

Divergence and gene flow history at two large chromosomal inversions underlying ecotype differentiation in the long-snouted seahorse

Laura Meyer¹  | Pierre Barry^{1,2} | Florentine Riquet³  | Andrew Foote⁴  |
Clio Der Sarkissian⁵ | Regina L. Cunha⁶ | Christine Arbiol¹ | Frédérique Cerqueira¹ |
Erick Desmarais¹ | Anaïs Bordes¹ | Nicolas Bierne¹ | Bruno Guinand¹  |
Pierre-Alexandre Gagnaire¹

¹ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

²CIBIO-InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos Universidade do Porto, Vairão, Portugal

³Ifremer, RBE-ASIM, Station de La Tremblade, La Tremblade, France

⁴Centre for Ecological and Evolutionary Synthesis (CEES), Department of Biosciences, University of Oslo, Oslo, Norway

⁵Centre for Anthropobiology and Genomics of Toulouse, CNRS, University of Toulouse Paul Sabatier, Toulouse, France

⁶Centre of Marine Sciences-CCMAR, University of Algarve, Faro, Portugal

Correspondence

Laura Meyer and Pierre-Alexandre Gagnaire, ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France.
Email: laura.meyer@umontpellier.fr and pierre-alexandre.gagnaire@umontpellier.fr

Funding information

Agence Nationale de la Recherche, Grant/Award Number: CoGeDiv ANR-17-CE02-0006-01

Handling Editor: Hannah Augustijnen

Abstract

Chromosomal inversions can play an important role in divergence and reproductive isolation by building and maintaining distinct allelic combinations between evolutionary lineages. Alternatively, they can take the form of balanced polymorphisms that segregate within populations until one arrangement becomes fixed. Many questions remain about how inversion polymorphisms arise, how they are maintained over the long term, and ultimately, whether and how they contribute to speciation. The long-snouted seahorse (*Hippocampus guttulatus*) is genetically subdivided into geographic lineages and marine-lagoon ecotypes, with shared structural variation underlying lineage and ecotype divergence. Here, we aim to characterize structural variants and to reconstruct their history and suspected role in ecotype formation. We generated a near chromosome-level genome assembly and described genome-wide patterns of diversity and divergence through the analysis of 112 whole-genome sequences from Atlantic, Mediterranean, and Black Sea populations. By also analysing linked-read sequencing data, we found evidence for two chromosomal inversions that were several megabases in length and showed contrasting allele frequency patterns between lineages and ecotypes across the species range. We reveal that these inversions represent ancient intraspecific polymorphisms, one likely being maintained by divergent selection and the other by pseudo-overdominance. A possible selective coupling between the two inversions was further supported by the absence of specific haplotype combinations and a putative functional interaction between the two inversions in reproduction. Lastly, we detected gene flux eroding divergence between inverted alleles at varying levels for the two inversions, with a likely impact on their dynamics and contribution to divergence and speciation.

KEYWORDS

ancestral recombination graph, ecotypes, evolutionary history, gene flux, inversions, whole-genome resequencing

1 | INTRODUCTION

How genetic differences accumulate between nascent species and contribute to the build-up of reproductive isolation remains a fundamental question in evolutionary biology (Westram, Stankowski, et al., 2022). Mechanisms that bring unlinked or distant sites into linkage disequilibrium and that maintain strong allelic associations in the presence of gene flow are key drivers of reproductive isolation (Butlin, 2005; Ortiz-Barrientos et al., 2016; Tigano & Friesen, 2016). Large structural variants (SVs) such as chromosomal inversions combine these two properties, since they both establish and maintain divergent haplotypes by reducing inter-haplotype recombination (Faria & Navarro, 2010; Hoffmann & Rieseberg, 2008; Mérot, Llaurens, et al., 2020; Wellenreuther et al., 2019; Zhang et al., 2021). For these reasons, large SVs have often received special attention in speciation genomics studies, particularly those investigating ecotype formation (e.g., Atlantic cod, Berg et al., 2016; rough periwinkles, Faria, Chaube, et al., 2019; deer mice, Harringmeyer & Hoekstra, 2022; seaweed flies, Mérot, Oomen, et al., 2020; sunflowers, Todesco et al., 2020; monkeyflowers, Lowry & Willis, 2010). Despite the growing number of studies reporting the importance of inversions for divergence between ecotypes in a variety of organisms, many questions remain about the conditions in which these inversions arise, how they are maintained over the long term, and how they contribute to reproductive isolation and the progress towards speciation (Faria, Johannesson, et al., 2019; Kirkpatrick, 2010; Wellenreuther & Bernatchez, 2018; Westram, Faria, et al., 2022).

The most immediate role of inversions is generally ascribed to recombination suppression, which protects combinations of locally advantageous or co-adapted alleles locked within an inverted region (Butlin, 2005; Kirkpatrick, 2010; Noor et al., 2001; Rieseberg, 2001). Several studies have directly linked inversions to traits that differ between ecotypes, with functional implications in local adaptation or reproductive isolation (Campbell et al., 2021; Funk et al., 2021; Gould et al., 2017; Hager et al., 2022; Jay et al., 2021; Lundberg et al., 2023). These tend to be complex phenotypes that likely involve several genes inherited as a single block (Lamichhaney et al., 2016; Matschiner et al., 2022; Schwander et al., 2014; Thompson & Jiggins, 2014). Large inversions spanning megabase-sized chromosomal segments can facilitate the emergence of such complex phenotypes by building linkage disequilibrium between multiple distant functional sites (Thompson & Jiggins, 2014). However, inversions can also be expected to carry other types of selected sites, including co-adapted gene complexes and deleterious mutations that are captured by chance (Berdan et al., 2021; Jay et al., 2021; Nei et al., 1967). These different types of selected loci, as well as the potential interactions between them, imply that the dynamics of an inversion can be driven by multiple simultaneous processes (Faria, Johannesson, et al., 2019; Guerrero et al., 2012). Furthermore, the allelic content of an inversion is not fixed and can change over time, for example, through the accumulation of new mutations or the exchange of genetic material by gene flux between inverted haplotypes (Cheng et al., 2012; Faria, Johannesson, et al., 2019; Matschiner et al., 2022;

Navarro et al., 1997; Schaeffer & Anderson, 2005). This can occur through double crossover events with the formation of an inversion loop, or through non-crossover gene conversion during the repair of double-stranded breaks (Korunes & Noor, 2019; Matschiner et al., 2022). The resulting changes in allelic content could affect the processes that control the frequency of an inversion (Berdan et al., 2021), and ultimately, its evolutionary trajectory and contribution to reproductive isolation.

Considering the evolutionary history of inversions provides a particularly useful framework for understanding their role in speciation (Faria, Johannesson, et al., 2019). This framework aims to disentangle the roles of different evolutionary processes acting at different stages – from the appearance of inversions to their maintenance and long-term fate – in order to go beyond a simple inventory of the inversions present in a given system. In their review of the evolution of inversions, Faria, Johannesson, et al. (2019) distinguished two main classes of inversion polymorphisms, which they refer to as Type I and Type II. Type I inversions are expected to show frequency differences and divergence between populations, with the possible presence of polymorphism within populations due to gene flow. This pattern may be caused by divergent selection acting on alternate haplotypes conferring local adaptation and/or a form of bistable selection involving frequency dependence, incompatibility selection, or assortative mating. A Type I polymorphism may show signs of underdominance either due to direct effects (loss of unbalanced recombinant gametes during meiosis) or due to the allelic contents of the inversion (Faria, Johannesson, et al., 2019; Kirkpatrick & Barton, 2006). Over time, Type I inversions likely contribute to the accumulation of Dobzhansky–Müller (DM) incompatibilities and reinforcement mechanisms via coupling with other reproductive isolation polymorphisms (Kulmuni et al., 2020; Navarro & Barton, 2003). Alternatively, inversions can be maintained by a form of balancing selection (Jay et al., 2021; Mérot, Llaurens, et al., 2020; Wellenreuther & Bernatchez, 2018; Yeaman, 2013; Yeaman & Whitlock, 2011). Faria, Johannesson, et al. (2019) recognize such Type II inversions as polymorphisms that are maintained by frequency dependence, disassortative mating, or antagonistic pleiotropy. Type II inversions might also contain overdominant loci and deleterious mutations in repulsion phase linkage (i.e., pseudo-overdominance), leading to heterokaryotype advantage (Marion & Noor, 2023). Indeed, some inversions can be expected to suffer from high mutation load, due to deleterious mutations hitchhiking with selected genes, or simply accumulating in low recombination regions (Hill & Robertson, 1966; Nei et al., 1967). These various processes associated with Type I and Type II polymorphisms are not necessarily mutually exclusive and might act in concert to shape the evolutionary dynamics of a given system. In particular, the role of potential interactions between different types of inversions remains a key question that needs to be addressed with empirical data.

Here, we aim to disentangle the different processes that affect the evolution of chromosomal inversions involved in ecotype differentiation in the long-snouted seahorse (*Hippocampus guttulatus*). Seahorses inhabit shallow coastal waters and littoral lagoons of the

Northeast Atlantic, Mediterranean, and Black Seas. Throughout their distribution, seahorse populations occupy a wide variety of habitats, spanning a gradient of environmental conditions from lagoons to marine habitats. These habitats notably differ in the composition of their species communities associated with the seagrass beds inhabited by seahorses (*Zostera*, *Cymodocea*, *Posidonia*), reflecting the ecological impact of variations in physical and chemical parameters across the marine-lagoon gradient (Perez-Ruzafa et al., 2011). The long-snouted seahorse is genetically subdivided into semi-isolated lineages separated by sharp eco-geographical boundaries (Riquet et al., 2019). In the Northeast Atlantic, seahorse populations are split into a northern and southern genetic lineage, while Mediterranean populations display ecotype divergence between marine and lagoon habitats. Although lagoon and marine ecotypes are morphologically similar, they show strong genetic differences concentrated in a large genomic island of divergence which also differentiates the two Atlantic lineages (Riquet et al., 2019). The genomic architecture of partial reproductive isolation between Mediterranean ecotypes on the one hand, and Atlantic lineages on the other hand, therefore probably relies on shared structural variation.

We introduce the first whole-genome study in *H. guttulatus*, aiming to characterize and understand the origin and evolution of the genomic landscape of divergence between lineages and ecotypes. We generate a near chromosome-level assembly of the *H. guttulatus* genome and describe genetic variation across habitats throughout the range distribution of the species. We found the presence of not a single, but two inversions of megabase size, which may have played an important role in ecotype formation in the Mediterranean Sea. We then attempt to reconstruct the evolutionary history of these two inversions, especially focusing on their origin, divergence, and subsequent evolution, including gene exchange between the inverted alleles. Our study highlights interactions between two inversions in the same system and the impact of gene flux between arrangements.

2 | MATERIALS AND METHODS

2.1 | Sampling and DNA extraction

A total of 112 samples from different habitats across the species range were used for whole-genome resequencing (Table S1). Different DNA extraction protocols were applied to three types of samples that either originated from published studies, our own laboratory collection, or museum collections. (i) Samples from published studies (Barry et al., 2022; Riquet et al., 2019) consisted of non-lethal fin or tail clips preserved in 95% ethanol at -20°C . Their genomic DNA was extracted using the Nucleospin Tissue Kit® (Macherey-Nagel, Germany) following the manufacturer's protocol. (ii) Dried seahorse samples were obtained from private collections by means of a call to the public in a local newspaper. The approximate date (1935–2010) and collection site were recorded based on information provided by the donors. The dorsal fin of each

dried seahorse was scratched to collect c.a. 20 µg of tissue powder. Genomic DNA was extracted using a standard Cetyltrimethyl Ammonium Bromide (CTAB) Chloroform:Isoamyl alcohol (24:1) protocol (Doyle & Doyle, 1987). (iii) Lastly, four alcohol-preserved museum samples (1856–1898) were provided by the National Museum of Natural History (Paris, France). Tissue fragments (1–2 mm) were obtained by internal needlepoint sampling (Haÿ et al., 2020) and subjected to overnight enzymatic digestion (40 µL proteinase K at 25 mg/mL for 500 µL volume, at 56°C). DNA was extracted using a phenol-chloroform method (Campos & Gilbert, 2012) in a dedicated clean lab facility located at the Institute of Evolutionary Science of Montpellier (ISEM, France).

2.2 | Assembly and annotation of the *H. guttulatus* reference genome

We performed high-coverage linked-read sequencing of an Atlantic *H. guttulatus* individual from Hossegor lagoon (Bay of Biscay, Hgutt_GA_13) to generate a high-quality reference genome assembly (hereafter referred to as Hgutt_V1). Fresh gill and muscle tissue were solubilized in a 25 mL solution of TNES-Urea (10 mM Tris-HCl, 120 mM NaCl, 10 mM EDTA, 0.5% SDS, 4 M urea, pH 8) during 4 weeks at 20°C . High-molecular-weight genomic DNA (HMW gDNA) was isolated using three phenol-chloroform followed by two chloroform extractions after digestion with proteinase K (150 µg/mL). Final precipitation was performed using two volumes of 100% ethanol. The resulting pellet was washed several times in 80% ethanol and resuspended in ultrapure water by heating to 65°C , before being kept at 40°C for 4 days. The length distribution of extracted DNA molecules was assessed by electrophoresis on a TapeStation Genomic DNA ScreenTape assay (Agilent Technologies). Single-stranded DNA damage was treated with the NEBNext FFPE DNA Repair mix and repaired DNA was then subjected to size selection to remove fragments shorter than 40 kb using a PippinHT instrument (Sage Science) with a 0.75% Agarose Gel Cassette. HMW gDNA was submitted to the 10x Genomics linked-read library preparation following the Chromium Genome Reagent Kit v2 protocol at the MGX sequencing facility (CNRS, Montpellier, France). The genome library was sequenced to $\sim 100\times$ on an S1 lane of an Illumina NovaSeq6000 in 150 bp paired-end mode by Genewiz, Inc. (USA), generating ~ 0.3 billion reads.

