Marine Management Organisation

Evaluating and advancing genetic methodologies to validate the origin of capture of Angler fish (*Lophius piscatorius*) (MMO1167)

MMO1167: Evaluating and advancing genetic methodologies to validate the origin of capture of Angler fish (*Lophius piscatorius*)
December 2019

Report prepared by: A. Rus Hoelzel, Molecular Ecology Group, Department of Biosciences. Durham University,

Project funded by: European Maritime and Fisheries Fund, grant number ENG3487

Version	Author	Note
0.1	RH	First draft
0.2	CG	MMO comments
1.0	RH/CG	Final version

Information about this publication and further copies are available from:

Marine Management Organisation
Lancaster House
Hampshire Court
Newcastle upon Tyne
NE4 7YH

Tel: 0300 123 1032
Email: info@marinemanagement.org.uk
Website: www.gov.uk/mmo

Disclaimer

This report contributes to the Marine Management Organisation (MMO) evidence base which is a resource developed through a large range of research activity and methods carried out by both MMO and external experts.

The opinions expressed in this report do not necessarily reflect the views of MMO nor are they intended to indicate how MMO will act on a given set of facts or signify any preference for one research activity or method over another. MMO is not liable for the accuracy or completeness of the information contained nor is it responsible for any use of the content.

When referencing this publication, please cite as:

MMO (2019). Evaluating and advancing methodologies to validate the origin of capture of Angler fish (*Lophius piscatorius*). A report produced for the Marine Management Organisation, MMO Project No: MMO1167, December 2019, 25pp

Contents

Figures

Tables

Defintions and Acronyms

Alleles	Each of two or more alternative forms of a gene that arise by mutation and are found at the same place on a chromosome
Allozyme	Allozymes are proteins used to study genetic variation and genetic differences between populations.
Deoxyribonucleic acid (DNA)	DNA is a molecule composed of two chains that coil around each other to form a double helix carrying genetic instructions for the development, functioning, growth and reproduction of all known organisms
Effective population size (Ne)	The number of individuals that effectively participate in producing the next generation, accounting for both number of breeders and genetic diversity
Genetic drift	Random processes such as variation in family size that affect the frequency of alleles in populations
Heterozygosity	The condition of having two different alleles at a locus Ranges from 0-1 and is often used as a measure of diversity.
Locus (Loci - Plural)	A specific, fixed position on a chromosome where a particular gene or genetic marker is located
Markers	A segment of DNA used to assess genetic diversity within and among populations
Neutral	Not subject to natural selection
Putative	Commonly belived or proposed but not confirmed
Sequencing	Determining the order of the four chemical building blocks - called "bases" - that make up the DNA molecule
A Single-Nucleotide Polymorphism (SNP)	An SNP is a substitution of a single nucleotide that occurs at a specific position in the genome
Stochastic processes	Systems or phenomena that seem to change in a random way

Executive summary

In this project we applied high resolution genomic methodologies to assess the population genetic structure of the Anglerfish (*Lophius piscatorius*) in North Atlantic and North Sea fisheries management areas. The objective was to identify a set of variable genetic loci ('markers') that could identify the population of origin for individual fish. Natural patterns of movement mean that a fish born in one region may have migrated into a new region during its lifetime, and may therefore have a genetic signature indicating membership to a population other than the one in which it was caught. Although the majority may not disperse far and therefore assign correctly to their region of capture, the possibility of catching migrants and instead assigning them to their native population means that genetic assignment is best applied in concert with other non-genetic methods, such as stable isotopic or otolith mineral composition analyses.

The method we applied was the development of genomic Single Nucleotide Polymorphisms (SNPs), which have been widely used to assign individuals. In particular, we used a method that sub-samples DNA regions from across the genome at locations where a specialised enzyme (called a 'restriction' enzyme) cleaves the DNA. Called 'Restriction Associated DNA' sequencing (RADseq) the method can generate thousands of informative SNP loci. A key requirement is that tissue for DNA extraction is preserved quickly after collection from a living or very recently dead fish, so that high molecular weight DNA can be obtained.

Applying the RADseq method we identified up to 8,742 informative SNP loci, though this number was only for the set of samples for which high molecular weight DNA could be extracted. We found that many of the samples obtained had likely remained too long out of preservative after death, during which time the DNA degraded. This observation was insightful to robustness under expected real world implementation of such an approach. However, even with the inclusion of samples of lower quality (reducing the number of loci that could be identified to 1,977) it was possible to identify population structure for this species.

