

RESOURCE ARTICLE

Management and conservation implications of cryptic population substructure for two commercially exploited fishes (*Merluccius* spp.) in southern Africa

Sarah Forde¹ | Sophie von der Heyden²  | Alan Le Moan³  | Erica S. Nielsen⁴  |
Deon Durholtz⁵  | Paulus Kainge⁶  | Johannes N. Kathena⁶  | Marek R. Lipinski⁷ |
Hilkka O. N. Ndjaula⁸  | Conrad A. Matthee²  | Romina Henriques¹ 

¹Marine Genomics Group, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Hatfield, South Africa

²Evolutionary Genomics Group, Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa

³CNRS-Sorbonne Université, Station Biologique de Roscoff, Place Georges Tessier, Roscoff, France

⁴University of California – Davis, Davis, California, USA

⁵Department of Forestry, Fisheries and Environment, Cape Town, South Africa

⁶National Marine Information and Research Centre, Ministry of Fisheries and Marine Resources, Swakopmund, Namibia

⁷Department of Ichthyology and Fisheries Science, Rhodes University, Makhanda, South Africa

⁸Sam Nujoma Marine and Coastal Resources Research Centre, University of Namibia, Henties Bay, Namibia

Correspondence

Romina Henriques, Marine Genomics Group, Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Private Bag X20, Hatfield 0028, South Africa.

Email: romina.henriques@up.ac.za

Funding information

Directorate-General for Development and Cooperation - EuropeAid, Grant/Award Number: 2010/222387; National Research Foundation, Grant/Award Number: 105949

Handling Editor: Alison Gonçalves Nazareno

Abstract

Genomic information can aid in the establishment of sustainable management plans for commercially exploited marine fishes, aiding in the long-term conservation of these resources. The southern African hakes (*Merluccius capensis* and *M. paradoxus*) are commercially valuable demersal fishes with similar distribution ranges but exhibiting contrasting life histories. Using a comparative framework based on Pool-Seq genome-wide SNP data, we investigated whether the evolutionary processes that shaped extant patterns of diversity and divergence are shared among these two congeneric fishes, or unique to each one. Our findings revealed that *M. capensis* and *M. paradoxus* show similar levels of genome-wide diversity, despite different census sizes and life-history features. In addition, *M. capensis* shows three highly structured geographic populations across the Benguela Current region (one in the northern Benguela and two in the southern Benguela), with no consistent genome–environment associations detected. In contrast, although population structure and outlier analyses suggested panmixia for *M. paradoxus*, reconstruction of its demographic history suggested the presence of an Atlantic–Indian Ocean subtle substructuring pattern. Therefore, it appears that *M. paradoxus* might be composed by two highly connected populations, one in the Atlantic and one in the southwest Indian Ocean. The reported similar low levels of genomic diversity, as well as newly discovered genetically distinct populations in

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

both hake species can thus assist in informing and improving conservation and management plans for the commercially important southern African *Merluccius*.

KEYWORDS

conservation genomics, fisheries management, genomic diversity, Pool-Seq, population dynamics

1 | INTRODUCTION

The conservation of marine biodiversity requires the sustainable management of fishery resources, as the majority of the world's stocks are currently over- or fully exploited (FAO, 2022). This, in turn, strongly relies on accurate knowledge of biological and demographic features, including population substructuring patterns, of targeted species. Molecular data can assist conservation and management actions, by providing vital information on evolutionary potential, genetic connectivity patterns and effective population sizes (Nielsen et al., 2023). In the last decade, novel sequencing approaches, such as genotype-by-sequencing (GBS), have been successfully used to inform fisheries management (Beacham et al., 2018; Clucas, Kerr, et al., 2019; Clucas, Lou, et al., 2019; Dahle et al., 2018; Hemmer-Hansen et al., 2019), as they can reveal subtle levels of population substructuring in otherwise high-gene flow species (reviewed in Gagnaire, 2020), while also allowing to detect local adaptation and selective pressures (Clucas, Kerr, et al., 2019; Clucas, Lou, et al., 2019; Diopere et al., 2018; Hoey & Pinsky, 2018; Lamichhaney et al., 2012; Nielsen et al., 2018; Riquet et al., 2019). For example, the use of genome-wide single-nucleotide polymorphisms (SNPs) has uncovered previously unrecognized patterns of population substructuring in commercially important species such as the American lobster (*Homarus americanus*; Benestan et al., 2015), Atlantic cod (*Gadus morhua*; Dahle et al., 2018), lesser sandeel (*Ammodytes marinus*; Jimenez-Mena et al., 2020), kingklip (*Genypterus capensis*; Schulze et al., 2020) or the European hake (*Merluccius merluccius*; Milano et al., 2014). Marine species are generally characterized by large effective population sizes and high levels of gene flow, which tend to minimize the effects of genetic drift and, subsequently, levels of population differentiation (Waples & Gaggiotti, 2006). Natural selection, via local adaptation, can thus counteract the effects of gene flow and lead to cryptic population substructuring patterns in marine species (Gagnaire et al., 2015). The management of fishery resources as one single population (or stock) when it is composed by multiple populations, with independent demographic and evolutionary histories, may lead to the depletion of the most vulnerable populations (Ovenden et al., 2015), and the loss of marine biodiversity. Therefore, the distinction between neutral markers (affected mainly by genetic drift and gene flow) and putatively adaptive markers (affected by natural selection) can provide valuable insights for the establishment of conservation and fishing practices that take into account the evolutionary dynamics of populations, an understanding crucial for the continued sustainable management of these

resources (Gagnaire et al., 2015). In fact, several empirical studies for marine species show that including neutral and outlier information can greatly enhance conservation and management outcomes (Nielsen et al., 2023; Schulze et al., 2020; Therkildsen et al., 2013).

The southern African hakes (*Merluccius capensis* and *M. paradoxus*) are the most commercially important fishery species in the Benguela Current region, in the south-eastern Atlantic (DFFE, 2020), with total capture weight reported as 265,172 tonnes for both species for 2020, a much higher volume than total capture weight for the European hake (*M. merluccius*) throughout Europe (98,238 tonnes) for the same year (https://www.fao.org/fishery/statistics-query/en/capture/capture_quantity). Both species are distributed throughout southern Africa, from northern Namibia to eastern South Africa (Jansen et al., 2015), and show similar external morphology; however, they exhibit key differences in their depth distribution, population structuring and migration patterns (Durholtz et al., 2015; Kathena et al., 2016; Strømme et al., 2016). *Merluccius capensis* inhabits waters ranging from 30 to 500m depth, whereas *M. paradoxus* occurs preferentially in waters of 110–1000m depth (Durholtz et al., 2015), and despite this overlap, as adults they occupy different niches, with adults *M. capensis* overlapping in depth with juveniles *M. paradoxus* (Durholtz et al., 2015; Kathena et al., 2016; Strømme et al., 2016). The southern African hakes appear to have diverged between 3 and 4.2 million years ago, and likely expanded into the region in two separate events (Campo et al., 2007).

Historically, high catch levels combined with data-deficient fishery management policies contributed towards severe population declines of both species during the 20th century (DFFE, 2020). The first genetic data to support management decisions were drawn from allozymes and mitochondrial DNA (mtDNA), confirming the presence of the two species in the region, as well as the possibility of species-specific population substructuring (Grant et al., 1988; von der Heyden et al., 2007). More recent work using microsatellite data supported the presence of two populations (northern and southern Benguela) of *M. capensis* (Henriques et al., 2016; Kapula et al., 2022) and revealed a panmictic population of *M. paradoxus* (Henriques et al., 2016) across the region. At present, management practices are country specific (Namibia vs. South Africa) and species-specific within South Africa, with both species showing signs of recovery and an increase in spawning biomass in South Africa (DFFE, 2020).

Based on a neutral data set, Henriques et al. (2016) suggested that the reported differences in population substructuring and genetic diversity levels in the southern African hakes were likely due to species-specific responses to contemporary neutral

microevolutionary processes (drift and gene flow), coupled with different historical exploitation patterns. However, the reported asymmetrical migration patterns between the two *M. capensis* populations, confirmed by Kapula et al. (2022), appeared to be associated with environmental differences across the Benguela Current system, as the northern Benguela subsystem is characterized by colder sea surface temperatures (SST), and frequent low oxygen water events (LOW), when compared to the southern Benguela subsystem (Hutchings et al., 2009). Similarly, Mbatha et al. (2019) revealed that the distribution of juvenile *M. paradoxus* is affected by dissolved oxygen and temperature. These findings thus suggest that local adaptation might also be an important evolutionary mechanism shaping the evolutionary history of the southern African hakes. The data previously collected, however, cannot provide insights into the interplay among environment, microevolutionary mechanisms and life-history features in these two economically valuable marine fishes. Therefore, the southern African hakes represent a good example of commercially exploited fishes for which molecular tools have had potentially important management outcomes, but for which new high-throughput data have the potential to translate molecular insights into more accurate fisheries outcomes that take into consideration the effects of local adaptation.

Given the previous findings on diversity and differentiation for the southern African hakes and their life-history features, as well as the environmental differences across the Benguela Current region, we hypothesize that genome-wide markers will allow us to identify cryptic population substructuring in these species due to adaptation to local environmental conditions. In this context, we used a population-based approach (Pool-Seq) to generate genome-wide SNPs for the southern African hakes to assess the mechanisms involved in shaping the evolution of these two species. In particular, we were interested in understanding if: (1) extant patterns of diversity and divergence are unique to each species; (2) contemporary genomic patterns are influenced by selection, genetic drift and/or gene flow; (3) both species show similar demographic histories; and (4) whether identified outlier SNPs are associated with specific environmental features. Our results showed the presence of cryptic populations in both species, as well as outlier loci, although no significant associations were observed between loci and environmental conditions. These findings provide further support to the need of incorporating genomic data for conservation of species (Hunter et al., 2018), and in particular for the management of commercially important marine fishes, to avoid further biodiversity loss.

2 | MATERIALS AND METHODS

2.1 | Sampling and Pool-Seq sequencing

Samples used in this study were collected in 2012 for the study of Henriques et al. (2016), from fisheries surveys by the Department of Forestry, Fisheries and Environment (DFFE – South Africa) and the Ministry of Fisheries (Namibia), and ethical clearance for data

usage was obtained from the University of Pretoria (NAS217/2021). The sampling design mimicked the main oceanographic features of the Benguela Current system, with samples collected from north and south of the perennial upwelling cell, as well as past Cape Point towards the east in South Africa (Figure 1a). For *M. capensis*, samples were obtained from northern Namibia to the southern West Coast of South Africa, as previous findings reported a single population break across the Orange River. For *M. paradoxus*, sampling took place from northern Namibia to south-western South Africa, as von der Heyden et al. (2007) reported a putative genetic break across Cape Point in South Africa, which was not detected by Henriques et al. (2016). A reduced representation, pooled sequencing approach was employed (ezRAD Pool-Seq–Toonen et al., 2013) to generate SNPs, as this methodology is a cost-effective way to accurately obtain population-based allele frequencies (Dorant et al., 2019).

Muscle tissue was sampled from 20 to 40 individuals for each pool per species, and their DNA was pooled based on average latitude and depth of the sampling sites, as well as the genotypes of *M. capensis* individuals obtained from microsatellite data (Henriques et al., 2016), meaning that for this species, northern and southern Benguela pools contained individuals with either a northern or southern genotype. Four *M. capensis* pools were obtained: northern Namibia (CNN), central Namibia (CCN), northern South African west coast (CWC) and southern South African west coast (CWC2). The five *M. paradoxus* pools included northern Namibia (PNN), central Namibia (PCN), northern South Africa west coast (PWC), southern South African west coast (PWC2) and South African southwest coast (PSW). DNA was extracted using the chloroform:isoamyl alcohol method of Winnepeninckx et al. (1993) and quantified using a broad-range QUBIT assay (Invitrogen) at the Central Analytical Facilities of Stellenbosch University. DNA samples were pooled in equal concentrations and sent for paired-end sequencing at the Hawai'i Institute of Marine Biology (HIIMB), following the protocol of Knapp et al. (2016). This method employs a high-frequency restriction enzyme (*DpnII*–5' GATC 3') to shear DNA and uses the Kapa Hyper Prep kit for library preparation. Sequencing was conducted on an Illumina MiSeq platform for 600 cycles, generating paired-end reads of 300 bp.

