Homoploid Hybrid Speciation in a Marine Pelagic Fish *Megalaspis cordyla* (Carangidae)

NOZOMU MUTO[1*], YONG-CHAO SU[2], HARUTAKA HATA[3], NGUYEN VAN QUAN[4], VEERA VILASRI[5], MAZLAN ABD. GHAFFAR[6] AND RICARDO P. BABARAN[7]

[1]Department of Marine Biosciences, Kitasato University, Sagamihara, Kanagawa 252-0373, Japan;

[2]Department of Biomedical Science and Environmental Biology, Kaohsiung Medical University, Kaohsiung City, 80708, Taiwan;

[3]National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA;

[4]Institute of Marine Environment and Resources, Vietnam Academy of Science and Technology, Hai Phong City, Vietnam;

[5]Office of Natural Science Research, National Science Museum, Pathum Thani 12120, Thailand;

[6]Institute of Climate Adaption and Marine Biotechnology, Universiti Malaysia Terengganu, Kuala Nerus 21030, Malaysia;

[7]College of Fisheries and Ocean Sciences, University of the Philippines Visayas, Miagao Iloilo 5023, Philippines

**Correspondence to be sent to: Department of Marine Biosciences, Kitasato University, Sagamihara, Kanagawa 252-0373, Japan; E-mail: muto.nozomu@kitasato-u.ac.jp / nzmutou@gmail.com.*

Abstract.—Homoploid hybrid speciation (HHS) is an enigmatic evolutionary process where new species arise through hybridization of divergent lineages without changes in chromosome number. Although increasingly documented in various taxa and ecosystems, convincing cases of HHS in marine fishes have been lacking. This study presents evidence of HHS in Torpedo scad *Megalaspis cordyla* based on comprehensive genomic, morphological, and ecological analyses. A Principal Component Analysis using thousands of SNPs identified three sympatric clusters in the western Pacific. Genome-wide differentiation between the clusters and the admixed nature of a cluster between the others were evident from population genomic analyses, species tree estimation, mitochondrial DNA divergence, and tests of introgression. Multiple statistical methods for hybrid detection also supported the admixed ancestry of this cluster. Moreover, model-based demographic inference favored a hybrid speciation scenario over introgression. Examination of occurrence data and ecologically relevant morphological characters suggested ecological differences between the clusters, potentially contributing to reproductive isolation and niche partitioning in sympatry. The clusters are morphologically distinguishable and thus can be taxonomically recognized as separate species. The hybrid cluster is restricted to the coasts of Taiwan and Japan, where all three clusters coexist. The parental clusters are additionally found in lower latitudes such as the coasts of the Philippines, Vietnam, Thailand, and Malaysia, where they display non-overlapping distributions. Given the geographical distributions, estimated times of the species formation, and patterns of historical demographic changes, we propose that the Pleistocene glacial cycles were the primary driver of HHS in

this system. Based on this argument, we develop an ecogeographic model of HHS in marine coastal ecosystems, including a novel hypothesis to explain the initial stages of HHS.

Key words: climate change; ecological divergence; hybridization; latitudinal gradient; morphology; Pleistocene; phylogeography; reproductive isolation]

Running head: HYBRID SPECIATION IN A MARINE PELAGIC FISH

Hybridization can drive adaptation and species diversification. One possible outcome of hybridization is hybrid speciation, where interbreeding between diverged species results in a new population evolving independently from the parental species (Coyne and Orr 2004; Mallet 2007; Peñalba et al. 2024). In allopolyploid hybrid speciation, a hybrid population arises with whole-genome duplication, reported frequently in plants (Wood et al. 2009). In contrast, homoploid hybrid speciation (HHS) occurs without changes in ploidy, with the nascent hybrid population possessing a recombinant genome of parental species (Mallet 2007). There is much debate about the criteria for a case to be regarded as HHS. An emerging view is that at least the first and second of the following criteria must be satisfied, whereas a more stringent definition requires the third as well: 1) the species arose via hybridization with detectable genomic signatures, 2) intrinsic or extrinsic reproductive isolation from parental species has been established, and 3) the reproductive barriers directly resulting from hybridization (Abbott et al. 2013; Schumer et al. 2014; Nieto Feliner et al. 2017; Long and Rieseberg 2024). Cases of HHS fulfilling the first two criteria are increasingly documented. However, it is still rare compared to other hybridization-related phenomena, such as the formation of hybrid zones and introgression. Moreover, reported cases are strongly biased towards terrestrial and freshwater organisms (Abbott et al. 2013; Yakimowski and Rieseberg 2014; Taylor and Larson 2019 and references therein).

A major challenge in HHS is the persistence of the nascent hybrid population. Immediately after formation, it faces ecological competition with parental species and the swamping effects of gene flow (Buerkle et al. 2000). The hybrid

population can overcome both and persist on an evolutionary timescale by acquiring a novel niche not occupied by the parental species. The acquisition of a novel niche drives HHS in several ways. First, it may directly cause extrinsic pre- or postzygotic reproductive isolation between the hybrid and parental species (Rieseberg et al. 2003; Schwarz et al. 2005). Second, it enables the hybrid and parental species to coexist in sympatry by alleviating competition through niche partitioning (Lamichhaney et al. 2018; Olave et al. 2022). Third, it may allow the hybrid species to occupy a spatially separated habitat from the parental species, facilitating allopatric or parapatric divergence (Donovan et al. 2010; Brennan et al. 2012; Lopes et al. 2023). Supporting the importance of the third role, the independence of the hybrid species in most compelling cases of homoploid hybrid speciation seems to rely on geographical separation from at least one parental species (Rieseberg et al. 2003; Mavárez and Linares 2008; Larsen et al. 2010; Brennan et al. 2012; Kang et al. 2013; Lamichhaney et al. 2018; Maier et al. 2019; Noguerales and Ortego 2022; but see Olave et al. 2022 and Rosser et al. 2024 for counterexamples). Theory suggests that the recombinant hybrid genome can express transgressive phenotypes via additive and/or epistatic effects of parental genetic variants, leading to the exploitation of novel niches (Rieseberg et al. 2003; Gross and Rieseberg 2005; Mallet 2007; Abbott et al. 2013). However, in what ecogeographic settings the niche divergence leads to reproductive or geographical isolation remains largely unknown.

The Pleistocene glacial cycles significantly impacted the evolutionary histories of organisms at and below the species level across various ecosystems and taxa.

In terrestrial ecosystems, ice sheets formed in high altitudes and latitudes during glacial periods disrupted existing habitats, while glacial retreat during interglacial periods created new, unoccupied habitats. Coupled with temperature changes, these habitat dynamics resulted in allopatric divergence, secondary contact, altitudinal and latitudinal range shifts, population size changes, and local adaptation (Hewitt 2004). Mountain systems are among the most well-studied (Hewitt 2000, 2004), including a few possible cases of homoploid hybrid speciation (Gompert et al. 2006; Maier et al. 2019). Marine coastal areas are another ecosystem strongly influenced by historical climate changes. Eustatic sea level changes cyclically disrupted and opened habitats for marine organisms in shallow waters, creating barriers or corridors to gene flow (Bowen et al. 2016). Changes in water temperature and associated environmental factors caused latitudinal range shifts, population size changes, and adaptive responses, similar to those ongoing under contemporary climate change (Kiessling et al. 2012; Neiva et al. 2014; Donelson et al. 2019). These, in turn, resulted in various hybridization-related phenomena in coastal organisms via secondary contact: introgression (Duranton et al. 2020), formation of stable hybrid zones (Hirase et al. 2021), and divergence with gene flow (Muto and Kai 2023). Nonetheless, HHS in this environment has not been documented to date.

Here, we provide genomic and phenotypic evidence for HHS in the torpedo scad *Megalaspis cordyla* (Linnaeus, 1758) in the western Pacific, representing the first compelling case of this phenomenon in a marine pelagic fish. The torpedo scad is a pelagic marine predator in the family Carangidae (Teleostei), usually reaching 40 cm in total length. Widely distributed in tropical to temperate

130 coastal waters throughout the Indian Ocean and the western Pacific, it is an
131 important fisheries resource in the surrounding countries (Smith-Vaniz 1999).
132 Previous phylogeographic studies based on mitochondrial DNA (mtDNA)
133 sequences suggested historical vicariance followed by secondary contact in the
134 western Pacific region. Specimens from the southwestern part of the region (the
135 Pacific coast of the Asian continent, from Vietnam to Malaysia) and the
136 southeastern part (the islands of the Philippines and Indonesia) represent
137 reciprocally monophyletic lineages, whereas specimens from the northern part
138 (Taiwan) represent both lineages (Su et al. 2020; Muto et al. 2021). In the
139 present study, we provided evidence for genome-wide divergence between the
140 two lineages. We also discovered a third genetic cluster restricted to locations
141 where the two lineages are sympatric. We tested the hypotheses that this third
142 cluster originated from the admixture of the other two and that they represent
143 separate species, using detailed genomic, ecological, and morphological
144 analyses. Based on the results, we suggest that the Pleistocene glacial cycles
145 primarily drove HHS in the present system. Furthermore, we develop an
146 ecogeographic model of HHS in coastal marine environments, where
147 demographic dynamics under climate change and adaptation across latitudinal
148 environmental gradients play critical roles. This model bridges a major gap in
149 the understanding of how the initial stages of HHS proceed.
150

151 **MATERIALS AND METHODS**

152 A full description of methods is provided in Online Appendix 1.
153

Specimens

We analyzed a total of 160 specimens of *M. cordyla* collected from various localities in the western Pacific (Fig. 1a), where two distinct mtDNA lineages have been discovered (Su et al. 2020; Muto et al. 2021). For genotyping, we used tissues stored in natural history collections in respective countries (Table S1). Total genomic DNA was extracted from the fin or muscle tissue of each specimen using the Favorgen Tissue Genomic DNA Extraction Mini Kit (Favorgen). The specimens from Vietnam, Thailand, Malaysia, and the Philippines were analyzed in the previous study (Muto et al. 2021).

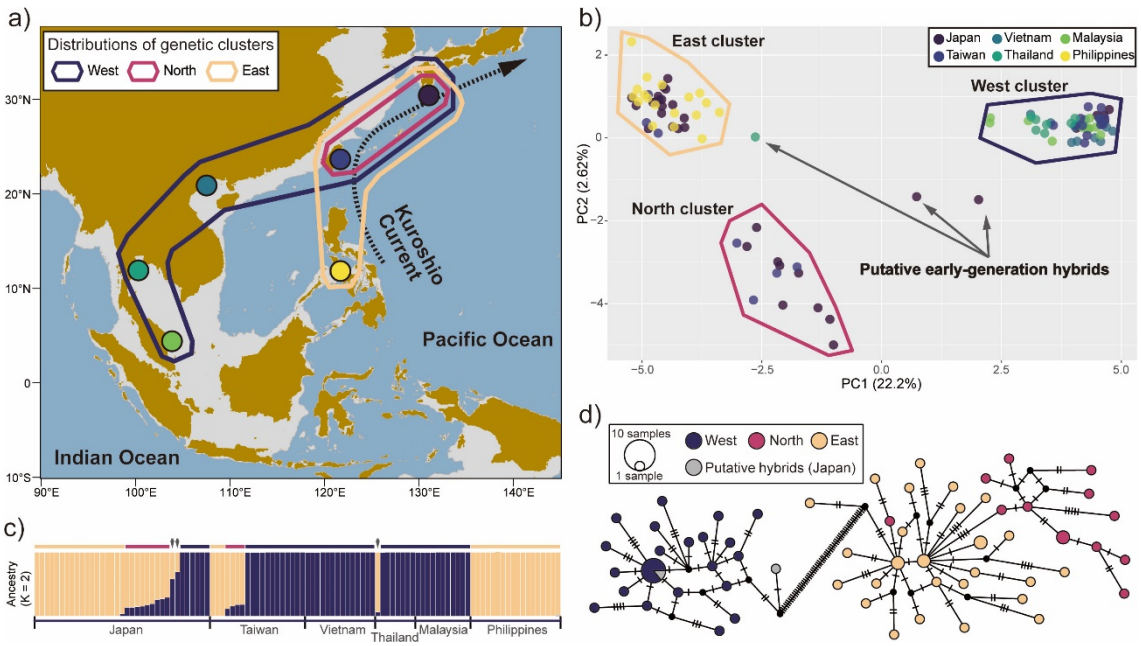

Fig. 1. a) Sampling sites (circles) and distributions of three genetic clusters (polygons) of *Megalaspis cordyla* in the western Pacific. The grey-shaded area represents regions currently shallower than 130 m, repeatedly exposed during Pleistocene glacial cycles. The dotted arrow indicates the approximate route of the Kuroshio Current. b) Scatter plots for the principal component analysis (PCA) based on the genome-wide SNPs, with polygons and arrows indicating the three genetic clusters and their putative early-generation hybrids, respectively. The clusters are named according to their geographical distributions. c) Result of the STRUCTURE analysis assuming the number of clusters (*K*) = 2, which was the optimal according to the delta *K* values. Horizontal bars and arrows indicate individual assignments to the three clusters and putative early-generation hybrids, respectively, based on the PCA. The color code for the horizontal bars matches that for the PCA plot polygons. Note that individuals in the North cluster and putative early-generation hybrids show