We used assignment and ordination (the representation of individual samples in graphical space as determined by their genetic composition) methods to cluster individuals according to their capture origins. As expected, there was some overlap consistent with migration and interbreeding, but the majority of individuals could be clearly assigned. Further analyses showed a pattern of population demographic decline over the period of the Holocene and continuing in recent times, and suggested a low level of hybridization in regions of overlap with *Lophius budegassa* (the black bellied anlgerfish, a more southerly species with overlapping distribution) though the data suggest that hybrids are likely infertile.

The objective of using genetics to identify population origin for this species is therefore likely to be successful, most useful in concert with other alternative methods. We also undertook some tests to determine the minimum number of SNP loci that may be required to provide useful assignments, and we discuss the practicalities of running these analyses for a large number of fish.

1. Introduction

As part of the Marine Management Organisation's (MMO) ongoing work towards enhancing traceability and identifying provenance, the MMO sought an assessment of methods for confirming where fish sold ashore were originally caught to potentially support and verify other systems for managing compliance and enforcement. This report concerns genetics, one of four technique classes previously identified as promising by the MMO along with lipids, trace elements and stable isotopes.

Natural populations differentiate genetically over time when connectivity between them is reduced or differential adaptation is strong. Differentiation may be due to genetic drift (stochastic processes leading to the differential loss or retention of alleles in isolated populations) or by natural selection. Natural selection can maintain different genotypes (the genetic constitution at a given locus) that are adapted to local environments despite gene flow and low or no differentiation at neutral loci. If loci can be identified that have differential allele frequencies in regional populations (by either mechanism), then an assignment method can be devised to match sample to population of origin. The precision of origin assignment depends on the level of genetic differentiation and the number of markers applied, and can be estimated empirically or using simulations.

The focus of this study was on the Anglerfish (*Lophius piscatorius*). *L. piscatorius* is distributed especially in the eastern North Atlantic waters around Iceland and the UK, in the North Sea and along the Scandinavian coasts through the Norwegian Sea and the Barents Sea. *L. piscatorius* is relatively rare in the Mediterranean Sea where another species in the same genus, *L. budegassa*, dominates. The two species overlap in their distributions especially in the eastern North Atlantic to the southwest of Ireland and the UK.

L. piscatorius is a highly commercial fish, taken by bottom trawls and gill nets and marketed both fresh and frozen. From the FAO (Food and Agriculture Organisation of the United Nations) catch list the take in 2011 [1]was 1,378 tonnes from Denmark, 2,270 from the Faroe Islands, 3,227 from Iceland, 3,794 from Ireland, 5,695 from Norway and 15,115 from the UK. There were also relatively minor catches from Portugal (173 tonnes) and the Netherlands (61 tonnes).

Little is currently known about stock structure of *L. piscatorius*. An early study based on 12 allozyme loci (enzyme electrophoresis) suggested some differentiation between Scotland and the Irish Sea (Crozier 1987), however a later study based on 9 microsatellite DNA loci found no structure among ICES (International Committee Exploration of the Sea) management areas (used for quotas) IVa (Northern North Sea), VIa (West of Scotland), VIb (Rockall) and VIIb (West of Ireland) (no information is provided on sample sizes). This latter study was a student project from the Piertney lab at the University of Aberdeen that was not fully published, but did appear as a published abstract (O'Sullivan et al. 2006).

[1] Supplied via Fishbase (2019)

Life history characteristics of *L. piscatorius* suggest the potential for local adaptation and differentiation by genetic drift. For example, the species is a benthic ambush predators feeding at a relatively high trophic level (including large fish and seabirds), and live in a variety of coastal habitats (sandy or muddy substrate, sometimes in kelp), and at depths of up to 1000m. Natural selection may be involved in maximising fitness for life in different habitats (as found for the roundnose grenadier, *Coryphaenoides rupestris*; Gaither et al. 2018) or in association with different local prey resources. Eggs are highly aggregated in a gelatinous 'veil', but eggs and larvae can drift for several weeks to several months (Russell 1976) facilitating gene flow (which could be reflected in measures of connectivity using neutral markers), but may also generate structure dependant on local current systems.

Our objective was to use high-resolution genomic methods to look for evidence of structure and any differential patterns when neutral loci are contrasted with loci putatively under natural selection. We then wanted to refine the analysis to find a minimum number of loci that would identify a fish to source population, in order to facilitate the screening process and reduce costs.

