Supporting the Blue Economy: Using Genomic Tools for Assessing Population Connectivity and Evolutionary History in the Cape Hakes

Romina Henriques, Veronica Kaleinasho, Melissa Schulze, Hilkka Ndjaula, Sophie von der Heyden, Conrad A. Matthee

Stellenbosch University and University of Namibia

2019

This report is confidential and cannot be used without the explicit consent of its authors

Disclaimer: The findings of this report have not been peer-reviewed, and are awaiting publication
The Cape Hakes

Merluccius capensis

Merluccius capensis, also known as shallow-water hake, is a hake species endemic to the Benguela region. Currently, two spawning grounds have been described for the species, one located in Namibia (northern Benguela) and the other located in South Africa (southern Benguela) (Jansen et al., 2015). Previous genetic work using nine microsatellites revealed the presence of two genetic populations, exhibiting limited gene flow (Henriques et al., 2016). The described populations appeared to coincide clearly with the physical and geopolitical boundaries in the region, and shallow-water hake was deemed to have a northern and southern population along the west coast of southern Africa (Henriques et al., 2016). No obvious geographic pattern was found when analysing the control region of the mitochondrial DNA, suggesting that the observed break of gene flow resulted from a recent split of the ancestral population (von der Heyden, Lipinski, & Matthee, 2007). The observed findings is further supported by previous work done using allozymes (Grant & Leslie, 2001).

Although, based on microsatellite loci, the genetic boundary between these two populations (northern and southern, respectively) remained stable throughout the years, the physical location of the boundary fluctuated temporarily. In 2012 it coincided roughly with the political Namibia-South Africa boundary, in the Orange river region (29°S), while in 2014 individuals of the northern stock were found further south, off Saint Helena Bay (32°S). Seascape analyses suggested a relationship between the presence of northern individuals in the south and environmental features, specifically upwelling events (Figure 1 - Henriques et al., 2016). Furthermore, southern individuals were rarely found north of 29°S, in all three years sampled.

Finally, the observed genetic differentiation was linked to a few microsatellite markers with elevated genetic differentiation levels (FST). Given that microsatellite are neutral markers, not under the direct
influence of natural selection, it is likely that these markers were linked to regions in the genome under selection (genetic hitchhiking). In addition, microsatellite loci suggested that a third population might exist along the South African south coast, but did not have enough statistical power to identify it properly (Henriques et al., 2016).

**Merluccius paradoxus**

*Merluccius paradoxus*, known as deep-water hake, is an endemic species of the Benguela region. The species exhibits a parasympatric distribution with *M. capensis*, co-occurring in the same region, but with a depth-related ontogenic distribution (Strømme, Lipinski, & Kainge, 2016). Younger *M. paradoxus* are found in shallower waters, and as individuals mature, they migrate deeper and northwards (Strømme et al., 2016). At shallower depths, *M. paradoxus* juveniles are found together with *M. capensis* adults, leading to high levels of predation of the younger age-classes of the former species (Strømme et al., 2016).

Biological surveys suggest the presence of one major spawning ground in the Benguela region, off the West coast of South Africa, and no spawning has been documented for Namibian waters (Strømme et al., 2016). However, there have been reports of ripe and running individuals off the south coast of South Africa (between Mossel Bay and Port Elizabeth), which might suggest the possibility of a second spawning unit within South African waters.

Previous genetic work using microsatellite loci corroborates the one-population hypothesis, as no significant differentiation was reported for three consecutive years (Figure 2 - Henriques et al., 2016). These results also corroborated previous findings using allozymes, where only one population was reported (Grant & Leslie, 2001). However, control region sequences of the mitochondrial DNA (CR mtDNA) suggested the presence of two populations, based on differences in haplotype frequencies (von der Heyden et al., 2007). These results were only reported for the adults, as juveniles showed evidence of panmixia (von der Heyden et al., 2007).

**Figure 2:** Main findings of Henriques et al. 2016 for *M. paradoxus*. From left to right: Principal Component Analysis based on nine microsatellite markers, and sampling map with the one population super-imposed.
A similar pattern of mtDNA-based differentiation was observed in the Henriques et al. (2016) study, but only for samples collected in 2013, and across the Cape Point boundary region of South Africa, not across the Orange River as suggested by von der Heyden et al. (2007).