admixed ancestry. d) Statistically parsimonious network of haplotypes of the concatenated control region and cytochrome b gene sequences of the mitochondrial genome. Circle/pie-chart size is proportional to haplotype frequency. Black dots and a hatch mark along branches represent unobserved hypothetical haplotypes and one mutational step, respectively.

Mitochondrial DNA Genotyping

We obtained partial sequences of the mitochondrial cytochrome b gene (CYTB: 873 bp) from all specimens. Additionally, we obtained partial sequences of the following seven loci from a subset of the specimens: cytochrome c oxidase subunits 1 (COI: 612 bp) and 2 (COII: 657 bp), NADH dehydrogenase subunits 1 (ND1: 828 bp), 2 (ND2: 531 bp), 4 (ND4: 828 bp), and 5 (ND5: 801 bp), and the control region (CR: 760 bp) (Table S1). We amplified the COI and CYTB sequences following Ivanova et al. (2007) and Cárdenas et al. (2005), respectively. The remaining loci were amplified using newly developed primers (Table S2). We edited and aligned the sequences using BioEdit v7.0.5.3 (Hall 1999) and CLUSTAL X (Larkin et al. 2007).

Genome-Wide SNPs Genotyping

We genotyped 115 specimens using MIG-seq (Suyama and Matsuki 2015; Suyama et al. 2022), a reduced representation sequencing method targeting SNPs in the inter-simple sequence repeat region (ISSR). We performed the first PCR using 12 multiplex primers to amplify various ISSR across the genome (Suyama and Matsuki 2015; Suyama et al. 2022). The first PCR used 12 multiplex primers to amplify various ISSRs across the genome. The second PCR added Illumina sequencing adapters and a unique dual index. We pooled

the second PCR products, size-selected for 400–800 bp, purified, quality-checked, quantified, and sequenced on an Illumina Novaseq 6000 in 150 bp paired-end mode at Novogen Co. Ltd., China, yielding 148M reads.

We processed the raw MIG-seq reads by the single-end mode of fastp v0.20.1 (Chen et al. 2018), merging the processed reads 1 and 2 into a single FASTQ file, and mapped them to the *Caranx ignobilis* reference genome (GenBank JAFHLA000000000.1) (Pickett et al. 2022) using BWA-mem v0.7.17 (Li 2013) with default settings. Unmapped or low-quality reads were removed using Samtools v1.12 (Danecek et al. 2021). The resulting BAM files were used to prepare 5 datasets for downstream analyses (see the next paragraph). We also downloaded the whole-genome shotgun reads of *C. ignobilis* (a different individual from the reference) and *Caranx melampygus* from the NCBI Short Read Archive (accession nos.SRR1002875 and SRR1002876) (Santos 2014) as outgroups. We processed these reads using the paired-end mode of fastp, mapped them to the reference genome using BWA-mem with default settings, and removed unmapped, low-quality, or duplicate reads using Samtools. The resulting BAM files were used for the analyses requiring outgroups (datasets 2–5: see below).

The following 5 datasets were prepared for downstream analyses: 1) 2,926 unlinked biallelic SNPs in all available (105) specimens of *M. cordyla* without an outgroup for population genetic analyses using various tools and ancestry estimation using HIest v2.0 (Fitzpatrick 2012); 2) 6,789 unlinked biallelic SNPs in 24 specimens (8 per genetic cluster) with *C. ignobilis* as an outgroup for admixture detection using HyDe v1.0.0 (Blischak et al. 2018) and

ADMIXTOOLS v7.0.2 (Patterson et al. 2012); 3) 3,179 unlinked biallelic SNPs in 9 specimens (3 per genetic cluster) with *C. ignobilis* as an outgroup for divergence time estimation using SNAPP v1.5.0 (Bryant et al. 2012); 4) 6,341 biallelic SNPs in the same 24 specimens, polarized using the *C. ignobilis* and *C. melampygus* genomes as outgroups, for demographic model selection using fastsimcoal v2.7.0.9 (Excoffier et al. 2013, 2021); and 5) 22,326 biallelic SNPs (linked and unlinked) and 920,184 monomorphic sites in the same 24 specimens, polarized using the *C. ignobilis* and *C. melampygus* genomes as outgroups, for demographic parameter estimation using fastsimcoal 2 and historical population size estimation using Stairway Plot v2.1.2 (Liu and Fu 2020). Note that monomorphic sites and linked SNPs in the 5th dataset are needed for direct parameter estimations using fastsimcoal 2, whereas only unlinked SNPs are used for the model selection (Excoffier et al. 2013). For each dataset, we called SNPs using Stacks v2.5.4 (Catchen et al. 2013) or *mpileup* implemented in bcftools v1.8 (Danecek et al. 2021). We then filtered the SNPs and created final vcf files using VCFtools v0.1.16 (Danecek et al. 2011) and R package 'SNPfiltR' v1.0.0 (DeRaad 2022) on R v4.2.2 (R Core Team 2022).

We used *easySFS.py* (https://github.com/isaacovercast/easySFS) to calculate the unfolded site frequency spectrum (SFS) from the 4th and 5th datasets, assuming the reference alleles fixed in both outgroups represent the ancestral state. The resulting one-dimensional SFS from the 5th dataset was used for the historical population size estimation by Stairway Plot 2, whereas multi-dimensional SFS was used for the demographic inference by fastsimcoal 2. For input file preparations for the other analyses, we used PGDSpider

v2.1.1.2 (Lischer and Excoffier 2012) to convert the vcf files to appropriate formats.

Population Genetic Analyses

We constructed minimum-spanning haplotype networks of mtDNA sequences using PopART v1.7 (Leigh and Bryant 2015).

The following population genetic analyses were conducted using 1st dataset of the genome-wide SNPs. We performed a principal component analysis (PCA) using the 'adegenet' package v2.1.4 in R (Jombart 2008). We used STRUCTURE v2.3.4 (Pritchard et al. 2000) to identify admixed individuals with 10 replicate runs for each predefined number of genetic clusters (K) = 1–8, each with 100,000 Markov Chain Monte Carlo (MCMC) generations after 100,000 generations of burn-in. These runs were parallelized using StrAuto v1.0 (Chhatre and Emerson 2017). Statistical support for K values was evaluated by calculating the ΔK value (Evanno et al. 2005) using Structure Harvester v0.6.94 (Earl and vonHoldt 2012). The results were summarized and visualized with 'pophelper' v2.3.1 in R (Francis 2017). We used SplitsTree v5.3.0 (Huson and Bryant 2006) to build an unrooted phylogenetic network based on the Neighbor-Net algorithm.

The population genetic analyses showed one genetic cluster was genetically intermediate to the other two clusters (Fig. 1b, c). This intermediate cluster is hereafter called the "North" cluster, while the others are the "East" and "West" clusters, according to their geographical distributions (Fig. 1a). Putative early-generation hybrids (F1, F2, or backcross) between these clusters were also

detected. Excluding these putative hybrids, we calculated locus-specific pairwise F_{ST} between the clusters using *populations* implemented in Stacks, and pairwise F_{ST} across all loci and associated *p*-values using Arlequin v3.5.2.2 (Excoffier and Lischer 2010).

We estimated historical population size changes in each cluster using Stairway Plot v2.1.2 (Liu and Fu 2020) based on the 5th dataset. We assumed a mutation rate of $1E^{-8}$ per site per generation (Tine et al. 2014) and a generation time of 1.2 years (Froese and Pauly 2023). Other settings were default or those recommended by the author of the software.

Divergence Time Estimation

We estimated the divergence times of the three genetic clusters using genome-wide SNP dataset 3 and mtDNA sequence data. The former was done under the multispecies coalescent model using the SNAPP v1.5.0 package (Bryant et al. 2012) of BEAST v2.7.1 (Bouckaert et al. 2014), following Stange et al. (2018). Time calibration was based on the estimated node age of 6.349 (95% CI: 4.434-8.698) million years ago (Ma) for the most recent common ancestor (MRCA) of *C. ignobilis* and *M. cordyla* (Glass et al. 2023), approximated by a lognormal distribution with a mean of 6.349 Ma, a standard deviation of 0.205, and an offset of -0.01. We performed two independent runs of 50,000,000 MCMC generations, sampling every 10,000 steps and discarding the first 10% of the trees as burn-in, then combined the posterior tree distributions from the two runs. We estimated a maximum clade credibility tree using TreeAnnotator v2.7.1 (Bouckaert et al. 2014) and visualized the tree using

FigTree v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/). The estimation based on the concatenated sequences of the 8 mtDNA loci was done using BEAST v2.7.1. We partitioned the sequences by locus and estimated the substitution model for each partition using the bModelTest v1.3.3 (Bouckaert and Drummond 2017), which averages over different substitution models in BEAST. We assumed the optimized relaxed clock model and the Yule model for the clock model and the tree prior, respectively. We used the mitogenome sequence of *C. ignobilis* (accession No.: NC_022932) as an outgroup. Time calibration, MCMC simulations, convergence check, tree estimation, and visualization were the same as described above.

Ancestry Estimation and Admixture Detection

We estimated the ancestries of the individuals in the North cluster and the putative early-generation hybrids using the 'HIest' package v2.0 (Fitzpatrick 2012) in R. This program estimates the ancestries of target individuals relative to predefined parents via log-likelihood maximization, using target genotypes and parental allele frequencies. First, we assumed the East and West clusters as potential parents and used 108 unlinked SNPs with allele frequencies differing by >0.5 between them. Second, we estimated the ancestries of putative hybrids only, assuming the three clusters as potential parents, using 130 SNPs with allele frequencies differing by >0.5 between one or more pairs of the clusters. In the first analysis, we estimated the ancestry index (*S*) and the inter-class heterozygosity (*H*) using the *HIest* function, employing the SANN method for searching the likelihood surface, 500 MCMC generations, and the

option surf = TRUE. *S* represents the proportion of an individual's alleles derived from one predefined parental population (here, the East cluster), equivalent to the hybrid index. *H* indicates the proportion of loci with one allele from each parental population. Expected *H* values vary between classes of early-generation hybrids (F1: 1.0; F2 and backcross: 0.5), decreasing if succeeding reproduction occurs exclusively among hybrids and alleles are lost by drift. We also estimated the likelihoods for six possible genotype classes resulting from 2 generations of inter- or intraspecific crosses (parental, F1, F2, and backcrosses) using the *HIclass* function. We compared the maximum likelihood of the six classes with that of the maximum likelihood estimation of *S* and *H* using the *HItest* function. For the second analysis, we estimated the ancestry indices S_{East}, S_{West}, and S_{North} of the putative early-generation hybrids using the *threeway* function. These indices are equivalent to *S*, representing the proportion of alleles derived from the three predefined parental populations. We also calculated the likelihoods for 15 possible genotype classes in the first two generations of hybridization between pairs of parental populations using the *thirdclass* function and compared the maximum likelihood with that obtained from the *threeway* analysis using the *HItest3* function.