Raw demultiplexed reads were deduplicated using *nubeam-dedup* (Dai & Guan, 2020) and processed with *process_10xReads.py* (<https://github.com/ucdavis-bioinformatics/proc10xG>) to extract the barcode sequence of each read pair and the number of read pairs associated to each barcode. The distribution of the number of read pairs per barcode was then analysed to identify rare barcodes potentially generated by sequencing errors and over-represented barcodes (Figure S1). A total of 143.7 million paired-end reads carrying 1.455 million retained barcodes were finally extracted using *proc10xG* scripts and used for linked-read-based de novo genome assembly using the *Supernova-2.1.1* software package

(Weisenfeld et al., 2017). Assembled scaffolds were outputted in pseudohap style, with a minimum size set to 1 kb. Assembly statistics of the Hgutt_V1 reference genome were computed and visualized with the *BlobToolKit* v3.5.2 software suite (Challis et al., 2020), combined with an assessment of genome assembly completeness with *BUSCO* 5.4.4 (Manni et al., 2021) using the *actinopterygii_odb10* fish dataset containing 3640 conserved genes. Whole-genome alignment was performed with *Minimap2* (Li, 2018) and visualized using *D-GENIES* (Cabannes & Klopp, 2018) to compare and anchor *H. guttulatus* scaffolds to the chromosome-scale assembly of the closely related *H. erectus* (Li et al., 2021), provided by the authors. We also used the genome sequence of a north Atlantic long-snouted seahorse (UK), which was assembled by Iridian Genomes by ordering and orienting pre-assembled contigs based on other fish reference genomes (Accession PRJNA481552, hereafter called HguttRefA).

We used *RepeatModeler2* (Flynn et al., 2020) for de novo repeat finding and identification of the unique transposable element families present in the seahorse genome. In addition, we searched for tandem repeats (TRs) following the strategy developed by Melters et al. (2013), using the same parameter values to run *Tandem Repeats Finder* v4.09.1 (Benson, 1999) within the *pyTanFinder* pipeline (Kirov et al., 2018). We then used *RepeatMasker* v4.0.5 (<http://repeatmasker.org>) to perform repeat annotation and masking of the identified repeat elements. Structural gene annotation was performed using the RNA-Seq pipeline in *Braker2* v2.1.6 (Br  na et al., 2021), which manages the training of the gene prediction tools *GeneMark-ET* (Lomsadze et al., 2014) and *AUGUSTUS* (Stanke et al., 2008). In brief, we used seven RNA-Seq libraries from NCBI's SRA (Accessions SRX565152-57, SRX20881937), totalling 27.8Gb of Illumina short reads, and mapped them to the soft-masked reference genome with *HISAT2* v2.0.4 (Kim et al., 2019). RNA-Seq alignment information was used to iteratively train *GeneMark-ET* to generate initial gene structures, which were passed to *AUGUSTUS* along with RNA-Seq mapping information to generate final gene predictions. Functional annotation of the predicted gene coding sequences was finally conducted using *eggNOG-mapper* v2 (Cantalapiedra et al., 2021), which relies on precomputed Orthologous Groups (OGs) to transfer functional information using phylogenetically refined orthology assignments.

2.3 | Library preparation and whole-genome resequencing

Whole-genome sequencing (WGS) libraries were prepared for 86 samples using the Ovation Ultralow System V2 library preparation kit (NuGEN/Tecan) from 100ng DNA input (when possible) following the manufacturer's instructions. We used unique dual indexing to minimize the effect of index-hopping, and PCR cycles were adapted to the amount of input DNA (9–20cycles). Libraries were pooled in equimolar ratio and sequenced to different coverage depths on a single S4 flow cell on a NovaSeq6000 instrument (Illumina) to generate 150bp paired-end reads.

2.4 | Sequence processing and alignment

To complement our dataset, WGS data for an additional 26 seahorse samples were obtained from Barry et al. (2022) (TruSeq DNA PCR-free libraries sequenced on Illumina NovaSeq6000, GenBank Sequence Read Archive, accession BioProject ID PRJNA777424). Therefore, our final combined dataset consisted of 112 samples (hereafter referred to as the “full dataset”, Table S1). Raw demultiplexed reads were processed using *fastp* (v0.23.1) (Chen et al., 2018) with the “--merge” option, in order to stitch together paired-end reads with overlapping sections. Both merged and unmerged reads were aligned to our reference genome using BWA-MEM (BWA v0.7.17; H. Li, 2013). Picard (v2.26.8) («Picard toolkit», 2019, Broad Institute) was used for sorting read alignments, marking duplicates, and adding read groups. DNA damage patterns in older dried and museum samples were visualized using PMDtools (v0.60) (Skoglund et al., 2014) (Figure S2).

2.5 | Variant calling of medium- to high-coverage samples

Forty-eight samples with sufficient coverage (~5–50x) were selected for variant calling in such a way that this subset contained five individuals in certain key locations (from now on referred to as the “GATK dataset”, Table S1, indicated by variant calling=Yes). Variants were called using the GATK best practices workflow (McKenna et al., 2010; Van der Auwera et al., 2013) without performing Variant and Base Quality Score Recalibration (VQSR and BQSR). First, individual GVCF files were created from bam files with HaplotypeCaller (GATK v.4.1.8.0). This information was then stored in a GVCF database using GenomicsDBImport, and VCF files (one file per scaffold) were generated with GenotypeGVCFs. After concatenation, the resulting VCF was filtered for indels, multiallelic SNPs, and missing data (“--max-missing 0.9”). Our detailed workflow and commands used are provided in File S2. An Rmarkdown report produced with flex dashboard (Sievert et al., 2022), showing detailed results across various bioinformatic steps and analyses, is provided in the Appendix S1.

2.6 | Population structure

A principal component analysis (PCA) was carried out on samples with sufficient coverage (>3x) from the full dataset without calling genotypes. To accomplish this, genotype likelihoods were calculated using ANGSD (v0.933) (Korneliussen et al., 2014) with the GATK model (“GL 2”) and the following parameters: “-doMajorMinor 1 -minMapQ 30 -minQ 20 -doMaf 1 -doCounts 1 -minMaf 0.05 -uniqueonly 1 -remove_bads 1 -C 50 -baq 1 -doCov 1 -doIBS 2 -makeMatrix 1 -ref referencegenome_Hgutt_V1.fa”. In addition, the first 5bp were trimmed off of reads (–trim 5) to account for DNA damage patterns due to cytosine deamination in historical samples.

The analysis was restricted to sites with a strong probability of being SNPs ("SNP_pval 1e-6"). Finally, sites were not considered if the total depth across samples was greater than twice the sum of mean coverage per sample ("setMaxDepth"), in order to avoid artefacts in problematic regions that received unexpectedly high coverage (low-complexity and duplicated regions), or if less than half of the individuals had data ("minInd"). The same filters were used in subsequent analyses conducted with ANGSD, unless specified otherwise. To perform a PCA while minimizing the effect of linkage disequilibrium, we used ANGSD with the "-sites" option and provided a file containing one randomly selected SNP per 10kb window in every 500kb interval. To complement this genotype likelihood-based approach, a PCA was also performed on called genotypes from the GATK dataset using the R package SNPRelate (v1.28.0) (Zheng et al., 2012). Lastly, we conducted local PCA on the GATK dataset to capture variation in population structure along the genome in order to identify outlying patterns indicating the presence of putative chromosomal inversions. To this end, local PCA was performed in non-overlapping windows of 5kb using the R package lostrust (v0.0.0.9, Li & Ralph, 2019).

2.7 | Analysis of large structural variants

Genomic regions that were identified in local PCA as being candidates for the presence of large SVs (i.e., three groups that persist over many consecutive windows) were more specifically tested for evidence of chromosomal inversions in linked-read sequencing data. First, we used abrupt signal shifts flanking outlying regions in the local PCA to determine the location of putative inversion breakpoints. We performed manual curation of the reference genome assembly in those regions following (Rhie et al., 2021), using the alignment to the genome assembly of *H. erectus* to confirm the suspected breakpoints. We then used MTG-Link v2.4.1 (Guichard et al., 2023) to perform local reassembly of the linked-reads mapping in the regions surrounding breakpoints. The 10X linked-reads data were preprocessed with EMA v0.6.2 (Shajii et al., 2018) and LRez v2.2.4 (Morisse, Lemaitre, & Legeai, 2021) before running MTG-Link with default settings. The locally reassembled contigs were finally aligned against the Hgutt_V1 reference genome with Minimap2 (H. Li, 2018) to check for local consistency between assemblies and identify possible mis-joins. We characterized repeat content around the inversion breakpoints, looking specifically for the presence of recombinogenic sequences such as long inverted repeats (LIRs). Finally, we used Leviathan v1.0.2 (Morisse, Legeai, & Lemaitre, 2021) for calling SVs using linked-read data information connecting distant regions that share a higher number of barcodes than expected based on their distance. We ran Leviathan on the 10X linked-read data from the Hgutt_V1 assembly mapped to the HguttRefA genome.

In addition, we followed the haplotagging library construction protocol described by Meier et al. (2021) to generate complementary linked-read sequencing data for alternate genotypes at the two

suspected inversions. Namely, we constructed haplotagging libraries for a DD/AA Mediterranean and a CC/AB Atlantic individual, and sequenced them with 2×50bp paired-end reads to an average coverage depth of 12×.

Finally, we performed a functional enrichment analysis of the gene sets contained in each inversion using ShinyGO v0.77 (Ge et al., 2020) with a false discovery rate (FDR) threshold of 0.05.

2.8 | Genomic landscape of divergence and introgression

We characterized the genomic landscape of divergence between localities and habitats using the GATK dataset. Genetic differentiation (F_{ST}), nucleotide diversity (π), and absolute genetic divergence (d_{XY}) were calculated in 25 kb windows using the popgenWindows.py script (https://github.com/simonhmartin/genomics_general). We then sought to characterize two highly divergent regions on Chr2 and Chr12, which represented suspected inversions. We used BAMscorer v1.4 (Ferrari et al., 2022) to assign inversion genotypes in all samples (including low-coverage samples). This first consisted of classifying all samples in the GATK dataset with regard to their haplotypes, based on PCA groupings and individual heterozygosity values calculated with VCFtools (v0.1.16) (Danecek et al., 2011). This reference database was then used to score alignments and to ascertain allelic state and haplotype for all samples.

In order to test for introgression with other seahorse species and to determine evolutionary relationships across the genome, we obtained whole-genome resequencing data for *H. erectus*, *H. hippocampus*, *H. zosterae*, *H. capensis*, and *H. comes* from Li et al. (2021) (NCBI BioProjects accession code PRJNA612146), including data for one individual from each species. Reads were processed and aligned to the Hgutt_V1 reference genome as described above. ANGSD was used to call genotypes and to produce a VCF ("-doBcf") containing seven individuals (two *H. guttulatus* that were homozygous for the four alternate inversion haplotypes, i.e., AA-DD and BB-CC, and five individuals from other species). This VCF was filtered to include only fixed sites (no heterozygous genotypes) and was used as input for performing topology weighting using the Twisst pipeline (Martin & Van Belleghem, 2017). This method performs iterative sampling of subtrees and quantifies the proportion of trees that match a certain taxon topology, allowing to study complex evolutionary relationships along the genome (e.g., patterns of introgression). We generated unrooted phylogenies using PhyML (Guindon et al., 2010) as implemented in the script *phyml_sliding_windows.py* (<https://github.com/simonhmartin/twisst>) for windows of 50 SNPs, setting a minimum of 25 non-missing genotypes per individual per window ("--minPerInd 25") and performed topology weighting using the *twisst.py* script. Additionally, we produced a genome-wide consensus tree, as well as consensus trees for the two inversions using the averageTree function from phytools (Revell, 2012).

2.9 | Inferring inversion history from the ancestral recombination graph

We used tsinfer (v0.3.0) (Kelleher et al., 2019) to describe the diversity of genetic relationships between sequenced individuals along the genome. This method takes ancestral recombination events into account to infer the genome-wide sequence of correlated local genealogies, called the ancestral recombination graph (ARG). Given accurate inference, the ARG provides a complete description of the available information on the evolutionary history of a set of related sequences, represented in terms of recombination and coalescence events. Inferring the ARG with tsinfer requires an input VCF file for a set of phased, diploid genomes, and the ancestral state of each mutation present in the INFO field. We carried out statistical phasing and missing data imputation of per-chromosome VCF files using SHAPEIT (v4.2.2) (Delaneau et al., 2019), assuming constant recombination rate of 1 cM per Mb (Stapley et al., 2017) and an effective population size (N_e) estimated based on genome-wide diversity values (Charlesworth, 2009). To infer ancestral states, we used BLAST searches to determine the allelic state of each SNP position in each of three outgroup species (*H. erectus*, *H. kuda* [Accession GCA_901007745.1], *H. comes* [Accession GCA_001891065.2]) using the BLAST+ package (v2.2.28) (Camacho et al., 2009). Flanking regions of 100bp on either side of each SNP were blasted to the reference genome of each outgroup using *blastn*. The top hit was inspected to determine the state of the homologous position in each outgroup using a program which was developed for this purpose (<https://gitlab.mbb.cnrs.fr/ibonnicci/snom>). Thereafter, *estsfs* (v2.04) (Keightley & Jackson, 2018) was run using the Kimura 2-parameter model to infer ancestral state probabilities, and phased VCFs were annotated with the most likely ancestral variant using BCFtools (v1.9) (Danecek et al., 2021).

We reconstructed the ARG by running tsinfer on our phased, oriented VCF containing all chromosomes. Sample objects were created using the CYVCF2 library (v0.30.18; Pedersen & Quinlan, 2017) as in the tsinfer tutorial (<https://tskit.dev/tsinfer/docs/stable/tutorial.html>). The tree sequence was then inferred using a constant recombination rate of $1e-8$ per bp and a mismatch ratio of 1. The ages of the ancestral nodes in the inferred trees were estimated using tsdate (v0.1.5) (Wohns et al., 2022) using parameters " $N_e = 200,000$ ", "*timepoints* = 100", and "*mutation_rate* = $1e-8$ ". The N_e value used was calculated from the nucleotide diversity of the most divergent region of the genome (i.e., Chr12 inversion). This choice did not affect the distribution of inferred coalescent times in the genome background, while enabling tsdate to infer ancient node ages within inversions.