At the time, Henriques et al. (2016) explained the observed changes in haplotype frequencies as likely to result from a cohort effect – genetic chaotic patchiness (GCP Johnson & Black, 1982). Small and transient genetic differentiation levels have been observed in other fishes, including *M. merluccius* (Lundy, Rico, & Hewitt, 2000), and appear to have resulted from fluctuations in the survival rates of different cohorts among years. As reported in Henriques et al. (2016):

“In order to accurately distinguish between the drivers of GCP (Selkoe et al. 2006; Hogan et al. 2010), it is necessary to evaluate genetic differentiation among regions and also among larvae, juveniles and adults (Hogan et al. 2010). Although all the individuals analysed for Henriques et al. (2016) were adults, von der Heyden et al. (2007) showed that differentiation was only observed among mature *M. paradoxus*. This suggests that adult reproductive sweepstakes and fluctuations of the origin of the larval pool can be rejected as possible explanations. Therefore, the observed GCP in *M. paradoxus* is likely to reflect selective pressures at either pre- or postsettlement phases, resulting in differential survival of recruits, which is supported by the occurrence of years of poor recruitment (DAFF 2014).”

However, it is still possible that the microsatellite loci used in the previous study did not have enough statistical power to detect subtle levels of differentiation in *M. paradoxus*. Although this is not likely, as the same loci were successfully used in *M. capensis*, it remains a possibility that needed to be addressed.

Taking all of the above into consideration, the current project used a combination of microsatellite loci and high-throughput genomic markers (Single Nucleotide Polymorphisms - SNPs) to:

- Investigate seasonal patterns of transboundary migration in *Merluccius capensis*, by sampling the Benguela region in both summer and winter of 2017;
- Investigate genome-wide levels of genetic diversity and divergence in the Cape hakes using a Pool-Seq Reduced Representation sequencing approach.
Seasonal Patterns of Transboundary Migration of *Merluccius capensis* across the Benguela Region

Master Thesis of Veronica Kaleinasho Kapula – University of Namibia and Stellenbosch University

A total of 503 fishes were sampled, ranging from northern Namibia to Cape Town, South Africa. Fishes were collected during the summer (January-March) and winter months (July-August) in 2017, either from scientific surveys from the Department of Agriculture, Forestry and Fisheries (DAFF – South Africa) or from commercial fishing operations (Tunacor Fisheries, Namibia, and CapFish, South Africa). Distribution of samples per latitude and season can be seen in Table 1.

**Table 1:** Sampling sites distribution across the Benguela region per season, and number of individuals sampled

<table>
<thead>
<tr>
<th></th>
<th>North (NN)</th>
<th>Central (CN)</th>
<th>South (SN)</th>
<th>Orange River (NWC)</th>
<th>Central West Coast (CWC)</th>
<th>Southern West Coast (SWC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Namibia</td>
<td>50</td>
<td>50</td>
<td>19</td>
<td>39</td>
<td>43</td>
<td>69</td>
</tr>
<tr>
<td>South Africa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total DNA was extracted using a standard chlorophorm:isopropanol protocol (Winnepenninckx, Backeljau, & Dewachter, 1993), and eight microsatellite loci and the protocol employed in Henriques et al. (2016) were used. Locus MP8450 was not used in this study as it consistently failed to amplify. Microsatellite fragments were genotyped on an ABI-377 sequencer (CAF, Stellenbosch University), using a LIZ500® as an internal size marker. Microsatellite alleles were scored based on size (bp).

Obtained microsatellite loci were checked for amplification errors (large allele drop out and stuttering) in MICROCHECKER (van Oosterhout, Weetman, & Hutchinson, 2006), while the presence of null alleles was estimated in FreeNA (Chapuis & Estoup, 2007). Deviations to the expectation of outcrossing and linkage disequilibrium were tested in Genepop 1.2 (Raymond & Rousset, 1995). Estimates of genetic diversity (expected and observed heterozygosity, allelic richness and number of alleles) were estimated in Arlequin 5.0 (Excoffier, Laval, & Schneider, 2005), while population differentiation levels were assessed using a pairwise F$_{ST}$ approach per sampling site and period, in FreeNA, with statistically significance assessed after 10 000 permutations. Finally, probability of individual assignment was assessed in STRUCTURE (Pritchard, Stephens, & Donnelly, 2000), using five independent runs, allowing...
for admixture and correlated allelic frequencies, with number of clusters (K) varying between 1 and 6, with a burnin of 250 000 MCMC, followed by 1 million MCMC steps. The most likely number of clusters was identified using the Delta K method of Evanno, Regnaut, & Goudet, (2005), in STRUCTURE Harvester (Dent & von Holdt, 2012).

