Genomics and the origin of marine species: final project report

O. Puebla[1]

[1]Leibniz Centre for Tropical Marine Research (ZMT), Fahrenheitstraße 6, 28359 Bremen, Germany

January 2026

This report follows the German Research Foundation guidelines, with sections 1-4 intended for the public.

1. General Information

DFG reference number: PU 571
Project number: 277213305.
Project title: Genomics and the origin of marine species
Name(s) of the applicant(s): Oscar Puebla
Official address(es): Leibniz Centre for Tropical Marine Research (ZMT), Fahrenheitstraße 6, 28359 Bremen, Germany
Name(s) of the co-applicants: N/A
Name(s) of the cooperation partners:
Dr. Owen McMillan (Smithsonian Tropical Research Institute (STRI), Panama)
Dr. Marc Höppner (Institute of Clinical Molecular Biology (IKMB), Kiel University, Germany)
Prof. Dr. Omar Dominguez (Universidad Michoacana de San Nicolás de Hidalgo, Mexico)
Dr. Yann Le Poul (Ludwig-Maximilians-Universität München)
Prof. Dr. David Parichy (University of Virginia, Charlottesville, USA)
Note that these are the partners that were identified early on and mentioned in the project proposal. Close to 30 co-authors contributed in the end.
Reporting period (entire funding period): 2 x 3 years of funding between 2016 and 2024, with a six-months gap between the two phases and two cost-neutral extensions.

DOI: https://doi.org/10.21244/zmt.2026.001

2. Summary

In this project we shed light on the genomic bases of diversification of the hamlets, a group of brightly coloured reef fishes from the Greater Caribbean that represents one of the fastest evolutionary radiation on Earth. We assembled a reference genome for the group, at a time when this was still a challenge for non-model species, and re-sequenced more than 300 genomes from all described species. Our results show that although these fishes present clear differences in colour pattern and are reproductively isolated through strong assortative mating, they are nonetheless extremely similar genetically throughout most of the genome, except for a few narrow genomic regions that underlie functional divergence among species (essentially colour pattern and mate choice). For this we notably teamed up with a marine engineer and an image analysis specialist. Together we developed a portable underwater photo studio to take standardized images of live fishes *in situ* and a method to analyse the images in a fully standardized and automated way. This allowed us to detect associations between colour pattern variation and genetic variation with pixel and DNA base pair resolution, respectively. The results of our research were published in high-profile scientific journals such as *Nature Ecology and Evolution*, *Proceedings of the National Academy of Sciences of the United States of America* and *Science Advances*, as well as solid disciplinary journals such as *Molecular Ecology*. And more are to come (submitted or in preparation).

Furthermore, we conducted a number of opportunistic side-projects related to the project but that go beyond what was specifically proposed. We notably re-described one overlooked species and described a new species from the Gulf of Mexico. We also took a conservation genetics approach to study the rare endemic Maya hamlet in Belize, whose conservation status was changed from Vulnerable to Endangered by the International Union for the Conservation of Nature (IUCN) on the basis of our study. In addition, we investigated the differences in chromosomal recombination between egg and sperm cells in simultaneous hermaphrodites, temporal variation in reef fish communities, and aggressive mimicry as a potential source of natural selection on colour pattern. Finally, we leveraged the fact that most of the DNA sequencing was done from gill tissue samples to conduct a large-scale metagenomic analysis of the gill microbiome. This allowed us to assemble 67 microbial genomes, most of which belong to new species and one of which could not be assigned to any known microbial family.

3. Progress report

Background. The hamlets are group of coral reef fishes from the greater Caribbean that differ essentially in terms of colour pattern (Figure 1). They are otherwise very similar ecologically and live in the same habitat, with up to nine species observed on a single reef. They pair and spawn on a daily basis throughout the year. The mating displays and spawnings can be observed on SCUBA, which provides a direct window on reproductive isolation. Different species can be observed pairing and spawning at the same time and in the same aera, but pairings and spawnings are strongly assortative (i.e. occur mostly between members of the same species). Inter-specific spawnings occur at a low frequency (< 2%), and our genetic data indicate that hybrids i. are viable and ii. backcross to parental species, i.e. are fertile (Hench *et al*. 2019, 2022). In this respect the hamlets have the characteristics of good species, i.e. they show clear morphological differences (colour pattern) and strong reproductive isolation. Nevertheless, they present exceptionally low levels of genetic divergence among species and form an extremely shallow radiation (Figure 2). They thereby provide a rare window on the early stages of genomic divergence, which allows to study speciation, the process through which new species arise, and the early stages of radiation, the process of biological diversification.