The inferred ARG was used for two purposes. First, we described variation in coalescence times across the genome using time to the most recent common ancestor (TMRCA) for randomly selected combinations of sample haplotypes (i.e., tree leaves) extracted from genealogies along the tree sequence. Second, we characterized local ancestry for each individual haplotype within inversion regions. The signal of an inversion was shown to appear in the ARG as clades of samples that persist over a more extended region of the genome

than would otherwise be expected (Ignatieva et al., 2023). Here, we assumed that, within an inversion, the two branches immediately below the oldest node of each local tree represent the two clades of non-inverted and inverted extended haplotypes. To assign local ancestry to each branch, a clade was considered to belong to non-inverted or inverted ancestry if it satisfied one of two empirically established conditions: (i) the clade contained more than 75% of the haplotype copies from all of the homokaryotypes of a given haplotype, and reciprocally for the alternate haplotype in the other clade; (ii) the clade contained 100% of the copies from all of the homokaryotypes of that haplotype, and the same clade contained less than 75% of the copies from the alternate haplotype. Local ancestry switches along individual haplotypes were used to identify inter-haplotype recombination events within inversions, indicative of gene conversion or introgression.

3 | RESULTS

3.1 | Reference genome

We obtained a near chromosome-level genome assembly of the *Hippocampus guttulatus* Hgutt_V1 reference genome. The assembly contained 3878 scaffolds spanning 451 Mb (424 Mb in scaffolds of at least 10kb, longest scaffold 28.8Mb, scaffold N50 = 18.1Mb, N90 = 5.65Mb, Figure S3), which is close to the genome size of 424Mb predicted by GenomeScope (Barry et al., 2022). Genome completeness was very high (Busco C: 95.6% [S: 94.4%, D: 1.2%], F: 1.6%, M: 2.8%, n: 3640) and comparable to the HguttRefA and the *H. erectus* assemblies (Figure S4a,b). All scaffolds showed strong homology and conserved syntenic relationships with the chromosome-level genome assembly of *H. erectus* (Figure S5), suggesting that Hgutt_V1 is a high-quality assembly. Scaffolds anchoring to the *H. erectus* chromosome assembly produced 22 pseudomolecules that were used to characterize the genomic landscape of divergence (Figure 2).

The repeat landscape showed that interspersed repeats add up to 33% of the genome, with 13% being occupied by unknown repeats (Figure S6). Among these unknown repeats, the clustering analysis of similar tandem repeats revealed a highly abundant TR of approximately 500bp in length, present in more than 5500 copies across the genome, with an accumulated abundance of 2.7Mb. The consensus monomer of this TR, hereafter called Hgutt_Tan9, is a short sequence of 37bp present in all chromosomes (Figure S7). Structural gene annotation based on RNA-Seq data predicted a total of 25,770 coding gene sequences, of which 13,346 unique gene names were identified by functional annotation.

3.2 | Overall genetic structure

We produced genome resequencing data of heterogeneous quality (ranging from 1x to ~50x, Table S1) and therefore characterized

overall genetic structure using a genotype likelihood-based approach after controlling for low-coverage individuals ($<3\times$) (Figure 1b,c). We found evidence for pronounced population structure, even between certain sampling locations that were in close geographic proximity (inset map, Figure 1a). This genetic structure could largely be ascribed to markers that are in tight linkage disequilibrium (LD), since PCA performed on unlinked markers revealed a different pattern (Figure 1c). Here, all Mediterranean marine

samples grouped together with Black Sea samples, separately from the different Mediterranean lagoon populations (Bz, Th, and Mu). In the Atlantic, we also discern a northern (Br and Ga) and a southern (Fa) cluster of samples. By contrast, when all markers are considered (Figure 1b), samples do not group together based on geographical origin. Instead, we observe that the individuals from a given location are grouped into either 1, 2, or 3 clusters that are organized along two orthogonal axes. The clusters from different locations were

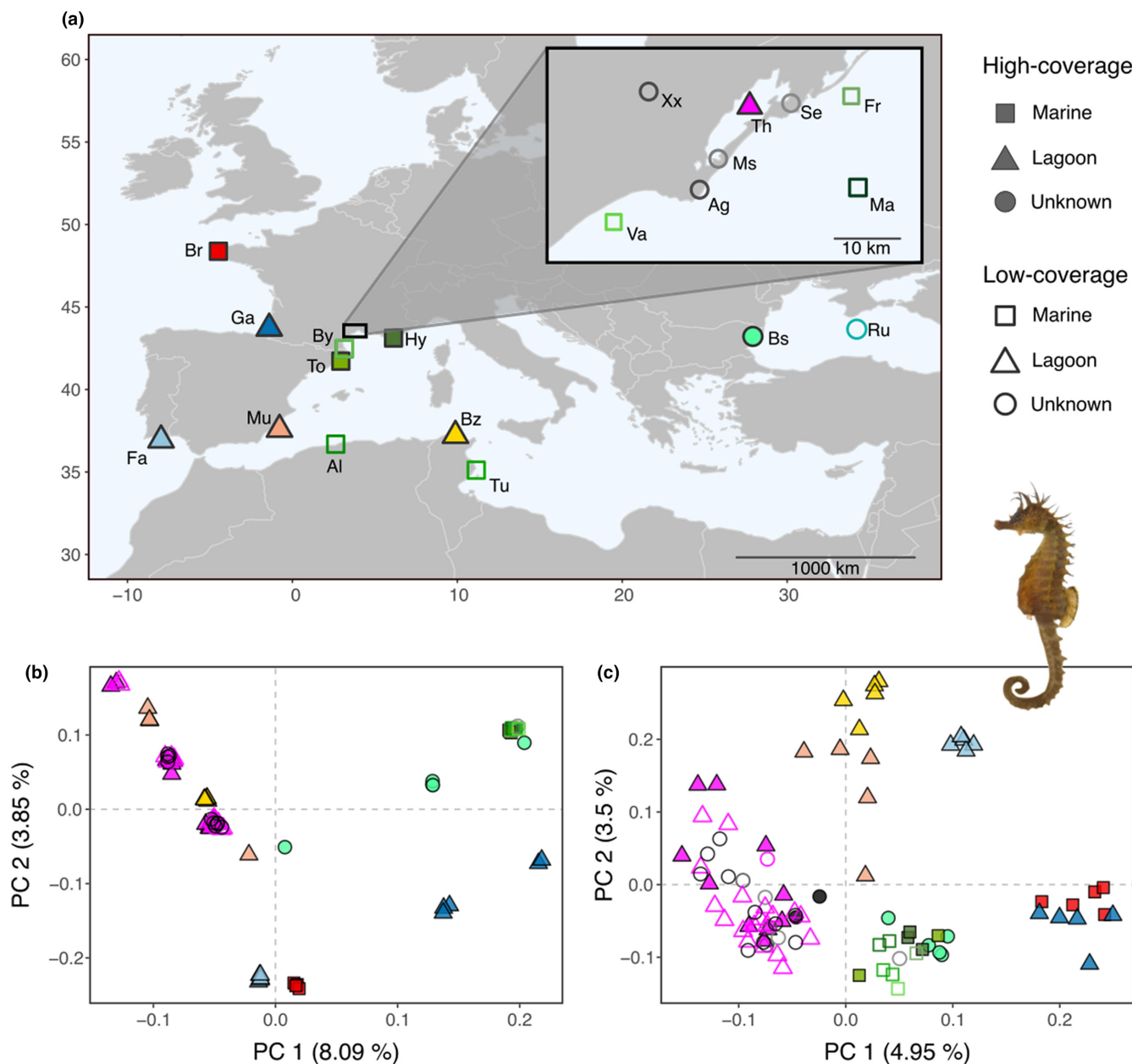


FIGURE 1 (a) Sampling map for the full dataset of *Hippocampus guttulatus* samples ($n=112$). Triangles: lagoon sites. Squares: marine sites. Circles: unknown habitat type. Filled shapes: medium- to high-coverage samples for which individual genotype calling was performed (the GATK dataset). Unfilled shapes: low-coverage samples for which a genotype likelihood-based approach was used. Br: Brest. Ga: Hossegor. Fa: Faro. Mu: Murcia. To: Tossa de mar. By: Banyuls. Va: Valras. Ag: Agde. Ms: Marseillan. Th: Thau lagoon. Se: Sète. Fr: Frontignan. Hy: Hyères. Ma: unknown Mediterranean marine site. Al: unknown Algerian site. Bz: Bizerte lagoon. Tu: unknown Tunisian site. Bs: Varna. Ru: unknown Russian site. Xx: unknown northwestern Mediterranean location. (b and c) PCA of 89 individuals with $>3\times$ coverage based on IBS distances calculated in ANGSD. PCA was either performed using all genome-wide markers (1,478,955 SNPs) (b), or using a subset of markers at linkage equilibrium (898 SNPs) (c). © Seahorse picture Iglésias 2013.

sometimes shifted away from each other (e.g., the three groups of Bs and Ma mirror those of Br and Ga), although they likely captured the same underlying variation. Observing these replicated three-cluster patterns in the PCA is consistent with the presence of large chromosomal inversions, with the three groups representing different inversion genotypes – homokaryotypes with two inverted alleles, heterokaryotypes carrying both haplotypes, and non-inverted homokaryotypes. The segregation of samples along two orthogonal axes, each showing replicated three-cluster patterns, is thus indicative of two polymorphic inversions containing large numbers of sites in strong LD.

In line with this finding, we observed heterogeneous landscapes of differentiation along the genome (Figure 2). Background genomic differentiation between northern and southern Atlantic lineages ranged between 0.06 and 0.15 (first and third quartiles of F_{ST} distribution) (Figure 2a). Differentiation was weaker between Mediterranean marine and lagoon locations, with F_{ST} values ranging between 0 and 0.03 (Figure 2b). For other similar contrasts (i.e., northern vs. southern Atlantic lineage, or Mediterranean marine vs. lagoon comparison), the values were not substantially affected by the choice of the specific locations being compared (Figures S8–S10). These patterns of relatively weak genome-wide differentiation contrasted with high F_{ST} and d_{XY} values found on two chromosomes – Chr2 (Figure 2c) and Chr12 (Figure 2d). Chr12 presented an 8.2 Mb long plateau of high F_{ST} and d_{XY} values in both Atlantic and Mediterranean comparisons, a pattern that has previously been associated with inversions segregating at different frequencies between populations. For Chr2, a differentiation plateau was only observed in the Mediterranean marine versus lagoon comparison. This pattern was also less clear, as the highest values did not occur in one contiguous block but were split across scaffolds and even showed discontinuities within scaffolds.

3.3 | Inversions differentiate ecotypes in the Mediterranean Sea

The presence of an inversion for the block of high differentiation on Chr12 was first supported by the alignment of Hgutt_V1 against HguttRefA (generated from the opposite homokaryotype, Figure S11), as well as with the *H. erectus* assembly. This analysis showed the presence of an 8.2 Mb long inverted segment between the two *H. guttulatus* assemblies, with a putative inversion breakpoint located near position 1.699 Mb on scaffold 14 of Hgutt_V1, which corresponds to the abrupt signal shift in divergence at the beginning of the block. Linked-read-based local reassembly of a 3 kb region centred on the putative breakpoint confirmed the contiguity of the Hgutt_V1 reference. In the middle of this 3 kb region, we found a long inverted repeat (LIR) consisting of two inverted arrays of the Tan9 monomer, with the internal spacer overlapping the inversion breakpoint (Figure S12). The other end of the inversion at the end of scaffold 14 (near 9.877 Mb) also contained a tandem repeat of Tan9 monomers (Figure S6). Linked-read sequencing data obtained from

three individuals showing the three possible genotypes (AA, AB, and BB) showed mapping patterns that were consistent with the presence of a large chromosomal inversion (Figure S12). Finally, the analysis of linked-reads from a BB genotype mapped to the HguttRefA assembly allowed the direct detection of an 8.2 Mb inversion at the breakpoints expected from all previous analyses.

Although we were not able to perform the same detailed analysis for Chr2, since the two *H. guttulatus* assemblies both carry the same haplotype, other patterns suggested that a second inversion is highly likely to be present on this chromosome. Performing PCA separately on Chr2 and Chr12, we observed three clusters along PC1 with the middle group presenting higher heterozygosity than the outer groups (Figure 3a,b). Using BAMscorer (Ferrari et al., 2022), we characterized inversion genotypes in all samples except for one extremely low-quality individual (Hgutt_Se_1898_94, see Appendix S1). These scores were consistent with the groupings observed from the PCA (Figure 3a,b) – that is, samples that were classified as heterokaryotypes were in the middle PCA group, and homokaryotypic samples were in the outer groups. These patterns confirm that the multiple blocks of high F_{ST} on Chr2 are in perfect LD, and that they collectively segregate as a single contiguous variant, despite the discontinuous signature observed in the divergence landscape (Figure 2c). Consequently, our results indicate that there are two polymorphic inversions segregating in seahorse populations and that these play an important role in the differentiation between lineages and ecotypes. As for their gene content, the 8.2 Mb inverted segment on Chr12 contained a total of 409 annotated genes showing a significant 3.7-fold enrichment for the oestrogen signalling pathway (Enrichment FDR=0.042, 11 of 134 annotated genes in the pathway). The high-LD region on Chr2 had a total of 419 annotated genes and showed a significant 2.5-fold enrichment for the neuroactive ligand–receptor interaction pathway (Enrichment FDR=0.047, 20 of 372 annotated genes in the pathway).