Individual assignments were then used to classify individuals to the northern and southern population, per latitude and season, creating a distribution cline. Individuals were assigned to one cluster or the other, based on a probability assignment of $q = 0.75$, to account for the possibility of incomplete lineage sorting and admixture. As such, individuals with $q > 0.75$ were assigned to the northern cluster and with $q < 0.25$ were assigned to the southern cluster. Individuals with $0.25 < q < 0.75$ were considered “hybrids” between the two populations, as per Henriques et al. (2016).

Findings

As in Henriques et al. (2016), no evidence of large allele drop out, stuttering or linkage disequilibrium was found. Some loci exhibited null alleles, but all were below 25%, for both seasons. Therefore, all loci were retained for further analyses. Unfortunately, it was not possible to sample the Orange River region during the winter months, and thus all estimates for this area are limited to the summer.

Sites exhibited significant deviations to Hardy-Weinberg equilibrium due to a heterozygote deficit, with the exception of northern Namibia (Table 2). Overall genetic diversity was low, regardless of the site and period ($0.484 < H_e < 0.595$), with allelic richness varying between $7.290 < AR < 7.561$ (Table 2).

<table>
<thead>
<tr>
<th></th>
<th>Summer</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NN</td>
<td>CN</td>
</tr>
<tr>
<td>N</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>NA</td>
<td>10,000</td>
<td>10,375</td>
</tr>
<tr>
<td>$H_e$</td>
<td>0.570</td>
<td>0.544</td>
</tr>
<tr>
<td>$H_o$</td>
<td>0.567</td>
<td>0.497</td>
</tr>
<tr>
<td>$F_{IS}$</td>
<td>0.009</td>
<td>0.087</td>
</tr>
</tbody>
</table>

Genetic differentiation levels were relatively high, with global $F_{ST} = 0.160$, and significantly different from zero ($p<0.005$). As before, two clusters emerged: sites in the northern Benguela were significantly differentiated from sites in southern Benguela, but no differentiation was observed within regions (Table 3). A similar pattern was observed in the individual-based assignment analyses, with individuals collected in the northern Benguela being consistently differentiated from individuals collected in southern Benguela (Figure 3).
Table 3: Estimates of pairwise $F_{ST}$ for *M. capensis* across sampling sites and seasons: summer (below diagonal) and winter (above diagonal). Statistical significant results in bold (p<0.001). Samples locations as per Table 1

<table>
<thead>
<tr>
<th></th>
<th>NN</th>
<th>CN</th>
<th>SN</th>
<th>NWC</th>
<th>CWC</th>
<th>SWC</th>
</tr>
</thead>
<tbody>
<tr>
<td>NN</td>
<td>-</td>
<td>0.001</td>
<td>0.010</td>
<td>-</td>
<td>0.172</td>
<td>0.200</td>
</tr>
<tr>
<td>CN</td>
<td>0.001</td>
<td>-</td>
<td>0.004</td>
<td>-</td>
<td>0.170</td>
<td>0.201</td>
</tr>
<tr>
<td>SN</td>
<td>0.000</td>
<td>0.007</td>
<td>-</td>
<td>-</td>
<td>0.125</td>
<td>0.150</td>
</tr>
<tr>
<td>NWC</td>
<td>0.153</td>
<td>0.157</td>
<td>0.144</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CWC</td>
<td>0.145</td>
<td>0.150</td>
<td>0.147</td>
<td>0.009</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SWC</td>
<td>0.177</td>
<td>0.180</td>
<td>0.175</td>
<td>0.004</td>
<td>0.005</td>
<td>-</td>
</tr>
</tbody>
</table>