Figure 1. The 18 unambiguously recognized hamlet species: *a. Hypoplectrus unicolor* (Florida Keys) *b. H. castraguirrei* (Veracruz) *c. H. providencianus* (Belize) *d. H. puella* (Statia) *e. H. floridae* (Florida) *f. H. ecosur* (Alacran Bank) *g. H. liberte* (Haiti) *h. H. nigricans* (Belize) *i. H. atlahua* (Veracruz) *j. H. chlorurus* (Bonaire) *k. H. affinis* (Florida Keys) *l. H. randallorum* (Roatan) *m. H. aberrans* (US Virgin Islands) *n. H. guttavarius* (US Virgin Islands) *o. H. gummigutta* (Tobago) *p. H. indigo* (Belize). *q. H. gemma* (Florida) *r. H. maya* (Belize). From Puebla *et al*. 2022. All photographs by CJE and AME except *c* (with permission from Reef Fish Identification, New World Publications, © 2002, Paul Humann), *g* (with permission from Ken Marks) and *r* (with permission from Lisa Carne, © Fragments of Hope 2018, http://fragmentsofhope.org/). From Puebla *et al*. (2022).

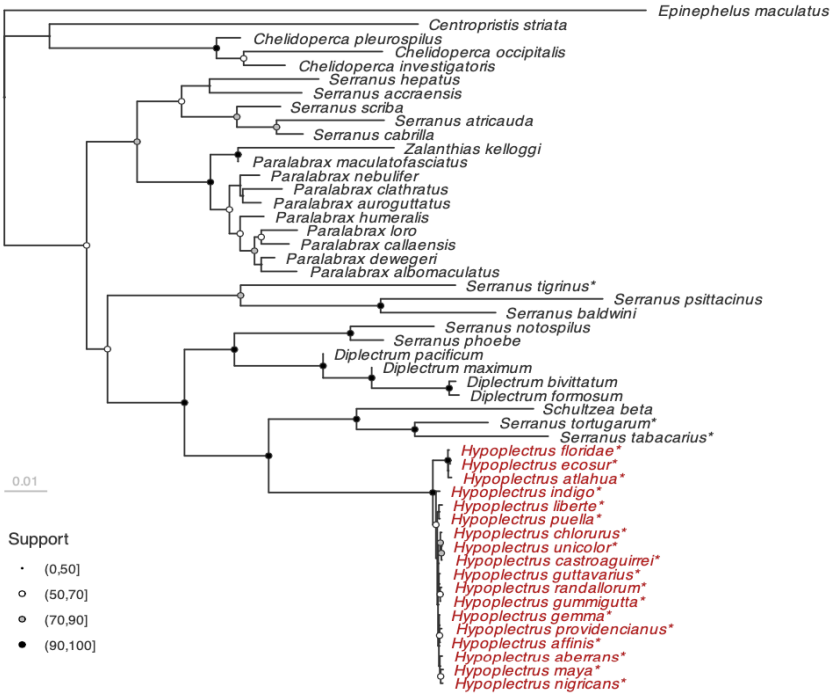

Figure 2. Maximum-likelihood phylogeny of the Serraninae subfamily based on 23 nuclear and mitochondrial genes. The hamlet radiation is highlighted in red and the species considered in our phylogeny are marked with an asterisk. Gene sequences for the other species were obtained from the Fish Tree of Life. Node point size and fill represent node support derived from 200 nonparametric bootstrap replicates. From Helmkampf *et al*. (2025).

Objective. The overarching objective of the project was to study the genomic bases of speciation and radiation using the hamlets as a model system.