For each inversion, we characterized the relative frequencies of the three alternative genotypes across sampling locations and habitats (Figure 3c,f). We found that the inversion on Chr12 (B12) was present in one of the two homozygous states (either AA or BB) in almost all locations (Figure 3f). Atlantic samples from the northern part of Biscay (Br), the Gulf of Cadiz (Fa), and samples from Mediterranean lagoons (Th, Mu, and Bz) exclusively presented homokaryotypes for the A allele, while Mediterranean marine sites only presented homokaryotypes for the B allele. Fine-scale variation in inversion genotypes was especially pronounced in the Mediterranean Sea, since samples that were collected only a few kilometres apart showed differential fixation for B12 between marine and lagoon habitats (inset map, Figure 3f). Samples for which precise location information was not available (grey circles) were either AA or BB homokaryotypes, with A alleles probably hailing from lagoon habitats in these locations and B alleles from the sea. Samples of unknown origin (“Xx”) mostly carried the lagoon haplotype, probably reflecting frequent incidental catches in lagoon habitats during artisanal and non-professional fishing activities. There were only two sampling locations which were polymorphic for B12: the Hossegor

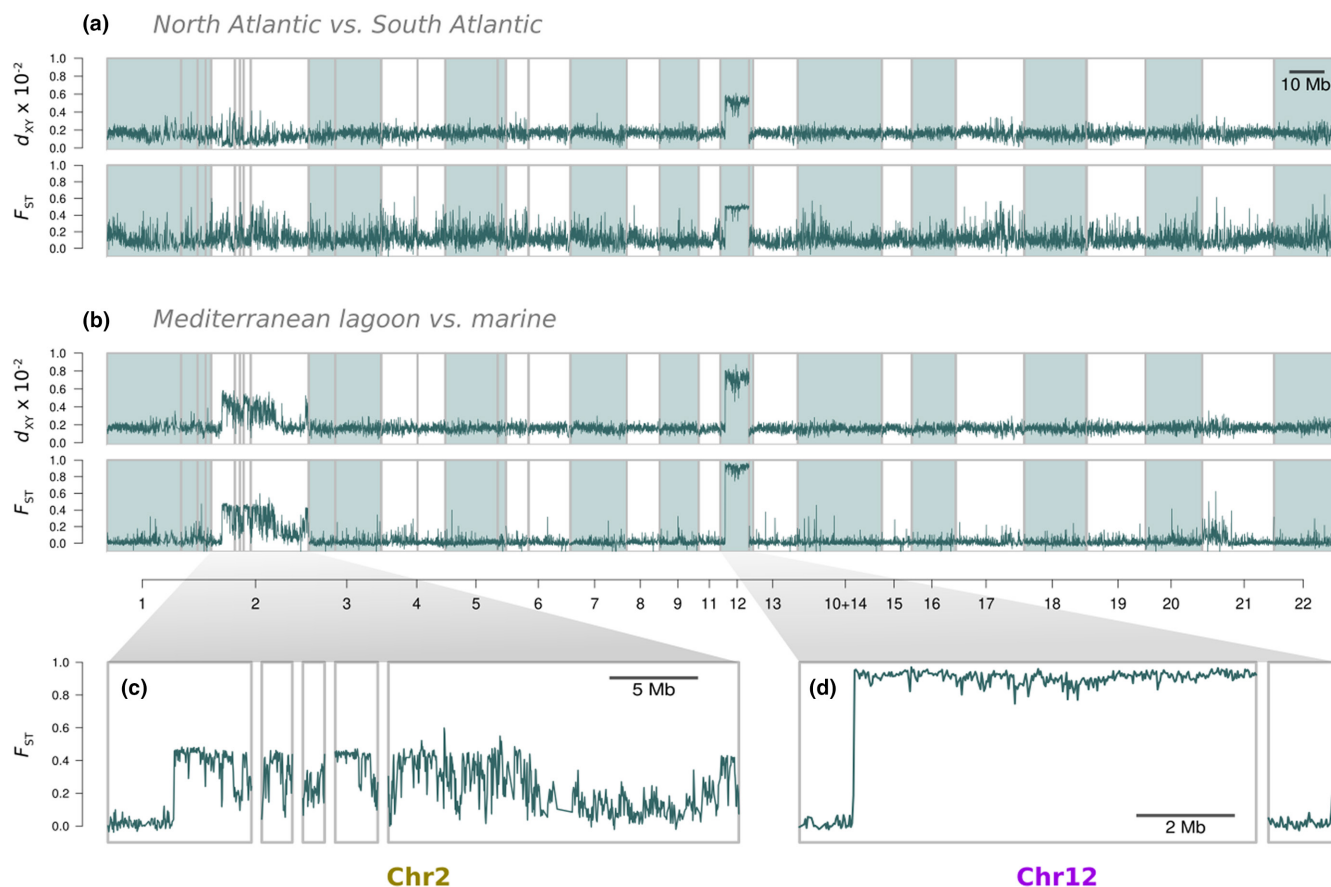


FIGURE 2 Genomic landscape of absolute divergence ($d_{xy} \times 10^{-2}$) and differentiation (F_{ST}) calculated in 25 kb non-overlapping sliding windows between a north Atlantic (Ga) and a south Atlantic (Fa) population (a) and between a Mediterranean lagoon (Mu) and a Mediterranean marine (To and Hy) population (b), using five high-coverage samples per population. *Hippocampus guttulatus* scaffolds (grey rectangles) were aligned to the chromosome-level assembly of *H. erectus* and are displayed according to their homology with the 22 *H. erectus* chromosomes (marked by alternating white and blue rectangles). Regions of high divergence clustered on two chromosomes: Chr2 (~32 Mb) (c) and Chr12 (~11 Mb) (d), which each carry a large chromosomal inversion. Exact ordering and orientation of *H. guttulatus* scaffolds within chromosomes 2 and 12 of *H. erectus* could not be determined due to multiple rearrangements between species (Figure S5).

marine lake in the Bay of Biscay (Ga) and Varna in the Black Sea (Bs), which presented samples from each of the three inversion genotypes (AA, AB, and BB). The inversion located on Chr2 (B2) showed different distribution patterns compared to that of B12, since this polymorphism was only found in Mediterranean lagoons (Figure 3c). Both in the Atlantic and in the Black Sea, all samples were fixed for haplotype C. The alternative D haplotype was found at varying frequencies (28–70%) in Thau, Bizerte, and Murcia lagoons. All samples caught in the Bizerte lagoon were CD heterokaryotypes ($n=5$), while three out of five samples from Murcia were DD homokaryotypes. All samples carrying one or two D alleles had been classified as an AA homokaryotype for B12 (including those from unknown habitats), indicating that the Chr2 polymorphism is private to Mediterranean lagoon populations.

Our results thus indicate that two large chromosomal inversions covering roughly 9% of the seahorse genome largely drive ecotype differentiation in the Mediterranean Sea. Lagoon populations are fixed at B12 for the A haplotype, whereas populations in the sea only carry the alternate B haplotype. In addition, there is a second

inversion on Chr2 (with haplotypes C or D) which differentiates marine and lagoon ecotypes based on a polymorphism which is private to the lagoon ecotype. Even though the B12 inversion is also polymorphic in the Atlantic, it does not show differential fixation between habitat types in this zone. Instead, Atlantic populations to the north of Hossegor carry the A allele, while both A and B alleles are present from Hossegor to the south of Portugal (see also Riquet et al., 2019). Interestingly, the B haplotype on B12 and the D haplotype on B2 were never found together in the same individual (Figure 3g), suggesting a possible negative interaction between the two alleles. In what follows, we sought to further characterize the respective histories of the two inversions, including their origins and ages.

3.4 | Evolutionary history of the two inversions

An important question regarding the origin of the inversions is whether the heterogeneous divergence landscape has resulted from

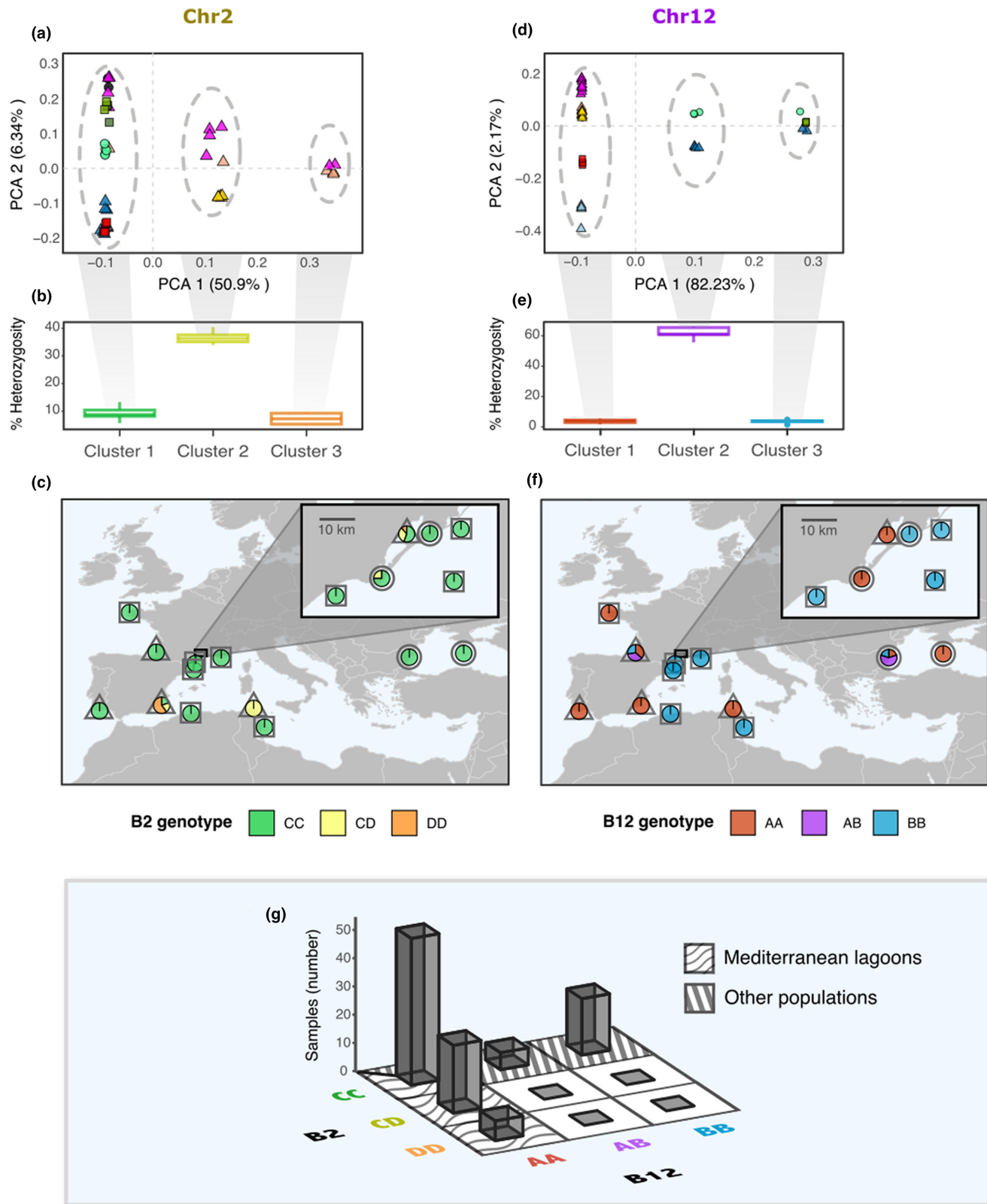


FIGURE 3 Molecular diversity patterns observed for Chr2 (a–c) and Chr12 (d–f), which carry large inversions (B2 and B12). (a and d) Chromosome-wide Principal Component Analysis (PCA) of 48 medium- to high-coverage samples (the GATK dataset) using the same symbols as in Figure 1. The three clusters for Chr2 (a) and Chr12 (d) are illustrated by grey ellipses. (b and e) Boxplots of observed heterozygosity (percentage of heterozygous sites) for each cluster. Heterozygosity was calculated at a chromosome-wide scale and averaged per sample. (c and f) Maps of inversion genotype frequencies with respect to habitat type. Pie charts represent inversion genotypes, and grey symbols indicate marine (square), lagoon (triangle), or uncharacterized (circle) habitats. (g) Observed genotypic combinations at inversions B2 and B12, showing distinct associations between Mediterranean lagoons (B12 fixed for the A haplotype and B2 polymorphic) and other populations (B2 fixed for the C haplotype, B12 locally polymorphic), and the complete absence of four genotype combinations.

the differential erosion of past genome-wide divergence, or alternatively, from the emergence of newly divergent regions (i.e., secondary contact versus primary divergence). To address this question, we studied the distribution of coalescence times along the genome as inferred by tsdate (v0.1.5). Outside the inversion regions, we did not find a strong signature of ancient coalescence in the form of deep time to the most recent common ancestor (TMRCA) (Figure 4e). Only the two chromosomes carrying inversions showed the presence of TMRCA values that were up to five times older than in the genome background, which also corresponded to their high d_{XY} values (Figure 2). What is more, except for the inversions, regions of high F_{ST} were not shared between Atlantic lineages and Mediterranean ecotype contrasts, suggesting a lack of parallelism in the genome background (Figure S13).

Topology weighting conducted using Twisst (Martin & Van Belleghem, 2017) revealed that the most predominant topologies – both in the genome background and in the inversions – were consistent with the currently accepted genome-wide phylogeny of the genus (Li et al., 2021; Stiller et al., 2022). However, they simultaneously showed high levels of incomplete lineage sorting (ILS) near the ancestral nodes of the genus. If one of the two haplotypes on either Chr2 or Chr12 had been introduced through introgression from a closely related species, we would have expected the tree topologies in this region to differ from genome-wide topologies. In this case, an introgressed haplotype carried by an *H. guttulatus* individual would have grouped closer to the donor species than with the alternate haplotype of *H. guttulatus*. The topologies that did not group alternate haplotypes together within *H. guttulatus* amounted to only 4.57% and 15.7% of all topologies that were observed in the B2 and B12 regions, respectively (see Figure 4c,d for the five most frequent of these topologies). Even when the two *H. guttulatus* haplotypes were not in the same grouping, they were generally not placed in distant positions (i.e., they still occurred in the same sub-branch). These topologies did not show any particular haplotype tending to group with a potential donor species, and both haplotypes showed shifted positions, most probably due to ILS. Given these findings, we conclude that the two inversion polymorphisms in *H. guttulatus* were not likely introduced through introgression with a closely related species.

Our results indicate that the B2 and B12 polymorphisms are most likely explained by chromosomal inversion events that took place within the *H. guttulatus* lineage. These inversions have been maintained as intraspecific polymorphisms for a long enough period of time for divergence to have accumulated between haplotypes. The divergence between A and B haplotypes on Chr12 was particularly high, since d_{XY} values for comparisons between opposite homokaryotypes ranged between 0.69% and 0.76% (first and third quartiles). Absolute divergence between opposite B2 homokaryotypes was slightly lower and showed more variance (first to third quartiles, 0.36–0.57%) (Figure S9). TMRCA inferred in the B12 region were also higher than for the B2 region, and consensus phylogenies constructed with PhyML showed a deeper split for haplotypes A and B (Figure 4c) than for C and D (Figure 4b). We conclude that

the inversions probably do not have the same age, and that the B12 polymorphism emerged before B2. Furthermore, it should be noted that divergence might have been underestimated in our study due to reference bias, since reads from AA individuals were mapped to our B reference (Hgutt_V1).