Figure 3: Assignment probability results for *M. capensis* based on cluster analyses. Summer: dark grey – northern population, light grey – southern population. Winter: light grey – northern population, dark grey – southern population. Samples locations as per Table 1
Assessment of the distribution of the two identified populations per latitude per season revealed contrasting patterns (Figure 4). In the summer, 5% of the northern population was found in South Africa as far south as Saint Helen Bay (32°S), while no southern individuals were found in Namibia (Figure 4). In winter, northern hakes were still found in South African waters (6% - 32°S), but a small proportion (6%) of southern hakes was detected in Namibia all the way to 27°S (Figure 4). It is possible to observe that at the extreme latitudes (northern Namibia and southern West Coast) in the summer/winter, the populations are composed by 100% of the regional individuals (no migrants identified).

**Figure 4:** Distribution clines of population composition across sampling sites, in the summer and winter

**Conclusions**

As before, two populations were observed across the Benguela region for *M. capensis*, roughly coinciding with the geo-political boundaries: a northern population appears to be mainly confined to the northern Benguela region, while a southern population appears to be confined to the southern Benguela region.
Migration seems to occur throughout the region, with a small percentage of northern individuals found throughout the year as far south as Saint Helena Bay, in South Africa (32°S). Interestingly, in the winter, a small percentage of southern individuals was found in Namibian waters, as far north as 27°S. As previously suggested, the Orange River area appears to be a mixing zone between the two populations, although the absence of samples in this region in the winter precludes drawing strong conclusions on the fine-scale movements throughout the physical geo-political border.

Despite the observed migration patterns, very few “hybrids” were found, suggesting a significant break in gene flow between the two *M. capensis* populations, and pointing to demographic independence of these populations. Interestingly, “hybrids” were mostly found in the southern latitudes.

These results suggest that *M. capensis* exhibits differences in seasonal transboundary movements. This might have implications for fisheries management, as it appears that a small proportion of northern individuals can be found off South African waters throughout the year, while the reverse is only true in the colder winter months.

We recommend that an annual monitoring programme should be established to track the movement of the two populations throughout the year, as a way of assessing the extent of mixing between the populations. These findings further suggest that genetic markers can be used for real-time tracking of, otherwise, cryptic stocks in order to aid management decisions.
Genome-Wide Patterns of Population Connectivity in *Merluccius capensis* across the Benguela Region

Postdoctoral project of Romina Henriques – Stellenbosch University

Microsatellite markers might not have enough statistical power to detect subtle breaks in gene flow, especially for marine fish, where historically large effective population sizes (\(N_E\)) tend to lead to high genetic diversity. Therefore, sequencing approaches that generate thousands of SNPs can greatly aid in fisheries management, by detecting previously overlooked fine-scale population structure (Gagnaire et al., 2015).

Although sequencing costs have decreased in the last decade, it remains expensive to sequence a large number of individuals, as required in fisheries genomics. Therefore, the Pool-Seq approach, where individuals are pooled together based on pre-defined criteria and sequenced as one unit, is a cost-effective method to generate genome-wide SNPs. Several studies have shown that Pool-Seq studies accurately reflects population allelic frequencies, and can be used for population-based inferences, if a sufficient number of individuals is pooled (12-40 per pool) (Anand et al., 2016; Schlötterer, Tobler, Kofler, & Nolte, 2014). Therefore, in this study each pool consisted of 20-40 individuals, grouped per species and based on latitude and depth. A total of nine pools were obtained, five for *M. paradoxus* (pnn, pcn, por, pwc and psw) and four for *M. capensis* (cnn, ccn, cwc and cwc2) (Figure 5). Samples were part of the original Henriques et al. (2016) study, and were collected in 2012. All individuals were sequenced for mtDNA CR and the majority was also genotyped in Henriques et al. (2016).

![Figure 5: Location of samples pooled for the genomic study on the Cape hakes. Samples were collected in 2012 for the Henriques et al. (2016) study. From north to south: nn – northern Namibia, cn – central Namibia, or – Orange river, wc – west coast of South Africa and sw – southwest coast of South Africa](image-url)
As we were mostly interested in understanding the microevolutionary processes shaping the population structure of *M. capensis*, pools were only obtained from either side of the mixing region between southern Namibia – central West Coast South Africa, to avoid collecting mixed population samples.

