and 0.097 (*H. nigricans* from Panama versus *H. indigo* from Belize, Table S2.2), 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.

Princpal Component 1 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 S2.4a, b). Principal Component 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.

#### Identification of hybrids and backcrosses

One individual out of the 80 included in the NewHybrids analyzes (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 inter-specific spawnings reported in wild populations, including the populations from Florida, Belize and Bocas del Toro considered in this study (Fischer, 1980; Puebla et al., 2007; Barreto and McCartney, 2008; Puebla et al., 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 S2.5).

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* color 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, i.e. 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%.

### 2.3.3 Genome-Wide Association Study

Principal Components 1 and 5 showed sharp univariate association peaks along the genome (Figure 2.3b, c). Principal Component 5 (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 2.3b). Principal Component 1 (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 2.3c). 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 2.4b). Principal Component 2 (which is associated with yellow caudal fin) showed a large number of relatively low association

peaks distributed all over the genome (Figure 2.4c). Principal Component 3 (which appears to capture artefactual variation) showed the fewest and lowest association peaks (Figure S2.3b). 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 an inversion—that has been previously identified in this LG (Theodosiou et al., 2016; Hench et al., 2019).

The multivariate GWAS considering the first five PCs altogether retrieved the same association peaks that were identified by the univariate analyzes, and the same higher levels of association in the low-recombining region of LG08 (Figure 2.5a). When considering each PC individually, the multivariate approach provided similar association patterns as the univariate analyzes (Figure S2.6). 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 color pattern variation for each peak, and characterize the multivariate phenotypic effect of this SNP (Figure 2.5b, c, 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 2.5b). 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 2.5c).

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 2.5d).

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 analyzed in detail due to the limited power provided by our sample size to characterize small-effect loci.

## 2.4 Discussion

Principal Component Analysis of the standardized images dissects color pattern variation into broadly interpretable color 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 color elements that could not be scored manually in previous analyses. The univariate GWAS allows then to dissect the genetic associations with each color pattern element independently. The color pattern elements identified differ in terms of what aspect of color pattern variation they capture, the proportion of color pattern variation they explain, and also in terms of genetic associations. For example, PC5 explains only 3.5% of color pattern variation but is associated with discrete and specific color pattern elements (vertical bars and caudal peduncle mark), and presents one outstanding association peak on LG12. On the other hand, PC1 explains 58% of color pattern variation, is associated with a more continuous color pattern element (dark to light ventral part, caudal fin and anal fin), and shows several associations peaks on different LGs.

The color pattern elements identified are independent from each other from an image analysis perspective, but this doesn't 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 color 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 (Moran et al., 2019; Hench 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 color pattern elements identified. Indeed, although the vertical bars are a conspicuous element of color pattern variation in the hamlets and show an outstanding association peak, they represent only 3.5% of variation in the color pattern PCA. In terms of gene annotation, this association peak points to a castor zinc finger transcription factor (*casz1*). Interestingly this gene has been

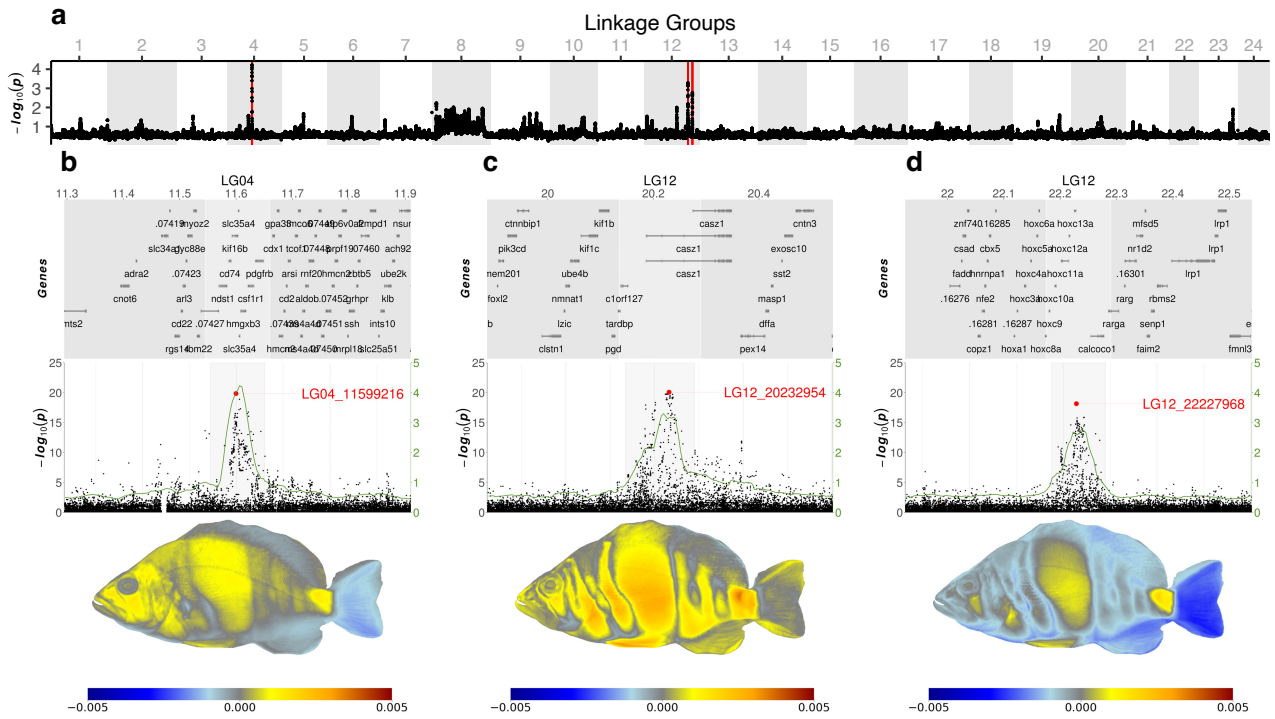



**Figure 2.5** — Results of the multivariate GWAS on the first five PCs of the PCA on color pattern variation. **a.** Manhattan plot showing the association *-log(p-value)* along the hamlet genome. The three major association peaks are highlighted in red. **b-d.** Close-ups on the three major association 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 50kb 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).

shown to be expressed in the photoreceptors of mice (Mattar et al., 2015, 2018) and zebrafish (Ogawa and 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 color 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 poly-

morphic 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 color pattern element from other color 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' to 5' organisation 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 (Thummel et al., 2004; Jakovlić and Wang, 2016). This is 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 color pattern variation. In the univariate analyses this association peak is strongest for PC1, which is associated with light to dark color. 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 color 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 color patterning. This is particularly true for PC2, which shows the highest associations on LG08. Chromosomal inversions have been shown to be associated with color 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 color pattern elements, as reported in *Heliconius* butterflies (Jay et al., 2022). Nevertheless, larger sam-

ple sizes will be needed to identify the specific color 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.

### 2.4.1 Phenotypic and genetic modularity

A striking aspect of the color pattern elements identified by the PCA (color 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 color pattern variation in the whole radiation. Indeed, the color patterns displayed by the different species are broadly different combinations of these basic elements. Furthermore, the GWAS indicates that each color 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 color 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 hybridisation and introgression in the group (Fischer 1980; Domeier 1994; Barreto and McCartney 2008; Puebla et al. 2012; Hench et al. 2019, 2022, 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 color pattern also plays a prominent role (Van Belleghem et al., 2017; Stryjewski and Sorenson, 2017; Campagna 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 color pattern elements.

### 2.4.2 Advantages and limitations

The major advantage of our approach is that it considers color pattern variation in its entirety, without *a priori* identification of color pattern elements, manual scoring, or clustering of pixel color values into discrete color categories. Furthermore, it generates results that are interpretable from both a color pattern and a genetic perspective. The PCA dissects color 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 color pattern elements and large-effect loci, it is too low to characterize more subtle elements of color pattern variation and small-effect loci. Such small-effect loci appear to also contribute to color 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 (Holt et al. 2011; Barreto and McCartney 2008; Puebla et al. 2012, 2014; Hench et al. 2022, this study) and iii. the fact that most major color 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, i.e. 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 color 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 color 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 color 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, e.g. in the UV (Siebeck et al., 2010; Mitchell et al., 2022), can be important in some reef fishes. This limitation may be addressed by 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 color patterns are perceived by e.g. conspecifics, congeners, predators or preys, which requires visual modeling (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 (*sws2a*$\alpha$, *sws2a*$\beta$, *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 color pattern from other traits that may also differ among species.

### 2.4.3 Perspectives

This study constitutes a proof-of-concept for the quantitative analysis of color pattern variation in reef fishes. The results illustrate the potential of our approach to dissect color pattern variation into interpretable color pattern elements and detect genetic associations with these color pattern elements, both independently with a univariate GWAS and altogether with a multivariate GWAS. The method may by further refined by 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 color pattern variation, not only among species but also within, providing a quantitative basis for study of the eco-evolutionary significance of color pattern in reef fishes.

**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 sup- port 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:** 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.

## 2.5 Supplementary Information

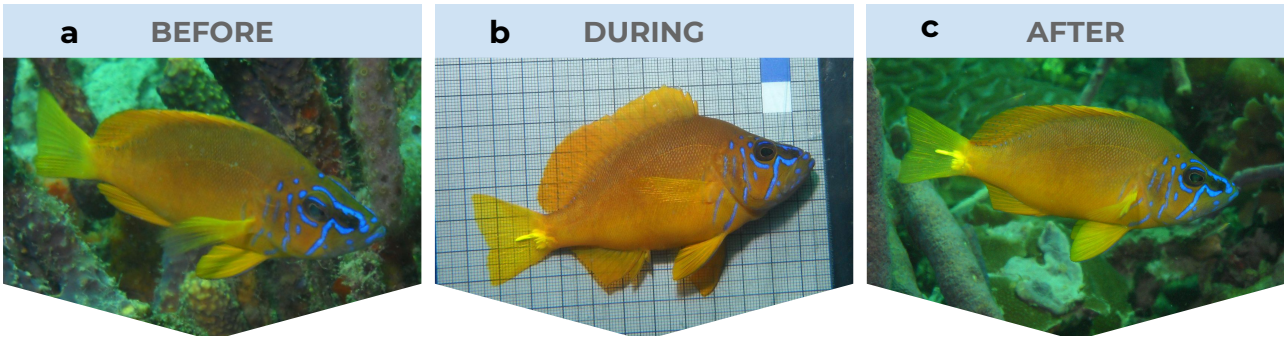



**Figure S2.1** — Image of the same individual taken **a.** just before the sampling procedure, **b.** during the sampling procedure, and **c.** just after the sampling procedure. Note that this fish was tagged on the caudal fin, which was not the case in this study, and the early version of the background and color standard, which were not used in this study either.

**Table S2.1** — List of samples used in this study. Locations correspond to Belize (bel), Panama (boc), Puerto Rico (pue) and Florida (flo) and species to *H. aberrans* (abe), *H. chlorurus* (chl), *H. floridae* (flo), *H. gemma* (gem), *H. gummigutta* (gum), *H. guttavarius* (gut), *H. indigo* (ind), *H. maya* (may), *H. nigricans* (nig), *H. puella* (pue), *H. randallorum* (ran), undescribed (sp.) and *H. unicolor* (uni).

| Nr | ID | Species | Location | Date | Latitude | Longitude | Coverage | Accesion Number |
|---|---|---|---|---|---|---|---|---|
| 1 | 28366 | gum | boc | 2017–02–04 | 9.332778 | –82.25472 | 20.9 | ERS8632035 |
| 2 | 28377 | uni | boc | 2017–02–05 | 9.332778 | –82.25472 | 19.6 | ERS14948427 |
| 3 | 28384 | pue | boc | 2017–02–06 | 9.332778 | –82.25472 | 34.2 | ERS14948442 |
| 4 | 28386 | nig | boc | 2017–02–06 | 9.332778 | –82.25472 | 26.9 | ERS14948430 |
| 5 | 28387 | nig | boc | 2017–02–06 | 9.332778 | –82.25472 | 35 | ERS14948426 |
| 6 | 28388 | uni | boc | 2017–02–06 | 9.332778 | –82.25472 | 33 | ERS14948436 |
| 7 | 28390 | nig | boc | 2017–02–06 | 9.332778 | –82.25472 | 34.2 | ERS14948425 |
| 8 | 28391 | uni | boc | 2017–02–07 | 9.332778 | –82.25472 | 35.7 | ERS14948444 |
| 9 | 28392 | uni | boc | 2017–02–07 | 9.332778 | –82.25472 | 28.2 | ERS14948440 |
| 10 | 28394 | nig | boc | 2017–02–07 | 9.332778 | –82.25472 | 37.3 | ERS14948432 |
| 11 | 28399 | nig | boc | 2017–02–07 | 9.332778 | –82.25472 | 27.8 | ERS14948433 |
| 12 | AG9RX_46 | nig | boc | 2017–02–06 | 9.332778 | –82.25472 | 31.5 | ERS14948429 |
| 13 | AG9RX_48 | pue | boc | 2017–02–07 | 9.332778 | –82.25472 | 29.9 | ERS14948431 |
| 14 | AG9RX_49 | nig | boc | 2017–02–07 | 9.332778 | –82.25472 | 30.2 | ERS14948424 |
| 15 | AG9RX_50 | nig | boc | 2017–02–07 | 9.332778 | –82.25472 | 33.9 | ERS14948437 |
| 16 | AG9RX_51 | pue | boc | 2017–02–07 | 9.332778 | –82.25472 | 31.1 | ERS14948434 |
| 17 | AG9RX_53 | pue | boc | 2017–02–07 | 9.332778 | –82.25472 | 28.7 | ERS14948435 |
| 18 | PL17_01 | uni | boc | 2017–02–07 | 9.332778 | –82.25472 | 23.9 | ERS14948439 |
| 19 | PL17_02 | pue | boc | 2017–02–07 | 9.332778 | –82.25472 | 26.5 | ERS14948428 |
| 20 | PL17_04 | pue | boc | 2017–02–07 | 9.332778 | –82.25472 | 28.6 | ERS14948438 |
| 21 | PL17_05 | pue | boc | 2017–02–07 | 9.332778 | –82.25472 | 24.5 | ERS14948441 |
| 22 | PL17_103 | pue | bel | 2017–10–04 | 16.765278 | –88.14417 | 27 | ERS14948415 |
| 23 | PL17_100 | ind | bel | 2017–09–04 | 16.765278 | –88.14417 | 29.2 | ERS14948407 |
| 24 | PL17_104 | nig | bel | 2017–10–04 | 16.765278 | –88.14417 | 21.9 | ERS14948408 |
| 25 | PL17_105 | pue | bel | 2017–10–04 | 16.765278 | –88.14417 | 37.8 | ERS14948419 |
| 26 | PL17_106 | nig | bel | 2017–10–04 | 16.765278 | –88.14417 | 21.9 | ERS14948401 |
| 27 | PL17_107 | pue | bel | 2017–10–04 | 16.765278 | –88.14417 | 31.5 | ERS14948420 |
| 28 | PL17_108 | nig | bel | 2017–10–04 | 16.765278 | –88.14417 | 24.6 | ERS14948411 |
| 29 | PL17_109 | pue | bel | 2017–10–04 | 16.765278 | –88.14417 | 25.3 | ERS14948417 |
| 30 | PL17_110 | pue | bel | 2017–10–04 | 16.765278 | –88.14417 | 23.6 | ERS14948422 |
| 31 | PL17_111 | ind | bel | 2017–11–04 | 16.765278 | –88.14417 | 26.4 | ERS14948399 |
| 32 | PL17_112 | nig | bel | 2017–11–04 | 16.765278 | –88.14417 | 18.8 | ERS14948405 |
| 33 | PL17_125 | ran | bel | 2017–13–04 | 16.765278 | –88.14417 | 30.7 | ERS14948421 |
| 34 | PL17_117 | pue | bel | 2017–12–04 | 16.765278 | –88.14417 | 24.5 | ERS14948404 |
| 35 | PL17_127 | ind | bel | 2017–14–04 | 16.765278 | –88.14417 | 25.4 | ERS14948402 |
| 36 | PL17_128 | ind | bel | 2017–14–04 | 16.765278 | –88.14417 | 21.1 | ERS14948406 |
| 37 | PL17_132 | ind | bel | 2017–14–04 | 16.765278 | –88.14417 | 24.3 | ERS14948416 |
| 38 | PL17_134 | uni | flo | 2017–05–07 | 24.752580 | –80.76065 | 21.8 | ERS14948448 |
| 39 | PL17_135 | uni | flo | 2017–06–07 | 24.752580 | –80.76065 | 25.8 | ERS14948455 |
| 40 | PL17_136 | uni | flo | 2017–06–07 | 24.752580 | –80.76065 | 38.4 | ERS14948452 |
| 41 | PL17_137 | uni | flo | 2017–07–07 | 24.752580 | –80.76065 | 34.5 | ERS14948453 |
| 42 | PL17_138 | uni | flo | 2017–07–07 | 24.752580 | –80.76065 | 24.9 | ERS14948451 |
| 43 | PL17_139 | pue | flo | 2017–08–07 | 24.752580 | –80.76065 | 29.5 | ERS14948450 |
| 44 | PL17_140 | uni | flo | 2017–08–07 | 24.752580 | –80.76065 | 25 | ERS14948456 |
| 45 | PL17_141 | uni | flo | 2017–08–07 | 24.752580 | –80.76065 | 32 | ERS14948457 |
| 46 | PL17_143 | uni | flo | 2017–09–07 | 24.752580 | –80.76065 | 30 | ERS14948454 |
| 47 | PL17_149 | nig | flo | 2017–12–07 | 24.752580 | –80.76065 | 30.6 | ERS14948447 |
| 48 | PL17_155 | pue | flo | 2017–14–07 | 24.752580 | –80.76065 | 29.6 | ERS14948449 |
| 49 | PL17_157 | pue | flo | 2017–15–08 | 24.752580 | –80.76065 | 24.9 | ERS14948446 |
| 50 | PL17_159 | pue | flo | 2017–16–08 | 24.752580 | –80.76065 | 26.7 | ERS14948445 |
| 51 | PL17_23 | sp. | pue | 2017–03–13 | 17.952820 | –67.05643 | 26.6 | ERS14948480 |
| 52 | PL17_35 | ind | pue | 2017–03–15 | 17.952820 | –67.05643 | 28.5 | ERS14948482 |
| 53 | PL17_37 | chl | pue | 2017–03–15 | 17.952820 | –67.05643 | 29.7 | ERS14948469 |
| 54 | PL17_38 | chl | pue | 2017–03–15 | 17.952820 | –67.05643 | 32.7 | ERS14948464 |
| 55 | PL17_39 | chl | pue | 2017–03–15 | 17.952820 | –67.05643 | 23.4 | ERS14948460 |
| 56 | PL17_40 | chl | pue | 2017–03–15 | 17.952820 | –67.05643 | 31.4 | ERS14948466 |
| 57 | PL17_41 | chl | pue | 2017–03–15 | 17.952820 | –67.05643 | 27.6 | ERS14948462 |
| 58 | PL17_42 | chl | pue | 2017–03–15 | 17.952820 | –67.05643 | 21.9 | ERS14948467 |
| 59 | PL17_43 | chl | pue | 2017–03–15 | 17.952820 | –67.05643 | 28.3 | ERS14948461 |
| 60 | PL17_44 | chl | pue | 2017–03–15 | 17.952820 | –67.05643 | 23.9 | ERS14948463 |
| 61 | PL17_50 | pue | pue | 2017–03–16 | 17.952820 | –67.05643 | 24.5 | ERS14948475 |
| 62 | PL17_53 | pue | pue | 2017–03–16 | 17.952820 | –67.05643 | 22.1 | ERS14948476 |
| 63 | PL17_55 | uni | pue | 2017–03–16 | 17.952820 | –67.05643 | 20 | ERS14948485 |
| 64 | PL17_54 | pue | pue | 2017–03–16 | 17.952820 | –67.05643 | 28.5 | ERS14948483 |
| 65 | PL17_56 | sp. | pue | 2017–03–17 | 17.952820 | –67.05643 | 28.6 | ERS14948484 |
| 66 | PL17_57 | pue | pue | 2017–03–17 | 17.952820 | –67.05643 | 28.4 | ERS14948479 |
| 67 | PL17_60 | pue | pue | 2017–03–17 | 17.952820 | –67.05643 | 30.4 | ERS14948481 |
| 68 | PL17_62 | chl | pue | 2017–03–17 | 17.952820 | –67.05643 | 30.2 | ERS14948465 |
| 69 | PL17_63 | uni | pue | 2017–03–18 | 17.952820 | –67.05643 | 30.3 | ERS14948487 |
| 70 | PL17_64 | ind | pue | 2017–03–18 | 17.952820 | –67.05643 | 38 | ERS14948471 |
| 71 | PL17_65 | pue | pue | 2017–03–18 | 17.952820 | –67.05643 | 26 | ERS14948490 |
| 72 | PL17_67 | uni | pue | 2017–03–18 | 17.952820 | –67.05643 | 23.3 | ERS14948493 |
| 73 | PL17_68 | gut | pue | 2017–03–20 | 17.952820 | –67.05643 | 27.2 | ERS14948470 |
| 74 | PL17_69 | pue | pue | 2017–03–21 | 17.952820 | –67.05643 | 28.5 | ERS14948486 |
| 75 | PL17_70 | uni | pue | 2017–03–21 | 17.952820 | –67.05643 | 27.3 | ERS14948489 |
| 76 | PL17_71 | sp. | pue | 2017–03–21 | 17.952820 | –67.05643 | 34.4 | ERS14948478 |
| 77 | PL17_72 | sp. | pue | 2017–03–21 | 17.952820 | –67.05643 | 42 | ERS14948472 |
| 78 | PL17_73 | uni | pue | 2017–03–21 | 17.952820 | –67.05643 | 33.3 | ERS14948491 |
| 79 | PL17_74 | uni | pue | 2017–03–21 | 17.952820 | –67.05643 | 30.7 | ERS14948492 |
| 80 | PL17_93 | nig | bel | 2017–08–04 | 16.765278 | –88.14417 | 36.6 | ERS14948412 |
| 81 | PL17_91 | nig | bel | 2017–08–04 | 16.765278 | –88.14417 | 26.7 | ERS14948400 |
| 82 | PL17_90 | pue | bel | 2017–08–04 | 16.765278 | –88.14417 | 30.1 | ERS14948418 |
| 83 | PL17_88 | abe | bel | 2017–07–04 | 16.765278 | –88.14417 | 37.2 | ERS14948398 |
| 84 | PL17_86 | chl | pue | 2017–03–24 | 17.952820 | –67.05643 | 31.1 | ERS14948468 |
| 85 | PL17_87 | nig | bel | 2017–05–04 | 16.765278 | –88.14417 | 32.1 | ERS14948413 |
| 86 | PL17_75 | abe | pue | 2017–03–22 | 17.952820 | –67.05643 | 33.8 | ERS14948458 |
| 87 | PL17_76 | sp. | pue | 2017–03–22 | 17.952820 | –67.05643 | 34.6 | ERS14948477 |
| 88 | PL17_77 | uni | pue | 2017–03–22 | 17.952820 | –67.05643 | 32.3 | ERS14948488 |
| 89 | PL17_82 | pue | pue | 2017–03–23 | 17.952820 | –67.05643 | 37 | ERS14948473 |
| 90 | PL17_85 | ind | pue | 2017–03–24 | 17.952820 | –67.05643 | 33.3 | ERS14948474 |
| 91 | PL17_97 | ind | bel | 2017–09–04 | 16.765278 | –88.14417 | 34.7 | ERS14948409 |
| 92 | PL17_94 | pue | bel | 2017–08–04 | 16.765278 | –88.14417 | 16.2 | ERS14948410 |
| 93 | PL17_96 | nig | bel | 2017–09–04 | 16.765278 | –88.14417 | 38.7 | ERS14948414 |
| 94 | PL17_99 | ind | bel | 2017–09–04 | 16.765278 | –88.14417 | 22 | ERS14948403 |
| 95 | 28383 | uni | boc | 2017–02–06 | 9.332778 | –82.25472 | 38 | ERS14948443 |
| 96 | 28385 | nig | boc | 2017–02–06 | 9.332778 | –82.25472 | 40.2 | ERS14948423 |
| 97 | 28389 | abe | boc | 2017–02–06 | 9.332778 | –82.25472 | 34.1 | ERS8632036 |
| 98 | PL17_66 | uni | pue | 2017–03–18 | 17.952820 | –67.05643 | 23 | ERS14948459 |
| 99 | PL17_89 | may | bel | 2017–07–04 | 16.765278 | –88.14417 | 16.8 | ERS2899590 |
| 100 | PL17_95 | may | bel | 2017–08–04 | 16.765278 | –88.14417 | 20.5 | ERS2899591 |
| 101 | PL17_119 | may | bel | 2017–13–04 | 16.765278 | –88.14417 | 17.2 | ERS2899593 |
| 102 | PL17_120 | may | bel | 2017–13–04 | 16.765278 | –88.14417 | 15.5 | ERS2899594 |
| 103 | PL17_121 | may | bel | 2017–13–04 | 16.765278 | –88.14417 | 16.2 | ERS2899595 |
| 104 | PL17_122 | may | bel | 2017–13–04 | 16.765278 | –88.14417 | 22.2 | ERS2899596 |
| 105 | PL17_123 | may | bel | 2017–13–04 | 16.765278 | –88.14417 | 17.2 | ERS2899597 |
| 106 | PL17_124 | may | bel | 2017–13–04 | 16.765278 | –88.14417 | 18.5 | ERS2899598 |
| 107 | PL17_126 | may | bel | 2017–13–04 | 16.765278 | –88.14417 | 18.1 | ERS2899599 |
| 108 | PL17_142 | gem | flo | 2017–08–07 | 24.752580 | –80.76065 | 20.4 | ERS2899137 |
| 109 | PL17_144 | gem | flo | 2017–09–07 | 24.752580 | –80.76065 | 24.3 | ERS2899138 |
| 110 | PL17_145 | gem | flo | 2017–09–07 | 24.752580 | –80.76065 | 19.6 | ERS2899139 |
| 111 | PL17_148 | gem | flo | 2017–11–07 | 24.752580 | –80.76065 | 17.7 | ERS2899140 |
| 112 | PL17_153 | gem | flo | 2017–13–07 | 24.752580 | –80.76065 | 21.9 | ERS2899141 |
| 113 | PL17_160 | flo | flo | _ | 24.752580 | –80.76065 | 19.6 | ERS4141277 |

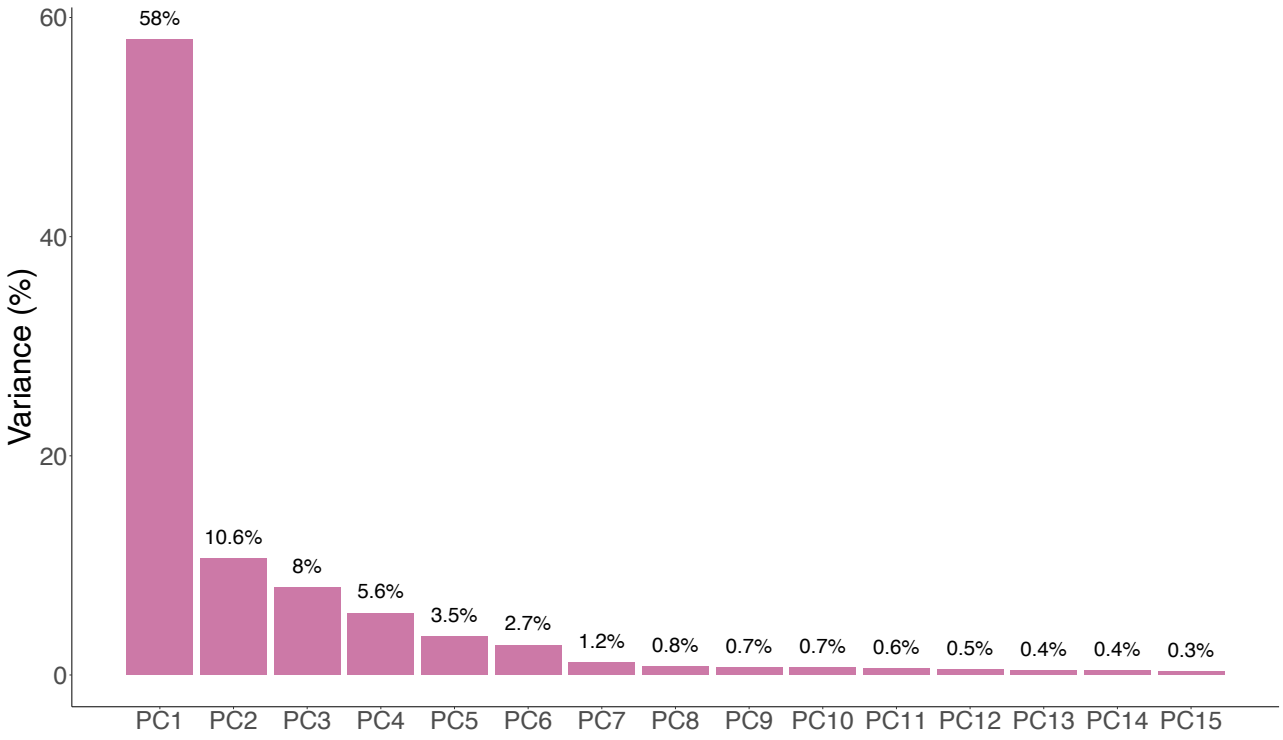



**Figure S2.2** — Percentage of variance explained by each individual Principal Component (PC) in the PCA on color pattern variation.

**Table S2.2** — Whole-genome $F_{ST}$ estimates between all pairs of samples (species/population) that have a sample size of at least seven individuals. Population names are composed of species (first 3 digits) and location (last 3 digits). Locations correspond to Belize (bel), Panama (boc), Puerto Rico (pue) and Florida (flo) and species to *H. chlorurus* (chl), *H. indigo* (ind), *H. maya* (may), *H. nigricans* (nig), *H. puella* (pue) and *H. unicolor* (uni).

| POPULATIONS | puebel | nigbel | indbel | maybel | pueboc | nigboc | puepue | unipue | chlpue | uniflo |
|---|---|---|---|---|---|---|---|---|---|---|
| **puebel** | _ | _ | _ | _ | _ | _ | _ | _ | _ | _ |
| **nigbel** | 0.007 | _ | _ | _ | _ | _ | _ | _ | _ | _ |
| **indbel** | 0.059 | 0.072 | _ | _ | _ | _ | _ | _ | _ | _ |
| **maybel** | 0.027 | 0.034 | 0.089 | _ | _ | _ | _ | _ | _ | _ |
| **pueboc** | 0.009 | 0.015 | 0.076 | 0.033 | _ | _ | _ | _ | _ | _ |
| **nigboc** | 0.028 | 0.023 | 0.097 | 0.052 | 0.026 | _ | _ | _ | _ | _ |
| **puepue** | 0.015 | 0.019 | 0.08 | 0.037 | 0.019 | 0.039 | _ | _ | _ | _ |
| **unipue** | 0.015 | 0.019 | 0.077 | 0.037 | 0.02 | 0.04 | 0.003 | _ | _ | _ |
| **chlpue** | 0.017 | 0.021 | 0.083 | 0.038 | 0.021 | 0.04 | 0.002 | 0.009 | _ | _ |
| **uniflo** | 0.015 | 0.02 | 0.072 | 0.038 | 0.02 | 0.04 | 0.015 | 0.014 | 0.021 | _ |

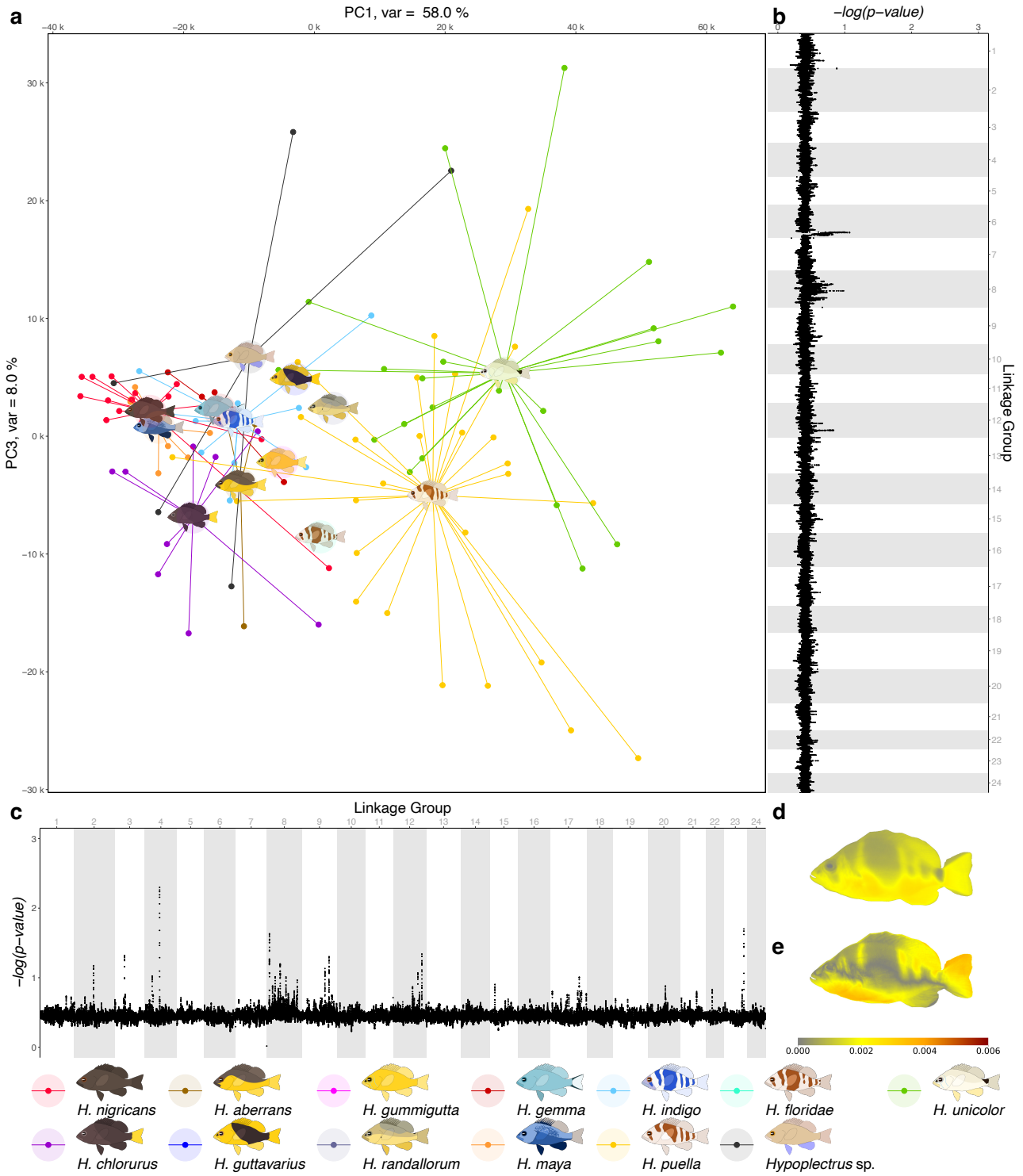



**Figure S2.3** — **a.** Principal Component (PC) 1 *versus* PC3 of the color pattern PCA. **b.** association *-log(p-value)* of the univariate GWAS for PC3 along the hamlet genome. **c.** association *-log(p-value)* of the univariate GWAS for PC1 along the hamlet genome. **d.** heatmap showing the contribution of each pixel to PC1. **e.** heatmap showing the contribution of each pixel to PC3.

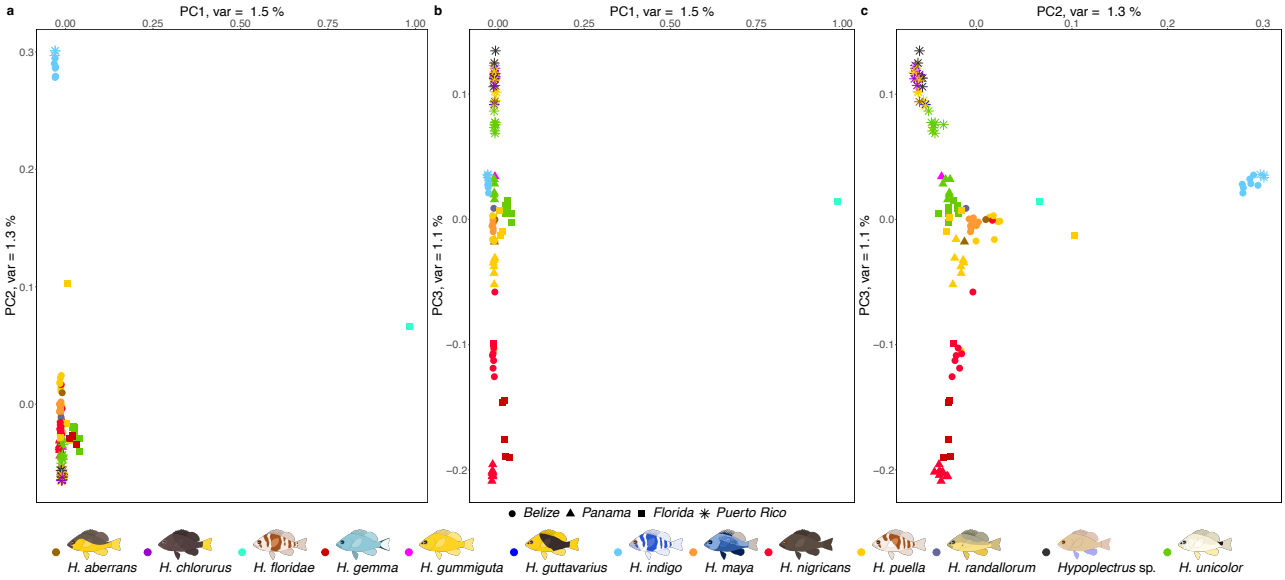



**Figure S2.4** — Principal Component Analysis (PCA) on 289,515 biallelic Single Nucleotide Polymorphism (SNP) markers from 113 *Hypoplectrus* samples spanning 13 species and four locations **a.** Principal Component (PC) 1 *vs.* PC2, **b.** PC1 *vs.* PC3, **c.** PC2 *vs.* PC3

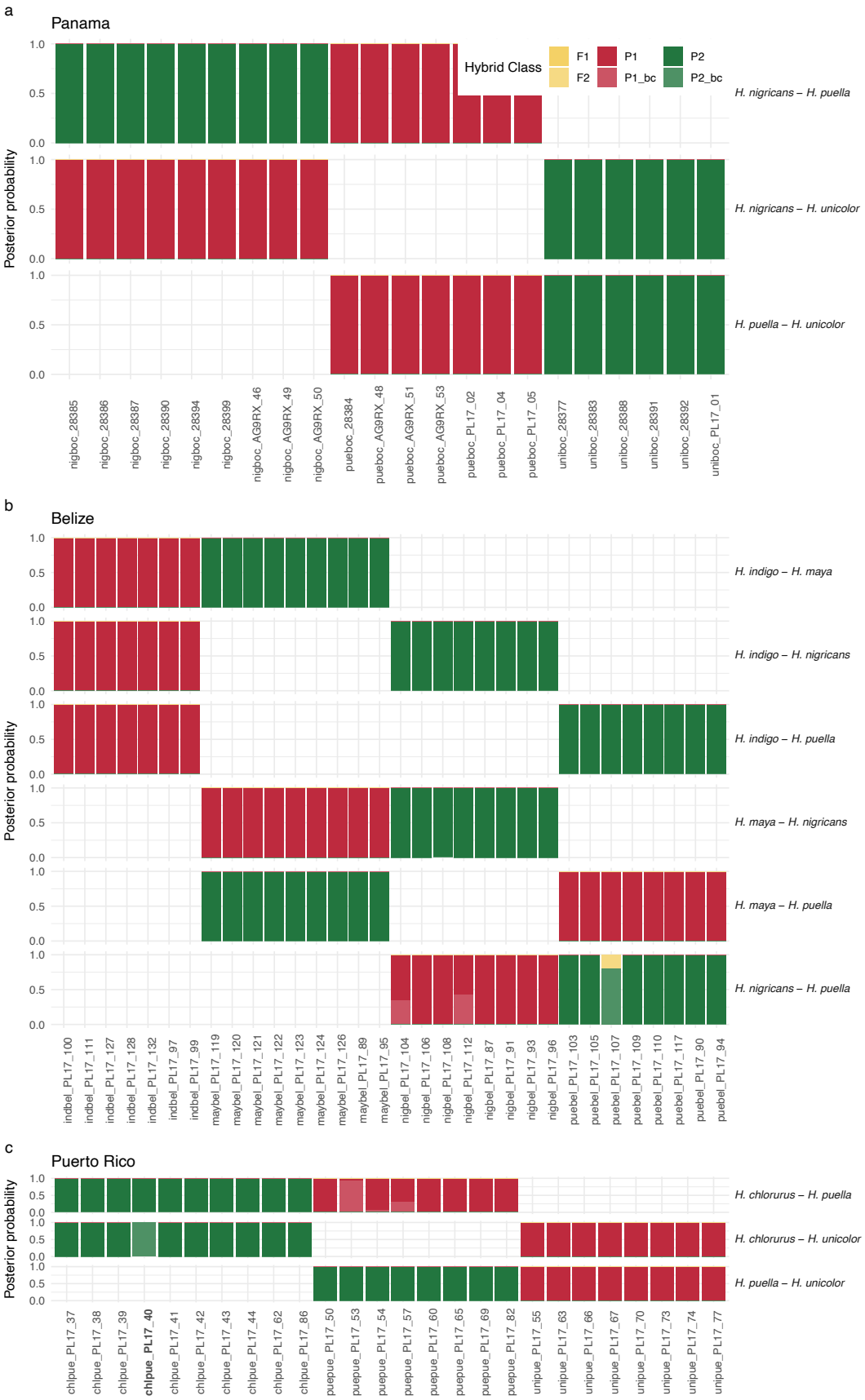



**Figure S2.5** — Identification of putative hybrids and backcrosses in Panama (a), Belize (b) and Puerto Rico (c). Bars indicate the posterior probability of assignment of each individual to the different hybrid classes, with each line corresponding to one pair of sympatric species. P: parental, F1 and F2: first- and second-generation hybrids, respectively, bc: backcrosses. One individual, highlighted in bold, was identified as high-probability backcross (posterior probability >0.99).

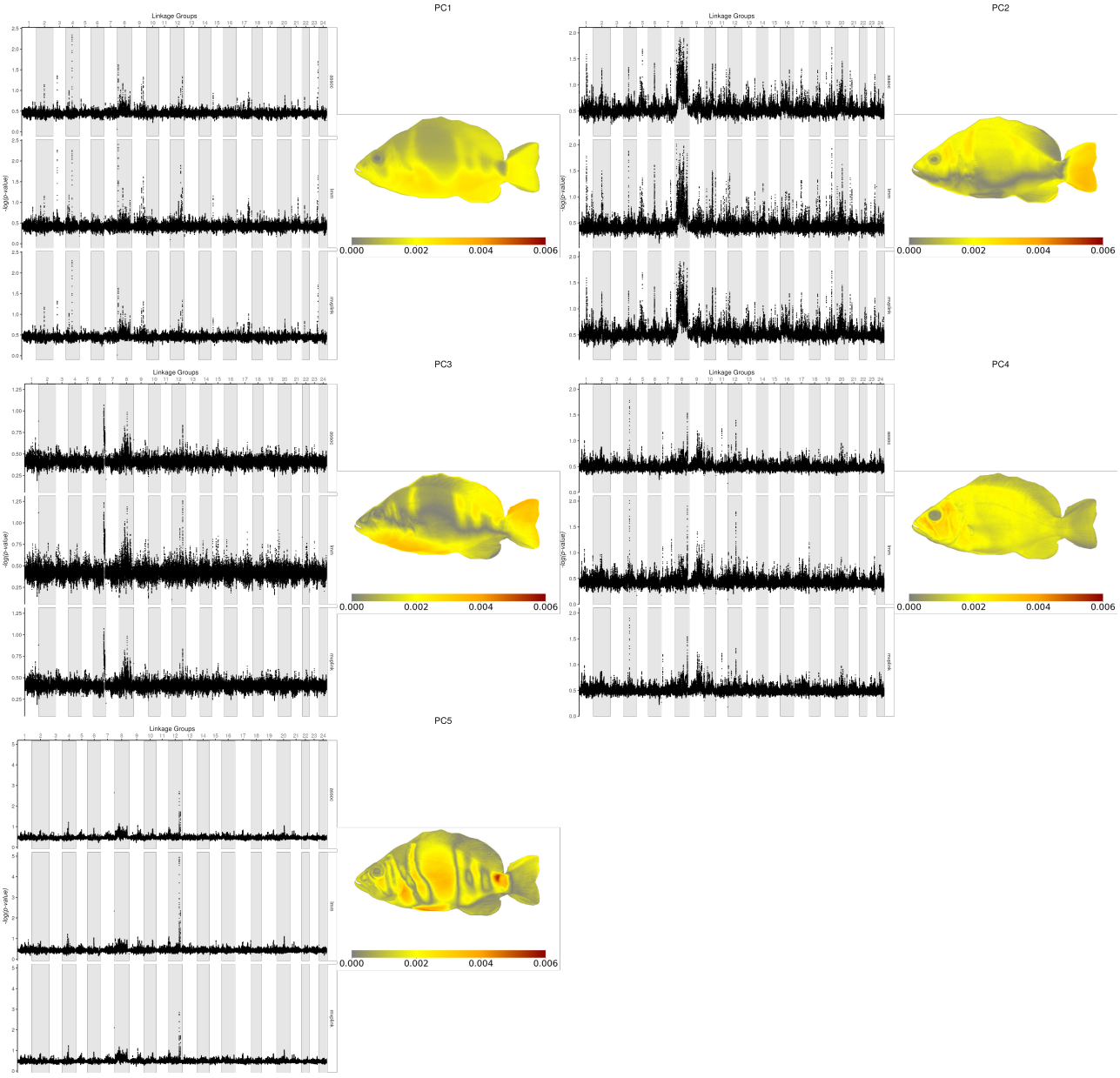



**Figure S2.6** — Comparison of results provided by the two univariate and the multivariate Genome-Wide Association Study (GWAS) methods for each Principal Component (PC) of the image Principal Component Analysis (PCA). Top: PLINK univariate, middle: GEMMA univariate, bottom PLINK multivariate (considering one PC at a time).

## 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, 2004.

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

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

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

Barreto, F. S. and 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. and 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. and 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. and 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. and 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. and Weatherbee, S. D. (2005). From DNA to diversity, molecular genetics and the evolution of animal design. 2 edition, Blackwell Publishing Ltd, Oxford.

Cheney, K. L., Grutter, A. S., Blomberg, S. P. and 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. and 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. et al. (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. et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics *43*, 491.

Domeier, M. L. (1994). Speciation in the serranid fish *Hypoplectrus*. Bull. Mar. Sci. *54*, 103–141.

Ferreira, M. A. and 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. and 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. and 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. and 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. and 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. and 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. and 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. and Puebla, O. (2022). Rapid radiation in a highly diverse marine environment. Proceedings of the National Academy of Sciences *119*.

Hench, K., Vargas, M., Höppner, M. P., McMillan, W. O. and 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. and 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. and 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. and Wang, W.-M. (2016). Expression of *hox* paralog group 13 genes in adult and developing *Megalobrama amblycephala*. Gene Expr. Patterns *21*, 63–68.

Jay, P., Leroy, M., Le Poul, Y., Whibley, A., Arias, M., Chouteau, M. and 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. and 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. and 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. and Joron, M. (2014). Evolution of dominance mechanisms at a butterfly mimicry supergene. Nature communications *5*, 1–8.

Li, H. and 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. and 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. and 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. Proc R Inst Gr Brit *39*, 382–296.

Marques, D. A., Meier, J. I. and 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. and 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. and Cayouette, M. (2015). A conserved regulatory logic controls temporal identity in mouse neural progenitors. Neuron *85*, 497–504.

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

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

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. et al. (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. and 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. and Cheney, K. L. (2022). Higher ultraviolet skin reflectance signals submissiveness in the anemonefish, Amphiprion akindynos. Behavioral Ecology *11*.

Moran, B. M., Hench, K., Waples, R. S., Höppner, M. P., Baldwin, C. C., McMillan, W. O. and 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. and Dasmahapatra, K. K. (2019). The genetic architecture of adaptation: convergence and pleiotropy in Heliconius wing pattern evolution. Heredity *123*, 138–152.

Ogawa, Y. and 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. and 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. and 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. and Wolf, J. B. W. (2015). Transcriptomics of colour patterning and coloration shifts in crows. Mol. Ecol. *24*, 4617–4628.

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

Puebla, O., Coulmance, F., Estapé, C. J., Estapé, A. M. and 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. and 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. et al. (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. and 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-TAIPEI- *44*, 299.

Randall, J. E. and Randall, H. A. (1960). Examples of mimicry and protective resemblance in tropical marine fishes. Bull. Mar. Sci. *10*, 444–480.

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

Rossi, M., Hausmann, A. E., Thurman, T. J., Montgomery, S. H., Papa, R., Jiggins, C. D., McMillan, W. O. and 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. and 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. and Frederich, B. (2019). Magic traits in magic fish: understanding color pattern evolution using reef fish. Trends in Genetics .

Siebeck, U. E., Parker, A. N., Sprenger, D., Mäthger, L. M. and 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. and Cowman, P. F. (2020). Trophic innovations fuel reef fish diversification. Nature communications *11*, 1–11.

Stryjewski, K. F. and 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. Macmillan Company.

Theodosiou, L., McMillan, W. and 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. and Godwin, A. R. (2004). Differences in expression pattern and function between zebrafish *hoxc13* orthologs: recruitment of *hoxc13b* into an early embryonic role. Dev. Biol. *274*, 318–333.

Tian, L., Rahman, S. R., Ezray, B. D., Franzini, L., Strange, J. P., Lhomme, P. and 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 *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. et al. (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. et al. (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. and Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. Evolution *38*, 1358–1370.

Whiteman, E., Côté, I. and 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. and 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. and 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. and Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics *28*, 3326–3328.

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

# 03

# PHENOTYPIC VARIATION WITHIN AND BETWEEN SPECIES DURING EARLY RADIATION STAGES

**Floriane Coulmance,** Melanie J. Heckwolf, Jakob Gismann, Tane Kafle, Omar Domínguez-Domínguez, Karim Awhida, Martin Helmkampf, W. Owen McMillan, Oscar Puebla

Intended submission in *Proceedings of the Royal Society B: Biological Science*.

**Coulmance, F.**, Heckwolf, M. J., Gismann, J., Kafle, T., Domínguez-Domínguez, O., Awhida, K., Helmkampf, M., McMillan, W. O. and Puebla, O. *(in prep).* Phenotypic variation within and between species during the early stages of radiation. Proceedings of the Royal Society B: Biological Sciences.

## Abstract

Recent radiations provide rare windows into the early stages of diversification. With the advent of second- and third-generation DNA sequencing technologies, it is now possible to study this process at the whole-genome level. However, few studies have taken a similarly comprehensive approach at the phenotypic level. This is what we do here, using the hamlets (*Hypoplectrus* spp.) as a model system. These reef fishes from the Greater Caribbean have diversified rapidly into 18+ species that differ essentially in terms of color pattern. Using a recently developed workflow, we analyzed *in situ* photographs of 571 fishes from 14 species at pixel resolution with a fully standardized and automated procedure. The results show that sympatric species form phenotypic clusters that are significantly different from each other, but still exhibit substantial within-species variation and between-species overlap. At the scale of the Greater Caribbean, geographic variation within species further contributes to this overlap, resulting in nearly continuous variation across the entire radiation. A complementary dataset of 327 genomes from 18+ species indicates that similar patterns are observed at the population genomic level. These results demonstrate that sympatric clusters are maintained by selection on a phenotypic and genomic substrate that is largely continuous and shared across species.

## Keywords

phenotypes, variation, genomic, sympatry, *Hypoplectrus*, diversification

## Author's contribution

FC, MJH, JG, TK, ODD, KA, WOM and OP coordinated and performed sample collection. FC, MH and OP conducted laboratory work. WOM and OP provided funds for sequencing. FC, MH and OP analyzed the data and visualized the results. FC and OP wrote the manuscript with input and feedback from all co-authors.