In addition to these inversions, other intra- and inter-chromosomal rearrangements were evidenced in alignments between our reference assembly (Hgutt_V1), HguttRefA and the *H. erectus* assembly. Due to the abundance of these structural rearrangements, we were not able to determine the ancestral arrangement for inversions B2 and B12 through comparison with the *H. erectus* assembly. These analyses showed that *H. erectus* chromosomes Chr11 and Chr12 were fused in the HguttRefA assembly to form chromosome JAOYMQ010000004.1 (Figure S11). This could potentially indicate that the A haplotype (carried by HguttRefA) on Chr12 is involved in a chromosomal fusion with Chr11 (contrary to the B haplotype), or alternatively, that this might represent a technical artefact of the HguttRefA assembly, which is based on short-read data. As for B2, we speculate that the discontinuous signal of high F_{ST} associated with this inversion (Figure 2c) might be due to additional rearrangements that have altered the collinearity between the C and D haplotypes. The D haplotype might present a different intra-chromosomal structure to the C haplotype, resulting in a heterogeneous divergence landscape when mapped to the C reference. However, our analyses did not allow us to confirm this hypothesis, since both *H. guttulatus* assemblies were CC homokaryotypes. Alternatively, regions of low F_{ST} and d_{XY} could be due to local erosion of divergence through recombination between inverted haplotypes.

We tested for gene flux between inverted and non-inverted haplotypes, which can be expected in the form of double crossover events or gene conversion. Plotting PC1 coordinates from local PCA along inversion regions allowed us to locate genome windows which were candidates for gene flux. This was based on local deviations from the three-cluster pattern typical of inversions (three groups of horizontal lines representing the three genotypes, Figure 5a,d). B12 showed the classical three-cluster pattern at a large scale, whereas B2 showed a more “choppy” landscape, with long stretches of “irregular” patterns separating three-cluster windows. For both inversions, these deviated windows were further examined for evidence of local gene flux. We used inferred ARG trees to determine local ancestry along both inversions (Figure 5b,c,e). We determined which trees showed the expected topology of reciprocal monophyly between non-inverted and inverted haplotypes grouping in two different clades (Figure 5b) and which trees showed a discordant topology (Figure 5c). Local trees grouping opposite haplotypes in the same clade indicated regions which were locally introgressed with a tract from the alternate haplotype. This allowed us to perform chromosome painting for each non-recombinant block associated with a particular tree. This approach revealed large-scale haplotypes corresponding to inverted and non-inverted ancestry, interspersed with local ancestry variation at a finer scale. Chromosome painting strongly reflected the patterns observed in the local PCA and F_{ST}

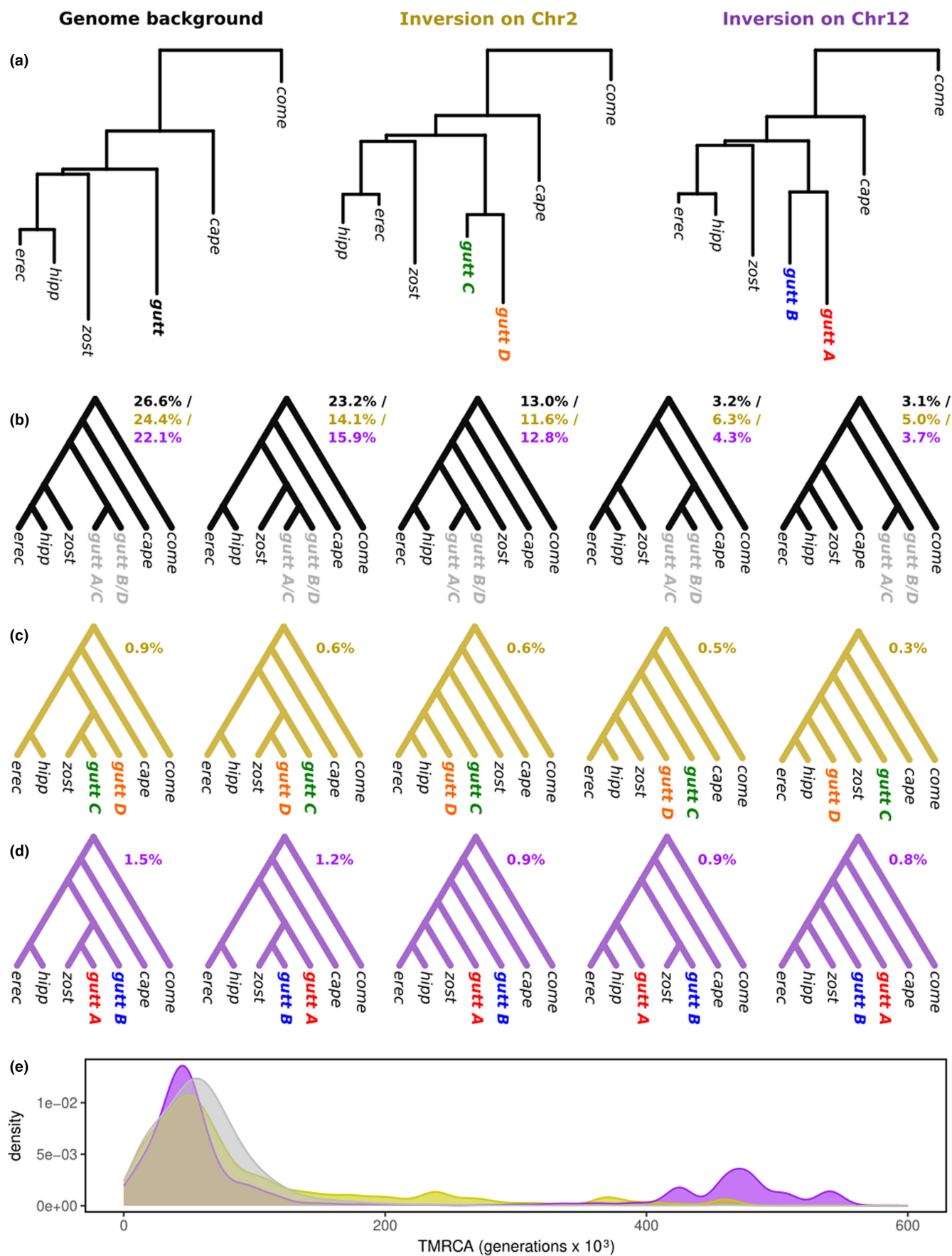


FIGURE 4 (a) Maximum-likelihood phylogenies for different regions of the seahorse genome. (b) The five most predominant seven-taxon topologies identified with Twisst in the genome background as well as in the inversions. Relative abundance among all topologies for each specific region is written as a percentage for the genome background (top), for inversion B2 (middle, leaves *gutt C* and *gutt D*), and for inversion B12 (bottom, leaves *gutt A* and *gutt B*). (c and d) Among topologies not placing alternate haplotypes as sister lineages, the five most common are shown for inversions B2 (c) and B12 (d). (e) Distribution of coalescent times estimated with *tsdate* (v0.1.5) in the genome background (grey), on Chr2 (yellow) and on Chr12 (purple). The TMRCA (in units of 1 k generations) was extracted for any random pair of haplotypes in every local genealogy along the genome. *Come*, *H. comes*; *Cape*, *H. capensis*; *Zost*, *H. zosterae*; *Hipp*, *H. hippocampus*; *Erec*: *H. erectus*.

landscapes (see [Figures S14](#) and [S15a,b](#) for painted chromosomes of the entire GATK dataset). For lower quality samples, we found evidence for phasing errors in the form of switches between maternal and paternal haplotypes, which were visible in heterokaryotic samples. These errors did not prevent us from locating introgressed tracts in homokaryotypes, which spanned up to ~100kb for B12 and up to ~500kb for B2.

4 | DISCUSSION

We present the first genome-wide study of intraspecific diversity in *H. guttulatus*, investigating multiple aspects of population structure related to geography, habitat type, and genome architecture. We further characterize divergence between two previously described geographical lineages in the Atlantic and between marine-lagoon ecotypes in the Mediterranean Sea (Riquet et al., 2019). Our results reveal a sharp contrast between low levels of genome background differentiation and the substantial haplotype frequency differences observed at two megabase-scale chromosomal inversions differentiating Mediterranean ecotypes. We characterize the origin of these inversions and discuss the possible mechanisms responsible for their long-term maintenance. Lastly, we reveal patterns suggestive of gene flux between inverted alleles and address how this process could shape the dynamics and evolutionary fate of seahorse inversions.

4.1 | Evolutionary origin of seahorse inversions

Our results indicate that inversions B2 and B12 have been maintained as polymorphisms for hundreds of thousands of generations, that is, well beyond the mean coalescence time inferred within populations (<100k generations, [Figure 4e](#)). Consistently, these regions also showed remarkably high levels of raw divergence compared to the rest of the genome (d_{xy} up to 1%, vs. maximum of 0.25% elsewhere). Moreover, the levels of nucleotide diversity associated with each haplotype were, given their respective frequencies, consistent with genome-wide average values. It is therefore likely that sufficient time has elapsed since the appearance of the inverted haplotypes for them to have progressed towards mutation-drift equilibrium. It should, however, be noted that estimating coalescence times with *tsdate* is known to be hampered by technical limitations (e.g., underestimation of deeper coalescence times; Brandt et al., 2022) that

prevented our ability to precisely determine and compare the ages of B2 and B12.

Our finding that the two seahorse inversions represent ancient polymorphisms thus raises the question of their origin. One way that a divergent inverted haplotype may be introduced is through hybridization with a closely related species (e.g., Hsieh et al., 2019; Jay et al., 2018). We did not find evidence for introgressed ancestry at either of the inversions, since topology analysis did not support similarity between any inversion haplotypes and a potential donor species ([Figure 4b–e](#)). Instead, all major sources of genealogical conflict could be explained by ILS occurring at deep nodes in the *Hippocampus* genus phylogeny. Alternatively, the inversions could have resisted re-homogenization by gene flow upon secondary contact between two divergent lineages, due to recombination suppression (Lundberg et al., 2023; Rafajlović et al., 2021; Yeaman, 2013). However, we did not find any regions of deep coalescence (i.e., expected remnants of recombined divergent ancestries) comparable to that found in the inverted regions. It is thus most likely that the inversions have emerged within the *H. guttulatus* lineage and that they have remained polymorphic at the species level for a long period of time.

Molecular mechanisms facilitating the emergence of inversions have been studied in detail using long-read sequencing in deer mice, showing that inversion breakpoints tend to occur in centromeric and telomeric regions and that they are sometimes flanked by LIRs (Harringmeyer & Hoekstra, 2022). Here, we were able to directly demonstrate, using linked-read data, that the 8.2Mb inversion B12 occurs near a chromosome extremity and that it has its breakpoints in an approximately 1kb LIR bearing a Tan9 monomer. Tandem repeats containing Tan9 are widespread throughout the *H. guttulatus* genome and may therefore facilitate ectopic recombination, leading to an increased rate of formation of new structural variants.

4.2 | Maintenance of polymorphisms B12 and B2

A chromosomal inversion might undergo a series of evolutionary changes from the time of its appearance to its loss or fixation (Faria, Johannesson, et al., 2019). We thus sought to distinguish between the mechanisms that might maintain the two inversion polymorphisms in the seahorse and to provide explanations for their respective patterns. Our results support that the B12 polymorphism is most comparable in its characteristics to Type I inversion polymorphisms (i.e., divergent between populations), while B2 probably calls

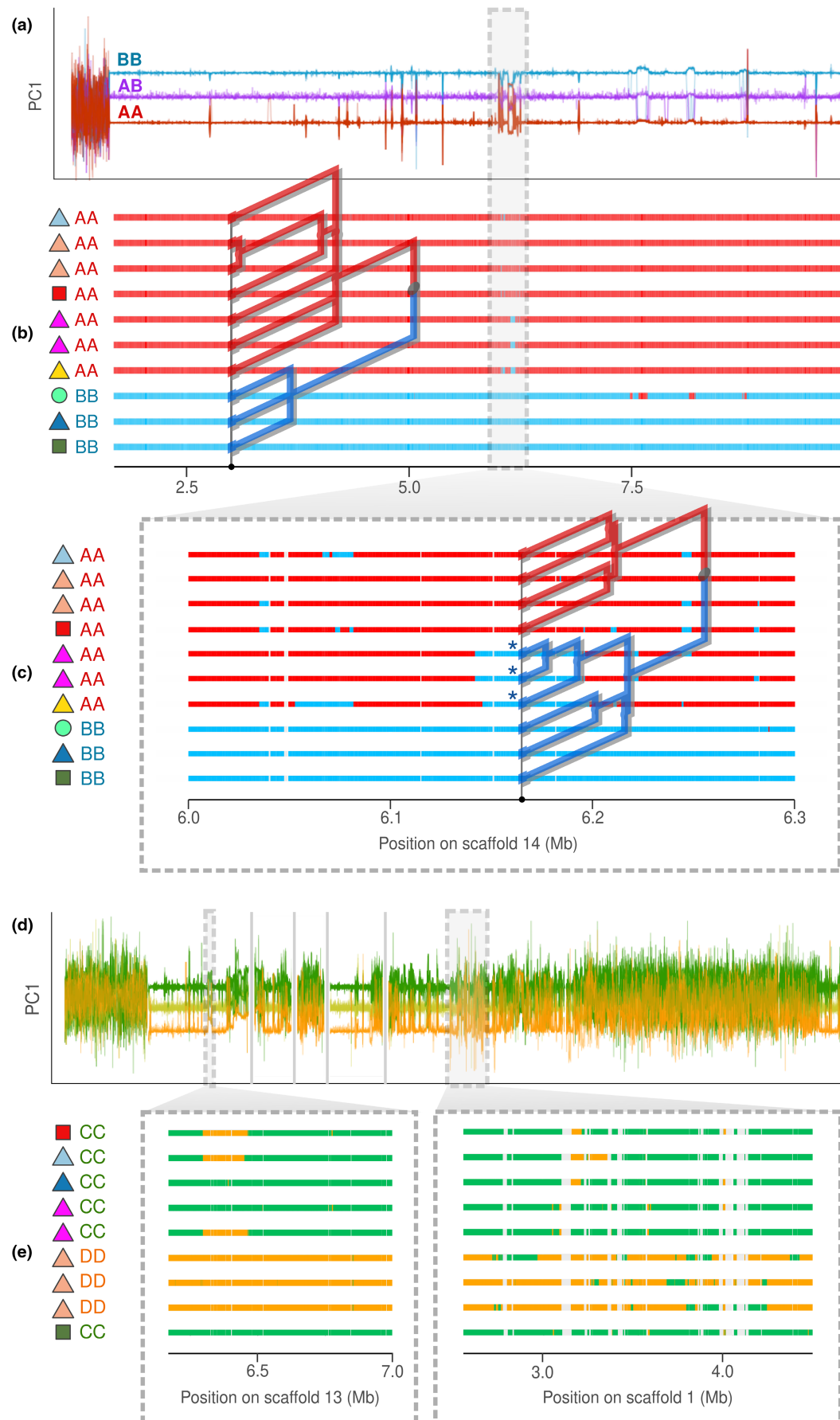


FIGURE 5 Gene flux between the inverted and non-inverted haplotypes on Chr12 (a–c) and on Chr2 (d and e). (a and d) PC1 coordinates from local PCA plotted for non-overlapping 5 kb window along the inversions. Individual lines represent samples and are coloured according to inversion genotype. Vertical grey lines indicate the edges of scaffolds composing Chr2. (b, c and e) Chromosome painting showing local ancestry within inversion regions as determined by our ARG-based approach. The superposed trees were constructed with *tsinfer* (v0.3.0) and illustrate how local ancestry at a given position was determined for each haplotype leaf. Horizontal coloured bars represent a subsample of all haplotypes and colours indicate inverted or non-inverted ancestry. Symbols on the left of each chromosome indicate sampling location and habitat type (see Figure 1a), as well as the large-scale inversion genotype for each individual. (c) Enlarged view of an introgressed region of B12, where certain A haplotypes (red) locally show B ancestry (blue). The tree at position 6.175 Mb shows that three recombinant A haplotypes (marked by asterisks) are grouped in the same clade as the B haplotypes. (e) Enlarged views of two regions on Chr2 showing different levels of introgression.

on the interaction of forces linked to both Type I and Type II (i.e., within-population variation) polymorphisms.

