ICES Journal of Marine Science

ICES Journal of Marine Science (2019), doi:10.1093/icesjms/fsz161

Genome-wide SNP based population structure in European hake reveals the need for harmonizing biological and management units

Agostino Leone[1], Paula Álvarez[1], Dorleta García[1], Fran Saborido-Rey[2], and Naiara Rodriguez-Ezpeleta [1]*

[1]*Marine Research Division, AZTI, Txatxarramendi ugartea z/g, 48395 Sukarrieta, Bizkaia, Spain*
[2]*IIM-CSIC, Institute of Marine Research, Eduardo Cabello 6, 36208 Vigo, Pontevedra, Spain*

**Corresponding author: tel: + 34 667 174 514; e-mail: nrodriguez@azti.es.*

Leone, A., Álvarez, P., García, D., Saborido-Rey, F., and Rodriguez-Ezpeleta, N. Genome-wide SNP based population structure in European hake reveals the need for harmonizing biological and management units. – ICES Journal of Marine Science, doi:10.1093/icesjms/fsz161.

Received 18 April 2019; revised 23 July 2019; accepted 30 July 2019.

Despite its economic importance, the population structure of the European hake, *Merluccius merluccius*, is unresolved, and the species is assessed based on two stocks (northern and southern) separated by the Capbreton Canyon. In order to shed light into the European hake population structure, we used Restriction-site-Associated DNA sequencing (RAD-seq) to discover and genotype thousands of genome-wide single nucleotide polymorphisms in more than a hundred samples. Our population genetic inferences confirm differentiation of Mediterranean and northeast Atlantic locations and reveal management relevant information within the latter. First, hake in the Norwegian Sea is genetically different from that of the rest of the locations under study and, second, samples from the eastern Bay of Biscay and the northwestern Iberian Peninsula are not genetically different. These results imply that samples from the northern stock belong to different genetic populations, and that samples belonging to locations included in the northern and southern stocks are part of a single genetically homogeneous population. Although the definition of the boundary between the northern and southern stocks and the potential need for additional stocks still requires further analyses, the mismatch between biological and management units should already be considered in further assessments of European hake.

Keywords: European Hake, fishery management, genetic connectivity, RAD-seq, SNPs, stock delimitation

Introduction

Determining spatio-temporal genetic variation is important for sustainable fishery management (Hauser and Carvalho, 2008). Yet, intrinsic characteristics of marine fish such as large population sizes, high dispersal capacity, and high fecundity often translate into low genetic differentiation even at interoceanic distances (Nielsen and Kenchington, 2001; Nielsen *et al.*, 2009). This often results in a mismatch between biological and management units (Reiss *et al.*, 2009), which can lead to reduced productivity, local reduction of populations, and in extreme cases, to local population collapse (Bonfil, 2005). An example of potential mismatch between biological and management units is the European hake (*Merluccius merluccius*; Reiss *et al.*, 2009), an important commercial fish that inhabits the Northeast Atlantic Ocean and adjacent seas (Korta *et al.*, 2015; ICES, 2018). Within the Northeast Atlantic Ocean, the European hake is managed by the International Council for the Exploration of the Sea (ICES) assuming two stocks separated by the Capbreton Canyon, the northern stock extending northward to Norway (ICES Division 3a, Subareas 2, 4, 6, and 7, and Divisions 8a, b, and d) and the southern stock, southward to Gibraltar Strait (ICES Divisions 8c and 9a; ICES, 2011, 2018). Yet, this separation is not based on

clear biological evidences other than the assumption of the barrier effect of the Capbreton Canyon (Casey and Pereiro, 1995). Additionally, in locations from the northern stock (e.g. Bay of Biscay), the main spawning season ranges from January to March, whereas in locations from the southern stock (Galicia and Portugal), spawning occurs all year-round with the presence of several other peaks in summer and autumn (Korta *et al.*, 2015; Dominguez-Petit, 2017).

There is still no consensus on the population structure of the European hake (Roldán *et al.*, 1998; Castillo *et al.*, 2005; Pita *et al.*, 2014). Studies based on otolith chemistry, which reflects where fish has been born and grown, have detected differences between samples from southern (represented by northern Portugal) and northern (represented by the Norwegian Sea and West of Scotland) stocks (Swan *et al.*, 2006), whereas others (Tanner *et al.*, 2012) found no significant differences between samples from the northern (represented by the Celtic Sea and the Armorican Shelf) and the southern stock (represented by samples from the Portugal and Galician Shelf). Early studies based on microsatellite genetic markers resulted in conflicting results. For example Lundy *et al.* (1999) found differentiation within the northern (e.g. Celtic Sea and Norwegian Sea) and southern (e.g. Cantabrian Sea and southern Portugal) stocks but no differentiation between the two stocks (e.g. Celtic Sea and Bay of Biscay). Castillo *et al.* (2005) confirmed the lack of differentiation between samples from the Bay of Biscay (ICES Divisions 8c vs. 8a, b, d) belonging to the two stocks but found no differentiation between the Cantabrian Sea and Portuguese coast samples (ICES Divisions 8 and 9). More recent studies using microsatellites have confirmed the lack of differentiation between Bay of Biscay samples belonging to northern and southern stocks and suggested a possible differentiation of North Sea samples (Pita *et al.*, 2011, 2014).