## 3.1 Introduction

Evolutionary radiations play an instrumental role in generating biodiversity, and can sometimes unfold rapidly in evolutionary terms (Schluter, 2000). An iconic example is Darwin's finches, which diversified into 15 different species in 2-3 million years on the Galapagos Islands (Grant and Grant, 2002). In contrast to this historical and rather slow radiation, the Lake Victoria cichlids recently and explosively diverged into ~500 species in the last 15,000-100,000 years (Santos et al., 2023; McGee et al., 2020). They have colonized several distinct ecological niches and make up most of the fish biodiversity in these lakes, representing 5% of all teleost fishes (Santos et al., 2023; McGee et al., 2020). Studying the origins and drivers of such rapid and recent evolutionary radiations therefore contributes to our general understanding of the diversification processes that shape present-day biodiversity.

Advances in the field of sequencing and the accessibility of genomic data have allowed scientists to study diversification mechanisms at the DNA level (Stroud and Losos, 2020). McGee et al. (2020) analyzed DNA sequences of all extant cichlid species using phylogenetics to infer past diversification processes. They showed that genetic variation consisting of many ancient haplotypes in Lake Victoria cichlids was responsible for their explosive radiation when combined with ecological opportunity and sexual selection (McGee et al., 2020). On a smaller scale, species and populations of three-spined sticklebacks have been used to understand the accumulation of reproductive isolation in speciation processes, providing insights into early diversification stages (Hendry et al., 2009; Reid et al., 2021). In these studies, genetics is the primary approach to the understanding of radiations, while the phenotypic perspective is often overlooked.

Phenotypes have been extensively studied in established radiations, allowing the phenotypic variation to be characterized at the family level. For example, Navalón et al. (2020) looked at the craniofacial morphology of 128 land bird families that diverged ~67 Ma (Jarvis et al., 2014), by subsampling one to a few species per family. They showed that radiating lineages such as honeycreepers and Darwin's finches had a tighter beak-skull morphology (Navalón et al., 2020). In the more recent East African cichlid radiations, multi-species studies have described variation and characterized the genetic basis of the repeated melanin stripes (Urban et al., 2022), the orange blotch pattern (Streelman et al., 2003), body and jaw morphology (Muschick et al., 2012) phenotypes. Studies on multi-species and intraspecific variation of early radiation stages are very limited. Gören and Kaymak (2024) showed intraspecific variation in body shape and size between carp populations of a Turkish river, but this was restricted to two species. Overall, we lack a comprehensive understanding of the phenotypic variation substrate for selection at early diversification stages. In this study, we propose to characterize the inter- and intra-specific phenotypic variation of the entire and recently diverged hamlet (*Hypoplectrus* spp.) radiation.

The hamlets provide a rare opportunity to evaluate phenotypic variation in early radiation stages. This genus of small

seabasses from the wider Caribbean has 18 currently described species characterized by their striking variation in color patterns (Lobel, 2011; Puebla et al., 2022; Coulmance et al., 2024). They form an exceptionally recent radiation that has diverged ~10,000 generations ago (Hench et al., 2022; Coulmance et al., *in review*). With up to nine species found in sympatry (Puebla et al., 2022), color patterns play a strong role in the hamlet assortative and visually based mating system and have been suggested to play a role in mimicry and camouflage (Randall and Randall, 1960; Thresher, 1978; Fischer, 1980; Domeier, 1994; Puebla et al., 2007; Barreto and McCartney, 2008; Puebla et al., 2012, 2018; Heckwolf et al., *in review*). Consistent with high levels of sympatry and the occurrence of gene flow, hamlets are highly similar genetically, morphologically and ecologically (Whiteman et al., 2007; Holt et al., 2008). The levels of genetic differentiation among species fall within the range that is typically found among populations within species (Puebla et al., 2007; Barreto and McCartney, 2008; McCartney et al., 2003; Puebla et al., 2008; Holt et al., 2011; Puebla et al., 2014). The hamlet genomic architecture shows sharp peaks of differentiation that stand out against a background of low differentiation (Hench et al., 2019, 2022). The same patterns underlie phenotypic diversity with a few large effect loci responsible for each color pattern element (Coulmance et al., 2024). This follows what is predicted at the early stages of speciation (Wu, 2001; Wu and Ting, 2004), suggesting that hamlets may still be diverging and rather in the very first stages of radiation.

We evaluate the intra- and inter-specific color pattern variation in hamlets using an extensive dataset of 571 *in situ* photographs and leveraging our newly developed pipeline for quantitative analysis of fish color patterns (Coulmance et al., 2024). We show that although sympatric species can form phenotypic clusters, variation is largely continuous at the scale of the entire radiation. We complement our study with a genomic dataset of 327 hamlets and show that similar patterns are observed at the genetic level.

## 3.2 Methods

### 3.2.1 Phenotyping

The phenotypic analyses are based on 571 hamlet photographs collected *in situ* over the course of eight fieldwork efforts in 2017, 2022 and 2023. Hamlets were collected at seven sites across their entire distribution range in Panama, the US Virgin Islands, Tobago, Mexico, Belize and the Florida Keys (Figures 3.1*a*, S3.1). Each photograph represents an individual, and a total of 14 hamlet species are represented in our dataset (Figures 3.1*a*, S3.1 and Table S3.1). This extensive dataset constitutes a major improvement with ~5× more photographs than before, hamlets from the *small clade* and the Gulf of Mexico that were previously missing (Coulmance et al., 2024; Coulmance et al., *in review*), as well as an increase in sample size for each species, allowing the investigation of within-species variation across the whole radiation.

#### 3.2.1.1 Collection sites

The seven sites are: the Keys Marine Laboratory in the Florida Keys (July 2017), the Carrie Bow Cay field station in Belize (May

2017), the Bocas del Toro field station in Panama (February 2017, 2022, 2023), the La Parguera field station in Puerto Rico (March 2017), the Tobago Pigeon Point Fish Facility in Trinidad and Tobago (November 2022), the Virgin Islands of the United States (Saint John, May 2023) and Mexico (Veracruz, June 2023). All field work was conducted under Smithsonian Tropical Research Institute (STRI) Institutional Animal Care and Use Committee (IACUC) protocols 2017-0101-2020-2 and SI-22010 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, SEX/A-35-17 and ARG-0051-2022 and Access and Benefit-Sharing Clearing-House identifier ABSCH-IRCC-PA-241203-1 (Panama), Departamento de Recursos Naturales y Ambientales research permit #2016-IC-127 (E) (Puerto Rico), Division of Food Security Natural Resources the Environment and Sustainable Development permit #_005/2022 (Trinidad and Tobago), United States Department of the Interior National Park Service Virgin Islands study #VIIS-23015 and permit #VIIS-2023-SCI-0011 and Department of Planning and Natural Resources permit DFW23026U (US Virgin Islands) and Dirección General de Ordenamiento Pesquero y Acuícola permit PPF/DGOPA-025/23 (Mexico).

#### 3.2.1.2 Photographs collection

Following the procedure and using the same equipment described in (Coulmance et al., 2024), live hamlet photographs were collected *in situ* maximizing distance, angle, and color standardization. Briefly, a color checker (X-Rite Mini ColorChecker Classic) was included with each photograph to allow for color correction. Photographs were shot at a fixed distance and angle to allow for further image alignment. All photographs were taken in *.CR2* raw format to prevent the irreversible alteration of true colors from *.jpg*. The process takes only a few minutes (typically three minutes) from the time the fish is caught on scuba to the time it is released. It is a highly selective method that does not affect the reef, does not affect the color patterns of the fish or their survival (O. Puebla, unpublished observations).

#### 3.2.1.3 Image processing

Following the image processing steps from Coulmance et al. (2024), photographs were first color corrected in MATLAB, then pre-aligned using a custom Python script, and finally aligned in an alignment software using a two-dimensional structure made of landmarks and Bézier curves (Le Poul et al., 2014; Coulmance et al., 2024). Aligned images were output in *.png* format. All image processing code is available in Coulmance et al. (2024) github folder. Since all images were aligned, a common fish body mask was created in the GNU Image Manipulation Program (GIMP) to remove background pixels. The images were slightly blurred with a 5-by-5 pixel convolution to remove fine-scale texture and converted to the LAB channels color space (L for luminosity, A and B for color). Each image was then flattened into a long one-dimensional vector. The resulting table has 571 rows corresponding to hamlet individuals and 1,030,452 columns containing color information from each LAB channels and corresponding to 343,484 pixels. This table was further

used for statistical analyses.

#### 3.2.1.4 Statistical analyses

In addition to the dataset with all 571 hamlets, two types of data were extracted for further statistical analyses: subsets per site for which ≧29 individuals were available (excluding Puerto Rico) and subsets per species for which ≧5 individuals were available in at least two sites (*H. puella*, *H. nigricans*, *H. unicolor*, *H. chlorurus*, *H. aberrans* and *H. indigo*; Figures 3.1*a*, S3.1).

**Principal Component Analysis (PCA)** was performed to reduce and visualize the hamlet phenotypic space into interpretable axes of phenotypic variation, the principal components (PC). PCA was performed using $sklearn$ in Python (Pedregosa et al., 2011) and PCs were limited to 15. Results were visualized in two PC dimension plots in R *v4.2.2* using $ggplot$ *v3.5.1* and related packages. Ellipses were drawn with $stat\_ellipse$ (Fox and Weisberg, 2018; Friendly et al., 2013) assuming a multivariate t-distribution for groups with ≧4 individuals and at a 0.95 confidence level. This allowed us to assess the presence of phenotypic clusters of species across the whole hamlet radiation. Heatmaps showing the contribution of pixels to variation along PC were generated using a custom Python script. This facilitates the interpretation of variation in color and pattern elements on the hamlet body.

**Permutational multivariate analysis of variance (PERMANOVA)** was used to test the overlap between phenotypic clusters of species for the entire dataset and per-site subsets. For per-specie subsets, phenotypic clusters were compared between sites. Over all groups and pairwise PERMANOVA were performed with $adonis2$ (Oksanen et al., 2024) and $pairwise.adonis$ respectively (Martinez Arbizu, 2020). Euclidean distance matrices were computed between points taking into account the 15 PCs and, the number of permutations was set to 10,000. Groups with <5 individuals were excluded.

**Hierarchical clustering** was used to assess the presence of phenotypic clusters for >2 axes of phenotypic variation. Euclidean pairwise distance matrices were computed using the 15 PCs and then used to perform clustering with the Ward method.

### 3.2.2 Population genetics

The population genetic analyses are based on a partially overlapping dataset of 327 individual hamlet genomes gathered for a previous study (Coulmance et al., *in review*). The dataset covers 15 sites across the entire hamlet distribution range (Figure 3.1*b*; Coulmance et al., *in review*). It represents 18+ hamlet species (Figure 3.1*b*; Puebla et al., 2022; Puebla et al., *submitted*). All subsequent computational analyses were performed on the high performance computing cluster ROSA (University of Oldenburg) using Linux shell scripts and Slurm *v23.02.4*. Results were visualized locally in R *v4.2.2* using $ggplot$ *v3.5.1* and related packages.

#### 3.2.2.1 Genotyping

Genomic DNA extraction, sequencing and genotyping were available from previous studies (Hench et al., 2019; Puebla et al., 2022; Hench et al., 2022; Coulmance et al., 2024; Coulmance et al., *in*

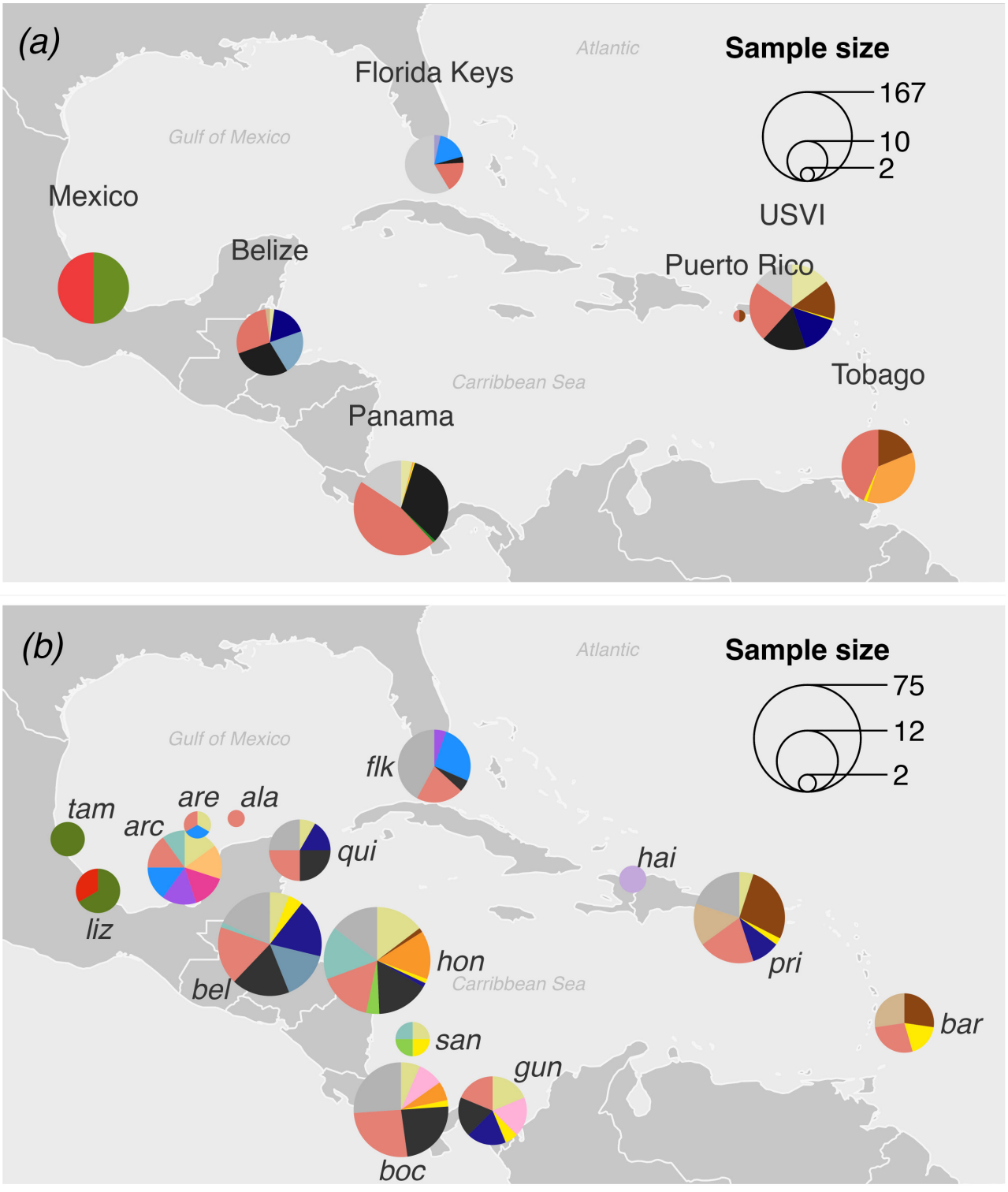



**Figure 3.1** — Map of the sampling design for *(a)* the phenotypic dataset with *n=571* hamlet individuals collected at 7 sites: Panama, the US Virgin Islands, Tobago, Mexico, Belize and the Florida Keys, and *(b)* the genetic dataset with *n=327* hamlet individuals collected at 15 sites covering the Gulf of Mexico [Tamiahua (tam), Antón Lizardo (liz), Cayos Arcas (arc), Cayo Arenas (are), Alacranes Reef (ala) and the Florida Keys (flk)], the Western Caribbean [Quintana Roo (qui), Belize (bel), Honduras (hon), San Andrés (san), Bocas del Toro (boc) and Guna Yala (gun)], and the Eastern Caribbean [Haiti (hai), Puerto Rico (pri) and Barbados (bar)].

*review*). Briefly, genomic DNA was extracted from fin clips and gills stored in 100% ethanol at –20°C with the MagAttract high-molecular-weight DNA extraction kit (Qiagen). Illumina DNA Prep libraries were prepared and sequenced on three NovaSeq6000 S4 lanes (2 × 151 bp) at the Institute of Clinical Molecular Biology (IKMB, Kiel University). Genotyping was performed jointly using GATK *v4.1.9* and following its best practice guidelines (McKenna et al., 2010), and the procedure described in Coulmance et al. (*submitted*). The resulting genotyping file was filtered for minimum allele count ≧2 and reduced to retain only bialleic Single Nucleotide Polymorphisms (SNP) using VCFtools (Danecek et al., 2011).

#### 3.2.2.2 PCA

The SNP dataset was retrieved from Coulmance et al. (*submitted*) and used in a PCA to assess the presence of genetic clusters of species across the entire hamlet radiation and in sympatry at sites for which ≧40 individuals were available (Honduras, Belize, Panama and Puerto Rico; Figures 3.1*b*, S4.1). Linkage disequilibrium was accounted for in the PCAs by recursively removing SNPs within a sliding window for which the pairwise correlation >1 (Zheng et al., 2012).

## 3.3 Results

### 3.3.1 Phenotypic PCA

#### 3.3.1.1 Sympatric cluster of species

We first evaluate the presence of species clusters in the phenotypic space at each site separately. The PC combinations with the best species clustering are shown (Figure 3.2).

In Panama, PC1 and PC3 account for 68.4% and 5.7% of the phenotypic variation, respectively, and show clusters of *H. unicolor*, *H. puella*, and *H. nigricans* (Figure 3.2*a*). Further PCs allow the significant separation of *H. unicolor*, *H. puella*, *H. nigricans* and *H. aberrans* clusters with pairwise PERMANOVA p-value=0.0006 for all pairs except for *H. aberrans vs. H. puella*, where p-value=0.0012 (Figures S3.2*a*, S3.3*a* and Table S3.2).

In the US Virgin Islands, PC1 and PC4 account for 62.9% and 2.8% of the color pattern variation, respectively. The combination of PC1 and PC4 and the hierarchical clustering reveal phenotypic clusters of species (Figures 3.2*b* and S3.2*b*), and the pairwise PERMANOVA confirms that these clusters are significantly non-overlapping (p-value=0.0015; Table S3.2, Figure S3.3*b*).

In Tobago, PC1 and PC3 explain 40.8% and 17.5% of the phenotypic variation in color patterns, respectively. *H. puella*, *H. chlorurus* and *H. gummigutta* form distinct phenotypic clusters and are significantly separated (pairwise PERMANOVA p-value=0.0003; Figures 3.2*c*, S3.2*c*, S3.3*c* and Table S3.2).

In Mexico, PC1 and PC2 account for 80.6% and 11.5% of the phenotypic variation, respectively, and show distinct clusters corresponding to *H. atlahua* and *H. castroaguirrei* (Figure 3.2*d*). These clusters are further confirmed by hierarchical clustering and pairwise PERMANOVA (p-value=9.9e-05; Figures S3.2*d*, S3.3*d* and Table S3.2).

In Belize, PC1 and PC2 represent 67.8% and 14.0% of the color pattern variation, respectively, and cluster *H. indigo*, *H.*

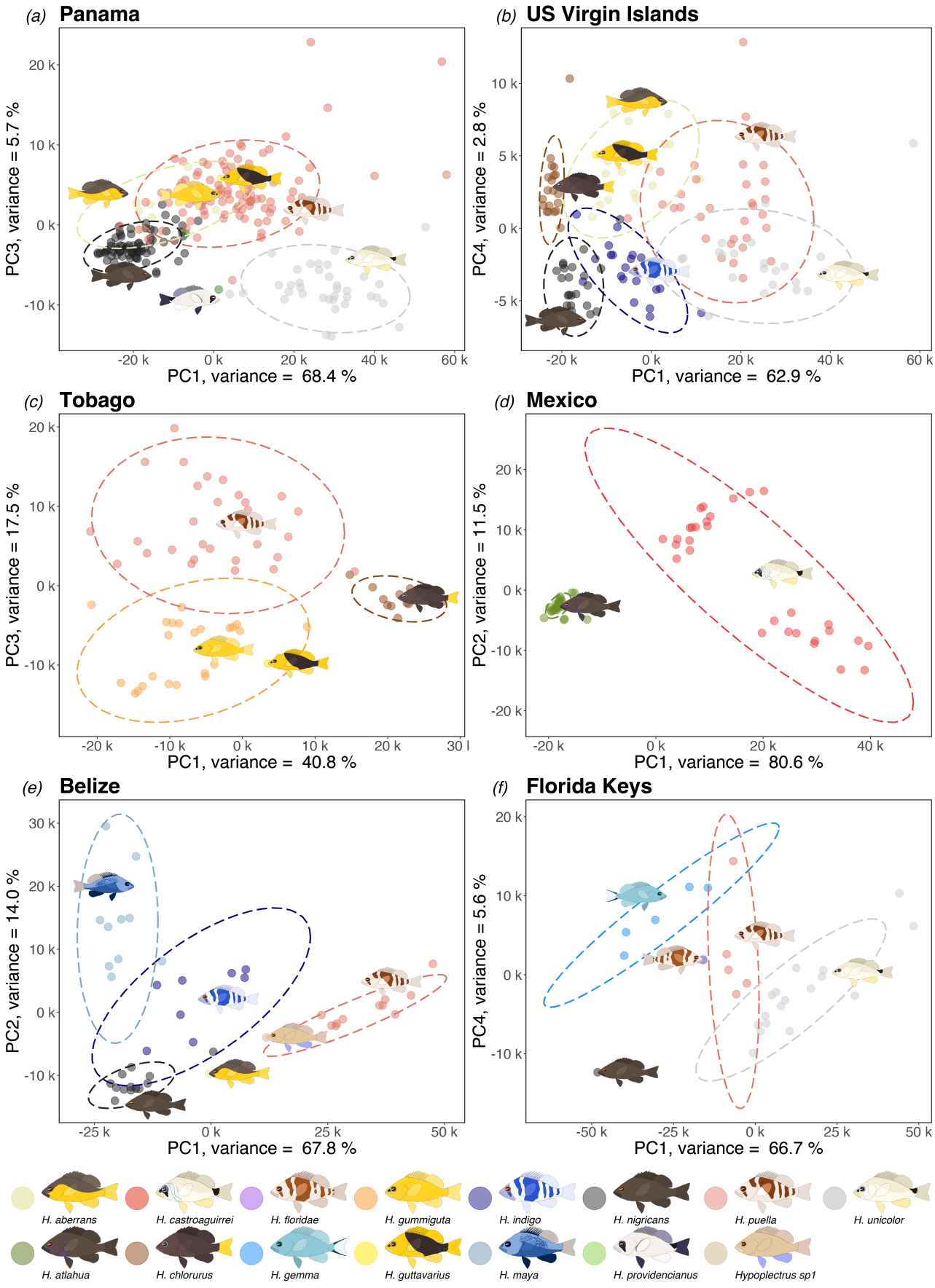



**Figure 3.2** — Phenotype PCA for *(a)* Panama, *(b)* US Virgin Islands, *(c)* Tobago, *(d)* Mexico, *(e)* Belize, and *(f)* the Florida Keys.

*maya*, *H. nigricans*, and *H. puella* (Figure 3.2*e*). While the hierarchical clustering over 15 PCs shows a mixture of clusters with some individuals of *H. puella* and *H. maya* scattered throughout the tree (Figure S3.2*e*), the pairwise PERMANOVA significantly separates these species (p-value=0.0006; Table S3.2, Figure S3.3*e*).

In the Florida Keys, PC1 and PC4 account for 66.7% and 5.6% of the phenotypic variation, respectively, and show clusters of *H. gemma* and *H. unicolor* (Figure 3.2*f*). The hierarchical clustering over 15 PCs shows a mixture of *H. gemma*, *H. puella* and *H. unicolor* (Figure S3.2*f*). The pairwise PERMANOVA significantly separates *H. gemma* from *H. puella* (p-value=0.0264), *H. puella* from *H. unicolor* (p-value=0.00024), and *H. gemma* from *H. unicolor* (p-value=0.0006; Table S3.2, Figure S3.3*f*). The over all groups PERMANOVA shows a higher significant separation of species (p-value=9.9e-05) as compared to the pairwise results between sympatric species (Table S3.2).

#### 3.3.1.2 The whole hamlet radiation

Considering the entire hamlet phenotypic dataset (*n=571*, 14 species and 7 sites), PC1 and PC4 best visualized species clusters, explaining 62.1% and 4.6% of the color pattern variation, respectively (Figure 3.3*a*). While the hierarchical clustering seems to show phenotypic clusters of species (Figure 3.3*b*), we observe large overlap between species in the phenotypic space (Figure 3.3*a*). Most pairwise comparisons show significant non-overlap between species clusters with a p-value=0.0055. *H. castroaguirrei vs. H. gemma* (p-value=0.011), *H. castroaguir-*

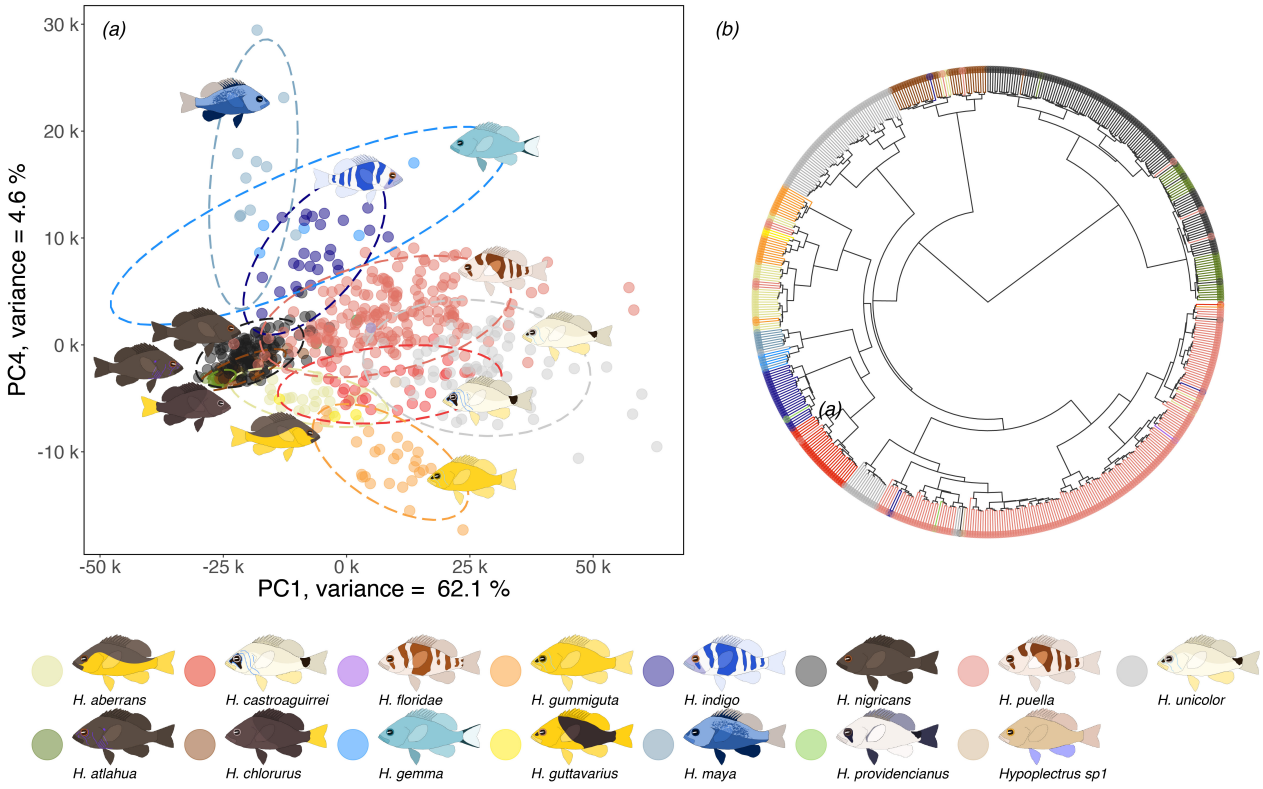



**Figure 3.3** — Phenotype PCA over all sites *(a)* PCA clusters of phenotypes grouped by species, x-axis represents variation along PC1 and y-axis represents variation along PC4, and *(b)* hierarchical clustering based on 15 PCs and colored by species.

*rei vs. H. gummigutta* (p-value=0.039), *H. gemma vs. H. indigo* (p-value=0.022), and *H. gemma vs. H. puella* (p-value=0.022) cluster significantly less than other pairs (Table S3.3, Figure S3.5). *H. gemma vs. H. maya* (p-value=0.071) do not significantly separate into distinct clusters (Table S3.3, Figure S3.5). The PERMANOVA on all confounded groups shows a significant species separation (p-value=9.9e-05; Table S3.3).

### 3.3.2 Genetic PCA

#### 3.3.2.1 Sympatric cluster of species

We evaluate the presence of species clusters in the genetic space at each site separately. The combination of PCs showing the best species grouping is shown (Figure 3.4).

In Honduras, PC1 and PC2 explain 1.7% and 1.5% of the genetic variation, respectively, and separate *H. gummigutta* and *H. indigo* from the other hamlet species (Figure 3.4*a*). PC3 further separates *H. providencianus* and accounts for 1.4% of the genetic variation (Figure 3.4*b*).

In Belize, PC1 and PC2 account for 2.0% and 1.8% of the variation in genomes, respectively, and separate *H. indigo*, *H. maya*, and *H. nigricans* from the other species (Figure 3.4*c*). PC3 and PC4, which explain 1.7% and 1.6% of the genetic variation, respectively, confirm these clusters (Figure 3.4*d*), but do not allow the separation of other hamlet species.

In Panama, PC1 and PC2 represent 2.5% and 2.4% of the genetic variation, respectively, and separate most of the species: *H. gummigutta*, *H. affinis*, *H. nigricans*, *H. puella*, and *H. unicolor* (Figure 3.4*e*).

In Puerto Rico, *H. indigo* is the only species separated from the others by PC1 (variance=2.8%; Figure 3.4*g*).

#### 3.3.2.2 The whole hamlet radiation

Considering the entire hamlet genetic dataset (*n=327*, 18+ species and 15 sites), the first six PCs allow the separation of different hamlet species (Figure 3.5). PC1, which accounts for 1.1% of the genetic variation, separates species of the *small clade* (*H. atlahua*, *H. ecosur*, and *H. floridae*) from the rest of the hamlet species (Figure 3.5*a*; Coulmance et al., *in review*). PC2, which accounts for 0.9% of the genetic variation, clusters individuals of *H. indigo* together (Figure 3.5*a*). PC3 with 0.7% of the genetic variation separates individuals of *H. gummigutta* (Figure 3.5*b*). PC4 (variance=0.6%) clusters the Panamanian population of *H. nigricans* away from the rest of the dataset (Figure 3.5*b*). PC5 separates *H. maya* from the rest with 0.5% of genetic variation explained (Figure 3.5*c*).

## 3.4 Discussion

The large phenotypic (*n=571*) and genomic (*n=327*) datasets in this study allowed us to examine the substrate of variation across the entire hamlet radiation, spanning both local and their whole distribution range. We found that hamlets form phenotypic clusters of sympatric species at the local scale, with substantial overlap between species due to intraspecific variation. Our population genetic analyses revealed similar patterns, where there were sympatric clusters of species with varying degrees of overlap. At the scale of the entire radiation, these patterns were partially lost for both phenotypes and genotypes.

### 3.4.1 Phenotypic and genetic variation depends on scale

Hamlets form phenotypic clusters in sympatry (Figures 3.2, S3.2, S3.3 and Table S3.2). The maintenance of these clusters may be due to the interaction of both sexual and natural selection on hamlet color patterns at the local scale. Color pattern is involved in the visual mating system of hamlets (Barreto and McCartney, 2008; Puebla et al., 2012) and has been suggested to play a role in their predatory behavior (Puebla et al., 2007). While little to no ecological differences have been recorded in hamlets (Whiteman et al., 2007; Holt et al., 2008), differences in ecological niche and predatory behavior could represent potential for divergence. In Mexico, *H. atlahua* and *H. castroaguirrei* occupy shallower rocky reef plateau (10 meters) and deeper colder reefs (10-20 meters), respectively (observations by F. Coulmance), suggesting subtle habitat differences in sympatry. In Panama, Heckwolf et al. (*under revision*) showed that *H. puella* and *H. nigricans* have different behavioral response to predators and that this response is related to camouflage. Such selective forces have been highlighted as driving multiple adaptive radiations ranging from birds to fish (Grant and Grant, 2007; Wagner et al., 2012). In Darwin's finches, natural selection on beak size and shape, driven by food constraints, has contributed to their divergence across different islands (Grant and Grant, 2007). The interaction of sexual and natural selection with ecological opportunity was also key to the diversification of the African lake cichlids (Wagner et al., 2012). While the adaptive aspect of the hamlet radiation remains speculative (Picq et al., 2016; Hench et al., 2017), its recent origin and speciation rate (Hench et al., 2022; Coulmance et al., *in review*) suggest that phenotypic diversity in the hamlets was generated rather rapidly. Given the estimates of large effective population sizes in the hamlets (Coulmance et al., *in review*), it is unlikely that long neutral processes, such as mutation accumulation and genetic drift could explain the observed diversity (Marques et al., 2019). Therefore, the role of selection in maintaining local divergence during the early stages of the hamlet rapid radiation cannot be ignored.

While clusters of sympatric species are maintained locally, we found a substantial amount of within-species variation and overlap between species. The degree of overlap varied, depending on site and species. This result is consistent with previously reported spatial variation within species (Thresher, 1978; Aguilar-Perera, 2003). At the global scale, most phenotypic clusters were lost and there was an increased amount of overlap between species (Figures 3.3*a*, S3.5*b* and Table S3.3). The varying degrees of spatial variation at local scales add to the within-species variation and contribute to the increasing overlap between species at the global scale. Jordan and Evermann (1896) noted that "*nearly all the forms of* Hypoplectrus *constitute but a single species, subject to almost endless variations in color*". This is true at the whole radiation level, where clusters are lost, but false at the sympatric level, where clusters are maintained. The species with the largest overlap always involved a hamlet with a vertical barred phenotype – *H. puella* or *H. indigo* (Figure S3.3). This is consistent with both the ancestral hamlet species being barred (Coulmance et al., *in re-*

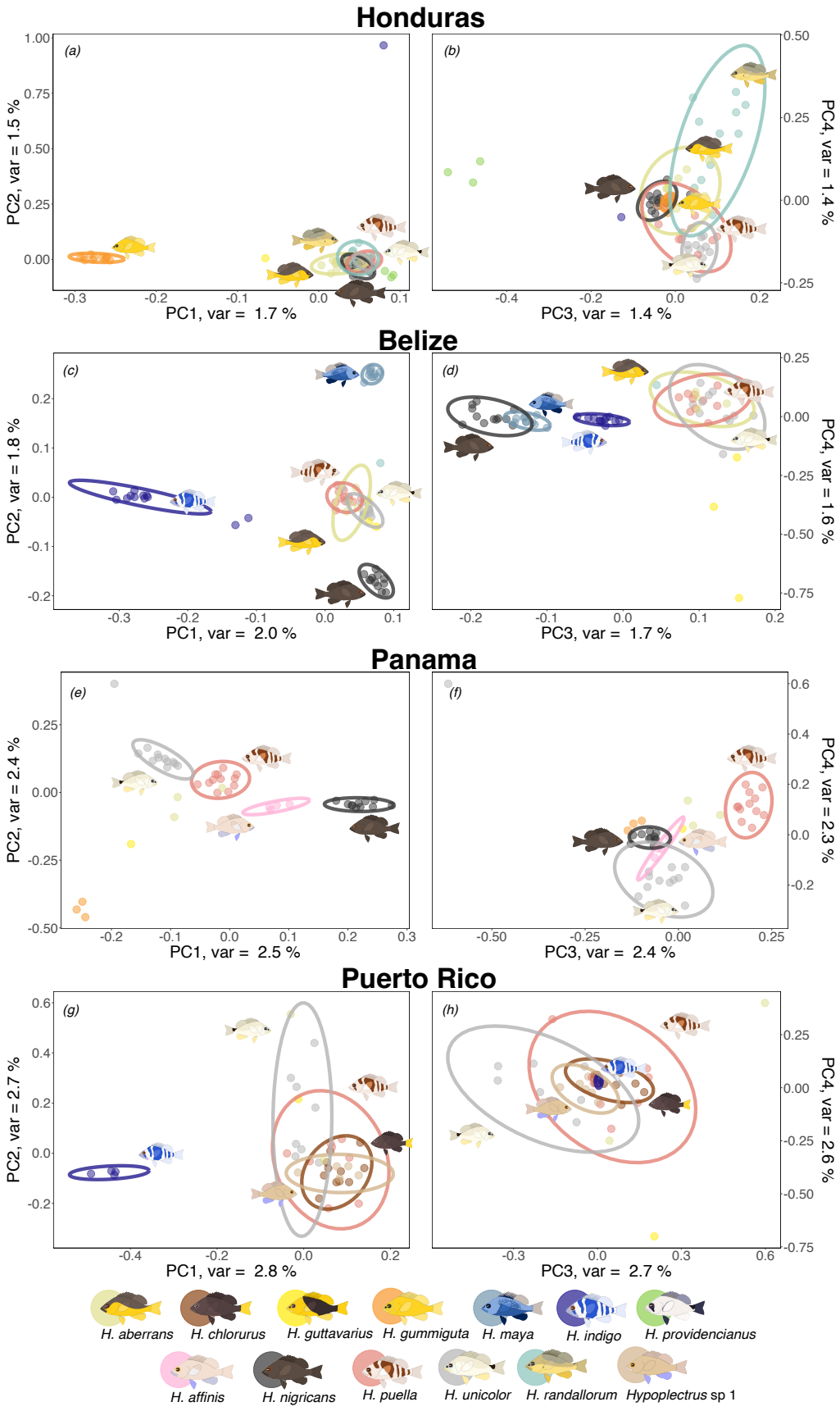



**Figure 3.4** — Genetic PCA per site for *(a), (b)* Honduras with 75 individual hamlets and based on 100,382 SNPs, *(c), (d)* Belize with 66 individual hamlets and based on 84,072 SNPs, *(e), (f)* Panama with 46 individual hamlets and based on 54,080 SNPs, and *(g), (h)* Puerto Rico with 40 individual hamlets and based on 45,954 SNPs. Each dot represents genetic data from one individual hamlet and ellipse are drawn for hamlet species where ≥4 individuals are available.

*view*), and *H. puella* being the most abundant species (Aguilar-Perera, 2003) and therefore the most phenotypically variable. Similar spatial within-species variation in introduced and common carp of Turkish rivers has been shown to be a result of rapid local adaptation (Gören and Kaymak, 2024).

At the population genetic level, some well-defined clusters of sympatric species were recovered – *H. indigo* and *H. gummigutta* – and other species tended to overlap greatly. Again, within-species spatial variation can contribute to these overlaps. Within *H. nigricans*, spatial genetic variation has been reported (Figure 3.5*b*; Coulmance et al., *in review*). In fishes such as Atlantic killifish, sockeye salmon, and rainbow trout, intraspecific genetic and phenotypic variation related to warming temperature tolerance is heterogeneous among sites (McKenzie et al., 2021). One observation is that species from the *large clade* and *small clade*, such as *H. nigricans* and *H. atlahua*, were not clearly separated in the phenotypic space (Figure 3.3), whereas they were in the genetic space (Figure 3.5; Coulmance et al., *in review*). Similar patterns have been observed in *Heliconius* butterflies with their Müllerian mimicry, where genetically divergent species have converged on similar wing patterns (Kronforst and Papa, 2015). Overall, these results suggest that local processes have shaped the observed phenotypic and genetic variation in hamlets.

### 3.4.2 The substrate for selection and species concepts

In the context of speciation, our results describe in detail how phenotypic and genetic variation serve as substrates on which selection can act upon during early stages of a radiation. While we have divergence between sympatric species, within-species variation contributes to species overlap on a broader scale. Ancestral variation (standing genetic variation) and hybridization are mechanisms that have been shown to influence phenotypic and genetic variation in Hawaiian silverswords, African lake cichlids, and Darwin's finches (Barrier et al., 1999; Meier et al., 2017; Selz et al., 2014; Marques et al., 2019). In African lake cichlids, such processes increased genetic and phenotypic diversity, allowing new species to emerge and fueling their explosive radiation, especially combined with new ecological opportunities (Meier et al., 2017; Selz et al., 2014; Marques et al., 2019) . Hamlets have a recent origin and their demographic history shows both ancestral and ongoing gene flow (Coulmance et al., *in review*). Combined with the modular architecture of color patterns at both the genetic and trait levels, such similar mechanisms may be driving the hamlet radiation.

Our results indicate the importance of local processes in generating variation rapidly at the scale of an entire radiation. Variation between sites contributes to within-species variation (Figure S3.6). Hamlets have a pelagic larval stage that allows them to disperse. However, dispersal is limited and constrained by oceanic factors, making it unlikely that populations of the same species are completely homogenized at such large scales (Salzburger, 2008). In this respect, hamlets can be compared to species of Darwin's finches present on different islands, which have evolved separately and ex-

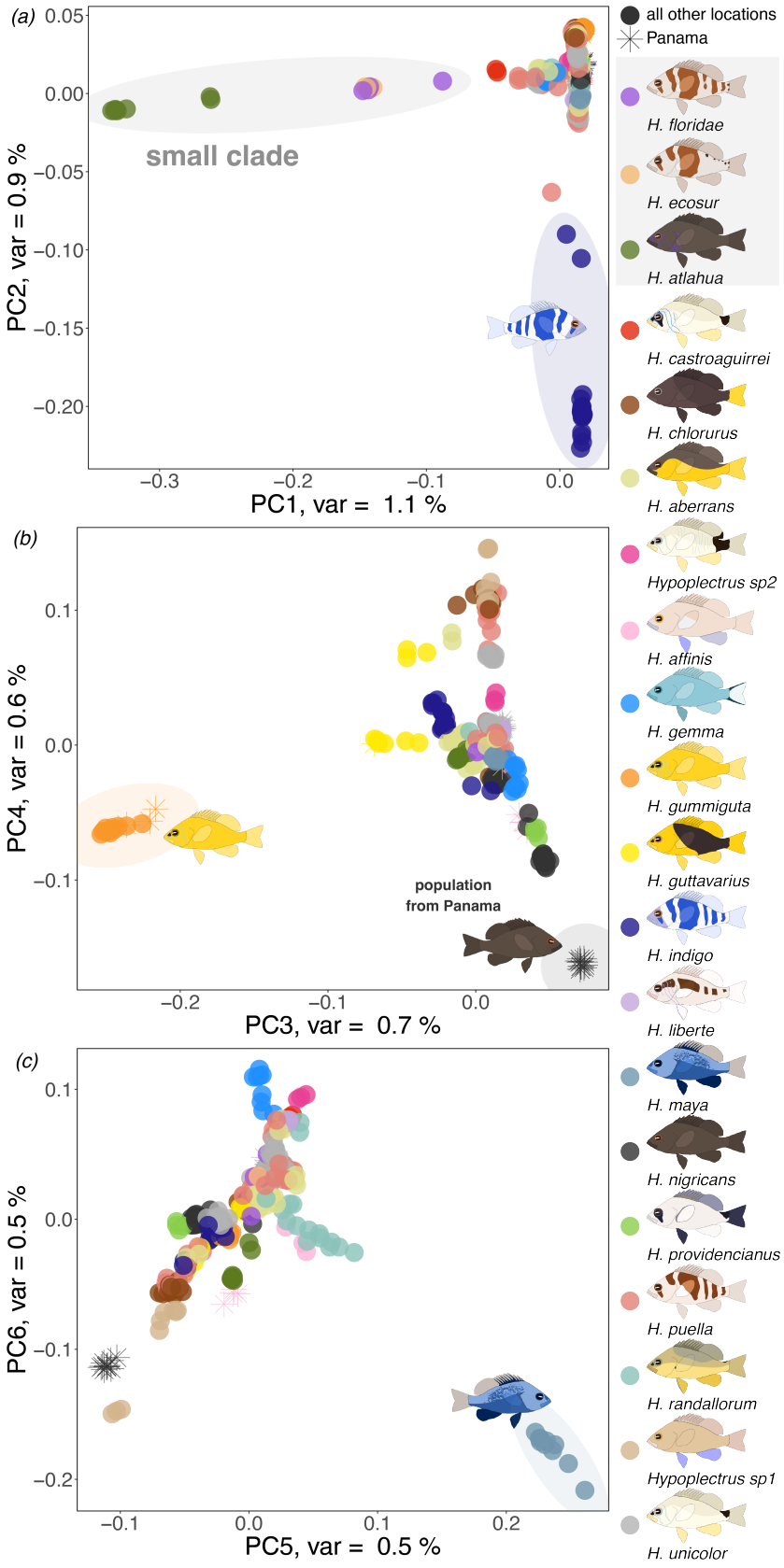



**Figure 3.5** — Genetic PCA based on 620,539 SNPs, each dot represent 1 of the 327 individual hamlets. *(a)* PC1 *vs.* PC2 separates species from the small and large clade and *H. indigo*, *(b)* PC3 *vs.* PC4 separates *H. gummigutta* and the *H. nigricans* population from Bocas del Toro in Panama, *(c)* PC5 *vs.* PC6 separates *H. maya*.

hibit intra-specific variation. Another implication of local processes is the convergent evolution of similar phenotypes. It is possible that the barred phenotypes—*H. puella, H. floridae, H. ecosur* and *H. liberte*— or the butter phenotypes—*H. unicolor, H. castroaguirrei* and *H.* sp2— have convergently evolved across sites (Coulmance et al., *in review*; Puebla et al., *submitted*). At the scale of the entire radiation, repeated phenotypic evolution between the *large clade* and *small clade* has resulted in similar black phenotypes in *H. nigricans* and *H. atlahua* (Figures 3.3, 3.5; Coulmance et al., *in review*). The convergent evolution of melanin stripes in cichlids or wing patterns in *Heliconius* butterflies are examples where local processes have shaped sympatric speciation, while contributing to variation at the scale of an entire radiation (Urban et al., 2021; Kronforst and Papa, 2015).

Our results provide a clear example of how species concepts that make sense at the local scale can become irrelevant when considered at a larger geographic scale (Mallet, 2001). While both the phenotypic (Sokal and Crovello, 1970) and genetic cluster view of species (Mallet, 1995) hold in the case of sympatric speciation, they do not hold at the larger scale for hamlets. Similarly, the Phylogenetic Concept of Species (Cracraft, 1983) does not apply to hamlets at the scale of the entire radiation (Coulmance et al., *in review*; Losos and Glor, 2003; Hahn and Nakhleh, 2016). In general, hamlets represent an extremely early stage along the speciation continuum (Stankowski and Ravinet, 2021; Bolnick et al., 2023), and fill a previously unoccupied spot among radiations, being characterized by little genetic divergence but with marked reproductive isolation and phenotypic divergence (Mallet, 2001; Losos and Glor, 2003; Salzburger, 2008; Puebla et al., 2008).

## Conclusion

Using a comprehensive and high-resolution approach, we showed that the phenotypic variation basis for selection at early radiation stages is largely continuous in hamlets, and that this continuum is driven by within-species variation. Similar patterns were observed at the population genetic level. Despite the presence of sympatric species clusters, the phenotypic and genetic variation substrates remain rather continuous. Our results highlight the importance of local processes in maintaining phenotypic and genetic clusters of species.

**Acknowledgments:** We would like to thank D. Ross Robertson for his immense help in organizing the fieldwork. We are also grateful to the permit team at the Smithsonian Tropical Research Institute, Kelly Kingon, Shomari Charles, Randy Davis, Irwin Nicholls (Maestro), Allison & Carlos Estapé, Andre M Hernandez, Serenity Mitchell, Alasdair Dunlap-Smith, Sara & Lee Richter, Cesar Gutierrez, Liam Brister and Bogdan for their assistance with fieldwork logistics.

**Funding:** This study was funded by an individual grant from the German Research Foundation (DFG): OP625 (PU571/1–2). **FC** was independently supported by the Smithsonian Tropical Research Institute with the D. Ross Robertson Fellowship award, made possible by a donation from Dr. Robertson to the Smithsonian through the Entropy Foundation. Data analyses were performed on

the high performance computing cluster ROSA627 (University of Oldenburg, Germany), funded by the DFG through its Major Research Instrumentation Programme (INST 184/225-1 FUGG) and the Ministry of Science and Culture (MWK) of the state of Lower Saxony.

**Conflict of interests:** The authors declare that they have no conflicts of interest.

**Data Accessibility:** Raw photographs will be available upon publication. Sequence data were deposited in the European Nucleotide Archive (ENA) under project accession numbers PRJEB75494, PRJEB74501, PRJEB76552. Individual sample accession numbers are listed in Table S4.1 and Table S4.3 (Coulmance et al., *in review*). The code used for the phenotypic and genetic analyses is novel and publicly available at chapter3 github folder.

**Ethics:** This work required ethical approval from the STRI Animal Care and Use Committee (ACUC), which was obtained under the number SI-22010. All research permits are listed in the methods section (3.2).

**Declaration of AI use:** We did not use AI-assisted technologies to create this article.

## 3.5 Supplementary Material

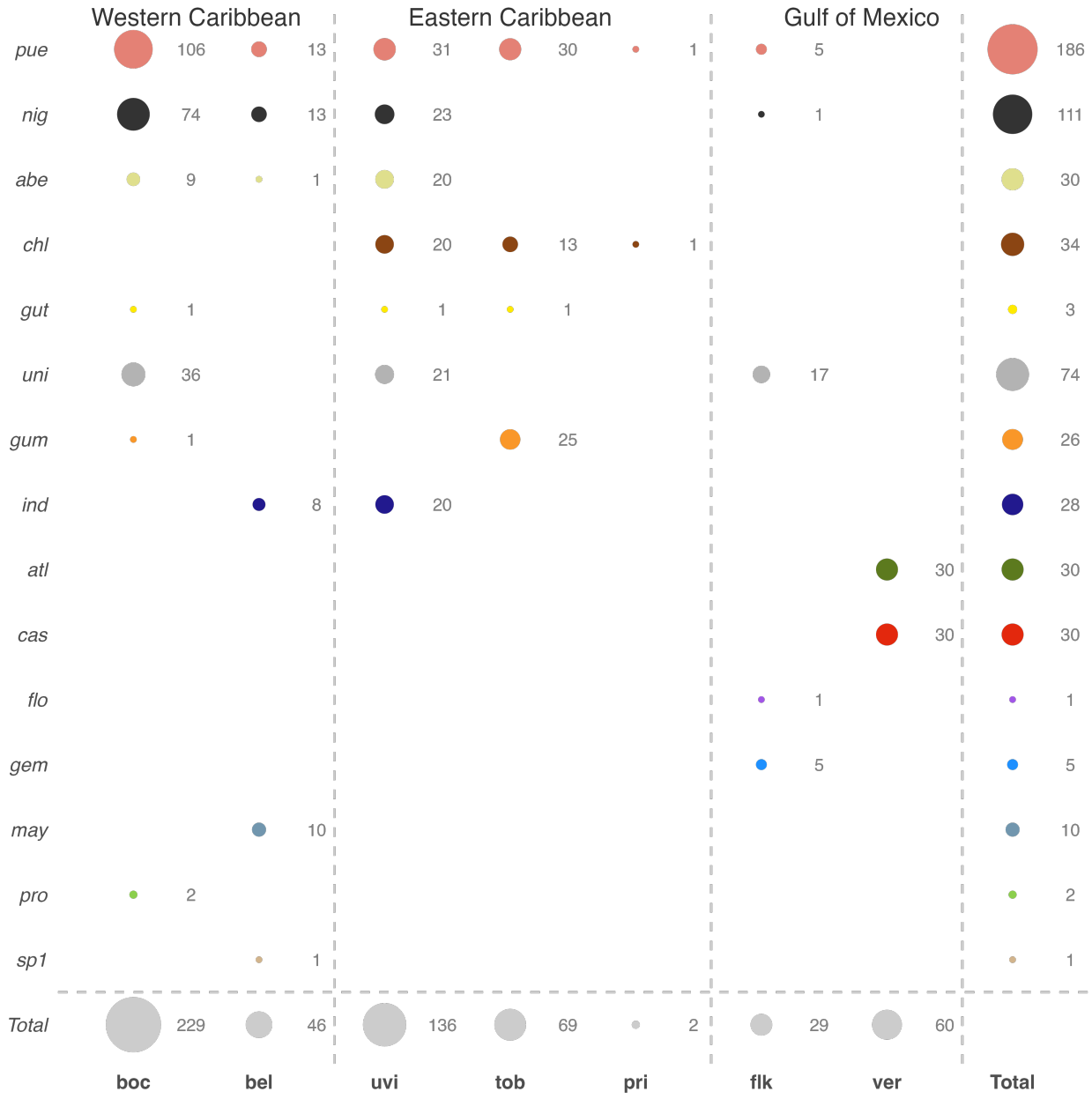



**Figure S3.1** — Overview of the sampling design for phenotypes, comprising 571 samples. pue: *Hypoplectrus puella*, uni: *Hypoplectrus unicolor*, nig: *Hypoplectrus nigricans*, abe: *Hypoplectrus aberrans*, ind: *Hypoplectrus indigo*, chl: *Hypoplectrus chlorurus*, gum: *Hypoplectrus gummigutta*, gut: *Hypoplectrus guttavarius*, may: *Hypoplectrus maya*, gem: *Hypoplectrus gemma*, sp1: *Hypoplectrus sp. 1*, atl: *Hypoplectrus atlahua*, flo: *Hypoplectrus floridae*, pro: *Hypoplectrus providencianus*, cas: *Hypoplectrus castroaguirrei*. Samples were collected at 7 sites covering the Gulf of Mexico [Veracruz (ver) and Florida Keys (flk)], the Western Caribbean [Belize (bel) and Bocas del Toro (boc)], and the Eastern Caribbean [Puerto Rico (pri), US Virgin Islands (uvi), and Tobago (tob)].