2. Methodology

The samples received for further analysis are shown in Table 1. DNA was extracted by standard protocols using phenol-chloroform (see Hoelzel 1998). DNA quality was assessed by visualisation on an agarose gel, and using a spectrophotometer and a Qubit fluorimeter. Quality control was applied to select only those samples with sufficient high molecular weight DNA.

Table 1- Samples included at the start of the project. Numbers lower than 30 were limited by availability, while the cap of 30 was associated with the available budget.

Area	Location	Number
VIa	Hebrides	30
IVb	NE England	30
IVa	Shetland/Orkney	30
VIIj	SW Ireland	30
VIIh	Cornwall	30
VIIg	Celtic Sea	7
Va	Iceland	26
Total		183

Trial DNA restriction enzyme digestions confirmed that a combination the research group had used for fish species earlier, MluCI and SphI-HF, would be suitable for this species. Library construction followed the double digest restriction associated DNA method described in Peterson et al. (2012). Briefly, DNA digested with one frequent (MluCI) and one less frequent (SphI-HF) cutting restriction enzyme is ligated to adaptors that include an index sequence at one end, barcode at the other, and Illumina sequence primer sites. Pools of samples uniquely barcoded (to later identify individual samples bioinformatically) were size selected on a PippinPrep (a

machine which facilitates DNA library construction), checked for size distribution on a TapeStation (an automated electrophoresis tool for DNA and RNA sample quality control), and the DNA quantity per pool checked and standardised. Combined pools representing the library were checked for quantity and quality by qPCR, and the concentration adjusted for sequencing on an Illumina HiSeq 2500. Sequencing was undertaken on two lanes of the HiSeq sequencer applying paired end 125bp sequencing.

SNPs were identified using the software package Stacks2 (Catchen et el. 2011), which filters out low quality reads, identifies SNPs, and generates output genotype data that can be applied to various types of analyses. Quality filters were set at a minimum read depth of 4-fold, and 75% retention of individuals per putative population. For most analyses only samples with greater than one million (1M) reads were retained.

Four different approaches were tried for the identification of outlier loci putatively under natural selection, with the objective of choosing the method that identified the largest number for further assessment. These were: *PCAdapt* (Duforet-Frebourg *et al.* 2014), *BayesEnv2* (Günther & Coop 2013), *Lositan* (Antao *et al.* 2008) and *BayeScan* (Foll & Gaggiotti 2008). The objective was to find those loci most useful for population assignment. We focused primarily on ordination methods for various subsets of the data. This included factorial correspondence analysis (FCA) run in Genetix (Belkhir et al. 2002), principal component analysis (PCA) and Discriminant Analysis of Principal Components (DAPC; Jombart *et al.* 2010). The method best identifying clusters associated with management areas was then chosen for analysis on multiple subsets of the data to identify a minimum number of informative loci. An assignment method based on minor allele frequencies (after Paetkau et al. 1995) set in a Bayesian context (after Prichard et al. 2000) was also applied (see deJong 2019). Historical demography was assessed using the 'stairway' plot method as described in Liu & Fu (2015).