Despite the advent of High Throughput Sequencing technologies and their demonstrated potential to study population structure in marine fish (Bernatchez *et al.*, 2017), so far only one study has applied this technique to study population connectivity in the European hake (Milano *et al.*, 2014), and is only based on 381 single nucleotide polymorphisms (SNPs), a very small number compared with the tens of thousands used in other studies aimed at resolving population structure of marine fish (Rodríguez-Ezpeleta *et al.*, 2016, 2017). Milano *et al.* (2014) did not find differentiation between northern and southern stocks when using neutral markers but found differentiation within the southern (between Galicia and north Portugal) and within the northern (between North Sea and the other locations) stocks when using seven outlier loci. This result was further confirmed by another study based on a subset of this dataset (Westgaard *et al.*, 2017); yet, neutral markers did not show the differentiation observed with outlier loci, even when using a bigger sample size, meaning that additional and more powerful analyses are needed to better resolve the European hake population structure.

Here, we have discovered and genotyped genome-wide SNP markers in European hake to (i) define the population structure of this commercially important species along the Mediterranean–Norwegian gradient, (ii) investigate the possibility of seasonal genetic changes within southern stock samples, and (iii) test if the current assessment and management units and population structure of European hake coincide. In order to reach these goals, we generated Restriction-site-Associated DNA sequencing (RAD-seq) data from 111 European hake individuals that span four geographic areas where this species inhabits: the Norwegian Sea, the eastern Bay of Biscay, the northwestern Iberian Peninsula (including samples from both spawning seasons) and the Mediterranean Sea (Balearic Sea), enabling the opportunity to explore population genetic structure and scan for selective divergences in this species. Our study reveals a mismatch between biological populations and management units in the European hake, which should be considered in the assessment of this commercially important species.

Material and methods

Sampling, DNA extraction, and RAD-seq library preparation

European hake samples from the Mediterranean Sea, the northwestern Iberian Peninsula (southern stock), the eastern Bay of Biscay (northern stock), and the Norwegian Sea (northern stock) were obtained from scientific surveys and commercial fishery (Figure 1; Supplementary Table S1). Samples include fish in spawning and prespawning condition, assumed to be at or close to where they will spawn, and juveniles of age 0 (<20 cm), assumed to be close to where they were born (de Pontual *et al.*, 2013). From each specimen, a ~1 cm^3 muscle sample was excised and stored in 96% molecular grade ethanol at −20°C prior to DNA extraction. Genomic DNA was extracted from about 20 mg of muscle tissue using the Wizard® Genomic DNA Purification kit (Promega, WI, USA) following manufacturer's instructions. Extracted DNA was suspended in Milli-Q water and concentration was determined with the Quant-iT dsDNA HS assay kit using a Qubit® 2.0 Fluorometer (Life Technologies). DNA integrity was assessed by electrophoresis. RAD-seq libraries were prepared following the methods of Etter *et al.* (2011). Between 300 and 400 ng of genomic DNA per sample was digested using the *SbfI* restriction enzyme then ligated to modified Illumina P1 adapters containing 5 bp sample-specific barcodes. Pools of 32 DNA samples were sheared using the Covaris® M220 Focused-ultrasonicator™ Instrument (Life Technologies) and size-selected to 300–500 pb by cutting DNA fragments migrated though an agarose gel. Following the ligation of the Illumina P2 adaptors, each genomic library was amplified using 12 PCR cycles. Finally, the pools were paired-end sequenced (100 bp) on an Illumina HiSeq2000. Raw sequences are available at NCBI SRA Bioproject PRJNA556115.

RAD-tag assembly and SNP calling

Generated RAD-tags were analysed using *Stacks* version 2.1 (Catchen *et al.*, 2013). Quality filtering and demultiplexing was performed using the *Stacks* module *process_radtags*, truncating all reads to 90 nucleotides after sequence quality verification in FastQC (version 0.11.7; Andrews, 2010) in order to avoid lower quality bases at the end of the read. The module *clone_filter* was applied to reads whose forward and reverse pairs passed quality filtering to remove PCR duplicates, and only samples with at least 400 000 clone and quality-filtered reads remaining were kept. The module *ustacks* was then used to assemble orthologous tags (stacks) per individual, with a minimum coverage depth required to create a stack (parameter -m) of five, and a maximum nucleotide mismatches allowed between stacks (parameter -M) of two. RAD-loci were then assembled using the module *cstacks* with a maximum of two mismatches allowed between sample tags when generating the catalogue (parameter -n). Matches to the catalogue for each sample were searched using *sstacks*. The module