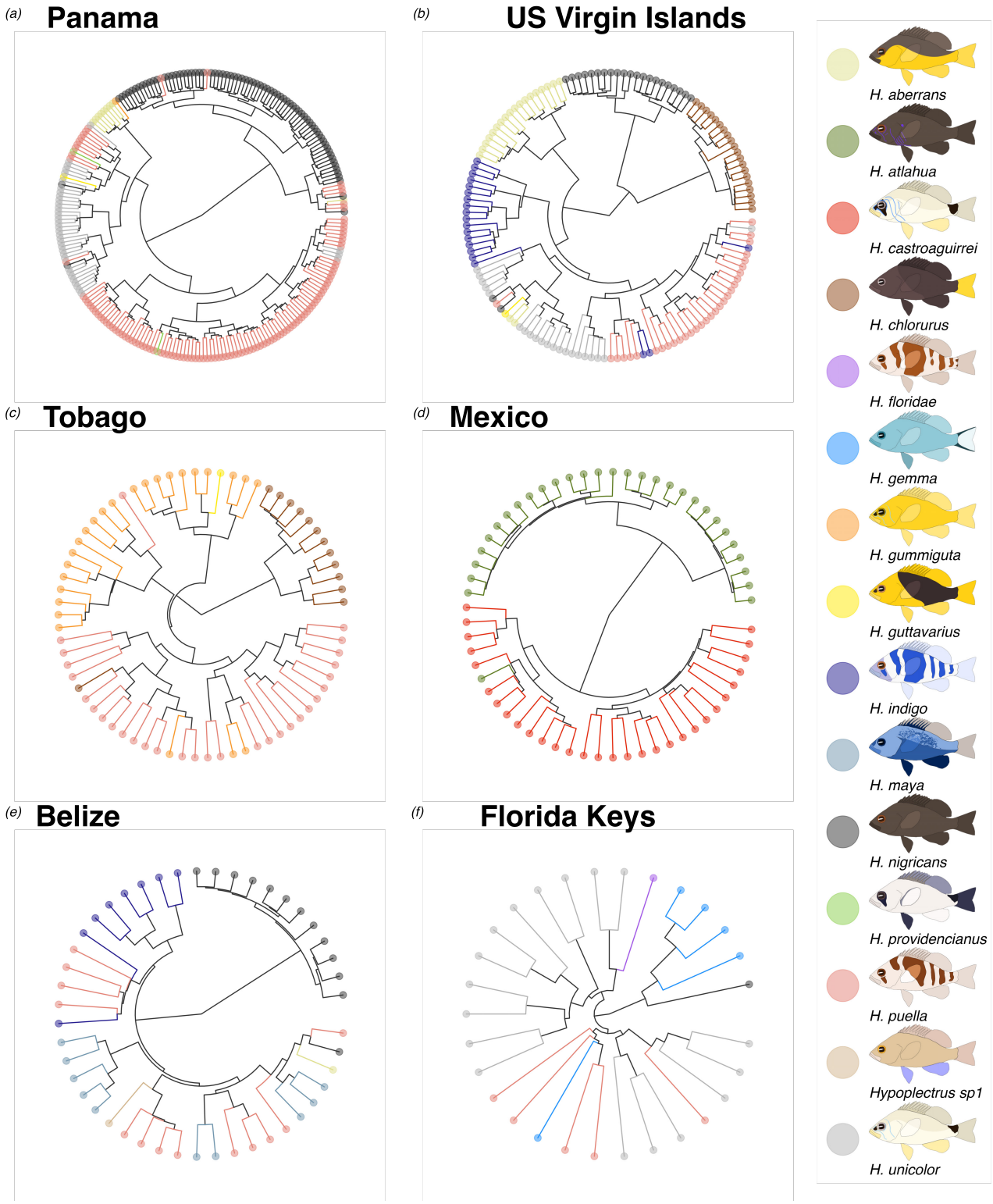



**Figure S3.2** — Hierarchical clustering based on 15 PCs for *(a)* Panama, *(b)* US Virgin Island, *(c)* Tobago, *(d)* Veracruz, *(e)* Belize and *(f)* the Florida Keys.

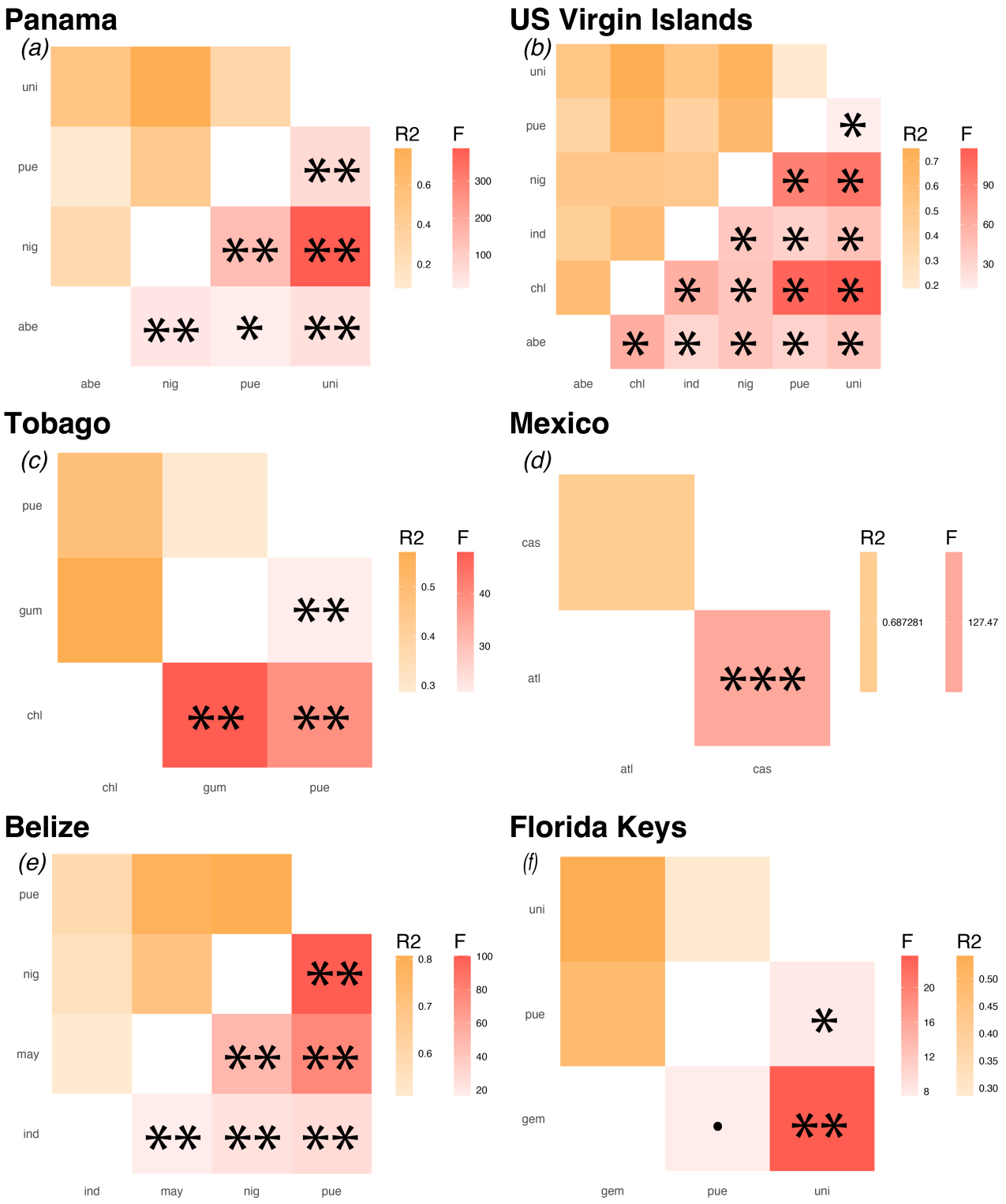



**Figure S3.3** — Pairwise PERMANOVA based on 15 PCs for hamlet species with $\geqq 5$ individuals at each site. Significance indicated as . for $p.adjusted \leqq 0.05$, * for $p.adjusted \leqq 0.01$, ** for $p.adjusted \leqq 0.001$ and *** for $p.adjusted \leqq 0.0001$. The F statistic is reported in red and the residual R2 in orange for *(a)* Panama, *(b)* US Virgin Island, *(c)* Tobago, *(d)* Veracruz, *(e)* Belize and *(f)* the Florida Keys.

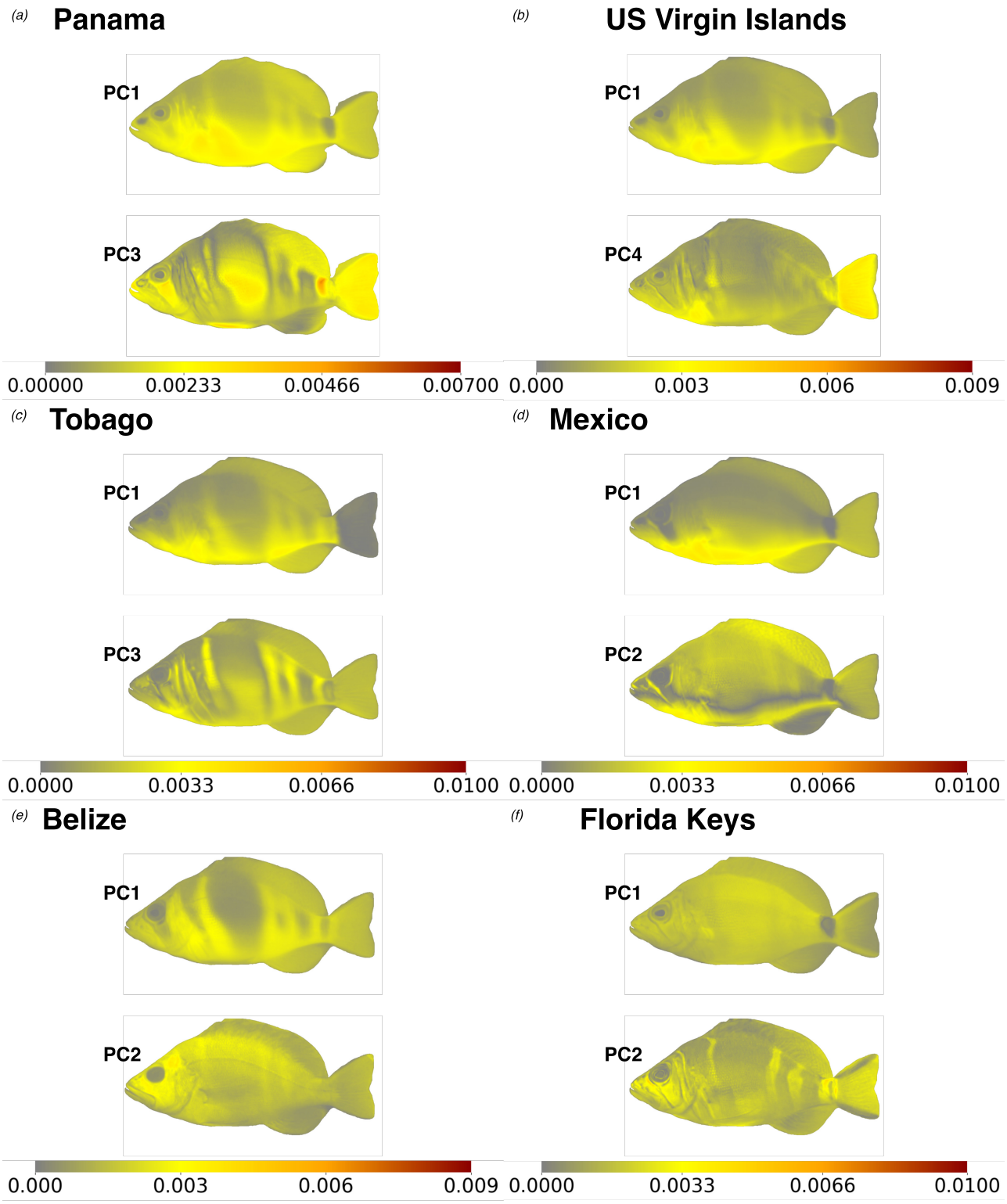



**Figure S3.4** — Heatmaps of pixel contribution to variation along the PC axes from Figure 3.2. *(a)* Panama, *(b)* US Virgin Island, *(c)* Tobago, *(d)* Veracruz, *(e)* Belize and *(f)* the Florida Keys.

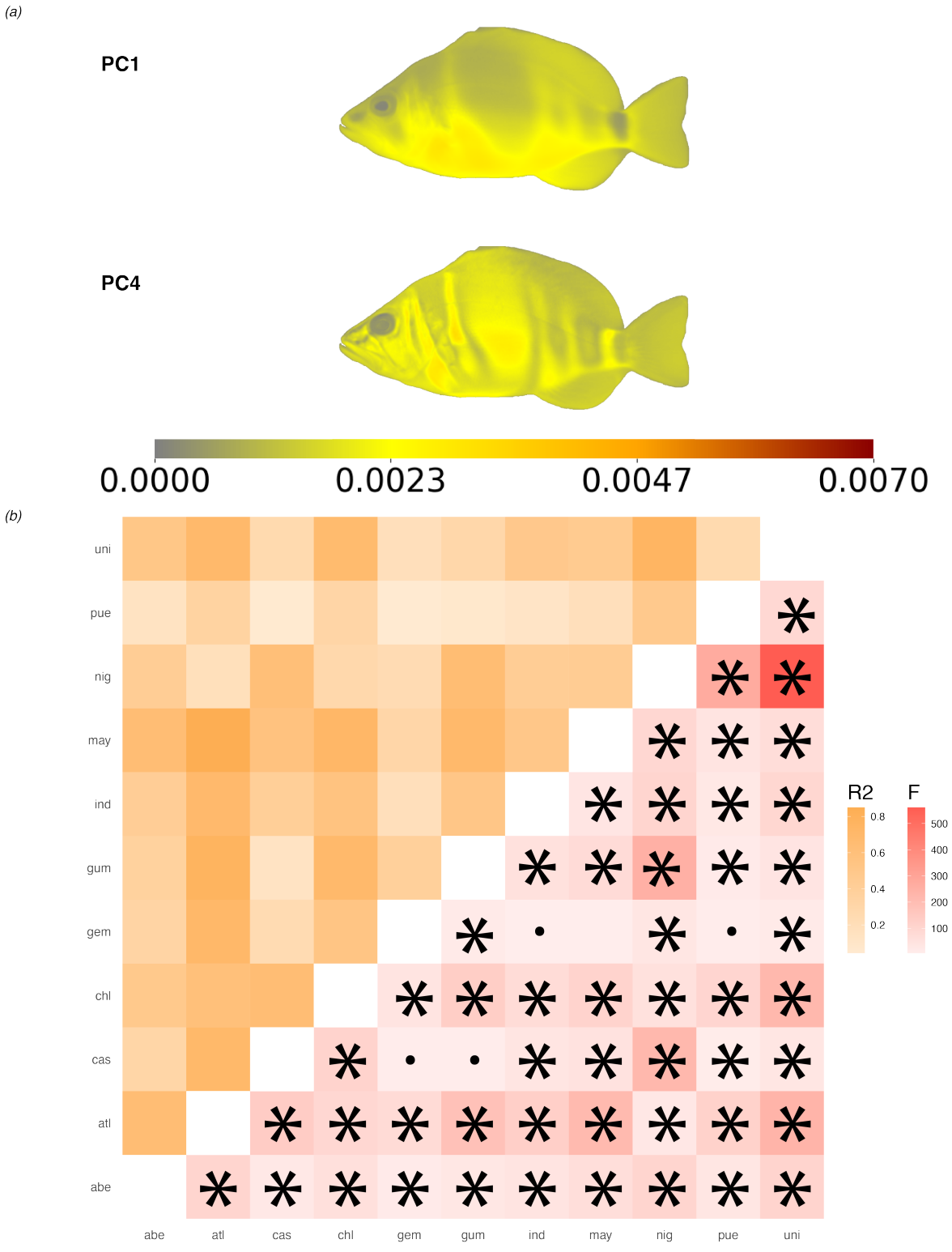



**Figure S3.5** — *(a)* Heatmap of the pixel contribution to variation of PC1 and PC4 from PCA overall hamlet species and sites. *(b)* Pairwise PERMANOVA based on 15 PCs for hamlet species with $\geqq 5$ individuals at the scale of the whole radiation, all sites confounded. Significance is indicated as . for $p.adjusted \leqq 0.05$, * for $p.adjusted \leqq 0.01$, ** for $p.adjusted \leqq 0.001$ and *** for $p.adjusted \leqq 0.0001$. The F statistic is reported in red and the residual R2 in orange.

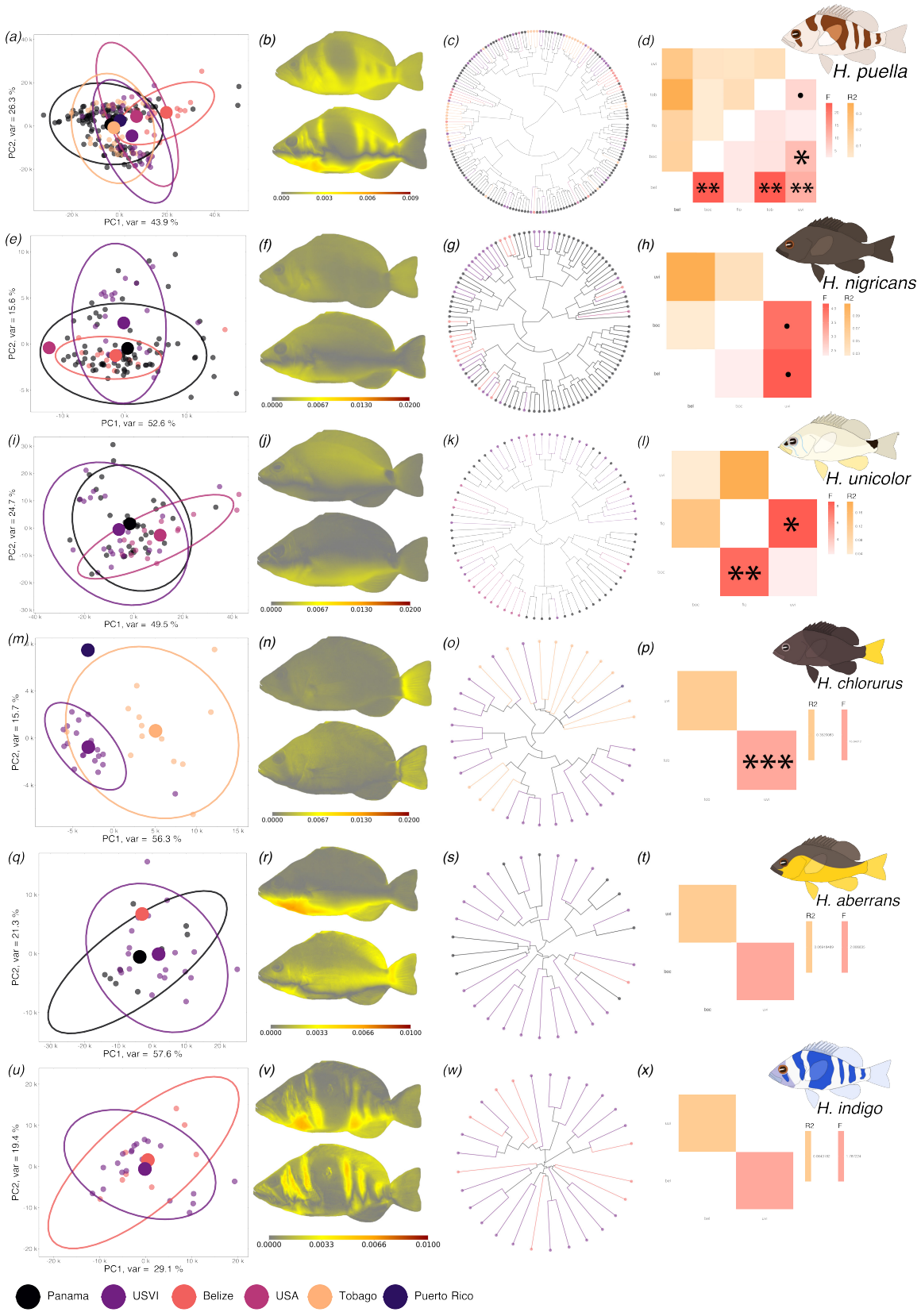



**Figure S3.6** — Spatial phenotypic differences per species for *H. puella*, *H. nigricans*, *H. unicolor*, *H.chlorurus*, *H. aberrans* and *H. indigo*. *(a), (e), (i), (m), (q), (u)* PCA clusters of phenotypes grouped by site, *(b), (f), (j), (n), (r), (v)* heatmaps of pixel contribution to variation along PC1 (top) and PC2 (bottom), *(c), (g), (k), (o), (s), (w)* hierarchical clustering based on 15 PCs and *(d), (h), (l), (p), (t), (x)* pairwise PERMANOVA of PCA clusters between sites with significance indicated as . for $p.adjusted \leqq 0.05$, * for $p.adjusted \leqq 0.01$, ** for $p.adjusted \leqq 0.001$ and *** for $p.adjusted \leqq 0.0001$, the F statistic is reported in red and the residual R2 in orange.

**Table S3.1** — Metadata for the phenotypic dataset (*n=571*). Due to large size, the table was deposited in the chapter3/metadata/ github folder.

**Table S3.2** — Statistics from pairwise PERMANOVA of phenotypes between hamlet species of each individual site. Significance is indicated by . for $p.adjusted \leqq 0.05$, * for $p.adjusted \leqq 0.01$, ** for $p.adjusted \leqq 0.001$ and *** for $p.adjusted \leqq 0.0001$. The overall PERMANOVA for each site is reported in the first row.

| **PANAMA** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 3 | 0.58 | 103.7 | *9.9e-05*** | |
| *H. aberrans vs. H. nigricans* | 0.26 | 28.90 | 0.0001 | 0.0006** |
| *H. aberrans vs. H. puella* | 0.08 | 9.71 | 0.0002 | 0.0012* |
| *H. aberrans vs. H. unicolor* | 0.51 | 44.73 | 0.0001 | 0.0006** |
| *H. nigricans vs. H. puella* | 0.45 | 144.11 | 0.0001 | 0.0006** |
| *H. nigricans vs. H. unicolor* | 0.78 | 388.21 | 0.0001 | 0.0006** |
| *H. puella vs. H. unicolor* | 0.32 | 64.45 | 0.0001 | 0.0006** |

| **USVI** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 5 | 0.68 | 56.06 | *9.9e-05*** | |
| *H. aberrans vs. H. chlorurus* | 0.62 | 61.02 | 9.9e-05 | 0.0015* |
| *H. aberrans vs. H. indigo* | 0.44 | 30.17 | 9.9e-05 | 0.0015* |
| *H. aberrans vs. H. nigricans* | 0.51 | 42.00 | 9.9e-05 | 0.0015* |
| *H. aberrans vs. H. puella* | 0.40 | 32.06 | 9.9e-05 | 0.0015* |
| *H. aberrans vs. H. unicolor* | 0.52 | 43.00 | 9.9e-05 | 0.0015* |
| *H. chlorurus vs. H. indigo* | 0.61 | 60.36 | 9.9e-05 | 0.0015* |
| *H. chlorurus vs. H. nigricans* | 0.51 | 42.53 | 9.9e-05 | 0.0015* |
| *H. chlorurus vs. H. puella* | 0.70 | 112.02 | 9.9e-05 | 0.0015* |
| *H. chlorurus vs. H. unicolor* | 0.75 | 117.62 | 9.9e-05 | 0.0015* |
| *H. indigo vs. H. nigricans* | 0.50 | 41.03 | 9.9e-05 | 0.0015* |
| *H. indigo vs. H. puella* | 0.40 | 33.19 | 9.9e-05 | 0.0015* |
| *H. indigo vs. H. unicolor* | 0.54 | 44.97 | 9.9e-05 | 0.0015* |
| *H. nigricans vs. H. puella* | 0.64 | 93.66 | 9.9e-05 | 0.0015* |
| *H. nigricans vs. H. unicolor* | 0.71 | 100.87 | 9.9e-05 | 0.0015* |
| *H. puella vs. H. unicolor* | 0.19 | 11.58 | 9.9e-05 | 0.0015* |

| **TOBAGO** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|

| | | | | |
|---|---|---|---|---|
| # group = 2 | 0.50 | 32.52 | *9.9e-05**** | |
| *H. chlorurus vs. H. gummigutta* | 0.57 | 47.82 | 9.9e-05 | 0.0003** |
| *H. chlorurus vs. H. puella* | 0.49 | 38.85 | 9.9e-05 | 0.0003** |
| *H. gummigutta vs. H. puella* | 0.29 | 21.31 | 9.9e-05 | 0.0003** |

| **MEXICO** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 1 | 0.69 | 127.47 | *9.9e-05**** | |
| *H. atlahua vs. H. castroaguirrei* | 0.69 | 127.47 | 9.9e-05 | 9.9e-05*** |

| **BELIZE** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 3 | 0.80 | 53.42 | *9.9e-05**** | |
| *H. indigo vs. H. maya* | 0.51 | 16.68 | 9.9e-05 | 0.0006** |
| *H. indigo vs. H. nigricans* | 0.56 | 23.73 | 9.9e-05 | 0.0006** |
| *H. indigo vs. H. puella* | 0.59 | 27.00 | 9.9e-05 | 0.0006** |
| *H. maya vs. H. nigricans* | 0.70 | 49.00 | 9.9e-05 | 0.0006** |
| *H. maya vs. H. puella* | 0.79 | 77.05 | 9.9e-05 | 0.0006** |
| *H. nigricans vs. H. puella* | 0.81 | 100.18 | 9.9e-05 | 0.0006** |

| **FLORIDA KEYS** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 2 | 0.54 | 14.18 | *9.9e-05**** | |
| *H. gemma vs. H. puella* | 0.48 | 7.51 | 0.0088 | 0.0264 . |
| *H. gemma vs. H. unicolor* | 0.54 | 23.67 | 0.0002 | 0.0006** |
| *H. puella vs. H. unicolor* | 0.29 | 8.02 | 0.0008 | 0.0024* |

**Table S3.3** — Statistics from pairwise PERMANOVA of phenotypes between hamlet species at the scale of the whole radiation, including all confounded sites. Significance is indicated by . for $p.adjusted \leqq 0.05$, * for $p.adjusted \leqq 0.01$, ** for $p.adjusted \leqq 0.001$ and *** for $p.adjusted \leqq 0.0001$. The overall PERMANOVA is reported in the first row.

| **ALL HAMLETS & SITES** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 10 | 0.63 | 93.91 | *9.9e-05**** | |
| *H. aberrans vs. H. atlahua* | 0.64 | 103.49 | 0.0001 | 0.0055* |
| *H. aberrans vs. H. castroaguirrei* | 0.31 | 26.15 | 0.0001 | 0.0055* |
| *H. aberrans vs. H. chlorurus* | 0.49 | 59.24 | 0.0001 | 0.0055* |

| | | | | |
|---|---|---|---|---|
| *H. aberrans vs. H. gemma* | 0.34 | 16.64 | 0.0001 | 0.0055* |
| *H. aberrans vs. H. gummigutta* | 0.37 | 31.53 | 0.0001 | 0.0055* |
| *H. aberrans vs. H. indigo* | 0.42 | 41.34 | 0.0001 | 0.0055* |
| *H. aberrans vs. H. maya* | 0.64 | 68.41 | 0.0001 | 0.0055* |
| *H. aberrans vs. H. nigricans* | 0.43 | 103.50 | 0.0001 | 0.0055* |
| *H. aberrans vs. H. puella* | 0.14 | 34.90 | 0.0001 | 0.0055* |
| *H. aberrans vs. H. unicolor* | 0.52 | 109.36 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. castroaguirrei* | 0.71 | 139.36 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. chlorurus* | 0.60 | 94.09 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. gemma* | 0.71 | 80.82 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. gummigutta* | 0.77 | 185.18 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. indigo* | 0.70 | 127.62 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. maya* | 0.85 | 214.17 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. nigricans* | 0.18 | 31.02 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. puella* | 0.35 | 117.58 | 0.0001 | 0.0055* |
| *H. atlahua vs. H. unicolor* | 0.70 | 235.73 | 0.0001 | 0.0055* |
| *H. castroaguirrei vs. H. chlorurus* | 0.64 | 111.23 | 0.0001 | 0.0055* |
| *H. castroaguirrei vs. H. gemma* | 0.24 | 10.47 | 0.0002 | 0.011 . |
| *H. castroaguirrei vs. H. gummigutta* | 0.14 | 8.45 | 0.0007 | 0.039 . |
| *H. castroaguirrei vs. H. indigo* | 0.41 | 38.83 | 0.0001 | 0.0055* |
| *H. castroaguirrei vs. H. maya* | 0.57 | 51.14 | 0.0001 | 0.0055* |
| *H. castroaguirrei vs. H. nigricans* | 0.62 | 222.98 | 0.0001 | 0.0055* |
| *H. castroaguirrei vs. H. puella* | 0.05 | 11.44 | 0.0001 | 0.0055* |
| *H. castroaguirrei vs. H. unicolor* | 0.26 | 36.33 | 0.0001 | 0.0055* |
| *H. chlorurus vs. H. gemma* | 0.54 | 43.92 | 0.0001 | 0.0055* |

| | | | | |
|---|---|---|---|---|
| *H. chlorurus vs. H. gummigutta* | 0.70 | 134.17 | 0.0001 | 0.0055* |
| *H. chlorurus vs. H. indigo* | 0.56 | 77.26 | 0.0001 | 0.0055* |
| *H. chlorurus vs. H. maya* | 0.73 | 111.17 | 0.0001 | 0.0055* |
| *H. chlorurus vs. H. nigricans* | 0.28 | 56.92 | 0.0001 | 0.0055* |
| *H. chlorurus vs. H. puella* | 0.32 | 102.14 | 0.0001 | 0.0055* |
| *H. chlorurus vs. H. unicolor* | 0.68 | 223.37 | 0.0001 | 0.0055* |
| *H. gemma vs. H. gummigutta* | 0.39 | 18.82 | 0.0001 | 0.0055* |
| *H. gemma vs. H. indigo* | 0.19 | 7.16 | 0.0004 | 0.022 . |
| *H. gemma vs. H. maya* | 0.31 | 5.75 | 0.0013 | 0.071 |
| *H. gemma vs. H. nigricans* | 0.24 | 35.68 | 0.0001 | 0.0055* |
| *H. gemma vs. H. puella* | 0.05 | 8.85 | 0.0004 | 0.022 . |
| *H. gemma vs. H. unicolor* | 0.19 | 17.67 | 0.0001 | 0.0055* |
| *H. gummigutta vs. H. indigo* | 0.52 | 56.39 | 0.0001 | 0.0055* |
| *H. gummigutta vs. H. maya* | 0.70 | 78.55 | 0.0001 | 0.0055* |
| *H. gummigutta vs. H. nigricans* | 0.65 | 253.16 | 0.0001 | 0.0055* |
| *H. gummigutta vs. H. puella* | 0.08 | 17.18 | 0.0001 | 0.0055* |
| *H. gummigutta vs. H. unicolor* | 0.29 | 40.72 | 0.0001 | 0.0055* |
| *H. indigo vs. H. maya* | 0.51 | 36.97 | 0.0001 | 0.0055* |
| *H. indigo vs. H. nigricans* | 0.43 | 104.57 | 0.0001 | 0.0055* |
| *H. indigo vs. H. puella* | 0.12 | 28.24 | 0.0001 | 0.0055* |
| *H. indigo vs. H. unicolor* | 0.49 | 96.37 | 0.0001 | 0.0055* |
| *H. maya vs. H. nigricans* | 0.44 | 95.31 | 0.0001 | 0.0055* |
| *H. maya vs. H. puella* | 0.18 | 43.92 | 0.0001 | 0.0055* |
| *H. maya vs. H. unicolor* | 0.46 | 69.80 | 0.0001 | 0.0055* |
| *H. nigricans vs. H. puella* | 0.48 | 271.85 | 0.0001 | 0.0055* |
| *H. nigricans vs. H. unicolor* | 0.75 | 560.60 | 0.0001 | 0.0055* |
| *H. puella vs. H. unicolor* | 0.26 | 90.82 | 0.0001 | 0.0055* |

**Table S3.4** — Statistics from pairwise PERMANOVA of phenotypes between different sites for each individual hamlet species. Significance is indicated by . for $p.adjusted \leqq 0.05$, * for $p.adjusted \leqq 0.01$, ** for $p.adjusted \leqq 0.001$ and *** for $p.adjusted \leqq 0.0001$. The overall PERMANOVA for each species is reported in the first row.

| ***H. puella*** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 5 | 0.18 | 9.80 | *9.9e-05**** | |
| Belize *vs.* Panama | 0.17 | 23.78 | 0.0001 | 0.001** |
| Belize *vs.* Florida Keys | 0.16 | 3.26 | 0.0096 | 0.096 |
| Belize *vs.* Tobago | 0.37 | 23.75 | 0.0001 | 0.001** |
| Belize *vs.* US Virgin Islands | 0.22 | 12.17 | 0.0001 | 0.001** |
| Panama *vs.* Florida Keys | 0.03 | 3.26 | 0.0309 | 0.309 |
| Panama *vs.* Tobago | 0.03 | 3.95 | 0.0120 | 0.120 |
| Panama *vs.* US Virgin Islands | 0.07 | 9.92 | 0.0002 | 0.002* |
| Florida Keys *vs.* Tobago | 0.10 | 3.76 | 0.0148 | 0.148 |
| Florida Keys *vs.* US Virgin Islands | 0.06 | 2.40 | 0.0779 | 0.779 |
| Tobago *vs.* US Virgin Islands | 0.09 | 6.00 | 0.0026 | 0.026 . |

| ***H. nigricans*** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 2 | 0.06 | 3.27 | *0.009** | |
| Belize *vs.* Panama | 0.03 | 2.31 | 0.0804 | 0.241 |
| Belize *vs.* US Virgin Islands | 0.11 | 4.18 | 0.0070 | 0.021 . |
| Panama *vs.* US Virgin Islands | 0.04 | 3.95 | 0.0163 | 0.049 . |

| ***H. unicolor*** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 2 | 0.14 | 5.86 | *9.9e-05**** | |
| Panama *vs.* Florida Keys | 0.14 | 7.98 | 0.0002 | 0.0006** |
| Panama *vs.* US Virgin Islands | 0.04 | 2.19 | 0.0870 | 0.2607 |
| Florida Keys *vs.* US Virgin Islands | 0.19 | 8.54 | 0.0008 | 0.0024* |

| ***H. chlorurus*** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 1 | 0.35 | 16.84 | *9.9e-05**** | |
| Tobago *vs.* US Virgin Islands | 0.35 | 16.84 | 9.9e-05 | 9.9e-05*** |

| ***H. aberrans*** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 1 | 0.07 | 2.00 | *0.1208* | |
| Panama *vs.* US Virgin Islands | 0.07 | 2.01 | 0.1145 | 0.1145 |

| ***H. indigo*** | *R2* | *F* | p-value | p-adjusted |
|---|---|---|---|---|
| # group = 1 | 0.06 | 1.79 | *0.08659* | |
| Belize *vs.* US Virgin Islands | 0.06 | 1.79 | 0.0911 | 0.0911 |

## References


Aguilar-Perera, A. (2003). Abundance and distribution of hamlets (Teleostei: Hypoplectrus) in coral reefs off southwestern Puerto Rico: Support for the multiple-species theory. Caribbean Journal of Science *39*, 147–150.

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

Barrier, M., Baldwin, B. G., Robichaux, R. H. and Purugganan, M. D. (1999). Interspecific hybrid ancestry of a plant adaptive radiation: allopolyploidy of the Hawaiian silversword alliance (Asteraceae) inferred from floral homeotic gene duplications. Molecular biology and evolution *16*, 1105–1113.

Bolnick, D. I., Hund, A. K., Nosil, P., Peng, F., Ravinet, M., Stankowski, S., Subramanian, S., Wolf, J. B. and Yukilevich, R. (2023). A multivariate view of the speciation continuum. Evolution *77*, 318–328.

Coulmance, F., Akkaynak, D., Le Poul, Y., Höppner, M. P., McMillan, W. O. and Puebla, O. (2024). Phenotypic and genomic dissection of colour pattern variation in a reef fish radiation. Molecular Ecology *33*, e17047.

Cracraft, J. (1983). Species concepts and speciation analysis. Current ornithology *1*, 159–187.

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. et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

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

Fischer, E. A. (1980). Speciation in the hamlets (Hypoplectrus: Serranidae): a continuing enigma. Copeia *(1980)4*, 649–659.

Fox, J. and Weisberg, S. (2018). An R companion to applied regression. Sage publications.

Friendly, M., Monette, G. and Fox, J. (2013). Elliptical insights: understanding statistical methods through elliptical geometry.

Gören, C. M. and Kaymak, N. (2024). Inter-and intraspecific phenotypic variation in co-occurring invasive and introduced cyprinid fish species. Saudi Journal of Biological Sciences *31*, 103943.

Grant, P. R. and Grant, B. R. (2002). Adaptive radiation of Darwin's finches: Recent data help explain how this famous group of Galapagos birds evolved, although gaps in our understanding remain. American scientist *90*, 130–139.

Grant, P. R. and Grant, B. R. (2007). How and why species multiply: the radiation of Darwin's finches. Princeton University Press.

Hahn, M. W. and Nakhleh, L. (2016). Irrational exuberance for resolved species trees. Evolution *70*, 7–17.

Hench, K., Helmkampf, M., McMillan, W. O. and Puebla, O. (2022). Rapid radiation in a highly diverse marine environment. Proceedings of the National Academy of Sciences *119*, e2020457119.

Hench, K., Mcmillan, W. O., Betancur-R, R. and Puebla, O. (2017). Temporal changes in hamlet communities (Hypoplectrus spp., Serranidae) over 17 years. Journal of fish biology *91*, 1475–1490.

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

Hendry, A. P., Bolnick, D. I., Berner, D. and Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. Journal of fish biology *75*, 2000–2036.

Holt, B. G., Côté, I. M. and 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. and 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.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., Ho, S. Y., Faircloth, B. C., Nabholz, B., Howard, J. T. et al. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. Science *346*, 1320–1331.

Jordan, D. S. and Evermann, B. W. (1896). The fishes of North and Middle America: a descriptive catalogue of the species of fish-like vertebrates found in the waters of North America, north of the Isthmus of Panama. Number 47, US Government Printing Office.

Kronforst, M. R. and Papa, R. (2015). The functional basis of wing patterning in Heliconius butterflies: the molecules behind mimicry. Genetics *200*, 1–19.

Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V. and Joron, M. (2014). Evolution of dominance mechanisms at a butterfly mimicry supergene. Nature communications *5*, 5644.

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

Losos, J. B. and Glor, R. E. (2003). Phylogenetic comparative methods and the geography of speciation. Trends in Ecology & Evolution *18*, 220–227.

Mallet, J. (1995). A species definition for the modern synthesis. Trends in Ecology & Evolution *10*, 294–299.

Mallet, J. (2001). Species, concepts of. Encyclopedia of biodiversity *5*, 427–440.

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

Martinez Arbizu, P. (2020). pairwiseAdonis: Pairwise multilevel comparison using adonis. R package version 0.4 *1*.

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

McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., Kishe, M. A., O'Meara, B., Bruggmann, R., Excoffier, L. et al. (2020). The ecological and genomic basis of explosive adaptive radiation. Nature *586*, 75–79.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research *20*, 1297–1303.

McKenzie, D. J., Zhang, Y., Eliason, E. J., Schulte, P. M., Claireaux, G., Blasco, F. R., Nati, J. J. and Farrell, A. P. (2021). Intraspecific variation in tolerance of warming in fishes. Journal of Fish Biology *98*, 1536–1555.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L. and Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nature communications *8*, 14363.

Muschick, M., Indermaur, A. and Salzburger, W. (2012). Convergent evolution within an adaptive radiation of cichlid fishes. Current biology *22*, 2362–2368.

Navalón, G., Marugán-Lobón, J., Bright, J. A., Cooney, C. R. and Rayfield, E. J. (2020). The consequences of craniofacial integration for the adaptive radiations of Darwin's finches and Hawaiian honeycreepers. Nature Ecology & Evolution *4*, 270–278.

Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Borman, T., Carvalho, G., Chirico, M., De Caceres, M., Durand, S., Evangelista, H. B. A., FitzJohn, R., Friendly, M., Furneaux, B., Hannigan, G., Hill, M. O., Lahti, L., McGlinn, D., Ouellette, M.-H., Ribeiro Cunha, E., Smith, T., Stier, A., Ter Braak, C. J. and Weedon, J. (2024). vegan: Community Ecology Package. R package version 2.7-0, https://github.com/vegandevs/vegan.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M. and Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research *12*, 2825–2830.

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

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

Puebla, O., Coulmance, F., Estape, C. J., Estape, A. M. and 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. and 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.

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

Reid, K., Bell, M. A. and Veeramah, K. R. (2021). Threespine stickleback: A model system for evolutionary genomics. Annual review of genomics and human genetics *22*, 357–383.

Salzburger, W. (2008). To be or not to be a hamlet pair in sympatry.

Santos, M. E., Lopes, J. F. and Kratochwil, C. F. (2023). East African cichlid fishes. EvoDevo *14*, 1.

Schluter, D. (2000). The ecology of adaptive radiation. OUP Oxford.

Selz, O. M., Lucek, K., Young, K. A. and Seehausen, O. (2014). Relaxed trait covariance in interspecific cichlid hybrids predicts morphological diversity in adaptive radiations. Journal of evolutionary biology *27*, 11–24.

Sokal, R. R. and Crovello, T. J. (1970). The biological species concept: a critical evaluation. The American Naturalist *104*, 127–153.

Stankowski, S. and Ravinet, M. (2021). Defining the speciation continuum. Evolution *75*, 1256–1273.

Streelman, J. T., Albertson, R. C. and Kocher, T. D. (2003). Genome mapping of the orange blotch colour pattern in cichlid fishes. Molecular Ecology *12*, 2465–2471.

Stroud, J. T. and Losos, J. B. (2020). Bridging the process-pattern divide to understand the origins

and early stages of adaptive radiation: a review of approaches with insights from studies of Anolis lizards. Journal of Heredity *111*, 33–42.

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

Urban, S., Gerwin, J., Hulsey, C. D., Meyer, A. and Kratochwil, C. F. (2022). The repeated evolution of stripe patterns is correlated with body morphology in the adaptive radiations of East African cichlid fishes. Ecology and Evolution *12*, e8568.

Urban, S., Nater, A., Meyer, A. and Kratochwil, C. F. (2021). Different sources of allelic variation drove repeated color pattern divergence in cichlid fishes. Molecular Biology and Evolution *38*, 465–477.

Wagner, C. E., Harmon, L. J. and Seehausen, O. (2012). Ecological opportunity and sexual selection together predict adaptive radiation. Nature *487*, 366–369.

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

Wu, C.-I. (2001). The genic view of the process of speciation. Journal of evolutionary biology *14*, 851–865.

Wu, C.-I. and Ting, C.-T. (2004). Genes and speciation. Nature Reviews Genetics *5*, 114–122.

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

# 04

# RADIATION WITH REPRODUCTIVE ISOLATION IN THE NEAR-ABSENCE OF PHYLOGENETIC SIGNAL

**Floriane Coulmance,** Martin Helmkampf, Melanie J. Heckwolf, Arturo Acero P., Alice Balard, Iliana Bista, Omar Domínguez-Domínguez, Paul B. Frandsen, Montserrat Torres-Oliva, Aintzane Santaquiteria, Jose Tavera, Benjamin C. Victor, D. Ross Robertson, Ricardo Betancur-R., W. Owen McMillan, Oscar Puebla

This manuscript is currently being reviewed in *Science Advances*, * shared first authorship.

**Coulmance, F.***, Helmkampf, M.*, Heckwolf, M. J., Acero, P. A., Balard, A., Bista, I., Domínguez-Domínguez, O., Frandsen, P. B., Torres-Oliva, M., Santaquiteria, A., Tavera, J., Victor, B. C., Robertson, D. R., Betancur-R., R., McMillan, W. O. and Puebla, O. *(in review).* Radiation with reproductive isolation in the near-absence of phylogenetic signal. Science Advances.

## Abstract

According to the genic view, species are characterized by the genes that underlie functional divergence. Here, we take a phylogenomic approach to assess this view at the scale of a whole radiation. The hamlets (*Hypoplectrus* spp.) represent a recent radiation of reef fishes from the Greater Caribbean that are reproductively isolated through assortative mating. A total of 335 genomes from 15 locations revealed a single well-supported phylogenetic split among species, with a large share of the radiation unresolved. The polytomic nature of the hamlet radiation is extreme compared to other recent radiations such as Lake Victoria cichlids. At the gene-tree level we identified just one genomic region, centered around the *casz1* transcription factor, with a topology that reflects species differences. These results show that phenotypic diversification and reproductive isolation—two major attributes of species—may unfold in the near-absence of phylogenetic signal, both genome-wide and at the gene-tree level.

## Keywords

phylogenomic, genic view of species, reproductive isolation, rapid radiation, reef fishes

## Author's Contribution

**FC**, MJH, AAP, ODD, JT, BCV, DRR, RBR, WOM and OP coordinated and performed sample collection. **FC**, MJH, RBR, PBF, WOM and OP conducted laboratory work. PBF, MTO, RBR, WOM and OP provided funds for sequencing. MH, **FC**, MJH, AB, IB, PBF, AS, RBR and OP analyzed the data and visualized the results. MH, **FC**, MJH and OP wrote the manuscript with input and feedback from all co-authors.

## 4.1 Introduction

Species are the foundational units of biodiversity, but how they form and what exactly characterizes them remain open questions. Furthermore, few theoretical frameworks and empirical systems provide the opportunity to address these two questions jointly. In this regard the genic view of species and speciation (hereafter "genic view") is of particular interest. This framework was articulated by Wu with a specific focus on animals (Wu, 2001; Wu and Ting, 2004). However, the idea predates Wu (Mallet, 1995; Feder, 1998), also applies to plants (Rieseberg and Burke, 2001), and is not necessarily incompatible with alternative views of species and speciation (Rieseberg and Burke, 2001; Mayr, 2001; De Queiroz, 2007). A fundamental tenet of the genic view is that species are essentially formed and characterized by the genes that underlie adaptation to ecological and sexual environments, with reproductive isolation emerging as a by-product of this functional divergence. From a genomic perspective, it is in the presence of gene flow that the genic view manifests itself most clearly because gene flow tends to reveal the genomic regions underlying functional divergence by homogenizing genetic variation throughout the rest of the genome. Nevertheless, an understanding of the nature and origin of species requires more than just a genomic perspective. It also entails knowledge of the traits that underlie adaptation, how they relate to ecological and sexual environments, and how this functional divergence results in reproductive isolation.

The hamlets (*Hypoplectrus* spp., Serranidae) provide a rare opportunity to empirically assess the genic view. These reef fishes from the Greater Caribbean differ mainly in terms of color pattern (Puebla et al., 2022), which is ecologically relevant for camouflage and mimicry (Randall and Randall, 1960; Thresher, 1978; Fischer, 1980; Puebla et al., 2007, 2018), but are otherwise morphologically and ecologically very similar (Whiteman et al., 2007; Holt et al., 2008). They also differ in terms of distribution and abundance (Domeier, 1994; Aguilar-Perera, 2003; Holt et al., 2010; Aguilar-Perera and González-Salas, 2010; Hench et al., 2017) but are highly sympatric, with up to nine species encountered on a single dive survey (Puebla et al., 2022). Hamlet courtship and spawning behavior can be observed on a daily basis throughout the year, which provides a direct window on prezygotic reproductive isolation. Different species are commonly observed spawning at the same time and in the same area, but spawning is strongly assortative with respect to color pattern (Fischer, 1980; Domeier, 1994; Puebla et al., 2007; Barreto and McCartney, 2008; Puebla et al., 2012). Assortative mate choice is maintained in experimental conditions and also when fishes are kept in different compartments without water exchange (Domeier, 1994), indicating that it relies heavily on visual cues. Interspecific spawnings are nonetheless observed at a low frequency ($< 2$ %) in natural populations (Fischer, 1980; Domeier, 1994; Puebla et al., 2007; Barreto and McCartney, 2008; Puebla et al., 2012). There are no barriers to fertilization among species (Whiteman and Gage, 2007) and the available evidence indicates that hybrids develop normally. In the few cases where hybrids were raised to the adult stage, they developed color patterns that were

intermediate between the two parental species (Domeier, 1994). Furthermore, high-probability hybrids and backcrosses have been identified in natural populations through genetic analysis (Hench et al., 2019, 2022; Coulmance et al., 2024), indicating that hybridization and introgression are ongoing.

In agreement with the occurrence of gene flow, hamlet species are genetically highly similar. Levels of genetic differentiation among species fall within the range that is typically encountered among populations within species (Mccartney et al., 2003; Puebla et al., 2007, 2008; Holt et al., 2011; Puebla et al., 2012, 2014). The genomic architecture of species differences is characterized by sharp peaks of differentiation that stand out against a background of low differentiation (Hench et al., 2019, 2022), exactly as envisioned by the genic view during the early stages of speciation (Wu, 2001; Wu and Ting, 2004). As a group, the hamlets form an exceptionally shallow radiation of 18 described species (Figure 4.1A) that appear to have diverged very recently [<10,000 generations (Hench et al., 2022)], and is largely unresolved from a phylogenetic perspective (Ramon et al., 2003; Mccartney et al., 2003; Garcia-Machado et al., 2004; Victor, 2012; Tavera and Acero, 2013; Hench et al., 2022). This raises the question whether phenotypic diversification and reproductive isolation, two major attributes of species, may unfold in the absence of phylogenetic signal. Following the genic view, a small part of the genome—the genomic regions underlying functional divergence—are expected to show an evolutionary history that relates to species differences. Nevertheless, previous phylogenetic studies of the hamlet radiation were based on a small number of markers and/or a limited subset of the species collected at few locations (Ramon et al., 2003; Mccartney et al., 2003; Garcia-Machado et al., 2004; Victor, 2012; Tavera and Acero, 2013; Hench et al., 2022).

Here, we use genome-wide data to i. provide a first comprehensive phylogenomic perspective of the hamlet radiation, ii. explore whether diversification may develop in a genomic context that is overwhelmingly dominated by incomplete lineage sorting and introgression, and iii. provide an illustration of what species can be minimally (i.e. in a backdrop of minimal genetic divergence) and how they may arise and persist in this context. Following the genic view, we hypothesized that specific genomic regions may exhibit a phylogenetic signal that relates to species differences, and that these genomic regions may point to the genomic basis of functional divergence among species.