2.2 | Bioinformatic pipeline

Raw reads were received in fastq format and checked for quality in FASTQC (Andrews, 2010). Three trimming and quality checks were done sequentially to improve read quality in TrimGalore! 0.6.5, namely, reads with a low Phred quality score ($Q < 25$), shorter than 50 base pairs (bp) and containing uncalled bases ('N') were removed from the data sets. A final trimming of 5 bp from both 5' and 3' ends of all reads was performed to improve per base sequence content. The cleaned and trimmed reads were merged using FLASH 1.2.11 (Magoč & Salzberg, 2011) with default parameters, as this significantly improves read alignment to reference genomes (Li & Durbin, 2009).

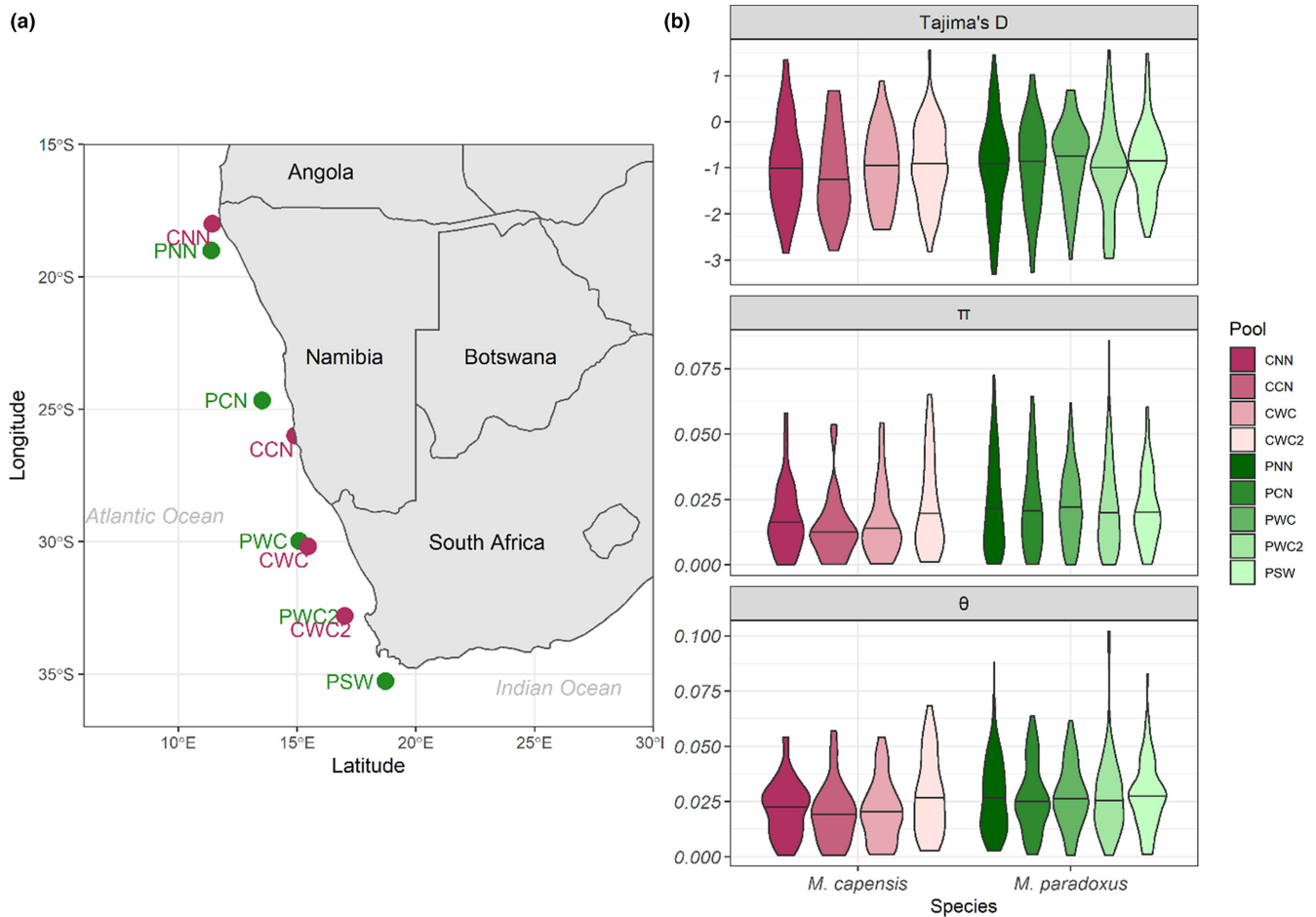


FIGURE 1 (a) Sampling strategy for *Merluccius capensis* (C) and *Merluccius paradoxus* (P) across the Benguela region: NN—northern Namibia, CN—central Namibia, CWC—north West Coast South Africa, WC2—south West Coast south Africa; SW—Southwest Coast South Africa. (b) The variation and densities of Tajima's D , nucleotide diversity (π) and Watterson's estimator (θ) calculated in a 500 bp window, across each pool for *M. capensis* (1271 SNPs) and *M. paradoxus* (3494 SNPs), with mean represented by the black bar, and variance represented by the width of the violin plot.

To create a nuclear DNA (nDNA) data set for further downstream analyses, mitochondrial reads were identified and removed. As there are no available mitochondrial genomes for either of the study species, the cleaned reads were mapped to the mitochondrial genome of the European hake *M. merluccius* (accession number FR751402), using BWA-MEM 0.7.17 alignment program with default parameters (Li, 2013; Li & Durbin, 2009). The resulting SAM files were used to extract reads that mapped to the reference mitogenome in samtools 1.9 (Li et al., 2009; Li, 2021), using the 'filterbyname' script in BBmap (Barnett et al., 2011). The nDNA-only data sets were mapped to the available reference genome of *M. capensis* (accession number GCA_900312945.1, total size 414.3Mb) in BWA-MEM, with resulting SAM files sorted, indexed, merged (combined and non-combined FLASH reads) and converted into BAM files. In order to decrease the possibility of secondary alignments (i.e. where reads map to more than one location in the reference), the BAM files were filtered further, retaining only reads with mapping quality (MAPQ) \geq 20, and subsampled all bam files in samtools 1.9 (Li et al., 2009; Li, 2021) to

the minimum number of reads across pools to ensure equal representation of each pool in downstream analyses.

SNP calling was first performed in samtools 0.1.18, generating an mpilup file per species, which was then converted to sync files using the 'mpilup2sync.py' script from PoPoolation (Kofler, Orozco-terWengel, et al., 2011). Pooling and sequencing error biases are accounted for when using PoPoolation as it is designed specifically for Pool-seq data. Further SNP filtering was done using the 'snp-frequency-diff.pl' script from PoPoolation2 (Kofler, Pandey, & Schlötterer, 2011), and SNPs were retained if they had a minimum allele count of 3, minimum coverage of 20 reads and a maximum coverage of 300 reads, to minimize the possibility of called SNPs resulting from sequencing errors as well as over-merging, that is, collapsing reads of different RAD-loci into a single RAD-locus. This approach ensured that for a SNP to be called, all pools had to have at least three reads in that region. Therefore, if a SNP had an allele frequency of zero in any particular pool, it would result from that pool having the reference allele and not from missing data. Further filtering was conducted using custom

scripts (Data S1) to retain only biallelic SNPs and SNPs between populations for *M. paradoxus*, to avoid capturing SNPs against the *M. capensis* reference in this data set. Given the type of data and the aim to understand if natural selection may play a role in shaping the evolutionary history of these two species, the data were not filtered for linkage disequilibrium.

2.3 | Estimates of genomic diversity and population differentiation

To characterize genome-wide levels of diversity for each pool and species, several metrics were calculated (Table S1). First, the minor allele frequencies were obtained from the sync files using a custom-built command (Data S1), to obtain the number of total and private SNPs (as a measure of uniqueness) per pool and species, in R Core Team (2021). Second, nucleotide diversity (π), Watterson's theta estimator (θ) and Tajima's D were calculated for each species and each pool in PoPoolation (Kofler, Orozco-terWengel, et al., 2011). These metrics were calculated in 500bp windows, as recommended by Kofler, Orozco-terWengel, et al. (2011) for pools with low-sequencing coverage ($<50\times$), with a minimum read count of 2, minimum coverage of 20 reads, maximum coverage of 300 reads, minimum percentage of sites having sufficient coverage in the given window of 60% and a minimum mapping quality of 20, using the 'Variance-sliding.pl' script of Popoolation, per pool and species, and their average calculated for intervals where SNPs were present (number of SNPs different from zero and metric different from zero, to avoid calculating diversity for SNPs that might be fixed against the reference). The pool size was specified as the median number of gene copies (Table S2). In order to understand if obtained metrics of diversity were significantly different between species, a Mann-Whitney-Wilcoxon test was performed in R Core Team (2021).

To assess patterns of population substructuring in the two species, principal component analyses (PCA) were performed based on the exact allele counts generated by PoPoolation2 (Kofler, Pandey, & Schlötterer, 2011), using the packages 'dplyr', 'ade4', 'FactoMineR', 'factoextra', 'gridExtra' and 'utils' in R (Auguie, 2017; Kassambara & Mundt, 2021; Lê et al., 2008; R Core Team, 2021; Thioulouse et al., 2018; Wickham et al., 2021). Preliminary PCAs for *M. paradoxus*, and to a lesser extent *M. capensis*, revealed a clustering pattern which coincided with the grouping of pools for sequencing, that is, a batch effect. Batch effects are variation patterns observed between subgrouped data due to technical variables, and can drastically influence the conclusions drawn from a study (Leek et al., 2010; Lou & Therikildsen, 2021). They can occur at three different stages: during DNA extraction, library preparation and sequencing. Sequencing batch effects are likely due to differences in sequencing chemistry, read type, sequencing run, read length, DNA degradation level and sequencing depth (Lou & Therikildsen, 2021). During the DNA extraction and pooling step, DNA extraction protocols were standardized, pooling samples with similar DNA concentrations and qualities across pools, in order to minimize variability. Although it is not

possible to discard the possibility that problems arose during library preparation, the same protocol has been used for different studies on different taxa by our group (Nielsen et al., 2018, 2020; Schulze et al., 2020) without similar effects being observed. The most likely explanation might be linked with differences in DNA degradation and/or sequencing chemistry across pools, as the observed effect accurately mimicked the order in which the samples were sequenced particularly for *M. paradoxus*. Lou and Therikildsen (2021) noticed a similar effect when comparing low coverage whole genome data from different studies, and suggested developing batch effect resistant bioinformatic pipelines, particularly during the SNP calling phase. In this study, to assess the possibility of a batch effect, the contribution of each SNP to the differentiation along the PC 1 was computed from the PCAs using the entire data sets and extracted into a data frame. The top 1%, 5%, 10% and 15% of SNPs causing the differentiation observed in the individual PCAs were identified. These four sets of SNPs were sequentially removed from the original data sets to evaluate their impact in the population substructuring patterns identified. Results indicated that the top 5% for *M. capensis* and the top 10% *M. paradoxus* contributing SNPs should be removed to adequately eliminate the observed batch effect (Data S3), while still retaining enough information for population-based inference, and this was completed using the R package 'dplyr'. The filtered data sets were then used in all downstream analyses and are hereafter named the final data sets.