The B12 polymorphism corresponds to the genomic island identified by Riquet et al. (2019) (Figure S16) as the main region differentiating geographical lineages in the Atlantic and ecotypes in the Mediterranean Sea. B12 shows local fixation for a given haplotype within almost all populations (Figure 3f), in accordance with the pattern expected for Type I polymorphisms (Faria, Johannesson, et al., 2019). In this scenario, within-population polymorphism is not maintained unless there is sufficient gene flow, but among-population variation is conferred by divergent selection on alternate haplotypes. Given the differential fixation of B12 haplotypes between adjacent lagoon (A haplotype) and marine (B haplotype) sites in the Mediterranean Sea, it is possible that the inversion is either underdominant, under negative epistasis, or affected by selection against locally disadvantageous haplotypes. The general lack of recombination between inverted haplotypes ensures that the genes in the A and B haplotypes are inherited together, collectively forming a barrier to gene flow. Over a long period of time, the accumulation of new mutations between inverted haplotypes leads to high levels of divergence, as is reflected by high d_{xy} values observed between AA and BB homokaryotypes (Figure 2b). Although there is little opportunity for gene flow between alternate haplotypes, recombination can proceed normally within every population that is fixed for a given haplotype, since crossing over formation is unimpeded in homokaryotypes. In this way, the accumulation of mutation load in inverted regions is not expected to be substantially higher than in the collinear genome (Berdan et al., 2021). Empirical examples are provided by inversions in deer mice that show local fixation and lack of mutation load (Harrington & Hoekstra, 2022), as well as inversions in sunflower populations fixed for one arrangement having lower mutation load compared to polymorphic populations (Huang et al., 2022).

If inversion B12 is a Type I polymorphism under divergent selection, a strong association of each haplotype with either the lagoon or marine habitats would be expected across the entire range. However, a few seahorse samples from Atlantic (Hossegor, Bay of Biscay) and Black Sea (Varna) populations showed a breakdown of the strong association between B12 genotype and habitat type. In these populations, both haplotypes A and B were present and heterokaryotypes were locally common (4 out of 11 and 3 out of 5 samples, respectively), as already observed by Riquet et al. (2019). A possible explanation for this surprising pattern involves a predominantly

intrinsic barrier effect of B12 and possible epistatic interactions with selected loci present on other chromosomes, such as the B2 inversion (see below).

In contrast to B12, the B2 inversion is maintained as a polymorphism which is private to Mediterranean lagoons. Classical work that has addressed the long-term maintenance of inversion polymorphisms reported an important role of balancing selection in maintaining within-population variation (e.g., reviewed in Llaurens et al., 2017; Wellenreuther & Bernatchez, 2018). This results in Type II polymorphisms *sensu* Faria, Johannesson, et al. (2019), which could explain the distribution of the B2 polymorphism in the Mediterranean Sea. We observed intermediate haplotype frequencies and relatively high frequencies of CD genotypes in Mediterranean lagoons (Th: 12 out of 36; Mu: 3 out of 5; Bz: 5 out of 5), possibly pointing to a form of heterozygote advantage. The B2 inversion is nearly three times longer than B12, making it more likely to have captured multiple recessive deleterious mutations (especially due to reduced effective population size) that are inherited as a single block (Connallon & Olito, 2022). For inversions that suffer from high mutation load, theory predicts that polymorphism could be maintained by pseudo-overdominance (Berdan et al., 2021), which is also supported by empirical studies in insects (Jay et al., 2021; Yang et al., 2002). For these reasons, it is likely that compensation of the recessive mutation load in heterokaryotypes could contribute to the persistence of the B2 polymorphism in Mediterranean lagoons.

The simultaneous action of forces associated with both Type I and Type II polymorphisms has been shown to potentially give rise to a range of different inversion equilibrium frequencies (Faria, Johannesson, et al., 2019). For instance, selection for heterokaryotypes combined with local adaptation can result in polymorphism in one environment and fixation in the other. A similar commixture of forces could be at work to explain the fixation of B2 haplotype C in marine *H. guttulatus* populations, while polymorphism is maintained in Mediterranean lagoons. If the B2 polymorphism is only maintained by pseudo-overdominance, we could ask why the D haplotype is not present outside of Mediterranean lagoons. The answer may be connected to local adaptation alleles that are carried in the haplotypes of B2, or in epistatic association with B12. If D provides local advantage in lagoons, or if D/A combinations are favoured over C/A, it could help maintain polymorphism in these environments. On the other hand, D would be outcompeted elsewhere (e.g., in Mediterranean marine habitats) due to lack of selective advantage. Although current knowledge does not allow us to answer all

questions regarding the maintenance of the B2 polymorphism, we suspect that it involves an interaction of forces which could include local selection and a form of balancing selection.

4.3 | Long-term fate of inversion polymorphisms in the long-snouted seahorse

Theoretical work (reviewed in, e.g., Hoffmann & Rieseberg, 2008) and empirical studies (e.g., Huang et al., 2020; Lohse et al., 2015; Noor et al., 2001) have found that, under certain conditions, the presence of inversion polymorphisms may facilitate speciation. In fish, multiple empirical studies have highlighted the role of chromosomal inversions in local adaptation, ecotype formation, and speciation (Berg et al., 2016; Cayuela et al., 2020; Jones et al., 2012; Le Moan et al., 2021; Matschiner et al., 2022; Pettersson et al., 2019; Tigano et al., 2021). Mediterranean ecotypes of long-snouted seahorse show differential fixation at inversion B12, which represents a Type I polymorphism that could facilitate speciation between marine and lagoon ecotypes. We found evidence for very low levels of gene flux taking place in this inversion, suggesting that gene exchange in heterokaryotypes is either rare or selected against. Since exchanged segments were generally longer than a few kilobases and occurred in the middle of the inversion, they most likely resulted from double crossovers rather than gene conversion. This contrasts with what was found for cod inversions where gene conversion predominated (Matschiner et al., 2022) but is in line with theoretical expectations for long inversions (Navarro et al., 1997). The introgressed tracts found within B12 across the species range were generally small in size (up to 50kb), suggesting that introgression took place a long time ago and that the remaining tracts represent recombined segments that have passed the filter of selection. This is consistent with the view that multiple selected mutations have accumulated within the inversion over the long term (e.g., DM incompatibilities; Navarro & Barton, 2003). A key question remains whether the barrier effect of B12 also reduces effective gene flow between lagoon and marine ecotypes at unlinked neutral loci, which may be facilitated by coupling with premating isolation mechanisms. Some level of genome-wide differentiation has been observed between Mediterranean ecotypes, thus indicating a significant reduction in effective gene flow at a small spatial scale relative to dispersal. However, the contribution of B12 to gene flow reduction is more uncertain in the Atlantic and the Black Sea. Although a weak association has been observed between B12 and habitat type along the Portuguese coasts, no associated genetic structure was detected in the genomic background (Riquet et al., 2019).

The B12 polymorphism could facilitate speciation between seahorse ecotypes over the long term, but the fate of the B2 polymorphism is less straightforward to predict. Since the frequencies of this inversion may result from a balance between local and balancing selection, it is unclear whether B2 will eventually undergo differential fixation between habitats, or universal fixation of one arrangement. In contrast to B12, we found ample evidence for the erosion

of divergence in B2 through gene flux, as illustrated by the detection of many introgressed segments and "suspension bridge" patterns in the F_{ST} landscape (Figure 2c). The intensity of gene flux was heterogeneous along this chromosome, since we found evidence for relatively recent double crossover events (i.e., larger introgressed segments than in B12) as well as one region where low F_{ST} values (comparable to the genome background) are bordered by high differentiation segments, indicating that divergence has been eroded over time. Gene flux is therefore likely to impact the fate of this inversion by purging some of its mutation load (Berdan et al., 2021) and weakening the pseudo-overdominance that could be maintaining polymorphism. Because the large-scale assembly of Chr2 is still only partly resolved, we cannot rule out the possibility that additional chromosomal rearrangements have affected its divergence landscape and evolutionary trajectory. For example, multiple inversions occurring in the same region might have extended or restructured the block of high LD, as is the case for adjacent (Jay et al., 2021) or nested inversions (Maggiolini et al., 2020).

The framework laid out by Faria, Johannesson, et al. (2019) does not specifically consider the dynamics of multiple inversions co-existing in the same system but leaves such interactions as an outstanding question. The presence of other polymorphic inversions could indeed impact the establishment, maintenance, and long-term fate of a given inversion, for example, through epistatic interactions or coupling. The two inversion polymorphisms segregating in *H. guttulatus* might present a case study for such potential interactions. Epistatic interactions may explain why there is a strong association between B12 and habitat type in the Mediterranean, while the absence of the B2 polymorphism in the Atlantic would only produce weak associations of B12 with habitat. Interestingly, the two molecular pathways that were enriched within the inversions have interconnected functions in reproduction. The oestrogen signalling pathway enriched in B12 has been shown to play a role in seahorse sexual dimorphism, gonad development, and stimulation of parturition in pregnant seahorses (Qin et al., 2019; Whittington et al., 2015; Zhang et al., 2022). This involves the regulation of expression of the two oestrogen receptor genes, *esr1* and *esr2*, which are both present in B12. In mice, estrogenic regulation by *esr1* induces expression changes in a population of neurons that mediate oestrogen feedback mechanisms affecting the neuroactive ligand-receptor interaction pathway (Göcz et al., 2022), corresponding to the pathway that was enriched in B2. Functional incompatibilities among genes that interact in the regulation of these pathways may potentially disrupt the reproductive function in incompatible genotypes at the two inversions involved in seahorse ecotype differentiation. Future directions for research could address these aspects by focusing on polymorphic populations, attempting to characterize the fitness of different karyotypes in different environments, and by quantifying gene expression and mutation load. Studying gene flux may also reveal more about the loci contained in each arrangement, as we should expect erosion of divergence in regions carrying selectively neutral or disadvantageous mutations, and maintenance in regions that are under divergent selection.

AUTHOR CONTRIBUTIONS

LM, BG, NB, and P-AG conceived the study. LM wrote the manuscript with inputs from all co-authors. Sampling was conducted by NB, FR, PB, RC, and P-AG. CA performed HMW DNA extraction and preparation for the reference genome 10X library. DNA extraction for WGS libraries was performed by LM, PB, and FR. Aspects of degraded/historical DNA extraction and analysis were handled by AF, CDS, BG, and LM. WGS library preparation was done by LM and FC. FC, AB, and ED were responsible for haplotagging library construction and sequencing. P-AG produced the reference genome and analysed linked-read sequencing data. PB provided scripts for genotyping high-coverage samples. LM performed all other bioinformatics and population genomic analyses. P-AG managed financial support.

ACKNOWLEDGEMENTS

The data were partly produced and analysed with the support of the GenSeq genotyping and sequencing platform and the MBB Montpellier Bioinformatics Biodiversity platform, both being supported by ANR program "Investissements d'avenir" (ANR-10-LABX-04-01). We thank Rémy Dérnat and Khalid Belkhir for their assistance in data storage, management, and processing, and Iago Bonnici for support on variant orientation. DNA extraction of historical samples was carried out in a clean lab at ISEM, Montpellier (Plateforme d'ADN dégradé, LabEx CeMEB). We thank the Montpellier GenomiX platform for constructing the reference genome's 10x Chromium library. We are grateful to the National Museum of Natural History (MNHN Paris, Agnès Dettai) and colleagues who provided us with samples, as well as those who facilitated or participated in sampling: Jorge Palma and Rita Castilho (CCMAR, Portugal), Cristina Mena (Hippocampus association, Spain), Patrick Louisy (peau-bleue association, CPIE Bassin de Thau, France), Lucy Woodall (University of Oxford, UK), and citizens from the Sète region who provided dried seahorses samples. We thank Huixian Zhang of the Key Laboratory of Tropical Marine Bio-Resources and Ecology, South China Sea Institute of Oceanology, for providing the *Hippocampus erectus* reference genome. Finally, we thank Thomas Broquet for his inputs on the manuscript. This work was supported by the ANR grant CoGeDiv ANR-17-CE02-0006-01 to PAG and by a Languedoc-Roussillon Region "Chercheur(se)s d'avenir" grant to NB (Connect7 project), with the support of the Occitanie Regional Council's program «Key challenge BiodivOc». Mobility during the project was partly funded by an ESEB Godfrey Hewitt Mobility Award as well as a Laura Corrigan conservation grant.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Sequence reads for the reference genome assembly and whole-genome resequencing data have been deposited in the GenBank Sequence Read Archive under the accession code BioProject ID

PRJNA777424. File S2 contains scripts and commands used for all bioinformatic analyses. Full detailed metadata and results may be consulted in the supplied Rmarkdown reports in the Appendix S1.