## 4.2 Results

### The hamlet radiation is characterized by a single deep phylogenetic split

To test our hypothesis, we compiled a dataset of 335 genomes including all recognized hamlet species from 15 locations across their geographical range (Figures 4.1B, S4.1 and Table S4.1), and generated genome-wide phylogenies using three *Serranus* species as outgroup. Multiple phylogenetic reconstruction methods based on different datasets identified just one consistent and well-supported phylogenetic split among species in the whole radiation. This split separated three species from the Gulf of Mexico (*H. floridae*, *H. ecosur* and *H. atlahua*, hereafter *small clade*) from

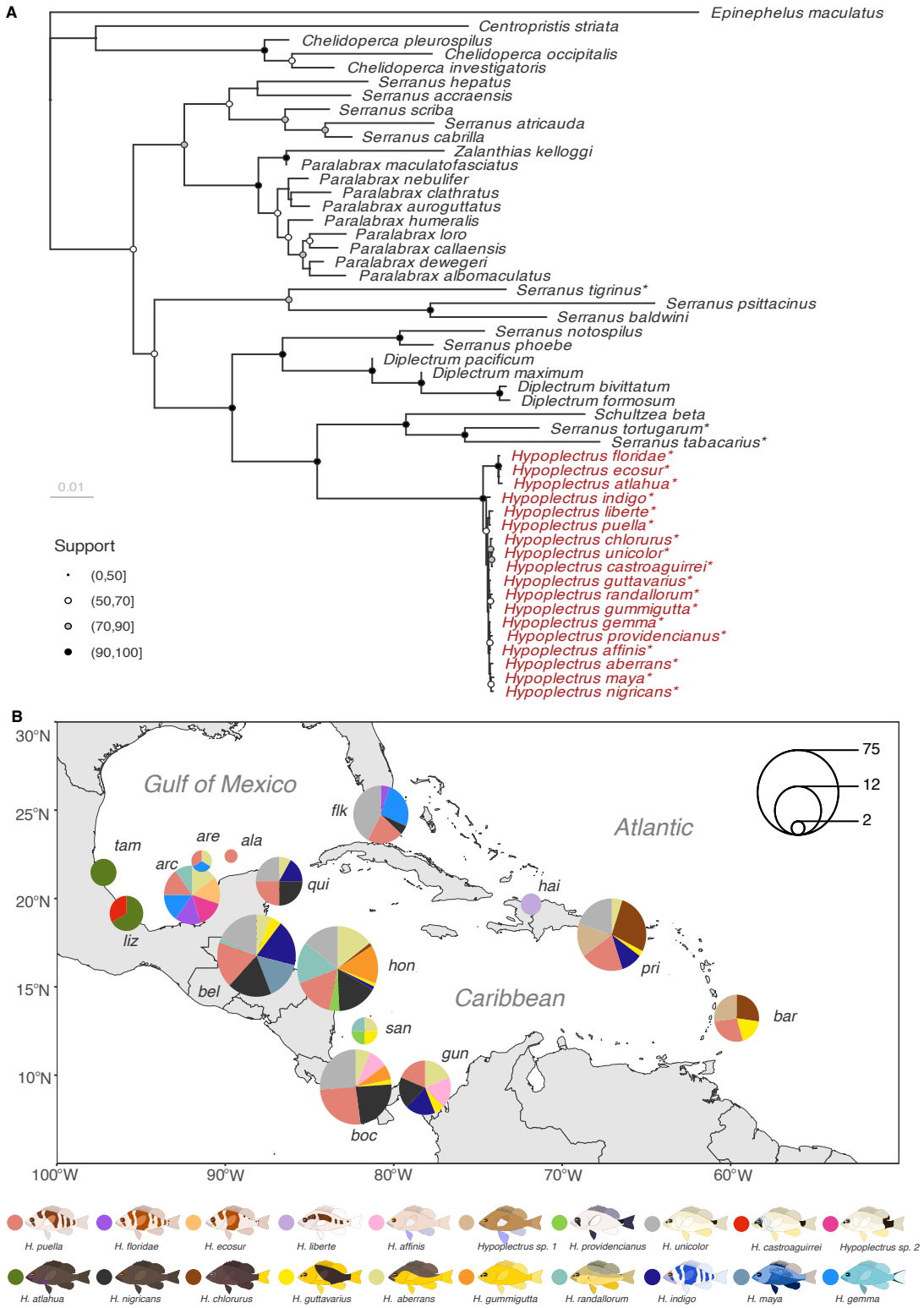



**Figure 4.1** — Phylogenetic context of the hamlet radiation and sampling design. (**A**) Maximum-likelihood phylogeny of the Serraninae subfamily based on 23 nuclear and mitochondrial genes. The hamlet radiation (genus *Hypoplectrus*) is highlighted in red and the species considered in this study are marked with an asterisk. Gene sequences for the other species were obtained from the Fish Tree of Life (Rabosky et al., 2018). Node point size and fill represent node support derived from 200 nonparametric bootstrap replicates. (**B**) Sampling design. The dataset consists of 327 hamlet (*Hypoplectrus*) and eight *Serranus* outgroup genomes that include all described hamlet species from 15 locations covering the Gulf of Mexico [Tamiahua (tam), Antón Lizardo (liz), Cayos Arcas (arc), Cayo Arenas (are), Alacranes Reef (ala) and the Florida Keys (flk)], the Western Caribbean [Quintana Roo (qui), Belize (bel), Honduras (hon), San Andrés (san), Bocas del Toro (boc) and Guna Yala (gun)], and the Eastern Caribbean [Haiti (hai), Puerto Rico (pri) and Barbados (bar)].

the rest of the radiation (hereafter *large clade*; Figures 4.2A, S4.2, S4.3; also apparent in Figure 4.1A). Using the multiple sequentially Markovian coalescent [MSMC2, (Schiffels and Durbin, 2014)], we estimated that these two lineages started to diverge 60,000 generations ago (Figure S4.4). The three species constituting the *small clade* are largely restricted to the Gulf of Mexico, but the phylogenetic split does not reflect complete geographic isolation. Indeed, several species from the *large clade* are also present in the Gulf of Mexico, often in sympatry with members of the *small clade* (Figures 4.2A, 4.2C).

The phylogenetic split between the two clades was also recovered in the mitochondrial genome (Figure S4.5). However, we identified two individuals with a nuclear genome from the large clade and a mitochondrial genome from the small clade, and two individuals with the opposite pattern. These mitonuclear discordances suggest that mitonuclear incompatibilities do not play a major role in the isolation of the two clades and more generally in reproductive isolation. They also point to introgression between the two clades. Analyses of SNP-derived ancestry proportions (Alexander et al., 2009) and chromosome-scale ancestral recombination graphs (Kelleher et al., 2019) indicate that the two clades are admixed (Figure S4.6). Admixture is largely restricted to the Gulf of Mexico, where the two clades are found in sympatry. Furthermore, the species from the *small clade* that exhibits the least admixture with the *large clade* is *H. atlahua*, which is restricted to the western-most part of the Gulf of Mexico. These two geographic patterns suggest that gene flow and introgression, not just incomplete lineage sorting, contribute to admixture. The *D* statistic and related estimates (Patterson et al., 2012) confirm that introgression occurred between the two clades (Figure S4.7). Furthermore, demographic inference with linked selection using Approximate Bayesian Computation (Fraïsse et al., 2021) indicates that the genomic data are more consistent with a scenario of continuous gene flow between the two clades—i.e. non-allopatric divergence—than scenarios without gene flow or with secondary contact (Table S4.4).

Thus, the hamlet radiation is characterized by a single deep phylogenetic split that distinguishes three species from the Gulf of Mexico from the rest of the radiation, but these two clades are not completely geographically and genetically isolated. Another striking aspect of this split is that the species forming the small clade, in particular *H. floridae* and *H. ecosur*, present the highest levels of nucleotide diversity ($\pi$) in the whole radiation (Figure 4.2B). This suggests that the hamlet radiation originated in the Gulf of Mexico from a hamlet population similar to *H. floridae* and *H. ecosur*. Furthermore, maximum likelihood biogeographic analysis indicates that despite the lack of phylogenetic signal for species boundaries, the dataset is well-structured biogeographically (Gulf of Mexico, Western Caribbean and Eastern Caribbean), and that the large clade originated in the Gulf of Mexico (Figure S4.8 and Table S4.5). The emerging picture is therefore an origin of the hamlet radiation in the Gulf of Mexico, where two major lineages diverged with continuous gene flow, followed by expansion and further diversification of the large clade in the Caribbean.

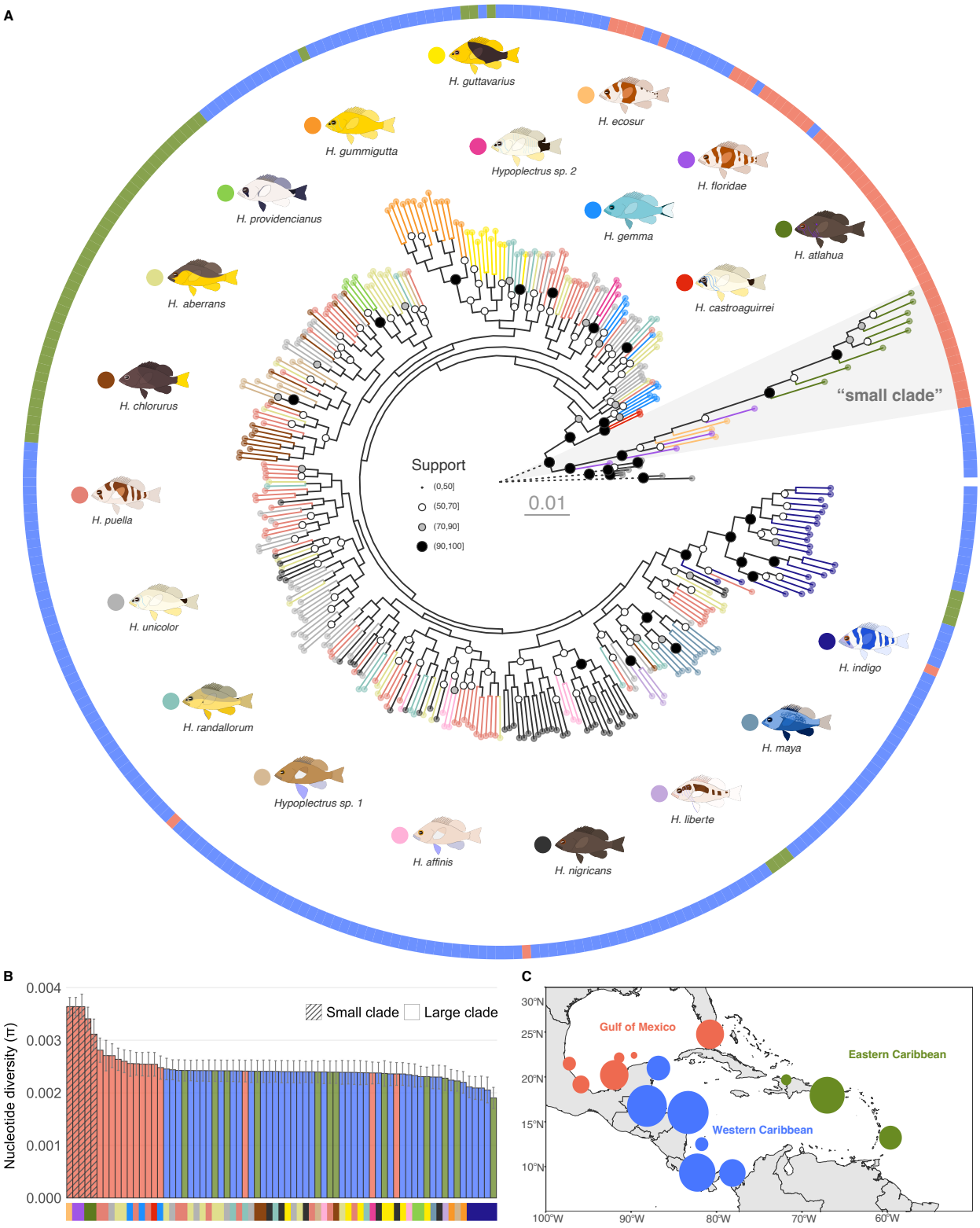



**Figure 4.2** — Phylogeny and nucleotide diversity of the hamlet radiation. (**A**) IQ-TREE phylogeny based on approximately 110,000 Single Nucleotide Polymorphism (SNP) markers distributed across the genome. Branch and tip point colors indicate the hamlet species as shown by the icons surrounding the tree, and outer ring colors correspond to the three biogeographic regions depicted in panel **C**. Node point size and fill represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation. (**B**) Nucleotide diversity ($\pi$) of the 71 populations (species / location combinations) considered in this study. Bar colors correspond to the three biogeographic regions depicted in panel **C**, and x-axis labels indicate species as in panel **A**. Small clade species are additionally highlighted with a hatched pattern. Bar height and errors bars represent mean and standard deviation, respectively, across the 24 linkage groups, for each of which ($\pi$) was calculated separately. (**C**) Map showing the grouping of the 15 locations into three biogeographic regions: Gulf of Mexico (orange), Western Caribbean (blue) and Eastern Caribbean (green).

### A large share of the radiation is unresolved at the genome-wide level

In contrast to the major phylogenetic split described above, none of the other deep nodes of the radiation were highly supported or consistently recovered by the different phylogenetic reconstruction methods that we implemented. Even at the shallow levels of the radiation, just six or seven of the 18 described species were recovered as strictly monophyletic lineages with high support by the different methods applied to the nuclear genome (and this number may have been inflated by small samples size for some of these species), and none according to the mitochondrial genome. The species recovered as strictly monophyletic with high support include members of both clades, with a high proportion of microendemic species (*H. atlahua*, *H. castroaguirrei*, *H. maya* and *H. liberte*). These species are endemic to specific areas (Moran et al., 2019; Tavera and Acero, 2013; Flores et al., 2011; Lobel, 2011), a single bay in the extreme case of *H. liberte* (Victor, 2012). This suggests that increased genetic drift and lineage sorting due to low population size contributed to their divergence. The same may be said about geographic isolation, but to a limited extent since similar to the *small clade* in the Gulf of Mexico, the microendemic species are often found in sympatry with other hamlets.

On the other hand, *H. indigo* and *H. gummigutta*, two species that are extensively distributed across the Greater Caribbean, were often recovered as monophyletic and were the most diverged species of the *large clade* (Figure 4.2A, S4.2, S4.3). These two species are characterized by bright blue and yellow/orange color, respectively, which are thought to be particularly effective for camouflage and communication in reef fishes (Marshall, 2000). It is tempting to speculate that stronger natural and sexual selection contributed to the increased divergence of these two species, but they do not show stronger prezygotic isolation than the other hamlets (Puebla et al., 2007, 2012). Interestingly, five populations were also recovered as monophyletic lineages. These include a tan-colored hamlet from Barbados whose taxonomic status is unclear [*Hypoplectrus sp. 1* (Puebla et al., 2022)], *H. nigricans* from Bocas del Toro in Panama, *H. randallorum* from Cayo Arcas in the Gulf of Mexico, another hamlet from Cayo Arcas whose taxonomic status is unclear (*Hypoplectrus sp. 2*), and *H. gemma* from the Florida Keys. Some of these populations are phenotypically distinct and may represent undescribed microendemic species.

All in all, at best just seven of the 18 currently recognized species were recovered as monophyletic lineages, and even then their phylogenetic relationships are largely unresolved since the deeper nodes of the radiation have low support values. Phylogenetic resolution did not noticeably improve as we considered more SNPs or genomic windows, and the general outcome was robust to all the SNP and genomic window selection strategies we tried for phylogenetic reconstruction. This leaves at least 11 species that are phylogenetically unresolved at the whole-genome level, and notably include *H. puella*, *H. nigricans* and *H. unicolor*, the three most abundant and widely distributed species in the whole radiation. In fact, as illustrated by the SVDQuartets tree (Figure S4.3), the hamlet phylogeny is in large part a polytomy. Ancient polytomies are well doc-

umented (Suh, 2016; Braun et al., 2019; Scherz et al., 2022), but the hamlet polytomy extends largely until present. Following the genic view of species and speciation, we hypothesized that specific genomic regions may be characterized by a phylogenetic signal that relates to species differences, and that these genomic regions may point to the genomic basis of functional divergence.

## A single genomic region relates to species differences

In order to explore the phylogeny of specific genomic regions, we first examined the 2,000 regions considered for the window-based phylogenetic inference (Figure S4.2). We hypothesized that windows with higher mean support might resolve species better, but this was not the case. All the species were mixed and none were resolved in the local phylogeny with highest mean support (Figure S4.9), illustrating the complexity of local phylogenies. This complexity, coupled with the large number of possible species/location topologies and the fact that a large proportion of the local phylogenies that we examined did not align with species/location boundaries, precluded the use of a topology-based approach to analyze the local phylogenies.

We then posited that if a genomic region resolves species or groups of species better than the rest of the genome, it is expected to show an increased association with species identity. Based on this principle, we conducted a Genome-Wide Association Study (GWAS) on species identity to uncover such regions. In agreement with previous association studies on a subset of the radiation (Hench et al., 2022; Coulmance et al., 2024), the results revealed sharp association peaks along the genome (Figure 4.3A). The number of association peaks recovered by the GWAS is remarkably small considering that this analysis integrates the differences among all the described species. This confirms that a small number of large-effect loci contribute to the hamlet radiation.

In particular, one major association peak was recovered in the *casz1* gene region of Linkage Group (LG, putative chromosome) 12 (Figure 4.3B). The phylogeny of this region failed to resolve any of the species with high bootstrap support, but it distinguishes the species that have brown vertical bars (*H. floridae*, *H. ecosur*, *H. puella* and *H. liberte*), blue vertical bars (*H.indigo*) and no vertical bars (Figure 4.3C). The same pattern was obtained when considering the *casz1* gene only (Figure S4.10), but it faded away when only exons were considered (Figure S4.11), indicating that this signal is essentially non-coding. The branching pattern suggests a scenario in which the ancestral hamlets had brown vertical bars, from which blue-barred and non-barred color patterns evolved and diversified rapidly. This is consistent with the high levels of nucleotide diversity of *H. floridae* and *H. ecosur*, which have brown vertical bars (Figure 4.2B), and with the observation that most members of the subfamily Serraninae display cryptic color patterns such as brown vertical bars.

While the gene genealogy of the *casz1* region is striking, it is nonetheless characterized by moderate to low bootstrap support values and it is not perfectly clear-cut, with seven *H. puella* and two *H.*

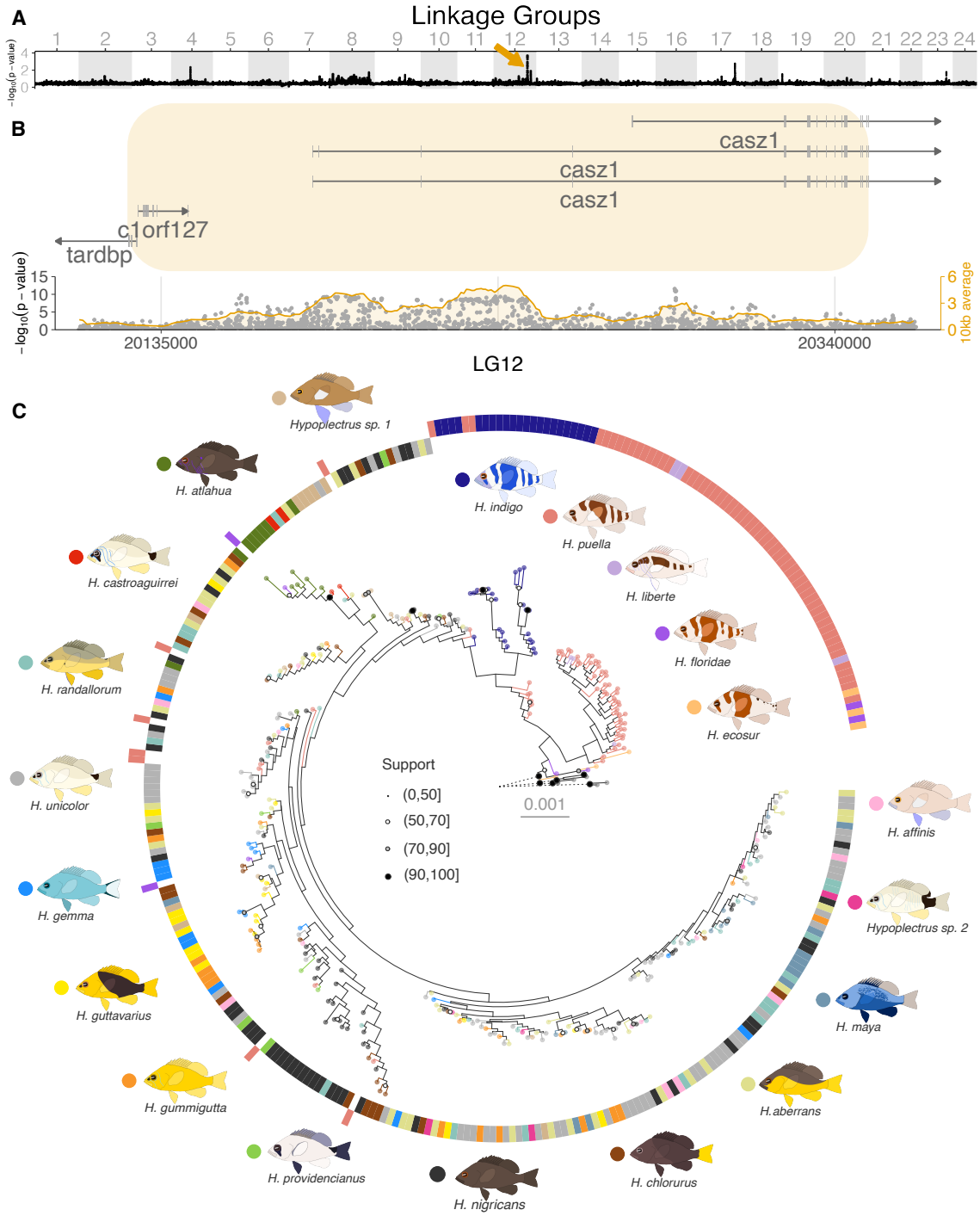



**Figure 4.3 — GWAS for species identity and *casz1* gene tree.** (**A**) Results of the genome-wide association study (GWAS) for species identity. The gray and white blocks represent the 24 linkage groups (LG, putative chromosomes) and the black dots are averages of the GWAS log-transformed *p*-value over 50 kb windows with 5 kb increments. (**B**) close-up on the major association peak in the *casz1* genomic region on LG12. The gray dots represent the GWAS log-transformed *p*-value for each individual SNP and the orange line its average over 10 kb windows. The x-axis shows the position on LG12 (in bp) and the orange rectangle highlights the region that shows an association >1.5 considering 50 kb windows. (**C**) Phylogeny of the *casz1* gene region highlighted in orange in panel **B**, which spans 0.2 Mb. Ring, branch and tip point colors indicate the hamlet species as shown by the icons surrounding the tree. The outer ring corresponds to species that have a barred color pattern and the inner ring the species that have a non-barred color pattern. Node point size and fill represent node support. The eight *Serranus* outgroup samples are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualisation of the hamlet radiation. The same pattern was obtained when considering the *casz1* gene only (Figure S4.10). The phylogenies of the other major association peaks are shown in Figures S4.12–S4.18.

*floridae* samples grouping with non-barred samples (Figure 4.3C). The phylogenies of the other major association peaks are even more diffuse (Figures S4.12–S4.18). Some of these local phylogenies recover the small and large clades or the most diverged species, but beyond that they do not clearly relate to species, geography or color pattern except for a few trends. For example, the association peak on LG04 tends to group samples from the species that are blue together (*H. indigo*, *H. maya* and *H. gemma*; Figure S4.12). This is consistent with previous analyses showing that this genomic region is associated with body color (Coulmance et al., 2024). One genomic region on LG08 tends to cluster *H. unicolor* and *Hypoplectrus sp. 2* individuals, which are white (Figure S4.13), and two genomic regions on LG19 and 23 tend to group *H. nigricans* samples, which are black (Figures S4.17, S4.18). Some regions also show reduced genetic variation, with almost the same haplotype shared by several individuals from different species and locations (e.g. the association peak on LG17, which contains short- and long-wave sensitive opsin genes; Figure S4.16). In these cases, selection is likely to be the cause of the association peak. Species are broadly better recovered when considering the eight high-association regions jointly (Figure S4.19), but here again the deep nodes have low support values. Just four species are recovered as monophyletic in this tree (*H. guttavarius*, *H. maya*, *H. gemma*, and *H. liberte*), and among these only *H. gemma* with high bootstrap support.

In sum, we identified a single genomic interval with a phylogenetic signal that relates to species differences. This region is centered around the *casz1* gene and its tree topology is linked to color pattern, the major functional and ecological trait of the radiation. *Casz1* encodes a zinc-finger transcription factor that orchestrates cell differentiation in retinal and neural progenitors, T helper cells and during cardiac morphogenesis (Mattar et al., 2018, 2021; Bhaskaran et al., 2018; Liu et al., 2014), suggesting a role as a housekeeping gene. In line with this, we found *casz1* to be co-expressed with a large number of genes in the skin, retina and brain, three tissues that relate to the formation and perception of color pattern (Figure S4.20). Interestingly, *casz1* expression was orders of magnitude higher in the retina than in the skin and brain (Figure S4.21). This is consistent with the fact that *casz1* plays a role in photoreceptor development (Mattar et al., 2015, 2018, 2021), and that the visual system of fishes grows continuously and therefore expresses developmental genes. The role of *casz1* for the establishment and maintenance of photoreceptor gene expression profiles (Mattar et al., 2015, 2018, 2021) led us to look into the genes involved in the phototransduction cascade specifically, and we found that a large share of them were co-expressed with *casz1* in the retina (Figure S4.22). Furthermore, some of these genes, such as *PDE6* and *RDH*, were also co-expressed in the skin and brain in non-visual contexts (Figure S4.22), suggesting possible pleiotropic effects of *casz1* in different tissues. Considering that the association peak is centered around the *casz1* transcription start site (Figure 4.3B), we further tested for differential expression and splicing of *casz1* among species and found no differences (Figure S4.21, S4.23). We also examined the expression of transposable elements in

the *casz1* region and found neither differential expression among species (Figure S4.24) nor elevated transposon expression compared to the rest of the genome (Figure S4.25). Finally, we assembled hamlet genomes *de novo* using long-read sequencing to look into structural variation in the *casz1* region and identified no major structural differences among species (Figure S4.26). Taken together, considering the association of *casz1* with color pattern and its strong expression in the retina, we speculate that it might pleiotropically regulate both color pattern and mate choice. This would provide a mechanism for the explosive radiation of the hamlets, paralleling the situation in *Heliconius* butterflies where a gene linked to male approach behavior is in close proximity to a gene that modulates color pattern (Kronforst et al., 2006; Rossi et al., 2024).

## 4.3 Discussion

We took a phylogenomic perspective to address the genic view of species and speciation at the scale of a whole radiation and uncovered just one well-supported phylogenetic split among species, and few unequivocally monophyletic species. Thus, a large share of the hamlet radiation is phylogenetically unresolved, even with genome-wide data and state-of-the-art phylogenetic reconstruction methods. This lack of phylogenetic signal is striking considering that the hamlets present clear phenotypic differences among species and are reproductively isolated to a large extent through assortative mating. In comparison, the cichlids from Lake Victoria, the most rapid east African cichlid radiation and one of the few other systems that compare to the hamlets in terms of speciation rate, can be phylogenetically resolved with a fraction of the amount of data that we used (Wagner et al., 2013). However, the observation that genome-wide data do not resolve species well should not come as a surprise under the view that species are essentially formed and characterized by the genes that underlie adaptation to ecological and sexual environments, which may represent a small fraction of the genome. This was recently illustrated by another striking case, an instance of homoploid hybrid speciation in *Heliconius* butterflies where one of the two parental species contributed just 1% of the genome that includes genomic regions linked to functional divergence (Rosser et al., 2024).

Following the genic view, we hypothesized that the genomic regions underlying functional divergence would show a phylogenetic signal that relates to species differences, and this is exactly what we found for the *casz1* region that distinguishes brown-barred, blue-barred and non-barred species. Nevertheless, this was less clearly or not the case at all for the other genomic regions that we identified. These results are consistent with a recent high-resolution analysis of color pattern variation in a subset of the radiation (Coulmance et al., 2024), which indicates that the presence/absence of vertical bars is essentially a discrete trait that associates almost exclusively with the *casz1* region, while other traits such as body color vary more continuously and associate with several genomic regions. Many of the regions that associate with body color were identified by our GWAS on species identity, but did not show a phylogenetic signal that clearly reflects species differences. The fact that the pres-

ence/absence of vertical bars is a discrete trait that is essentially governed by a single locus explains why this is the only genomic region that exhibits a phylogenetic signal that reflects species differences. Under polygenic control, as seems to be the case for body color and other species differences, phylogenetic patterns are expected to be diffuse at each individual genomic region since variation at other loci may also affect the trait, either individually or in interaction. Thus, our results are consistent with the genic view but they illustrate that while phylogenetic patterns are clear-cut when functional divergence is controlled by a single locus, they erode when the traits underlying functional divergence have a polygenic basis. This aligns with the observation that species are broadly better recovered when considering the eight high-association regions jointly, but the fact that this phylogeny still poorly recovers species suggests that large-effect genes are only part of the story. Considering that many functional traits may have a polygenic basis (Rockman, 2012), the near-absence of phylogenetic signal at the gene-tree level should not come as a surprise either, and an even more extreme situation in which phenotypic diversification and reproductive isolation develop in the total absence of any phylogenetic signal, both genome-wide and at the gene-tree level, does not seem far-fetched. In this respect our results highlight the limits of the phylogenomic approach to address adaptation, speciation and rapid radiation at both the species and trait levels.

An in-depth discussion as to whether the hamlets should be considered species is provided in the latest taxonomic review of the group (Puebla et al., 2022). For all practical purposes they are recognized as species by the ichthyological community and beyond, but this ultimately depends on the species definition that is considered. In light of our results one may be tempted to conclude that there are just two hamlet species corresponding to the two lineages that we identified, but this view would be completely at odds with the patterns of assortative mating and color variation in the group. In this regard it is worth noting that although the genic view is not a species definition, the genomic regions that underlie functional divergence and the occurrence of linkage disequilibrium among these regions (Hench et al., 2019) may be used to identify species, which relates to the genotypic cluster species definition (Mallet, 1995). Anyhow, the hamlet radiation is characterized by phenotypic differentiation and strong reproductive isolation, two fundamental attributes of species. We show herein that these attributes may unfold and persist in the near-absence of phylogenetic signal, and this conclusion remains valid regardless of the species status of the hamlets. This implies that the phylogenetic signal that we observe in older radiations may have largely developed after phenotypic diversification and prezygotic reproductive isolation, with little relation to the initial evolution of these traits.

## 4.4 Materials and Methods

### Sample collection and sequencing

This study is based on a total of 335 genomes from 327 hamlet (*Hypoplectrus* spp.) and eight outgroup (*Serranus* spp.) samples (Table S4.1). This constitutes a major improvement over the last genome-

wide phylogenetic analysis of the hamlets that included just eight species from three Western Caribbean locations (Hench et al., 2022). The samples were collected between 2004 and 2017 across 15 locations that cover nearly the entire range of the hamlet radiation and include the 18 currently recognized hamlet species (Figures 4.1B, S4.1 and Table S4.1). Part of these genomes were available from previous studies (Hench et al., 2019; Puebla et al., 2022; Hench et al., 2022; Coulmance et al., 2024), three had been sequenced previously but were unpublished, and 104 were sequenced anew for this study. Genomic DNA of these was extracted from fin clips stored in 100% ethanol at –20°C with the MagAttract high-molecular-weight DNA extraction kit (Qiagen). Illumina DNA Prep libraries were prepared and sequenced on three NovaSeq6000 S4 lanes (2 × 151 bp) at the Institute of Clinical Molecular Biology (IKMB, Kiel University).

In addition, genomes of three hamlet species were assembled *de novo* to look into structural variation within the *casz1* region (Table S4.2). High-molecular-weight genomic DNA was extracted from gill tissue using a Qiagen genomic tip DNA extraction kit. DNA was sheared to 18 kb using a Diagenode Megaruptor and size-selected for fragments >10 kb on a SAGE BluePippin. PacBio CCS libraries were prepared according to the SMRTbell Express Template Prep Kit 2.0. Libraries were then sequenced at the IKMB on a single 30-hour SMRT cell on the PacBio Sequel II instrument in CCS mode (Table S4.2).

Finally, gene expression was analyzed from 99 tissue samples. These include 24 previously published retina samples (Hench et al., 2019), as well as 54 and 21 new brain and skin samples, respectively (Table S4.3). These samples included three hamlet species (*H. puella, H. nigricans* and *H. unicolor*), collected in reefs around Isla Colón (Bocas del Toro, Panama). The collection and dissection procedures were randomized and standardized to minimize bias. For the skin, a 1 × 1 cm square was dissected from the area where the barred species have their central bar, with the upper edge parallel to the lateral line. The brain was dissected into three brain regions (telencephalon, diencephalon and optic tectum), which were individually stored, extracted and sequenced, resulting in three brain samples per individual. RNA was extracted from brain and skin tissue using the PureLink RNA Mini Kit with an additional TRIzol lysis step (Invitrogen by Thermo Fisher). Sequencing libraries were prepared using the Illumina TruSeq Stranded mRNA HT Sample Prep Kit Illumina, which includes a ployA enrichment step for mRNA purification, and sequenced on a Illumina NovaSeq platform (2 × 100 bp) at the IKMB (Table S4.3).

### Computational analyses

The following data processing steps and analyses were performed on the high-performance computing cluster ROSA (University of Oldenburg) using Linux shell or Python scripts and Slurm *v23.02.4*, unless noted otherwise. Results were visualized locally in R *v3.4.1* using *ggplot v3.4.4* and related packages.

### Genotyping

Previously available and new datasets were genotyped together using GATK

*v4.1.9* (McKenna et al., 2010). After marking adapters for removal, datasets were individually mapped to the *H. puella* reference genome assembly (Hench et al., 2019) with bwa *v0.7.17*. Read data of the same sample sequenced across multiple lanes were merged, duplicates marked, and samples with a mean coverage of less than 10× removed from further analysis (the coverage of all samples after filtering averaged 22×). After calculating haplotype likelihoods, per-sample VCF files were combined into a single cohort GVCF file, and genotypes called jointly on all 24 reference linkage groups (LG, putative chromosomes). During this step, the workflow was split into parallel tracks including all samples (*phylo*) and hamlet samples only (*phyps*). In both cases, only variant sites were called (*snp*), in addition to all callable sites except indels (*all*). These four datasets were then hard-filtered based on missing data (10% cutoff) and various quality metrics as recommended by the Broad Institute for germline short variants. Finally, the SNP datasets were limited to bi-allelic sites with a minor allele count of at least 2 using VCFtools *v0.1.16* (Danecek et al., 2011). Where applicable, linked sites were removed by either applying a physical distance filter of 5 kb, beyond which very little physical linkage remains in the hamlets (Hench et al., 2019; Moran et al., 2019), or by accepting a maximum correlation coefficient of 0.5 along 50 kb windows by a combination of VCFtools and BCFtools. A phased version of the *phylo-snp* dataset (without minor allele count filter) was produced based on phase-informative reads with extractPIRs v1 and SHAPEIT v2 (Delaneau et al., 2013).

#### 4.4.0.1 Phylogenetic context

Before examining our whole-genome data in detail, we put the radiation into a broader taxonomic and phylogenetic context. First, we extracted the sequences of 23 nuclear and mitochondrial genes used by Fish Tree of Life (Rabosky et al., 2018) from the *phylo-all* dataset (or re-genotyped contigs where necessary). For each of the 18 described hamlet species, the sample with the highest coverage was chosen. These sequences were then combined with the corresponding sequences of the Serraninae subfamily represented in FToL. Alignment and phylogenetic inference followed (Hench et al., 2022), with the exception of employing 200 nonparametric bootstrap replicates.

### Species tree inference

Multiple genome-wide, conventional and coalescent-aware approaches were pursued to reconstruct the phylogeny of the hamlet radiation itself. First, 110 k genome-wide SNPs – the *phylo-snp* dataset filtered by physical linkage – were converted to Fasta format and treated as a conventional alignment containing 70,505 parsimony-informative sites. A thorough Nearest Neighbor Interchange (NNI) search was conducted with IQ-TREE *v2.2.2.7* (Minh et al., 2020) under the GTR+ASC model for ascertainment bias correction (Lewis, 2001). Branch support values were obtained by ultrafast bootstrap approximation (Hoang et al., 2018) with 1000 NNI-optimized replicates.

Second, the same dataset was converted to Nexus format and subjected to an SVDQuartets (Chifman and Kubatko, 2014) analysis implemented in PAUP *v4.0a* (Swofford and et al., 2002). The optimal

quartet-based tree was identified by considering 490 million quartets (95% of the total number of distinct quartets). In the case of heterozygous (ambiguous) sites, counts were distributed over all compatible site patterns. A 50% majority consensus tree was calculated from 200 bootstrap replicates each considering 5 million (1%) quartets.

Third, 2000 non-overlapping windows of 5 kb near-continuous genomic sequence were randomly extracted from the *phylo-all* dataset. A window size of 5 kb was considered because very little linkage remains beyond this distance in the hamlets (Hench et al., 2019; Moran et al., 2019). Extracted sites were converted to Fasta format using a custom Perl script (`github.com/JinfengChen/vcf-tab-to-fasta`), and aligned with MAFFT v7 (Katoh and Standley, 2013). Genes trees corresponding to each window were inferred using IQ-TREE *v2.1.2* based on the best-fit model according to the built-in ModelFinder. A summary tree was computed from these gene trees using ASTRAL *v5.15.5* (Zhang et al., 2018). Branch support values for gene trees were obtained by 1000 ultrafast bootstrap replicates.

### Mitochondrial tree

To investigate the phylogeny of the hamlet mitochondrial genome, we called the genotypes mapping to reference linkage group M, the mitochondrial genome identified in (Hench et al., 2019), from the GVCF above (see Genotyping). Only hamlet samples were considered, and all callable sites taken into account except indels (16,998 bp in total). The resulting VCF files were converted to near-continuous sequences in Fasta format as above. Sequences were checked for equal length to ensure alignment, and then subjected to a combined maximum likelihood tree search and nonparametric bootstrap analysis in RAxML-NG *v1.0.3* (Kozlov et al., 2019). Settings included the GTR+G substitution model, 20 each of random and parsimony starting trees, and 200 bootstrap replicates.

### Admixture analysis

Admixture between the large and small clades was first assessed with Admixture *v1.3.0* using a version of the *phyps* dataset without missing data, filtered by genetic linkage (see above) and converted to BED format with PLINK *v1.90b*. The number of clusters *k* was set to 2 and the clustering unsupervised, i.e. without assigning the samples to the two clades *a priori*.

### Nucleotide diversity ($\pi$) and genealogical nearest neighbors (GNN)

We used an ancestral recombination graph-based approach to estimate nucleotide diversity and GNN proportions (Kelleher et al., 2019). The latter quantifies the amount of shared ancestry among individuals with respect to *a priori* defined groups using local topologies. For this we relied on the unfiltered and phased version of the *phylo-snp* dataset. Ancestral alleles were identified with est-sfs *v2.04* (Keightley and Jackson, 2018) based on VCFtools-derived raw allele counts and the allele identities in all three outgroup species. Sites with missing data in any samples were removed with VCFtools, the data were converted from VCF to tsinfer samples file format in Python *v3.11.4*, and tree sequences inferred with tsinfer *v0.3.1* (Kelleher et al., 2019). Nei's nucleotide diversity (Nei and Li, 1979) was then calculated for each linkage group (LG, puta-

tive chromosomes) with the tskit library (*v0.5.5*) and averaged across LGs. GNN proportions were computed with respect to clade identity ($small$ and $large$), also using tskit(Kelleher et al., 2019), and taken as averages across LG02 and both haplotypes.

### History of effective population size ($N_e$) and divergence

The demographic history of effective population size and divergence among species was inferred with MSMC2 (Schiffels and Durbin, 2014). This analysis was executed through a Nextflow *v20.10.0* pipeline and based on the phased version of the SNP dataset (*phylo-snp*). After removing the outgroup species, the density of heterozygous sites was extracted for each individual, taking into account individual sequencing depth and mapping quality to the *H. puella* reference genome assembly [masking files were obtained from deduplicated BAM files and with the help of a calling script adapted by (Hench et al., 2022)]. For $N_e$, individuals were randomly placed into groups of three or four samples from the same species and location using a custom R script (without replacement; *H. castroaguirrei* was excluded because only two samples were available). For species divergence, the analysis was restricted to the Gulf of Mexico where species from the two clades co-occur. For each pair of species, two individuals were randomly chosen per species (with replacement between comparisons). In both cases, MSMC input files were created for each group, and MSMC run with a time segmentation pattern of $1 \times 2 + 25 \times 1 + 1 \times 2 + 1 \times 3$ and the average of Watterson's estimator across input datasets ($\theta = 2.55 \times 10^{-3}$). In the case of population divergence, within- and between-species coalescence rates were estimated separately and then combined into a single output file. Times and rates were interpreted assuming a mutation rate of $\mu = 3.7 \times 10^{-8}$ (estimated in the threespine stickleback by Liu et al., 2016).

### *D*-statistics

We used Patterson's *D* and the *f*4-ratio statistic (Patterson et al., 2012) to quantify gene flow among species / locations. Linked sites were first removed with the correlation coefficient threshold described above from a version of the *phylo-snp* dataset containing no missing data (936 k sites). *D* was then calculated with Dtrios in the Dsuite package (Malinsky et al., 2021) from all 67,525 possible trios, with the *Serranus* samples serving as outgroup. Both BBAA and Dmin topologies were considered, and P values were corrected for multiple testing using Bonferroni's method. For each pair of populations, the highest *D*- and *f*4-values as well as associated p-values were retrieved from all trios with these populations in the P2 and P3 positions.

### Demographic inference with linked selection

For this analysis, 935 non-coding, unlinked windows were selected out of the 2000 windows of 5 kb near-continuous genomic sequence described above. This was accomplished by removing windows that overlapped with coding sequence of gene models identified in the reference genome by MAKER (Hench et al., 2019). Two pseudo-haplotypes were generated for each window with VCFtools, SAMtools and BCFtools without relying on phase information. All windows with both haplo-

types were then concatenated and used as input for the Demographic Inferences with Linked Selection software *v1.0* [DILS (Fraïsse et al., 2021)]. The analysis focused on the history of divergence between the small and large clades identified by the phylogenetic analyses. DILS was run with default parameters considering a per-site mutation rate $\mu$ of 3.7 $\times 10^{-8}$ following (Liu et al., 2016). Migration *vs*. isolation models were compared first. Next, submodels of the best model (here migration) were compared, including secondary contact (SC, the daughter populations evolve initially in isolation and exchange alleles upon secondary contact) and isolation with migration (IM, the two daughter populations continuously exchange alleles). Then, models with homogeneous (Nhomo) *vs*. heterogeneous (Nhetero) effective population size were compared, and finally with homogeneous (Mhomo) *vs*. heterogeneous (Mhetero) migration. The posterior probability corresponded to 1 – error rate, which is based on 140,000 simulations.

### Biogeographic analysis

We estimated ancestral ranges for the hamlet species using the R package BioGeoBEARS (Matzke, 2013), considering three regions: Gulf of Mexico, Western Caribbean, and Eastern Caribbean (Figure 4.2B). The author of this program recommends that: (i) operational taxonomic units (OTUs) should ideally be phylogenetic lineages (i.e., genetically isolated populations), and (ii) an ultrametric tree should be used as input. To satisfy the first requirement, we used two alternative pruning schemes: first, by keeping only one individual per monophyletic species/region (Scheme 1, 169 OTUs), and second, by keeping only one individual per monophyletic region (Scheme 2, 58 OTUs). To meet the second requirement, we estimated relative ages for the ancestral nodes in the SNP-based whole-genome tree (Figure 4.2A) using the least-square method implemented in IQ-TREE [LSD2; (To et al., 2016)]. We built a presence/absence matrix by coding each individual based on the region where they were collected, and evaluated six biogeographic models including DEC (Ree and Smith, 2008), DIVA (Ronquist, 1997), and BAYAREA (Landis et al., 2013), with and without jump-dispersal or founder-speciation event (*j*) (Matzke, 2014). We then assessed their fit using Akaike weights as estimated with Akaike Information Criterion scores corrected for small sample size (AICc). As all three models incorporating the *j* parameter demonstrate equivalent best fits (AICc weights = 0.33) for both schemes (Table S4.5), we provide results for six competing analyses (three models per scheme).

### Genome-wide association study (GWAS) for species identity

Association between gene variants and species identity was examined based on the *phyps-snp* dataset using a linear mixed model with GEMMA (Zhou and Stephens, 2012). This approach takes population structure into account by considering a matrix of relatedness among individuals. In order to remove the effect of the major phylogenetic split, this GWAS was first conducted on the large clade only, which contains 15 of the 18 named species. The dataset was transformed to PLINK format using VCFtools and PLINK. G $\times$ P association was calculated on a per-SNP basis for species identity, coded from 1 to

17 and corresponding to the 15 named hamlet species and 2 undescribed species from the large clade. The association between genotype and species identity was tested with a Wald test for each SNP with GEMMA. The results were averaged over 50 kb windows (note that Wald test P values were $-\log 10$ transformed before averaging, so $-\log 10$(p) is reported for every window). The analysis was then repeated with all species and identified the same genomic regions of high association.

### Region-specific trees

Nuclear genomic sequences characterized by a significant association with species identity (50 kb window average $-\log(\text{pvalue}) > 1.5$) were extracted from the *phylo-all* dataset using VCFtools. Files were converted to Fasta format as above, and a combined maximum likelihood tree search and nonparametric bootstrap analysis was conducted using RAxML-NG *v1.0.3* (Kozlov et al., 2019). Settings included the GTR+G substitution model, 20 each of random and parsimony starting trees, and 100 bootstrap replicates. This analysis was repeated with the eight high-association regions concatenated.

### Genome assembly

Following sequencing, HiFi reads with a minimum quality score of Q = 20 were extracted using PacBio SMRT Link. HiFi reads were then assembled with Hifiasm *v.0.19.5* (Cheng et al., 2021) into primary and haplotype assemblies. The initial assemblies (v1) were screened for contamination using the following approach: For each contig, the putative source organism was identified with BLAST+ (Camacho et al., 2009) according to the National Center for Biotechnology Information's (NCBI) nt database downloaded on Apr 04, 2022. Mean coverage per contig was determined by mapping HiFi reads to the contigs with minimap *v2.22* (Li, 2018) and SAMtools *v1.9* (Danecek et al., 2021). Results were visualized using BlobToolKit *v3.3.4* (Laetsch and Blaxter, 2017), and contigs identified as anything other than "Chordata" or "no hit" removed from the assemblies. The filtered primary assemblies were anchored with the help of two RAD-tag based *H. nigricans* recombination maps (Hench et al., 2019) using ALLMAPS (Tang et al., 2015), creating version 2 of the assemblies.

### RNA read filtering, alignment and counting

De-multiplexed and converted Fastq files were quality-checked with FastQC *v0.11.5* (Andrews et al., 2010; Table S4.6). Illumina adapters were removed and low-quality reads were trimmed with Trimmomatic *v0.36* (Bolger et al., 2014) using a sliding-window procedure. Filtered reads were aligned to the *H. puella* reference genome (Hench et al., 2019) using HISAT2 *v2.1.0* (Kim et al., 2015). HTSeq *v0.13.5* (Anders et al., 2015) was used to quantify the number of reads unambiguously mapped per gene, and read counts were filtered and normalized using the standard DESeq2 *v1.42* analysis pipeline (Love et al., 2014). A sample clustering analysis on regularized log-transformed expression values identified two outliers samples (one *H. puella* retina and one *H. nigricans* brain), which were removed from subsequent analyses.

### Co-expression network analysis

Co-expression networks were constructed for each tissue with WGCNA (Langfelder

and Horvath, 2008). Genes with low expression counts were filtered and variance-stabilizing transformation was carried out with DESEq2 (Love et al., 2014). A total of 22,046, 18,249 and 18,423 genes remained after filtering in the brain, retina and skin tissue, respectivley. A soft thresholding power was chosen for each tissue (4 for brain, 3 for retina and 6 for skin) based on the criterion of approximate scale-free topology (Zhang and Horvath, 2005). Based on this power, a weighted correlation network was constructed for each tissue. Modules, i.e. clusters of genes whose expression correlates highly with each other, were identified, as well as the genes belonging to the same module as *casz1*. We then identified over-represented biological processes, molecular functions and cellular components of genes co-expressed with *casz1* in at least two tissues with GOstats *v2.68* (Falcon and Gentleman, 2007) and GSEABase *v1.64* (Morgan et al., 2023). To correct for multiple testing, we used the false discovery rate method implemented in goEnrichment *v1.0* (Hallab, 2015).

### *casz1* splicing analysis

Considering that the strongest association signal within the *casz1* transcription factor was located in the region between the first and the fifth exon, which harbours large introns, we tested for differential splicing of *casz1* among species (*H. puella*, *H. nigricans* and *H. unicolor*) across the skin, brain and retina tissues. Exon-specific reads were counted and tested for differential exon usage with DEXseq *v1.48* (Anders et al., 2012).

### Annotation and analysis of transposable elements

Repeat libraries were generated *de novo* for the *H. puella* reference genome. This was achieved with RepeatModeler *v.2.0* (Flynn et al., 2020), relying on the LTRharvest (Ellinghaus et al., 2008) and LTR_retriever (Ou and Jiang, 2018) options to enhance LTR detection. The libraries were then used to annotate the genome with RepeatMasker *v.4.0.1* (Smit et al., 2013). The resulting repeat annotation and a curated version of the reference gene annotation were used to build a genome index and map our RNA-sequencing reads with STAR *v2.7.10* (Dobin and Gingeras, 2015). Reads were then counted and normalized, and tested for differential expression in transposable elements among species with TEtranscripts *v2.2.1* (Jin et al., 2015).

### Structural variation in the *casz1* region

We extracted the *casz1* region from the *H. puella* reference genome, from the start of the association region (LG12 position 20,135,000) until the end of the *casz1* gene (LG12 position 20,347,811). We then retrieved this same region from our six PacBio haplotype assemblies using the megablast function in BLAST+ *v2.14.1* (Camacho et al., 2009) and SAMtools *v1.18* faidx (Danecek et al., 2021). Finally, pairwise alignments were constructed using Minimap2 *v2.26* (Li, 2018) and checked for structural rearrangements with SyRI *v1.6.3* (Goel et al., 2019). The results were visualized with plotsr *v1.1.1* (Goel and Schneeberger, 2022).

**Ethics statement:** This study builds on a sampling effort conducted over 18 years

(2004–2022) at 15 locations across the Greater Caribbean region. It was developed and conducted in collaboration with local institutions and scientists, following specific Institutional Animal Care and Use Committee (IACUC) protocols (notably STRI 2013-0301-2016, 2017-0101-2020-2, SI-21007, NU 17-0206 and UPR 01009-02-16-2015). It includes co-authors who are or were permanently based in Mexico, Panama, Colombia and Puerto Rico.

**Supplementary Materials include:**

- Supplementary Figures S4.1 to S4.26
- Supplementary Tables S4.1 to S4.6

**Acknowledgements:** We thank Jim Mallet for feedback on an early version of the manuscript and Emma Elena Evenz for input on the MSMC analyses. We are also grateful to the Harte Research Institute for Gulf of Mexico Studies, the BDMY academic group, and Dr. Quetzalli Hernandez and Dr. Nuno Simoes for their support in Campeche Bank field expeditions.

**Funding:** This study was funded by an individual German Research Foundation (DFG) grant to OP (PU571/1–2). MJH was independently supported by the German Research Foundation (DFG, HE 8763/1-1 and 8763/2-1). Data analyses were performed on the high-performance computing cluster ROSA (University of Oldenburg, Germany), funded by the DFG through its Major Research Instrumentation Programme (INST 184/225-1 FUGG) and the Ministry of Science and Culture (MWK) of the state of Lower Saxony.

**Competing interests:** The authors declare that they have no competing interests.

**Data and materials availability:** Newly sequenced raw data were deposited in the European Nucleotide Archive (ENA) under project accession numbers PRJEB75494 (resequenced genomes and PacBio assemblies), PRJEB74501 (skin RNA-sequencing) and PRJEB76552 (brain RNA-sequencing). Individual sample accession numbers are listed in Tables S4.1–S4.3. VCF files and other large datasets were deposited at the European Variation Archive (EVA).

**Code availability:** All workflows and detailed script parameters are available on GitHub at `https://github.com/mhelmkampf/hamlet_phylogeny`.