We tested if the North cluster originated from an admixture of the East and West clusters by three methods based on the 3rd dataset of the genome-wide SNPs. First, we used HyDe v1.0.0 (Blischak et al. 2018), which assesses the evidence for hybridization under a model of hybrid speciation and estimates the associated ancestry proportions in a putative hybrid population from two parental populations (γ and $1-\gamma$). We tested for hybridization at the population

level for each cluster, leaving the remaining clusters as putative parental populations, using the *run_hyde.py* script. We then performed 500 bootstrap resampling using the *bootstrap_hyde.py* script to detect individual variation in r within a hybrid population. Second, we calculated the f_3 statistic using 'admixr' v0.9.1 (Petr et al. 2019). The f_3 statistic is a three-population statistic based on branch lengths between populations. A significantly negative f_3 statistic indicates that a predefined outgroup population is a mixture of two ingroup populations (Patterson et al. 2012). We calculated the f_3 statistics designating each cluster as an outgroup. Third, we calculated the *D*-statistic using 'admixr' to test for introgression between the clusters. The *D*-statistic detects introgression between non-sister taxa under the user-defined topology by asking if the allele-sharing pattern between the taxa deviates from what is expected under incomplete lineage sorting without introgression (Patterson et al. 2012). Given the possible hybrid origin of the North cluster (see Results), two strictly bifurcating topologies were possible for the user-defined topology, i.e., one with a sister relationship between the East and North clusters, i.e., (East,North),West), and another between the West and North clusters, (West,North),East). These were referred to as topologies 1 and 2, respectively. We calculated the *D*-statistics for both.

Demographic Model Selection and Parameter Estimation

We inferred the demographic history of *M. cordyla* by testing alternative scenarios using fastsimcoal v2.7.0.9 (Excoffier et al. 2021). Given the admixed nature of the North cluster, we aimed to determine if it originated from an

369 admixture of the East and West clusters (“hybrid speciation” scenario) or an
370 instantaneous introgression event following divergence (“pulse introgression”).
371 We also considered a null scenario of strict bifurcation between the clusters
372 without a pulse introgression event (“bifurcation”), following Noguerales and
373 Ortego (2022). Alternative topologies 1 and 2, i.e., (East,North),West) and
374 (West,North),East), were assumed for the pulse introgression and bifurcation
375 scenarios, resulting in five scenarios. These scenarios were simulated with or
376 without gene flow between the clusters before (ancestral gene flow) and after
377 (contemporary gene flow) the formation of the North cluster, resulting in 20
378 models (Fig. S1).

379 We performed 100 independent fastsimcoal runs for each demographic
380 model using the 4th dataset for model selection. Each run comprised 100
381 expectation-conditional maximization (ECM) cycles with 100,000 coalescent
382 simulations. The run with the smallest difference between the maximum
383 expected likelihood and the maximum observed likelihood was considered the
384 best fit. For model selection, we compared 1) Akaike’s information criterion
385 (AIC) based on the best-fit run of each model, and 2) the likelihood distributions
386 from 100 expected SFS, generated by 1,000,000 coalescent simulations under
387 the parameters estimated from the best-fit run (Meier et al. 2017). Under the
388 selected model, we then estimated demographic parameters using the 5th
389 dataset. To obtain 95% confidence intervals (CIs) for each parameter, we
390 generated 100 parametric bootstrap replicates of SFS under the maximum
391 likelihood parameters of the best-fit run following the program manual. The
392 parameters were re-estimated using 100 independent runs for each

bootstrapped SFS with the same settings, except that the initial values were set to the maximum likelihood parameters of the best-fit run, and the number of ECM cycles was set to 50. We rescaled the estimated parameters assuming a mutation rate of $1E^{-8}$ per site per generation and a generation time of 1.2 years (see above).

Morphology

We examined 37 linear measurements and 17 countable characters in 37 specimens from Japan, Taiwan, Vietnam, and Thailand (Tables S1, S3). This analysis aimed to determine if the genetic clusters 1) differ in ecologically relevant morphological characters, suggesting ecological divergence, and 2) can be recognized as separate species in a taxonomic sense. Linear measurements were made to the nearest 0.01 mm using a digital vernier caliper. Methods for measuring and counting generally followed Marr and Schaefer (1949), Gibbs and Collette (1967), and Muto et al. (2016). Linear measurements are expressed as % of the Standard length (SL). To illustrate the variation in overall morphology, we performed a PCA of 24 characters with no missing data using the R package 'FactoMineR' v2.11 (Lê et al. 2008). All linear measurements were corrected for allometry using the R package 'GroupStruct' v0.1.0 (Chan and Grismer 2022) before performing the PCA.

Ecology

To investigate potential ecological divergence between the genetic clusters, we re-analyzed the ecological data of *M. cordyla* in Taiwan reported by Su et al.

(2020). The data include sampling season and location, individual body length and weight, gonad weight, gonad somatic index (GSI: [gonad weight/body weight] × 100), and sex (Table S4), along with the mtCR sequences of the specimens. Su et al. (2020) compared the ecological factors between two major lineages of mtCR sequences. In the present study, we show that one of these two major lineages corresponded to the West cluster, while the other corresponded to the East and North clusters (see Results). Additionally, we demonstrate that the East and North clusters can be distinguished by a minor mtCR divergence within the major lineage. These observations allowed us to reclassify the specimens examined by Su et al. (2020) into the three clusters based solely on the mtCR sequences reported by the authors, albeit with some uncertainty. First, we tested the association between the ecological factors and genetic clusters under the Random Forest model using the R package ‘ranger’ v0.16.0 (Wright and Ziegler 2017). We assessed the importance of each ecological factor using the mean decrease in accuracy and estimated the significance of this importance metric with a permutation test using the R package ‘rfPermute’ v2.5.2 (Archer 2021). Second, we performed a hierarchical cluster analysis to create a dendrogram of the specimens based on the ecological data. We calculated the pairwise Gower’s distance between each specimen and constructed Ward’s hierarchical agglomerative clustering using the R package ‘stats’ v4.2.2 (R Core Team 2022). Third, we performed a Factor Analysis of Mixed Data (FAMD) using the R package ‘FactomineR’.

RESULTS

Population Genetic Analyses

The first two PCA axes separated the specimens into three distinct clusters: 1) individuals from Japan to Malaysia along the east coast of the Eurasian continent, 2) individuals from Japan to the Philippines along the Kuroshio Current, and 3) individuals from Japan and Taiwan (Fig. 1a, b). These clusters are referred to as the West, East, and North clusters, respectively, based on their geographical distributions. The East and West clusters occupied opposite positions along PC 1, with the North cluster in an intermediate position. Principal component 2 further separated the North cluster from the others. Two individuals from Japan were intermediate between the North and West clusters and are hereinafter referred to as putative early-generation hybrids. One individual from Thailand was plotted near the East cluster, while all other individuals from Thailand were included in the West cluster. This individual from Thailand is also regarded as a putative early-generation hybrid. In the STRUCTURE analysis, the most likely number of *K* was 2 (Fig. S2). Assuming *K* = 2, individuals of the North cluster showed signatures of admixture between the East and West clusters, with a greater ancestry contribution from the former (Fig. 1c). The putative early-generation hybrids from Japan showed intermediate ancestry between the North and West clusters, while that from Thailand was intermediate between the North and East clusters. Assuming *K* = 3–8, individuals in the North cluster showed admixed ancestry from three clusters: two corresponding to the East and West clusters, and an additional minor cluster (Fig. S2). The Neighbor-Net network showed similar patterns regarding the separation of the East, West, and North clusters, the intermediate

465 nature of the North cluster, and the positions of the putative early-generation
466 hybrids (Fig. S3). Pairwise F_{ST} across all SNPs was greatest in the East-West
467 comparison, followed by the West-North and East-North comparisons (0.2989,
468 0.2167, and 0.0861, respectively); all comparisons were significant (Table S5).
469 Significant locus-by-locus pairwise F_{ST} was most frequently observed in the
470 East-West comparison, followed by the West-North and East-North
471 comparisons (Fig. S4). Several SNPs showed significant differentiation in both
472 East-North and West-North comparisons, highlighting the unique evolutionary
473 history of the North cluster at these loci. In terms of allele frequency, a total of
474 130 SNPs were highly differentiated (i.e., differences >0.5) between one or
475 more cluster pairs (Fig. S5; Table S6). The majority of these SNPs (95 loci)
476 were highly differentiated in the East-West comparison, with the North cluster
477 showing intermediate allele frequencies. For 56 of these 95 SNPs, the North
478 cluster's allele frequency was closer to the East cluster, while for others, it was
479 closer to the West cluster, suggesting that alleles from the East and West
480 clusters were alternatively sorted in the North cluster following admixture. In
481 contrast, the North cluster was highly differentiated from both the East and West
482 clusters at 4 SNPs, due to the alleles reaching high frequency in the North
483 cluster while being almost absent in the East and West (0.56–0.67 in the North
484 vs. 0–0.05 in East and West).
485 The population sizes of the East and West clusters were inferred to have
486 declined approximately 70–100 thousand years ago (Ka), while the North
487 cluster remained stable (Fig. S6). Subsequently, all three clusters expanded in
488 population size towards the present time.

mtDNA Differentiation

The minimum spanning network based on concatenated sequences of CYTB and CR (873 + 760 bp) consisted of three major haplogroups, each corresponding to one of the three genetic clusters (Fig. 1d). We refer to these haplogroups as the East, West, and North haplogroups, corresponding to their respective clusters. The East-West and East-North divergences of the haplogroups were equivalent to 58 and 4 mutations, respectively. However, the East and North clusters were not reciprocally exclusive: one individual of the North cluster was included in the East haplogroup. Moreover, the East-North divergence disappeared when only the CYTB sequences or concatenated sequences of the seven protein-coding genes were analyzed (Fig. S7), indicating the divergence is due to variation in CR sequences. The putative early-generation hybrids from Japan were included in the West haplogroup (Fig. 1d). We were unable to obtain the CR sequence for the putative hybrid from Thailand; this individual was included in the “East + North” haplogroup in the CYTB-only network (Fig. S7).

Ancestry Estimation and Admixture Detection

In the *HIest* analysis, assuming the East and West clusters as potential parents, *S* and *H* for individuals in the North cluster were 0.51–0.76 and 0.10–0.35, respectively (Fig. 2a). The values of *S* suggest a greater ancestry contribution from the East cluster, while those of *H* suggest that the individuals are later-generation hybrids (>F_2), having exclusively reproduced within the

same cluster. *HIclass* assigned each individual in the North cluster to either backcross to the East cluster or F_2, but the likelihood ratio test by *HItest* rejected these simple assignments in favor of the more nuanced results of *HIest* (Table S7). Compared to the North cluster individuals, the putative early-generation hybrids from Japan showed lower *S* (i.e., more closely related to the West cluster) and higher *H*, whereas the hybrid from Thailand showed greater *S* and lower *H*. *HIclass* assigned them to F_2 and backcross to the East cluster, respectively, but *HItest* again rejected these assignments (Fig. 2a). In the *threeway* analysis, where the three clusters were assumed as potential parents, the putative early hybrids from Japan and Thailand showed mixed ancestries predominantly between the North and West clusters and between the North and East clusters, respectively (Table S8). *Thirdclass* assigned them to F_2 between the North and West clusters and F_2 between the North and East clusters, respectively; *HItest3* did not reject these assignments.