Overall and pairwise F_{ST} were calculated for each species using the R package 'Poolstat' (Hivert et al., 2018). The filtered sync files were converted to a 'pooldata' object using 'popsync2pooldata', with additional filtering steps including exact pool size (haploid number of individuals per pool), and a minimum minor allele frequency (MAF) of 1%. Overall F_{ST} was calculated using the command 'computeFST', while pairwise F_{ST} were calculated using the 'compute.pairwiseFST' command, specifying ANOVA as method for the analysis of variance and with a block-jackknife of 2 SNPs per block, given the number of contigs present in the reference genome (Hivert et al., 2018). In order to understand if using different number of individuals per pools would significantly impact our estimates of divergence, F_{ST} was estimated for (i) original number of individuals per pool; (ii) minimum number of individuals per pool (20 for *M. capensis* and 35 for *M. paradoxus*); and (iii) using the median number of individuals per pool (35 for both species). Statistical significance among pairwise estimates for each scenario was determined using a Kruskal-Wallis test in R.

Finally, to assess genome-wide patterns of F_{ST} for each of the Southern African hakes, the sliding F_{ST} values were calculated using 'fst-sliding.pl' in PoPoolation2 (Kofler, Pandey, & Schlötterer, 2011). F_{ST} was calculated per SNP only for variable SNPs, with the same parameters as before. Visualization of the genome-wide distribution of SNP-specific pairwise F_{ST} was done using a scatterplot in the R package 'ggplot2' (Wickham, 2016) for only the most genetically (*M. capensis*) or geographically (*M. paradoxus*) distant pools of each species as previously identified (i.e. CNN vs. CWC2 and PNN vs. PSW2, as per Figure 1).

2.4 | Demographic history

Reconstruction of demographic history for the southern African hakes was performed based on two simple models, a single panmictic population (PAN) or derived from two structured populations connected by gene flow (IM), using the demographic modelling approach implemented in *moments* (Jouganous et al., 2017). This software uses a likelihood approximation approach to infer the demographic history between two sets of individuals from the joint site frequency spectrum (jSFS). This statistic only requires the allele counts sampled in a given location and does not use individual-based information, making it perfectly suited for Pool-Seq data. In the absence of an outgroup like in this case, the jSFS is computed in its folded version using only the allelic count obtained from the overall MAF. To bypass the effect of variable depth of sequencing across SNPs within and across all pools, allele counts were standardized as if 20 chromosomes were sequenced (~10 diploid individuals) prior to computing the jSFS, based on allele counts generated in Poolfstat as above, without enforcing a minor allele frequency. Then, the ability of the two models to reproduce the observed jSFS was compared: A situation where an ancestral population went through changes in population size sometime in the past without any population split (PAN model), and a classic island model with one ancestral population split in two derived populations, which have then diverged for a certain period of time (IM model; Figure S1). The two models were then adjusted to the data using the parameter optimization procedure initially developed by Portick et al. (2017) for dadi and adapted to *moments* in Momigliano et al. (2021) and Le Moan et al. (2021). This routine uses a four-step optimization procedure to infer the best parameters of each model. Each of the optimization steps was replicated 20, 20, 40 and 40 times, respectively, and the optimization routine was run 20 times for each model to control for convergence. The 10 best runs were kept for each model, and the best model was then selected based on the Akaike information criterion (AIC). A stringent threshold of four points of AIC differences with the best model (Δ AIC) was used to assess the significance of a model. For each species, the model comparison was run between each possible pair of pools, for a total of six and 10 pairwise comparison in *M. capensis* and *M. paradoxus*, respectively.

2.5 | Environmental association analyses

Assessment of possible local adaptation and association of observed population substructuring patterns with environmental features were performed employing an environmental association analysis (EAA) in BayPass 2.31, which is specifically designed for Pool-Seq data (Gautier, 2015). Environmental data were obtained from Bio-Oracle (Assis et al., 2017), using the R packages 'sdmpredictors' (Bosch et al., 2022) and 'leaflet' (Cheng et al., 2021). Given the previous findings of Henriques et al. (2016), the environmental layers included: mean chlorophyll a concentration (mg/m^3); current velocity (mean at mean depth; m/s); mean dissolved oxygen concentration at mean depth (mmol/m^3), mean sea water temperature (SST) at mean depth ($^{\circ}\text{C}$); mean primary production at mean depth ($\text{g}/\text{m}^3/\text{day}$);

maximum sea water salinity at mean depth (PSS); mean sea surface temperature ($^{\circ}\text{C}$); and average depth of the seafloor (m). Collinearity between environmental variables was assessed using the 'dudi.pca' function and the 'usdm' R package (Naimi et al., 2014; Thioulouse et al., 2018), and variables that were found to be collinear were removed, based on PCA correlation circles, a variance inflation factor (VIF) > 10 and a Pearson correlation > 0.7.

The 'pooldata' objects generated in the section above were then used to generate 'genobypass' files in Poolfstat for BayPass 2.31. These were run through the standard covariate model of BayPass which allows to assess the extent that a population covariable (environmental variable, in this case) is associated with each SNP (Gautier, 2015), and the recommended model for studies with low number of samples (Gautier, personal comm.). Five independent runs were conducted per species, using the 'scalecov' option, to scale each covariable in the environmental database. Given the small number of pools and SNPs, the medians of resulting metrics were calculated for the five runs (Gautier, personal comm.). Detection of outliers in BayPass 2.31 were based in two different approaches: First, based on the XtX values, a metric analogous to F_{ST} corrected for the scaled covariance of allele frequencies of the populations; second, based on environmental associations. In the first approach, a pseudo-observed (POD) data set was created to estimate the posterior predictive distribution of the XtX values. In this case, outlier SNPs were selected if they fell within the 99% quantile of the POD XtX distribution (Gautier, 2015; Nielsen et al., 2020). In the second approach, environmental associated SNPs were obtained from the relationship between environmental covariates and allele frequencies, with significance assessed based on calculated Bayes Factors (BF) for each locus, where loci with a \log_{10} BF (db) > 20 were identified as outliers (Nielsen et al., 2020).

In addition, redundancy analyses (RDA) were performed using the 'vegan' R package, to calculate linear regressions between allele frequencies and the environmental variables at the sites of interest while simultaneously being constrained using a PCA (Forester et al., 2018; Oksanen et al., 2020). Allele frequencies were obtained from the 'genobypass' files generated above and subjected to a Hellinger transformation (Legendre & Gallagher, 2001). The environmental variables were centred and scaled using the 'scale' function (Forester et al., 2018). Using both the adjusted R^2 value and an ANOVA, the significance was tested. Putative adaptive loci were considered to have a loading score of ± 3 standard deviation of the average loading for the first two axes (Forester et al., 2018). Preliminary results for *M. paradoxus* suggested the presence of a batch effect still, and thus, a partial RDA was conducted for this data set, using sequencing lane as a conditional categorical variable.

3 | RESULTS

3.1 | SNP calling and filtering

The average number of raw reads per pool was ~3,239,433 for *M. capensis* (four pools) and ~3,406,445 for *M. paradoxus* (five pools), and after quality control ~2,960,270 reads were retained

for *M. capensis* and ~2,950,904 for *M. paradoxus* (Table S1). Very few reads mapped to the mtDNA of *M. merluccius*, with *M. paradoxus* showing, on average, a larger number of reads mapped than *M. capensis* (Table S1). The nuclear-only data sets for each species mapped to the reference genome with a success of 49.88% for *M. capensis* and 46.27% for *M. paradoxus*, for MAPQ >20. After subsampling to the lowest number of mapped reads per species, 6030 SNPs were retained for *M. capensis* and 4963 SNPs were retained for *M. paradoxus*, with average coverages of 16.81× and 26.58× respectively. Only 512 SNPs were common between both species, representing 8.5% of the *M. capensis* data set and 10.3% of the *M. paradoxus* data set.

3.2 | Genomic diversity metrics and population substructuring

Overall, *M. capensis* showed a lower level of genomic diversity ($\theta=0.019$, $\pi=0.015$), compared to *M. paradoxus* ($\theta=0.026$, $\pi=0.021$; Figure 1b, Table S2). *Merluccius capensis* had a lower number of private SNPs, ranging from 522 to 2923 per pool, than *M. paradoxus* pools (1181–3493 private SNPs; Table S2). Within each species, diversity metrics did not vary greatly across pools, but Watterson's θ was consistently higher than π (Figure 1b, Table S2). Tajima's D was similar across species, with $D=-0.839$ and $D=-0.913$ for *M. capensis* and *M. paradoxus*, respectively (Table S2). Tajima's D per pool ranged from $D=-0.951$ (CNN) to $D=-0.818$ (CWC) in *M. capensis* and from $D=-0.764$ (PWC) to $D=-0.983$ (PWC2) in *M. paradoxus*

(Table S2). The distribution of all metrics was similar across the southern African hakes but *M. paradoxus* pools did show the greatest variance (particularly π and θ , Figure 1b). Only the comparisons between nucleotide diversity were significantly different from zero between species ($p < .05$).

Patterns of population substructuring between the pools of each species were first visualized and assessed using a PCA (Figure 2a,b,d,e). A clear pattern emerged for *M. capensis*, where the Namibian pools (CNN and CCN) clustered together separately from the South African pools (CWC and CWC2), while the South African pools appeared isolated from each other (Figure 2a). This structuring pattern of the *M. capensis* pools remained the same even when loci that may have been involved in the batch effect were accounted for (305 loci removed, Figure 2b). On the contrary, using the complete data set resulted in a clear grouping of *M. paradoxus* pools which matched the sequencing lanes (PNN+PWC2 and PCN+PWC, with PSW somewhat isolated from the rest—Figure 2d). When these loci were removed (498 loci), there were no obvious geographical clustering patterns (Figure 2e).

Further filtering in poolstat resulted in 3602 SNPs for *M. capensis* and 2388 SNPs for *M. paradoxus* (MAF > 0.01). No significant differences were observed between estimates of pairwise F_{ST} between scenarios (i)–(iii) for pool size for *M. paradoxus*. For *M. capensis*, only scenario (ii) all pools of minimum size 40 (20 individuals) were significantly different from the remaining scenarios (Data S3). Overall values of F_{ST} were not significantly different from 0 for either species ($F_{ST}=0.001$, 95% CI = -0.001 to 0.004, for *M. capensis* and $F_{ST}=-0.007$ for *M. paradoxus*, 95% CI = -0.009 to 0.005), suggesting