Results. Our results show that the hamlets are extremely similar genetically throughout most of the genome, except for a few narrow regions (Hench *et al*. 2019, 2022; Helmkampf *et al*. 2025). Within these regions we identified so-called candidate “speciation genes” that underlie functional divergence among species (essentially colour pattern and mate choice). Our results also show that the hamlet radiation is on par with the fastest radiations on Earth, and shed light on the evolutionary process through which species can diversify (Hench *et al*. 2019, 2022; Helmkampf *et al*. 2025). These results largely confirm our working hypotheses, but they also show that although we identified the handful of large-effect genes that contribute to the hamlet radiation, these genes do not explain everything and we are still missing a potentially large number of small-effect genes. The details are provided in the project publications listed below. A specific objective of the project was to build a comprehensive phylogeny for the hamlets, which is now published (Helmkampf *et al.* 2025). This study is the result of a major effort with no less than 15 collaborators and includes 335 genomes from all named hamlet species. It does not only include resequenced genomes but also *de novo* assembled genomes using long read sequencing technology and transcriptomes from brain, retina and skin tissue. The results show that diversification and reproductive isolation can arise in the near-absence of phylogenetic signal, both genome-wide and in any part of the genome. This implies that the phylogenetic signal that we observe in older radiations may have little to do with the initial burst of diversification and reproductive isolation.

Following the project proposal we also teamed up with Dr. David Parichy from the University of Virginia. He knocked down our major candidate gene in zebrash using CRISPR-Cas but observed no major effect, which we think is due to the considerable phylogenetic distance between zebrafish and the hamlets. We are now in the process of establishing the breeding of reef fishes in our experimental facilites at ZMT to do this in closely related species. I finally note that the last batch of samples (from Bermuda) was collected towards the end of the project. These have been sequenced with the remaining funds from the project and will be the focus of an additional, stand-alone manuscript.

Contributions. This is a single-investigator project and I have led all the publications as either first or last author, but this does not imply that this work is a single-person effort, quite the contrary. The project-related publications include close to 30 co-authors, not counting the people who contributed in one way or another but are not co-authors (listed in the Acknowledgements sections of the publications). It would be too long to list the individual contribution of each collaborator here but it is important to highlight the key contribution of Dr. Kosmas Hench and Floriane Coulmance, the two PhD students hired on this project, Dr. Martin Helmkampf, the bioinformatician and Senior Scientist in my workgroup, Dr. Owen McMillan, our main collaborator at the Smithsonian Tropical Research Institute in Panama, as well as all our other collaborators in Panama, Colombia, Puerto Rico and Mexico.

Data handling and accessibility. All the genomes generated in this project were uploaded in the European Nucleotide Archive and are publicly available (accession numbers provided in the publications). The scripts that we used for data analysis are also publicly available, allowing to repeat our analyses from raw data to figures (links to the script repositories provided in the publications). Fish photographs were uploaded in Pangea.

Method development. One limitation of our early studies was that colour pattern is a complex trait that is difficult to quantify objectively. Furthermore, the colour pattern of the hamlets are altered when the fishes are kept in aquaria or preserved. In order to address this issue, we teamed up with Dr. Derya Akkaynak, an engineer trained at the Massachusetts Institute of Technology (MIT) and Woods Hole Oceanographic Institution (WHOI) specialized in the analysis of colour pattern in the marine environment, and Dr. Yann Lepoul, an image analysis specialist, to develop a procedure to analyse colour pattern in a fully automated and standardized way. This involved the development of a portable underwater photo studio to take standardized photographs *in situ*, a procedure to standardize colour across photographs, and a method to align all the images exactly so that each pixel of each fish is comparable to the same pixel in other fishes (Figure 3). This allowed us to detect associations between colour

pattern variation and genetic variation with pixel and DNA base pair resolution, respectively (Coulmance *et al*. 2024).