DNA concentration of each individual used in the pool was normalized to equal molarities to ensure equal individual contributions during sequencing. Final concentration of the pools was 2000ng/ul. Pools were sent to the Hawaiian Institute for Marine Biology for preparation of ezRAD libraries (Toonen et al., 2013).

Raw reads were received from HIMB in the fastq format. Quality was assessed using fastqc (available at https://www.bioinformatics.babraham.ac.uk/projects/fastqc), and reads were trimmed for Illumina adapters, read quality (Q > 20) and minimum length (l=50 bp), in trimmomatic (Bolger, Lohse, & Usadel, 2014). Trimmed reads were mapped using the Burrows-Wheeler aligner of BWA-MEM (available at http://bio-bwa.sourceforge.net/bwa.shtml) against the cod genome (*Gadus morhua: gadMor2* – (Tørresen et al., 2017), available at https://osf.io/4qsdw), as there are no available genomes for Merlucidae. Mapped reads were filtered based on mapping quality (q=10) and only primary alignments were retained (Fx0004), in samtools v.1.0.18 (Li et al., 2009). Resulting bam files were further sorted and sub-sampled to normalize the number of reads proportionally to the lowest pool. Variant calling was conducted using the ‘mpileup’ command in samtools v.1.0.18, and converted to a ‘sync’ file in Popoolation2 (Kofler, Pandey, & Schlotterer, 2011). SNPs were called from this ‘sync’ file using the ‘snp-frequency-diff.pl’ and based on minimum count 2, minimum number of reads 5 and maximum number of reads 100, in Popoolation2. Additional filtering steps were performed to keep only biallelic, non-fixed (SNPs against the cod genome), and no singletons SNPs, resulting in two SNP datasets for each species (“capensis” and “paradoxus”), as well as one final dataset with both species combined (“hakes”).

Genetic diversity measures were estimated as number of SNPs (Nb) and genetic diversity (d) based on the obtained allelic frequencies per pool, using a custom script in R (R Core Team, 2018). Genetic diversity was estimated as Nei’s d (\(h_j=1-p^j-q^j\)), where \(p\) and \(q\) are the major and minor allele frequencies, respectively), while patterns of genetic differentiation were calculated as pairwise \(F_{ST}\) per bp in Popoolation2, using the same criteria as to call SNPs. Obtained values were plotted per Linkage Group in R, to assess the genomic distribution of genetic divergence. Furthermore, Principal Component Analyses (PCAs) were performed in R, using the estimated allelic frequencies, per species and for both species combined.

Assessment of the composition of each pool in terms of total length was conducted in R, in order to account for the possibility of a cohort effect, as suggested by Henriques et al. (2016) for the observed annual differences in allelic frequencies in *M. paradoxus*. Data was visualized via boxplots, and differences in size between pools were tested for statistical significance via a paired t-test, in R.

**Findings**

A total of 26 544 916 reads were obtained for *M. capensis*, and 27 498 973 reads were obtained for *M. paradoxus*. After trimming, mapping and filtering, the final dataset consisted of 52 354 SNPs for both species, distributed across the 23 Linkage Groups of the cod genome, with 41 189 SNPs found in *M. capensis* and 36 189 SNPs found in *M. paradoxus*, as some SNPs were fixed in one species, but variable in the other, while others were variable in both (Figure 6).
Overall, mean genetic diversity did not differ between *M. capensis* and *M. paradoxus* (d=0.082), but there were some differences between pools (Figure 7). For *M. capensis*, mean d did not vary greatly, ranging from 0.080 < d < 0.084. For *M. paradoxus*, though, the central Namibian pool (pcn) exhibited the highest value (d=0.086), while the southwestern South African pool (psw) exhibited the lowest value (d=0.075) (Figure 7).

Principal component analyses between the two species allowed to clearly identify each species as being unique (Figure 8). The first two axis explained 39.9% of the total variation, with the differentiation between the species being mostly found along PC1. While *M. capensis* clustered in a small area of the plot, *M. paradoxus* pools appear widespread along PC2 (Figure 8).
When focusing on each species alone, the first two axes of the PCA for *M. capensis* explained 69.9% of variation, and allowed to distinguish three major groups: the northern pools (cnn and ccn), and two South African pools (cwc and cwc2, respectively) (Figure 9). Isolation of the cwc pool appeared to be found mostly along the 1st axis, while cwc2 pool appears closer to the northern pools, being differentiated along the 2nd axis (Figure 9).