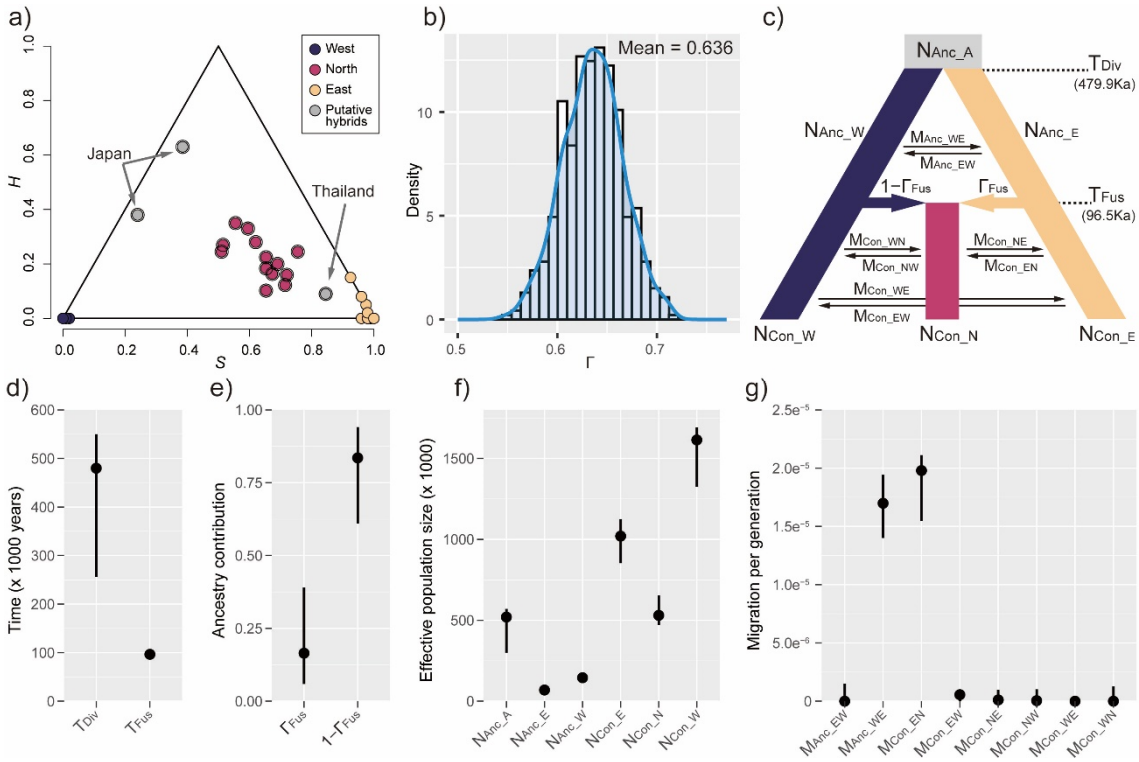

Fig. 2. a) Estimates of ancestry index (*S*), equivalent to hybrid index, and inter-class heterozygosity (*H*) by HIest, assuming the East and West clusters as potential parents. Individual assignments to the three clusters and putative early-generation hybrids are based on the PCA (Fig. 1b). Double circles indicate specimens for which simple classification as one of 6 genotype classes within 2 generations of inter- or intraspecific crosses (parentals, F1, F2, and backcrosses) was rejected by *HItest*. See Table S8 for results assuming the three clusters as potential parents. b) Density plot of estimated gamma (Γ) values for the North cluster across 500 individual bootstraps in HyDe. Γ and 1-Γ represent estimated ancestry contributions from the East and West clusters, respectively. c) A schematic diagram of the best-fit demographic model, i.e., hybrid speciation with ancestral and contemporary gene flow between clusters, tested using Fastsimcoal 2. Parameters include; N: ancestral ($_{Anc}$) or contemporary ($_{Con}$) effective population size of the common ancestor ($_{_A}$) or the three genetic clusters ($_{_E}$ / $_{_W}$ / $_{_N}$); T: time of divergence ($_{Div}$) or fusion ($_{Fus}$) of the East and West clusters; Γ: ancestry contribution from the East (Γ_{Fus}) or West ($1-\Gamma_{Fus}$) cluster upon fusion; M: ancestral ($_{Anc}$) or contemporary ($_{Con}$) gene flow between clusters, with the letters following the underscore indicating the direction of the gene flow (e.g., M_{Con_EW} represents contemporary gene flow from the East to West cluster forward in time). See Fig. S1 for all 20 models tested and Fig. S9 and Table S10 for model evaluation by likelihood and Akaike's Information Criterion values. d)–g) Demographic parameters estimated using Fastsimcoal 2 under the best-fit model. Dots and vertical bars represent maximum likelihood point estimates and 95% confidence intervals based on 100 bootstrap replicates. Timings of demographic events are in units of years. Effective population sizes are in units of the number of alleles. Migration rates are in units of proportions of alleles moving per generation. See Table S11 for tabulated values.

Admixture detection using HyDe was significant only for the North cluster (*Z*-score = 2.584, *p* = 0.0049; Table S9), supporting its hybrid origin. Bootstrap resampling of the individuals in the North cluster showed an unimodal distribution of γ values, indicating uniform admixture across individuals (Fig. 2b). This uniform admixture is consistent with hybrid speciation rather than ongoing introgression (Blischak et al. 2018). The departure of the γ values from 0.5 (mean = 0.636; 95% CI = 0.633–0.638) suggested a greater ancestry contribution from the East cluster. The f_3 statistic was negative when the North cluster was designated as an outgroup, supporting its admixed ancestry (f_3 = -0.023, *Z*-score = -4.338; Table S9). Finally, the *D*-statistics detected gene flow between both West-North and East-North combinations (West-North: *D* = 0.185, *Z*-score = 7.079; East-North: *D* = 0.296, *Z*-score = 11.389; Table S9).

Divergence Time Estimation

The East and North clusters formed a single clade, leaving the West cluster in another clade, in both the species tree based on genome-wide SNPs and the mtDNA phylogeny (Fig. S8). The MRCA age of the three clusters was estimated to be 0.269 Ma (95% CI: 0.168–0.387) based on genome-wide SNPs and 1.333 Myr (0.244–3.208) based on mtDNA phylogeny. The divergence between the North and East clusters was estimated at 0.126 Ma (0.075–0.184) and 0.196 Ma (0.015–0.574), respectively. All estimated divergence ages fell within the Pleistocene. Note that the East and North clusters were not reciprocally monophyletic in the mtDNA phylogeny; the divergence time mentioned above is

the node age for two clades approximately corresponding to the East and North clusters.

Demographic Model Selection and Parameter Estimation

The hybrid speciation model with gene flow between clusters both before and after the formation of the North cluster (i.e., ancestral and contemporary gene flow) was the best fit among the 20 models tested, based on the highest log-likelihood and lowest AIC (Fig. 2c; Fig. S9; Table S10). The other three hybrid speciation models, differing in ancestral and contemporary gene flow, were not better supported than the pulse-introgression models assuming a sister relationship between the West and North clusters (topology 2) (Fig. S9; Table S10). The models assuming topology 2 were favored over the alternative topology 1 if accompanied by a pulse introgression event or contemporary gene flow.

Under the best-fit model, the divergence and fusion of the East and West clusters were estimated to have occurred 479.9 Ka (95% CI: 256.1–549.9) and 96.5 Ka (94.9–106.4), respectively (Fig. 2d; Table S11), both within the Pleistocene. The ancestry contribution from the East cluster (Γ_{Fus}) upon the fusion was less than that from the West cluster ($1-\Gamma_{Fus}$) (Fig. 2e). Effective population size was greater in the West than in the East cluster, both before and after the fusion, with both clusters expanding over time (Fig. 2f). Contemporary population size was smallest in the North cluster, consistent with its restricted geographical range. Gene flow between the clusters was highly asymmetric; contemporary gene flow from the East to North cluster was orders of magnitude

599 greater than that in the opposite direction and other cluster combinations (Fig.
600 2g; Table S11). Similarly, ancestral gene flow from the West to East cluster was
601 significantly greater than that in the opposite direction.
602
603 *Ecology*
604 A NJ tree based on the 592 bp alignment of the mtCR sequences, including
605 both the present and Su et al's. (2020) sequences, showed that 20, 55, and 15
606 haplotypes reported by Su et al. (2020) are most likely identified as the East,
607 West, and North clusters, respectively (Fig.S10; Table S4). Based on this
608 identification, we tested possible ecological divergence between the three
609 clusters. The Random Forest model showed that sampling location and season
610 were significant predictors of the cluster ($p < 0.05$) (Fig. 3a), suggesting that the
611 clusters tend to be found in different locations and seasons along the coast of
612 Taiwan. This implies habitat segregation or differences in migration patterns
613 contribute to reproductive isolation between the clusters. The Random Forest
614 proximity scores of the three clusters, reduced to two dimensions through
615 multidimensional scaling, largely overlapped (Fig. 3b), whereas the individual
616 assignments to the clusters predicted from the Random Forest model were
617 more accurate for the North cluster than for the other two clusters (Fig. 3c). This
618 suggests that the ecological divergence is more pronounced between the North
619 and the other clusters. In contrast, the dendrogram constructed through the
620 hierarchical cluster analysis could not fully explain the split of expected East,
621 West, and North clusters (Fig. S11a). Similarly, the three genetic clusters were

not clearly separated on the first four dimensions (explaining 61.5% of the total variation) derived from the FAMD (Fig. S11b–g).