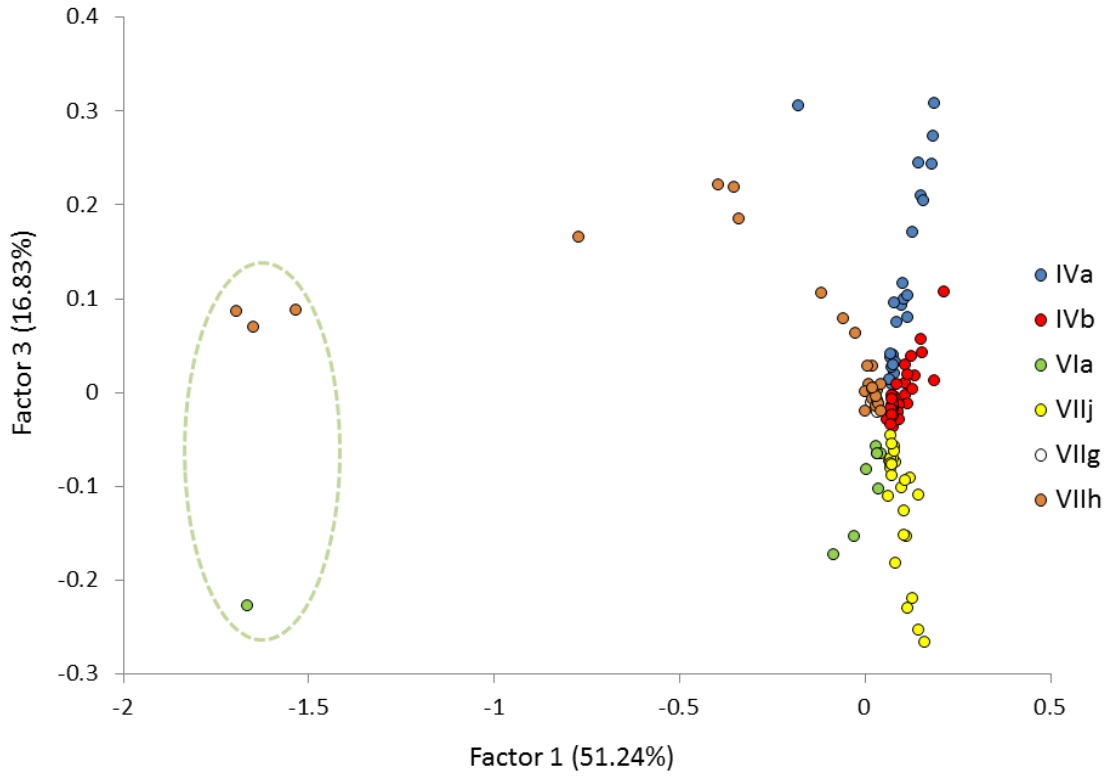
3. Results

After initial analysis based on visual assessment by electrophoresis, spectrophotometric tests of quality based on 260/280 ratios, and fluorescent quantification it was clear that the quality of samples received were highly variable. The issue was degradation, likely associated with the time that elapsed between capture and preservation. Therefore although 180 of the 183 samples were included in further processing, we were aware that some of the more degraded samples may provide fewer sequence reads than the better preserved samples. All were included because it was not possible to know in advance what degree of degradation would be too much. Degradation information would also be informative for any further implementation of the approaches used here. The management regions with the largest proportion of highly degraded DNA were areas Va and VIa.

After de-multiplexing and processing, only one sample from area Va had sufficient read depth for further analysis, and so that putative population was not included for any further assessments. Samples from area VIa had reduced read depth, but a subset were sufficient for SNP detection at a reduced level. It was therefore decided to move forward with two analyses, one including and one excluding samples from

area VIa. When six regions were retained (IVa, IVb, VIa (problematic), VIIj, VIIg and VIIh), 1,977 SNPs could be identified. The sample size from area VIIg was small (Table 1), and after initial ordination analyses (see Figure 1) it was evident that the geographically proximate areas VIIg and VIIh could not be differentiated based on available materials, and therefore they were combined for further analyses. After filtering for quality control, there were 24 samples used from IVa, 29 from IVa, 9 from VIa, 25 from VIIj, and 31 from VIIg&h. Four samples (within the dashed circle in Figure 1) showed unexpectedly high heterozygosity (see below), and are possible F1 hybrids of the two *Lophius* species.

Figure 1 - FCA analysis based on all 1,977 SNPs from the five-population analysis



Samples from the five putative populations could be separated by FCA analysis (Figure 1), though F_{ST} values were low and only the comparison between IVa and VIIgh was significantly greater than zero (Table 2 comparisons excluded potential hybrids).