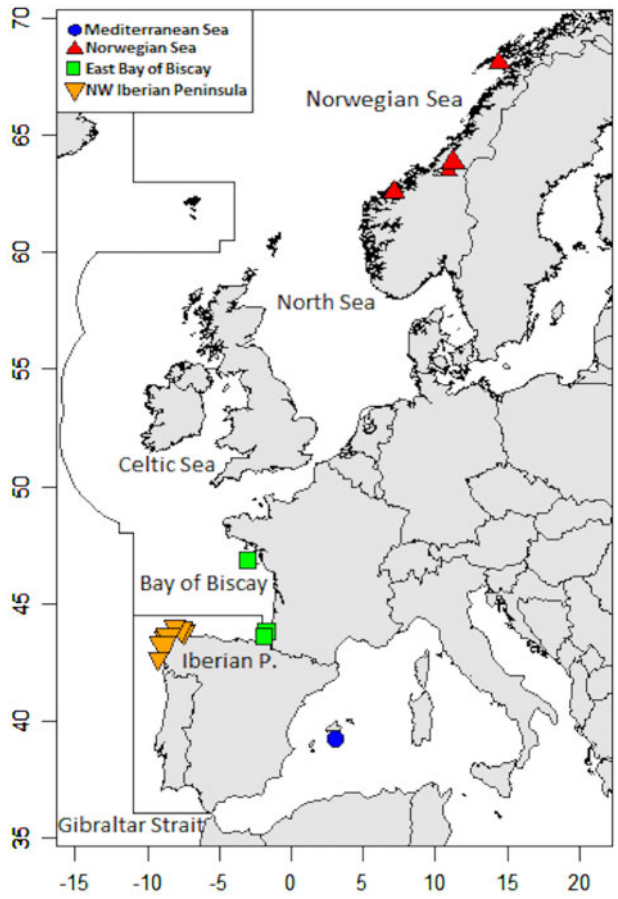

Figure 1. Sampling sites of European hake and ICES delimitation for the northern (top) and southern (bottom) stock management units. The border between northern and southern stocks approximates the Capbreton Canyon.

populations were used to export from the catalogue the SNPs present in RAD-loci found in at least 75% of the individuals. The software *PLINK* version 1.07 (Purcell *et al.*, 2007) was used to further filter the data. The initial filtering steps were based on a maximum of 10% of missing values per locus, and a maximum of 25% of missing values per individual. Subsequently, SNPs with a minimum allele frequency (MAF) smaller than 0.05 and which failed the Hardy–Weinberg equilibrium (HWE) test at $p < 0.05$ in at least two areas of study were excluded from downstream analyses. Only the first filtered SNP of each RAD-tag was kept. Additionally, an alternative dataset with $M = 4$ and $n = 6$ (keeping the rest of the parameters unchanged) and another with MAF threshold of 0.1 (keeping the rest of the parameters unchanged) were also explored for consistency as suggested (Díaz-Arce and Rodríguez-Ezpeleta, 2019). The genotype files are available at https://sites.google.com/site/naiararodriguezezpeleta/resources.

Identification of outlier loci

Outlier loci, potentially under selection, were identified using the independent approaches used by *Bayescan* version 2.0 (Foll and Gaggiotti, 2008) and *Arlequin* version 3.5.2.2 (Excoffier and Lischer, 2010) using the whole dataset and only the Bay of Biscay (Western and Eastern) samples. In *Bayescan*, 20 pilot runs of 5000 iterations each were run, with a burn-in of 50 000 iterations, 5000 samples with a thinning interval of 10 and a prior odd increased to 100 in order to minimize potentially false positives in a dataset with >10 000 loci. Using these parameters, loci with an α-value significantly >0 (indicative of diversifying selection) and q-values <0.05 were considered "outliers." In addition, in the hierarchical Fdist model implemented in *Arlequin*, loci with significantly higher fixation index (FST) values ($p < 0.001$) were considered "outliers." These stringent parameters were used to account for potential false positive outliers discovery (Narum and Hess, 2011; Ahrens *et al.*, 2018). Outputs from *Bayescan* and *Arlequin* were intersected in order to create a consensus panel of shared loci potentially under selection, which were annotated by matching the SNP flanking regions against non-redundant database of GenBank (www.ncbi.nlm.nih.gov/genbank/) using BLAST (Altschul *et al.*, 1990). In addition, a multivariate approach to outlier loci discovery was also applied using the package *pcadapt* (Luu *et al.*, 2016) in R version 3.5.0 (R Core Team, 2018).