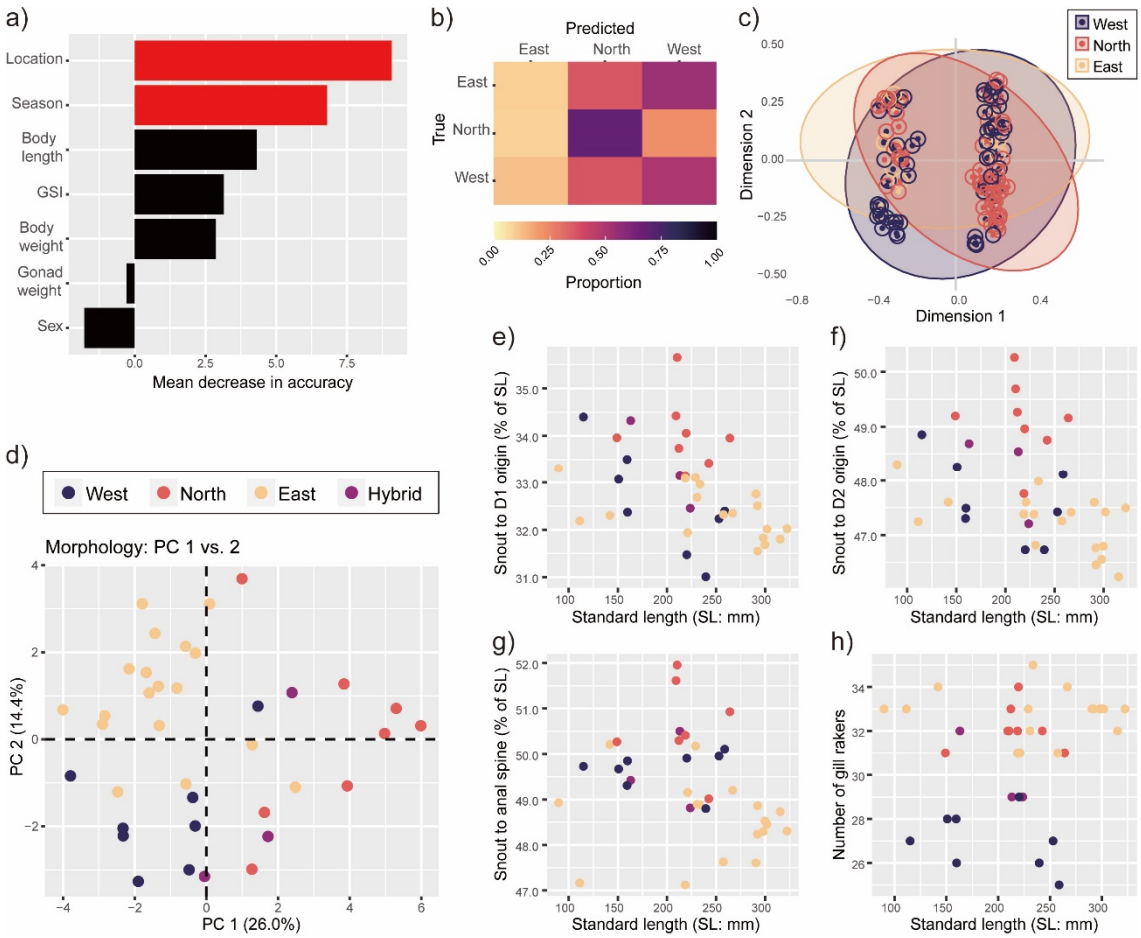

Fig. 3. Results of the ecological (a–c) and morphological (d–h) analyses. a) Importance of ecological factors in the Random Forest model, shown by the mean decrease in accuracy. Red bars indicate significant results (permutation test, $p < 0.05$). GSI: gonad somatic index. b) Confusion matrix of the Random Forest model comparing true (genetic) individual assignments to the clusters with those predicted by the model. c) Proximity scores reduced to two multidimensional scaling dimensions, showing ecological similarity between samples as determined by the Random Forest model. Inner points represent true (genetic) assignments to the clusters, and outer circles indicate those predicted by the model. d) Principal component analysis of 24 morphological characters. The proportion of variance explained by each principal component is in parentheses. e)–h) Plots of selected morphological characters. e) The distance between the tip of the snout and the origin of the 1st dorsal fin. f) The distance between the tip of the snout and the origin of the 2nd dorsal fin g) The distance between the tip of the snout and the origin of the 1st detached anal-fin spine. h) The number of gill rakers.

Morphology

Principal components (PCs) 1–4 explained 26.0%, 14.4%, 12.6%, and 6.7% of the total variation in overall morphology, respectively (Fig. S12a). The North cluster was separated from the East and West clusters on PC 1, suggesting its distinct morphology (Fig. 3d). This reflects the North cluster's relatively elongated anterior body: it is distinguishable from the East and West clusters by the distances from the tip of the snout to the origins of the first-dorsal, second-dorsal, and anal fins (Fig. 3e–g), with the former two characters strongly contributing to PC 1 (Fig. S12b–d). The East and West clusters were separated from each other on PCs 2 and 3, to which gill-raker number and head width strongly contributed, respectively (Figs. 3d, S12b–e). The West cluster had fewer gill-rakers (24–29) compared to the East and North clusters (31–35 and 31–34, respectively) (Fig. 3h), implying a distinct trophic niche, as gill-raker number generally correlates with diet size and shape in fishes (e.g., Gerking 1994). Other characters did not differ considerably between the three clusters (Fig. S13). The putative hybrids occupied intermediate positions between the East and North clusters on PC 1. The discovery of distinct morphological characteristics, coupled with genetic and ecological differences, would allow us to describe the three clusters as taxonomically distinct species. See Appendix 2 for taxonomic implications.

DISCUSSION

Reproductive Isolation and Gene Flow

We demonstrated that the North cluster is an evolutionarily independent lineage of hybrid origin. The genome-wide differentiation (Figs. 1b, S3–S5),

664 coupled with the presence of loci where the North cluster showed significant
665 differentiation from both the East and West clusters (Figs. 1d, S4, S5), provides
666 strong evidence that the North cluster is evolving separately from the others in
667 sympatry. Eco-morphological differences further support its ecological and
668 evolutionary independence (Figs. 3, S11–13). These results highlight that the
669 North cluster can be recognized as a distinct species under multiple species
670 concepts (reviewed in Coyne and Orr, 2004), such as the biological, genotypic,
671 evolutionary, and cohesion concepts. Moreover, the genetic analyses
672 consistently indicated the intermediate and admixed nature of the North cluster
673 (Figs. 1c, 2a–b, S3; Tables S7–S9), and demographic model selection favored
674 the hybrid speciation scenario over the introgression scenario (Figs. 2c, S9;
675 Table S10). To our knowledge, this is the most compelling example of
676 homoploid hybrid speciation in marine fishes.
677 The present system represents a relatively rare case of HHS where parental
678 and hybrid species coexist in sympatry. This pattern could most readily be
679 realized by ecological divergence between the species, which facilitates
680 extrinsic pre- or postzygotic reproductive isolation while reducing competition
681 (Buerkle et al. 2000). Indeed, our Random Forest analysis suggested spatio-
682 temporal segregation of the North cluster from the others, i.e., the seasonal and
683 locational factors strongly associated with the predictions of clusters, and the
684 distinct number of gill-rakers in the West cluster implied its unique trophic niche
685 (Fig. 3). We suggest that these ecological differences maintain the eco-
686 evolutionary independence of the sympatric species. However, the ecological
687 divergence is likely a by-product of allopatry under historical climate changes

(rather than a direct outcome of hybridization), with this allopatry being the primary driver of HHS, given the estimated demographic histories and ecogeographical setting of the present system (see the next section "*Ecogeographic Scenario of Hybrid Speciation in M. cordyla*").

It remains uncertain whether the reproductive isolation of the North cluster from the sympatric parental clusters is complete: the results of our genomic analyses were equivocal. We observed three putative F_2s between the North and parental clusters (Table S8), indicating incomplete pre- and postzygotic isolation before the F_2 generation. Nevertheless, postzygotic isolation due to hybrid breakdown can mainly manifest in F_2 or later generations, effectively preventing gene flow (Orr 1995; Stelkens et al. 2015; Lollar et al. 2023). The demographic parameter estimation suggested the presence of "contemporary gene flow" (Fig. 2g); however, gene flow may be no longer ongoing, as the estimation represents an average over generations from the formation of the North cluster until the present (Fig. 2c). Another result, the uniform ancestry composition across the individuals in the North cluster (Fig. 2a–b), rather supports exclusive reproduction among the members of the same cluster, i.e., the lack of gene flow.

We observed unequal parental genomic contributions in the North cluster (Figs. 1b–c, 2a–b, S3–S5), which is typical in HHS (e.g., Elgvin et al. 2017; Barrera-Guzmán et al. 2018; Noguerales and Ortego 2022; Lopes et al. 2023; Rosser et al. 2024). Relative parental contributions in hybrid species are determined by initial contributions upon fusion, subsequent gene flow, and differential fixation of ancestry blocks via drift and selection (genomic

712 stabilization) (Baack and Rieseberg 2007; Runemark et al. 2019). In the present
713 system, the predominance of East cluster ancestry seems to largely reflect
714 asymmetric gene flow following the fusion; the demographic parameter
715 estimation suggested that the West cluster initially made a greater contribution
716 ($1-\Gamma_{Fus} > 0.5$: Fig. 2e), and subsequent gene flow was predominantly from the
717 East to the North cluster (Fig. 2f). Additional support for the initial closer
718 relatedness of the West and North clusters and the importance of subsequent
719 gene flow comes from comparing suboptimal demographic models, where the
720 models assuming both a sister relationship between West and North (topology
721 2) and subsequent gene flow or pulse introgression were generally better
722 supported than those without (Fig. S9, Table S10). The observed asymmetric
723 gene flow may have resulted from asymmetry in the strength of reproductive
724 isolation (Rosser et al. 2024). Alternatively, it may reflect unequal frequencies of
725 repeated secondary contacts under the Pleistocene climatic oscillations (see
726 the next section).
727 The limited number of SNPs available constrained our analytical power. This
728 limitation led us to compromise the complexity of the demographic models,
729 excluding events like population size changes, multiple pulses of introgression,
730 or temporal shifts in the magnitude of gene flow, even though such events are
731 plausible (e.g., De Jode et al. 2023; Muto and Kai 2023). Additionally, the
732 preference for the hybrid speciation scenario over the pulse introgression was
733 not robust, depending on the presence of both ancestral and contemporary
734 gene flow in the model (Fig. S9). This lack of robustness might also be

attributed to the limited number of SNPs. We plan a future study to address these limitations using whole genome sequencing.

Ecogeographic Scenario of Hybrid Speciation in M. cordyla

The Pleistocene glacial cycles led to changes in sea levels and seawater temperatures, causing allopatric divergence, admixture, local adaptation, and population size changes in coastal marine organisms (Bowen et al. 2016). Based on our results and the geographic setting, we propose that multiple cycles of these processes primarily drove HHS in *M. cordyla*. Below, we outline a possible scenario in detail. Moreover, based on this scenario, we propose a model for HHS in marine coastal ecosystems, including a novel hypothesis on the initial persistence of nascent hybrid populations (see the next section "*Hybrid Speciation in Coastal Marine Environments*").

In the Indo-West Pacific, a vast shallow continental shelf, the Sunda Shelf (Fig. 1a), strongly limited marine connectivity between the Indian and Pacific Oceans during the Pleistocene low sea-level (glacial) periods. This "Indo-Pacific barrier" drove allopatric divergence in various marine organisms, evidenced by concordant distribution patterns of sister species and phylogeographic breaks around the Sunda Shelf (Carpenter et al. 2011; Gaither and Rocha 2013). The distributions of the East and West clusters of *M. cordyla* in the tropical West Pacific align with this pattern (Fig. 1a), and their estimated divergence time fell within the Pleistocene, suggesting that they originated from the allopatric divergence across the Indo-Pacific Barrier (Fig. 4a).