## 4.5 Supplementary Figures

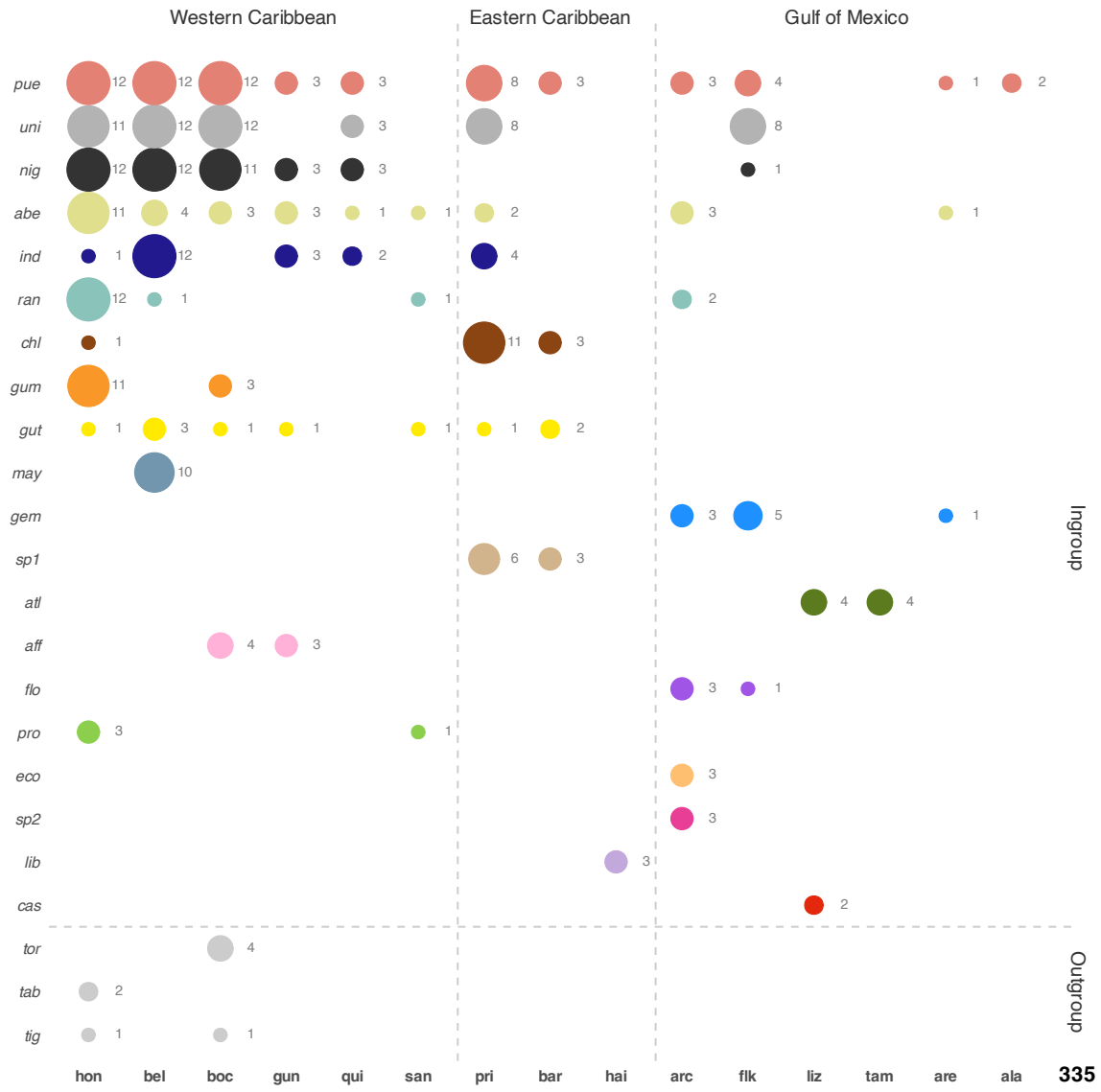



**Figure S4.1 — Overview of the sampling design, comprising 327 samples of all described hamlet species.** pue: *Hypoplectrus puella*, uni: *Hypoplectrus unicolor*, nig: *Hypoplectrus nigricans*, abe: *Hypoplectrus aberrans*, ind: *Hypoplectrus indigo*, ran: *Hypoplectrus randallorum*, chl: *Hypoplectrus chlorurus*, gum: *Hypoplectrus gummigutta*, gut: *Hypoplectrus guttavarius*, may: *Hypoplectrus maya*, gem: *Hypoplectrus gemma*, sp1: *Hypoplectrus sp. 1*, atl: *Hypoplectrus atlahua*, aff: *Hypoplectrus affinis*, flo: *Hypoplectrus floridae*, pro: *Hypoplectrus providencianus*, eco: *Hypoplectrus ecosur*, sp2: *Hypoplectrus sp. 2*, lib: *Hypoplectrus liberte*, cas: *Hypoplectrus castroaguirrei*. In addition, three outgroup species were included – tor: *Serranus tortugarum*, tab: *Serranus tabacarius* and tig: *Serranus tigrinus*. Samples were collected at 15 locations covering the Gulf of Mexico [Tamiahua (tam), Antón Lizardo (liz), Cayos Arcas (arc), Cayo Arenas (are), Alacranes Reef (ala) and the Florida Keys (flk)], the Western Caribbean [Quintana Roo (qui), Belize (bel), Honduras (hon), San Andrés (san), Bocas del Toro (boc) and Guna Yala (gun)], and the Eastern Caribbean [Haiti (hai), Puerto Rico (pri) and Barbados (bar)].

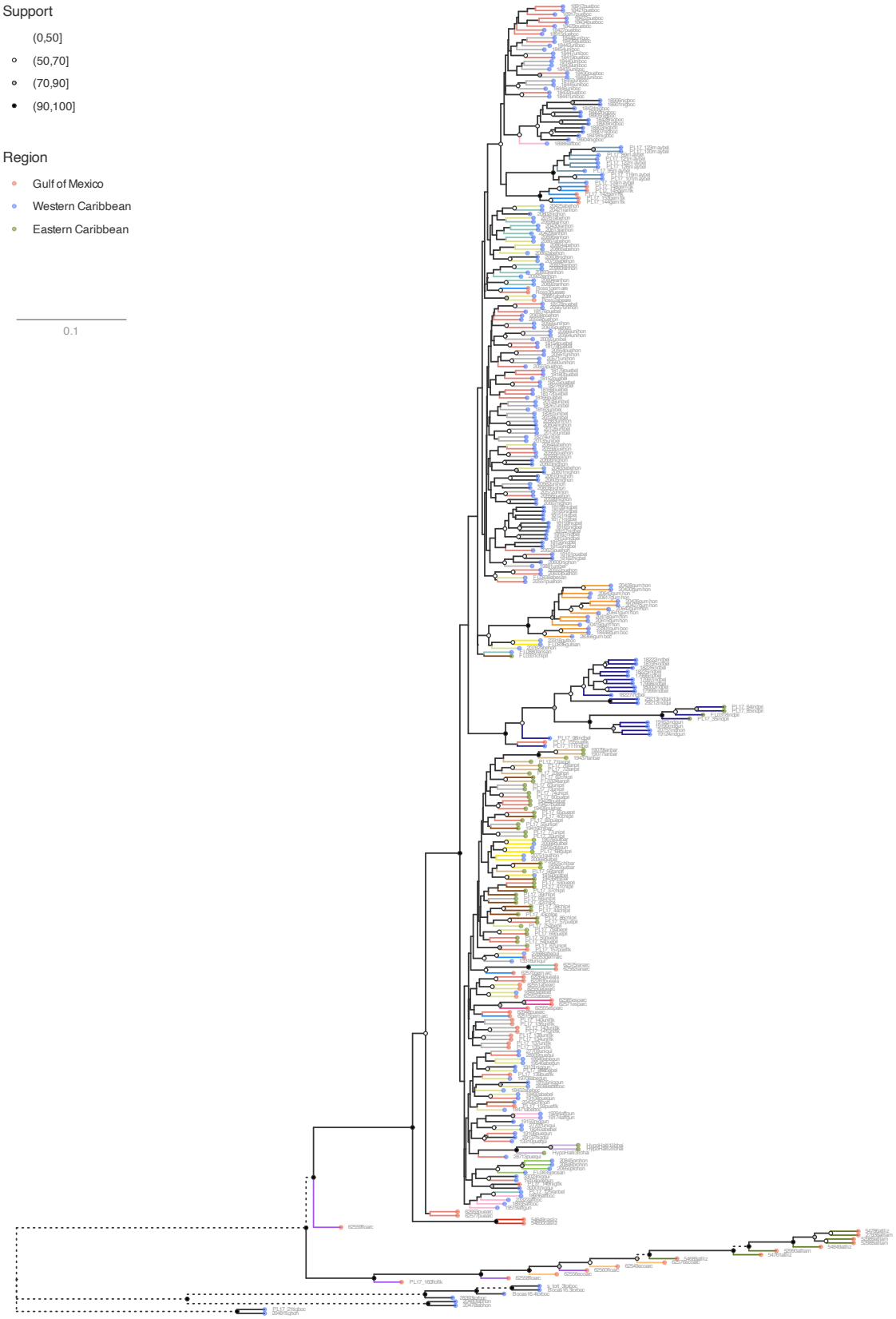



**Figure S4.2 — Summary tree inferred with ASTRAL-III from 2000 local trees.** Each local tree was based on a randomly chosen 5 kb genomic window. Local trees were reconstructed using IQ-TREE v2, the best-fit substitution model according to ModelFinder, and 1000 ultrafast bootstrap replicates. Internal branches reflect concordance among local trees (measured in coalescent units; terminal branches were set to an arbitrary constant). Dashed branches are displayed at 5% of their original value to improve visualization. Values at the nodes denote quartet (as opposed to bipartition) support, expressed in local posterior probabilities. Branch colors indicate species as defined in Figures 4.1 and S4.1. The last six letters of the sample names refer to the species (first three letters) and location (last three letters) as in Figure S4.1.

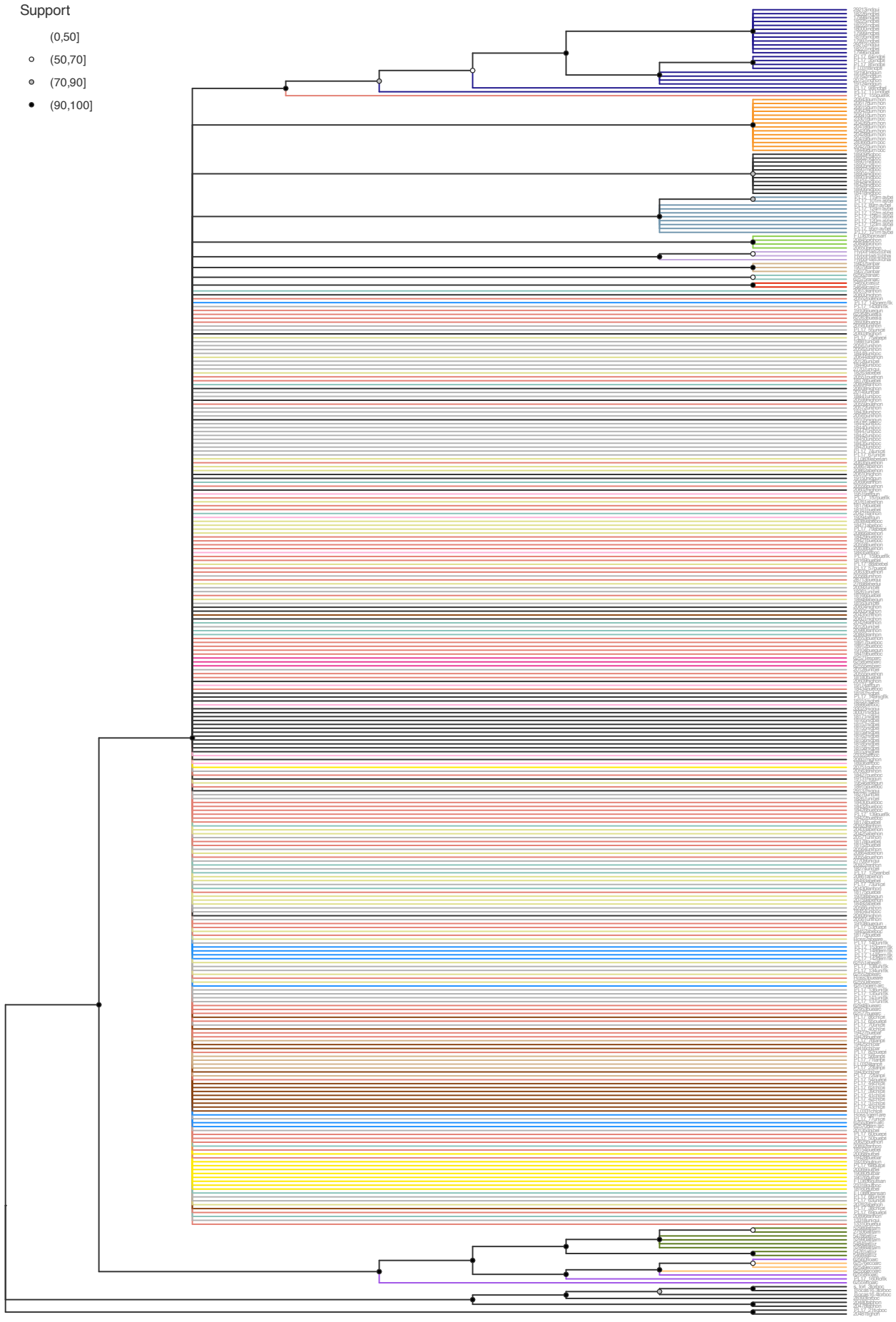



**Figure S4.3 — SVDQuartets 50% majority consensus tree based on approximately 110 k SNPs.** The consensus was derived from 200 bootstrap replicates each considering 5 million quartets (1% of all possible quartets). Branch colors indicate species as defined in Figures 4.1 and S4.1. The last six letters of the sample names refer to the species (first three letters) and location (last three letters) as in Figure S4.1.

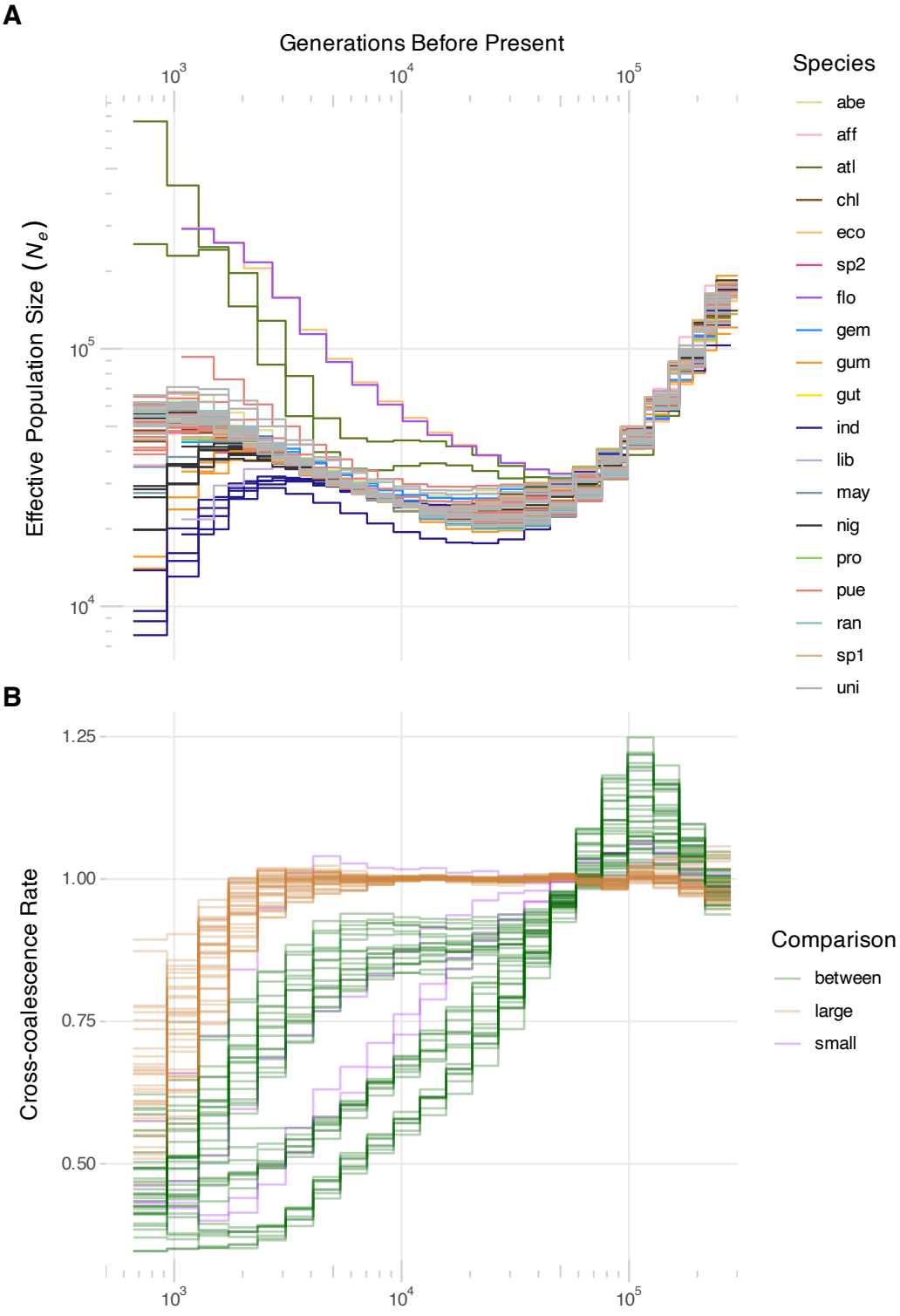



**Figure S4.4 — Effective population size and divergence history as inferred by MSMC2.** (**A**) Effective population size over time, with each line representing a set of three or four genomes of the same species (randomly chosen without replacement) *H. castroaguirrei* was excluded from this analysis because only two genomes were available for this species. (**B**) Cross-coalescence rates between pairs of species from the Gulf of Mexico, where species from the two clades co-occur. Each line represents an independent comparison of two genomes each from two species of the same location within the Gulf of Mexico. Comparisons between small and large clade species are shown in green, within large clade comparisons in orange, and within small clade comparisons in purple. Rates of one and zero indicate completely shared and no shared ancestry, respectively. Time in generations was based on a per-site mutation rate of $\mu = 3.7 \times 10^{-8}$. The four most recent and most ancient time segments were considered unreliable and excluded from plotting. The results suggest that the small and large clades separated between 50,000 and 70,000 generations ago (corresponding to 50,000–210,000 years on the basis of an estimated generation time of 1–3 years).

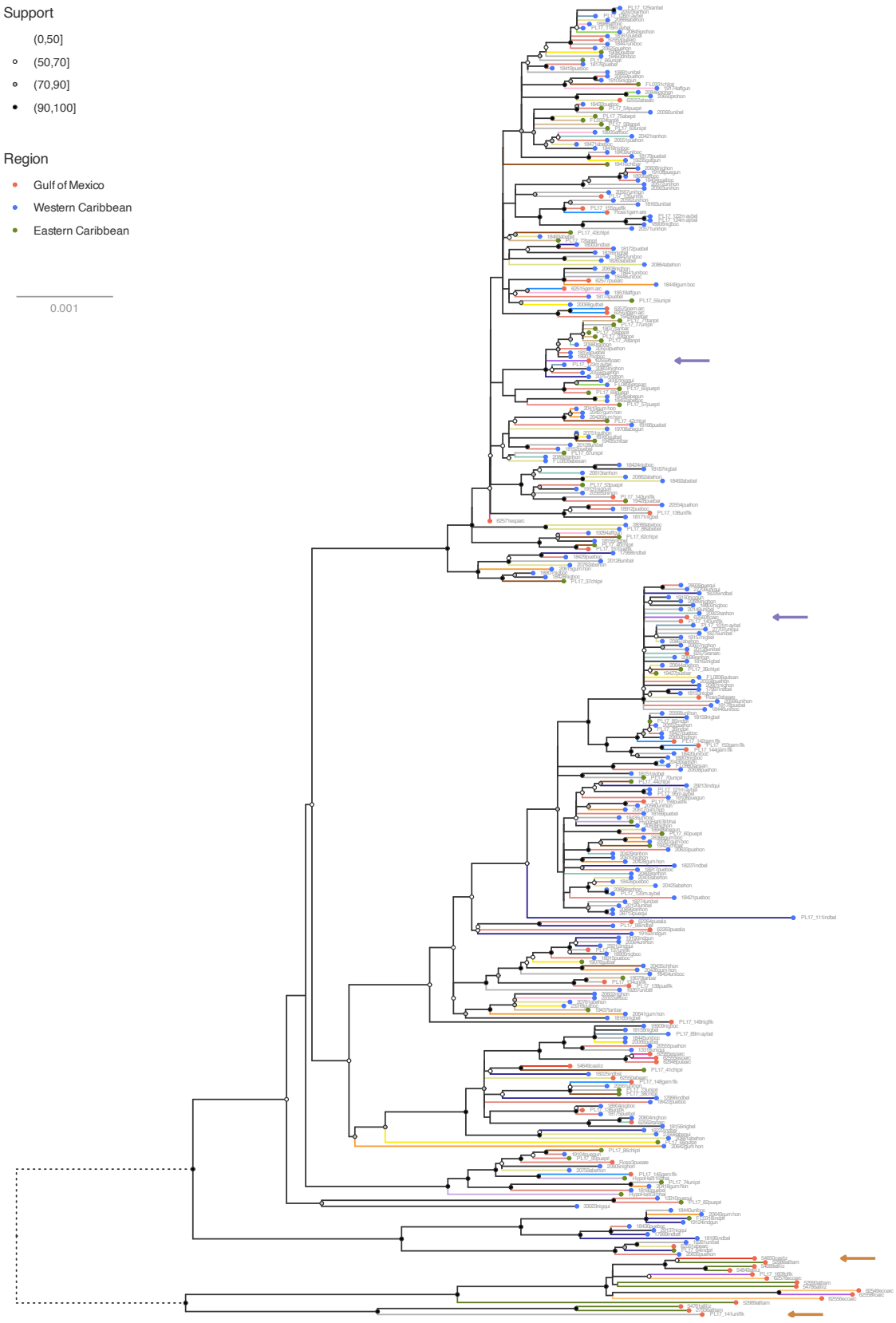



**Figure S4.5 — Mitochondrial genome phylogeny generated with RAxML-NG.** The analysis is based on the hamlet-only dataset including all mitochondrial sites (approximately 17 kb), the GTR+G model and thorough search parameters. Support values at the nodes were obtained from 200 nonparametric bootstrap replicates. Dashed branches are displayed at 10% of their original value to improve visualization. Branch colors indicate species as defined in Figures 4.1 and S4.1. The last six letters of the sample names refer to the species (first three letters) and location (last three letters) as in Figure S4.1. Colored arrows highlight individuals with mitonuclear discordance (nuclear genome from the large clade and mitochondrial genome from the small clade, in orange, and vice versa, in purple).

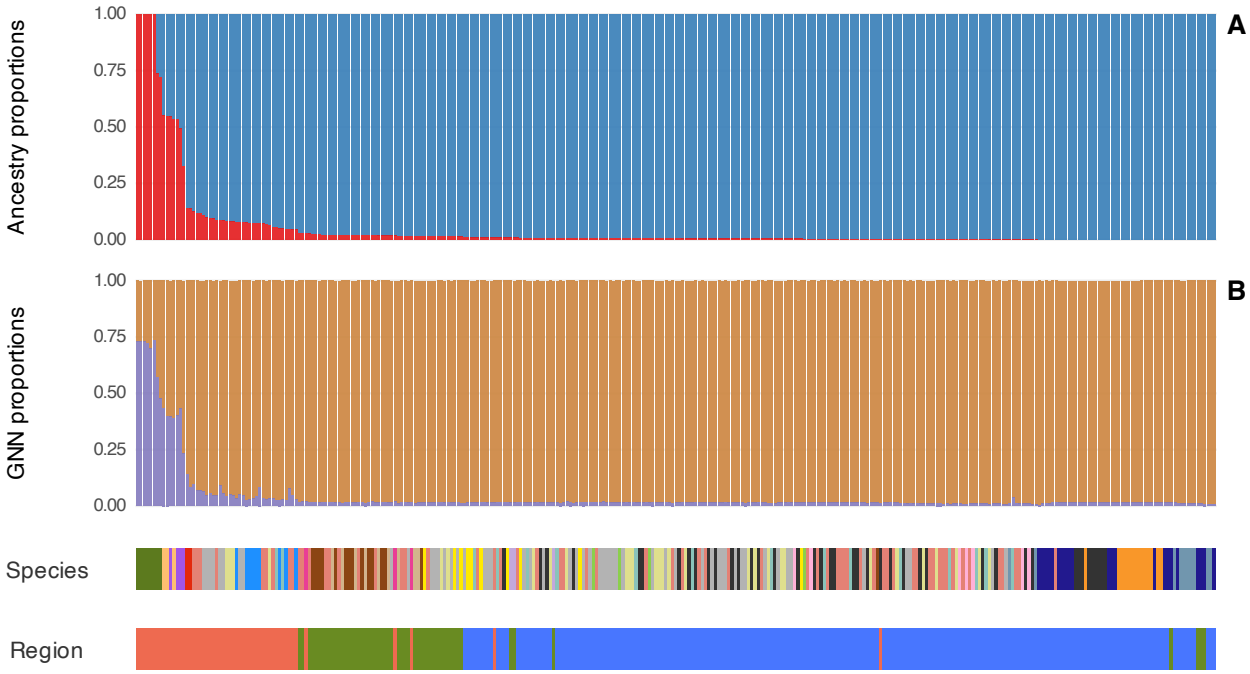



**Figure S4.6 — Admixture between the large and small clades.** (**A**) Admixture plot for *k* (number of ancestral components) = 2 based on the genome-wide SNP dataset (*phyps-snp*). Each individual is represented by a vertical bar broken up into two colored segments whose lengths are proportional to the contributions of the two ancestral components to the genome of the individual. The two ancestral components match the two clades identified by the phylogeny, with evidence of admixture between the two. (**B**) Average genealogical nearest neighbors (GNN) proportions for all individuals. Purple and orange bar segments indicate GNN proportions with respect to the small and large clade, respectively. This analysis is based on both haplotypes of putative chromosome LG02 and was performed with tskit. Individuals are arranged in the same order as in the admixture plot above. Assuming that LG02 is representative of the whole genome, the plot reflect the amount of shared ancestry among individual haplotypes with respect to the two clades. Here again there is evidence of admixture between the two clades. The colored bars below the plots indicate species and region following Figures 4.1, 4.2 and S4.1.

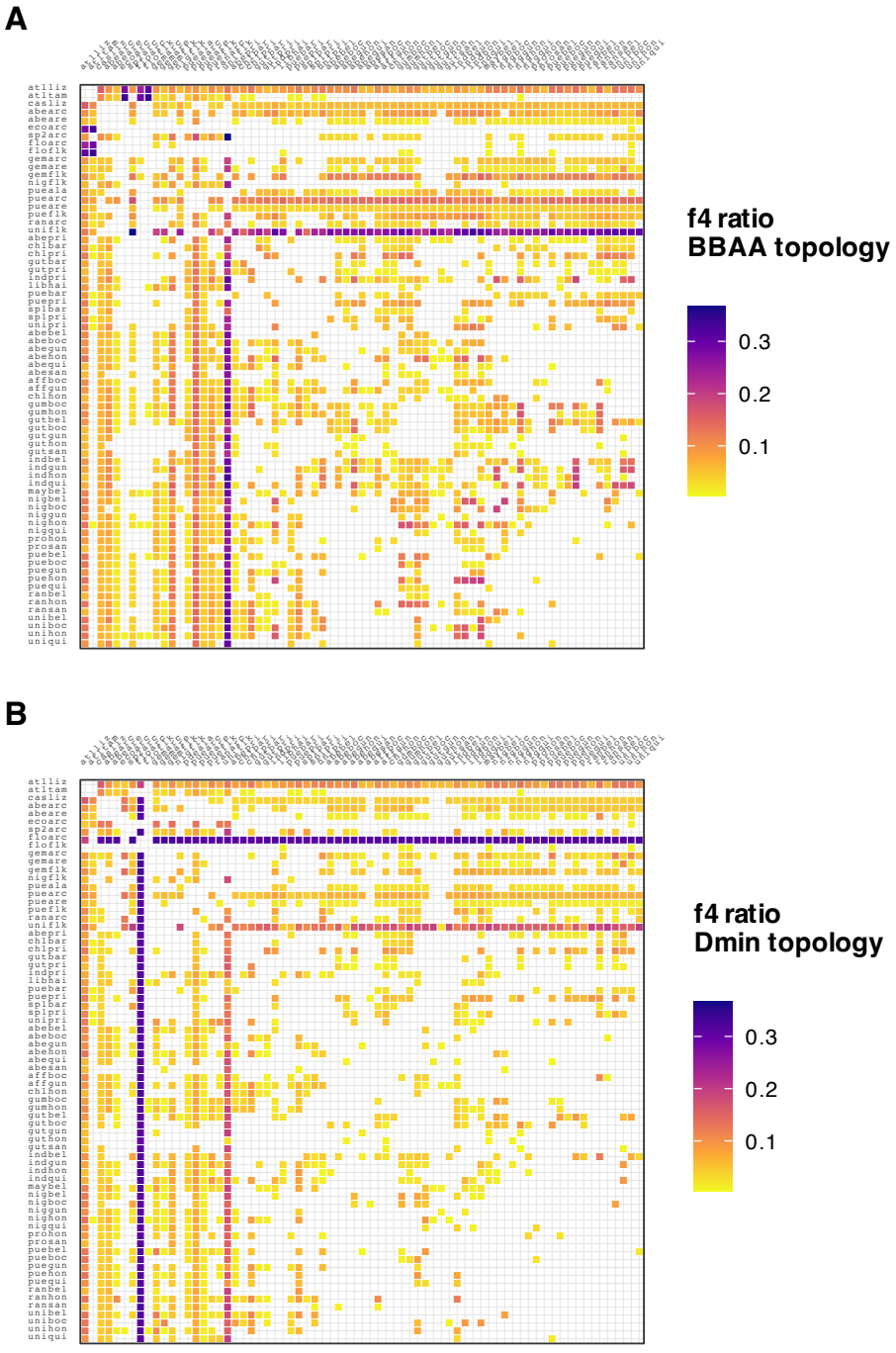



**Figure S4.7 — Introgression between the two clades.** $f_4$-ratios calculated by Dtrios (Dsuite package) between all population pairs, considering (**A**) BBAA and (**B**) Dmin topologies. The x- and y-axes represent positions P2 and P3, respectively. For each pair, only the trio with the lowest Bonferroni-corrected *p*-value regarding *D* is considered (same-pair trios differ at the P1 position). Tile colors indicate $f_4$-ratios corresponding to *D*-values that were significant at the 0.05 level, with white tiles representing pairs without significant *D*-values. The analysis was based on 936 k SNPs (*phyps-snp* dataset with linked sites removed and containing no missing data), and the *Serranus* samples serving as outgroup. The $f_4$-ratio estimates the proportion of introgressed DNA in admixed populations. Thus, gene flow between populations seems to be widespread among hamlets, including between clades, and affect substantial parts of the genome.

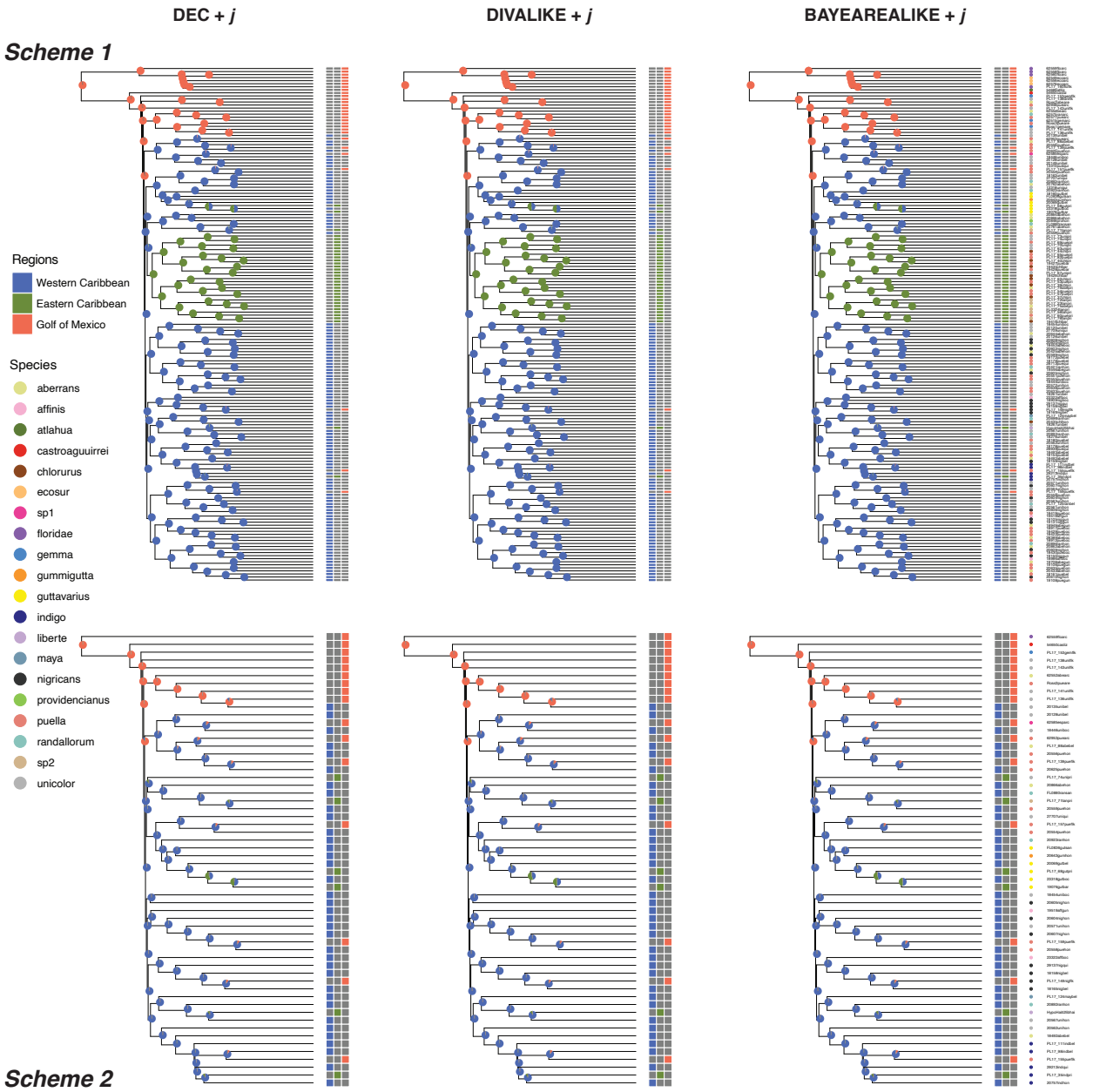



**Figure S4.8 — Inference of ancestral ranges.** This analysis was conducted with BioGeoBEARS based on the ultrametric whole-genome tree generated with IQ-TREE, pruned to retain only one individual per monophyletic species/region (Scheme 1) or only one individual per monophyletic region (Scheme 2). Results shown depict ancestral ranges according to three competing models (see Table S4.5). Branch lengths are relative ages from 0 at the root to 1 at the tips, estimated by the least-square method implemented in IQ-TREE based on the ancestral nodes in the SNP-based whole-genome tree (Figure 4.2A).

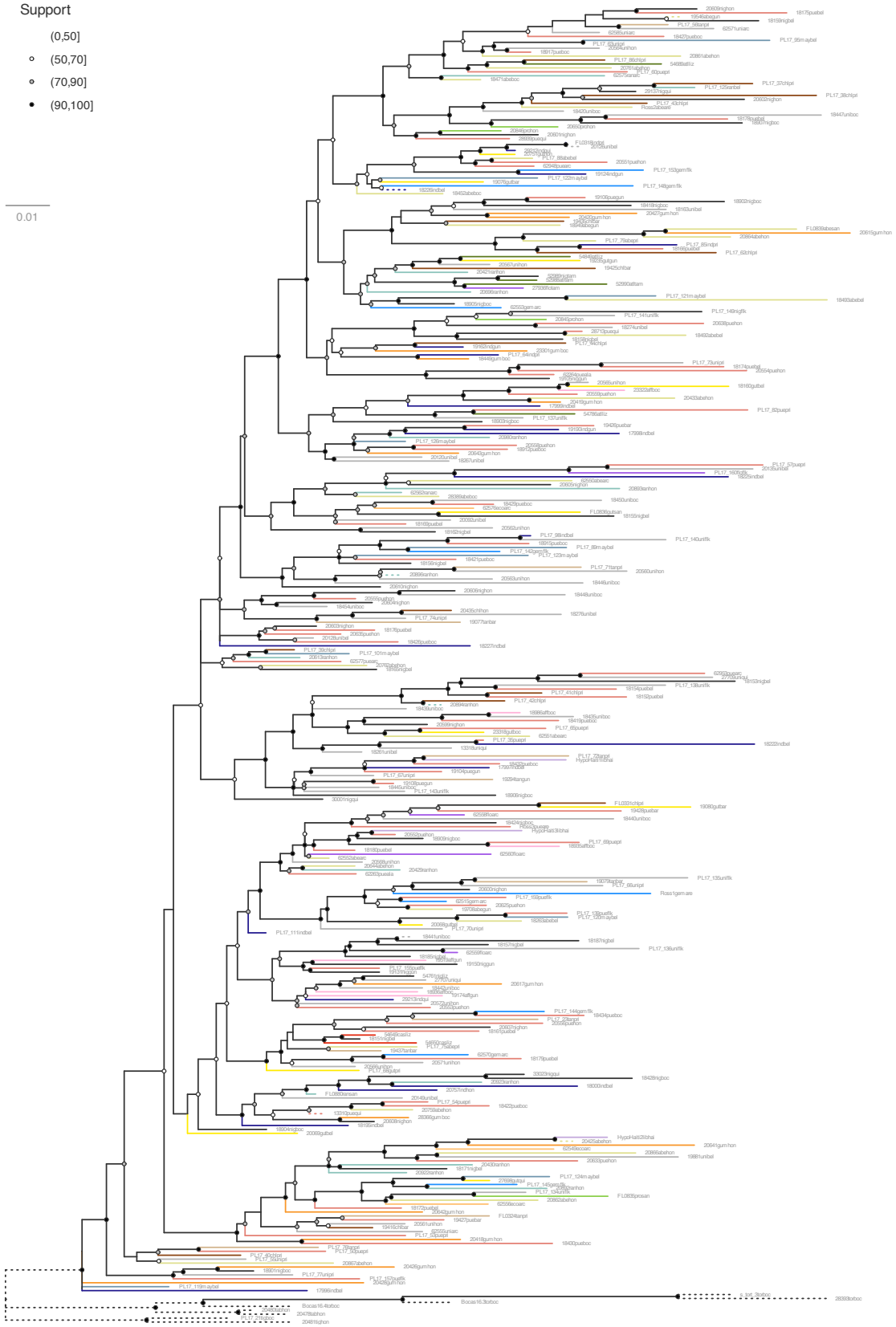



**Figure S4.9 — Local phylogeny with highest mean support**. The highest mean support is calculated across nodes among the 2000 genomic windows of 5 kb length that were considered. Support values were obtained by ultrafast bootstrap approximation with 1000 replicates implemented in IQ-TREE. Dashed branches are displayed at 5% of their original value to improve visualization. Branch colors signify species as defined in Figures 4.1 and S4.1. The last six letters of the sample names refer to the species (first three letters) and location (last three letters) as in Figure S4.1.

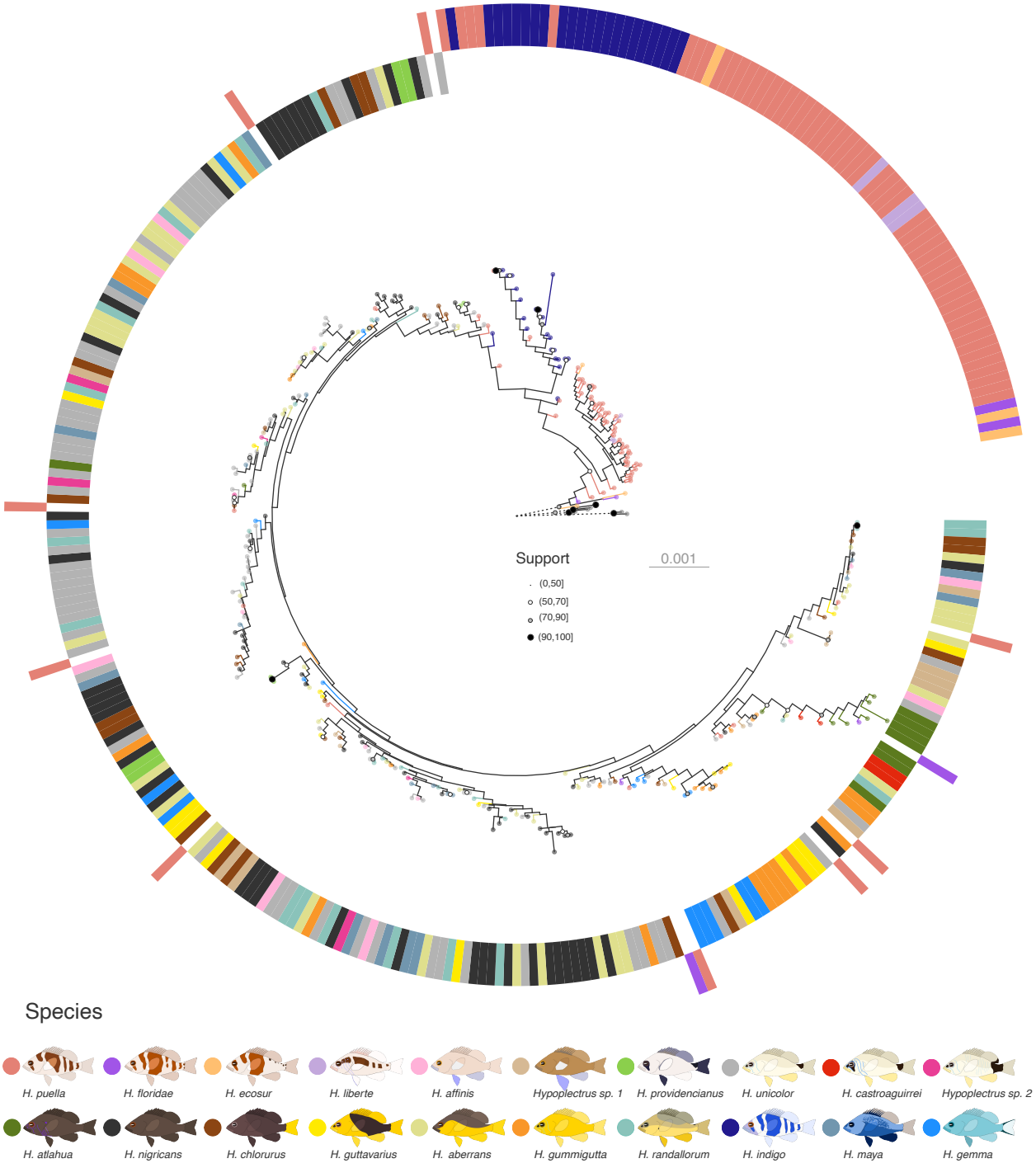



**Figure S4.10 — Phylogeny of the *casz1* gene including both introns and exons, spanning >152 kb.** The rings, branch and tip point colors indicate hamlet species as shown by the icons below the tree. The outer ring corresponds to species that have a barred color pattern and the inner ring the species that have a non-barred color pattern. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation.

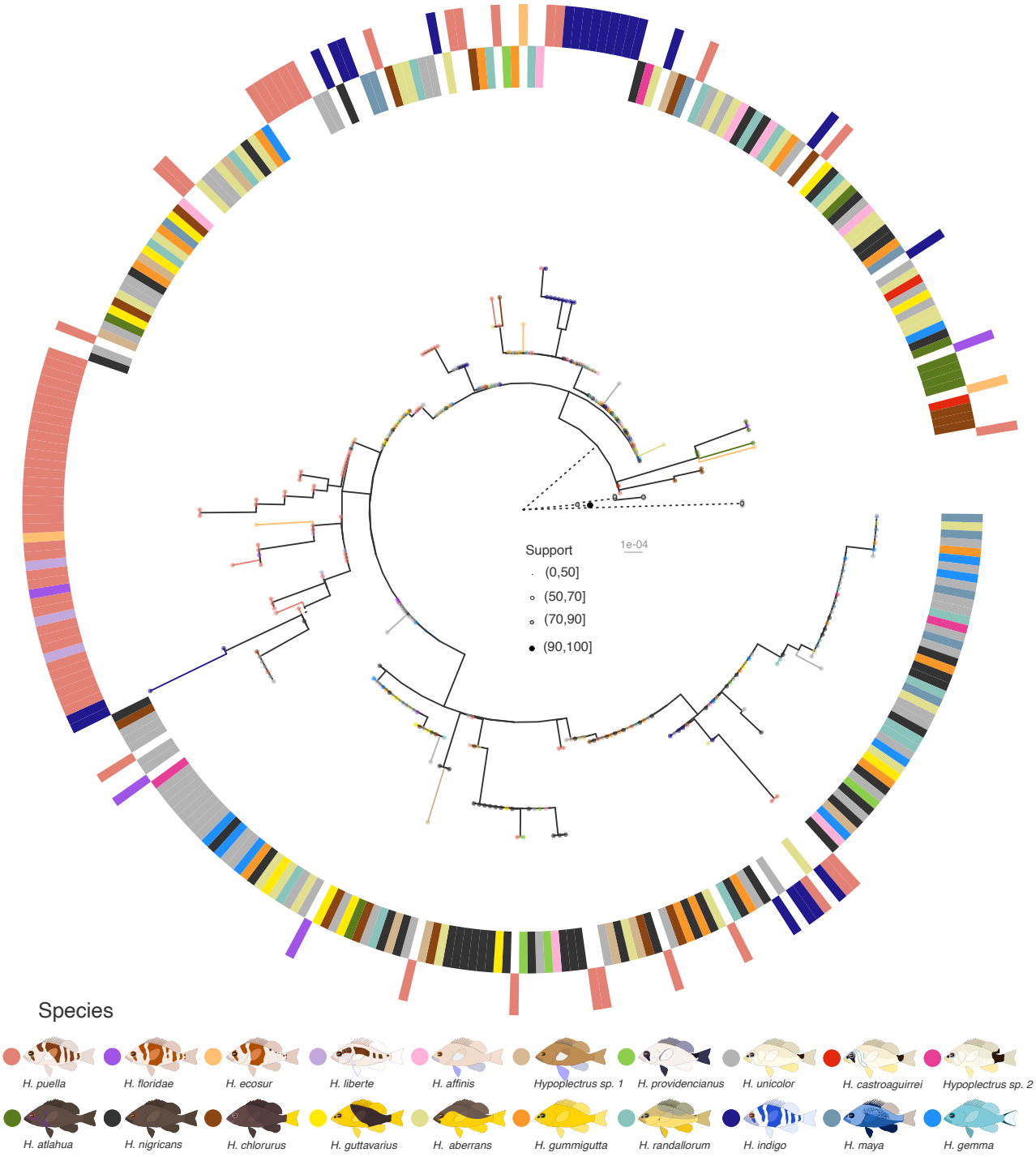



**Figure S4.11 — Phylogeny of the *casz1* gene including exons only, spanning 6689 bp.** The outer ring, branch and tip point colors indicate hamlet species as shown by the icons below the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation.

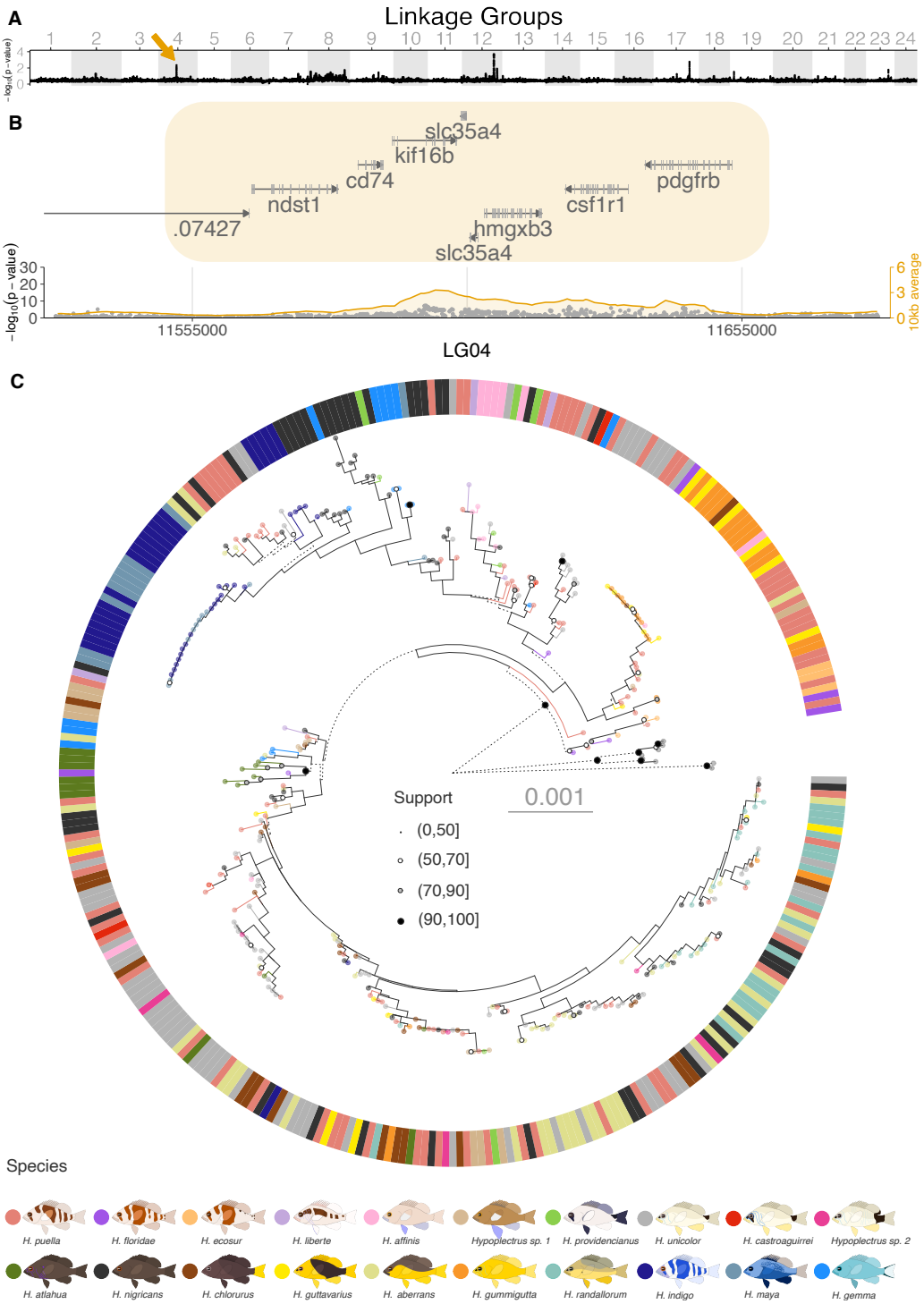



**Figure S4.12 — GWAS for species identity and phylogeny of the LG04 association peak.** (**A**) Results of the genome-wide association study (GWAS) for species identity. The gray and white blocks represent the 24 linkage groups (LG, putative chromosomes) and the black dots are averages of the log-transformed *p*-value of the GWAS association over 50 kb windows with 5 kb increments. (**B**) Close-up on the LG04 association peak. The x-axis shows the position on LG04 (in bp) and the orange rectangle highlights the region that shows an association >1.5 considering 50 kb windows. Grey dots represent the GWAS log-transformed *p* value for each individual SNP, and the orange line its average over 10 kb windows. (**C**) Phylogeny of the LG04 region, covering 91,382 bp centered on the GWAS association peak highlighted in orange in **B**. The outer ring, branch and tip point colors indicate hamlet species as shown by the icons below the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation.