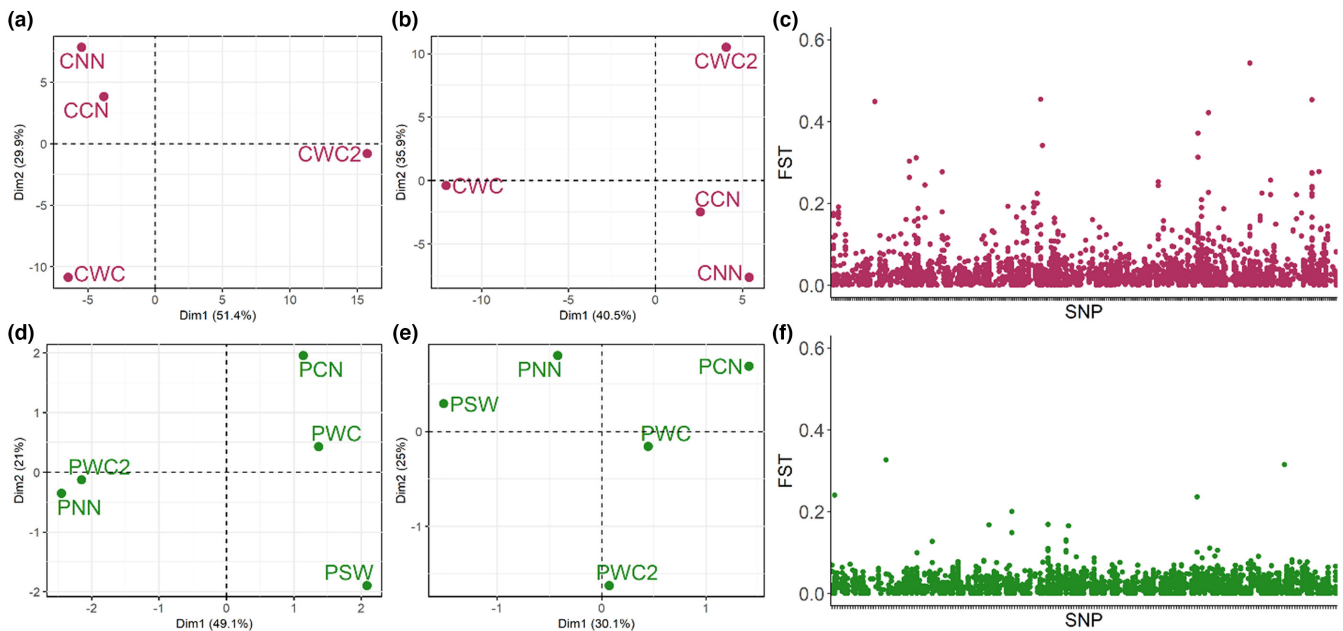


FIGURE 2 Principal component analyses (PCAs) of the genomic pools of southern African hakes' samples to graphically represent the genetic variation observed between the two species: (a) and (b) for *Merluccius capensis* and (d) and (e) for *Merluccius paradoxus*. The PCAs represent the entire data sets (a, d) and the data sets without the top percentage of SNPs associated with a batch effect (b, e). Pairwise F_{ST} values (0–1) calculated for each SNP of the most geographically distant pools of (c) *M. capensis* (CNN vs. CWC2) and (f) *M. paradoxus* (PNN vs. PSW), for the final data sets. Pool labels are shown in Figure 1a.

a higher within-pool than across-pool variability. This is most likely due to the fact that if there are no strong geographical structuring patterns, such as expected from marine fishes, then most of the genetic variability will be present within each pool, and private alleles in a given pool will be rare. For *M. capensis*, pairwise F_{ST} were only significantly different from zero for the CNN and CWC2 comparison ($F_{ST}=0.013$, 95% CI=0.0075–0.0189), despite a small divergence also observed for the CWC and CWC2 comparison ($F_{ST}=0.002$, 95% CI=–0.0023 to 0.0070; Table S3). For *M. paradoxus*, no pairwise comparison was significantly different from zero (Table S4). The baseline F_{ST} for each identified SNP within each species was around zero (Figure 2c,f). However, there were SNPs with elevated values of F_{ST} in both species, with more observed for *M. capensis* (Figure 2c). The maximum SNP-specific F_{ST} was higher in *M. capensis* ($F_{ST}=0.543$) compared to *M. paradoxus* ($F_{ST}=0.326$; Figure 2c,f).

3.3 | Demographic history

The jSFS is a half-matrix that allows to visualize how well the allelic count of each SNP in a given pool correlate with the allelic count of another pool. Each cell of the matrix is coloured according to the number of SNPs it contains. The cells on lower and left side of the matrix are associated with private alleles (allele only found in one pool), while the cells on the diagonal $x=y$ are associated with SNPs in equal frequency in the two pools. Thus, the more concentrated around the diagonal the jSFS is, $x=y$, the more similar the pools are. Here, the jSFS showed clear differences between *M. capensis* and *M. paradoxus* (Figure 3a,b, Tables 1 and 2). Indeed, the jSFS between *M. paradoxus* pools was more restricted to the diagonal $x=y$ than

in *M. capensis*, suggesting more pronounced differences in the latter. In accordance with this observation, most of the *M. paradoxus* inferences showed similar support for IM and PAN model (4 of 10, Table 2) or supported a PAN model (2 of 10, Table 2), suggesting a lack of structure (on the more spatially robust level) in this species. Only the pairwise comparison involving the south-eastern most pool, PSW, showed consistent support for the IM model (3 of 4 pairwise comparison with $\Delta AIC > 4$ between IM and PAN, Table 2). In contrast, all but one inference showed support for an IM model in *M. capensis*, with only the two Namibian pools (CNN and CCN) showing support for the PAN model (Table 1). The details about the parameters estimated by each model are provided in Table S5.

3.4 | Environmental associations

After controlling for collinearity and VIF, only two environmental variables were retained for *M. capensis*: mean dissolved oxygen (DOmean) and mean sea surface temperature (SSTmean), while three variables were retained for *M. paradoxus*: mean chlorophyll a concentration (CAmean), mean dissolved oxygen (DOmean) and mean sea salinity (SSmean; Tables S6–S8).

The population scaled covariance matrices (Ω) indicated some differentiation between pools of both species (Figure 4). *Merluccius capensis* pools which are geographically closer exhibited a higher correlation of scaled covariance; however, as before, there was no clear geospatial pattern observed for *M. paradoxus*. The hierarchical clustering trees showed three distinct clusters (one Namibian and two South African) observed for *M. capensis* and no obvious geographical pattern for *M. paradoxus*, as seen for the PCAs (Figure S2).

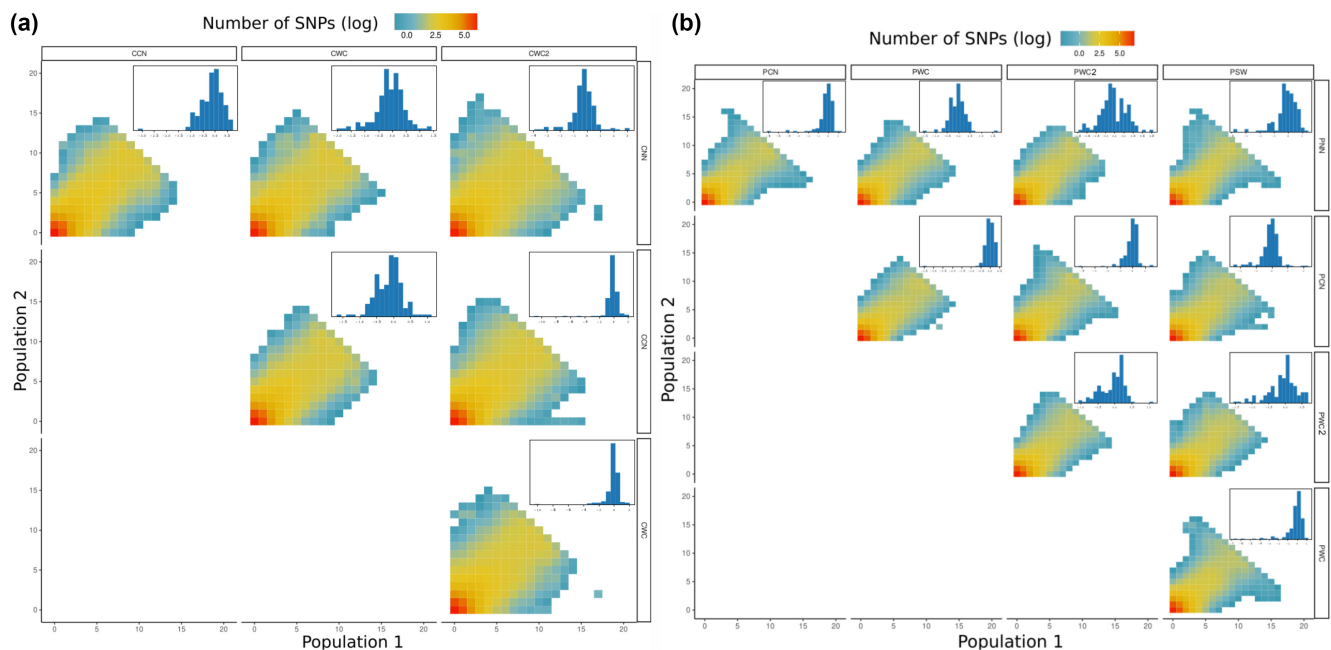


FIGURE 3 Joint site frequency spectrum for (a) *Merluccius capensis* and (b) *Merluccius paradoxus*, with inserts depicting the distribution of residuals obtained from the best model (SFS obs–SFS theoretical). Pool labels are shown in Figure 1a.

Although 37 of 3602 SNPs were indicated to be outliers for *M. capensis* based on the XtX metric, none exhibited a significant association with either of the environmental variables tested. For *M. paradoxus*, 24 of 2388 SNPs were identified as outliers based on the XtX metric, but again no significant associations were observed between these and the environmental variables used.

The RDAs showed adjusted R^2 values of .020 ($p = .417$) and .250 ($p = .008$) for *M. capensis* and *M. paradoxus* respectively. Therefore, no statistically significant environmental associations were found for *M. capensis*. On the contrary, although weak, *M. paradoxus* RDA appeared to show statistically significant associations between environmental conditions and allele frequencies, with 98 SNPs identified

as outliers. However, when a partial RDA was conducted using sequencing lane as a conditional variable, the model was no longer statistically significant (adjusted $R^2 = -.064$, $p > .05$), suggesting that observed associations were still with sequencing lane and not environment. Redoing the analyses for the data set with 15% of batch effect loci removed did not change these findings (adjusted $R^2 = .243$, $p < .05$; conditional RDA adjusted $R^2 = -.078$, $p > .05$).

4 | DISCUSSION

4.1 | Genome-wide diversity levels for the southern hakes

Genome-wide diversity is fundamental for the long-term persistence of species. Marine fish tend to exhibit high levels of diversity due to historically high effective population sizes, reproductive potential and connectivity levels. Based on the neutral coalescence theory, the level of genomic diversity (θ) is proportional to the mutation rate (μ) and the effective population size (N_e), such that $\theta = \mu 2N_e$, assuming an equilibrium has been reached between mutation and genetic drift (Kingman, 1983). However, recent work has shown that other factors, such as habitat and life-history features, can influence genetic diversity levels in marine fishes (Barry et al., 2021; Martinez et al., 2018). In particular, age at maturity, fecundity and variance in reproductive success (e.g. age-specific

TABLE 1 Summary of the pairwise inferences from *Merluccius capensis*.

	CCN	CCN	CWC	CWC2
CCN	x	IM=PAN	IM	IM
CNN	1.12	x	IM	IM
CWC	19.91	7.70	x	IM
CWC2	166.23	87.40	65.34	x

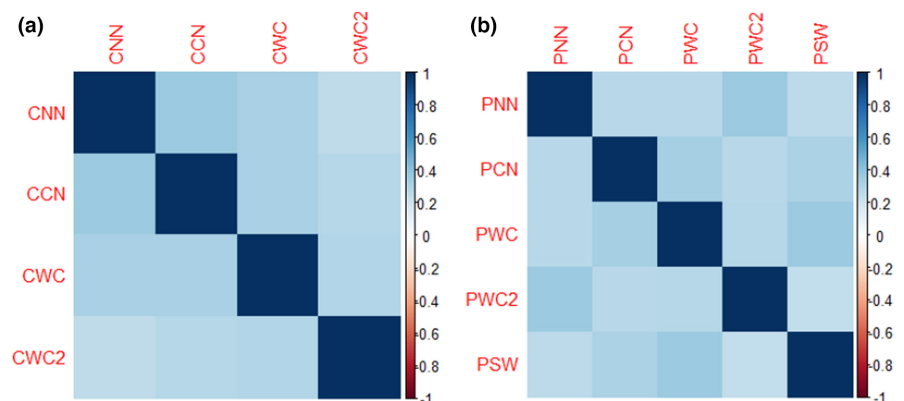
Note: The name of the best model in each pairwise comparison is showed above the diagonal, and the Δ AIC between the IM and the PAN model is showed below the diagonal. A Δ AIC <4 shows an equal support for both IM and PAN, suggesting that the effect of the population split has no significant effect. Pool labels are shown in Figure 1a. Statistically significant values ($p < .05$) in bold.

TABLE 2 Summary of the pairwise inferences from *Merluccius paradoxus*.