Similarly, the first two axes of the PCA for *M. paradoxus* explained 63.1% of the total variation, and three clusters were identified: northern Namibia (pnn), west coast South Africa (pwc) and remaining
pools: central Namibia, Orange River and southwestern South Africa (pcn, por, psw), mostly along the 1st axis (Figure 10). No obvious geographical structure was retrieved for *M. paradoxus*.

**Figure 10:** PCA of allelic frequencies of 36 189 SNPs identified for *M. paradoxus*

Analyses of pairwise $F_{ST}$ revealed higher mean estimates for *M. paradoxus* ($F_{ST} = 0.050$) than *M. capensis* ($F_{ST} = 0.040$). Within species, pairwise $F_{ST}$ levels varied between $F_{ST} = 0.037 – 0.043$ for *M. capensis* and $F_{ST} = 0.036 – 0.059$ for *M. paradoxus* (Tables 4 and 5). For *M. capensis*, the pairwise comparison between the northern pools was the lowest, while the comparison between the two southern pools was the highest (Table 4). For *M. paradoxus*, pairwise $F_{ST}$ levels replicated the observed patterns of the PCA, with pools pcn, por and psw showing the lowest values, and comparisons involving pnn, pwc and psw showing the highest (Table 5).

**Table 4:** Pairwise $F_{ST}$ values for *M. capensis*

<table>
<thead>
<tr>
<th></th>
<th>cnn</th>
<th>ccn</th>
<th>cwc</th>
<th>cwc2</th>
</tr>
</thead>
<tbody>
<tr>
<td>cnn</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ccn</td>
<td>0.037</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>cwc</td>
<td>0.041</td>
<td>0.042</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>cwc2</td>
<td>0.040</td>
<td>0.039</td>
<td>0.043</td>
<td>x</td>
</tr>
</tbody>
</table>

**Table 5:** Pairwise $F_{ST}$ values for *M. paradoxus*

<table>
<thead>
<tr>
<th></th>
<th>pnn</th>
<th>pcn</th>
<th>por</th>
<th>pwc</th>
<th>psw</th>
</tr>
</thead>
<tbody>
<tr>
<td>pnn</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pcn</td>
<td>0.057</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>por</td>
<td>0.057</td>
<td>0.036</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pwc</td>
<td>0.048</td>
<td>0.051</td>
<td>0.052</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>psw</td>
<td>0.059</td>
<td>0.042</td>
<td>0.038</td>
<td>0.058</td>
<td>x</td>
</tr>
</tbody>
</table>
Assessment of size ranges for *M. capensis* revealed a large difference between the northern and southern individuals, with average size for the northern pools ranging from 40 – 43 cm, and for the northern pools ranging from 55 – 61 cm (Figure 11). No significant differences were observed between average pool sizes within each region (t=0.98, p>0.05; t=1.38, p>0.05 for northern and southern pools). The differences in size between northern and southern populations of *M. capensis*, thus, do not fully account for the observed PCA patterns and pairwise F<sub>ST</sub> differences, as the southern pools were strongly genetically differentiated from each other in both instances, but do not have significant changes in size.

![Figure 11: Boxplots with the distribution of sizes of individuals pooled per region for M. capensis. Average depth of stations pooled above the boxplots](image)

Distribution of sizes for *M. paradoxus*, on the contrary, showed very similar patterns to those retrieved for the pairwise F<sub>ST</sub> and PCA (Figure 12). The northern most pool (pnn) contained individuals that were significantly smaller than the remaining pools (mean TL = 40 cm; all comparisons had p<0.0001, with the exception of pcn that was not significantly different), while the west coast pool (pwc) was significantly larger than remaining pools (mean TL = 49.9 cm; all comparisons p<0.0001). The three pools that appeared closely related in the PCA (pcn, por and psw) were either not significantly different from each other (por x psw – t=-6.28, p>0.05), or slightly different (pcn x psw – t= -3.3716, p<0.001).
Interestingly, pool pcn that appears closely related to pools por and ps w based on allelic frequencies (Figure 10), was closer related to pool pnn based on mean size (Figure 12). However, the variance of individual sizes was larger in pcn than pnn (standard variance = 6.70 vs standard variance = 5.24, respectively), suggesting that pcn is composed by multiple cohorts.