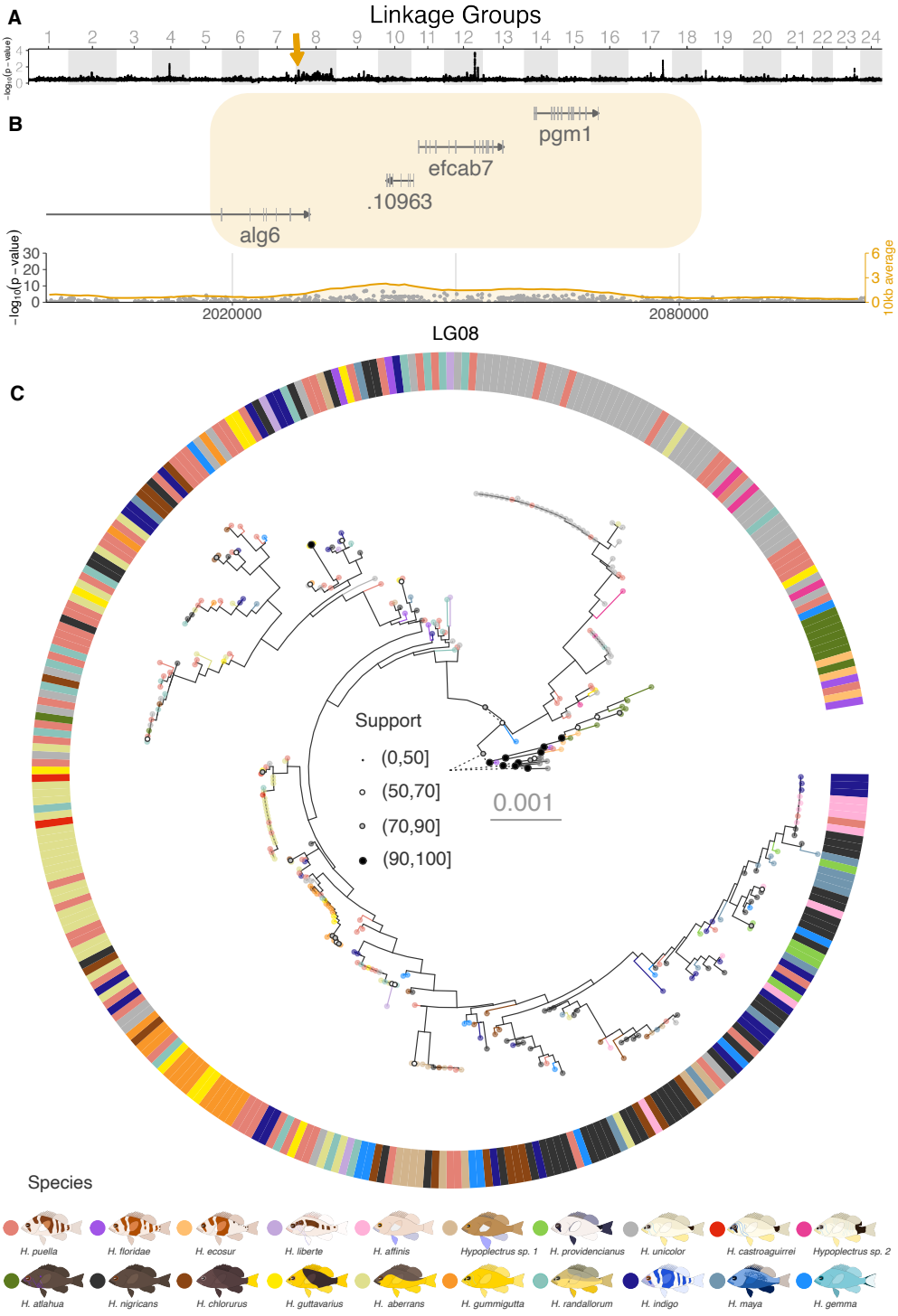



**Figure S4.13 — GWAS for species identity and phylogeny of the first LG08 association peak.** (**A**) Results of the genome-wide association study (GWAS) for species identity. The gray and white blocks represent the 24 linkage groups (LG, putative chromosomes) and the black dots are averages of the log-transformed *p*-value of the GWAS association over 50 kb windows with 5 kb increments. (**B**) Close-up on the first LG08 association peak. The x-axis shows the position on LG08 (in bp) and the orange rectangle highlights the region that shows an association >1.5 considering 50 kb windows. Grey dots represent the GWAS log-transformed *p* value for each individual SNP, and the orange line its average over 10 kb windows. (**C**) Phylogeny of the LG08 region, covering 55,355 bp centered on the GWAS association peak highlighted in orange in **B**. The outer ring, branch and tip point colors indicate hamlet species as shown by the icons below the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 2% of their actual length to improve the visualization of the hamlet radiation.

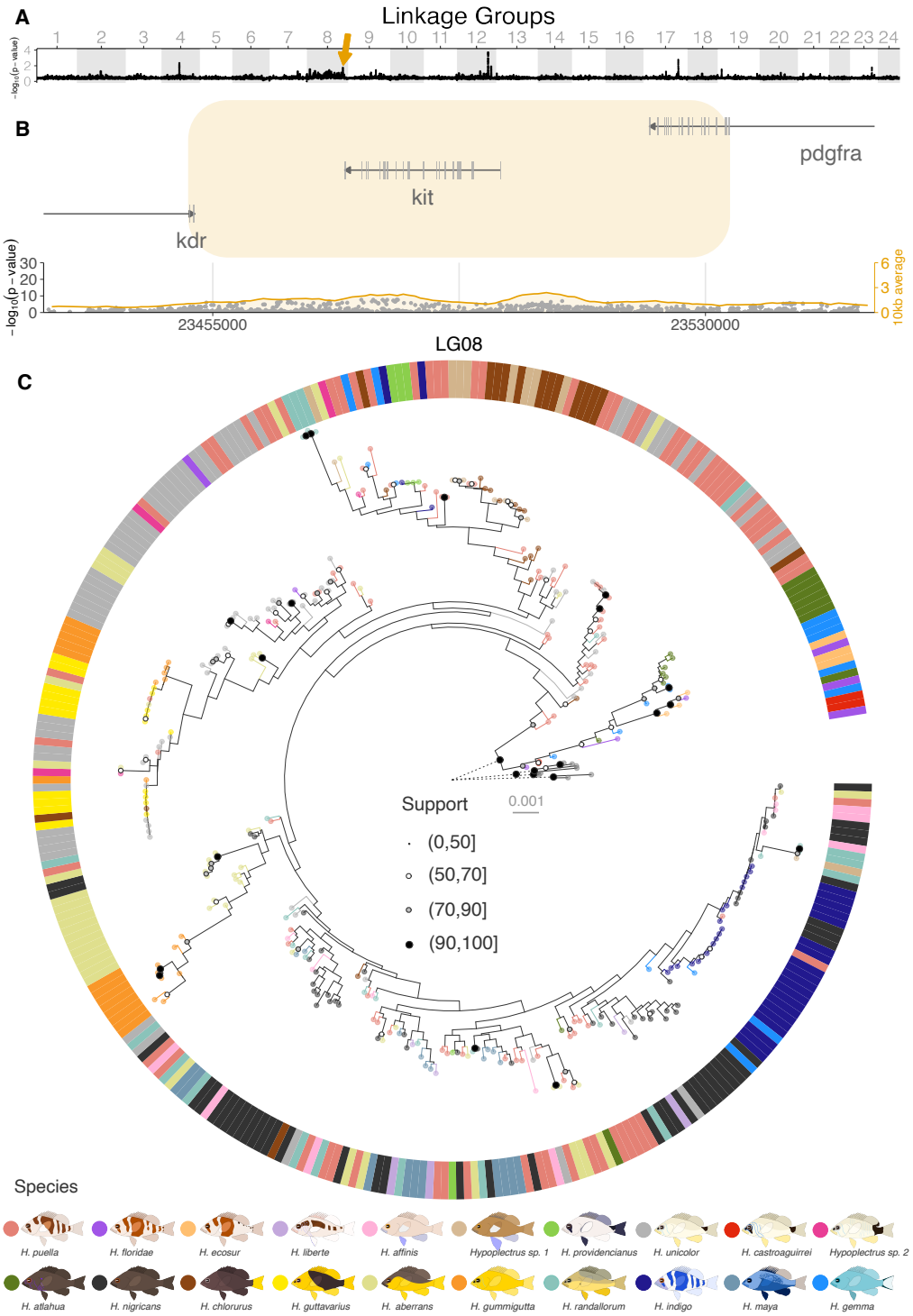



**Figure S4.14 — GWAS for species identity and phylogeny of the second LG08 association peak.** (**A**) Results of the genome-wide association study (GWAS) for species identity. The gray and white blocks represent the 24 linkage groups (LG, putative chromosomes) and the black dots are averages of the log-transformed *p*-value of the GWAS association over 50 kb windows with 5 kb increments. (**B**) Close-up on the second LG08 association peak. The x-axis shows the position on LG08 (in bp) and the orange rectangle highlights the region that shows an association >1.5 considering 50 kb windows. Grey dots represent the GWAS log-transformed *p* value for each individual SNP, and the orange line its average over 10 kb windows. (**C**) Phylogeny of the second LG08 region, covering 67,783 bp centered on the GWAS association peak highlighted in orange in **B**. The outer ring, branch and tip point colors indicate hamlet species as shown by the icons below the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation.

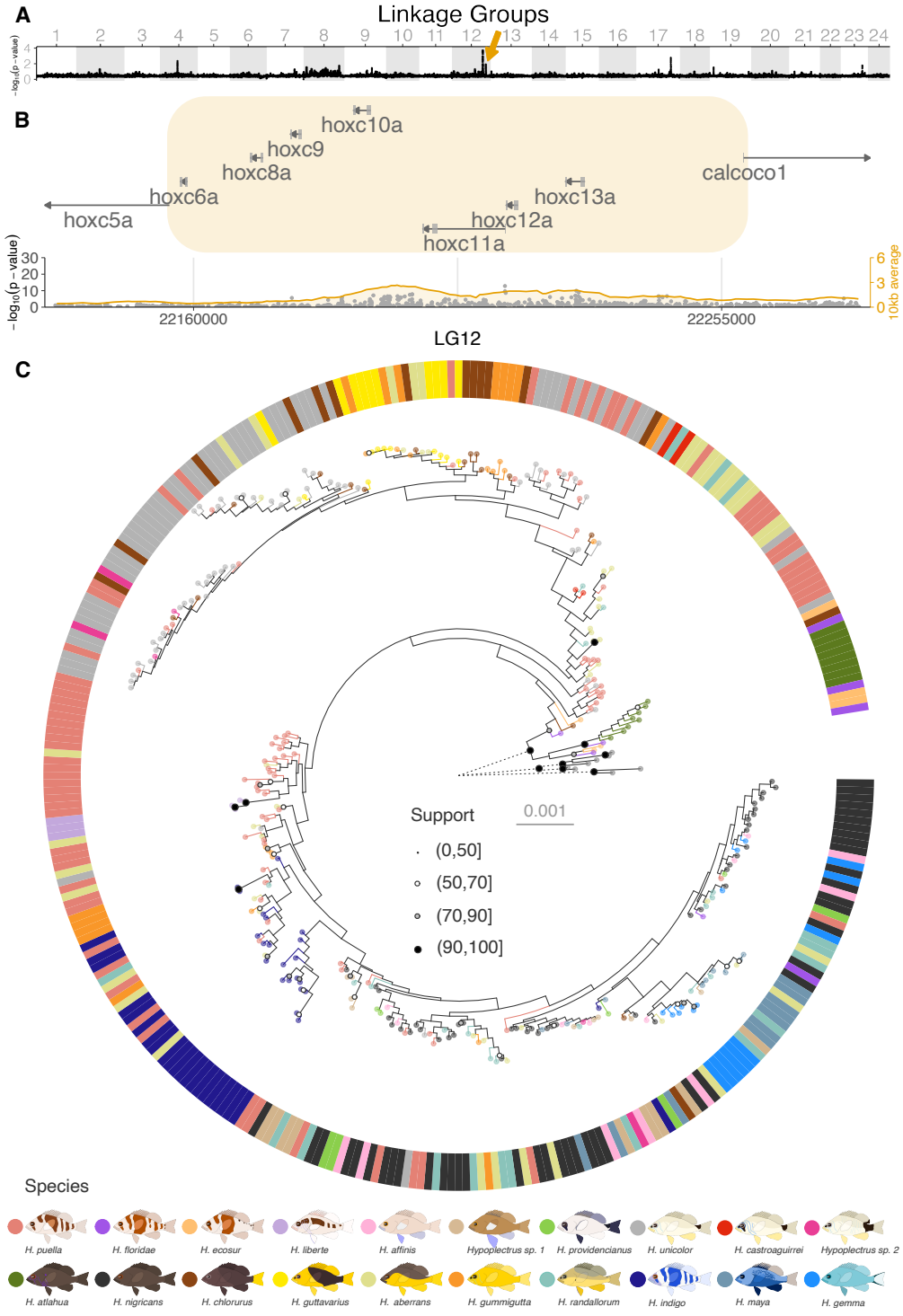



**Figure S4.15 — GWAS for species identity and phylogeny of the second LG12 association peak.** (**A**) Results of the genome-wide association study (GWAS) for species identity. The gray and white blocks represent the 24 linkage groups (LG, putative chromosomes) and the black dots are averages of the log-transformed *p*-value of the GWAS association over 50 kb windows with 5 kb increments. (**B**) Close-up on the second LG12 association peak. The x-axis shows the position on LG12 (in bp) and the orange rectangle highlights the region that shows an association >1.5 considering 50 kb windows. Grey dots represent the GWAS log-transformed *p* value for each individual SNP, and the orange line its average over 10 kb windows. (**C**) Phylogeny of the second LG12 region, covering 88,334 bp centered on the GWAS association peak highlighted in orange in **B**. The outer ring, branch and tip point colors indicate hamlet species as shown by the icons below the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation.

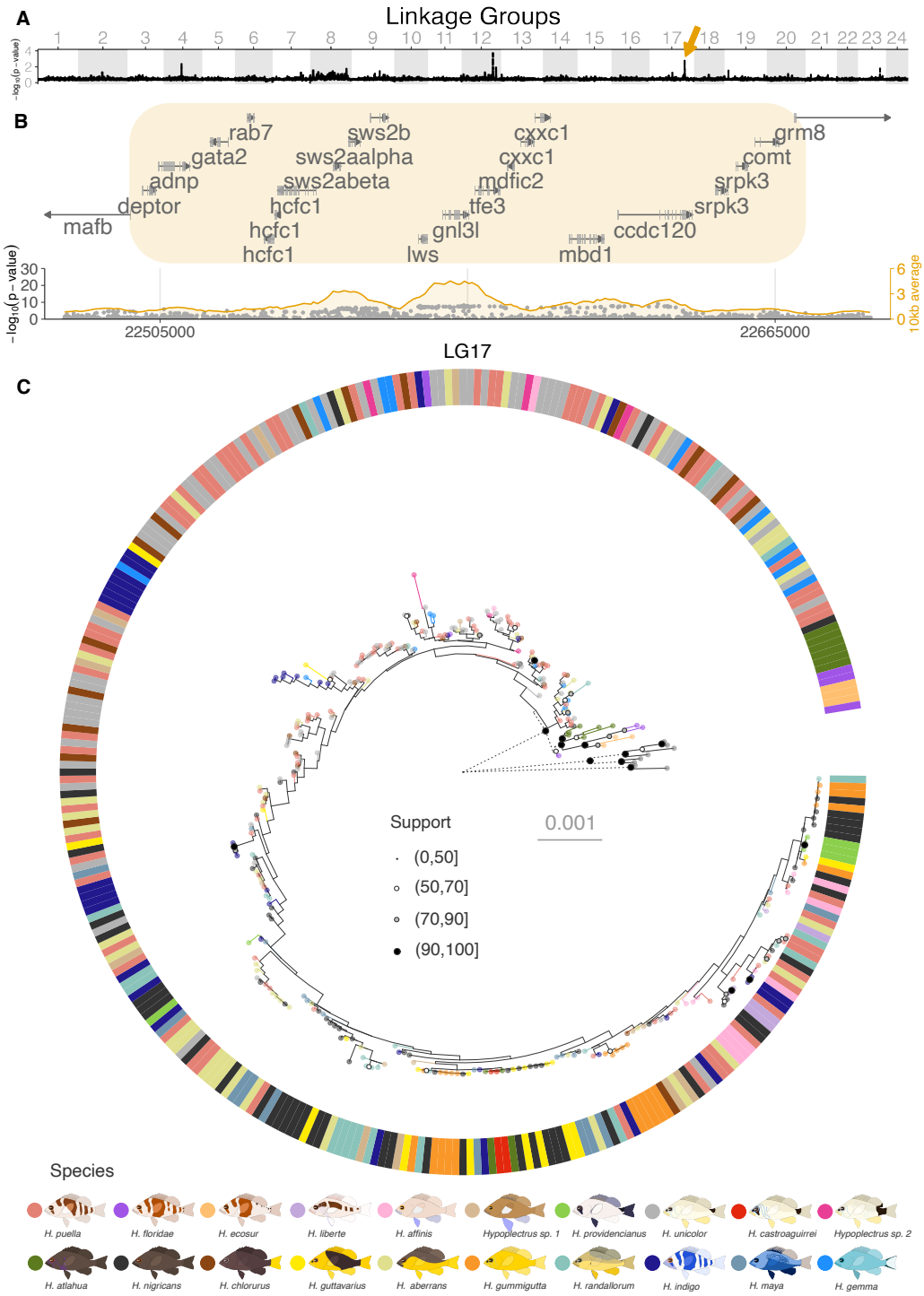



**Figure S4.16 — GWAS for species identity and phylogeny of the LG17 association peak.** (**A**) Results of the genome-wide association study (GWAS) for species identity. The gray and white blocks represent the 24 linkage groups (LG, putative chromosomes) and the black dots are averages of the log-transformed *p*-value of the GWAS association over 50 kb windows with 5 kb increments. (**B**) Close-up on the LG17 association peak. The x-axis shows the position on LG17 (in bp) and the orange rectangle highlights the region that shows an association >1.5 considering 50 kb windows. Grey dots represent the GWAS log-transformed *p* value for each individual SNP, and the orange line its average over 10 kb windows. (**C**) Phylogeny of the LG17 region, covering 146,194 bp centered on the GWAS association peak highlighted in orange in **B**. The outer ring, branch and tip point colors indicate hamlet species as shown by the icons below the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation.

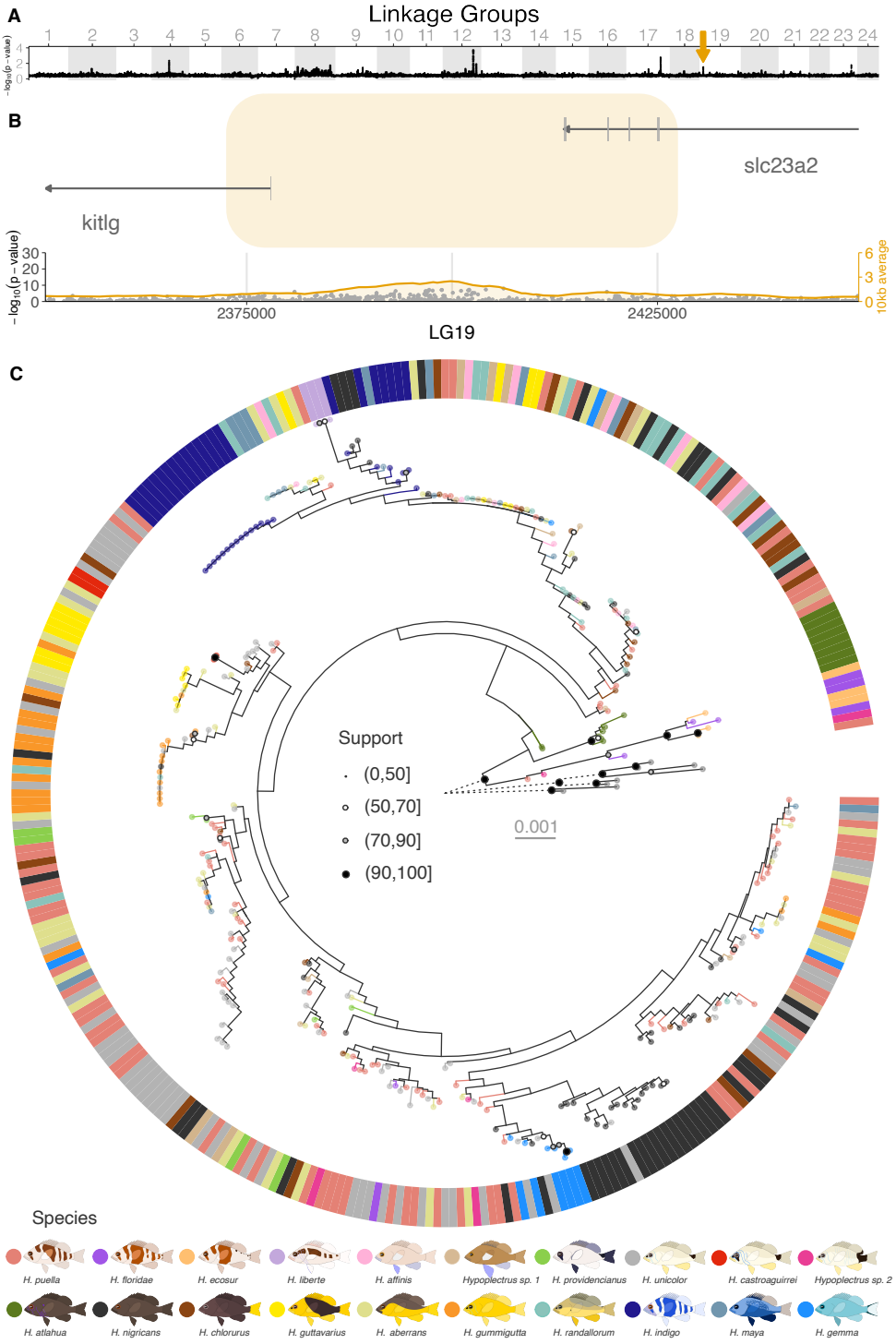



**Figure S4.17 — GWAS for species identity and phylogeny of the LG19 association peak.** (**A**) Results of the genome-wide association study (GWAS) for species identity. The gray and white blocks represent the 24 linkage groups (LG, putative chromosomes) and the black dots are averages of the log-transformed *p*-value of the GWAS association over 50 kb windows with 5 kb increments. (**B**) Close-up on the LG19 association peak. The x-axis shows the position on LG19 (in bp) and the orange rectangle highlights the region that shows an association >1.5 considering 50 kb windows. Grey dots represent the GWAS log-transformed *p* value for each individual SNP, and the orange line its average over 10 kb windows. (**C**) Phylogeny of the LG19 region, covering 44,506 bp centered on the GWAS association peak highlighted in orange in **B**. The outer ring, branch and tip point colors indicate hamlet species as shown by the icons below the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation.

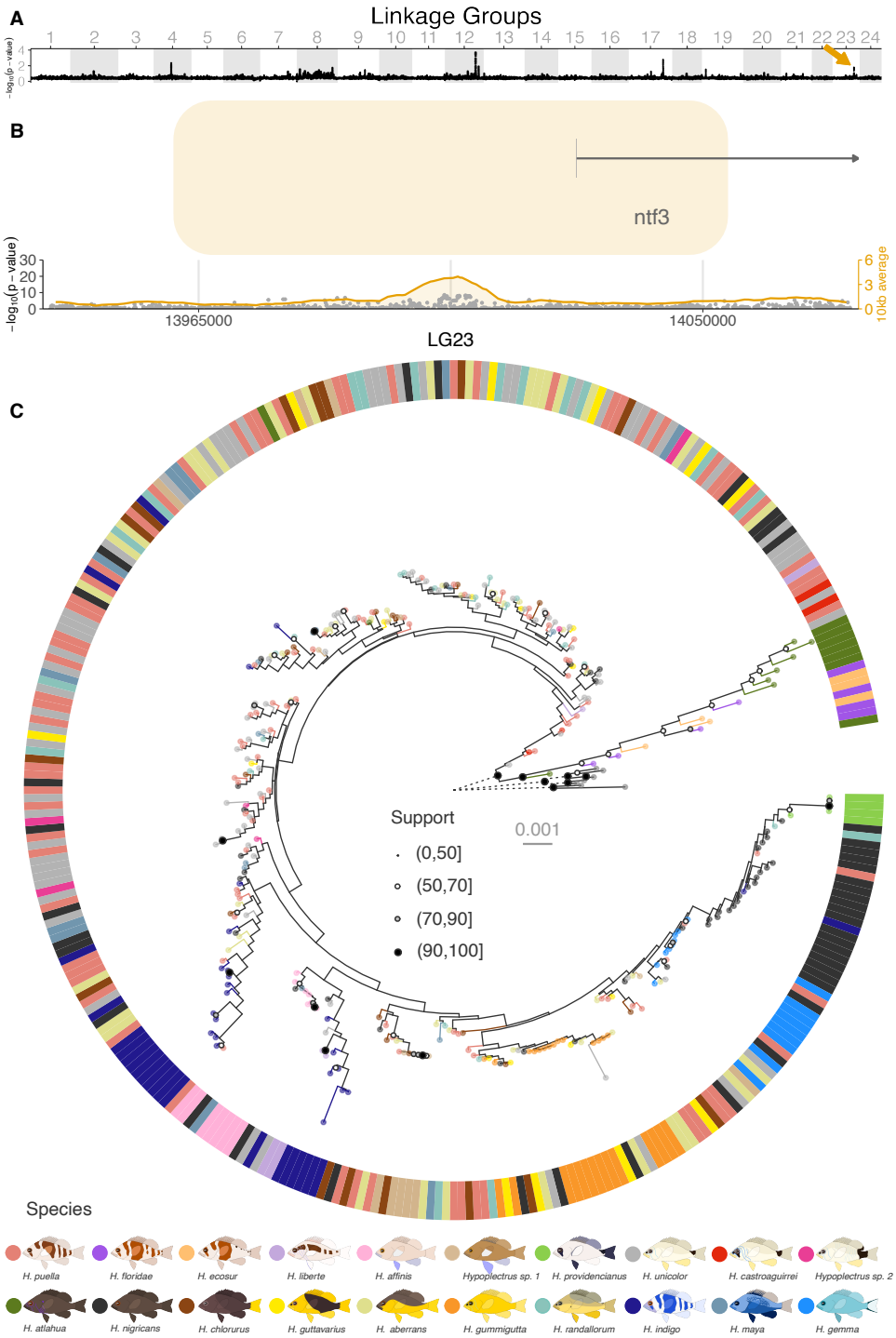



**Figure S4.18 — GWAS for species identity and phylogeny of the LG23 association peak.** (**A**) Results of the genome-wide association study (GWAS) for species identity. The gray and white blocks represent the 24 linkage groups (LG, putative chromosomes) and the black dots are averages of the log-transformed *p*-value of the GWAS association over 50 kb windows with 5 kb increments. (**B**) Close-up on the LG23 association peak. The x-axis shows the position on LG23 (in bp) and the orange rectangle highlights the region that shows an association >1.5 considering 50 kb windows. Grey dots represent the GWAS log-transformed *p* value for each individual SNP, and the orange line its average over 10 kb windows. (**C**) Phylogeny of the LG23 region, covering 75,943 bp centered on the GWAS association peak highlighted in orange in **B**. The outer ring, branch and tip point colors indicate hamlet species as shown by the icons below the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 5% of their actual length to improve the visualization of the hamlet radiation.

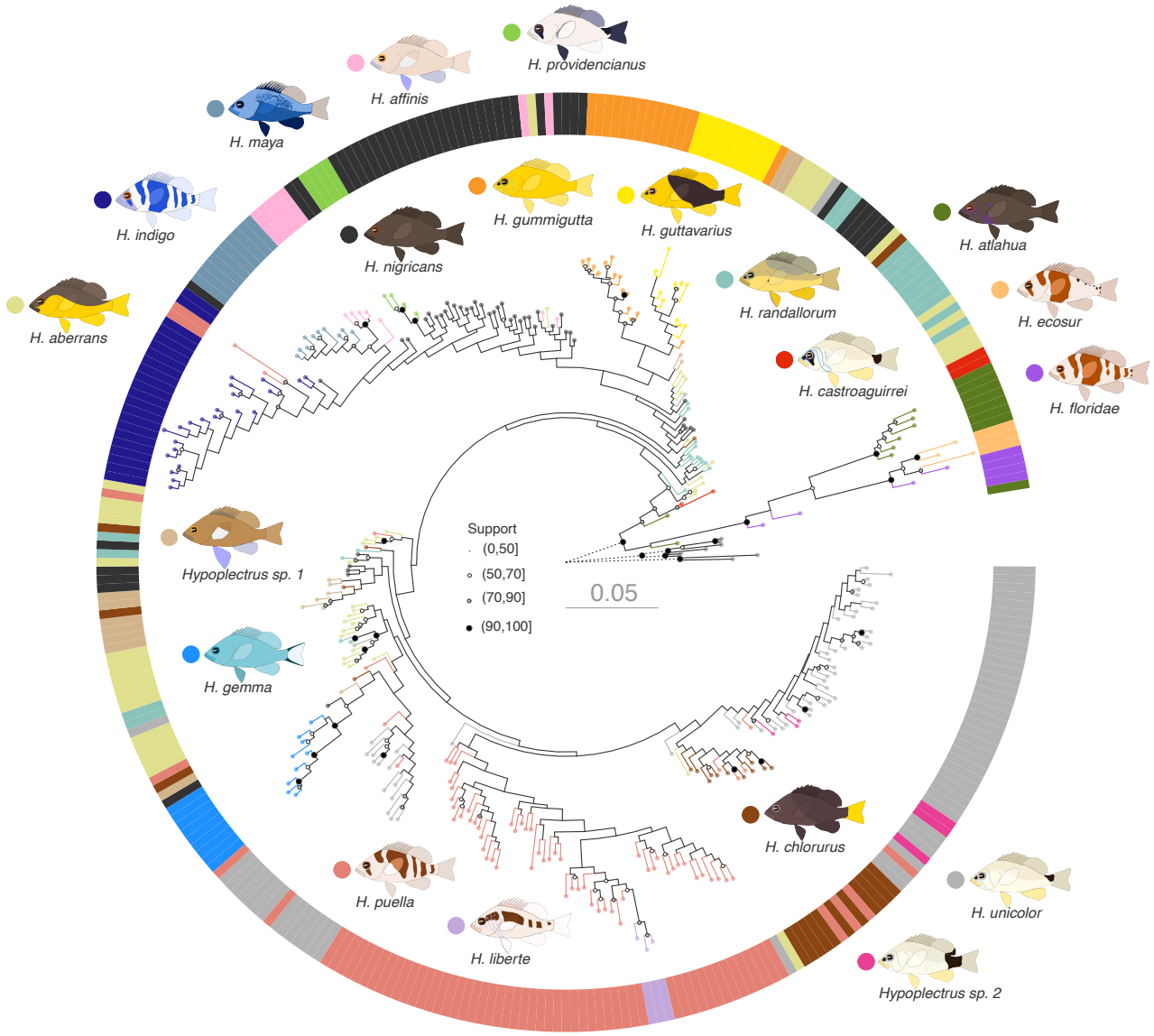



**Figure S4.19 — Phylogeny of the eight concatenated regions with high association to species identity.** These are the eight genomic regions with an association -log(pvalue) >1.5 considering 50 kb windows (Figures S4.12-S4.18), covering a total of 762,681 bp. The outer ring, branch and tip point colors indicate hamlet species as shown by the icons around the tree. Node point size and color represent node support. The eight *Serranus* outgroup individuals are shown in gray and the length of their branches, represented with dashed lines, are reduced to 0.5% of their actual length to improve the visualization of the hamlet radiation.

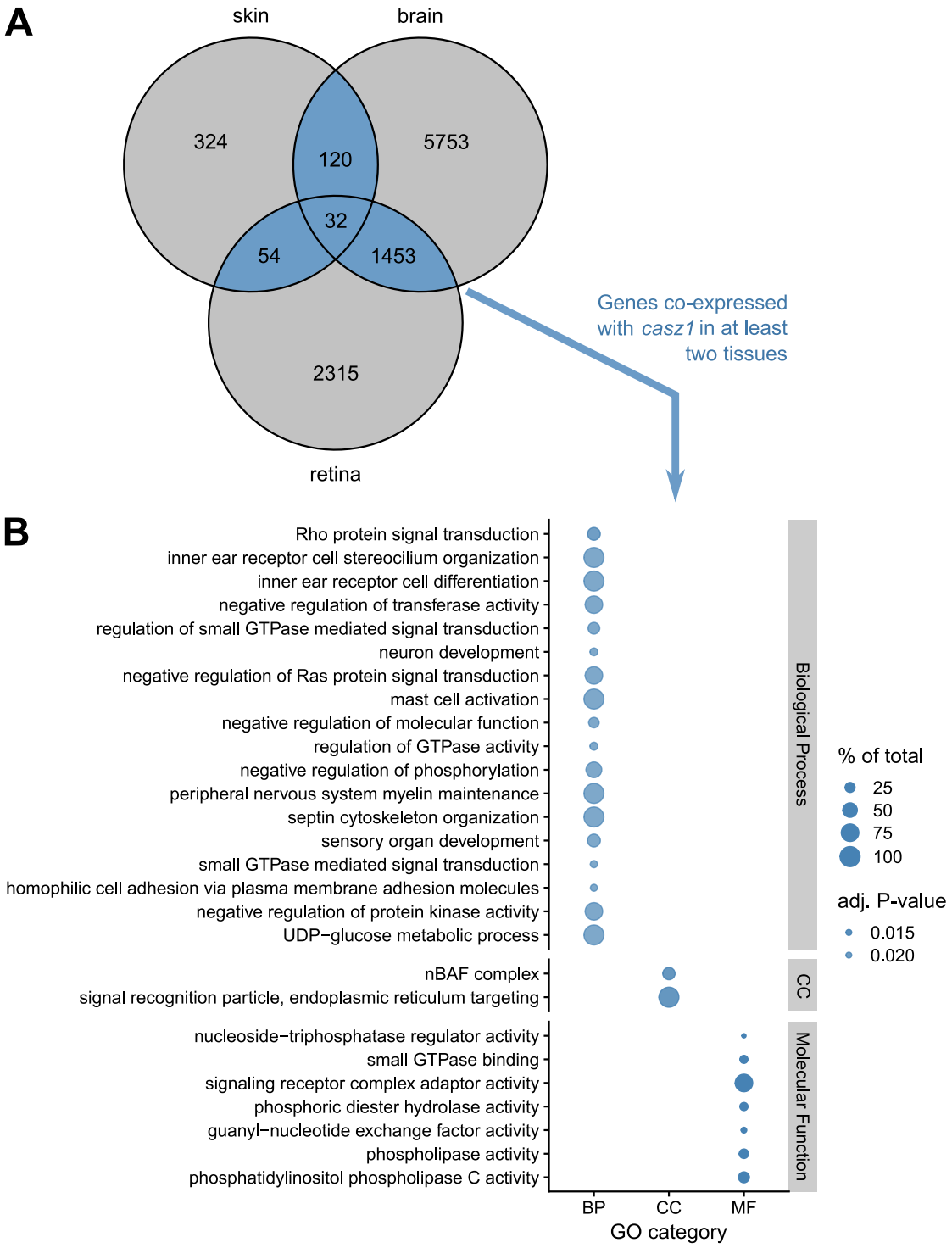



**Figure S4.20 — Genes co-expressed with *casz1*.** (**A**) Venn diagram showing the number of genes co-expressed with *casz1* in brain, retina and skin tissue. (**B**) Top 10% enriched Gene Ontology (GO) terms for genes that are co-expressed with *casz1* in at least two tissues, ordered by increasing adjusted P-value per GO category. The circle size indicates the percentage of co-expressed genes relative to the number of genes that are annotated with this term. The GO categories are biological processes (BP), cellular components (CC) and molecular functions (MF).

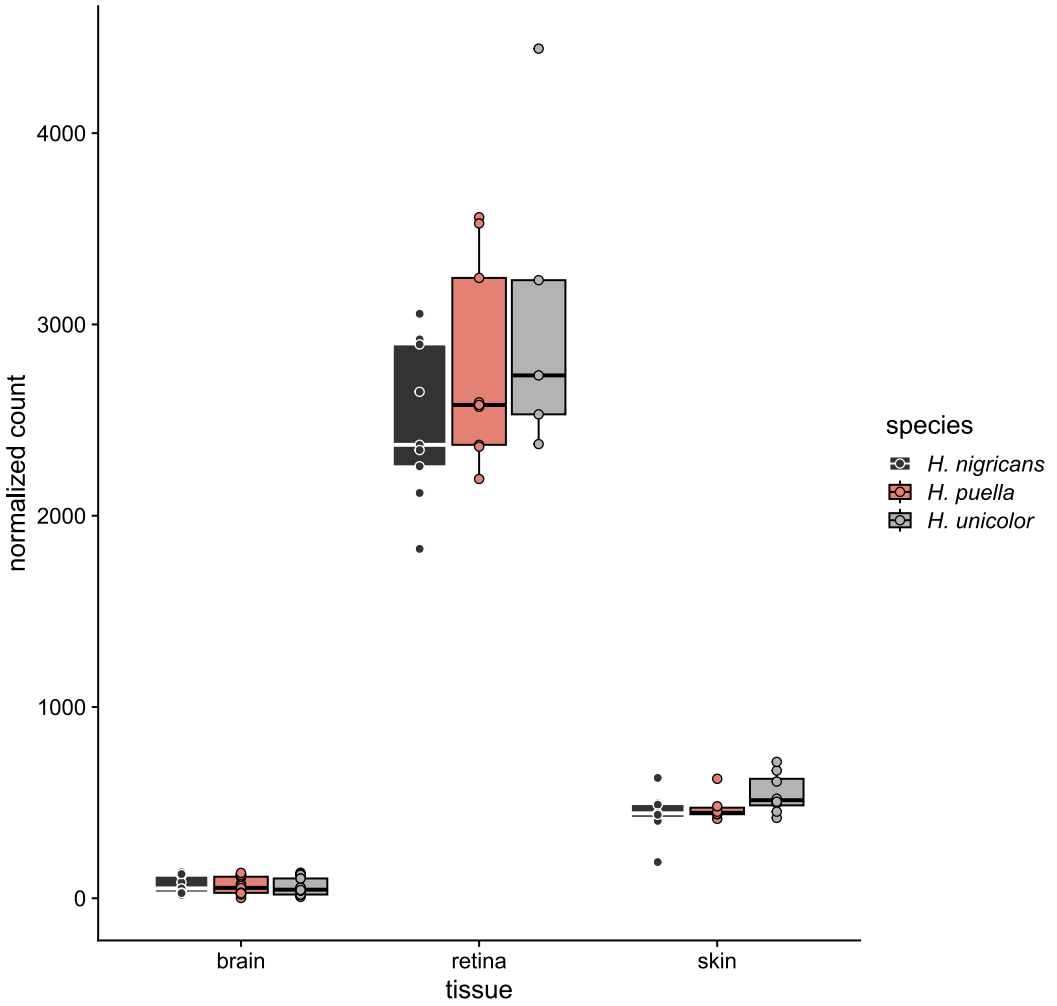



**Figure S4.21 — Normalized expression of *casz1* across tissues and species.** Boxplots display the median (center line), the 25th and 75th percentiles (box limits), the largest value within 1.5× the interquartile range (whiskers) and outliers (points) with sample points plotted on top. The sample sizes are: seven skin samples per species; 18, 18 and 17 brain samples from *H. puella*, *H. unicolor* and *H. nigricans*, respectively; and 9, 5 and 9 retinal samples from *H. puella*, *H. unicolor* and *H. nigricans*, respectively.

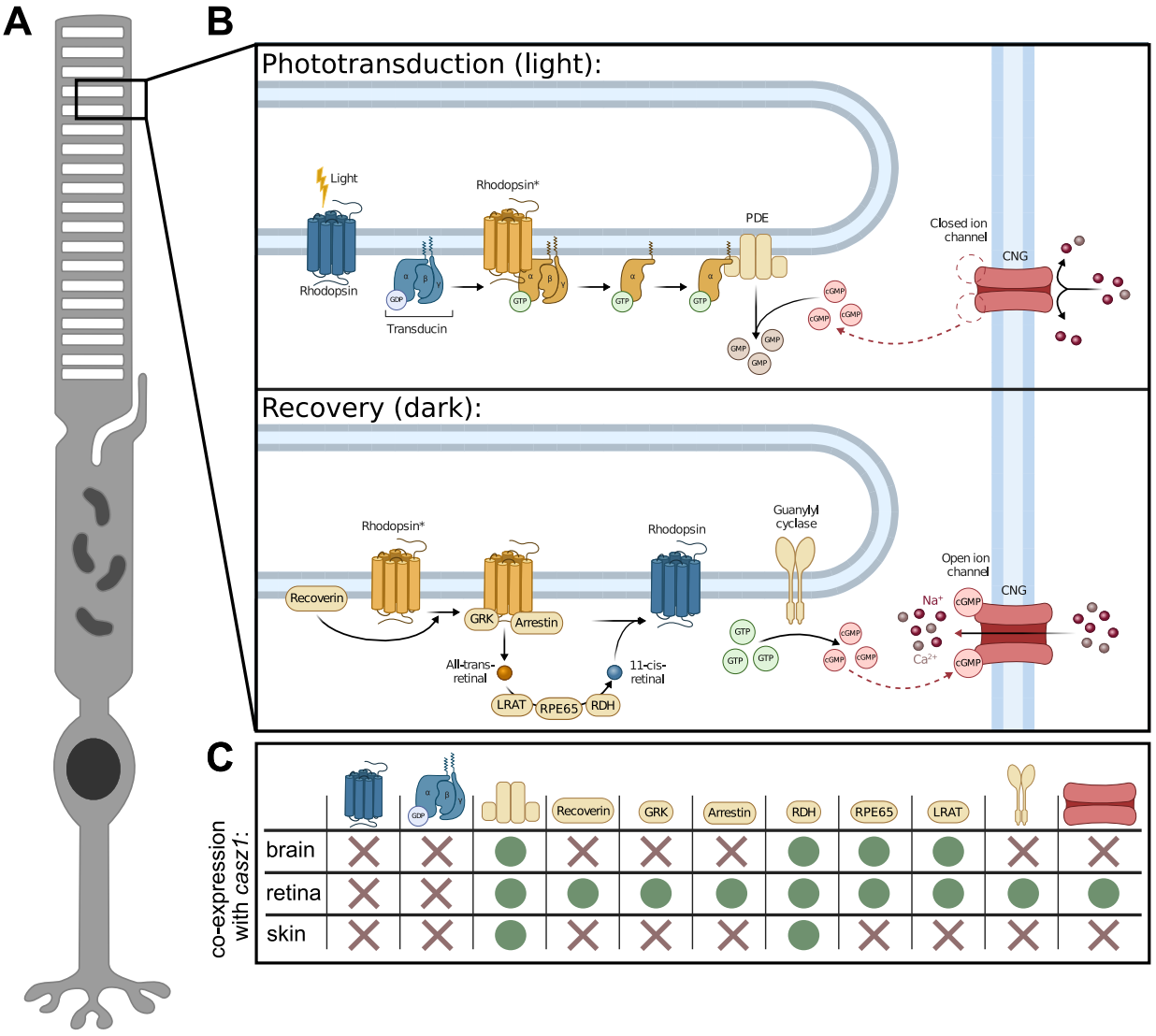



**Figure S4.22 — Phototransduction cascade.** (**A**) Illustration of a rod cell in the retinal tissue. (**B**) Genes involved in the phototransduction cascade during light and recovery phase. (**C**) Genes that are co-expressed with *casz1* across the three different tissues are highlighted with a green dot.

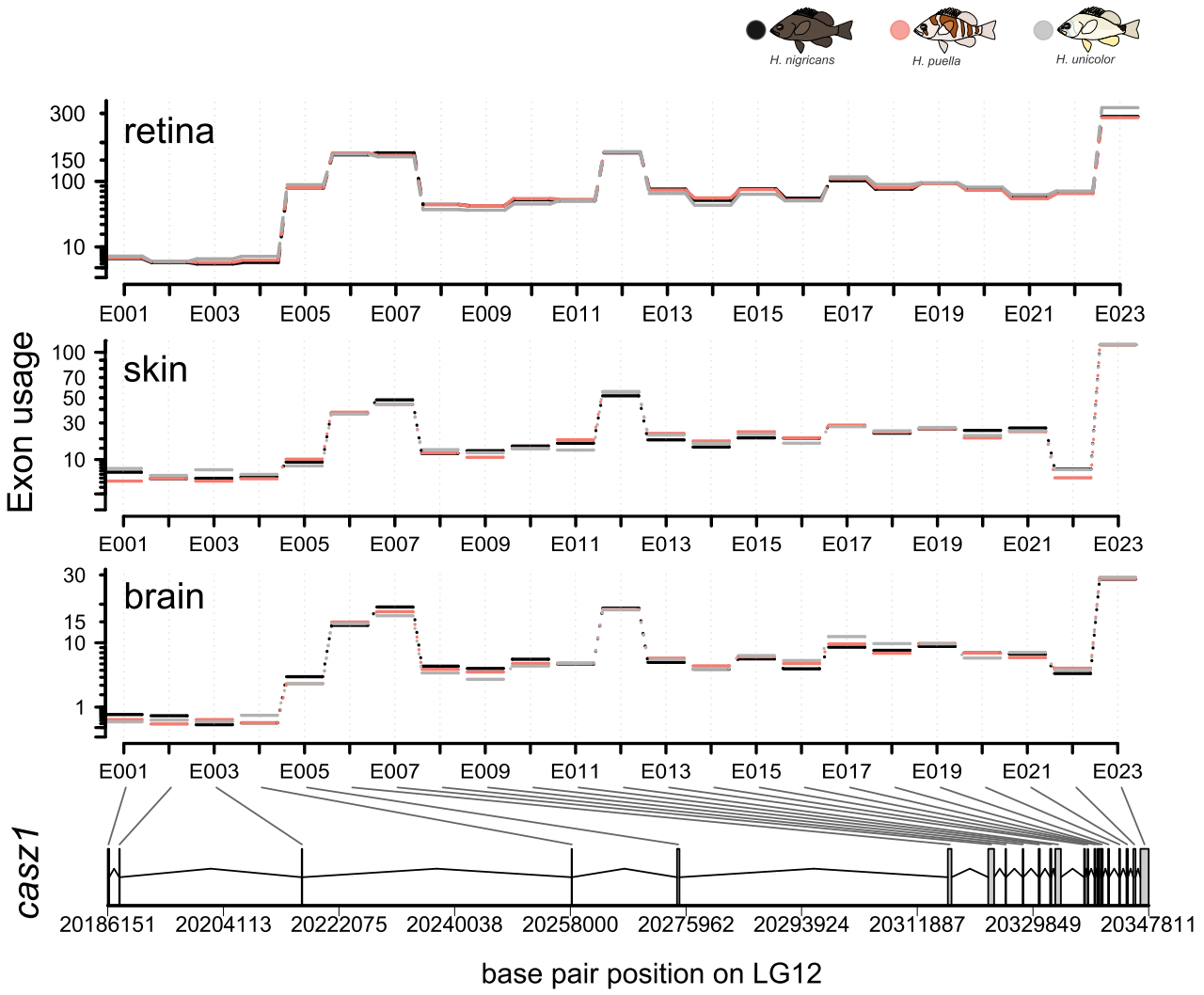



**Figure S4.23 — *casz1* exon usage.** The three plots show the normalized expression of *casz1* exons relative to the gene mean.

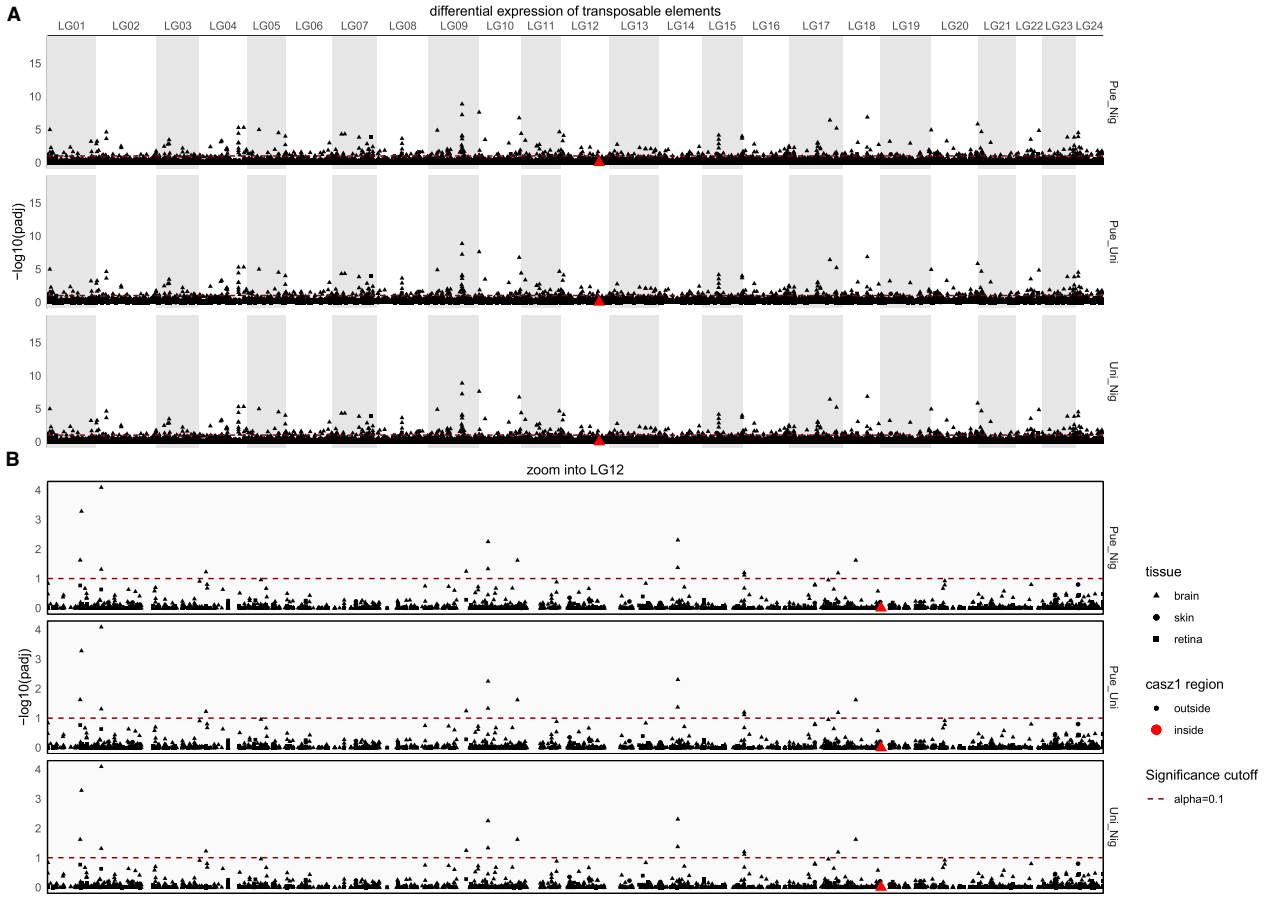



**Figure S4.24 — Differential expression of transposable elements between species**. Pue: *H. puella*, Nig: *H. nigricans*, Uni: *H. unicolor*). Adjusted log transformed P-values were plotted along the genome (**A**) and for LG12 (**B**). The *casz1* region is highlighted in red.

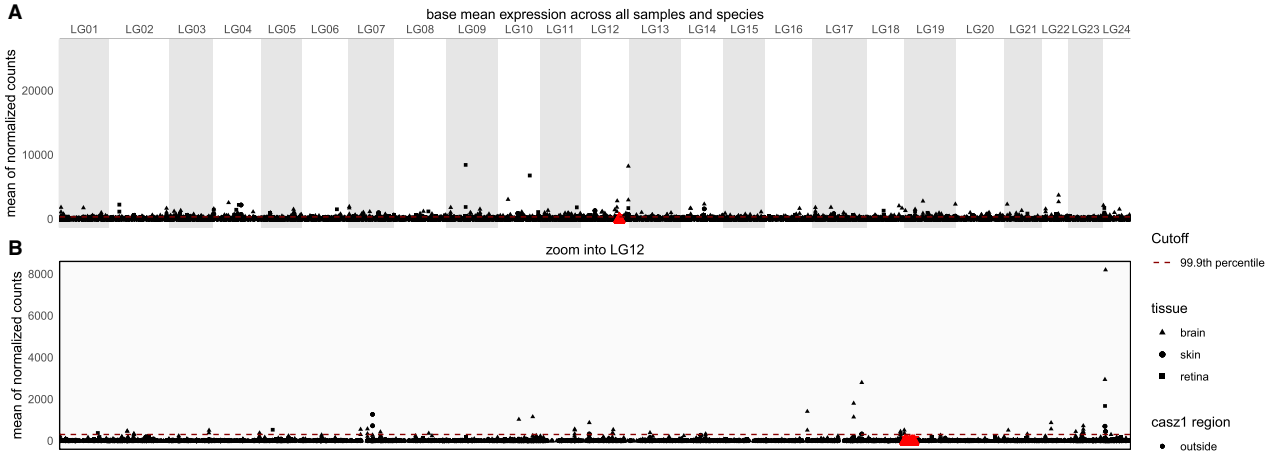



**Figure S4.25 — Normalized expression of transposable elements.** (**A**) Along the genome and (**B**) in LG12 across species (*H. puella, H. nigricans, H. unicolor*). High expression values indicate mobilized transposable elements. The *casz1* region is highlighted in red.