	PNN	PCN	PWC	PWC2	PSW
PNN	x	IM=PAN	IM=PAN	PAN	IM
PCN	2.79	x	PAN	IM	IM
PWC	2.88	7.39	x	IM=PAN	IM=PAN
PWC2	4.72	5.09	2.26	x	IM
PSW	12.6	11.6	0.37	23.6	x

Note: The name of the best model in each pairwise comparison is shown above the diagonal, and the Δ AIC between the IM and the PAN model is shown below the diagonal. A Δ AIC <4 shows an equal support for both IM and PAN, suggesting that the effect of the population split has no significant effect. Pool labels are shown in Figure 1a. Statistically significant values ($p < .05$) in bold.

FIGURE 4 Population scaled covariance matrices (Ω) obtained with the standard covariate model with scaled covariables in BayPass 2.31 for the pools of (a) *Merluccius capensis* (3602 SNPs) and (b) *Merluccius paradoxus* (2388 SNPs). Pool labels are shown in Figure 1a.



mortality) can all impact the ratio between census size (N) and N_e , leading to deviations of genomic diversity levels from theoretical predictions (Barry et al., 2021).

Henriques et al. (2016) first showed reduced contemporary genetic diversity levels for the southern African hakes, especially when compared with also the heavily exploited European hake (*M. merluccius*—Castillo et al., 2005): $H_E=0.518$ for *M. capensis*; $H_E=0.692$ for *M. paradoxus* and $H_E=0.871$ for *M. merluccius*. Our present work further supports the findings of low genome-wide diversity for the southern African hakes, with *M. capensis* consistently showing lower values than *M. paradoxus* over 500bp non-overlapping intervals ($\theta=0.019$, $\pi=0.015$ vs. $\theta=0.026$, $\pi=0.021$ respectively). However, only the comparisons involving nucleotide diversity between species were significantly different from zero ($p<.05$). As diversity metrics were calculated based on 500bp intervals, it was not possible to extract batch effect loci, which may have affected the obtained estimates, particularly for *M. paradoxus*. Furthermore, since the reference used for mapping and SNP calling was *M. capensis*, it is possible that for calculation of windows-based metrics of diversity, *M. paradoxus* retained loci against the reference, instead of only population-based SNPs, inflating its estimates of diversity. It is thus possible that estimates for *M. paradoxus* may be overall lower than here documented, and we cannot at this time confidently reject the hypothesis that *M. capensis* and *M. paradoxus* have similar levels of genome-wide diversity. The observed differences between pairwise nucleotide diversity and Watterson's theta, as well as the negative Tajima's D , suggest an increase in rare alleles (there are more polymorphic sites than sequence diversity), pointing to a population expansion in both species. However, this expansion has likely occurred in the last thousands of years, and not in recent ecological time. Immediately after a bottleneck, Tajima's D is expected to be positive (loss of rare variants), and its value will decrease as the percentage of singletons increase with population growth (Tajima, 1989). Previous studies based on mtDNA data estimated population expansion to have occurred after the end of the Last Glacial Maximum for *M. paradoxus* (~6000 years, von der Heyden et al., 2010) and between 4000 and 75000 years for *M. capensis*, depending on the mtDNA locus, mutation rate and dating method used (Henriques et al., 2016; von der Heyden et al., 2007). Although different values can be observed between mtDNA and nDNA due to their different effective population sizes (Gattepaille et al., 2013), in this study, values of D were similar in range to the ones reported for mtDNA (*M. capensis* $D=-1.687$, *M. paradoxus* $D=-1.036$, Henriques et al., 2016), thus suggesting an historical expansion. Given the estimated time frame and life-history features of the species, it is likely that a historical population growth resulted from a range expansion as environmental conditions changed in the Benguela Current region during the Last Glacial Maximum (von der Heyden et al., 2007).

Although caution is warranted, these results, in combination with previous findings of Henriques et al. (2016), suggest that N_e for *M. capensis* should be similar in range to that of *M. paradoxus*. Interestingly, based on fishery-independent abundance indices, overall abundance levels of *M. capensis* are higher than *M. paradoxus*

(DFFE, 2020). For example, in Namibia, the abundance of *M. capensis* is roughly three times higher than that of *M. paradoxus*, while in South Africa, although *M. paradoxus* is more abundant in the west coast, *M. capensis* dominates in the south and east coasts (DFFE, 2020). *M. paradoxus* also experiences intense predation of its juveniles by *M. capensis* adults (as they overlap in depth), leading to high natural mortality levels at this life stage (Ross-Gillespie, 2016), as well as years of poor recruitment levels (DFFE, 2020), both of which contribute to a higher variance in survivability for this species. Furthermore, although both southern African hakes experienced appreciable population declines between 1950 and 1970, with the biomass of both species declining to below maximum sustainable yield levels (MSYL), *M. capensis* recovered relatively quickly and its biomass has generally been well above MSYL since the 1980s (DFFE, 2020). The abundance of *M. paradoxus*, on the contrary, remained below MSYL until the 1990s (DFFE, 2020), only showing an increase in recruitment success between 2005 and 2010, which was not observed for *M. capensis* (Strømme et al., 2016).

These demographic and life-history features would thus be expected to result in lower population sizes (Barry et al., 2021) and consequently significantly lower genome-wide genetic diversity levels for *M. paradoxus* when compared to *M. capensis*, which does not appear to be the case based on our findings. However, while *M. paradoxus* has a life cycle showing one panmictic population that migrates and expands over a vast area, with juveniles having relatively good conditions for feeding, growth and survival (Mbatha et al., 2019; Strømme et al., 2016), *M. capensis* appears to be composed by multiple demographic units across southern Africa, with different life-history features and limited connectivity in between (Jansen et al., 2015; Strømme et al., 2016). Smaller populations that show a degree of genetic isolation are likely to be more vulnerable to loss of diversity due to genetic drift (Wright, 1931), which may explain the observed genome-wide estimates of the two southern African hakes.

Interestingly, observed values of genome-wide diversity of *M. capensis* are similar to those of another highly structured southern African demersal fish, kingklip (*G. capensis*—Schulze et al., 2020). Using the same sequencing and SNP-calling methodology, and mapping to a *de novo* assembly of kingklip, Schulze et al. (2020) reported genome-wide diversity levels of $\theta=0.017$, $\pi=0.017$. Kingklip is currently caught as by-catch in the hake-directed fisheries, as there is a fishing ban for this species since 1994, due to an extreme population decline driven by overharvesting (Punt & Japp, 1994). It is thus possible that observed levels of genome-wide genetic diversity of the southern African hakes reflect a complex interplay between demographic and evolutionary history as well as historical exploitation levels. In order to properly assess the impact of historical fishing levels in the demographic history of the southern African hakes, it would thus be necessary to (i) evaluate the true extension of mapping *M. paradoxus* to a different species reference genome in estimates of genetic diversity; and (ii) obtain samples that predate the population collapse to compare genomic diversity and population structure levels through time (Manuzzi et al., 2022).

4.2 | Patterns of population structuring of the southern African hakes across the Benguela region

Estimates of population structure between the two southern hakes revealed contrasting patterns and differentiation levels. After removal of loci that were likely associated with the sequencing lane (batch effect loci), three populations were observed for *M. capensis*, while no obvious geographic structuring pattern was retrieved for *M. paradoxus*. These results were consistent in the PCA analyses, pairwise F_{ST} (Tables S3 and S4), but also in the baseline level of SNP-specific F_{ST} , which showed higher variation and more outliers in *M. capensis*, compared to *M. paradoxus*. Although these differences in population structure between the two species were also shown in previous work (Henriques et al., 2016), the presence of a previously un-identified population within the southern Benguela region of *M. capensis* is new.

For *M. capensis*, overall pairwise F_{ST} levels were low and non-significant, except for the northern and southern pools (CNN vs. CWC2, $F_{ST}=0.013$, 95% CI=0.0075–0.0189), while the PCA retrieved three major clusters: northern Benguela region (CNN and CCN) and clearly differentiated between CWC (northern west coast) and CWC2 (southern west coast). Visualization of SNP-specific F_{ST} values for this comparison revealed four major areas with elevated levels across the genome, which might suggest the presence of barriers to gene flow in this species. Similar findings have been reported for loci putatively under selection in kingklip, with three populations also described across the region: one in northern Namibia and two in South Africa, albeit being a clear west coast versus east coast South Africa differentiation pattern (Schulze et al., 2020). These findings further support the hypothesis of multiple spawning populations of *M. capensis* in southern Africa, proposed by Jansen et al. (2015) based on a gonadosomatic indices. Here, the authors identified the presence of two spawning populations within the west coast of South Africa (31–32.5°S and 34.5–36°S), as well as spawning females on the south coast of South Africa (east of 20°E). Our results showed two populations both found within the west coast, albeit at different latitudes (30°S vs. 33°S) than those identified by Jansen et al. (2015). It is thus likely that the southern-most population here identified (CWC2) may represent the western edge of the spawning population occurring at 34.5–36°S, as proposed by Jansen et al. (2015). As the main aim of this study was to evaluate the genomic component of the previously described north-south Benguela differentiation in *M. capensis*, where the break was located off the Orange River in the border of Namibia and South Africa and no previous genetic structure was observed within South African waters, we did not include individuals caught on the south and east coasts of the country. Our new findings suggest that this will be necessary to truly understand the geographical distribution of these two populations. Furthermore, although we pooled individuals based on their northern-southern microsatellite genotypes (Henriques et al., 2016), the presence of a third population within this system requires an individual-based sequencing approach in future population genomic surveys.

Unexpectedly, the data showed an indication of spatial differences for *M. paradoxus*. However, there was no reasonable geographical pattern which could explain the observed grouping. Preliminary PCA analyses grouped the northern most pool (PNN) with the west coast of South Africa (PWC2) and the south-western most pool (PSW) with the middle-range pools (PCN and PWC). After controlling for a sequencing-driven batch effect, the PCA and pairwise F_{ST} comparisons suggested a single population throughout the Benguela region. Contrary to what was observed of *M. capensis*, visualization of SNP-specific pairwise F_{ST} between PNN and PSW did not reveal major areas of differentiation across the genome, with only a few, sporadic loci exhibiting higher F_{ST} levels, suggesting that genetic drift and gene flow are the major microevolutionary processes shaping the population structure of *M. paradoxus*.

4.3 | Distinct evolutionary trajectories of co-distributed hake species

Based on the results obtained for population differentiation, we assessed two simple models of demographic history for the southern African hakes: panmixia (PAN) and isolation with gene flow (IM). As expected, *M. capensis* showed a deeply structured jSFS, where only the comparison involving the two northern Benguela pools (CNN and CCN) supported the PAN model. All other comparisons supported the IM model, thus further corroborating the hypothesis of three populations: one in northern Benguela and two within South Africa. On the contrary, *M. paradoxus* showed broad evidence of panmixia, with allele frequencies highly correlated across all pools. The only exception were the comparisons involving the south-eastern-most pool (SWC), and to a lesser extent PCN and PWC2. These results suggest that, as reported by von der Heyden et al. (2007), there might be a subtle population substructuring in *M. paradoxus*, between Atlantic (PNN, PCN, PWC, PWC2) and Indian Ocean sites (PSW). The southwest coast of South Africa is a known phylogeographic break within the region, with some species showing evidence of population substructuring across this area (Dalongeville et al., 2022; Teske et al., 2011). In fact, analyses of life-history features for *M. paradoxus* based on survey data suggest the presence of a single spawning ground off the west coast of South Africa, from 31° to 34°S, but large fish are also found off the East coast of South Africa between 26° and 27°E, which may indicate a small spawning ground in that area (Strømme et al., 2016). *Merluccius paradoxus* is not known to spawn in Namibia (Jansen et al., 2015), which further supports the hypothesis of an Atlantic-Indian Ocean separation. These results, combined with the reported PCA and F_{ST} levels, suggest that divergence between putative *M. paradoxus* populations is quite small, either due to gene flow, a recent isolation event or a combination of both (Andre et al., 2011). Marine species traditionally have large population sizes and high levels of gene flow, which can mask subtle levels of population isolation (Hauser & Carvalho, 2008). High-throughput sequencing, by generating thousands of genome-wide SNP markers, can thus assist in identifying

subtle levels of differentiation in high gene flow species, as seen for both southern African hakes.