Genetic diversity and structure analyses

Analyses were performed on the whole SNP dataset, on the neutral SNP dataset, and on the outlier SNP dataset. Expected and observed heterozygosity per group (He and Ho respectively), pairwise average FST among groups and distributions of FST and inbreeding coeficient (FIS) values were computed with *Arlequin* version 3.5.2.2 (Excoffier and Lischer, 2010). Principal component analyses (PCAs) and a three-cluster based discriminant analyses of principal components (DAPC; Jombart *et al.*, 2010) were performed without any a priori assignment of individuals to populations using the package *adegenet* (Jombart and Ahmed, 2011) in R version 3.5.0 (R Core Team, 2018). The best number of PCs to retain for the DAPC analysis was inferred using the cross-validation test. The genetic ancestry of each individual was estimated using the admixture model as implemented in the Bayesian clustering approach in STRUCTURE version 2.3.4 (Pritchard *et al.*, 2000) without any prior population assignment. Estimations were obtained from the 300 000 iterations that followed a burn-in period of 100 000 iterations. Each value of potential ancestral population (K from 1 to 5) was examined with CLUMPP (Jakobsson and Rosenberg, 2007) to identify common modes, and results were plotted using DISTRUCT (Rosenberg, 2003). The value of K that better fits the data were identified according to the Evanno method (Evanno *et al.*, 2005) as implemented in StructureHarvester (Earl and vonHoldt, 2012). Additionally, 10 subsets of 300 randomly selected SNPs were analysed with $K = 3$ in order to investigate the effect of using a smaller dataset of genome-wide representative SNPs.

Results

SNP datasets

Of the 111 analysed samples, only one was removed for having too few reads. Considering the remaining samples, between 402 599 and 3 053 042 quality-filtered RAD-tags per individual were obtained (average of 1 365 352). Ninety per cent of the reads were used to build the stacks, with a range per individual between 81 and 95%, and the average effective per sample coverage is 34.1× (stdev = 12.8×, min = 11.3×, max = 72.9×). The number of RAD-loci present in at least 75% of the individuals was 29 090, totalling 452 702 SNPs. After filtering for missing data, minor allele frequency and HWE and keeping a single SNP per tag, the number of RAD-loci and SNPs remaining for analysis was 14 120

Table 1. Observed and expected heterozygosities per group of samples when using all or only outlier SNPs.

POP	*n*	14 120 total SNPs		116 outlier SNPs	
		Ho	He	Ho	He
MED	14	0.25732	0.26521	0.29418	0.30206
EBOB	18	0.24744	0.24918	0.27476	0.27110
NWIPs	30	0.23193	0.24219	0.27043	0.27541
NWIPw	30	0.2348	0.24144	0.25000	0.24945
NOR	18	0.25252	0.25221	0.31828	0.31676

MED, Mediterranean Sea; EBOB, Eastern Bay of Biscay; NWIPs, Northwestern Iberian Peninsula (summer); NWIPw, Northwestern Iberian Peninsula (winter); NOR, Norwegian Sea.

(Supplementary Figure S1). The amount of missing data was <5% overall, and datasets with 0 and 1% of missing data, albeit composed of fewer SNPs produced equivalent results (not shown). Of the remaining SNPs, 147 (1% of the total) and 217 SNPs (1.5% of the total) were considered outliers according to the Bayesian approach implemented in *Bayescan* and the hierarchical Fdist model implemented in *Arlequin* when using the whole dataset (Supplementary Figure S2), and 116 of them were common to both approaches. No outlier loci were found when using only the northwestern Iberian Peninsula and the Eastern Bay of Biscay samples. Additionally, 347 SNPs (2.4% of the total) were considered outliers according to the multivariate approach implemented in *pcadapt*. Of these loci, 108 were included in the 116 shared outlier datasets by *Bayescan* and *Arlequin*. Annotation of each of the outlier SNPs is available in Supplementary Table S2.

Genetic diversity and population structure

Ranges of observed and expected heterozygosity per groups were similar, both when using the whole or the only outlier SNP dataset, indicating no deficits of heterozygous genotypes in all samples (Table 1). STRUCTURE analyses using the best *K* values according to the Evanno method for the whole or the outlier only datasets (Supplementary Table S3) resulted in three well differentiated clusters: (i) the Mediterranean Sea samples, (ii) the northwestern Iberian Peninsula and the eastern Bay of Biscay samples, and (iii) the Norwegian Sea samples (Figure 2), with no further structure observed when including only the eastern Bay of Biscay and the northwestern Iberian Peninsula samples. Analyses based on higher values of *K* resulted in no further differentiation (Supplementary Figure S3). Analyses based on 10 subsets of 300 randomly selected SNPs datasets provided varying results (Supplementary Figure S4): same differentiation pattern (five subsets), no differentiation within the Atlantic (four subsets) differentiation between the Mediterranean plus Norwegian Sea samples and the northwestern Iberian Peninsula plus the eastern Bay of Biscay samples (one subset).The Analysis of Principal Components (PCA) on the whole dataset were congruent with the STRUCTURE analyses, revealing the same three well differentiated groups (Figure 3), even with alternative RAD-tag assembly parameter and MAF threshold values (Supplementary Figure S5), and with the Discriminant Analysis of Principal Component (DAPC) when using three clusters (Supplementary Figure S6).