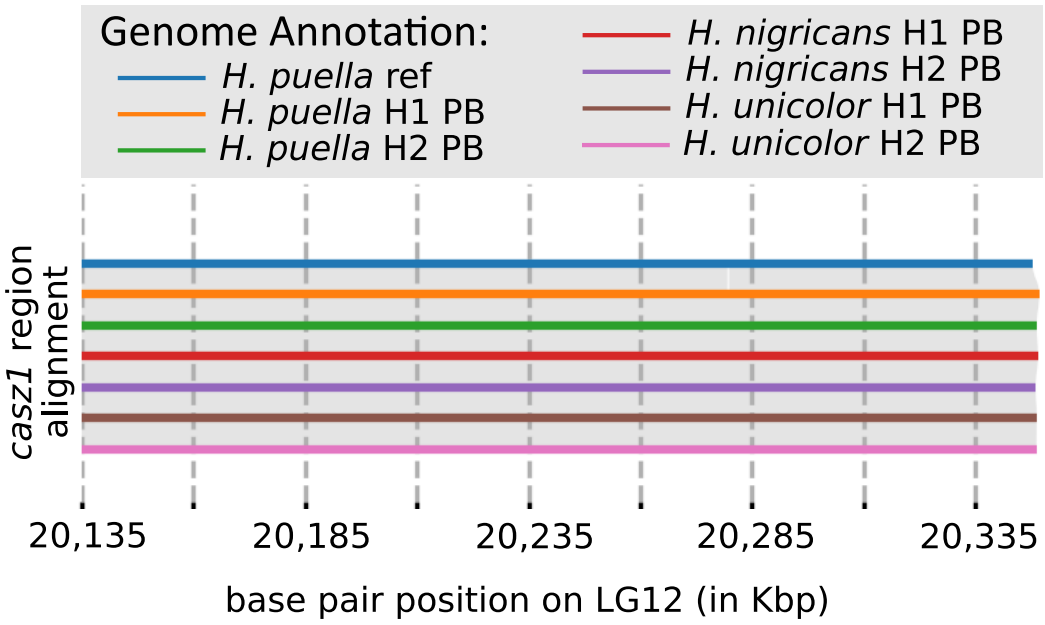



**Figure S4.26 — Structural variation in the *casz1* region.** Each line represents a sequence from six PacBio genomes (two haplotypes per genome, H1 PB and H2 PB) and the *H. puella* reference genome. The base pair position on the y-axis is relative to the reference genome assembly (blue). The grey areas between the sequences (colored lines) indicate that they align without major gaps or rearrangements. A Fasta file with the *casz1* region alignment can be found in the GitHub repository (github.com/mhelmkampf/hamlet_phylogeny).

## 4.6 Supplementary Tables

**Table S4.1 — List of resequenced genomes used in this study**, including 327 hamlet and eight outgroup samples. Date, latitude and longitude refer to collection dates and coordinates. Accession gives the sample record in the European Nucleotide Archive (ENA) for each sample, with asterisks indicating samples for which new data has been generated and submitted to ENA (note that data from previous sequencing runs may be attached to some samples, but only data from the latest runs were used in this study). Col.: Colombia, Mex.: Mexico, Pan.: Panama.

| Sample ID | Species | Location | Date | Latitude | Longitude | Accession |
|---|---|---|---|---|---|---|
| 13310 | *H. puella* | Quintana, Mex. | 2007-01-08 | 20.45344 | -87.26467 | ERS20924417* |
| 13318 | *H. unicolor* | Quintana, Mex. | 2007-01-08 | 20.45633 | -87.26217 | ERS20924418* |
| 17996 | *H. indigo* | Belize | 2004-07-27 | 16.80083 | -88.07889 | ERS4141229 |
| 17997 | *H. indigo* | Belize | 2004-07-27 | 16.80083 | -88.07889 | ERS4141230 |
| 17998 | *H. indigo* | Belize | 2004-07-27 | 16.80083 | -88.07889 | ERS4141231 |
| 17999 | *H. indigo* | Belize | 2004-07-27 | 16.80083 | -88.07889 | ERS4141232 |
| 18000 | *H. indigo* | Belize | 2004-07-27 | 16.80083 | -88.07889 | ERS4141233 |
| 18151 | *H. nigricans* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619600 |
| 18152 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619637 |
| 18153 | *H. nigricans* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619601 |
| 18154 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619638 |
| 18155 | *H. nigricans* | Belize | 2004-07-25 | 16.80083 | -88.07889 | ERS2619602 |
| 18156 | *H. nigricans* | Belize | 2004-07-25 | 16.80083 | -88.07889 | ERS2619603 |
| 18157 | *H. nigricans* | Belize | 2004-07-25 | 16.80083 | -88.07889 | ERS2619604 |
| 18158 | *H. nigricans* | Belize | 2004-07-25 | 16.80083 | -88.07889 | ERS2619605 |
| 18159 | *H. nigricans* | Belize | 2004-07-25 | 16.80083 | -88.07889 | ERS2619606 |
| 18160 | *H. guttavarius* | Belize | 2004-07-27 | 16.78389 | -88.07667 | ERS20924419* |
| 18161 | *H. puella* | Belize | 2004-07-26 | 16.80583 | -88.07917 | ERS2619639 |
| 18162 | *H. nigricans* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619607 |
| 18163 | *H. unicolor* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619674 |
| 18165 | *H. nigricans* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619608 |
| 18166 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619640 |
| 18169 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619641 |
| 18171 | *H. nigricans* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619609 |
| 18172 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619642 |
| 18174 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619643 |
| 18175 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619644 |
| 18176 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619645 |
| 18178 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619646 |
| 18179 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619647 |
| 18180 | *H. puella* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619648 |
| 18185 | *H. nigricans* | Belize | 2004-07-26 | 16.80583 | -88.07917 | ERS2619610 |
| 18187 | *H. nigricans* | Belize | 2004-07-26 | 16.80583 | -88.07917 | ERS2619611 |
| 18195 | *H. indigo* | Belize | 2004-07-26 | 16.80583 | -88.07917 | ERS4141234 |
| 18222 | *H. indigo* | Belize | 2004-07-27 | 16.80083 | -88.07889 | ERS4141235 |
| 18225 | *H. indigo* | Belize | 2004-07-27 | 16.80083 | -88.07889 | ERS4141236 |
| 18226 | *H. indigo* | Belize | 2004-07-27 | 16.78389 | -88.07667 | ERS4141237 |
| 18227 | *H. indigo* | Belize | 2004-07-27 | 16.78389 | -88.07667 | ERS4141238 |
| 18261 | *H. unicolor* | Belize | 2004-07-24 | NA | NA | ERS2619675 |
| 18263 | *H. aberrans* | Belize | 2004-07-24 | NA | NA | ERS20924420* |
| 18267 | *H. unicolor* | Belize | 2004-07-24 | NA | NA | ERS2619676 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 18274 | *H. unicolor* | Belize | 2004-07-24 | NA | NA | ERS2619677 |
| 18276 | *H. unicolor* | Belize | 2004-07-25 | 16.76528 | -88.14417 | ERS2619678 |
| 18418 | *H. nigricans* | Bocas del Toro, Pan. | 2004-05-12 | 9.37750 | -82.30389 | ERS2619625 |
| 18419 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619661 |
| 18420 | *H. unicolor* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619698 |
| 18421 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619662 |
| 18422 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 9.37750 | -82.30389 | ERS2619663 |
| 18424 | *H. nigricans* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619626 |
| 18426 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619664 |
| 18427 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619665 |
| 18428 | *H. nigricans* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619627 |
| 18429 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619666 |
| 18430 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619667 |
| 18432 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 10.23917 | -83.17306 | ERS2619668 |
| 18434 | *H. puella* | Bocas del Toro, Pan. | 2004-05-12 | 9.37750 | -82.30389 | ERS2619669 |
| 18435 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-23 | 9.33278 | -82.25472 | ERS2619699 |
| 18439 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-25 | 9.33278 | -82.25472 | ERS2619700 |
| 18440 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-25 | 9.33278 | -82.25472 | ERS2619701 |
| 18441 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-25 | 9.33278 | -82.25472 | ERS2619702 |
| 18442 | *H. unicolor* | Bocas del Toro, Pan. | 2004-07-08 | 9.29833 | -82.28944 | ERS2619703 |
| 18445 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-28 | 9.33278 | -82.25472 | ERS2619704 |
| 18446 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-29 | 9.33278 | -82.25472 | ERS2619705 |
| 18447 | *H. unicolor* | Bocas del Toro, Pan. | 2004-07-09 | 9.28944 | -82.25889 | ERS2619706 |
| 18448 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-28 | 9.33278 | -82.25472 | ERS2619707 |
| 18449 | *H. gummigutta* | Bocas del Toro, Pan. | 2004-07-03 | 9.37750 | -82.30389 | ERS2619710 |
| 18450 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-25 | 9.33278 | -82.25472 | ERS2619708 |
| 18452 | *H. aberrans* | Bocas del Toro, Pan. | 2004-07-02 | 9.33083 | -82.21444 | ERS20924421* |
| 18454 | *H. unicolor* | Bocas del Toro, Pan. | 2004-06-30 | 9.34806 | -82.26333 | ERS2619709 |
| 18471 | *H. aberrans* | Bocas del Toro, Pan. | 2004-07-10 | 9.37750 | -82.30389 | ERS20924422* |
| 18492 | *H. aberrans* | Belize | 2004-07-24 | 16.89389 | -88.05806 | ERS20924423* |
| 18493 | *H. aberrans* | Belize | 2004-07-24 | NA | NA | ERS20924424* |
| 18901 | *H. nigricans* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619629 |
| 18902 | *H. nigricans* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619630 |
| 18903 | *H. nigricans* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619631 |
| 18904 | *H. nigricans* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619632 |
| 18905 | *H. nigricans* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619633 |
| 18906 | *H. nigricans* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619634 |
| 18907 | *H. nigricans* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619635 |
| 18909 | *H. nigricans* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619636 |
| 18912 | *H. puella* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619670 |
| 18915 | *H. puella* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619671 |
| 18917 | *H. puella* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS2619672 |
| 18935 | *H. affinis* | Bocas del Toro, Pan. | 2005-03-27 | 9.25081 | -82.13061 | ERS20924425* |
| 18936 | *H. affinis* | Bocas del Toro, Pan. | 2005-03-27 | 9.25081 | -82.13061 | ERS20924426* |
| 18949 | *H. aberrans* | Guna Yala, Pan. | 2005-05-04 | 9.54875 | -78.95061 | ERS20924427* |
| 18986 | *H. affinis* | Bocas del Toro, Pan. | 2005-03-25 | 9.29833 | -82.28944 | ERS8632037 |
| 19076 | *H. guttavarius* | Barbados | 2005-06-09 | 13.22619 | -59.65286 | ERS20924428* |
| 19077 | *H. sp1* | Barbados | 2005-06-09 | 13.22619 | -59.65286 | ERS20924429* |
| 19079 | *H. sp1* | Barbados | 2005-06-10 | 13.13375 | -59.64072 | ERS20924430* |

| | | | | | | |
|---|---|---|---|---|---|---|
| 19080 | *H. guttavarius* | Barbados | 2005-06-12 | 13.17264 | -59.64675 | ERS20924431* |
| 19104 | *H. puella* | Guna Yala, Pan. | 2005-05-08 | 9.57592 | -78.72794 | ERS20924432* |
| 19105 | *H. nigricans* | Guna Yala, Pan. | 2005-05-08 | 9.57592 | -78.72794 | ERS20924433* |
| 19106 | *H. puella* | Guna Yala, Pan. | 2005-05-08 | 9.57592 | -78.72794 | ERS20924434* |
| 19108 | *H. puella* | Guna Yala, Pan. | 2005-05-08 | 9.57592 | -78.72794 | ERS20924435* |
| 19124 | *H. indigo* | Guna Yala, Pan. | 2005-05-08 | 9.57592 | -78.72794 | ERS20924436* |
| 19131 | *H. nigricans* | Guna Yala, Pan. | 2005-05-09 | 9.47431 | -78.52164 | ERS20924437* |
| 19150 | *H. nigricans* | Guna Yala, Pan. | 2005-05-09 | 9.47431 | -78.52164 | ERS20924438* |
| 19162 | *H. indigo* | Guna Yala, Pan. | 2005-05-07 | 9.58019 | -78.69406 | ERS20924439* |
| 19174 | *H. affinis* | Guna Yala, Pan. | 2005-05-07 | 9.58019 | -78.69406 | ERS20924440* |
| 19190 | *H. indigo* | Guna Yala, Pan. | 2005-05-08 | 9.57592 | -78.72794 | ERS20924441* |
| 19235 | *H. guttavarius* | Guna Yala, Pan. | 2005-05-15 | 9.19967 | -77.96972 | ERS20924442* |
| 19294 | *H. affinis* | Guna Yala, Pan. | 2005-05-22 | 9.57647 | -78.72912 | ERS20924443* |
| 19416 | *H. chlorurus* | Barbados | 2005-06-06 | 13.22369 | -59.64506 | ERS20924444* |
| 19425 | *H. chlorurus* | Barbados | 2005-06-06 | 13.22369 | -59.64506 | ERS20924445* |
| 19426 | *H. puella* | Barbados | 2005-06-06 | 13.22369 | -59.64506 | ERS20924446* |
| 19427 | *H. puella* | Barbados | 2005-06-06 | 13.22369 | -59.64506 | ERS20924447* |
| 19428 | *H. puella* | Barbados | 2005-06-06 | 13.22369 | -59.64506 | ERS20924448* |
| 19435 | *H. chlorurus* | Barbados | 2005-06-06 | 13.22369 | -59.64506 | ERS20924449* |
| 19437 | *H. sp1* | Barbados | 2005-06-06 | 13.22369 | -59.64506 | ERS20924450* |
| 19519 | *H. affinis* | Guna Yala, Pan. | 2005-05-14 | 9.29072 | -78.14039 | ERS20924451* |
| 19546 | *H. aberrans* | Guna Yala, Pan. | 2005-05-14 | 9.29072 | -78.14039 | ERS20924452* |
| 19708 | *H. aberrans* | Guna Yala, Pan. | 2005-05-06 | 9.54875 | -78.95061 | ERS20924453* |
| 19881 | *H. unicolor* | Belize | 2005-08-16 | 16.70781 | -87.85981 | ERS2619679 |
| 20068 | *H. guttavarius* | Belize | 2005-08-13 | NA | NA | ERS20924454* |
| 20069 | *H. guttavarius* | Belize | 2005-08-13 | NA | NA | ERS20924455* |
| 20092 | *H. unicolor* | Belize | 2005-08-15 | 16.89364 | -88.12256 | ERS2619680 |
| 20120 | *H. unicolor* | Belize | 2005-08-11 | 16.80083 | -88.07889 | ERS2619681 |
| 20126 | *H. unicolor* | Belize | 2005-08-12 | 16.89364 | -88.12256 | ERS2619682 |
| 20128 | *H. unicolor* | Belize | 2005-08-12 | 16.89364 | -88.12256 | ERS2619683 |
| 20135 | *H. unicolor* | Belize | 2005-08-12 | 16.89364 | -88.12256 | ERS2619684 |
| 20149 | *H. unicolor* | Belize | 2005-08-12 | 16.89364 | -88.12256 | ERS2619685 |
| 20418 | *H. gummigutta* | Honduras | 2006-06-03 | 16.03000 | -83.32861 | ERS4141241 |
| 20419 | *H. gummigutta* | Honduras | 2006-06-03 | 16.03000 | -83.32861 | ERS4141242 |
| 20420 | *H. gummigutta* | Honduras | 2006-06-03 | 16.03000 | -83.32861 | ERS4141243 |
| 20421 | *H. randallorum* | Honduras | 2006-06-03 | 16.03000 | -83.32861 | ERS4141244 |
| 20425 | *H. aberrans* | Honduras | 2006-06-03 | 16.03000 | -83.32861 | ERS4141245 |
| 20426 | *H. gummigutta* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS4141246 |
| 20427 | *H. gummigutta* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS4141247 |
| 20428 | *H. gummigutta* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS4141248 |
| 20429 | *H. randallorum* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS4141249 |
| 20430 | *H. randallorum* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS4141250 |
| 20433 | *H. aberrans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS4141251 |
| 20435 | *H. chlorurus* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS20924456* |
| 20478 | *S. tabacarius* | Honduras | 2006-06-06 | 15.25000 | -82.61700 | ERS4141252 |
| 20480 | *S. tabacarius* | Honduras | 2006-06-07 | 15.25000 | -82.61700 | ERS20924457* |
| 20481 | *S. tigrinus* | Honduras | 2006-06-07 | 15.25000 | -82.61700 | ERS20924458* |
| 20551 | *H. puella* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619649 |
| 20552 | *H. puella* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619650 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 20553 | *H. puella* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619651 |
| 20554 | *H. puella* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619652 |
| 20555 | *H. puella* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619653 |
| 20556 | *H. puella* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619654 |
| 20558 | *H. puella* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619655 |
| 20559 | *H. puella* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619656 |
| 20560 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619686 |
| 20561 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619687 |
| 20562 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619688 |
| 20563 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619689 |
| 20564 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619690 |
| 20565 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619691 |
| 20566 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619692 |
| 20567 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619693 |
| 20568 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619694 |
| 20571 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619695 |
| 20572 | *H. unicolor* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619696 |
| 20599 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619612 |
| 20600 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619613 |
| 20601 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619614 |
| 20602 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619615 |
| 20603 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619616 |
| 20604 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619617 |
| 20605 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619618 |
| 20606 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619619 |
| 20607 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619620 |
| 20608 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619621 |
| 20609 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619622 |
| 20610 | *H. nigricans* | Honduras | 2006-06-04 | 15.95583 | -83.29306 | ERS2619623 |
| 20613 | *H. randallorum* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS4141253 |
| 20615 | *H. gummigutta* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS4141254 |
| 20617 | *H. gummigutta* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS4141255 |
| 20625 | *H. puella* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS2619657 |
| 20633 | *H. puella* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS2619658 |
| 20635 | *H. puella* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS2619659 |
| 20638 | *H. puella* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS2619660 |
| 20641 | *H. gummigutta* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS4141256 |
| 20642 | *H. gummigutta* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS4141257 |
| 20643 | *H. gummigutta* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS4141258 |
| 20644 | *H. aberrans* | Honduras | 2006-06-05 | 15.95583 | -83.29306 | ERS4141259 |
| 20650 | *H. providencianus* | Honduras | 2006-06-06 | 15.25000 | -82.61667 | ERS20924459* |
| 20696 | *H. randallorum* | Honduras | 2006-06-12 | 16.11028 | -86.95389 | ERS4141260 |
| 20751 | *H. guttavarius* | Honduras | 2006-06-06 | 15.25000 | -82.61667 | ERS20924460* |
| 20757 | *H. indigo* | Honduras | 2006-06-06 | 15.25000 | -82.61667 | ERS20924461* |
| 20759 | *H. aberrans* | Honduras | 2006-06-06 | 15.25000 | -82.61667 | ERS4141261 |
| 20761 | *H. aberrans* | Honduras | 2006-06-06 | 15.25000 | -82.61667 | ERS4141262 |
| 20762 | *H. aberrans* | Honduras | 2006-06-06 | 15.25000 | -82.61667 | ERS4141263 |
| 20845 | *H. providencianus* | Honduras | 2006-06-07 | 15.25000 | -82.61667 | ERS20924462* |
| 20846 | *H. providencianus* | Honduras | 2006-06-07 | 15.25000 | -82.61667 | ERS20924463* |

| | | | | | | |
|---|---|---|---|---|---|---|
| 20861 | *H. aberrans* | Honduras | 2006-06-07 | 15.25000 | -82.61667 | ERS4141264 |
| 20862 | *H. aberrans* | Honduras | 2006-06-07 | 15.25000 | -82.61667 | ERS4141265 |
| 20864 | *H. aberrans* | Honduras | 2006-06-07 | 15.25000 | -82.61667 | ERS4141266 |
| 20866 | *H. aberrans* | Honduras | 2006-06-07 | 15.25000 | -82.61667 | ERS4141267 |
| 20867 | *H. aberrans* | Honduras | 2006-06-07 | 15.25000 | -82.61667 | ERS4141268 |
| 20892 | *H. randallorum* | Honduras | 2006-06-08 | 16.44500 | -85.87500 | ERS4141269 |
| 20893 | *H. randallorum* | Honduras | 2006-06-08 | 16.44500 | -85.87500 | ERS4141270 |
| 20894 | *H. randallorum* | Honduras | 2006-06-08 | 16.44500 | -85.87500 | ERS4141271 |
| 20896 | *H. randallorum* | Honduras | 2006-06-08 | 16.44500 | -85.87500 | ERS4141272 |
| 20922 | *H. randallorum* | Honduras | 2006-06-09 | 16.47361 | -85.92389 | ERS4141273 |
| 20923 | *H. randallorum* | Honduras | 2006-06-09 | 16.47361 | -85.92389 | ERS4141274 |
| 20980 | *H. randallorum* | Honduras | 2006-06-10 | 16.49750 | -85.90278 | ERS4141275 |
| 23301 | *H. gummigutta* | Bocas del Toro, Pan. | 2009-04-19 | NA | NA | ERS20924464* |
| 23318 | *H. guttavarius* | Bocas del Toro, Pan. | 2009-12-09 | NA | NA | ERS20924465* |
| 23322 | *H. affinis* | Bocas del Toro, Pan. | 2009-09-17 | NA | NA | ERS20924466* |
| 27698 | *H. aberrans* | Quintana, Mex. | 2010-08-18 | 20.97881 | -86.80011 | ERS20924467* |
| 27707 | *H. unicolor* | Quintana, Mex. | 2010-08-19 | 20.84180 | -86.87781 | ERS20924468* |
| 27709 | *H. unicolor* | Quintana, Mex. | 2010-08-19 | 20.84180 | -86.87781 | ERS20924469* |
| 27936 | *H. atlahua* | Tamiahua, Mex. | 2010-10-02 | 21.47032 | -97.22894 | ERS20924470* |
| 28366 | *H. gummigutta* | Bocas del Toro, Pan. | 2017-02-04 | 9.33222 | -82.22188 | ERS8632035 |
| 28389 | *H. aberrans* | Bocas del Toro, Pan. | 2017-02-06 | 9.31810 | -82.22180 | ERS8632036 |
| 28393 | *S. tortugarum* | Bocas del Toro, Pan. | 2017-02-07 | 9.30140 | -82.29410 | ERS4141276 |
| 28713 | *H. puella* | Quintana, Mex. | 2010-10-15 | 20.97689 | -86.81816 | ERS20924471* |
| 28939 | *H. puella* | Quintana, Mex. | 2010-11-30 | 20.84180 | -86.87781 | ERS20924472* |
| 29137 | *H. nigricans* | Quintana, Mex. | 2010-12-04 | 20.84180 | -86.87781 | ERS20924473* |
| 29212 | *H. indigo* | Quintana, Mex. | 2010-12-04 | 20.86840 | -86.84730 | ERS20924474* |
| 29213 | *H. indigo* | Quintana, Mex. | 2010-12-04 | 20.86840 | -86.84730 | ERS20924475* |
| 30001 | *H. nigricans* | Quintana, Mex. | 2010-12-04 | 20.84180 | -86.87781 | ERS20924476* |
| 33023 | *H. nigricans* | Quintana, Mex. | 2010-12-04 | 20.84993 | -86.87318 | ERS20924477* |
| 52988 | *H. atlahua* | Tamiahua, Mex. | 2013-06-25 | 21.47588 | -97.22700 | ERS20924478* |
| 52989 | *H. atlahua* | Tamiahua, Mex. | 2013-06-25 | 21.47588 | -97.22700 | ERS20924479* |
| 52990 | *H. atlahua* | Tamiahua, Mex. | 2013-06-25 | 21.47588 | -97.22700 | ERS20924480* |
| 54649 | *H. castroaguirrei* | Anton Lizardo, Mex. | 2013-07-02 | 19.15555 | -95.86390 | ERS20924481* |
| 54650 | *H. castroaguirrei* | Anton Lizardo, Mex. | 2013-07-02 | 19.15555 | -95.86390 | ERS20924482* |
| 54689 | *H. atlahua* | Anton Lizardo, Mex. | 2013-07-02 | 19.17068 | -95.87361 | ERS20924483* |
| 54761 | *H. atlahua* | Anton Lizardo, Mex. | 2013-06-29 | 19.08144 | -95.96852 | ERS20924484* |
| 54786 | *H. atlahua* | Anton Lizardo, Mex. | 2013-06-29 | 19.09803 | -95.98555 | ERS20924485* |
| 54849 | *H. atlahua* | Anton Lizardo, Mex. | 2013-06-30 | 19.05436 | -95.83570 | ERS20924486* |
| 62263 | *H. puella* | Arr. Alacranes, Mex. | 2015-10-04 | 22.39949 | -89.48939 | ERS20924487* |
| 62264 | *H. puella* | Arr. Alacranes, Mex. | 2015-10-04 | 22.39949 | -89.48939 | ERS20924488* |
| 62515 | *H. gemma* | Cayo Arcas, Mex. | 2018-06-22 | 20.20523 | -91.96972 | ERS20924489* |
| 62549 | *H. ecosur* | Cayo Arcas, Mex. | 2018-06-24 | 20.21088 | -91.97582 | ERS20924490* |
| 62550 | *H. aberrans* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924491* |
| 62551 | *H. aberrans* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924492* |
| 62552 | *H. aberrans* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924493* |
| 62553 | *H. gemma* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924494* |
| 62555 | *H. sp2* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924495* |
| 62556 | *H. ecosur* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924496* |
| 62558 | *H. floridae* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924497* |

| | | | | | | |
|---|---|---|---|---|---|---|
| 62559 | *H. floridae* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924498* |
| 62560 | *H. floridae* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924499* |
| 62562 | *H. randallorum* | Cayo Arcas, Mex. | 2018-06-24 | 20.20394 | -91.97077 | ERS20924500* |
| 62570 | *H. gemma* | Cayo Arcas, Mex. | 2018-06-25 | 20.20354 | -91.97664 | ERS20924501* |
| 62571 | *H. sp2* | Cayo Arcas, Mex. | 2018-06-25 | 20.20354 | -91.97664 | ERS20924502* |
| 62575 | *H. randallorum* | Cayo Arcas, Mex. | 2018-06-25 | 20.20354 | -91.97664 | ERS20924503* |
| 62576 | *H. ecosur* | Cayo Arcas, Mex. | 2018-06-25 | 20.20354 | -91.97664 | ERS20924504* |
| 62577 | *H. puella* | Cayo Arcas, Mex. | 2018-06-25 | 20.20354 | -91.97664 | ERS20924505* |
| 62585 | *H. sp2* | Cayo Arcas, Mex. | 2018-06-25 | 20.19568 | -91.95947 | ERS20924506* |
| 62948 | *H. puella* | Cayo Arcas, Mex. | 2018-06-19 | 20.20492 | -91.97831 | ERS20924507* |
| 62953 | *H. puella* | Cayo Arcas, Mex. | 2018-06-19 | 20.20492 | -91.97831 | ERS20924508* |
| Bocas16.3 | *S. tortugarum* | Bocas del Toro, Pan. | 2016-02 | NA | NA | ERS20924509* |
| Bocas16.4 | *S. tortugarum* | Bocas del Toro, Pan. | 2016-02 | NA | NA | ERS20924510* |
| FL0318 | *H. indigo* | Puerto Rico | 2020-02-05 | 18.47923 | -67.00593 | SRR19070349 |
| FL0324 | *H. sp1* | Puerto Rico | 2020-02-05 | 18.47923 | -67.00593 | ERS20924511* |
| FL0331 | *H. chlorurus* | Puerto Rico | 2020-02-05 | 18.47923 | -67.00593 | SRR18184334 |
| FL0835 | *H. providencianus* | San Andres, Col. | 2016-06-06 | 12.50162 | -81.73137 | SRR17839752 |
| FL0836 | *H. guttavarius* | San Andres, Col. | 2016-06-06 | 12.50162 | -81.73137 | SRR18184176 |
| FL0839 | *H. aberrans* | San Andres, Col. | 2016-06-06 | 12.50162 | -81.73137 | SRR19070321 |
| FL0880 | *H. randallorum* | San Andres, Col. | 2016-06-08 | 12.83287 | -82.22055 | SRR17839729 |
| HypoHaiti1 | *H. liberte* | Haiti | 2015-08-05 | 19.67700 | -71.84300 | ERS20924512* |
| HypoHaiti2 | *H. liberte* | Haiti | 2015-09-04 | 19.67700 | -71.84300 | ERS20924513* |
| HypoHaiti3 | *H. liberte* | Haiti | 2015-09-04 | 19.67700 | -71.84300 | ERS20924514* |
| PL17_101 | *H. maya* | Belize | 2017-06-04 | 16.77096 | -88.16366 | ERS2899592 |
| PL17_111 | *H. indigo* | Belize | 2017-11-04 | 16.89555 | -88.06139 | ERS14948399 |
| PL17_119 | *H. maya* | Belize | 2017-04-13 | 16.48726 | -88.27183 | ERS2899593 |
| PL17_120 | *H. maya* | Belize | 2017-04-13 | 16.48726 | -88.27183 | ERS2899594 |
| PL17_121 | *H. maya* | Belize | 2017-04-13 | 16.48726 | -88.27183 | ERS2899595 |
| PL17_122 | *H. maya* | Belize | 2017-04-13 | 16.48726 | -88.27183 | ERS2899596 |
| PL17_123 | *H. maya* | Belize | 2017-04-13 | 16.51037 | -88.25462 | ERS2899597 |
| PL17_124 | *H. maya* | Belize | 2017-04-13 | 16.51037 | -88.25462 | ERS2899598 |
| PL17_125 | *H. randallorum* | Belize | 2017-04-13 | 16.51037 | -88.25462 | ERS14948421 |
| PL17_126 | *H. maya* | Belize | 2017-04-13 | 16.51037 | -88.25462 | ERS2899599 |
| PL17_134 | *H. unicolor* | Florida Keys, USA | 2017-05-07 | 24.84680 | -80.62296 | ERS14948448 |
| PL17_135 | *H. unicolor* | Florida Keys, USA | 2017-06-07 | 24.75258 | -80.76065 | ERS14948455 |
| PL17_136 | *H. unicolor* | Florida Keys, USA | 2017-06-07 | 24.75258 | -80.76065 | ERS14948452 |
| PL17_137 | *H. unicolor* | Florida Keys, USA | 2017-07-07 | 24.73489 | -80.80080 | ERS14948453 |
| PL17_138 | *H. unicolor* | Florida Keys, USA | 2017-07-07 | 24.73489 | -80.80080 | ERS14948451 |
| PL17_139 | *H. puella* | Florida Keys, USA | 2017-08-07 | 24.80618 | -80.67677 | ERS14948450 |
| PL17_140 | *H. unicolor* | Florida Keys, USA | 2017-08-07 | 24.80618 | -80.67677 | ERS14948456 |
| PL17_141 | *H. unicolor* | Florida Keys, USA | 2017-08-07 | 24.80618 | -80.67677 | ERS14948457 |
| PL17_142 | *H. gemma* | Florida Keys, USA | 2017-08-07 | 24.80618 | -80.67677 | ERS2899137 |
| PL17_143 | *H. unicolor* | Florida Keys, USA | 2017-09-07 | 24.81247 | -80.66971 | ERS14948454 |
| PL17_144 | *H. gemma* | Florida Keys, USA | 2017-09-07 | 24.81247 | -80.66971 | ERS2899138 |
| PL17_145 | *H. gemma* | Florida Keys, USA | 2017-09-07 | 24.81247 | -80.66971 | ERS2899139 |
| PL17_148 | *H. gemma* | Florida Keys, USA | 2017-11-07 | 24.76937 | -80.72791 | ERS2899140 |
| PL17_149 | *H. nigricans* | Florida Keys, USA | 2017-12-07 | 24.89978 | -80.61716 | ERS14948447 |
| PL17_153 | *H. gemma* | Florida Keys, USA | 2017-07-13 | 24.80696 | -80.67675 | ERS2899141 |
| PL17_155 | *H. puella* | Florida Keys, USA | 2017-07-14 | 25.04367 | -80.36930 | ERS14948449 |

| | | | | | | |
|---|---|---|---|---|---|---|
| PL17_157 | *H. puella* | Florida Keys, USA | 2017-07-15 | 24.72062 | -80.83116 | ERS14948446 |
| PL17_159 | *H. puella* | Florida Keys, USA | 2017-07-16 | 24.68493 | -80.92229 | ERS14948445 |
| PL17_160 | *H. floridae* | Florida Keys, USA | 2017-08-17 | 24.50768 | -81.57143 | ERS4141277 |
| PL17_21 | *S. tigrinus* | Bocas del Toro, Pan. | 2017-02-07 | 9.30140 | -82.29410 | ERS20924515* |
| PL17_23 | *H. sp1* | Puerto Rico | 2017-03-13 | 17.93425 | -67.01878 | ERS14948480 |
| PL17_35 | *H. indigo* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948482 |
| PL17_37 | *H. chlorurus* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948469 |
| PL17_38 | *H. chlorurus* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948464 |
| PL17_39 | *H. chlorurus* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948460 |
| PL17_40 | *H. chlorurus* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948466 |
| PL17_41 | *H. chlorurus* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948462 |
| PL17_42 | *H. chlorurus* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948467 |
| PL17_43 | *H. chlorurus* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948461 |
| PL17_44 | *H. chlorurus* | Puerto Rico | 2017-03-15 | 17.95282 | -67.05643 | ERS14948463 |
| PL17_50 | *H. puella* | Puerto Rico | 2017-03-16 | 17.95505 | -67.05325 | ERS14948475 |
| PL17_53 | *H. puella* | Puerto Rico | 2017-03-16 | 17.95505 | -67.05325 | ERS14948476 |
| PL17_54 | *H. puella* | Puerto Rico | 2017-03-16 | 17.95505 | -67.05325 | ERS14948483 |
| PL17_55 | *H. unicolor* | Puerto Rico | 2017-03-16 | 17.95505 | -67.05325 | ERS14948485 |
| PL17_56 | *H. sp1* | Puerto Rico | 2017-03-17 | 17.95787 | -67.05804 | ERS14948484 |
| PL17_57 | *H. puella* | Puerto Rico | 2017-03-17 | 17.95787 | -67.05804 | ERS14948479 |
| PL17_60 | *H. puella* | Puerto Rico | 2017-03-17 | 17.95787 | -67.05804 | ERS14948481 |
| PL17_62 | *H. chlorurus* | Puerto Rico | 2017-03-17 | 17.95787 | -67.05804 | ERS14948465 |
| PL17_63 | *H. unicolor* | Puerto Rico | 2017-03-18 | 17.97003 | -67.04642 | ERS14948487 |
| PL17_64 | *H. indigo* | Puerto Rico | 2017-03-18 | 17.97003 | -67.04642 | ERS14948471 |
| PL17_65 | *H. puella* | Puerto Rico | 2017-03-18 | 17.97003 | -67.04642 | ERS14948490 |
| PL17_66 | *H. unicolor* | Puerto Rico | 2017-03-18 | 17.97003 | -67.04642 | ERS14948459 |
| PL17_67 | *H. unicolor* | Puerto Rico | 2017-03-18 | 17.93425 | -67.01878 | ERS14948493 |
| PL17_68 | *H. guttavarius* | Puerto Rico | 2017-03-20 | 17.89242 | -67.01418 | ERS14948470 |
| PL17_69 | *H. puella* | Puerto Rico | 2017-03-21 | 17.89483 | -67.01714 | ERS14948486 |
| PL17_70 | *H. unicolor* | Puerto Rico | 2017-03-21 | 17.89483 | -67.01714 | ERS14948489 |
| PL17_71 | *H. sp1* | Puerto Rico | 2017-03-21 | 17.89483 | -67.01714 | ERS14948478 |
| PL17_72 | *H. sp1* | Puerto Rico | 2017-03-21 | 17.89483 | -67.01714 | ERS14948472 |
| PL17_73 | *H. unicolor* | Puerto Rico | 2017-03-21 | 17.89032 | -67.01794 | ERS14948491 |
| PL17_74 | *H. unicolor* | Puerto Rico | 2017-03-21 | 17.89032 | -67.01794 | ERS14948492 |
| PL17_75 | *H. aberrans* | Puerto Rico | 2017-03-22 | NA | NA | ERS14948458 |
| PL17_76 | *H. sp1* | Puerto Rico | 2017-03-22 | NA | NA | ERS14948477 |
| PL17_77 | *H. unicolor* | Puerto Rico | 2017-03-22 | NA | NA | ERS14948488 |
| PL17_79 | *H. aberrans* | Puerto Rico | 2017-03-22 | NA | NA | ERS20924516* |
| PL17_82 | *H. puella* | Puerto Rico | 2017-03-23 | NA | NA | ERS14948473 |
| PL17_85 | *H. indigo* | Puerto Rico | 2017-03-24 | NA | NA | ERS14948474 |
| PL17_86 | *H. chlorurus* | Puerto Rico | 2017-03-24 | 17.95282 | -67.05643 | ERS14948468 |
| PL17_88 | *H. aberrans* | Belize | 2017-07-04 | 16.65361 | -88.20097 | ERS14948398 |
| PL17_89 | *H. maya* | Belize | 2017-07-04 | 16.66000 | -88.18500 | ERS2899590 |
| PL17_95 | *H. maya* | Belize | 2017-08-04 | 16.77096 | -88.16366 | ERS2899591 |
| PL17_98 | *H. indigo* | Belize | 2017-09-04 | 16.80126 | -88.07942 | ERS20924517* |
| Rare1 | *H. gemma* | Cayo Arenas, Mex. | 2017-01-01 | 22.11528 | -91.39833 | ERS20924518* |
| Rare2 | *H. aberrans* | Cayo Arenas, Mex. | 2017-01-01 | 22.11528 | -91.39833 | ERS20924519* |
| Rare3 | *H. puella* | Cayo Arenas, Mex. | 2017-01-01 | 22.11528 | -91.39833 | ERS20924520* |
| s_tort_3 | *S. tortugarum* | Bocas del Toro, Pan. | 2016-02 | NA | NA | ERS4141278 |

**Table S4.2 — Samples used for *de novo* genome assembly with long reads.** All samples were collected in Bocas del Toro, Panama. Date, latitude and longitude refer to sampling dates and coordinates. Accession specifies GenBank assembly accessions numbers, at which assembly statistics can be found.

| Species | Assembly | Date | Latitude | Longitude | Accession |
|---|---|---|---|---|---|
| *H. puella* | HypPue2.2 | 2022-02-10 | 9.252250 | -82.130233 | ERS20388951 |
| *H. nigricans* | HypNig1.2 | 2022-03-01 | 9.302050 | -82.205533 | ERS20388952 |
| *H. unicolor* | HypUni1.2 | 2022-03-01 | 9.302050 | -82.205533 | ERS20388953 |

**Table S4.3 — Samples used for DNA expression analysis.** All samples were collected in Bocas del Toro, Panama. Date, latitude and longitude refer to sampling dates and coordinates.

| Sample ID | Species | Tissue | Date | Latitude | Longitude | Accession |
|---|---|---|---|---|---|---|
| Pue_10 | *H. puella* | retina | 2017-02-09 | 9.367 | -82.291 | ERR2750999 |
| Uni_04 | *H. unicolor* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751000 |
| Nig_09 | *H. nigricans* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751001 |
| Pue_05 | *H. puella* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751002 |
| Nig_08 | *H. nigricans* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751003 |
| Pue_04 | *H. puella* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751004 |
| Pue_08 | *H. puella* | retina | 2017-02-09 | 9.367 | -82.291 | ERR2751005 |
| Pue_09 | *H. puella* | retina | 2017-02-09 | 9.367 | -82.291 | ERR2751006 |
| Uni_03 | *H. unicolor* | retina | 2017-02-06 | 9.318 | -82.222 | ERR2751007 |
| Nig_05 | *H. nigricans* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751008 |
| Nig_03 | *H. nigricans* | retina | 2017-02-06 | 9.318 | -82.222 | ERR2751009 |
| Uni_02 | *H. unicolor* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751010 |
| Nig_04 | *H. nigricans* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751011 |
| Nig_07 | *H. nigricans* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751012 |
| Pue_03 | *H. puella* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751013 |
| Pue_02 | *H. puella* | retina | 2017-02-06 | 9.318 | -82.222 | ERR2751014 |
| Nig_06 | *H. nigricans* | retina | 2017-02-06 | 9.318 | -82.222 | ERR2751015 |
| Pue_07 | *H. puella* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751016 |
| Uni_05 | *H. unicolor* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751017 |
| Pue_06 | *H. puella* | retina | 2017-02-07 | 9.301 | -82.294 | ERR2751018 |
| Uni_01 | *H. unicolor* | retina | 2017-02-06 | 9.318 | -82.222 | ERR2751019 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Nig_01 | *H. nigricans* | retina | 2017-02-06 | 9.318 | -82.222 | ERR2751020 |
| Pue_01 | *H. puella* | retina | 2017-02-06 | 9.318 | -82.222 | ERR2751021 |
| Nig_02 | *H. nigricans* | retina | 2017-02-06 | 9.318 | -82.222 | ERR2751022 |
| S52 | *H. unicolor* | skin | 2022-07-11 | 9.299082 | -82.207186 | * |
| S53 | *H. puella* | skin | 2022-07-10 | 9.299082 | -82.207186 | * |
| S54 | *H. unicolor* | skin | 2022-07-15 | 9.279863 | -82.218015 | * |
| S55 | *H. puella* | skin | 2022-07-14 | 9.279863 | -82.218015 | * |
| S56 | *H. unicolor* | skin | 2022-05-02 | 9.322150 | -82.220150 | * |
| S57 | *H. unicolor* | skin | 2022-07-19 | 9.332653 | -82.199490 | * |
| S58 | *H. nigricans* | skin | 2022-07-16 | 9.332653 | -82.199490 | * |
| S59 | *H. nigricans* | skin | 2022-07-14 | 9.279863 | -82.218015 | * |
| S60 | *H. nigricans* | skin | 2022-07-15 | 9.279863 | -82.218015 | * |
| S61 | *H. puella* | skin | 2022-07-14 | 9.279863 | -82.218015 | * |
| S62 | *H. nigricans* | skin | 2022-05-02 | 9.322150 | -82.220150 | * |
| S63 | *H. puella* | skin | 2022-07-19 | 9.332653 | -82.199490 | * |
| S64 | *H. unicolor* | skin | 2022-07-09 | 9.299082 | -82.207186 | * |
| S65 | *H. unicolor* | skin | 2022-07-14 | 9.279863 | -82.218015 | * |
| S66 | *H. nigricans* | skin | 2022-07-09 | 9.299082 | -82.207186 | * |
| S67 | *H. nigricans* | skin | 2022-07-11 | 9.299082 | -82.207186 | * |
| S68 | *H. unicolor* | skin | 2022-05-02 | 9.322150 | -82.220150 | * |
| S69 | *H. puella* | skin | 2022-07-09 | 9.299082 | -82.207186 | * |
| S70 | *H. puella* | skin | 2022-07-16 | 9.332653 | -82.199490 | * |
| S71 | *H. unicolor* | skin | 2022-07-19 | 9.332653 | -82.199490 | * |
| S72 | *H. nigricans* | skin | 2022-07-19 | 9.332653 | -82.199490 | * |
| S73 | *H. puella* | optic tectum | 2022-07-10 | 9.299082 | -82.207186 | * |
| S74 | *H. unicolor* | diencephalon | 2022-07-11 | 9.299082 | -82.207186 | * |
| S75 | *H. unicolor* | diencephalon | 2022-07-19 | 9.332653 | -82.199490 | * |
| S76 | *H. nigricans* | diencephalon | 2022-07-14 | 9.279863 | -82.218015 | * |
| S77 | *H. unicolor* | diencephalon | 2022-07-15 | 9.279863 | -82.218015 | * |
| S78 | *H. nigricans* | optic tectum | 2022-07-09 | 9.299082 | -82.207186 | * |
| S79 | *H. unicolor* | telencephalon | 2022-07-14 | 9.279863 | -82.218015 | * |
| S80 | *H. puella* | optic tectum | 2022-07-09 | 9.299082 | -82.207186 | * |
| S81 | *H. puella* | telencephalon | 2022-07-14 | 9.279863 | -82.218015 | * |
| S82 | *H. puella* | diencephalon | 2022-07-10 | 9.299082 | -82.207186 | * |
| S83 | *H. unicolor* | diencephalon | 2022-07-09 | 9.299082 | -82.207186 | * |

| | | | | | | |
|---|---|---|---|---|---|---|
| S84 | *H. unicolor* | diencephalon | 2022-07-14 | 9.279863 | -82.218015 | * |
| S85 | *H. nigricans* | optic tectum | 2022-07-15 | 9.279863 | -82.218015 | * |
| S86 | *H. unicolor* | diencephalon | 2022-07-19 | 9.332653 | -82.199490 | * |
| S87 | *H. unicolor* | optic tectum | 2022-07-19 | 9.332653 | -82.199490 | * |
| S88 | *H. nigricans* | telencephalon | 2022-07-16 | 9.332653 | -82.199490 | * |
| S89 | *H. nigricans* | telencephalon | 2022-07-15 | 9.279863 | -82.218015 | * |
| S90 | *H. unicolor* | optic tectum | 2022-07-14 | 9.279863 | -82.218015 | * |
| S91 | *H. unicolor* | telencephalon | 2022-07-19 | 9.332653 | -82.199490 | * |
| S92 | *H. puella* | diencephalon | 2022-07-14 | 9.279863 | -82.218015 | * |
| S93 | *H. puella* | optic tectum | 2022-07-16 | 9.332653 | -82.199490 | * |
| S94 | *H. nigricans* | telencephalon | 2022-07-11 | 9.299082 | -82.207186 | * |
| S95 | *H. unicolor* | telencephalon | 2022-07-11 | 9.299082 | -82.207186 | * |
| S96 | *H. unicolor* | telencephalon | 2022-07-19 | 9.332653 | -82.199490 | * |
| S97 | *H. unicolor* | optic tectum | 2022-07-15 | 9.279863 | -82.218015 | * |
| S98 | *H. nigricans* | telencephalon | 2022-07-14 | 9.279863 | -82.218015 | * |
| S99 | *H. puella* | telencephalon | 2022-07-10 | 9.299082 | -82.207186 | * |
| S100 | *H. nigricans* | diencephalon | 2022-07-09 | 9.299082 | -82.207186 | * |
| S101 | *H. puella* | telencephalon | 2022-07-14 | 9.279863 | -82.218015 | * |
| S102 | *H. nigricans* | optic tectum | 2022-07-11 | 9.299082 | -82.207186 | * |
| S103 | *H. nigricans* | telencephalon | 2022-07-09 | 9.299082 | -82.207186 | * |
| S104 | *H. nigricans* | optic tectum | 2022-07-19 | 9.332653 | -82.199490 | * |
| S105 | *H. puella* | telencephalon | 2022-07-19 | 9.332653 | -82.199490 | * |
| S106 | *H. nigricans* | optic tectum | 2022-07-16 | 9.332653 | -82.199490 | * |
| S107 | *H. puella* | optic tectum | 2022-07-19 | 9.332653 | -82.199490 | * |
| S108 | *H. unicolor* | telencephalon | 2022-07-09 | 9.299082 | -82.207186 | * |
| S109 | *H. unicolor* | optic tectum | 2022-07-09 | 9.299082 | -82.207186 | * |
| S110 | *H. puella* | telencephalon | 2022-07-09 | 9.299082 | -82.207186 | * |
| S111 | *H. unicolor* | optic tectum | 2022-07-11 | 9.299082 | -82.207186 | * |
| S112 | *H. puella* | telencephalon | 2022-07-16 | 9.332653 | -82.199490 | * |
| S113 | *H. nigricans* | telencephalon | 2022-07-19 | 9.332653 | -82.199490 | * |
| S114 | *H. nigricans* | diencephalon | 2022-07-15 | 9.279863 | -82.218015 | * |
| S115 | *H. nigricans* | diencephalon | 2022-07-16 | 9.332653 | -82.199490 | * |
| S116 | *H. puella* | optic tectum | 2022-07-14 | 9.279863 | -82.218015 | * |
| S117 | *H. puella* | diencephalon | 2022-07-16 | 9.332653 | -82.199490 | * |
| S118 | *H. nigricans* | diencephalon | 2022-07-11 | 9.299082 | -82.207186 | * |

| | | | | | | |
|---|---|---|---|---|---|---|
| S119 | *H. puella* | optic tectum | 2022-07-14 | 9.279863 | -82.218015 | * |
| S120 | *H. unicolor* | telencephalon | 2022-07-15 | 9.279863 | -82.218015 | * |
| S121 | *H. puella* | diencephalon | 2022-07-14 | 9.279863 | -82.218015 | * |
| S122 | *H. puella* | diencephalon | 2022-07-09 | 9.299082 | -82.207186 | * |
| S123 | *H. puella* | diencephalon | 2022-07-19 | 9.332653 | -82.199490 | * |
| S124 | *H. nigricans* | diencephalon | 2022-07-19 | 9.332653 | -82.199490 | * |
| S125 | *H. unicolor* | optic tectum | 2022-07-19 | 9.299082 | -82.207186 | * |
| S126 | *H. nigricans* | optic tectum | 2022-07-14 | 9.279863 | -82.218015 | * |

**Table S4.4 — Demographic Inference with Linked Selection (DILS) results.** Models with migration are favoured over models with isolation (first column). Within models with migration, isolation with migration (IM, the two daughter populations continuously exchange alleles) is favored over a model with secondary contact (SC, the daughter populations evolve initially in isolation and then exchange alleles upon secondary contact, second column). Models with homogeneous effective population size and migration are favored over models with heterogeneous effective population size and migration (third and fourth column, respectively). PP: posterior probability.

| | **Migration vs. Isolation** | **IM vs. SC** | **Mhomo vs. Mhetero** | **Nhomo vs. Nhetero** |
|---|---|---|---|---|
| Best model | migration | IM | Mhomo | Nhomo |
| Vote counts | 860 \| 140 | 788 \| 212 | 846 \| 154 | 619 \| 381 |
| PP | 0.879 | 0.830 | 0.867 | 0.770 |

**Table S4.5 — Summary statistics of the six biogeographic models implemented in BioGeoBEARS** based on Schemes 1 and 2 (first and second block separated by horizontal line; see Methods for details). LnL: Log-Likelihood, npar: number of parameters, d: dispersal, e: extinction, j: founder-speciation, w: dispersal matrix power exponential, AICc: corrected Akaike Information Criterion.

| **Model** | **LnL** | **npar** | **d** | **e** | **j** | **w** | **AICc** | **AICc weight** |
|---|---|---|---|---|---|---|---|---|
| DEC | -108.5 | 2 | 0.028 | 1.00E-12 | 0 | 1 | 221.1 | 2.40E-17 |
| DEC+j | -70.29 | 3 | 1.00E-12 | 1.00E-12 | 0.025 | 1 | 146.7 | 0.33 |
| DIVALIKE | -101.7 | 2 | 0.034 | 1.00E-12 | 0 | 1 | 207.5 | 2.20E-14 |
| DIVALIKE+j | -70.29 | 3 | 1.00E-12 | 1.00E-12 | 0.025 | 1 | 146.7 | 0.33 |
| BAYAREALIKE | -143.5 | 2 | 0.033 | 0.077 | 0 | 1 | 291 | 1.60E-32 |
| BAYAREALIKE+j | -70.29 | 3 | 1.00E-07 | 1.00E-07 | 0.025 | 1 | 146.7 | 0.33 |
| DEC | -92.36 | 2 | 0.076 | 0.085 | 0 | 1 | 188.9 | 1.80E-18 |
| DEC+j | -51.48 | 3 | 1.00E-12 | 1.00E-12 | 0.1 | 1 | 109.4 | 0.33 |
| DIVALIKE | -86.54 | 2 | 0.087 | 0.03 | 0 | 1 | 177.3 | 6.10E-16 |
| DIVALIKE+j | -51.48 | 3 | 1.00E-12 | 1.00E-12 | 0.099 | 1 | 109.4 | 0.33 |
| BAYAREALIKE | -104.3 | 2 | 0.072 | 0.3 | 0 | 1 | 212.8 | 1.20E-23 |
| BAYAREALIKE+j | -51.49 | 3 | 1.00E-07 | 1.00E-07 | 0.094 | 1 | 109.4 | 0.33 |

**Table S4.6 — General read statistics and number of samples per tissue for the DNA expression analysis.** Asterisks denote datasets number of samples after removing outliers. M: million.

| | **Retina** | **Brain** | **Skin** |
|---|---|---|---|
| No. of samples, *H. puella* | 9* | 18 | 6 |
| No. of samples, *H. nigricans* | 9 | 17* | 7 |
| No. of samples, *H. unicolor* | 5 | 18 | 8 |
| Raw read length | 150 bp PE | 100 bp PE | 100 bp PE |
| Sum raw reads | 157.7 M | 1,676.7 M | 353.1 M |
| Mean ± sd raw reads | 6.6 ± 0.7 M | 31.1 ± 8.4 M | 16.8 ± 4.5 M |
| Mean ± sd trimmed read length | 129 ± 9 bp | 94 ± 1 bp | 95 ± 1 bp |
| Sum trimmed reads | 137.4 M | 1,293.1 M | 252.8 M |
| Mean ± sd trimmed reads | 5.7 ± 0.7 M | 24.0 ± 6.5 M | 12.0 ± 3.6 M |
| Sum uniquely aligned reads | 78.1 M | 603.3 M | 161.0 M |
| Mean ± sd uniquely aligned reads | 3.3 ± 0.4 M | 11.2 ± 2.9 M | 7.7 ± 2.5 M |

## References


Aguilar-Perera, A. (2003). Abundance and distribution of hamlets (Teleostei: Hypoplectrus) in coral reefs off southwestern Puerto Rico: Support for the multiple-species theory. Caribbean Journal of Science *39*, 147–150.