The observed differences in population substructuring patterns further suggest that *M. capensis* and *M. paradoxus* are not experiencing the same evolutionary trajectories despite being congeneric and sharing similar geographical distribution ranges (Irwin et al., 2016; Teske et al., 2019; van Doren et al., 2017). Interestingly, these differences may also assist in contextualizing the observed differences in genomic diversity levels between species. As *M. capensis* appears to be composed of at least three demographically independent populations, its N_e is likely to be smaller than expected in comparison to species with panmictic populations. Therefore, *M. capensis* might be more vulnerable to loss of genomic diversity due to genetic drift (Gagnaire, 2020) than *M. paradoxus*, which appears to be composed by two populations with highly correlated allele frequencies, implying higher genetic connectivity. These results suggest an evolutionary response that is species-specific and possibly related to the different environmental niches that the species occupy (Irwin et al., 2016; Teske et al., 2019; van Doren et al., 2017), with similar findings reported among four flatfish species sampled along the Baltic Sea transition zone (Le Moan et al., 2019), as well as stickleback species sampled along a latitudinal gradient (Reeve et al., 2022).

4.4 | Environmental association analyses: The relative irrelevance of the environment in shaping the evolutionary history of the southern African hakes?

Genome-wide patterns of population differentiation varied between *M. capensis* and *M. paradoxus*. In particular, *M. capensis* exhibited not only overall higher levels of differentiation, but also four areas of the genome with elevated F_{ST} , while for *M. paradoxus* only a few loci were found with elevated levels of F_{ST} . These results suggest that adaptation to local environmental features might be driving the observed patterns of differentiation in the former, while genetic drift and gene flow might be the dominant microevolutionary forces shaping the evolutionary history of the latter. In fact, studies have reported associations between local oceanographic features and patterns of demersal fish population structuring (White et al., 2010). For example, in North Atlantic roundnose grenadier (*Coryphaenoides rupestris*), population differentiation is in part driven by a locus under depth-mediated selection (White et al., 2010). The oceanographic features of the Benguela system are stratified by depth so that different depths experience varying water temperatures and current movement. The southern African hakes are only partially sympatric with a preferential depth overlap at 110–500m for all life stages (Durholtz et al., 2015; Henriques et al., 2016). Therefore, we hypothesized that species-specific population substructuring is likely influenced by different local oceanographic conditions, which however did not seem to be supported by the results for the gene–environment association analyses.

Despite the observed elevated F_{ST} in regions of the genome of *M. capensis*, and the detection of outlier loci in BayPass, there

was no correlation between these and the environmental variables. Similarly, the RDA did not show a significant association between allele frequencies and dissolved oxygen and SST, suggesting that these are not mediating the observed genetic divergence across the Benguela current in this species. A smaller number of outlier loci were identified for *M. paradoxus* in BayPass, but as seen for *M. capensis*, no significant associations were found neither with BayPass nor with the RDA once sequencing lane (i.e. batch effect) was used as a conditional variable. This is somewhat unexpected given that those loci were removed from the data set and the demographic history results suggesting an Atlantic–Indian Ocean divergence. Nevertheless, our results do not support strong gene–environment associations for either species.

These results seem to be in contradiction with previous findings for these species, as dbRDA analyses conducted by Henriques et al. (2016) based on microsatellite data suggested that upwelling events (high concentration of Chlorophyll a and low dissolved oxygen) might explain the differentiation patterns between northern and southern Benguela population of *M. capensis*, while Mbatha et al. (2019) demonstrated a clear impact of environmental variables in the distribution patterns of juvenile *M. paradoxus*. In addition, Nielsen et al. (2020) using the same genomic methodology, and a smaller number of SNPs, detected significant associations between SST and SSS for three inshore invertebrates across the South African coastline. It is thus possible that our data sets did not contain enough geographical resolution to detect significant associations between genotypes and environmental conditions.

4.5 | Implications for conservation and sustainable management of the southern African hakes

The assessment of stock structure, where a stock represents a demographically cohesive unit that can be exploited (Benestan, 2019), is the central tenet of fisheries management for sustainable conservation of resources. Molecular data, with its ability to identify demographically independent populations, has provided vital information for stock management and conservation for several commercially important species. The traditional example is the ‘real-time’ genomic monitoring of salmonids in the North America for allocating fishing quotas (Waples et al., 2008). This approach has recently been implemented in Norway to manage the very abundant North East Arctic cod and the vulnerable Norwegian Coastal cod (*Gadus morhua*) stocks—where genotyping of landed samples allowed to establish fishing quotas that are actively being used in the management of this mixed-stock fishery (Dahle et al., 2018). The description of previously unknown winter and spring spawning populations in the Atlantic cod in the Gulf of Maine (Clucas, Kerr, et al., 2019; Clucas, Lou, et al., 2019), which do not match current fishing regulations, as well as the physical mixing of two genetically divergent Atlantic cod populations in the Baltic Sea indicating the presence of a mixed-fishery (Hemmer-Hansen

et al., 2019), demonstrates that molecular data can successfully lead to improvements of current fisheries management strategies.

It is well known that harvesting mixed populations as a single unit can lead to depletion of the least abundant (more vulnerable) population (Ovenden et al., 2015), and also result in losses of genomic diversity (Pinsky & Palumbi, 2014) and differentiation (Gandra et al., 2020). Prior to the 1970s, the southern African hakes were initially commercially exploited as a single unit throughout their distribution range. The resources are now managed separately in South African and Namibian waters. Within South Africa, the hake resources have been managed primarily with a species-combined total allowable catch (TAC) restriction, calculated using an operational management plan (OMP) approach since 1991 (Rademeyer et al., 2008). As a result of the substantial overlap in distribution and the difficulty of distinguishing between the two hake species, species-specific catch and effort data are not available from the commercial fishery, and the two species were initially assessed and managed as a single resource. However, algorithms to apportion the commercial hake between the two species were developed during 2005 using research survey data, enabling the development of species-disaggregated assessment models. In Namibia, the two hake species are still treated as a single stock, and the assessment procedure includes information about biomass supporting the maximum sustainable yield (BSMY), the replacement yield and the catch corresponding to the harvest control rule (HCR).

Our findings suggest that *M. capensis* is composed of one population in Namibia, but two populations within South African jurisdictional waters, and thus point to a possible mixed-stock scenario in this region, which should be addressed with new fishing regulations. If the species is indeed composed of two populations that are mixed in their distribution, but have different demographic histories, then harvesting them as a single unit may well contribute towards maintaining low levels of genomic diversity, as observed in this study. Genomic diversity (both neutral and adaptive) is fundamental for the long-term persistence of species, particularly in a climate change context (Nielsen et al., 2021). Therefore, accurate fishing management policies are fundamental to ensure the conservation of *M. capensis*. Further studies should be conducted using individual-based sequencing approaches to assess the geographical (and possibly temporal) boundaries of these populations, taking into consideration possible migration patterns, in order to understand if they might be geographically based (west vs. south coast) or constitute a true mixed stock.

Similarly, the reconstruction of demographic history in *M. paradoxus* revealed the presence of two highly connected populations, with a break detected across Cape Point, as first reported for mtDNA in von der Heyden et al. (2007). The Atlantic *M. paradoxus* putative stock appears to extend from Cape Town, South Africa, to northern Namibia, suggesting a transboundary stock. Within South African waters, the genomic break is already considered to a certain extent in the assessment of the species (west and south coast-specific input data) and hence in subsequent management measures, but further research on the extent of mixing between the various populations/stocks, followed by an evaluation of the impacts of this

on stock dynamics and exploitation is required to establish whether or not the current management strategies need to be reconsidered. In either case, the current country-wide OPM and TAC might need to be revised to reflect the presence of two previously undescribed populations of southern hakes in South Africa.

AUTHORS' CONTRIBUTION

RH, SvdH, HONN and CAM designed the study; SF, ALM, ESN and RH analysed the data and prepared the results; RH performed the DNA extractions and quantification; DD, JK, PK obtained the samples; DD, JK, PK and ML provided information regarding fisheries management regulations; RH and SF led the manuscript writing, and all authors contributed to the writing of the manuscript.

ACKNOWLEDGEMENTS

We would like to thank the survey personnel at the Ministry of Fisheries and Marine Resources (NatMIRC—Namibia), the Department of Forest, Fisheries and Environment (DFFE—South Africa), the Benguela Current Commission and the Nansen programme for sample collection, during the Ecofish project (EuropeAid grant 2010/222387). This work was funded by a National Research Foundation South Africa—Namibia Bilateral grant (Grant No. 105949).

CONFLICT OF INTEREST STATEMENT

There are no conflict of interest to declare.

DATA AVAILABILITY STATEMENT

The code used in this manuscript is supplied as Data S1. Genomic data (raw reads) as well as filtered sync files are available in DRYAD (<https://doi.org/10.5061/dryad.sn02v6x8n>).

BENEFIT-SHARING STATEMENT

Benefits generated: This work results from a long-standing collaboration between scientists of South Africa and Namibia, where all collaborators are included as co-authors, and the results of the research are shared with local communities, government officials and the broader scientific community. The samples were obtained from the Department of Forestry, Fisheries and Environment in South Africa, and the Ministry of Fisheries in Namibia, in compliance with local regulations. Our group has a long-standing goal of local capacity development and training in the southern African region.