The locus-by-locus FST and FIS were distributed around zero (Supplementary Figure S7) suggesting a balanced amount of

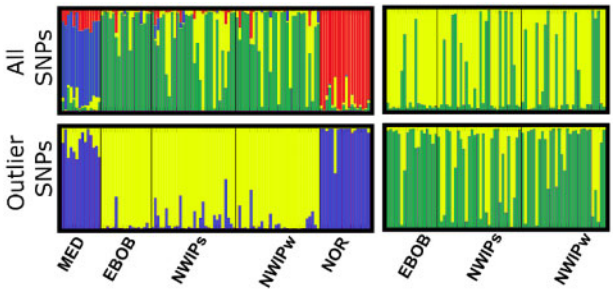

Figure 2. Graphical representation of the Bayesian clustering approach implemented in STRUCTURE when using all (top) or only outlier SNPs (bottom) and all (left) or only the eastern Bay of Biscay and northwestern Iberian Peninsula samples (right) for the best *K* value according to the Evanno method, which is $K = 2$ for all dataset except for the all SNPs all samples dataset, for which it is $K = 4$. Refer to Table 1 for location codes.

heterozygosity and homozygosity along the loci, which is congruent with the small number of outliers detected. Pairwise FST values among the three differentiated areas were significant ranging from 0.023 to 0.058, from 0.443 to 0.498, and from 0.014 to 0.042 (Table 2, Supplementary Tables S4 and S5) when using all, only the outlier SNPs and only the neutral SNPs, respectively. Low and not significant differentiation was detected within samples from southern Europe, even when using only adult samples or only outlier SNPs suggesting that all samples from the northern Spanish coast are part of a single population with no seasonal demographic changes.

Discussion

Genome-wide population structure of European hake does not support current stock delineation

Our genome-wide marker-based population structure inference of European hake supports differentiation between Mediterranean and Northeast Atlantic hake. Within the Northeast Atlantic, our study showed a clear differentiation between the Norwegian Sea and southern Europe, but did not find differences between samples of the Bay of Biscay and northwestern Iberian Peninsula, belonging to the northern and southern stocks. These results confirm previous findings using microsatellites that found no differentiation between samples within the Bay of Biscay and Celtic Sea, but found differentiation between these regions and the Norwegian Sea (Lundy *et al.*, 1999; Pita *et al.*, 2014). Interestingly, the only previous study based on SNP markers (Milano *et al.*, 2014) failed to detect population structure within the Northeast Atlantic when using neutral markers, which could be because of a lack of evolutionarily informative signal in the few (299) markers used. Yet, when we base our inferences in 300 randomly selected SNPs, we still observe population structure, but only in half of the 10 subsets. This suggests that using a few selected hundreds SNPs as in previous studies (Milano *et al.*, 2014) are not enough for a reliable inference of hake population structure. Using a small subset of outlier SNPs ($n = 7$ or $n = 6$) both Milano *et al.* (2014) and Westgaard *et al.* (2017) showed differentiation between the Bay of Biscay and the North Sea and/or Norwegian Sea samples and argued that the differences obtained using neutral and outlier loci are because of adaptation to local conditions. Although it is not unusual to find adaptive divergences in fish even when neutral markers suggest panmixia at local

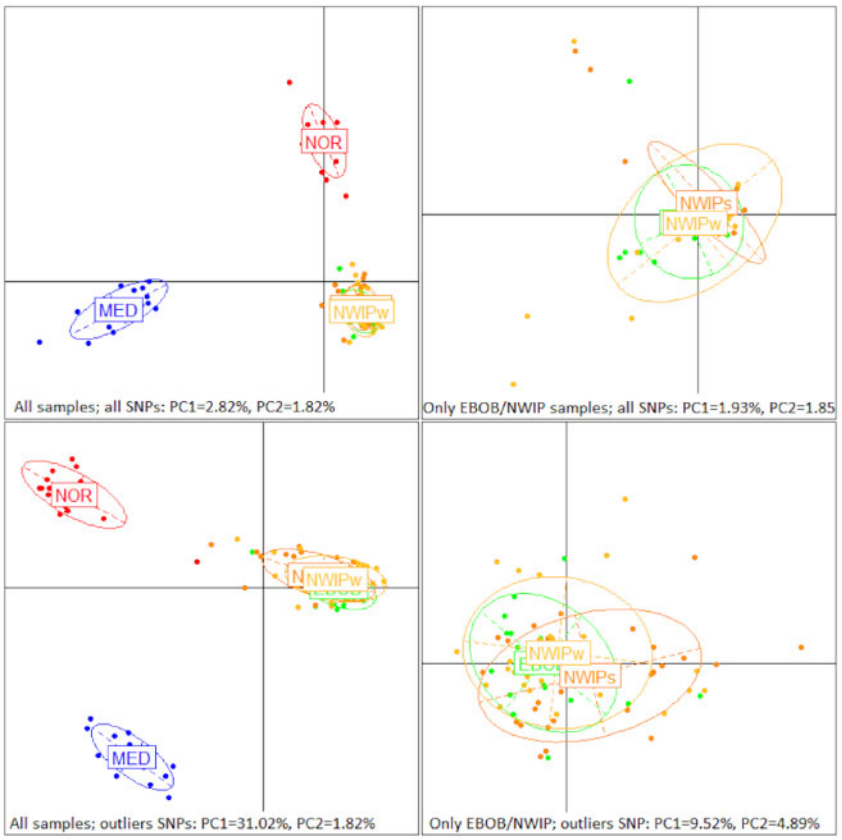

Figure 3. Principal component analysis (PCA) when using all (top) or only outlier SNPs (bottom) and all (left) or only the eastern Bay of Biscay and northwestern Iberian Peninsula samples (right). The variance explained by each axis of the PCA reflects the amount of variation observed when using the total or the outlier SNPs dataset.