Aguilar-Perera, A. and González-Salas, C. (2010). Distribution of the genus Hypoplectrus (Teleostei: Serranidae) in the Greater Caribbean Region: support for a color-based speciation. Marine ecology *31*, 375–387.

Alexander, D. H., Novembre, J. and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome research *19*, 1655–1664.

Anders, S., Pyl, P. T. and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Anders, S., Reyes, A. and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Nature Precedings , 1–1.

Andrews, S. et al. (2010). FastQC: a quality control tool for high throughput sequence data.

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

Bhaskaran, N., Liu, Z., Dong, L., Zelenka, P., Zheng, L., Bletsos, V., Harrington, B. and Pandiyan, P. (2018). Identification of Casz1 as a regulatory protein controlling T helper cell differentiation, inflammation, and immunity. Frontiers in immunology *9*, 318439.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Braun, E. L., Cracraft, J. and Houde, P. (2019). Resolving the avian tree of life from top to bottom: the promise and potential boundaries of the phylogenomic era. Avian genomics in ecology and evolution: from the lab into the wild , 151–210.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009). BLAST+: architecture and applications. BMC bioinformatics *10*, 1–9.

Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nature Methods *18*, 170–175.

Chifman, J. and Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. Bioinformatics *30*, 530.

Coulmance, F., Akkaynak, D., Le Poul, Y., Höppner, M. P., McMillan, W. O. and Puebla, O. (2024). Phenotypic and genomic dissection of colour pattern variation in a reef fish radiation. Molecular Ecology *33*, mec.17047.

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. and 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M. et al. (2021). Twelve years of SAMtools and BCFtools. Gigascience *10*, giab008.

De Queiroz, K. (2007). Species concepts and species delimitation. Systematic biology *56*, 879–886.

Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F. and Marchini, J. (2013). Haplotype Estimation Using Sequencing Reads. The American Journal of Human Genetics *93*, 687–696.

Dobin, A. and Gingeras, T. R. (2015). Mapping RNA-seq reads with STAR. Current protocols in bioinformatics *51*, 11–14.

Domeier, M. L. (1994). Speciation in the serranid fish *Hypoplectrus*. Bull. Mar. Sci. *54*, 103–141.

Ellinghaus, D., Kurtz, S. and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC bioinformatics *9*, 1–14.

Falcon, S. and Gentleman, R. (2007). Using GOstats to test gene lists for GO term association. Bioinformatics *23*, 257–258.

Feder, J. L. (1998). The apple maggot fly, Rhagoletis pomonella: flies in the face of conventional wisdom. Endless forms species and speciation .

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

Flores, L. F. D. M., Tello-Musi, J. L. and Martínez-Pérez, J. A. (2011). Descripción de una nueva especie del género Hypoplectrus (Actinopterigy: Serranidae) del Sistema Arrecifal Veracruzano, suroeste del Golfo de México. Revista de Zoología *22*, 1–10.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C. and Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences *117*, 9451–9457.

Fraïsse, C., Popovic, I., Mazoyer, C., Spataro, B., Delmotte, S., Romiguier, J., Loire, E., Simon, A., Galtier, N., Duret, L. et al. (2021). DILS: Demographic inferences with linked selection by using ABC. Molecular Ecology Resources *21*, 2629–2644.

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

Goel, M. and Schneeberger, K. (2022). Plotsr: visualizing structural similarities and rearrangements between multiple genomes. Bioinformatics *38*, 2922–2926.

Goel, M., Sun, H., Jiao, W.-B. and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome biology *20*, 1–13.

Hallab, D. A. (2015). goEnrichment: Helper functions to compute GO enrichment tests using GOstats and GSEABase. R package version 1.0.

Hench, K., Helmkampf, M., McMillan, W. O. and Puebla, O. (2022). Rapid radiation in a highly diverse marine environment. Proceedings of the National Academy of Sciences *119*, e2020457119.

Hench, K., Mcmillan, W. O., Betancur-R, R. and Puebla, O. (2017). Temporal changes in hamlet communities (Hypoplectrus spp., Serranidae) over 17 years. Journal of fish biology *91*, 1475–1490.

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

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. and Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution *35*, 518–522.

Holt, B., Côté, I. and Emerson, B. (2010). Signatures of speciation? Distribution and diversity of Hypoplectrus (Teleostei: Serranidae) colour morphotypes. Global Ecology and Biogeography *19*, 432–441.

Holt, B. G., Côté, I. M. and 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. and 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.

Jin, Y., Tam, O. H., Paniagua, E. and Hammell, M. (2015). TEtranscripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. Bioinformatics *31*, 3593–3599.

Katoh, K. and Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution *30*, 772–780.

Keightley, P. and Jackson, B. (2018). Inferring the Probability of the Derived vs. the Ancestral Allelic State at a Polymorphic Site. Genetics *209*, 897–906.

Kelleher, J., Wong, Y., Wohns, A. W., Fadil, C., Albers, P. K. and McVean, G. (2019). Inferring whole-genome histories in large population datasets. Nature Genetics *51*, 1330–1338.

Kim, D., Langmead, B. and Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature methods *12*, 357–360.

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453–4455.

Kronforst, M. R., Young, L. G., Kapan, D. D., McNeely, C., O'Neill, R. J. and Gilbert, L. E. (2006). Linkage of butterfly mate preference and wing

color preference cue at the genomic location of wingless. Proceedings of the National Academy of Sciences *103*, 6575–6580.

Laetsch, D. R. and Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. F1000Research *6*, 1287.

Landis, M. J., Matzke, N. J., Moore, B. R. and Huelsenbeck, J. P. (2013). Bayesian analysis of biogeography when the number of areas is large. Systematic biology *62*, 789–804.

Langfelder, P. and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics *9*, 1–13.

Lewis, P. O. (2001). A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. Systematic Biology *50*, 913–925.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094–3100.

Liu, S., Hansen, M. M. and 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*, 5187–5202.

Liu, Z., Li, W., Ma, X., Ding, N., Spallotta, F., Southon, E., Tessarollo, L., Gaetano, C., Mukouyama, Y.-s. and Thiele, C. J. (2014). Essential role of the zinc finger transcription factor Casz1 for mammalian cardiac morphogenesis and development. Journal of Biological Chemistry *289*, 29801–29816.

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

Love, M., Anders, S. and Huber, W. (2014). Differential analysis of count data–the DESeq2 package. Genome Biology *15*, 10–1186.

Malinsky, M., Matschiner, M. and Svardal, H. (2021). Dsuite - Fast D-statistics and related admixture evidence from VCF files. Molecular Ecology Resources *21*, 584–595.

Mallet, J. (1995). A species definition for the modern synthesis. Trends in Ecology & Evolution *10*, 294–299.

Marshall, J. N. (2000). Communication and camouflage with the same 'bright'colours in reef fishes. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences *355*, 1243–1248.

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

Mattar, P., Jolicoeur, C., Dang, T., Shah, S., Clark, B. S. and Cayouette, M. (2021). A Casz1–NuRD complex regulates temporal identity transitions in neural progenitors. Scientific reports *11*, 3858.

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

Matzke, N. J. (2013). BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. R package, version 0.2 *1*, 2013.

Matzke, N. J. (2014). Model selection in historical biogeography reveals that founder-event speciation is a crucial process in island clades. Systematic biology *63*, 951–970.

Mayr, E. (2001). Wu's genic view of speciation. Journal of Evolutionary Biology *14*, 866–867.

Mccartney, M. A., Acevedo, J., Heredia, C., Rico, C., Quenoville, B., Bermingham, E. and 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. and DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research *20*, 1297–1303.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A. and Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution *37*, 1530–1534.

Moran, B. M., Hench, K., Waples, R. S., Höppner, M. P., Baldwin, C. C., McMillan, W. O. and Puebla, O. (2019). The evolution of microendemism in a reef fish (Hypoplectrus maya). Molecular Ecology *28*, mec.15110.

Morgan, M., Falcon, S. and Gentleman, R. (2023). GSEABase: Gene set enrichment data structures and methods. R package version 1.64.0.

Nei, M. and Li, W.-H. (1979). Mathematical Model for Studying Genetic Variation in Terms of Restriction Endonucleases. Proceedings of the National Academy of Sciences of the United States of America *76*, 5269–5273.

Ou, S. and Jiang, N. (2018). LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant physiology *176*, 1410–1422.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T. and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065–1093.

Puebla, O., Bermingham, E. and 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. and Guichard, F. (2012). Pairing dynamics and the origin of species. Proc. R. Soc. B-Biol. Sci. *279*, 1085–1092.

Puebla, O., Bermingham, E., Guichard, F. and Whiteman, E. (2007). Colour pattern as a single trait driving speciation in *Hypoplectrus* coral reef fishes? Proc. R. Soc. B-Biol. Sci. *274*, 1265–1271.

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

Puebla, O., Coulmance, F., Estape, C. J., Estape, A. M. and 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. and 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.

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M. and Alfaro, M. E. (2018). An inverse latitudinal gradient in speciation rate for marine fishes. Nature *559*, 392–395.

Ramon, M. L., Lobel, P. S. and 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. and Randall, H. A. (1960). Examples of mimicry and protective resemblance in tropical marine fishes. Bulletin of Marine Science *10*, 444–480.

Ree, R. H. and Smith, S. A. (2008). Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic biology *57*, 4–14.

Rieseberg, L. H. and Burke, J. (2001). A genic view of species integration. Journal of Evolutionary Biology *14*, 883–886.

Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution *66*, 1–17.

Ronquist, F. (1997). Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Systematic biology *46*, 195–203.

Rosser, N., Seixas, F., Queste, L. M., Cama, B., Mori-Pezo, R., Kryvokhyzha, D., Nelson, M., Waite-Hudson, R., Goringe, M., Costa, M. et al. (2024). Hybrid speciation driven by multilocus introgression of ecological traits. Nature *628*, 811—-817.

Rossi, M., Hausmann, A. E., Alcami, P., Moest, M., Roussou, R., Van Belleghem, S. M., Wright, D. S., Kuo, C.-Y., Lozano-Urrego, D., Maulana, A. et al. (2024). Adaptive introgression of a visual preference gene. Science *383*, 1368–1373.

Scherz, M. D., Masonick, P., Meyer, A. and Hulsey, C. D. (2022). Between a rock and a hard polytomy: phylogenomics of the rock-dwelling Mbuna Cichlids of Lake Malaŵi. Systematic Biology *71*, 741–757.

Schiffels, S. and Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. Nature genetics *46*, 919–925.

Smit, A., Hubley, R. and Green, P. (2013). RepeatMasker Open-4.0.

Suh, A. (2016). The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. Zoologica Scripta *45*, 50–62.

Swofford, D. and et al. (2002). Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, MA. Version 4.

Tang, H., Wang, X., Bowers, J. E., Jin, H. M., Tang, M. and Paterson, A. R. (2015). ALLMAPS: robust scaffold ordering based on multiple maps. Genome Biology *16*, 3.

Tavera, J. and 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*, 29–38.

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

To, T.-H., Jung, M., Lycett, S. and Gascuel, O. (2016). Fast Dating Using Least-Squares Criteria and Algorithms. Systematic Biology *65*, 82–97.

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.

Wagner, C. E., Keller, I., Wittwer, S., Selz, O. M., Mwaiko, S., Greuter, L., Sivasundar, A. and Seehausen, O. (2013). Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Molecular ecology *22*, 787–798.

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

Wu, C. (2001). The genic view of the process of speciation. Journal of Evolutionary Biology *14*, 851–865.

Wu, C.-I. and Ting, C.-T. (2004). Genes and speciation. Nature Reviews Genetics *5*, 114–122.

Zhang, B. and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Statistical applications in genetics and molecular biology *4*.

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

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

# 05

# SYNTHESIS & PERSPECTIVES

The main objective of my thesis was to expand our knowledge on the processes shaping rapid diversification, and specifically evaluate the basis of diversity for selection at both the phenotypic and genetic levels. I characterize and investigate color pattern in hamlet fishes (*Hypoplectrus* spp.), a particularly relevant trait for assortative mating among species and for their predatory behavior. I find that phenotypic diversity is generated by the modularity of color pattern elements at the trait level and of alleles in a few key regions on the hamlet genome. As a result of this modularity, I show that hamlets' phenotypic variation is relatively continuous and that intraspecific variation plays an important role in the observed phenotypic continuity. Taking a phylogenetic approach to the hamlet radiation, I show that the observed reproductive isolation does not translate into a phylogenetic signal at the whole-genome level or at the scale of genomic regions associated with functional differences between species. This is expected, given the recent origin of the radiation and hints towards a more diluted polygenic architecture of color pattern elements.

This thesis showcases the value of quantitatively evaluating phenotypes in evolutionary studies. The pipeline developed in ***manuscript 1*** allows for the direct and unbiased measurement of variation in color patterns in fishes, which, in this thesis, is applied to hamlets. This level of resolution and multidimensionality is unprecedented in reef fish studies and allowed for the detailed, pixel-by-pixel analysis of color patterns in hamlets.

The main findings of this thesis are ***1)*** that gradual genetic and phenotypic diversity are the basis for selection (***manuscript 2***), ***2)*** that this diversity is generated by the combinations of different phenotypic elements and of underlying alleles at large-effect pleiotropic loci (***manuscript 1***), ***3)*** but that these large-effect loci do not solely participate in diversification and that smaller-effect and polygenic loci participate equally or more, while being less detectable (***manuscript 3***).

## 5.1 Hamlet diversification

### Combinatorial speciation

Results from ***manuscript 2*** show that the raw materials on which selection acts, whether at the phenotypic or genotypic level, is made up of continuous, gradual variation across species, built upon within-species variation. Such patterns of variation, whether it is genetic or phenotypic, are not unique to hamlets and have been demonstrated in other organisms. Examples of phenotypes best described as gradual variation include the mosaic melanin patterning of plumage patches in munias and southern capuchinos, and the striking variation in wing color patterns observed in *Heliconius* butterflies (Funk and Taylor, 2019; Stryjewski and Sorenson, 2017; Campagna et al., 2017; Van Belleghem et al., 2017, 2021; McMillan et al., 2020). The hamlets represent an extreme case of overlap between species in the genetic

space, fueled by intraspecific variation. Such patterns of genetic variation have been suggested to drive speciation (Marques et al., 2019). In particular, genetic diversity generated by the recombination of ancestral genetic variants through introgressive hybridization has been shown to fuel the rapid radiations of munias, southern capuchinos, and monkeyflowers, *Heliconius* butterflies, and Lake Victoria cichlids (Stryjewski and Sorenson, 2017; Campagna et al., 2017; Stankowski and Streisfeld, 2015; Nelson et al., 2021; Pardo-Diaz et al., 2012; Enciso-Romero et al., 2017; Meier et al., 2017; Marques et al., 2019). Extended gene flow and the fact that genomic variants associated with color pattern traits predate the radiation (Figures S4.4, S4.7 and Table S4.4; ***manuscript 3***, Hench et al., 2022) suggest that similar mechanisms played a role in the rapid diversification of hamlets.

## Modularity at large-effect loci drive rapid radiations

In these rapid radiations, phenotypic diversity is driven by the allelic recombination of a few large-effect loci. Modularity in the alleles of *ASIP*, *KITLG* and *MC1R* are responsible for the phenotypic diversity in melanin patterning of munias and southern capuchinos (Funk and Taylor, 2019; Stryjewski and Sorenson, 2017; Campagna et al., 2017). In *Heliconius* butterflies, the combined evolution of *WntA*, *cortex* and *optix* genes is associated with black patterning and red formation (Concha et al., 2019; Saenko et al., 2019; Martin et al., 2014; Lewis et al., 2019; Van Belleghem et al., 2017; McMillan et al., 2020). Hamlet phenotypic diversity is also generated by the combinations of alleles at a few large-effect loci, specifically three, that each influence different elements of color patterns (***manuscript 1***). Among these, the *casz1* region has the largest effect, playing a key role in determining the discrete bar patterning seen in hamlets.

The research presented in this thesis therefore aligns with a combinatorial view of sympatric speciation, where large-effect loci introduced from standing genetic variation and introgressive hybridization provide a mechanism by which a group of organisms can diversify rapidly (Marques et al., 2019).

## Species or not?

Within speciation biology, definitions and concepts of species vary depending on the field of research or the organisms studied, making it difficult to establish a universal definition of what a species is (Stankowski and Ravinet, 2021b). Because hamlet species are very closely related, are genetically very similar, have extensive gene flow, and can hybridize, their status as separate species has long been debated (Puebla et al., 2022). From the phylogenetic perspective in ***manuscript 3*** and following the Phylogenetic Species Concept (Cracraft, 1983), which is based on supported monophyly, one might consider the hamlets as two species corresponding to the two identified lineages. However, reproductive isolation through assortative mating has been widely documented both in the wild and experimentally (Fischer, 1980; Domeier, 1994; Puebla et al., 2007; Barreto and McCartney, 2008; Puebla et al., 2012, 2022). This is consistent with the Biological Species Concept (Mayr, 1942) and with predictions from the genic view of speciation

(Wu, 2001), which suggests that reproductive isolation can occur prior to detectable phylogenetic signals. Hybridization is relatively common in nature, with examples even occurring between distantly related species that are not classified within the same genus (Mallet, 2005). For example, blue and fin whales are known to interbreed, yet this does not disqualify their classification as different species (Bérubé and Aguilar, 1998). Detailed descriptions of why hamlets are considered distinct species have been reviewed by Puebla et al. (2022) (Appendix). This work outlines how these fishes align with several species concepts, including the phenetic (Sokal and Crovello, 1970), biological (Mayr, 1942), recognition (Paterson, 1985), and genotypic cluster concepts (Mallet, 1995). Regardless of their species status, hamlets are characterized by phenotypic differentiation and strong reproductive isolation, and these are present in the near absence of a phylogenetic signal (***manuscript 3***). Further insights into how this process can occur would allow us to better understand how diversity arises.

## 5.2 Many complex processes shape biodiversity

### Nuancing the role of large-effect loci in phenotypic diversification

Large-effect loci play a role in phenotypic diversification, however their role may be less significant than previously thought (Wellenreuther and Hansson, 2016; Barghi et al., 2020). Polygenic evolution, in which quantitative traits are underpinned by many small-effect loci, has long been described in theoretical frameworks, but limitations in measurement methods have hindered their detection (Slatkin, 1970; Wellenreuther and Hansson, 2016; Barghi et al., 2020). Human, medical, agricultural, and domestication research, where large sample sizes are available and direct applications are valued, have applied these polygenic frameworks and explored the polygenic basis of traits (Quillen et al., 2019; Pavan and Sturm, 2019; Zhang et al., 2022). For example, human skin pigmentation is a polygenic trait involving dozens of genes, including *SLC24A5*, *SLC45A2*, *MC1R*, and *KITLG*, and has received much attention in scientific studies (Quillen et al., 2019; Pavan and Sturm, 2019). Another example where a polygenic basis underlie a trait is the aggression phenotype of domesticated Betta fish, *Betta splendens*, which is influenced by neural-related genes such as *esyt2*, *apbb2*, and *pank2* (Zhang et al., 2022). This example is particularly interesting as it highlights the potential to study the polygenic basis of behavioral phenotypes, such as mate choice in hamlets. Because of these methodological limitations, mainly due to sample size and statistical power, studies of the genetic basis of diversification have been biased toward the detection of large-effect loci (Wellenreuther and Hansson, 2016; Barghi et al., 2020). I believe that the results of my thesis suggest that diversification is also driven by nuanced polygenic signal, involving small-effect loci. Indeed, hamlet body coloration is a polygenic trait, and several smaller association peaks were identified in the GWAS for color pattern (***manuscript 1***; unpublished results). Furthermore, concatenating the genetic sequences from the eight largest

peaks associated with species identity better resolved the species tree than each of them individually (***manuscript 3***). These observations and results highlights the importance of using advanced methods to detect subtle polygenic signals, which are critical for understanding the complexity of diversification in radiations.

## Importance of geographic context in speciation studies

Gavrilets and Vose (2005) predicted that in the early stages of adaptive radiations, the spatial structure within a species might be more pronounced than the spatial genetic structure overall. Although the open ocean and pelagic larval stages of many fish species suggest high connectivity, widespread dispersal (Palumbi, 1994; Kinlan and Gaines, 2003; Kokko and López-Sepulcre, 2006; Hernández et al., 2023), and imply reduced level of reproductive isolation, reef fishes inhabit fragmented coral reefs (Syms and Jones, 2000; Bonin et al., 2011; Leprieur et al., 2021). These fragmented habitats resemble island systems, suggesting that coral reef fish diversification is influenced by geographic factors, as much as terrestrial islands or lake radiations (Salzburger, 2008; Leprieur et al., 2021). Geographic constraints have been shown to shape diversification in systems such as Galapagos finches, *Anolis* lizards, Hawaiian *Drosophila*, silverswords, and East African cichlids (Schluter, 2000; Stroud and Losos, 2020; Poe et al., 2017; Salzburger and Meyer, 2004). The repeated evolution of similar phenotypes in organisms, such as birds and butterflies, further underscores the adaptive nature of these radiations (Grant, 2017; Lawson and Petren, 2017; Urban et al., 2021). For example, beak morphology in Darwin's finches has evolved independently across islands, driven by diet-based natural selection (Grant, 2017; Lawson and Petren, 2017), while wing pattern evolution in *Heliconius* butterflies has converged in distinct regions (Urban et al., 2021). In hamlets, comparison of genetic and phenotypic variation across their entire distribution range highlighted the importance of local processes in driving sympatric speciation and maintaining species clusters (***manuscript 2***). The decoupling between population-level evolutionary patterns and phylogenetic signals (***manuscript 3***) reinforces the spatial structure view of Gavrilets and Vose (2005) and confirms that considering geographic scale in recent radiation studies is important. This suggests that species concepts such as the phylogenetic (Cracraft, 1983), the phenetic (Sokal and Crovello, 1970), and the genomic cluster species concepts (Mallet, 1995) may be effective at local scales, but become problematic at larger geographic scales (Mallet, 2001; Losos and Glor, 2003; Salzburger, 2008). These problems likely arise due to evolutionary processes such as introgression, hybridization, and incomplete lineage sorting, which have been observed in the phylogenies of species like bats (Hahn and Nakhleh, 2016). As predicted by Losos and Glor (2003) and echoed in Gavrilets and Vose (2005) and Gillespie et al. (2020), phylogenies often fail to incorporate the geographic context of speciation. These findings underscore the need for caution in interpreting evolutionary radiations across broad geographic scales and the critical role of local processes in maintaining species clusters.

## The continuum of reproductive isolation beyond sequences

The concept of speciation as a continuum of reproductive isolation has been widely discussed from a genetic perspective (Stankowski and Ravinet, 2021a; Shaw and Mullen, 2014; Barley et al., 2024), but multiple axes of divergence also play a role in such processes and should not be overlooked. In my thesis, I highlighted genetic mechanisms involved in the genetic similarities among species (Synthesis & Perspectives 5.2, paragraph 1). The genic view of speciation predicts such similarities, specifically during the early stages of radiation (Wu, 2001). Furthermore, I showed that reproductive isolation can evolve rapidly, driven by strong reproductive barriers, long before genome-wide divergence accumulates (***manuscript 3***). Consequently, measures of genetic divergence may not reliably capture the extent of reproductive isolation. Bolnick et al. (2023) propose a multivariate extension of the speciation continuum, namely the speciation hypercube, building on the concept already proposed by Dieckmann (2004). The original framework presented three axes of divergence: ecological differentiation, reproductive isolation (originally mating differentiation), and spatial differentiation (Dieckmann, 2004). While the three dimensions provide a unifying framework for speciation research, they do not reflect the whole complexity of processes involved. Bolnick et al. (2023) proposed a flexible alternative to account for additional dimensions. Research on cichlids applied the hypercube framework to test how the evolution of reproductive isolation depends on genetic, ecological, and phenotypic divergence (Weber et al., 2021). They found that continuous morphological differentiation and two genomic differentiation patterns are involved in the parallel evolution of cichlids (Weber et al., 2021). In my thesis, the hamlet research provides a clear example of what Bolnick et al. (2023) defined as the speciation hypercube, with axes representing genetic divergence, phenotypic divergence, and pre-zygotic reproductive isolation. In the second paragraph of Synthesis & Perspectives 5.2, I emphasized the importance of integrating the geographic context of the study clade when evaluating reproductive isolation. In the context of the highly multivariate divergence hypercube, hamlets have reproductive isolation with phenotypic and spatial divergence, but little genetic divergence (Figure 5.1). My research underscores the need for such flexible and multidimensional frameworks, because understanding the multiple facets of speciation is essential to unraveling the origins of species and, ultimately, the mechanisms that shape biodiversity.

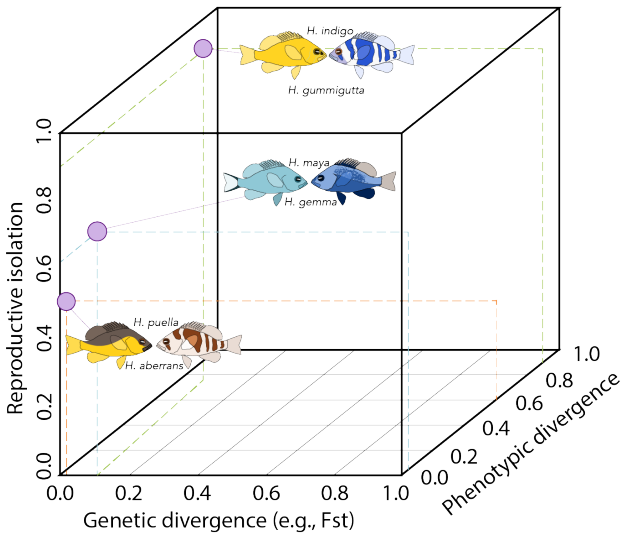



**Figure 5.1** — Hamlets in the speciation hypercube. Three hamlet species pairs are represented corresponding to varying degrees of reproductive isolation, genetic and phenotypic divergence (purple dots). Adapted from Bolnick et al. (2023), *Oxford University Press* license *5917620719420*.

## 5.3 Perspectives

### Comparative studies across lineages

Under the unifying framework for speciation research discussed above, comparative approaches across lineages could look at the different modes of diversification across the Tree of Life (Stankowski and Ravinet, 2021a). At the DNA level, comparative genomics allows for the investigation of structural rearrangements, gene family expansions or gene expression between lineages. Such mechanisms have played a role in the radiation of butterflies and fishes. For example, chromosomal inversions drove wing pattern diversification in the *Heliconius* radiation (Merrill et al., 2015; Jay and Joron, 2022) and loss of immune genes in angler fishes facilitated their bathypelagic transition (Brownstein et al., 2024). At a much broader scale, duplications and expansions of visual opsin gene families contributed to the teleosts diversification (Musilova et al., 2021), and the expression of *tbx3a* is essential for the posterior limb development of many vertebrates ranging from *Prionotus* walking fishes to humans (Herbert et al., 2024; Khan et al., 2020). The effect of such rearrangements, expansions or differential expressions could be further investigated at the phenotypic level with comparative phenomics. For example, the angler fish immunogenomic degradation is paired with sexual dimorphism, which has allowed reproductive success through sexual parasitism and further diversification in the bathypelagic zone (Brownstein et al., 2024). The quantitative image analysis pipeline developed in ***manuscript 1*** could allow for comparative color pattern evolution studies across reef fish species. In hamlets, the most closely related genus *Serranus* did not undergo a radiation, and comparisons between the two groups would help us understand why some lineages are more likely to radiate than others. Considering that vision is important for mate choice in hamlets and that vision is tightly linked to brain activity, parts and sizes (Bauchot et al., 1977; Howell et al., 2021), research avenues could couple genomic comparisons at loci that influence color pattern, vision and mate choice with phenotyping of color pattern, visual systems and brains. Overall, comparative studies between lineages and across the multiple axes of divergence within speciation frameworks would expand our knowledge on diversification modes and processes.

### The need for experiments

While comparative studies are great to deepen our knowledge on the multiple factors influencing reproductive isolation, they cannot establish causal links (Stankowski and Ravinet, 2021a). For example, one can describe genes and phenotypes involved in the hamlet radiation, but cannot know its direct effects on mate choice or foraging success. Empirical studies are therefore needed to validate findings from descriptive-based studies (Stankowski and Ravinet, 2021a; Bolnick et al., 2023). CRISPR–Cas9 knockouts of red pigmentation genes in zebrafish allowed the dissection of the cell mechanisms involved in the red coloration formation of teleost fishes (Huang et al., 2021). Knocking out the *Csf1* gene in male medaka fishes, *Oryzias woworae*, reduced their red coloration and showed that they were more preferred by females, drawing

causative conclusions on the role of *csf1* in reproductive success and its sexual selection (Ansai et al., 2021). With such approaches phenotypes can directly be evaluated, causal links between genotypes and phenotypes validated and direct fitness of mutants tested in experimental settings. For hamlets, experiments could first focus on knocking out the large-effect *casz1* gene involved in the vertical bar pattern, followed by other genes known to be involved in color patterning, local adaptation (e.g. *Tpm4*; Picq et al., 2016) or mate choice. Further breeding would allow the testing of pure, hybrid, backcrossed and mutant hamlets in mate choice and behavioral experiments and verify hypotheses about their sexual and natural selection. These experimental approaches have the potential to bridge the gap between descriptive studies and causal validation, offering deeper insights into the mechanisms driving reproductive isolation and diversification in hamlets or in any group of organisms.

## Conclusion

Overall, the hamlet research presented in this thesis has shed light on the multiple and complex mechanisms driving diversification in marine fishes inhabiting a largely open ocean environment. The modes of diversification in hamlets support a combinatorial view of speciation both at the genetic and phenotypic level, involving underlying large-effect loci—a pattern that has been recorded in many other marine and non-marine organisms. The importance of large-effect loci should be approached with caution, as the contribution of smaller-effect loci is more challenging to detect. The findings of my thesis support the genic view of speciation, showing that reproductive isolation may arise prior to any detectable genomic signal, especially when investigating recent radiations and when smaller-effect loci are involved. Together, this work shows that biodiversity is shaped by complex mechanisms and involves multiple axes of variation, each contributing to the accumulation of reproductive isolation. To comprehensively understand how diversity arises, a framework such as the speciation hypercube is required. It not only allows the extension of these multiple dimensions of variation for specific lineages, but also provides a unifying means to compare modes of diversification across lineages. The work presented in my thesis featured extensive phenotypic and genetic datasets, with sampling spanning the entire geographic range of hamlets, to explore their radiation. This approach provides a robust foundation for applying such frameworks to other reef fishes and opens pathways for future comparative studies on the modes of diversification in reef fishes and teleosts more generally.

## References


Ansai, S., Mochida, K., Fujimoto, S., Mokodongan, D. F., Sumarto, B. K. A., Masengi, K. W., Hadiaty, R. K., Nagano, A. J., Toyoda, A., Naruse, K. et al. (2021). Genome editing reveals fitness effects of a gene for sexual dichromatism in Sulawesian fishes. Nature communications *12*, 1350.

Barghi, N., Hermisson, J. and Schlötterer, C. (2020). Polygenic adaptation: a unifying framework to understand positive selection. Nature Reviews Genetics *21*, 769–781.

Barley, A. J., Nieto-Montes de Oca, A., Manríquez-Morán, N. L. and Thomson, R. C. (2024). Understanding Species Boundaries that Arise from Complex Histories: Gene Flow Across the Speciation Continuum in the Spotted Whiptail Lizards. Systematic Biology , syae040.

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

Bauchot, R., Bauchot, M., Platel, R. and Ridet, J. (1977). Brains of Hawaiian tropical fishes; brain size and evolution. Copeia *Vol. 1977, No. 1*, 42–46.

Bérubé, M. and Aguilar, A. (1998). A new hybrid between a blue whale, Balaenoptera musculus, and a fin whale, B. physalus: frequency and implications of hybridization. Marine Mammal Science *14*, 82–98.

Bolnick, D. I., Hund, A. K., Nosil, P., Peng, F., Ravinet, M., Stankowski, S., Subramanian, S., Wolf, J. B. and Yukilevich, R. (2023). A multivariate view of the speciation continuum. Evolution *77*, 318–328.

Bonin, M. C., Almany, G. R. and Jones, G. P. (2011). Contrasting effects of habitat loss and fragmentation on coral-associated reef fishes. Ecology *92*, 1503–1512.

Brownstein, C. D., Zapfe, K. L., Lott, S., Harrington, R. C., Ghezelayagh, A., Dornburg, A. and Near, T. J. (2024). Synergistic innovations enabled the radiation of anglerfishes in the deep open ocean. Current Biology *34*, 2541–2550.

Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L. and Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. Science advances *3*, e1602404.

Concha, C., Wallbank, R. W., Hanly, J. J., Fenner, J., Livraghi, L., Rivera, E. S., Paulo, D. F., Arias, C., Vargas, M., Sanjeev, M. et al. (2019). Interplay between developmental flexibility and determinism in the evolution of mimetic Heliconius wing patterns. Current Biology *29*, 3996–4009.

Cracraft, J. (1983). Species concepts and speciation analysis. Current ornithology *1*, 159–187.

Dieckmann, U. (2004). Adaptive speciation. Cambridge University Press.

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

Enciso-Romero, J., Pardo-Díaz, C., Martin, S. H., Arias, C. F., Linares, M., McMillan, W. O., Jiggins, C. D. and Salazar, C. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. Molecular ecology *26*, 5160–5172.

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

Funk, E. R. and Taylor, S. A. (2019). High-throughput sequencing is revealing genetic associations with avian plumage color. The Auk *136*, ukz048.

Gavrilets, S. and Vose, A. (2005). Dynamic patterns of adaptive radiation. Proceedings of the National Academy of Sciences *102*, 18040–18045.

Gillespie, R. G., Bennett, G. M., De Meester, L., Feder, J. L., Fleischer, R. C., Harmon, L. J., Hendry, A. P., Knope, M. L., Mallet, J., Martin, C. et al. (2020). Comparing adaptive radiations across space, time, and taxa. Journal of Heredity *111*, 1–20.

Grant, P. R. (2017). Ecology and Evolution of Darwin's Finches (Princeton Science Library Edition). Princeton University Press.

Hahn, M. W. and Nakhleh, L. (2016). Irrational exuberance for resolved species trees. Evolution *70*, 7–17.

Hench, K., Helmkampf, M., McMillan, W. O. and Puebla, O. (2022). Rapid radiation in a

highly diverse marine environment. Proceedings of the National Academy of Sciences *119*, e2020457119.

Herbert, A. L., Allard, C. A., McCoy, M. J., Wucherpfennig, J. I., Krueger, S. P., Chen, H. I., Gourlay, A. N., Jackson, K. D., Abbo, L. A., Bennett, S. H. et al. (2024). Ancient developmental genes underlie evolutionary novelties in walking fish. Current Biology *34*, 4339–4348.

Hernández, C. M., Paris, C. B., Vaz, A. C., Jones, B. T., Kellner, J. B., Richardson, D. E., Sponaugle, S., Cowen, R. K. and Llopiz, J. K. (2023). Diverse patterns of larval coral reef fish vertical distribution and consequences for dispersal and connectivity. Coral Reefs *42*, 453–465.

Howell, K. J., Beston, S. M., Stearns, S. and Walsh, M. R. (2021). Coordinated evolution of brain size, structure, and eye size in Trinidadian killifish. Ecology and Evolution *11*, 365–375.

Huang, D., Lewis, V. M., Foster, T. N., Toomey, M. B., Corbo, J. C. and Parichy, D. M. (2021). Development and genetics of red coloration in the zebrafish relative Danio albolineatus. Elife *10*, e70253.

Jay, P. and Joron, M. (2022). The double game of chromosomal inversions in a neotropical butterfly. Comptes Rendus. Biologies *345*, 57–73.

Khan, S. F., Damerell, V., Omar, R., Du Toit, M., Khan, M., Maranyane, H. M., Mlaza, M., Bleloch, J., Bellis, C., Sahm, B. D. et al. (2020). The roles and regulation of TBX3 in development and disease. Gene *726*, 144223.

Kinlan, B. P. and Gaines, S. D. (2003). Propagule dispersal in marine and terrestrial environments: a community perspective. Ecology *84*, 2007–2020.

Kokko, H. and López-Sepulcre, A. (2006). From individual dispersal to species ranges: perspectives for a changing world. Science *313*, 789–791.

Lawson, L. P. and Petren, K. (2017). The adaptive genomic landscape of beak morphology in Darwin's finches. Molecular ecology *26*, 4978–4989.

Leprieur, F., Pellissier, L., Mouillot, D. and Gaboriau, T. (2021). Influence of historical changes in tropical reef habitat on the diversification of coral reef fishes. Scientific Reports *11*, 20731.

Lewis, J. J., Geltman, R. C., Pollak, P. C., Rondem, K. E., Van Belleghem, S. M., Hubisz, M. J., Munn, P. R., Zhang, L., Benson, C., Mazo-Vargas, A. et al. (2019). Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. Proceedings of the National Academy of Sciences *116*, 24174–24183.

Losos, J. B. and Glor, R. E. (2003). Phylogenetic comparative methods and the geography of speciation. Trends in Ecology & Evolution *18*, 220–227.

Mallet, J. (1995). A species definition for the modern synthesis. Trends in Ecology & Evolution *10*, 294–299.

Mallet, J. (2001). Species, concepts of. Encyclopedia of biodiversity *5*, 427–440.

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

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

Martin, A., McCulloch, K. J., Patel, N. H., Briscoe, A. D., Gilbert, L. E. and Reed, R. D. (2014). Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. EvoDevo *5*, 1–14.

Mayr, E. (1942). Systematics and the origin of species–Columbia Univ. Press, New York , 99–107.

McMillan, W. O., Livraghi, L., Concha, C. and Hanly, J. J. (2020). From patterning genes to process: unraveling the gene regulatory networks that pattern Heliconius wings. Frontiers in Ecology and Evolution *8*, 221.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L. and Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. Nature communications *8*, 14363.

Merrill, R. M., Dasmahapatra, K. K., Davey, J., Dell'Aglio, D., Hanly, J., Huber, B., Jiggins, C. D., Joron, M., Kozak, K., Llaurens, V. et al. (2015). The diversification of Heliconius butterflies: what have we learned in 150 years? Journal of Evolutionary Biology *28*, 1417–1438.

Musilova, Z., Salzburger, W. and Cortesi, F. (2021). The visual opsin gene repertoires of teleost fishes: evolution, ecology, and function. Annual review of cell and developmental biology *37*, 441–468.

Nelson, T. C., Stathos, A. M., Vanderpool, D. D., Finseth, F. R., Yuan, Y.-w. and Fishman, L. (2021). Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (Mimulus section Erythranthe). PLoS Genetics *17*, e1009095.

Palumbi, S. R. (1994). Genetic divergence, reproductive isolation, and marine speciation. Annual review of ecology and systematics *25*, 547–572.

Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O. and Jiggins, C. D. (2012). Adaptive introgression across species boundaries in Heliconius butterflies. PLoS genetics *8*, e1002752.

Paterson, H. E. H. (1985). The Recognition Concept of Species. Species and Speciation .

Pavan, W. J. and Sturm, R. A. (2019). The genetics of human skin and hair pigmentation. Annual review of genomics and human genetics *20*, 41–72.

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

Poe, S., Nieto-Montes de Oca, A., Torres-Carvajal, O., De Queiroz, K., Velasco, J. A., Truett, B., Gray, L. N., Ryan, M. J., Köhler, G., Ayala-Varela, F. et al. (2017). A phylogenetic, biogeographic, and taxonomic study of all extant species of Anolis (Squamata; Iguanidae). Systematic biology *66*, 663–697.

Puebla, O., Bermingham, E. and 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. and 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., Coulmance, F., Estape, C. J., Estape, A. M. and 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.

Quillen, E. E., Norton, H. L., Parra, E. J., Lona-Durazo, F., Ang, K. C., Illiescu, F. M., Pearson, L. N., Shriver, M. D., Lasisi, T., Gokcumen, O. et al. (2019). Shades of complexity: New perspectives on the evolution and genetic architecture of human skin. American Journal of Physical Anthropology *168*, 4–26.

Saenko, S. V., Chouteau, M., Piron-Prunier, F., Blugeon, C., Joron, M. and Llaurens, V. (2019). Unravelling the genes forming the wing pattern supergene in the polymorphic butterfly Heliconius numata. EvoDevo *10*, 1–12.

Salzburger, W. (2008). To be or not to be a hamlet pair in sympatry.

Salzburger, W. and Meyer, A. (2004). The species flocks of East African cichlid fishes: recent advances in molecular phylogenetics and population genetics. Naturwissenschaften *91*, 277–290.

Schluter, D. (2000). The ecology of adaptive radiation. OUP Oxford.

Shaw, K. L. and Mullen, S. P. (2014). Speciation continuum. Journal of Heredity *105*, 741–742.

Slatkin, M. (1970). Selection and polygenic characters. Proceedings of the National Academy of Sciences *66*, 87–93.

Sokal, R. R. and Crovello, T. J. (1970). The biological species concept: a critical evaluation. The American Naturalist *104*, 127–153.

Stankowski, S. and Ravinet, M. (2021a). Defining the speciation continuum. Evolution *75*, 1256–1273.

Stankowski, S. and Ravinet, M. (2021b). Quantifying the use of species concepts. Current Biology *31*, R428–R429.

Stankowski, S. and Streisfeld, M. A. (2015). Introgressive hybridization facilitates adaptive divergence in a recent radiation of monkeyflowers. Proceedings of the Royal Society B: Biological Sciences *282*, 20151666.

Stroud, J. T. and Losos, J. B. (2020). Bridging the process-pattern divide to understand the origins and early stages of adaptive radiation: a review

of approaches with insights from studies of Anolis lizards. Journal of Heredity *111*, 33–42.

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

Syms, C. and Jones, G. P. (2000). Disturbance, habitat structure, and the dynamics of a coral-reef fish community. Ecology *81*, 2714–2729.

Urban, S., Nater, A., Meyer, A. and Kratochwil, C. F. (2021). Different sources of allelic variation drove repeated color pattern divergence in cichlid fishes. Molecular Biology and Evolution *38*, 465–477.

Van Belleghem, S. M., Lewis, J. J., Rivera, E. S. and Papa, R. (2021). Heliconius butterflies: a window into the evolution and development of diversity. Current opinion in genetics & development *69*, 72–81.

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. et al. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. Nature ecology & evolution *1*, 0052.

Weber, A. A.-T., Rajkov, J., Smailus, K., Egger, B. and Salzburger, W. (2021). Speciation dynamics and extent of parallel evolution along a lake-stream environmental contrast in African cichlid fishes. Science Advances *7*, eabg5391.

Wellenreuther, M. and Hansson, B. (2016). Detecting polygenic evolution: problems, pitfalls, and promises. TRENDS in Genetics *32*, 155–164.

Wu, C.-I. (2001). The genic view of the process of speciation. Journal of evolutionary biology *14*, 851–865.

Zhang, W., Wang, H., Brandt, D. Y., Hu, B., Sheng, J., Wang, M., Luo, H., Li, Y., Guo, S., Sheng, B. et al. (2022). The genetic architecture of phenotypic diversity in the Betta fish (Betta splendens). Science advances *8*, eabm4955.

# APPENDIX

# A review of 263 years of taxonomic research on *Hypoplectrus* (Perciformes: Serranidae), with a redescription of *Hypoplectrus affinis* (Poey, 1861)

Oscar Puebla, **Floriane Coulmance,** Carlos J. Estapé,, Allison Morgan Estapé, D. Ross Robertson

**Keywords:** Caribbean, reef fish, hamlet, bluelip

**Abstract:** The hamlets (*Hypoplectrus* spp., Perciformes: Serranidae) constitute a distinctive model system for the study of a variety of ecological and evolutionary processes including the evolution and maintenance of simultaneous hermaphroditism and egg trading, sex allocation, sexual selection, social-trap, mimicry, dispersal, speciation, and adaptive radiation. Addressing such fundamental and complex processes requires a good knowledge of the taxonomy and natural history of the hamlets. Here, we review the taxonomy of the hamlets, from early ichthyological studies to the most recent species description in 2018. We report a total of 72 different binomial names for *Hypoplectrus*, synonymized or invalidated down to 17 unambiguously recognized species today. In addition, we redescribe *Hypoplectrus affinis* (Poey, 1861) as a valid species. In Bocas del Toro (Panama), this hamlet is distinct from eight sympatric congeners in terms of color pattern, body size and behavior. Whole-genome analysis and spawning observations indicate that it is genetically distinct from sympatric congeners and reproductively isolated through assortative pairing. Based on the color pattern we detail in its redescription, live-fish photographs, videos, and earlier reports, *H. affinis* occurs in Panama, Nicaragua, Mexico, the Florida Keys, Cuba, Grand Cayman, Jamaica, the Dominican Republic, Los Roques (Venezuela), Bonaire, and Tobago. We conclude with a discussion of pending taxonomic issues in this group and the species status of the hamlets in general.

**Author's Contribution:** **FC** conducted the genetic PCA, whole genome phylogeny and COI cladogram analyses, wrote and provided review on the manuscript.

Puebla, O., **Coulmance, F.**, Estape, C. J., Estape, 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(2), 101-141.

# List of Figures

# List of Supplementary Figures

## List of Supplementary Tables

# List of Publications

- Puebla, O., **Coulmance, F.**, Estape, C. J., Estape, 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(2), 101-141.
- **Coulmance, F.**, Akkaynak, D., Le Poul, Y., Höppner, M. P., McMillan, W. O., & Puebla, O. (2024). Phenotypic and genomic dissection of colour pattern variation in a reef fish radiation. Molecular Ecology, 33(4), e17047.

# Declaration of Author Contributions

## Manuscript 1

**FC** contributed to the development of the project, conducted all the analyses and wrote the first draft of the manuscript. DA contributed to the development of the camera system, part of the fieldwork and provided the scripts for color correction. YLP contributed to the image analysis and provided the scripts for the image alignment. MPH contributed to the sequencing. WOM contributed to development of the project and part of the fieldwork. OP 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.

## Manuscript 2

**FC**, MJH, JG, TK, ODD, KA, WOM and OP coordinated and performed sample collection. **FC**, MH and OP conducted laboratory work. WOM and OP provided funds for sequencing. **FC**, MH and OP analyzed the data and visualized the results. **FC** and OP wrote the manuscript with input and feedback from all co-authors.

## Manuscript 3

**FC**, MJH, AAP, ODD, JT, BCV, DRR, RBR, WOM and OP coordinated and performed sample collection. **FC**, MJH, RBR, PBF, WOM and OP conducted laboratory work. PBF, MTO, RBR, WOM and OP provided funds for sequencing. MH, **FC**, MJH, AB, IB, PBF, AS, RBR and OP analyzed the data and visualized the results. MH, **FC**, MJH and OP wrote the manuscript with input and feedback from all co-authors.

## Appendix: A review of 263 years of taxonomic research on *Hypoplectrus*

**FC** conducted the genetic PCA, whole genome phylogeny and COI cladogram analyses, wrote and provided review on the manuscript.

# Declarations

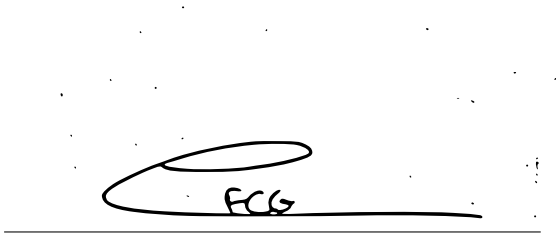
I, **Floriane Coulmance Gayrard**, hereby confirm that the following dissertation

*The phenotypic and genetic basis of a reef fish radiation*

was written independently by me, with the advice of my supervisor, is my own work in terms of content and form and no other sources and aids than those indicated were used.

This thesis has been written in accordance with the rules of good scientific practice of the Carl von Ossietzky University Oldenburg and the German Research Foundation and has not been submitted elsewhere as part of an examination procedure. Manuscripts that have been published or submitted for publication have been identified. No academic degree has been withdrawn.

Bremen, 05.12.2024



Floriane Coulmance Gayrard

## Acknowledgments

First of all, I would like to thank you, Oscar, for taking me on as your third Ph.D. student at such an unexpected time. What an adventure! Your mentorship, support and availability have been truly invaluable throughout this journey. It really feels like yesterday that we were sharing our first ideas and discussions. Thank you for always guiding me, for giving me the opportunity to contribute to this amazing project, and especially for pushing me to do the fellowship in Panama. Looking back, it was really one of the best years of my life, so I really appreciate it! You have also always managed to put the right people on my path throughout my PhD, it has really made things easier and none of it would have been possible without you.

To all the past and present members of the *Puebla lab*... Wow! This group has really grown! You have made my experience as part of the group really enjoyable and each one of you has contributed to shaping my academic and personal growth. Kosmas, you were the ideal support at the beginning of my thesis, thank you for everything you taught me. Martin, for being the undisputed facilitator of this group. Melanie, more than one sentence is needed to describe how much your advice and dedicated support as a field partner, post-doc, but most of all as a friend means to me. Serra, I don't know how I would have survived Germany if I hadn't met you, your positivity and listening ear are greatly appreciated. Sabrin and Cretus, thank you for all the deep and light conversations, I really learned a lot from both of you. To the people of the ZMT and those I met during scientific diving over the years, you made my PhD experience wonderful.

Special thanks to the people of Panama, Tobago, the US Virgin Islands, and Mexico, who pushed our science in so many different ways. D. Ross Robertson, thank you for the immense support and expert eye you provided throughout the fellowship. W. Owen McMillan for the warm welcome in Panama, the laughs during the fieldwork, and the constant support. I am forever grateful to the nice people at the STRI research station in Bocas, the Tobago dream team, the photographers in the US Virgin Islands, and our Mexico team.

I would also like to thank Nicolas Salamin for hosting me at the University of Lausanne and all the members of his laboratory. Your hospitality and constructive feedback made my stay memorable and impactful.

Friends, thank you for always keeping me grounded and for always putting a

smile on my face, for caring, listening and sharing, and for our many laughs. Whether in Bremen, Lausanne or Paris, you all truly contribute to my happiness.

To my family, you are the reason I am here today and words cannot express the gratitude I feel for your constant support and love. Grand-Père Pierre et Grand-Mère Geneviève, j'aurais tellement aimé que vous soyez encore à mes côtés. Grand-Père Pierre, merci d'avoir éveillé mon intérêt pour la nature et Grand-Mère Josette d'être le rayon de soleil qui a toujours su me comprendre. Maman, Papa, votre soutien et votre amour m'ont porté tout au long de ce chemin. Rien n'aurait été possible sans vous. Mon amour pour vous est tout aussi inconditionnel. Viviane, les années passent, et ta présence dans ma vie reste un véritable trésor.

Tane, funny that you showed me the way by starting your PhD first. We have done so much together, lets see where the future takes us, it will be fun like it always has been. I cherish having you by my side, you provide such precious support on all possible fronts. Also, a special thanks to your family, their warmth and hospitality always make me feel at home.