BENEFIT-SHARING STATEMENT

Regarding the benefits generated, we reanalysed genetic samples from countries (other than France) that were published in a previous study (Riquet et al., 2019) where collaborators who provided samples were already included as co-authors. The results of this research have been shared with the sample providers and the broader scientific community, in particular, through the sharing of genomic sequences on NCBI public databases. Other genetic resources will be made available on request.

ORCID

Laura Meyer  <https://orcid.org/0009-0005-6172-1113>

Florentine Riquet  <https://orcid.org/0000-0001-7122-6275>

Andrew Foote  <https://orcid.org/0000-0001-7384-1634>

Bruno Guinand  <https://orcid.org/0000-0002-6934-1677>

REFERENCES

- Barry, P., Broquet, T., & Gagnaire, P.-A. (2022). Age-specific survivorship and fecundity shape genetic diversity in marine fishes. *Evolution Letters*, 6(1), 46–62.
- Benson, G. (1999). Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Research*, 27(2), 573–580. <https://doi.org/10.1093/nar/27.2.573>
- Berdan, E. L., Blanckaert, A., Butlin, R. K., & Bank, C. (2021). Deleterious mutation accumulation and the long-term fate of chromosomal inversions. *PLoS Genetics*, 17(3), e1009411. <https://doi.org/10.1371/journal.pgen.1009411>
- Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., Jakobsen, K. S., & Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports*, 6(1), 23246. <https://doi.org/10.1038/srep23246>
- Y C Brandt, D., Wei, X., Deng, Y., Vaughn, A. H., & Nielsen, R. (2022). Evaluation of methods for estimating coalescence times using ancestral recombination graphs. *Genetics*, 221(1), iyac044. <https://doi.org/10.1093/genetics/iyac044>
- Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M., & Borodovsky, M. (2021). BRAKER2: Automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genomics and Bioinformatics*, 3(1), lqaa108. <https://doi.org/10.1093/nargab/lqaa108>
- Butlin, R. K. (2005). Recombination and speciation. *Molecular Ecology*, 14(9), 2621–2635. <https://doi.org/10.1111/j.1365-294X.2005.02617.x>
- Cabanettes, F., & Klopp, C. (2018). D-GENIES: Dot plot large genomes in an interactive, efficient and simple way. *PeerJ*, 6, e4958. <https://doi.org/10.7717/peerj.4958>
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10(1), 421. <https://doi.org/10.1186/1471-2105-10-421>
- Campbell, M. A., Anderson, E. C., Garza, J. C., & Pearse, D. E. (2021). Polygenic basis and the role of genome duplication in adaptation to similar selective environments. *Journal of Heredity*, 112(7), 614–625. <https://doi.org/10.1093/jhered/esab049>

- Campos, P. F., & Gilbert, T. M. P. (2012). DNA extraction from formalin-fixed material. *Methods in Molecular Biology*, 840, 81–85. https://doi.org/10.1007/978-1-61779-516-9_11
- Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P., & Huerta-Cepas, J. (2021). eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Molecular Biology and Evolution*, 38(12), 5825–5829. <https://doi.org/10.1093/molbev/msab293>
- Cayuela, H., Rougemont, Q., Laporte, M., Mérot, C., Normandeau, E., Dorant, Y., Tørresen, O. K., Hoff, S. N. K., Jentoft, S., Sirois, P., Castonguay, M., Jansen, T., Praebel, K., Clément, M., & Bernatchez, L. (2020). Shared ancestral polymorphisms and chromosomal rearrangements as potential drivers of local adaptation in a marine fish. *Molecular Ecology*, 29(13), 2379–2398. <https://doi.org/10.1111/mec.15499>
- Challis, R., Richards, E., Rajan, J., Cochrane, G., & Blaxter, M. (2020). BlobToolKit – Interactive quality assessment of genome assemblies. *G3: Genes, Genomes, Genetics*, 10(4), 1361–1374. <https://doi.org/10.1534/g3.119.400908>
- Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3), 195–205. <https://doi.org/10.1038/nrg2526>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Cheng, C., White, B. J., Kamdem, C., Mockaitis, K., Costantini, C., Hahn, M. W., & Besansky, N. J. (2012). Ecological genomics of *Anopheles gambiae* along a latitudinal cline: A population-resequencing approach. *Genetics*, 190(4), 1417–1432. <https://doi.org/10.1534/genetics.111.137794>
- Connallon, T., & Olito, C. (2022). Natural selection and the distribution of chromosomal inversion lengths. *Molecular Ecology*, 31(13), 3627–3641. <https://doi.org/10.1111/mec.16091>
- Dai, H., & Guan, Y. (2020). The Nubeam reference-free approach to analyze metagenomic sequencing reads. *Genome Research*, 30(9), 1364–1375. <https://doi.org/10.1101/gr.261750.120>
- 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(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- 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(2), giab008. <https://doi.org/10.1093/gigascience/giab008>
- Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L., & Dermitzakis, E. T. (2019). Accurate, scalable and integrative haplotype estimation. *Nature Communications*, 10(1), 5436. <https://doi.org/10.1038/s41467-019-13225-y>
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 11–15.
- Faria, R., Chaube, P., Morales, H. E., Larsson, T., Lemmon, A. R., Lemmon, E. M., Rafajlović, M., Panova, M., Ravinet, M., Johannesson, K., Westram, A. M., & Butlin, R. K. (2019). Multiple chromosomal rearrangements in a hybrid zone between *Littorina saxatilis* ecotypes. *Molecular Ecology*, 28(6), 1375–1393. <https://doi.org/10.1111/mec.14972>
- Faria, R., Johannesson, K., Butlin, R. K., & Westram, A. M. (2019). Evolving inversions. *Trends in Ecology & Evolution*, 34(3), 239–248. <https://doi.org/10.1016/j.tree.2018.12.005>
- Faria, R., & Navarro, A. (2010). Chromosomal speciation revisited: Rearranging theory with pieces of evidence. *Trends in Ecology & Evolution*, 25(11), 660–669. <https://doi.org/10.1016/j.tree.2010.07.008>
- Ferrari, G., Atmore, L. M., Jentoft, S., Jakobsen, K. S., Makowiecki, D., Barrett, J. H., & Star, B. (2022). An accurate assignment test for extremely low-coverage whole-genome sequence data. *Molecular Ecology Resources*, 22(4), 1330–1344. <https://doi.org/10.1111/1755-0998.13551>
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences of the United States of America*, 117(17), 9451–9457. <https://doi.org/10.1073/pnas.1921046117>
- Funk, E. R., Mason, N. A., Pálsson, S., Albrecht, T., Johnson, J. A., & Taylor, S. A. (2021). A supergene underlies linked variation in color and morphology in a Holarctic songbird. *Nature Communications*, 12(1), Article 1. <https://doi.org/10.1038/s41467-021-27173-z>
- Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, 36(8), 2628–2629. <https://doi.org/10.1093/bioinformatics/btz931>
- Göcz, B., Takács, S., Skrapits, K., Rumpel, É., Solymosi, N., Pólska, S., Colledge, W. H., Hrabovszky, E., & Sárvári, M. (2022). Estrogen differentially regulates transcriptional landscapes of preoptic and arcuate kisspeptin neuron populations. *Frontiers in Endocrinology*, 13, 960769.
- Gould, B. A., Chen, Y., & Lowry, D. B. (2017). Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive differentiation and reproductive isolation. *Molecular Ecology*, 26(1), 163–177. <https://doi.org/10.1111/mec.13881>
- Guerrero, R. F., Rousset, F., & Kirkpatrick, M. (2012). Coalescent patterns for chromosomal inversions in divergent populations. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 367(1587), 430–438. <https://doi.org/10.1098/rstb.2011.0246>
- Guichard, A., Legeai, F., Tagu, D., & Lemaitre, C. (2023). MTG-link: Leveraging barcode information from linked-reads to assemble specific loci. *BMC Bioinformatics*, 24(284), 284. <https://doi.org/10.1186/s12859-023-05395-w>
- Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hager, E. R., Harringmeyer, O. S., Wooldridge, T. B., Theingi, S., Gable, J. T., McFadden, S., Neugeboren, B., Turner, K. M., Jensen, J. D., & Hoekstra, H. E. (2022). A chromosomal inversion contributes to divergence in multiple traits between deer mouse ecotypes. *Science*, 377(6604), 399–405. <https://doi.org/10.1126/science.abg0718>
- Harringmeyer, O. S., & Hoekstra, H. E. (2022). Chromosomal inversion polymorphisms shape the genomic landscape of deer mice. *Nature Ecology & Evolution*, 6, 1965–1979. <https://doi.org/10.1038/s41559-022-01890-0>
- Haï, V., Mennesson, M. I., Dettai, A., Bonillo, C., Keith, P., & Lord, C. (2020). Needlepoint non-destructive internal tissue sampling for precious fish specimens. *Cybiu: Revue Internationale d'Ichtyologie*, 44(1), 73–79. <https://doi.org/10.26028/cybiu/2020-441-010>
- Hill, W. G., & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetics Research*, 8(3), 269–294. <https://doi.org/10.1017/S0016672300010156>
- Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, 39, 21–42. <https://doi.org/10.1146/annurev.ecolsys.39.110707.173532>
- Hsieh, P., Vollger, M. R., Dang, V., Porubsky, D., Baker, C., Cantsilieris, S., Hoekzema, K., Lewis, A. P., Munson, K. M., Sorensen, M., Kronenberg, Z. N., Murali, S., Nelson, B. J., Chiatante, G., Maggolini, F. A. M., Blanché, H., Underwood, J. G., Antonacci, F., Deleuze, J.-F., & Eichler, E. E. (2019). Adaptive archaic introgression of copy

- number variants and the discovery of previously unknown human genes. *Science*, 366(6463), eaax2083. <https://doi.org/10.1126/science.aax2083>
- Huang, K., Andrew, R. L., Owens, G. L., Ostevik, K. L., & Rieseberg, L. H. (2020). Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Molecular Ecology*, 29(14), 2535–2549. <https://doi.org/10.1111/mec.15428>
- Huang, K., Ostevik, K. L., Elphinstone, C., Todesco, M., Bercovich, N., Owens, G. L., & Rieseberg, L. H. (2022). Mutation load in sunflower inversions is negatively correlated with inversion heterozygosity. *Molecular Biology and Evolution*, 39(5), msac101. <https://doi.org/10.1093/molbev/msac101>
- Ignatieva, A., Favero, M., Koskela, J., Sant, J., & Myers, S. R. (2023). The distribution of branch duration and detection of inversions in ancestral recombination graphs. *bioRxiv*. <https://doi.org/10.1101/2023.07.11.548567>
- Jay, P., Chouteau, M., Whibley, A., Bastide, H., Parrinello, H., Llaurens, V., & Joron, M. (2021). Mutation load at a mimicry supergene sheds new light on the evolution of inversion polymorphisms. *Nature Genetics*, 53(3), 288–293. <https://doi.org/10.1038/s41588-020-00771-1>
- Jay, P., Whibley, A., Frézal, L., Rodríguez de Cara, M. Á., Nowell, R. W., Mallet, J., Dasmahapatra, K. K., & Joron, M. (2018). Supergene evolution triggered by the introgression of a chromosomal inversion. *Current Biology*, 28(11), 1839–1845. <https://doi.org/10.1016/j.cub.2018.04.072>
- 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., ... Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, 484(7392), 55–61. <https://doi.org/10.1038/nature10944>
- Keightley, P. D., & Jackson, B. C. (2018). Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics*, 209(3), 897–906. <https://doi.org/10.1534/genetics.118.301120>
- Kelleher, J., Wong, Y., Wohns, A. W., Fadil, C., Albers, P. K., & McVean, G. (2019). Inferring whole-genome histories in large population datasets. *Nature Genetics*, 51(9), 1330–1338. <https://doi.org/10.1038/s41588-019-0483-y>
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*, 37(8), 907–915. <https://doi.org/10.1038/s41587-019-0201-4>
- Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, 8(9), e1000501. <https://doi.org/10.1371/journal.pbio.1000501>
- Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, 173(1), 419–434. <https://doi.org/10.1534/genetics.105.047985>
- Kirov, I., Gilyok, M., Knyazev, A., & Fesenko, I. (2018). Pilot satellitome analysis of the model plant, *Physcomitrella patens*, revealed a transcribed and high-copy IGS related tandem repeat. *Comparative Cytogenetics*, 12(4), 493–513. <https://doi.org/10.3897/CompCyto-gen.v12i4.31015>
- Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1), 356. <https://doi.org/10.1186/s12859-014-0356-4>
- Korunes, K. L., & Noor, M. A. F. (2019). Pervasive gene conversion in chromosomal inversion heterozygotes. *Molecular Ecology*, 28(6), 1302–1315. <https://doi.org/10.1111/mec.14921>
- Kulmuni, J., Butlin, R. K., Lucek, K., Savolainen, V., & Westram, A. M. (2020). Towards the completion of speciation: The evolution of reproductive isolation beyond the first barriers. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 375(1806), 20190528. <https://doi.org/10.1098/rstb.2019.0528>
- Lamichhane, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D. S., Hoepfner, M. P., Kerje, S., Gustafson, U., Shi, C., Zhang, H., Chen, W., Liang, X., Huang, L., Wang, J., Liang, E., Wu, Q., Lee, S. M.-Y., Xu, X., Höglund, J., ... Andersson, L. (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nature Genetics*, 48(1), Article 1. <https://doi.org/10.1038/ng.3430>
- Le Moan, A., Bekkevold, D., & Hemmer-Hansen, J. (2021). Evolution at two time frames: Ancient structural variants involved in post-glacial divergence of the European plaice (*Pleuronectes platessa*). *Heredity*, 126(4), 668–683. <https://doi.org/10.1038/s41437-020-00389-3>
- Li, C., Olave, M., Hou, Y., Qin, G., Schneider, R. F., Gao, Z., Tu, X., Wang, X., Qi, F., Nater, A., Kautt, A. F., Wan, S., Zhang, Y., Liu, Y., Zhang, H., Zhang, B., Zhang, H., Qu, M., Liu, S., ... Lin, Q. (2021). Genome sequences reveal global dispersal routes and suggest convergent genetic adaptations in seahorse evolution. *Nature Communications*, 12(1), 1094. <https://doi.org/10.1038/s41467-021-21379-x>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (arXiv:1303.3997). *arXiv*. <http://arxiv.org/abs/1303.3997>
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, 34(18), 3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li, H., & Ralph, P. (2019). Local PCA shows how the effect of population structure differs along the genome. *Genetics*, 211(1), 289–304. <https://doi.org/10.1534/genetics.118.301747>
- Llaurens, V., Whibley, A., & Joron, M. (2017). Genetic architecture and balancing selection: The life and death of differentiated variants. *Molecular Ecology*, 26(9), 2430–2448. <https://doi.org/10.1111/mec.14051>
- Lohse, K., Clarke, M., Ritchie, M. G., & Etges, W. J. (2015). Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. *Evolution*, 69(5), 1178–1190. <https://doi.org/10.1111/evo.12650>
- Lomsadze, A., Burns, P. D., & Borodovsky, M. (2014). Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Research*, 42(15), e119. <https://doi.org/10.1093/nar/gku557>
- Lowry, D. B., & Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology*, 8(9), e1000500. <https://doi.org/10.1371/journal.pbio.1000500>
- Lundberg, M., Mackintosh, A., Petri, A., & Bensch, S. (2023). Inversions maintain differences between migratory phenotypes of a songbird. *Nature Communications*, 14(1), 452. <https://doi.org/10.1038/s41467-023-36167-y>
- Maggiolini, F. A. M., Sanders, A. D., Shew, C. J., Sulovari, A., Mao, Y., Puig, M., Catacchio, C. R., Dellino, M., Palmisano, D., Mercuri, L., Bitonto, M., Porubský, D., Cáceres, M., Eichler, E. E., Ventura, M., Dennis, M. Y., Korbel, J. O., & Antonacci, F. (2020). Single-cell strand sequencing of a macaque genome reveals multiple nested inversions and breakpoint reuse during primate evolution. *Genome Research*, 30(11), 1680–1693. <https://doi.org/10.1101/gr.265322.120>
- Manni, M., Berkeley, M. R., Seppely, M., & Zdobnov, E. M. (2021). BUSCO: Assessing genomic data quality and beyond. *Current Protocols*, 1(12), e323. <https://doi.org/10.1002/cpz1.323>
- Marion, S. B., & Noor, M. A. F. (2023). Interrogating the roles of mutation–selection balance, heterozygote advantage, and linked selection in maintaining recessive lethal variation in natural populations. *Annual Review of Animal Biosciences*, 11(1), 77–91. <https://doi.org/10.1146/annurev-animal-050422-092520>
- Martin, S. H., & Van Belleghem, S. M. (2017). Exploring evolutionary relationships across the genome using topology weighting. *Genetics*, 206(1), 429–438. <https://doi.org/10.1534/genetics.116.194720>
- Matschiner, M., Barth, J. M. I., Tørresen, O. K., Star, B., Baalsrud, H. T., Briec, M. S. O., Pampoulie, C., Bradbury, I., Jakobsen, K. S., &