Conclusions

Previous genetic work conducted for the Cape hakes using microsatellite markers revealed the presence of two populations for *M. capensis* and one population for *M. paradoxus*. These findings seemed to mimic the known biological features of the species, as two spawning grounds have been described for *M. capensis* while only one has been described for *M. paradoxus* across the Benguela region (Jansen et al., 2015; Strømme et al., 2016).

The present findings have added to the knowledge of population connectivity in the Cape hakes, by revealing a cryptic third stock in *M. capensis* in the southern Benguela region, and a complex pattern of population connectivity in *M. paradoxus*. Using the same methods to generate millions of reads spread throughout the genome, and the exact same bioinformatic pipeline to identify and call SNPs,
the population structure patterns observed remained species-specific, as previously seen in Henriques et al. (2016).

For *M. capensis*, the northern – southern divide previously identified remained the strongest differentiation found, supporting the previous conclusion of homing behaviour to the northern and southern spawning sites, with limited gene flow between populations. However, one additional population was further identified within the southern system, without geographic isolation. Differential survivability and reproductive success of cohorts could be a possible explanation for this genetic pattern, but analyses of average size of individuals per pool did not reveal significant differences between the southern pools. Therefore, the most likely explanation is the presence of a second population of *M. capensis* within the southern system, with extensive migration and mixing. Interestingly, the hypothesis of three groups was also suggested as likely using microsatellite loci, as for K = 3, southern individuals appeared to be composed by two different genetic pools. However, no “pure” individual was found for these putative southern clusters, and the hypothesis was discarded at the time. It is thus likely that the microsatellite loci used in Henriques et al. (2016) did not have enough statistical power to accurately detect between the two southern populations. A similar finding of three populations was observed for kingklip (*Genypterus capensis*) across the Benguela Region (Schulze et al, personal communication).

For *M. paradoxus*, strong genetic differentiation was observed across the Benguela region, with pools pnn (northern Namibia), pwc (west coast South Africa) and psw (southwestern coast South Africa) consistently identified as highly differentiated. However, the pools sampled in central Namibia (pcn) and the Orange river region (por) were more genetically similar to psw, than to their closest geographic pools (pnn and pwc, respectively). This absence of a clear geographical structure suggests that factors other than a break in gene flow due to demographic independence is responsible for the observed patterns. In the previous work by Henriques et al. (2016) the authors suggested that differential survivability and reproductive success among cohorts could help explain the stochastic patterns of genetic differentiation found based on the mtDNA data. Present results appear to corroborate this explanation, as significant differentiation was observed in terms of size of the individuals included in the analysed pools, pointing to different cohorts being sampled. The northern pools (pnn and pcn) had consistently smaller sizes, while the pool from the west coast of South Africa (pwc) was significantly larger than all remaining pools, and the Orange River and Southwest coast of South Africa (por and psw) had somewhat intermediate sizes and were similar in size in comparison. These findings suggest that a cohort effect may indeed be contributing to the observed pattern of genetic differentiation. Further analyses will need to be conducted to assess, among others, the level of relatedness between individuals within each pool, to further explore the hypothesis of a strong cohort-effect.

In conclusion, the two Cape hakes exhibit significantly different evolutionary histories, that appear to have been strongly impacted by their different life history features and demographic history, namely the presence of multiple versus a single spawning grounds, the levels of historical effective population size, the reproductive success of spawning adults and the survivability of cohorts.

These findings point, thus, to the presence of three populations within *M. capensis*, and a single, complex population of *M. paradoxus* across the Benguela region.
References


multilocus genotype data. *Genetics, 155*(2), 945–959.


**Funding**

This work was funded partially by the European Union through the ECOFISH project (EuropeAid grant 2010/222387) and a collaborative NRF-NCREST SA-NAMIBIA Bilateral Grant (56311/29729/1771) awarded to Stellenbosch University and University of Namibia.