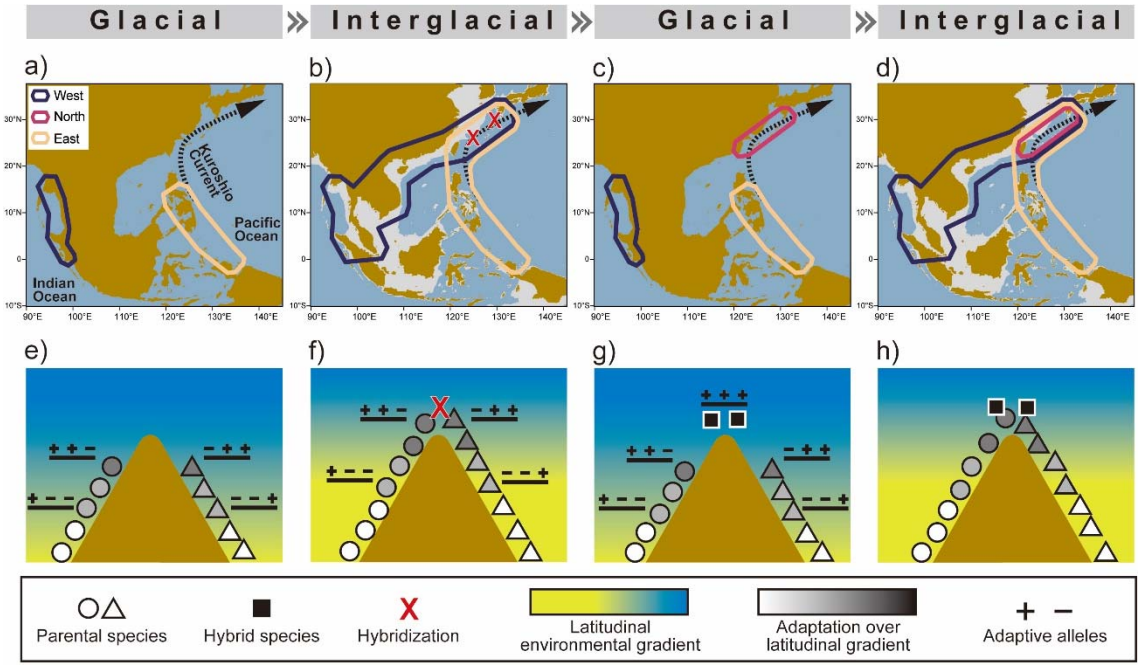

Fig. 4. a)–d) The proposed ecogeographic scenario of HHS in the present system. Pleistocene glacial-interglacial cycles drive a) allopatric divergence of the East and West clusters, b) their northward range expansions and secondary contact, leading to the formation of the North cluster via hybridization, c) range contractions of the East and West clusters resulting in the geographical isolation of the three clusters, during which ecological and other isolating barriers develops, and d) range overlap and coexistence in sympatry. e)–h) The ecogeographic model of hybrid speciation in coastal environments driven by the Pleistocene glacial cycles. Two coastlines stretching from low to high latitudes in parallel along a landmass, joining at high-latitude extreme, are shown. Gene flow can only occur along the coastline, not across the landmass. This process can also occur where topographic features are latitudinally reversed (i.e., two coastlines join at low-latitude extreme) or the barrier to gene flow is any other uninhabitable zone, such as an open ocean as in the present system. e) During a glacial period, high-latitude coastlines become uninhabitable due to extreme conditions (e.g., low temperature), resulting in allopatric divergence of parental species. Adaptation over latitudinal gradients in temperature and other factors occurs independently, based on different loci in each species. f) Environmental change from glacial to interglacial makes the high-latitude area habitable, leading to range expansions and secondary contact of parentals, forming a hybrid population. Hybridization occurs among individuals best adapted to high latitudes since the direction of the environmental change is aligned with the latitudinal gradient. g) In the next glacial period, the hybrids have greater local fitness in high-latitude areas due to additive/epistatic effects of adaptive variants from the parentals. Additionally, heterosis in hybrids via the combination of recessive deleterious alleles, having been accumulated in the expansion front of the parentals, can also contribute the greater fitness (not depicted). The greater fitness allows hybrids to persist in high latitudes while parentals contract to low-latitude refugia, resulting in geographical isolation. During this period of allopatry, isolating barriers such as pre- and post-mating barriers based on ecological divergence may

develop as by-products. h) With isolating barriers in place, the hybrid population maintains independence from parentals upon contact in subsequent interglacial periods.

Following the initial divergence, the ranges of the two clusters likely expanded and contracted in response to the glacial cycles, as demonstrated in many organisms (reviewed in Hewitt 2004; Hofreiter and Stewart 2009; Seersholm et al. 2020). Specifically, during warm interglacial periods, poleward range expansions likely occurred along mainland Asia for the West cluster and along the Kuroshio current for the East cluster (Kuriiwa et al. 2014; Motomura and Matsunuma 2022). One such poleward expansion may have resulted in their range overlap in the mid-latitude temperate area, similar to the present distribution pattern, triggering hybridization and the formation of the North cluster (Fig. 4b). In the subsequent glacial period, the North cluster survived in the temperate area with a stable population while the parentals contracted towards tropical refugia, as the Stairway Plot analysis implies (Fig. S6). This, in turn, resulted in their geographical isolation, securing the persistence of the nascent North cluster without being swamped by gene flow from the parentals (Fig. 4c). This allopatric period facilitated the development of inferred ecological difference and other unobserved isolating barriers, allowing the coexistence in sympatry during the next contact (Fig. 4d). The ancestral and contemporary gene flow inferred from the demographic modeling likely reflects relatively minor secondary contacts between the clusters.

We hypothesize that the North cluster's survival in higher latitudes during the glacial period (i.e., Fig. 4c) was driven by clinal adaptations in the parental clusters. Prior to hybridization, parallel adaptation along latitudinal gradients in

temperature and associated environmental factors likely occurred within each parental cluster, resulting in clinal patterns of adaptive genetic variants (Fig. 4e) (Jeffery et al. 2017; Teng et al. 2023). Subsequent hybridization probably took place at the high-latitude edge of the parental ranges, where these adaptive variants were most prevalent (Fig. 4f). In the North cluster, these variants recombined, enhancing fitness through additive and/or epistatic effects (Mallet 2007; Abbott et al. 2013), which enabled its survival in high latitudes during the subsequent glacial period (Fig. 4g). Conversely, deleterious alleles may have also accumulated at the expansion front of the parental clusters before hybridization, via allele surfing (Excoffier et al. 2009). However, if these deleterious alleles were recessive, they might have contributed to heterosis in hybrids, aiding their long-term persistence (MacPherson et al. 2022). Supporting our hypothesis, a recent simulation study demonstrated that admixture between two allopatric parental populations adapting to identical environments can result in hybrids that adapt more rapidly than the parentals to a novel environment (Kulmuni et al. 2024).

Hybrid Speciation in Coastal Marine Environments

Hybrid populations can express transgressive phenotypes via combinations of parental genetic variants, exploiting novel niches and thereby becoming geographically isolated from the parental populations (Rieseberg et al. 2003; Gross and Rieseberg 2005; Mallet 2007; Abbott et al. 2013; MacPherson et al. 2022; Kulmuni et al. 2024). Although this process may underlie HHS in various organisms (Mavárez and Linares 2008; Larsen et al. 2010; Brennan et al. 2012;

835 Kang et al. 2013; Lamichhaney et al. 2018; Maier et al. 2019; Noguerales and
836 Ortego 2022), the ecogeographic settings enabling the process are largely
837 unknown. The scenario described in the preceding section illustrates how this
838 process can occur in the ocean, including the necessary ecogeographic setting.
839 We propose this scenario as a general model of hybrid speciation in coastal
840 marine environments driven by the Pleistocene glacial cycles (Fig. 4e–h). This
841 model provides realistic, testable explanations for how parental adaptive genetic
842 variants are created and how hybrids' distinct niches lead to geographical
843 isolation, filling the gaps in the current understanding.

844 Under the model, hybrid speciation can occur where coastal marine habitats
845 in two parallel latitudinal arrays (e.g., coastlines or archipelagos) in a relevant
846 geographical scale, separated by an uninhabitable zone (e.g., land or open
847 ocean) but joined at latitudinal extremes (Fig. 4e). The uninhabitable zone
848 prevents gene flow between parental species, and hybridization occurs at the
849 latitudinal extreme under glacial maxima/minima (Fig. 4f). The following regions
850 not only fit these topographic conditions, but also are characterized by
851 secondary contact driven by the Pleistocene glacial cycles: temperate to tropical
852 Indo-west Pacific (Gaither and Rocha 2013); the Japanese Archipelago (Hirase
853 2022); the Aleutian Archipelago (Liu et al. 2012); southern Africa (Gaither et al.
854 2015); and the Florida peninsula (Soltis et al. 2006). We predict that these
855 regions harbor undiscovered cases of marine hybrid speciation.

856 Another key component of the model is the adaptation of the nascent hybrid
857 population to environmental conditions at the latitudinal extreme, leveraging the
858 preexisting adaptative genetic variants of the parents established through

parallel adaptation over a common latitudinal gradient (Fig. 4f–g). The parent's parallel adaptation may be based on similar phenotypes, but their genetic architectures are most likely *not* identical (e.g., Elmer and Meyer 2011; Jones et al. 2012), due to the lack of gene flow. Therefore, the recombinant hybrid genome can express transgressive phenotypes, conferring greater fitness than parents (Mallet 2007; Abbott et al. 2013). This greater fitness leads to local persistence under subsequent global cooling/warming and eventual allopatry from the parents, during which isolating barriers develop as by-products (Fig. 4h). These can be tested by investigating the eco-evolutionary forces underlying inter- and intraspecific genomic and phenotypic variations in parental and hybrid populations. For instance, Wang et al. (2021) employed a genome-wide association study, genomic cline analysis, and examinations of correlations between climate and genotype to investigate persistent hybrid populations of the *Setophaga* warbler complex. They demonstrated that these hybrids are differentiated from parental populations in genomic regions related to climatic adaptation, and that ancestry in these regions covaries with climatic variation.

ACKNOWLEDGEMENTS

We thank the following persons who facilitated the sampling and data collection: H. Motomura and members of The Kagoshima University Museum (KAUM) Fish Team; H-C Ho and J-F Huang (National Museum of Marine Biology and Aquarium, Taiwan); W-J Chen and M-Y Lee (National Taiwan University); M. Nakae and K. Matsuura (National Museum of Nature and Science, Japan); T. Nakabo, Y. Kai, and M. Matsunuma (Kyoto University

Museum); S. Kimura (Mie University); K. Koeda (University of the Ryukyus); M. I. Mahyam, R. H. Raja Bidin, O. Muda, W. M. Arshaad, M. Katoh, and O. Abe (Marine Fishery Resources Development and Management Department of Southeast Asian Fisheries Development Centre: SEAFDEC); Y. G. Seah (Universiti Malaysia Terengganu: UMT); A. Arshad (Universiti Putra Malaysia: UPM); S. Arnupapboon and K. Phuttharaksa (Eastern Marine Fisheries Research and Development Centre: EMDEC); N. D. The, P. V. Chien, D. V. Nhan, V. V. Chien, and P. T. Thu (Institute of Marine Environment and Resources, Vietnam: IMER); R. F. M. Traifalar, S. S. Garibay, late U. B. Alama, A. C. Gaje, and R. Cruz (University of the Philippines Visayas: UPV); Satoshi Ishikawa and members of the ‘Coastal Area Capability Enhancement in Southeast Asia’ project (Research Institute for Humanity and Nature: RIHN). We also thank Y. Kumazawa, F. Hibatullah, K. N. A. Kholil (Nagoya City University), W. Kusuma (Universitas Brawijaya), and J. R. Glass (University of Alaska Fairbanks) for their insightful discussions. The Philippines samples were collected under a Memorandum of Agreement for joint research made by and among the Department of Agriculture of the Republic of the Philippines (DA), UPV, KAUM, RIHN, and Tokai University, facilitated by S. L. Sanchez [Bureau of Fisheries and Aquatic Resources (BFAR), DA]. P. J. Alcala (DA) provided a Prior Informed Consent Certificate, and I. P. Cabacaba and S. M. S. Nolasco (BFAR, DA) provided a Fish Specimen Export Certificate (no. 2016-39812). The Malaysian samples were collected during the Japan Society for the Promotion of Science (JSPS) Asian Core Program “Establishment of Research and Education Network on Coastal Marine Science in Southeast Asia” and the JSPS

Core-to-Core Program B “Asia-Africa Science Platforms” supported by UPM, UMT, and the Ministry of Higher Education of the Government of Malaysia. The Thai samples were collected with support from the EMDEC, the Training Department of SEAFDEC, and the Office of Natural Science, National Science Museum, Pathum Thani. The Vietnamese samples were collected with the support of the IMER and the Ha Long Bay Management Department and with permission to use the samples from the Biodiversity Conservation Agency, Ministry of Natural Resources and Environment, Hanoi. Computations were performed, in part, on the NIG supercomputer at ROIS National Institute of Genetics, Japan.

FUNDING

This work was supported by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (No. JP26840131, 19K23691, 19KK0190, 23H02242, and 24K02087), JSPS Overseas Research Fellowships (No. 202160519), Sumitomo Fund (No.173246), Interdisciplinary Collaborative Research Program of Atmosphere and Ocean Research Institute, The University of Tokyo (No. JURCAOSIRG24-08), fund for integrated research activities (I.8b, 47 sub-project) from the Ministry of Agriculture and Rural Development, Vietnam, and Vietnam Academy of Science and Technology’s Grant in aid (VAST 06.04/18-19, NVCC 23.04/20-20).