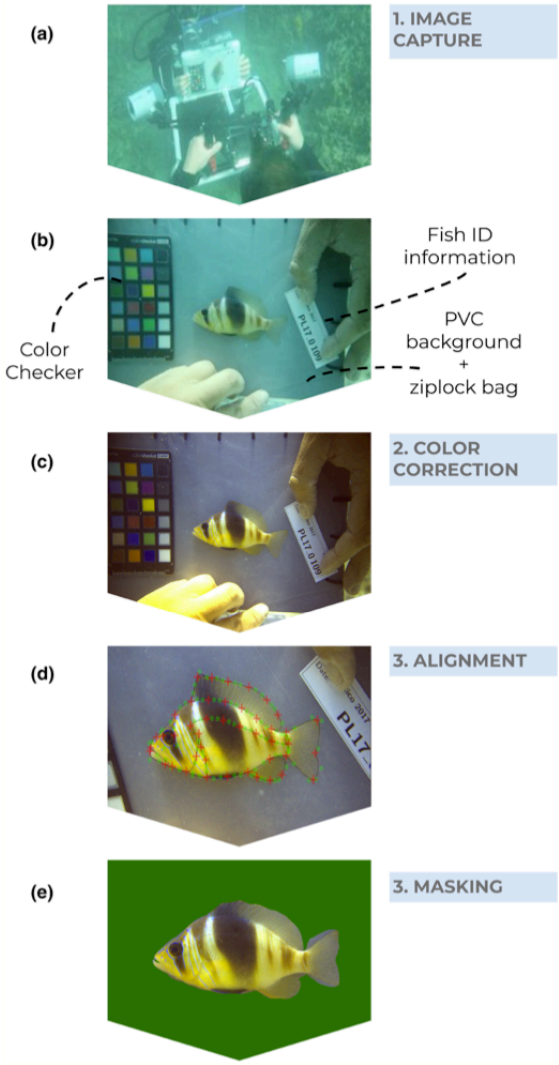

Figure 3. Overview of the image capture and standardization procedure. The fishes are captured *in situ*, photographed against a neutral background next to a colour checker and an ID label at a standardized distance and orientation from the camera lens. The fish images are then colour-standardized, aligned and masked. From Coulmance *et al*. (2024).

Furthermore, we conducted a number of opportunistic side-projects related to the project but that go beyond what was proposed specifically. We notably re-described one overlooked species (*Hypoplectrus affinis*, Puebla *et al*. 2022) and described a new species (*Hypoplectrus espinosaperezei*, Puebla *et al*. 2025) from the Gulf of Mexico (Figure 4). We also took a conservation genetics approach to study the rare endemic Maya hamlet in Belize (*Hypoplectrus maya*, Moran *et al*. 2019), whose conservation status was changed from Vulnerable to Endangered by the International Union for the Conservation of Nature (IUCN) on the basis of our study. In addition, we investigated the differences in chromosomal recombination between egg and sperm cells in simultaneous hermaphrodites (Theodosiou *et al*. 2016), temporal variation in reef fish communities (Hench et al. 2017), and aggressive mimicry as a potential source of natural selection on colour pattern (Puebla *et al*. 2018). Finally, we leveraged the fact that most of the DNA sequencing was done from gill tissue samples to conduct a large-scale metagenomic analysis of the gill microbiome (Abdelghany *et al*. 2025). This allowed us to assemble 67 microbial genomes, most of which belong to new species and one of which could not be assigned to any known microbial family.

Figure 3. Left (first two columns): the Campeche Bank hamlet (*H. espinosaperezei* sp. nov.). A black blotch covers the caudal peduncle and extends over the posterior part of the body, sometimes over the posterior border of the dorsal fin. Note the consistency of the thin vertical lines. Photographs from Alacranes reef by Carlos and Allison Estapé (*a*, *b*, *c*, *d*), Isai Dominguez Guerrero (*e*) and Alfonso Aguilar-Perera (*f*). From Puebla *et al*. (*under review*). Left (last two columns): *Hypoplectrus affinis* (Poey, 1861). *a*. neotype from Bocas del Toro, Panama *b*. Guna yala, Panama, *c*. Bocas del Toro Panama, *d*. Alacran reef, Campeche, Mexico *e*. & *f*. (juvenile) Florida Keys *g*. Grand Cayman (spawning pair, note blanched coloration) *h*. Bonaire. Photographs *a*, *b* and *c* by Oscar. Puebla, *d* by Isai Dominguez Guerrero, *e* and *f* by Carlos J. Estapé and Allison M. Estapé, *g* by Frank Krasovec and *h* by Kim White, all with permission. From Puebla *et al*. (2022).

Science communication. We plan to write a science communication piece on all this research, including what was done prior to, at the margin of, and outside of this specific project. But before this we still have 12 manuscripts in the pipeline for publication.