- Jentoft, S. (2022). Supergene origin and maintenance in Atlantic cod. *Nature Ecology & Evolution*, 6(4), Article 4. <https://doi.org/10.1038/s41559-022-01661-x>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Meier, J. I., Salazar, P. A., Kučka, M., Davies, R. W., Dréau, A., Aldás, I., Box Power, O., Nadeau, N. J., Bridle, J. R., Rolian, C., Barton, N. H., McMillan, W. O., Jiggins, C. D., & Chan, Y. F. (2021). Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proceedings of the National Academy of Sciences of the United States of America*, 118(25), e2015005118. <https://doi.org/10.1073/pnas.2015005118>
- Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J. G., Sebra, R., Peluso, P., Eid, J., Rank, D., Garcia, J. F., DeRisi, J. L., Smith, T., Tobias, C., Ross-Ibarra, J., Korf, I., & Chan, S. W. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology*, 14(1), R10. <https://doi.org/10.1186/gb-2013-14-1-r10>
- Mérot, C., Llaurens, V., Normandeau, E., Bernatchez, L., & Wellenreuther, M. (2020). Balancing selection via life-history trade-offs maintains an inversion polymorphism in a seaweed fly. *Nature Communications*, 11(1), 670. <https://doi.org/10.1038/s41467-020-14479-7>
- Mérot, C., Oomen, R. A., Tigano, A., & Wellenreuther, M. (2020). A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends in Ecology & Evolution*, 35(7), 561–572. <https://doi.org/10.1016/j.tree.2020.03.002>
- Morisse, P., Legeai, F., & Lemaitre, C. (2021). LEVIATHAN: Efficient discovery of large structural variants by leveraging long-range information from linked-reads data. *bioRxiv*. <https://doi.org/10.1101/2021.03.25.437002>
- Morisse, P., Lemaitre, C., & Legeai, F. (2021). LRez: A C++ API and toolkit for analyzing and managing linked-reads data. *Bioinformatics Advances*, 1(1), vbab022. <https://doi.org/10.1093/bioadv/vbab022>
- Navarro, A., & Barton, N. H. (2003). Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution*, 57(3), 447–459. <https://doi.org/10.1111/j.0014-3820.2003.tb01537.x>
- Navarro, A., Betrán, E., Barbadilla, A., & Ruiz, A. (1997). Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics*, 146(2), 695–709. <https://doi.org/10.1093/genetics/146.2.695>
- Nei, M., Kojima, K.-I., & Schaffer, H. E. (1967). Frequency changes of new inversions in populations under mutation-selection equilibria. *Genetics*, 57(4), 741–750. <https://doi.org/10.1093/genetics/57.4.741>
- Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences of the United States of America*, 98(21), 12084–12088. <https://doi.org/10.1073/pnas.221274498>
- Ortiz-Barrientos, D., Engelstädter, J., & Rieseberg, L. H. (2016). Recombination rate evolution and the origin of species. *Trends in Ecology & Evolution*, 31(3), 226–236. <https://doi.org/10.1016/j.tree.2015.12.016>
- Pedersen, B. S., & Quinlan, A. R. (2017). cyvcf2: Fast, flexible variant analysis with Python. *Bioinformatics*, 33(12), 1867–1869. <https://doi.org/10.1093/bioinformatics/btx057>
- Perez-Ruzafa, A., Marcos, C., Pérez-Ruzafa, I., & Pérez-Marcos, M. (2011). Coastal lagoons: “Transitional ecosystems” between transitional and coastal waters. *Journal of Coastal Conservation*, 15, 369–392. <https://doi.org/10.1007/s11852-010-0095-2>
- Pettersson, M. E., Rochus, C. M., Han, F., Chen, J., Hill, J., Wallerman, O., Fan, G., Hong, X., Xu, Q., Zhang, H., Liu, S., Liu, X., Haggerty, L., Hunt, T., Martin, F. J., Flicek, P., Bunikis, I., Folkvord, A., & Andersson, L. (2019). A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. *Genome Research*, 29(11), 1919–1928. <https://doi.org/10.1101/gr.253435.119>
- Picard toolkit. (2019). *GitHub repository*. Broad Institute. <https://broadinstitute.github.io/picard/>
- Qin, G., Luo, W., Tan, S., Zhang, B., Ma, S., & Lin, Q. (2019). Dimorphism of sex and gonad-development-related genes in male and female lined seahorse, *Hippocampus erectus*, based on transcriptome analyses. *Genomics*, 111(3), 260–266. <https://doi.org/10.1016/j.ygeno.2018.11.008>
- Rafajlović, M., Rambla, J., Feder, J. L., Navarro, A., & Faria, R. (2021). Inversions and genomic differentiation after secondary contact: When drift contributes to maintenance, not loss, of differentiation. *Evolution*, 75(6), 1288–1303. <https://doi.org/10.1111/evo.14223>
- Revell, L. J. (2012). Phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2), 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., ... Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, 592(7856), 737–746. <https://doi.org/10.1038/s41586-021-03451-0>
- Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16(7), 351–358. [https://doi.org/10.1016/s0169-5347\(01\)02187-5](https://doi.org/10.1016/s0169-5347(01)02187-5)
- Riquet, F., Liautard-Haag, C., Woodall, L., Bouza, C., Louisy, P., Hamer, B., Otero-Ferrer, F., Aublanc, P., Béduneau, V., Briard, O., Ayari, T. E., Hochscheid, S., Belkhir, K., Arnaud-Haond, S., Gagnaire, P.-A., & Bierne, N. (2019). Parallel pattern of differentiation at a genomic Island shared between clinal and mosaic hybrid zones in a complex of cryptic seahorse lineages. *Evolution*, 73(4), 817–835. <https://doi.org/10.1111/evo.13696>
- Schaeffer, S. W., & Anderson, W. W. (2005). Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. *Genetics*, 171(4), 1729–1739. <https://doi.org/10.1534/genetics.105.041947>
- Schwander, T., Libbrecht, R., & Keller, L. (2014). Supergenes and complex phenotypes. *Current Biology*, 24(7), R288–R294. <https://doi.org/10.1016/j.cub.2014.01.056>
- Shajii, A., Numanagić, I., & Berger, B. (2018). Latent variable model for aligning barcoded short-reads improves downstream analyses. *Research in Computational Molecular Biology: Annual International Conference, Proceedings*, 10812, 280–282.
- Sievert, C., Iannone, R., Allaire, J. J., & Borges, B. (2022). flexdashboard: R Markdown Format for Flexible Dashboards. <https://pkgs.rstudio.com/flexdashboard/>, <https://github.com/rstudio/flexdashboard/>
- Skoglund, P., Northoff, B. H., Shunkov, M. V., Derevianko, A. P., Pääbo, S., Krause, J., & Jakobsson, M. (2014). Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), 2229–2234. <https://doi.org/10.1073/pnas.1318934111>
- Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, 24(5), 637–644. <https://doi.org/10.1093/bioinformatics/btn013>
- Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Philosophical*

- Transactions of the Royal Society of London. Series B, Biological Sciences, 372(1736), 20160455. <https://doi.org/10.1098/rstb.2016.0455>
- Stiller, J., Short, G., Hamilton, H., Saarman, N., Longo, S., Wainwright, P., Rouse, G. W., & Simison, W. B. (2022). Phylogenomic analysis of Syngnathidae reveals novel relationships, origins of endemic diversity and variable diversification rates. *BMC Biology*, 20(1), 75. <https://doi.org/10.1186/s12915-022-01271-w>
- Thompson, M. J., & Jiggins, C. D. (2014). Supergenes and their role in evolution. *Heredity*, 113, 1–8. <https://doi.org/10.1038/hdy.2014.20>
- Tigano, A., Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, 25(10), 2144–2164. <https://doi.org/10.1111/mec.13606>
- Tigano, A., Jacobs, A., Wilder, A. P., Nand, A., Zhan, Y., Dekker, J., & Therikildsen, N. O. (2021). Chromosome-level assembly of the Atlantic silverside genome reveals extreme levels of sequence diversity and structural genetic variation. *Genome Biology and Evolution*, 13(6), evab098. <https://doi.org/10.1093/gbe/evab098>
- Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., ... Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, 584(7822), 602–607. <https://doi.org/10.1038/s41586-020-2467-6>
- Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altschuler, D., Gabriel, S., & DePristo, M. A. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43(1), 11.10.1–11.10.33. <https://doi.org/10.1002/0471250953.bi1110s43>
- Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research*, 27(5), 757–767. <https://doi.org/10.1101/gr.214874.116>
- Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, 33(6), 427–440. <https://doi.org/10.1016/j.tree.2018.04.002>
- Wellenreuther, M., Mérot, C., Berdan, E., & Bernatchez, L. (2019). Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. *Molecular Ecology*, 28(6), 1203–1209. <https://doi.org/10.1111/mec.15066>
- Westram, A. M., Faria, R., Johannesson, K., Butlin, R., & Barton, N. (2022). Inversions and parallel evolution. *Philosophical Transactions of the Royal Society, B: Biological Sciences*, 377(1856), 20210203. <https://doi.org/10.1098/rstb.2021.0203>
- Westram, A. M., Stankowski, S., Surendranadh, P., & Barton, N. (2022). What is reproductive isolation? *Journal of Evolutionary Biology*, 35(9), 1143–1164. <https://doi.org/10.1111/jeb.14005>
- Whittington, C. M., Griffith, O. W., Qi, W., Thompson, M. B., & Wilson, A. B. (2015). Seahorse brood pouch transcriptome reveals common genes associated with vertebrate pregnancy. *Molecular Biology and Evolution*, 32(12), 3114–3131. <https://doi.org/10.1093/molbev/msv177>
- Wohns, A., Wong, Y., Jeffery, B., Akbari, A., Mallick, S., Pinhasi, R., Patterson, N., Reich, D., Kelleher, J., & McVean, G. (2022). A unified genealogy of modern and ancient genomes. *Science*, 375, eabi8264. <https://doi.org/10.1126/science.abi8264>
- Yang, Y.-Y., Lin, F.-J., & Chang, H. (2002). Comparison of recessive lethal accumulation in inversion-bearing and inversion-free chromosomes in *drosophila*. *Zoological Studies*, 41(3), 271–282.
- Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19), E1743–E1751. <https://doi.org/10.1073/pnas.1219381110>
- Yeaman, S., & Whitlock, M. C. (2011). The genetic architecture of adaptation under migration-selection balance. *Evolution*, 65(7), 1897–1911. <https://doi.org/10.1111/j.1558-5646.2011.01269.x>
- Zhang, H., Liu, Y., Qin, G., & Lin, Q. (2022). Identification of neurohypophyseal hormones and the role of VT in the parturition of pregnant seahorses (*Hippocampus erectus*). *Frontiers in Endocrinology*, 13, 923234. <https://doi.org/10.3389/fendo.2022.923234>
- Zhang, L., Reifová, R., Halenková, Z., & Gompert, Z. (2021). How important are structural variants for speciation? *Genes*, 12(7), 1084. <https://doi.org/10.3390/genes12071084>
- Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, 28(24), 3326–3328. <https://doi.org/10.1093/bioinformatics/bts606>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Meyer, L., Barry, P., Riquet, F., Foote, A., Der Sarkissian, C., Cunha, R. L., Arbiol, C., Cerqueira, F., Desmarais, E., Bordes, A., Bierne, N., Guinand, B., & Gagnaire, P.-A. (2024). Divergence and gene flow history at two large chromosomal inversions underlying ecotype differentiation in the long-snouted seahorse. *Molecular Ecology*, 00, e17277. <https://doi.org/10.1111/mec.17277>