CONFLICT OF INTEREST

The authors declare no competing interests.

AUTHORS' CONTRIBUTIONS

NM conceived and designed the study; NM, HH, NVQ, VV, MAG, and RPB conducted sampling and sample preparation; NM conducted genetic analysis; HH and NM conducted morphological analysis; SYC conducted ecological analysis; NM wrote the manuscript with input from HH and SYC; all co-authors contributed to improving the manuscript.

DATA AVAILABILITY

Data will be available on an online data repository upon publication of the peer-reviewed version of this manuscript.

REFERENCES

Abbott R., Albach D., Ansell S., Arntzen J.W., Baird S.J.E., Bierne N., Boughman J., Brelsford A., Buerkle C.A., Buggs R., Butlin R.K., Dieckmann U., Eroukhmanoff F., Grill A., Cahan S.H., Hermansen J.S., Hewitt G., Hudson A.G., Jiggins C., Jones J., Keller B., Marczewski T., Mallet J., Martinez-Rodriguez P., Möst M., Mullen S., Nichols R., Nolte A.W., Parisod C., Pfennig K., Rice A.M., Ritchie M.G., Seifert B., Smadja C.M., Stelkens R., Szymura J.M., Väinölä R., Wolf J.B.W., Zinner D. 2013. Hybridization and speciation. J. Evol. Biol. 26:229–246.

Archer E. 2021. rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics. https://CRAN.R-project.org/package=rfPermute.

Baack E.J., Rieseberg L.H. 2007. A genomic view of introgression and hybrid speciation. Curr. Op. Genet. Dev. 17:513–518.

Barrera-Guzmán A.O., Aleixo A., Shawkey M.D., Weir J.T. 2018. Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. Proc. Natl. Acad. Sci. USA. 115:E218–E225.

Blischak P.D., Chifman J., Wolfe A.D., Kubatko L.S. 2018. HyDe: A Python package for genome-scale hybridization detection. Syst. Biol. 67:821–829.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.H., Xie D., Suchard M.A., Rambaut A., Drummond A.J. 2014. BEAST 2: A software platform for Bayesian evolutionary analysis. PLOS Comp. Biol. 10:1–6.

Bouckaert R.R., Drummond A.J. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol. Biol. 17:1–11.

Bowen B.W., Gaither M.R., DiBattista J.D., Iacchei M., Andrews K.R., Grant W.S., Toonen R.J., Briggs J.C. 2016. Comparative phylogeography of the ocean planet. Proc. Natl. Acad. Sci. USA. 113:7962–7969.

Brennan A.C., Barker D., Hiscock S.J., Abbott R.J. 2012. Molecular genetic and quantitative trait divergence associated with recent homoploid hybrid speciation: a study of *Senecio squalidus* (Asteraceae). Heredity. 108:87–95.

Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A., RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932.

Buerkle C.A., Morris R.J., Asmussen M.A., Rieseberg L.H. 2000. The likelihood of homoploid hybrid speciation. Heredity. 84:441–451.

Cárdenas L., Hernández C.E., Poulin E., Magoulas A., Kornfield I., Ojeda F.P. 2005. Origin, diversification, and historical biogeography of the genus *Trachurus* (Perciformes: Carangidae). Mol. Phylogenet. Evol. 35:496–507.

Carpenter K.E., Barber P.H., Crandall E.D., Ablan-Lagman M.C.A., Ambariyanto, Mahardika G.N., Manjaji-Matsumoto B.M., Juinio-Meñez M.A., Santos M.D., Starger C.J., Toha A.H.A. 2011. Comparative phylogeography of the coral triangle and implications for marine management. J. Mar. Biol. 2011:1–14.

Catchen J., Hohenlohe P.A., Bassham S., Amores A., Cresko W.A. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.

Chan K.O., Grismer L.L. 2022. GroupStruct: An R Package for Allometric Size Correction. Zootaxa. 5124:471–482.

Chen S., Zhou Y., Chen Y., Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 34:i884–i890.

Chhatre V.E., Emerson K.J. 2017. StrAuto: automation and parallelization of STRUCTURE analysis. BMC Bioinformatics. 18:192.

Coyne J.A., Orr H.A. 2004. Speciation. Sunderland, Massachusetts: Sinauer Associates.

Danecek P., Auton A., Abecasis G., Albers C.A., Banks E., DePristo M.A., Handsaker R.E., Lunter G., Marth G.T., Sherry S.T., McVean G., Durbin R., 1000 Genomes

Project Analysis Group. 2011. The variant call format and VCFtools. Bioinformatics. 27:2156–2158.

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M., Li H. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10:giab008.

De Jode A., Le Moan A., Johannesson K., Faria R., Stankowski S., Westram A., Butlin R., Rafajlović M., Fraïsse C. 2023. Ten years of demographic modelling of divergence and speciation in the sea. Evol. App. 16:542–559.

DeRaad D.A. 2022. snpfiltr: An R package for interactive and reproducible SNP filtering. Mol. Ecol. Resour. 22:2443–2453.

Donelson J.M., Sunday J.M., Figueira W.F., Gaitán-Espitia J.D., Hobday A.J., Johnson C.R., Leis J.M., Ling S.D., Marshall D., Pandolfi J.M., Pecl G., Rodgers G.G., Booth D.J., Munday P.L. 2019. Understanding interactions between plasticity, adaptation and range shifts in response to marine environmental change. Phil. Trans. R. Soc. B. 374:20180186.

Donovan L.A., Rosenthal D.R., Sanchez-Velenosi M., Rieseberg L.H., Ludwig F. 2010. Are hybrid species more fit than ancestral parent species in the current hybrid species habitats? J. Evol. Biol. 23:805–816.

Duranton M., Allal F., Valière S., Bouchez O., Bonhomme F., Gagnaire P.-A. 2020. The contribution of ancient admixture to reproductive isolation between European sea bass lineages. Evol. Lett. 4:226–242.

Earl D.A., vonHoldt B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv. Genet. Resour. 4:359–361.

Elgvin T.O., Trier C.N., Tørresen O.K., Hagen I.J., Lien S., Nederbragt A.J., Ravinet M., Jensen H., Sætre G.-P. 2017. The genomic mosaicism of hybrid speciation. Sci. Adv. 3:e1602996.

Elmer K.R., Meyer A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. Trends Ecol. Evol. 26:298–306.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14:2611–2620.

Excoffier L., Dupanloup I., Huerta-Sánchez E., Sousa V.C., Foll M. 2013. Robust demographic inference from genomic and SNP data. PLOS Genet. 9:e1003905.

Excoffier L., Foll M., Petit R.J. 2009. Genetic consequences of range expansions. Annu. Rev. Ecol. Evol. Syst. 40:481–501.

Excoffier L., Lischer H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10:564–567.

Excoffier L., Marchi N., Marques D.A., Matthey-Doret R., Gouy A., Sousa V.C. 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics. 37:4882–4885.

Fitzpatrick B.M. 2012. Estimating ancestry and heterozygosity of hybrids using molecular markers. BMC Evol. Biol. 12:131.

Francis R.M. 2017. pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. 17:27–32.

Froese R., Pauly D. 2023. Fishbase. Available from https://www.fishbase.se/popdyn/KeyfactsSummary_1.php?ID=384&GenusName=Megalaspis&SpeciesName=cordyla&vStockCode=398&fc=314.

Gaither M.R., Bernal M.A., Fernandez-Silva I., Mwale M., Jones S.A., Rocha C., Rocha L.A. 2015. Two deep evolutionary lineages in the circumtropical glasseye *Heteropriacanthus cruentatus* (Teleostei, Priacanthidae) with admixture in the south-western Indian Ocean. J. Fish Biol. 87:715–727.

Gaither M.R., Rocha L.A. 2013. Origins of species richness in the Indo-Malay-Philippine biodiversity hotspot: evidence for the centre of overlap hypothesis. J. Biogeogr. 40:1638–1648.

Gerking S. D. 1994. Feeding Ecology of Fish. San Diego: Academic Press.

Gibbs R.H., Collette B.B. 1967. Comparative anatomy and systematics of the tunas, genus *Thunnus*. Fish. Bull. 66:65–130.

Glass J.R., Harrington R.C., Cowman P.F., Faircloth B.C., Near T.J. 2023. Widespread sympatry in a species-rich clade of marine fishes (Carangoidei). Proc. R. Soc. B. 290:20230657.

Gompert Z., Fordyce J.A., Forister M.L., Shapiro A.M., Nice C.C. 2006. Homoploid hybrid speciation in an extreme habitat. Science. 314:1923–1925.

Gross B.L., Rieseberg L.H. 2005. The ecological genetics of homoploid hybrid speciation. J. Hered. 96:241–252.

Hall T.A. 1999. BioEdit : a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids. Symp. Ser. 41:95–98.

Hewitt G.M. 2000. The genetic legacy of the Quaternary ice ages. Nature. 405:907–913.

Hewitt G.M. 2004. Genetic consequences of climatic oscillations in the Quaternary. Phil. Trans. R. Soc. B. 359:183–195.

Hirase S. 2022. Comparative phylogeography of coastal gobies in the Japanese Archipelago: future perspectives for the study of adaptive divergence and speciation. Ichthyol. Res. 69:1–16.

Hirase S., Tezuka A., Nagano A.J., Sato M., Hosoya S., Kikuchi K., Iwasaki W. 2021. Integrative genomic phylogeography reveals signs of mitonuclear incompatibility in a natural hybrid goby population. Evolution. 75:176–194.

Hofreiter M., Stewart J. 2009. Ecological change, range fluctuations and population dynamics during the Pleistocene. Curr. Biol. 19:R584–R594.

Huson D.H., Bryant D. 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23:254–267.

Ivanova N.V., Zemlak T.S., Hanner R.H., Hebert P.D.N. 2007. Universal primer cocktails for fish DNA barcoding. Mol. Ecol. Notes. 7:544–548.

Jeffery N.W., Stanley R.R.E., Wringe B.F., Guijarro-Sabaniel J., Bourret V., Bernatchez L., Bentzen P., Beiko R.G., Gilbey J., Clément M., Bradbury I.R. 2017. Range-wide parallel climate-associated genomic clines in Atlantic salmon. Roy. Soc. Open Sci. 4:171394.

Jombart T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics. 24:1403–1405.

Jones F.C., Grabherr M.G., Chan Y.F., Russell P., Mauceli E., Johnson J., Swofford R., Pirun M., Zody M.C., White S., Birney E., Searle S., Schmutz J., Grimwood J., Dickson M.C., Myers R.M., Miller C.T., Summers B.R., Knecht A.K., Brady S.D., Zhang H., Pollen A.A., Howes T., Amemiya C., Lander E.S., Di Palma F., Lindblad-Toh K., Kingsley D.M. 2012. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 484:55–61.

Kang J.H., Schartl M., Walter R.B., Meyer A. 2013. Comprehensive phylogenetic analysis of all species of swordtails and platies (Pisces: Genus *Xiphophorus*) uncovers a hybrid origin of a swordtail fish, *Xiphophorus monticolus*, and demonstrates that the sexually selected sword originated in the ancestral lineage of the genus, but was lost again secondarily. BMC. Evol. Biol. 13:25.

Kiessling W., Simpson C., Beck B., Mewis H., Pandolfi J.M. 2012. Equatorial decline of reef corals during the last Pleistocene interglacial. Proc. Natl. Acad. Sci. USA. 109:21378–21383.

Kulmuni J., Wiley B., Otto S.P. 2024. On the fast track: hybrids adapt more rapidly than parental populations in a novel environment. Evol. Lett. 8:128–136.

Kuriiwa K., Chiba S.N., Motomura H., Matsuura K. 2014. Phylogeography of Blacktip Grouper, *Epinephelus fasciatus* (Perciformes: Serranidae), and influence of the Kuroshio Current on cryptic lineages and genetic population structure. Ichthyol. Res. 61:361–374.