Table 2. Pairwise FST value matrix when using all or only outlier SNPs (below and above diagonal respectively).

	MED	EBOB	NWIPs	NWIPw	NOR
MED	/	**0.464**	**0.443**	**0.480**	**0.470**
EBOB	**0.052**	/	0.008	0.005	**0.496**
NWIPs	**0.048**	0	/	0.005	**0.453**
NWIPw	**0.049**	0	0	/	**0.498**
NOR	**0.058**	**0.025**	**0.023**	**0.025**	/

Values in bold are significant after Bonferroni correction for multiple testing. MED, Mediterranean Sea; EBOB, Eastern Bay of Biscay; NWIPs, Northwestern Iberian Peninsula (summer); NWIPw, Northwestern Iberian Peninsula (winter); NOR, Norwegian Sea.

level (Hemmer-Hansen *et al.*, 2007; Freamo *et al.*, 2011), this does not seem to be the case for European hake. The analysis of a consensus panel of 116 outlier SNPs potentially under selective divergence confirmed the genetic structure observed with the total dataset with a stronger signal and failed in finding additional substructure, as observed before (Moore *et al.*, 2014), suggesting that the genetic divergence observed in the European hake is not only because of selection. This contradicts the views of Milano *et al.* (2014) and Westgaard *et al.* (2017), and further highlights the importance of using genome-wide based markers in order to reliable detect genetic differentiation and local adaptation.

European hake within the Bay of Biscay constitutes a seasonally stable homogeneous population

The lack of differentiation observed within the northwestern Iberian Peninsula and the eastern Bay of Biscay, with very low and not significant FST, suggests that all the specimens from the area are representative of a unique genetic population, partially solving the dubious genetic homogeneity present within the Bay of Biscay (Castillo *et al.*, 2005; Pita *et al.*, 2014, 2016). Additionally, in our study we have tested the seasonal stability of the European hake connectivity within the southern stock (northwestern Iberian Peninsula). Our analyses, based either on neutral or outlier loci do not show differentiation among samples collected in summer and winter within the northwestern Iberian Peninsula, supporting previous views (Lundy *et al.*, 2000; Pita *et al.*, 2016) and most likely explained by the fact that hake can spawn several times within the reproductive season (Murua *et al.*, 2006; Murua, 2010). Yet, further studies with bigger and seasonal representative datasets and including only spawning individuals are necessary to better investigate any potential spatio-temporal substructuring within the Bay of Biscay.

Management implications

In the Northeast Atlantic, assessment and management of the European hake is done independently for two stocks, the southern and the northern, separated by the Capbreton Canyon. The biomass of both stocks was in historical minimum and below

precautionary biological reference points in the last years of the previous century (ICES, 2018). At the beginning of the new century the stocks started recovering and now the biomass of both is above precautionary levels. In the case of the northern stock the recovery has been extraordinary and the current biomass level doubles the historical maximum estimated by the stock assessment model at the beginning of the 1980s. Furthermore, the fishing mortality is well below the maximum sustainable yield target. However, the recovery of the southern stock has not been sufficiently strong and although the biomass is above the reference levels the fishing mortality is still too high. This study suggests that the delimitation of the assessment and management areas should be revised. About the assessment, the southern stock and the southern part of the northern stock should be assessed jointly, and a separate stock should be considered in the northern part of the species distribution, which includes the Norwegian Sea and possibly the North Sea, although this has yet to be confirmed with additional analyses. Since 2006 the biomass of the northern stock has increased tenfold (ICES, 2018), resulting in an increased abundance of the stock in the North Sea during spring and summer (Baudron and Fernandes, 2015). Additionally, recent explicit stock assessment models (Vigier *et al.*, 2018) predict significant movement between the North Sea and the Celtic Sea and the Bay of Biscay suggesting that, under the hypothesis of a separate population in the northern part of the distribution, the North Sea could be a mixing area of both populations.

In sum, our results support the need for defining new assessment and management units for European hake. Yet, despite covering a wide range of the species distribution, our study does not provide information on where the new stock boundary should be located; therefore, future studies including samples from additional locations between the southern Bay of Biscay and the Norwegian Sea are necessary. Importantly, genetic analyses should also be complemented with additional evidence based on alternative techniques (morphometrics, vital rates, etc.) that would help providing biologically meaningful information for adequately defined stock boundaries.

Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (grant CTM2015-66676), and by the Department of Agriculture and Fisheries of the Basque and Government (grant GENPES). We wish to thank Arved Staby (Institute of Marine Research, Bergen), Sonia Rábade, Alex Alonso, Mariña Fabeiro, and Laura Casas (IIM-CSIC); Ibon Cancio (EHU/UPV); and Maria Korta (AZTI) for providing samples; and Naroa Aldanondo, Iñaki Mendibil, and Elisabete Bilbao (AZTI) for technical support. This manuscript is contribution 934 from the Marine Research Division of AZTI.

Authors contribution

PA, FSR, and NRE designed the study. AL and NRE analysed the data. AL, NRE, and DG interpreted the data and wrote the first draft of the paper. All authors performed critical reading of the first draft and contributed to the revisions.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

References

Ahrens, C. W., Rymer, P. D., Stow, A., Bragg, J., Dillon, S., Umbers, K. D. L., and Dudaniec, R. Y. 2018. The search for loci under selection: trends, biases and progress. Molecular Ecology, 27: 1342–1356.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. Journal of Molecular Biology, 215: 403–410.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Baudron, A. R., and Fernandes, P. G. 2015. Adverse consequences of stock recovery: European hake, a new "choke" species under a discard ban? Fish and Fisheries, 16: 563–575.

Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., and Maes, G. E. 2017. Harnessing the power of genomics to secure the future of seafood. Trends in Ecology & Evolution, 32: 665–680.

Bonfil, R. 2005. Management techniques for elasmobranch fisheries. *In* FAO Fisheries Technical Paper 474, pp. 6–14. Ed. by R. B. John Musick. Rome, Italy, FAO.

Casey, J., and Pereiro, J. 1995. European Hake (*M. merluccius*) in the Northeast Atlantic. *In* Jürgen Alheit, T.J. Pitcher Hake: Biology, Fisheries and Markets, pp. 125–147. Ed. by J. Alheit, and T.J. Pitcher. Chapman & Hall, London.

Castillo, A. G. F., Alvarez, P., and Garcia-Vazquez, E. 2005. Population structure of *Merluccius merluccius* along the Iberian Peninsula coast. ICES Journal of Marine Science, 62: 1699–1704.

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

de Pontual, H., Jolivet, A., Garren, F., and Bertignac, M. 2013. New insights on European hake biology and population dynamics from a sustained tagging effort in the Bay of Biscay. ICES Journal of Marine Science, 70: 1416–1428.

Díaz-Arce, N., and Rodríguez-Ezpeleta, N. 2019. Selecting RAD-seq data analysis parameters for population genetics: the more the better? Frontiers in Genetics, 10: 533.

Dominguez-Petit, R. 2017. Study of Reproductive Potential of Merluccius merluccius in the Galician Shelf. Vigo, Spain, University of Vigo, 253 pp.

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

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology, 14: 2611–2620.

Etter, P. D., Bassham, S., Hohenlohe, P. A., Johnson, E. A., and Cresko, W. A. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods in Molecular Biology, 772: 157–178.

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

Foll, M., and Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics, 180: 977–993.

Freamo, H., O'Reilly, P., Berg, P. R., Lien, S., and Boulding, E. G. 2011. Outlier SNPs show more genetic structure between two Bay of Fundy metapopulations of Atlantic salmon than do neutral SNPs. Molecular Ecology Resources, 11: 254–267.

Hauser, L., and Carvalho, G. R. 2008. Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. Fish and Fisheries, 9: 333–362.

Hemmer-Hansen, J., Nielsen, E. E., Frydenberg, J., and Loeschcke, V. 2007. Adaptive divergence in a high gene flow environment:

Hsc70 variation in the European flounder (*Platichthys flesus* L.). Heredity, 99: 592.

ICES. 2011. Report of the Working Group on the Assessment of Southern Shelf stocks of Hake, Monk and Megrim (WGHMM). ICES CM 2011/ACOM: 11. 625 pp.

ICES. 2018. Report of the Working Group for the Bay of Biscay and the Iberian Waters Ecoregion (WGBIE). ICES CM 2018/ACOM: 12. 585 pp.

Jakobsson, M., and Rosenberg, N. A. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23: 1801–1806.

Jombart, T., and Ahmed, I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics, 27: 3070–3071.

Jombart, T., Devillard, S., and Balloux, F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics, 11: 94.

Korta, M., García, D., Santurtún, M., Goikoetxea, N., Andonegi, E., Murua, H. M., Álvarez, P. *et al.* 2015. European hake (*Merluccius merluccius*) in the Northeast Atlantic Ocean. *In* Hakes. Ed. by H. Arancibia. Wiley-Blackwell, Oxford, UK.

Lundy, C. J., Moran, P., Rico, C., Milner, R. S., and Hewitt, G. M. 1999. Macrogeographical population differentiation in oceanic environments: a case study of European hake (*Merluccius merluccius*), a commercially important fish. Molecular Ecology, 8: 1889–1898.