ORCID

Sophie von der Heyden  <https://orcid.org/0000-0001-9166-976X>

Alan Le Moan  <https://orcid.org/0000-0002-9124-6844>

Erica S. Nielsen  <https://orcid.org/0000-0002-5439-571X>

Deon Durholtz  <https://orcid.org/0000-0002-3234-9798>

Paulus Kainge  <https://orcid.org/0000-0003-4498-9845>

Johannes N. Kathena  <https://orcid.org/0000-0002-1863-8673>

Hilkka O. N. Ndjaula  <https://orcid.org/0000-0003-2090-7307>

Conrad A. Matthee  <https://orcid.org/0000-0002-6836-069X>

Romina Henriques  <https://orcid.org/0000-0002-6544-5532>

REFERENCES

- Andre, C., Larsson, L. C., Laikre, L., Bekkevold, D., Brigham, J., Carvalho, G. R., Dahlgren, T. G., Hutchinson, W. F., Mariani, S., Mudde, K., Ruzzante, D. E., & Ryman, N. (2011). Detecting population structure in a high gene-flow species, Atlantic herring (*Clupea harengus*): Direct, simultaneous evaluation of neutral vs putatively selected loci. *Heredity*, 106(2), 270–280. <https://doi.org/10.1038/hdy.2010.71>
- Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Assis, J., Tyberghein, L., Bosh, S., Verbruggen, H., Serrão, E. A., & De Clerck, O. (2017). Bio-ORACLE v2.0: Extending marine data layers for bioclimatic modelling. *Global Ecology and Biogeography*, 27(3), 277–284.
- Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics*. <https://CRAN.R-project.org/package=gridExtra>
- Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: A C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics*, 27(12), 1691–1692. <https://doi.org/10.1093/bioinformatics/btr174>
- Barry, P., Broquet, T., & Gagnaire, P. A. (2021). Age-specific survivorship and fecundity shape genetic diversity in marine fishes. *Evolution Letters*, 6(1), 46–62. <https://doi.org/10.1002/evl3.265>
- Beacham, T. D., Wallace, C., Macconnachie, C., Jonsen, K., McIntosh, B., Candy, J. R., & Withler, R. E. (2018). Population and individual identification of Chinook salmon in British Columbia through parentage-based tagging and genetic stock identification with single nucleotide polymorphisms. *Canadian Journal of Fisheries and Aquatic Sciences*, 75(7), 1096–1105. <https://doi.org/10.1139/cjfas-2017-0168>
- Benestan, L. (2019). Population genomics applied to fishery management and conservation. In M. F. Oleksiak & O. P. Rajora (Eds.), *Population genomics: Marine organisms*. Springer Nature. https://doi.org/10.1007/13836_2019_66
- Benestan, L., Gosselin, T., Perrier, C., Sainte-Marie, B., Rochette, R., & Bernatchez, L. (2015). RAD genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine species, the American lobster (*Homarus americanus*). *Molecular Ecology*, 24(13), 3299–3315. <https://doi.org/10.1111/mec.13245>
- Bosch, S., Tyberghein, L., Clerck, O. D., Fernandez, S., & Schepers, L. (2022). *sdmPredictors: Species distribution modelling predictor datasets*. <https://cran.r-project.org/web/packages/sdmpredictors/index.html>
- Campo, D., Machado-Schiaffino, G., Perez, J., & Garcia-Vazquez, E. (2007). Phylogeny of the genus *Merluccius* based on mitochondrial and nuclear genes. *Gene*, 406(1–2), 171–179. <https://doi.org/10.1016/j.gene.2007.09.008>
- Castillo, A. G. F., Alvarez, P., & Garcia-Vazquez, E. (2005). Population structure of *Merluccius merluccius* along the Iberian Peninsula coast. *ICES Journal of Marine Science*, 62, 1699–1704. <https://doi.org/10.1016/j.icesjms.2005.06.001>
- Cheng, J., Karambelkar, B., & Xie, Y. (2021). *Leaflet: Create interactive web maps with the JavaScript 'Leaflet' library*. <https://cran.r-project.org/web/packages/leaflet/index.html>
- Clucas, G. V., Kerr, L. A., Cadrin, S. X., Zemeckis, D. R., Sherwood, G. D., Goethel, D., Whitener, Z., & Kovach, A. I. (2019). Adaptive genetic variation underlies biocomplexity of Atlantic cod in the Gulf of Maine and on Georges Bank. *PLoS One*, 14(5), e0216992. <https://doi.org/10.1371/journal.pone.0216992>
- Clucas, G. V., Lou, R. N., Therkildsen, N. O., & Kovach, A. I. (2019). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolutionary Applications*, 12(10), 1971–1987. <https://doi.org/10.1111/eva.12861>
- Dahle, G., Johansen, T., Westgaard, J. I., Aglen, A., & Glover, K. A. (2018). Genetic management of mixed-stock fisheries “real-time”: The case of the largest remaining cod fishery operating in the Atlantic in 2007–2017. *Fisheries Research*, 205, 77–85. <https://doi.org/10.1016/j.fishres.2018.04.006>
- Dalongeville, A., Nielsen, E. S., Teske, P. R., & von der Heyden, S. (2022). Comparative phylogeography in a marine biodiversity hotspot provides novel insights into evolutionary processes across the Atlantic-Indian Ocean transition. *Diversity and Distributions*, 28, 2622–2636. <https://doi.org/10.1111/ddi.13534>
- DFFE. (2020). *Status of South African marine fishery resources*.
- Diopere, E., Vandamme, S. G., Hablutzel, P. I., Cariani, A., Houddt, J. V., Rijnsdorp, A., Tinti, F., Volckaert, F. A. M., Maes, G. E., & FishPopTrace, C. (2018). Seascape genetics of a flatfish reveals local selection under high levels of gene flow. *ICES Journal of Marine Science*, 75(2), 675–689. <https://doi.org/10.1093/icesjms/fsx160>
- Dorant, Y., Benestan, L., Rougemont, Q., Normandeau, E., Boyle, B., Rochette, R., & Bernatchez, L. (2019). Comparing Pool-seq, rapture, and GBS genotyping for inferring weak population structure: The American lobster (*Homarus americanus*) as a case study. *Ecology and Evolution*, 9(11), 6606–6623. <https://doi.org/10.1002/ece3.5240>
- Durholtz, M. D., Singh, L., Fairweather, T. P., Leslie, R., Van Der Lingen, C. D., Bross, C. A. R., Hutchings, L., Rademeyer, R. A., Butterworth, D. S., & Payne, A. I. L. (2015). Fisheries, ecology and markets of south African hake. In H. Arancibia (Ed.), *Hakes: Biology and exploitation* (pp. 38–69). John Wiley and Sons.
- FAO. (2022). *The state of world fisheries and aquaculture 2022. Towards blue transformation*. FAO. <https://doi.org/10.4060/cc0461en>
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, 27, 2215–2233. <https://doi.org/10.1111/mec.14584>
- Gagnaire, P. A. (2020). Comparative genomics approach to evolutionary process connectivity. *Evolutionary Applications*, 13(6), 1320–1334. <https://doi.org/10.1111/eva.12978>
- Gagnaire, P.-A., Broquet, T., Aurelle, D., Viard, F., Souissi, A., Bonhomme, F., Arnaud-Haond, S., & Bierne, N. (2015). Using neutral, selected, and hitchhiker loci to assess connectivity of marine populations in the genomic era. *Evolutionary Applications*, 8(8), 769–786. <https://doi.org/10.1111/eva.12288>
- Gandra, M., Assis, J., Martins, M. R., & Abecasis, D. (2020). Reduced global genetic differentiation of exploited marine fish species. *Molecular Biology and Evolution*, 38(4), 1402–1412. <https://doi.org/10.1093/molbev/msaa299>
- Gattepaille, L. M., Jakobsson, M., & Blum, M. G. B. (2013). Inferring population size changes with sequence and SNP data: Lessons from human bottlenecks. *Heredity*, 110, 409–419. <https://doi.org/10.1038/hdy.2012.120>
- Gautier, M. (2015). Genome-wide scan for adaptive differentiation and association analysis with population-specific covariables. *Genetics*, 201(4), 1555–1579.
- Grant, W. S., Becker, I. I., & Leslie, R. W. (1988). Evolutionary divergence between sympatric species of southern African hakes, *Merluccius capensis* and *Merluccius paradoxus*. 1. Electrophoretic analysis of proteins. *Heredity*, 61, 13–20. <https://doi.org/10.1038/hdy.1988.86>
- Hauser, L., & Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, 9(4), 333–362. <https://doi.org/10.1111/j.1467-2979.2008.00299.x>
- Hemmer-Hansen, J., Hussy, K., Baktoft, H., Huwer, B., Bekkevold, D., Haslob, H., Herrmann, J. P., Hinrichsen, H. H., Krumme, U., Mosegaard, H., Nielsen, E. E., Reusch, T. B. H., Storr-Paulsen, M., Velasco, A., von Dewitz, B., Dierking, J., & Eero, M. (2019). Genetic analyses reveal complex dynamics within a marine fish management area. *Evolutionary Applications*, 12(4), 830–844. <https://doi.org/10.1111/eva.12760>

- Henriques, R., von der Heyden, S., Lipinski, M. R., Toit, N. D., Kainge, P., Bloomer, P., & Matthee, C. A. (2016). Spatio-temporal genetic structure and the effects of long-term fishing in two partially sympatric offshore demersal fishes. *Molecular Ecology*, 25(23), 5843–5861. <https://doi.org/10.1111/mec.13890>
- Hivert, V., Leblois, R., Petit, E. J., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from Pool-seq data. *Genetics*, 210(1), 315–330. <https://doi.org/10.1534/genetics.118.300900>
- Hoey, J. A., & Pinsky, M. L. (2018). Genomic signatures of environmental selection despite near-panmixia in summer flounder. *Evolutionary Applications*, 11, 1732–1747. <https://doi.org/10.1111/eva.12676>
- Hunter, M. E., Hoban, S. M., Bruford, M. W., Segelbacher, G., & Bernatchez, L. (2018). Next-generation conservation genetics and biodiversity monitoring. *Evolutionary Applications*, 11(7), 1029–1034. <https://doi.org/10.1111/eva.12661>
- Hutchings, L., Lingen, C. D. V. D., Shannon, L. J., Crawford, R. J. M., Verheye, H. M. S., Bartholomae, C. H., Plas, A. K. V. D., Louw, D., Kreiner, A., Ostrowski, M., Fidel, Q., Barlow, R. G., Lamont, T., Coetzee, J., Shillington, F., Veitch, J., Currie, J. C., & Monteiro, P. M. S. (2009). The Benguela current: An ecosystem of four components. *Progress in Oceanography*, 83(1–4), 15–32. <https://doi.org/10.1016/j.pocean.2009.07.046>
- Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H., & Owens, G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology*, 25(18), 4488–4507. <https://doi.org/10.1111/mec.13792>
- Jansen, T., Kainge, P., Singh, L., Wilhelm, M., Durholtz, D., Stromme, T., Kathena, J., & Erasmus, V. (2015). Spawning patterns of shallow-water hake (*Merluccius capensis*) and deep-water hake (*M. paradoxus*) in the Benguela current large marine ecosystem inferred from gonadosomatic indices. *Fisheries Research*, 172, 168–180. <https://doi.org/10.1016/j.fishres.2015.07.009>
- Jimenez-Mena, B., Le Moan, A., Christensen, A., van Deurs, M., Mosegaard, H., Hemmer-Hansen, J., & Bekkevold, D. (2020). Weak genetic structure despite strong genomic signal in lesser sandeel in the North Sea. *Evolutionary Applications*, 13(2), 376–387. <https://doi.org/10.1111/eva.12875>
- Jouganous, J., Long, W., Ragsdale, A. P., & Gravel, S. (2017). Inferring the joint demographic history of multiple populations: beyond the diffusion approximation. *Genetics*, 206(3), 1549–1567. <https://doi.org/10.1534/genetics.117.200493>
- Kapula, V. K., Ndjaula, H. O. N., Schulze, M. J., Durholtz, D., Japp, D., Singh, L., Matthee, C. A., von der Heyden, S., & Henriques, R. (2022). Genetic assessment of seasonal alongshore migration in *Merluccius capensis* in the Benguela region. *Fisheries Research*, 250, 106293. <https://doi.org/10.1016/j.fishres.2022.106293>
- Kassambara, A., & Mundt, F. (2021). *Factoextra: Extract and visualize the results of multivariate data analyses*. <https://CRAN.R-project.org/package=factoextra>
- Kathena, J., Nielsen, A., Thygesen, U. H., & Berg, C. W. (2016). Hake species (*Merluccius capensis* and *M. paradoxus*) assessment in the Benguela current large marine ecosystem. *Environment and Development*, 17(1), 193–201.
- Kingman, J. F. C. (1983). Mathematical genetics – Introductory-remarks. *Proceedings of the Royal Society B: Biological Sciences*, 219(1216), 221–222. <https://doi.org/10.1098/rspb.1983.0070>
- Knapp, I. S. S., Puritz, J., Bird, C., Whitney, J., Sudek, M., Forsman, Z., & Toonen, R. (2016). ezRAD – An accessible next-generation RAD sequencing protocol suitable for non-model organisms_v3.2. *protocols.io*, <https://doi.org/10.17504/protocols.io.e9pbh5n>
- Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlötterer, C. (2011). PoPoolation: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. *PLoS One*, 6(1), 1–9. <https://doi.org/10.1371/journal.pone.0015925>
- Kofler, R., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, 27(24), 3435–3436. <https://doi.org/10.1093/bioinformatics/btr589>
- Lamichhaney, S., Barrio, A. M., Rafati, N., Sundstrom, G., Rubin, C.-J., Gilbert, E. R., Berglund, J., Wetterbom, A., Laikre, L., Webster, M. T., Grabherr, M., Ryman, N., & Andersson, L. (2012). Population-scale sequencing reveals genetic differentiation due to local adaptation in Atlantic herring. *Proceedings of the National Academy of Sciences of the United States of America*, 109(47), 19345–19350. <https://doi.org/10.1073/pnas.1216128109>
- Le Moan, A., Gaggiotti, O., Henriques, R., Martinez, P., Bekkevold, D., & Hemmer-Hansen, J. (2019). Beyond parallel evolution: When several species colonize the same environmental gradient. *bioRxiv*. <https://doi.org/10.1101/662569>
- Le Moan, A., Roby, C., Fraise, C., Daguin-Thiebaut, C., Bierne, N., & Viard, F. (2021). An introgression breakthrough left by an anthropogenic contact between two ascidians. *Molecular Ecology*, 30(24), 6718–6732. <https://doi.org/10.1111/mec.16189>
- Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: A package for multivariate analysis. *Journal of Statistical Software*, 25, 1–8.
- Leek, J. T., Scharpf, R. B., Bravo, H. C., Simcha, D., Langmead, B., Johnson, W. E., Geman, D., Baggerly, K., & Irizarry, R. A. (2010). Tackling the widespread and critical impact of batch effects in high-throughput data. *Nature Reviews. Genetics*, 11(10), 733–739. <https://doi.org/10.1038/nrg2825>
- Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129(2), 271–280. <https://doi.org/10.1007/s0044200100716>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*, 1303.3997.
- Li, H. (2021). *Samtools*. <http://www.htslib.org/doc/samtools-mpile-up.html>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & Subgroup, G. P. D. P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lou, R. N., & Therikildsen, N. O. (2021). Batch effects in population genomic studies with low-coverage whole genome sequencing data: Causes, detection and mitigation. *Molecular Ecology Resources*, 22(5), 1675–2126. <https://doi.org/10.1111/1755-0998.13559>
- Magoč, T., & Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics*, 27(21), 2957–2963. <https://doi.org/10.1093/bioinformatics/btr507>
- Manuzzi, A., Jimenez-Mena, B., Henriques, R., Holmes, B. J., Pepperell, J., Edson, J., Bennett, M. B., Huvencers, C., Ovenden, J. R., & Nielsen, E. E. (2022). Retrospective genomics highlights changes in genetic composition of tiger sharks (*Galeocerdo cuvier*) and potential loss of a South-Eastern Australia population. *Scientific Reports*, 12(1), 6582. <https://doi.org/10.1038/s41598-022-10529-w>
- Martinez, A. S., Willoughby, J. R., & Christie, M. R. (2018). Genetic diversity in fishes is influenced by habitat type and life history variation. *Ecology and Evolution*, 8(23), 12022–12031. <https://doi.org/10.1002/ece3.4661>
- Mbatha, F. L., Yemane, D., Ostrowski, M., Moloney, C. L., & Lipinski, M. R. (2019). Oxygen and temperature influence the distribution of Deepwater cape hake *Merluccius paradoxus* in the southern Benguela: A GAM analysis of a 10-year time-series. *African Journal of Marine Science*, 41(4), 413–427.
- Milano, I., Babucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espineira, M., Fiorentino, F., Garofalo, G., Geffen, A. J., Hansen, J. H., Helyar, S. J., Nielsen, E. E., Ogden, R.,