4. Published Project Results

4.1 Published publications with scientific quality assurance

Coulmance F, Akkaynak D, Le Poul Y, Höppner MP, McMillan WO, **Puebla O** (2024) Phenotypic and genomic dissection of color pattern variation in a reef fish radiation. *Molecular Ecology* 33, e17047. https://doi.org/10.1111/mec.17047

Helmkampf M, Coulmance F, Heckwolf MJ, Acero AP, Balard A, Bista I, Domínguez-Domínguez O, Frandsen PB, Torres-Oliva M, Santaquiteria A, Tavera J, Victor BC, Robertson DR, Betancur-R R, McMillan WO, **Puebla O** (2025) Radiation with reproductive isolation in the near-absence of phylogenetic signal. *Science Advances* 11, eadt0973. https://doi.org/10.1126/sciadv.adt0973

Hench K, Helmkampf M, McMillan WO, **Puebla O** (2022) Rapid radiation in a highly diverse marine environment. *Proceedings of the National Academy of Sciences of the United States of America* 119, e202045711922. https://doi.org/10.1073/pnas.2020457119

Hench K, McMillan WO, Betancur-R R, **Puebla O** (2017) Temporal changes in hamlet communities (*Hypoplectrus* spp, Serranidae) over 17 years. *Journal of Fish Biology* 91, 1475–1490. https://doi.org/10.1111/jfb.13481

Hench K, Vargas M, Höppner MP, McMillan WO, **Puebla O** (2019) Inter-chromosomal coupling between vision and pigmentation genes during genomic divergence. *Nature Ecology & Evolution* 3, 657–667. https://doi.org/10.1038/s41559-019-0814-5

Moran BM, Hench K, Waples RS, Höppner MP, Baldwin CC, McMillan WO, **Puebla O** (2019) The evolution of microendemism in a reef fish (*Hypoplectrus maya*). *Molecular Ecology* 28, 2872–2885. https://doi.org/10.1111/mec.15110

Puebla O, Coulmance F, Estapé CJ, Estapé AM, Robertson DRR (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. https://doi.org/10.11646/zootaxa.5093.2.1

Puebla O, Aguilar-Perera A, Helmkampf M, Robertson DR, Estapé CJ, Estapé AM, Domínguez-Domínguez O (2025) *Hypoplectrus espinosai* sp. nov. (Teleostei: Serranidae), a new hamlet on coral reefs in the southwestern Gulf of Mexico. *Zootaxa* 5618, 509–524. https://doi.org/10.11646/zootaxa.5618.4.3

Puebla O, Picq S, Lesser JS, Moran B (2018) Social-trap or mimicry? An empirical evaluation of the *H. unicolor – C. capistratus* association in Bocas del Toro, Panama. *Coral Reefs* 37, 1127–1137. https://doi.org/10.1007/s00338-018-01741-0

Theodosiou L, McMillan WO, **Puebla O** (2016) Recombination in the eggs and sperm in a simultaneously hermaphroditic vertebrate. *Proceedings of the Royal Society B* 283, 20161821. https://doi.org/10.1098/rspb.2016.1821

Submitted manuscript

Abdelghany S, Helmkampf M, Schechter, Veseli IA, Leray M, Eren AM, **Puebla O** (2025) Proteobacteria with chemosynthetic potential are highly enriched in the gills of *Hypoplectrus* reef fishes. *bioRxiv*, July 2 2025. https://doi.org/10.1101/2025.07.01.662686

Manuscript in preparation (advanced)

Coulmance F, Heckwolf MJ, Gismann J, Kafle T, Domínguez-Domínguez O, Helmkampf M, McMillan WO, **Puebla O** (*in preparation*). Assortative mating, phenotypic variation and speciation.

5 Further information on the project, qualifications and outlook

The project went well. It took longer than anticipated but the cost-neutral extensions allowed to finalise pretty much everything that was proposed. Three years is way too short for a PhD though. We (both PhD students and advisor) really struggle at the end. The funding for PhDs should be 4 years. In this regard I note that the second PhD student hired on this project (Floriane Coulmance) obtained a Ross Robertson Fellowship from the Smithsonian Tropical Research Institute in Panama. She used this additional support to collect extra photographs and tissue samples in Panama, Tobago, the US Virgin Islands and Veracruz that will be the focus of another manuscript on phenotypic and genetic variation that goes beyond what was proposed in the project (Coulmance *et al*, *in prep*). She landed a postdoctoral position in the Nicolas Salamin group at the University of Lausanne in Switzerland to work on clownfishes.

5.1 Doctoral researchers involved:

Hench, Kosmas, m, graduated, March 2017 – September 2020, funded by this project March 2017 – March 2020

Coulmance, Floriane, f, thesis submitted, September 2020 – present (December 2024), funded by this project September 2020 – August 2024 with a one-year pause between September 2022 and August 2023.