Lundy, C. J., Rico, C., and Hewitt, G. M. 2000. Temporal and spatial genetic variation in spawning grounds of European hake (*Merluccius merluccius*) in the Bay of Biscay. Molecular Ecology, 9: 2067–2079.

Luu, K., Bazin, E., and Blum, M. G. B. 2016. pcadapt: an R package to perform genome scans for selection based on principal component analysis. Molecular Ecology Resources, 33: 67–77.

Milano, I., Babbucci, M., Cariani, A., Atanassova, M., Bekkevold, D., Carvalho, G. R., Espiñeira, M. *et al.* 2014. Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). Molecular Ecology, 23: 118–135.

Moore, J-S., Bourret, V., Dionne, M., Bradbury, I., O'Reilly, P., Kent, M., Chaput, G. *et al.* 2014. Conservation genomics of anadromous Atlantic salmon across its North American range: outlier loci identify the same patterns of population structure as neutral loci. Molecular Ecology, 23: 5680–5697.

Murua, H. 2010. The biology and fisheries of European hake, *Merluccius merluccius*, in the north-east Atlantic. Advances in Marine Biology, 58: 97–154.

Murua, H., Lucio, P., Santurtún, M., and Motos, L. 2006. Seasonal variation in egg production and batch fecundity of European hake *Merluccius merluccius* (L.) in the Bay of Biscay. Journal of Fish Biology, 69: 1304–1316.

Narum, S. R., and Hess, J. E. 2011. Comparison of FST outlier tests for SNP loci under selection. Molecular Ecology Resources, 11: 184–194.

Nielsen, E. E., Hemmer-Hansen, J., Larsen, P. F., and Bekkevold, D. 2009. Population genomics of marine fishes: identifying adaptive variation in space and time. Molecular Ecology, 18: 3128–3150.

Nielsen, E. E., and Kenchington, E. 2001. A new approach to prioritizing marine fish and shellfish populations for conservation. Fish and Fisheries, 2: 328–343.

Pita, A., Leal, A., Santafé-Muñoz, A., Piñeiro, C., and Presa, P. 2016. Genetic inference of demographic connectivity in the Atlantic European hake metapopulation (*Merluccius merluccius*) over a spatio-temporal framework. Fisheries Research, 179: 291–301.

Pita, A., Pérez, M., Balado, M., and Presa, P. 2014. Out of the Celtic cradle: the genetic signature of European hake connectivity in south-western Europe. Journal of Sea Research, 93: 90–100.

Pita, A., Pérez, M., Cerviño, S., and Presa, P. 2011. What can gene flow and recruitment dynamics tell us about connectivity between European hake stocks in the Eastern North Atlantic? Continental Shelf Research, 31: 376–387.

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J. *et al.* 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81: 559–575.

R Core Team. 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Reiss, H., Hoarau, G., Dickey-Collas, M., and Wolff, W. J. 2009. Genetic population structure of marine fish: mismatch between biological and fisheries management units. Fish and Fisheries, 10: 361–395.

Rodriguez-Ezpeleta, N., Álvarez, P., and Irigoien, X. 2017. Genetic diversity and connectivity in *Maurolicus muelleri* in the Bay of Biscay inferred from thousands of SNP markers. Frontiers in Genetics, 8: 195.

Rodríguez-Ezpeleta, N., Bradbury, I. R., Mendibil, I., Álvarez, P., Cotano, U., and Irigoien, X. 2016. Population structure of Atlantic mackerel inferred from RAD-seq-derived SNP markers: effects of sequence clustering parameters and hierarchical SNP selection. Molecular Ecology Resources, 16: 991–1001.

Roldán, M. I., García-Marín, J. L., Utter, F. M., and Pla, C. 1998. Population genetic structure of European hake, *Merluccius merluccius*. Heredity, 81: 327.

Rosenberg, N. A. 2003. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes, 4: 137–138.

Swan, S., Geffen, A., Moralesnin, B., Gordon, J., Shimmield, T., Sawyer, T., and Massuti, E. 2006. Otolith chemistry: an aid to stock separation of *Helicolenus dactylopterus* (bluemouth) and *Merluccius merluccius* (European hake) in the Northeast Atlantic and Mediterranean. ICES Journal of Marine Science, 63: 504–513.

Tanner, S., Vasconcelos, R., Cabral, H., and Thorrold, S. 2012. Testing an otolith geochemistry approach to determine population structure and movements of European hake in the Northeast Atlantic Ocean and Mediterranean Sea. Fisheries Research, 125–126: 198–205.

Vigier, A., Mahévas, S., and Bertignac, M. 2018. Towards a spatial integrated stock assessment model for European hake northern stock, Fisheries Research, 199: 158–170.

Westgaard, J-I., Staby, A., Aanestad Godiksen, J., Geffen, A. J., Svensson, A., Charrier, G., Svedäng, H. *et al.* 2017. Large and fine scale population structure in European hake (*Merluccius merluccius*) in the Northeast Atlantic. ICES Journal of Marine Science, 74: 1300–1310.

Handling editor: Lorenz Hauser