- Patarnello, T., Stagioni, M., FishPopTrace Consortium, Tinti, F., & Bargelloni, L. (2014). Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, 23(1), 118–135. <https://doi.org/10.1111/mec.12568>
- Momigliano, P., Florin, A. B., & Merila, J. (2021). Biases in demographic modeling affect our understanding of recent divergence. *Molecular Biology and Evolution*, 38(7), 2967–2985. <https://doi.org/10.1093/molbev/msab047>
- Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37, 191–203. <https://doi.org/10.1111/j.1600-0587.2013.00205.x>
- Nielsen, E. S., Hanson, J. O., Carvalho, S. B., Beger, M., Henriques, R., Kershaw, F., & von der Heyden, S. (2023). Molecular ecology meets systematic conservation planning. *Trends in Ecology & Evolution*, 38(2), 143–155. <https://doi.org/10.1016/j.tree.2022.09.006>
- Nielsen, E. S., Henriques, R., Beger, M., & Heyden, S. V. D. (2021). Distinct inter- and intra-specific vulnerability of coastal species to global change. *Global Change Biology*, 27, 3415–3431.
- Nielsen, E. S., Henriques, R., Beger, M., Toonen, R. J., & von der Heyden, S. (2020). Multi-model seascape genomics identifies distinct environmental drivers of selection among sympatric marine species. *BMC Evolutionary Biology*, 20(1), 121. <https://doi.org/10.1186/s12862-020-01679-4>
- Nielsen, E. S., Henriques, R., Toonen, R. J., Knapp, I. S. S., Guo, B., & von der Heyden, S. (2018). Complex signatures of genomic variation of two non-model marine species in a homogeneous environment. *BMC Genomics*, 19(1), 347. <https://doi.org/10.1186/s12864-018-4721-y>
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., Caceres, M. D., Durand, S., ... Weedon, J. (2020). Vegan: Community ecology package. <https://cran.r-project.org/web/packages/vegan/index.html>
- Ovenden, J. R., Berry, D. J., Welch Buckworth, R. C., & Dichmont, C. M. (2015). Ocean's eleven: A critical evaluation of the role of population, evolutionary and molecular genetics in the management of wild fisheries. *Fish and Fisheries*, 16(1), 125–159. <https://doi.org/10.1111/faf.12052>
- Pinsky, M. L., & Palumbi, S. R. (2014). Meta-analysis reveals lower genetic diversity in overfished populations. *Molecular Ecology*, 23(1), 29–39. <https://doi.org/10.1111/mec.12509>
- Portick, D. M., Leache, A. D., Rivera, D., Blackburn, D. C., Rodel, M. O., Barej, M. F., Burger, M., Hirschfeld, M., Rödel, M. O., Blackburn, D. C., & Fujita, M. K. (2017). Evaluating mechanisms of diversification in a Guineo-Congolian forest frog using demographic model selection. *Molecular Ecology*, 26, 5245–5263. <https://doi.org/10.1111/mec.14266>
- Punt, A. E., & Japp, D. W. (1994). Stock assessment of the kingklip *Genypterus capensis* off South-Africa. *South African Journal of Marine Science-Suid-Afrikaanse Tydskrif Vir Seewetenskap*, 14, 133–149. [://WOS:A1994PR32100013](https://doi.org/10.1111/mec.14266).
- R Core Team. (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rademeyer, R. A., Butterworth, D. S., & Plaganyi, E. (2008). A history of recent bases for management and the development of a species-combined operational management procedure for the south African hake resource. *African Journal of Marine Science*, 30(2), 291–310.
- Reeve, J., Li, Q. S., Lindtke, D., & Yeaman, S. (2022). Comparing genome scans among species of the stickleback order reveals three different patterns of genetic diversity. *Ecology and Evolution*, 12(1), e8502. <https://doi.org/10.1002/ece3.8502>
- Riquet, F., Liautard-Haag, C., Woodall, L., Bouza, C., Louisy, P., Hamer, B., Otero-Ferrer, F., Aublanc, P., Beduneau, V., Briard, O., El Ayari, T., 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>
- Ross-Gillespie, A. (2016). A comparison between the hake cannibalism and inter-species predation models presented in Bergh et al. (2016) and Ross-Gillespie (2016). DAFF Branch Fisheries document. FISHERIES/2016/MAR/SWG-DEM/08: 2pp.
- Schulze, M. J., von der Heyden, S., Japp, D., Singh, L., Durholtz, D., Kapula, V. K., Ndjaula, H. O. N., & Henriques, R. (2020). Supporting fisheries management with genomic tools: A case study of kingklip (*Genypterus capensis*) off Southern Africa. *Frontiers in Marine Science*, 7, 557146. <https://doi.org/10.3389/fmars.2020.557146>
- Strømme, T., Lipinski, M. R., & Kainge, P. (2016). Life cycle of hake and likely management implications. *Reviews in Fish Biology and Fisheries*, 26(2), 235–248. <https://doi.org/10.1007/s1160-015-9415-9>
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595. <https://doi.org/10.1093/genetics/123.3.585>
- Teske, P. R., Sandoval-Castillo, J., Golla, T. R., Emami-Khoyi, A., Tine, M., von der Heyden, S., & Beheregaray, L. B. (2019). Thermal selection as a driver of marine ecological speciation. *Proceedings of the National Academy of Sciences*, 286(1896), 1–7. <https://doi.org/10.1098/rspb.2018.2023>
- Teske, P. R., von der Heyden, S., McQuaid, C. D., & Barker, N. P. (2011). A review of marine phylogeography in southern Africa. *South African Journal of Science*, 107(5–6), 43–53. <https://doi.org/10.4102/sajs.v107i5/6.514>
- Therkildsen, N. O., Hemmer-Hansen, J., Hedeholm, R. B., Wisz, M. S., Pampoulie, C., Meldrup Bonanomi, S., Retzel, A., Olsen, S. M., & Nielsen, E. E. (2013). Spatiotemporal SNP analysis reveals pronounced biocomplexity at the northern range margin of Atlantic cod *Gadus morhua*. *Evolutionary Applications*, 6(4), 690–705. <https://doi.org/10.1111/eva.12055>
- Thioulouse, J., Dray, S., Dufour, A., Siberchicot, A., Jombart, T., & Pavoine, S. (2018). *Multivariate analysis of ecological data with ade4* (1st ed.). Springer.
- Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013). ezRAD: A simplified method for genomic genotyping in non-model organisms. *PeerJ*, 1, e203. <https://doi.org/10.7717/peerj.203>
- van Doren, B. M., Campagna, L., Helm, B., Illera, J. C., Lovette, I. J., & Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology*, 26(15), 3982–3997. <https://doi.org/10.1111/mec.14083>
- von der Heyden, S., Lipinski, M. R., & Matthee, C. A. (2007). Mitochondrial DNA analyses of the cape hakes reveal an expanding, panmictic population for *Merluccius capensis* and population structuring for mature fish in *Merluccius paradoxus*. *Molecular Phylogenetics and Evolution*, 42(2), 517–527. <https://doi.org/10.1016/j.ympev.2006.08.004>
- von der Heyden, S., Lipinski, M. R., & Matthee, C. A. (2010). Remarkably low mtDNA control region diversity in an abundant demersal fish. *Molecular Phylogenetics and Evolution*, 55(3), 1193–1188. <https://doi.org/10.1016/j.ympev.2009.09.018>
- Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular*

- Ecology*, 15(6), 1419–1439. <https://doi.org/10.1111/j.1365-294X.2006.02890.x>
- Waples, R. S., Punt, A. E., & Cope, J. M. (2008). Integrating genetic data into management of marine resources: How can we do it better? *Fish and Fisheries*, 9(4), 423–449. <https://doi.org/10.1111/j.1467-2979.2008.00303.x>
- White, T., Stamford, J., & Hoelzel, A. (2010). Local selection and population structure in a deep-sea fish, the roundnose grenadier (*Coryphaenoides rupes*). *Molecular Ecology*, 19, 216–226.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. <https://ggplot2.tidyverse.org>
- Wickham, H., François, R., Henry, L., & Müller, K. (2021). *Dplyr: A grammar of data manipulation*. <https://CRAN.R-project.org/package=dplyr>
- Winnepenninckx, B., Backeljau, T., & Dewachter, R. (1993). Extraction of high-molecular-weight DNA from mollusks. *Trends in Genetics*, 9(12), 407.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16(2), 97–159. <https://doi.org/10.1093/genetics/16.2.97>

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: Forde, S., von der Heyden, S., Le Moan, A., Nielsen, E. S., Durholtz, D., Kainge, P., Kathena, J. N., Lipinski, M. R., Ndjaula, H. O. N., Matthee, C. A., & Henriques, R. (2023). Management and conservation implications of cryptic population substructure for two commercially exploited fishes (*Merluccius* spp.) in southern Africa. *Molecular Ecology Resources*, 00, 1–17. <https://doi.org/10.1111/1755-0998.13820>