Lamichhaney S., Han F., Webster M.T., Andersson L., Grant B.R., Grant P.R. 2018. Rapid hybrid speciation in Darwin's finches. Science. 359:224–228.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., Mcgettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23:2947–2948.

Larsen P.A., Marchán-Rivadeneira M.R., Baker R.J. 2010. Natural hybridization generates mammalian lineage with species characteristics. Proc. Natl. Acad. Sci. USA. 107:11447–11452.

Lê S., Josse J., Husson F. 2008. FactoMineR: An R Package for Multivariate Analysis. J. Stat. Soft. 25:1–18.

Leigh J.W., Bryant D. 2015. POPART: Full-feature software for haplotype network construction. Methods Ecol. Evol. 6:1110–1116.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 1303.3997v2:1–3.

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

Liu M., Lin L., Gao T., Yanagimoto T., Sakurai Y., Grant W.S. 2012. What maintains the central North Pacific genetic discontinuity in Pacific Herring ? PLOS ONE. 7:e50340.

Liu X., Fu Y.-X. 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol. 21:280.

Lollar M.J., Biewer-Heisler T.J., Danen C.E., Pool J.E. 2023. Hybrid breakdown in male reproduction between recently diverged *Drosophila melanogaster* populations has a complex and variable genetic architecture. Evolution. 77:1550–1563.

Long Z., Rieseberg L.H. 2024. Documenting homoploid hybrid speciation. Mol. Ecol. n/a:e17412.

Lopes F., Oliveira L.R., Beux Y., Kessler A., Cárdenas-Alayza S., Majluf P., Páez-Rosas D., Chaves J., Crespo E., Brownell R.L., Baylis A.M.M., Sepúlveda M., Franco-Trecu V., Loch C., Robertson B.C., Peart C.R., Wolf J.B.W., Bonatto S.L. 2023. Genomic evidence for homoploid hybrid speciation in a marine mammal apex predator. Sci. Adv. 9:eadf6601.

MacPherson A., Wang S., Yamaguchi R., Rieseberg L.H., Otto S.P. 2022. Parental population range expansion before secondary contact promotes heterosis. Am. Nat. 200:E1–E15.

Maier P.A., Vandergast A.G., Ostoja S.M., Aguilar A., Bohonak A.J. 2019. Pleistocene glacial cycles drove lineage diversification and fusion in the Yosemite toad (*Anaxyrus canorus*). Evolution. 73:2476–2496.

Mallet J. 2007. Hybrid speciation. Nature. 446:279–283.

Marr J.C., Schaefer F.B. 1949. Definitions of body dimensions used in describing tunas. Fish. Bull. Fish. Wild. Serv. 51:241–244.

Mavárez J., Linares M. 2008. Homoploid hybrid speciation in animals. Mol Ecol. 17:4181–4185.

Meier J.I., Sousa V.C., Marques D.A., Selz O.M., Wagner C.E., Excoffier L., Seehausen O. 2017. Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. Mol. Ecol. 26:123–141.

Motomura H., Matsunuma M. 2022. Fish diversity along the Kuroshio Current. In: Yoshiaki K., Motomura H., Matsuura K., editors. Fish Diversity of Japan. Singapore: Springer Nature Singapore. p. 63–78.

Muto N., Alama U.B., Hata H., Guzman A.M.T., Cruz R., Gaje A., Traifalgar R.F.M., Kakioka R., Takeshima H., Motomura H., Muto F., Babaran R.P., Ishikawa S. 2016. Genetic and morphological differences among the three species of the genus *Rastrelliger* (Perciformes: Scombridae). Ichthyol. Res. 63:275–287.

Muto N., Kai Y. 2023. Allopatric origin, secondary contact and subsequent isolation of sympatric rockfishes (Sebastidae: *Sebastes*) in the north-western Pacific. Biol. J. Linn. Soc. 138:37–50.

Muto N., Kakioka R., Ghaffar M.A., Quan N.V., Vilasri V., Traifalgar R.F.M., Babaran R.P., Muto F., Takeshima H., Motomura H., Ishikawa S. 2021. Fine-scale phylogeography

of coastal fishes in the South China Sea: Possible roles of biological traits and geography. J. Biogeogr. 48:2785–2800.

Neiva J., Assis J., Fernandes F., Pearson G.A., Serrão E.A. 2014. Species distribution models and mitochondrial DNA phylogeography suggest an extensive biogeographical shift in the high-intertidal seaweed *Pelvetia canaliculata*. J. Biogeogr. 41:1137–1148.

Nieto Feliner G., Álvarez I., Fuertes-Aguilar J., Heuertz M., Marques I., Moharrek F., Piñeiro R., Riina R., Rosselló J.A., Soltis P.S., Villa-Machío I. 2017. Is homoploid hybrid speciation that rare? An empiricist's view. Heredity. 118:513–516.

Noguerales V., Ortego J. 2022. Genomic evidence of speciation by fusion in a recent radiation of grasshoppers. Evolution. 76:2618–2633.

Olave M., Nater A., Kautt A.F., Meyer A. 2022. Early stages of sympatric homoploid hybrid speciation in crater lake cichlid fishes. Nat. Commun. 13:5893.

Orr H.A. 1995. The population genetics of speciation: the evolution of hybrid incompatibilities. Genetics. 139:1805–1813.

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

Peñalba J.V., Runemark A., Meier J.I., Singh P., Wogan G.O.U., Sánchez-Guillén R., Mallet J., Rometsch S.J., Menon M., Seehausen O., Kulmuni J., Pereira R.J. 2024. The role of hybridization in species formation and persistence. Cold Spring Harb. Perspect. Biol. doi: 10.1101/cshperspect.a041445

Petr M., Vernot B., Kelso J. 2019. admixr—R package for reproducible analyses using ADMIXTOOLS. Bioinformatics. 35:3194–3195.

Pickett B.D., Glass J.R., Johnson T.P., Ridge P.G., Kauwe J.S.K. 2022. The genome of a giant (trevally): *Caranx ignobilis*. Gigabyte. 2022:1–16.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

R Core Team. 2022. R: A language and environment for statistical computing. https://www.r-project.org/.

Rieseberg L.H., Raymond O., Rosenthal D.M., Lai Z., Livingstone K., Nakazato T., Durphy J.L., Schwarzbach A.E., Donovan L.A., Lexer C. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. Science. 301:1211–1216.

Rosser N., Seixas F., Queste L.M., Cama B., Mori-Pezo R., Kryvokhyzha D., Nelson M., Waite-Hudson R., Goringe M., Costa M., Elias M., Mendes Eleres de Figueiredo C., Freitas A.V.L., Joron M., Kozak K., Lamas G., Martins A.R.P., McMillan W.O., Ready J., Rueda-Muñoz N., Salazar C., Salazar P., Schulz S., Shirai L.T., Silva-Brandão K.L., Mallet J., Dasmahapatra K.K. 2024. Hybrid speciation driven by multilocus introgression of ecological traits. Nature. 628:811–817.

Runemark A., Vallejo-Marin M., Meier J.I. 2019. Eukaryote hybrid genomes. PLOS Genet. 15:e1008404.

Schumer M., Rosenthal G.G., Andolfatto P. 2014. How common is homoploid hybrid speciation? Evolution. 68:1553–60.

Schwarz D., Matta B.M., Shakir-Botteri N.L., McPheron B.A. 2005. Host shift to an invasive plant triggers rapid animal hybrid speciation. Nature. 436:546–549.

Seersholm F.V., Werndly D.J., Grealy A., Johnson T., Keenan Early E.M., Lundelius E.L., Winsborough B., Farr G.E., Toomey R., Hansen A.J., Shapiro B., Waters M.R., McDonald G., Linderholm A., Stafford T.W., Bunce M. 2020. Rapid range shifts and megafaunal extinctions associated with late Pleistocene climate change. Nat. Commun. 11:2770.

Smith-Vaniz W.F. 1999. Carangidae. In: Carpenter K.E., Niem V.H., editors. THE LIVING MARINE RESOURCES OF THE WESTERN CENTRAL PACIFIC Volume 4: Bony fishes part 2 (Mugilidae to Carangidae). Rome: FAO. p. 2659–2756.

Soltis D.E., Morris A.B., McLACHLAN J.S., Manos P.S., Soltis P.S. 2006. Comparative phylogeography of unglaciated eastern North America. Mol. Ecol. 15:4261–4293.

Stange M., Sánchez-Villagra M.R., Salzburger W., Matschiner M. 2018. Bayesian divergence-time estimation with genome-wide single-nucleotide polymorphism data of sea catfishes (Ariidae) supports Miocene closure of the Panamanian Isthmus. Syst. Biol. 67:681–699.

Stelkens R.B., Schmid C., Seehausen O. 2015. Hybrid breakdown in cichlid fish. PLOS ONE. 10:e0127207.

Su Y.-C., Su S.-H., Li H.-Y., Wang H.-Y., Lee S.-C. 2020. Implication of single year seasonal sampling to genetic diversity fluctuation that coordinates with oceanographic dynamics in torpedo scads near Taiwan. Sci. Rep. 10:16829.

Suyama Y., Hirota S.K., Matsuo A., Tsunamoto Y., Mitsuyuki C., Shimura A., Okano K. 2022. Complementary combination of multiplex high-throughput DNA sequencing for molecular phylogeny. Ecol. Res. 37:171–181.

Suyama Y., Matsuki Y. 2015. MIG-seq: an effective PCR-based method for genome-wide single-nucleotide polymorphism genotyping using the next-generation sequencing platform. Sci. Rep. 5:16963.

Taylor S.A., Larson E.L. 2019. Insights from genomes into the evolutionary importance and prevalence of hybridization in nature. Nat. Ecol. Evol. 3:170–177.

Teng W., Fu H., Li Z., Zhang Q., Xu C., Yu H., Kong L., Liu S., Li Q. 2023. Parallel evolution in *Crassostrea* oysters along the latitudinal gradient is associated with variation in multiple genes involved in adipogenesis. Mol. Ecol. 32:5276–5287.

Tine M., Kuhl H., Gagnaire P.-A., Louro B., Desmarais E., Martins R.S.T., Hecht J., Knaust F., Belkhir K., Klages S., Dieterich R., Stueber K., Piferrer F., Guinand B., Bierne N., Volckaert F.A.M., Bargelloni L., Power D.M., Bonhomme F., Canario A.V.M., Reinhardt R. 2014. European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. Nat. Commun. 5:1–10.

Wang S., Ore M.J., Mikkelsen E.K., Lee-Yaw J., Toews D.P.L., Rohwer S., Irwin D. 2021. Signatures of mitonuclear coevolution in a warbler species complex. Nat. Commun. 12:1–11.

Wood T.E., Takebayashi N., Barker M.S., Mayrose I., Greenspoon P.B., Rieseberg L.H. 2009. The frequency of polyploid speciation in vascular plants. Proc. Natl. Acad. Sci. USA. 106:13875–13879.

Wright M.N., Ziegler A. 2017. ranger: a fast implementation of Random Forests for high dimensional data in C++ and R. J. Stat. Soft. 77:1–17.

Yakimowski S.B., Rieseberg L.H. 2014. The role of homoploid hybridization in evolution: A century of studies synthesizing genetics and ecology. Am. J. Bot. 101:1247–1258.

[dataset]* Santos S.R. 2014. Data from: “Complete mitochondrial and partial nuclear genomes for the jack species Caranx ignobilis (Forsskål, 1775) and C. melampygus (Cuvier, 1833) (Perciformes: Carangidae) from the High Hawaiian Islands” in Genomic Resources Notes accepted 1 October 2013 – 30 November 2013. Dryad. https://doi.org/10.5061/dryad.6gr7t.