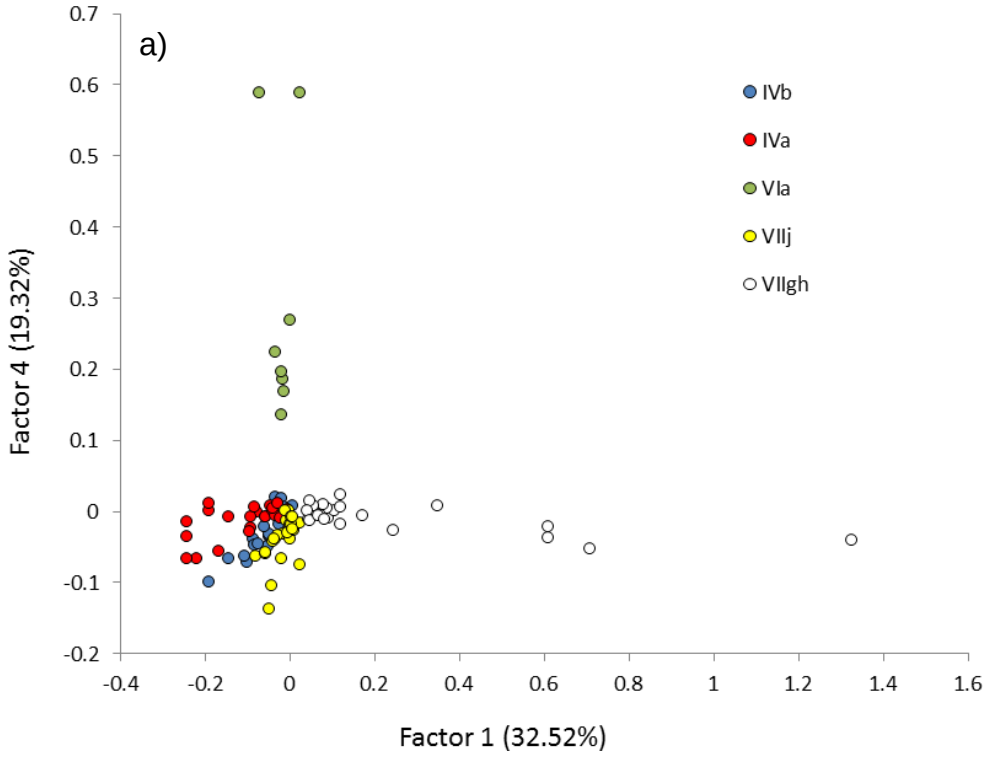
Table 2 - F_{ST} values for pairwise comparisons among management areas (bold = p≤0.05 after Bonferroni correction)

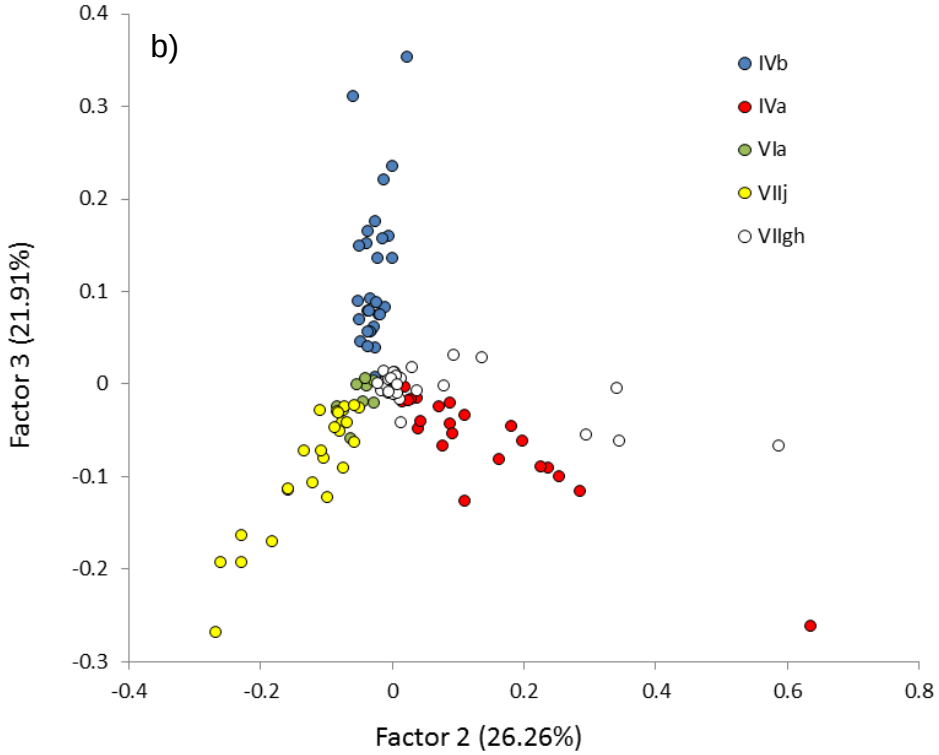
	IVb	IVa	VIa	VIIj
IVa	-0.00230			
VIa	0.00065	-0.02105		
VIIj	0.00488	0.00334	0.00225	
VIIg&h	0.00560	**0.00876**	-0.00289	0.00186

Further ordination analyses using DAPC and PCA resolved the clusters less well (data not shown), and so we focused on FCA for further analyses. We removed the

potential hybrid samples and re-assessed the FCA analyses (**Error! Reference source not found.**). All five clusters could then be better resolved, though it required two different combinations of factor comparisons to fully resolve all five putative populations (see **Error! Reference source not found.**).

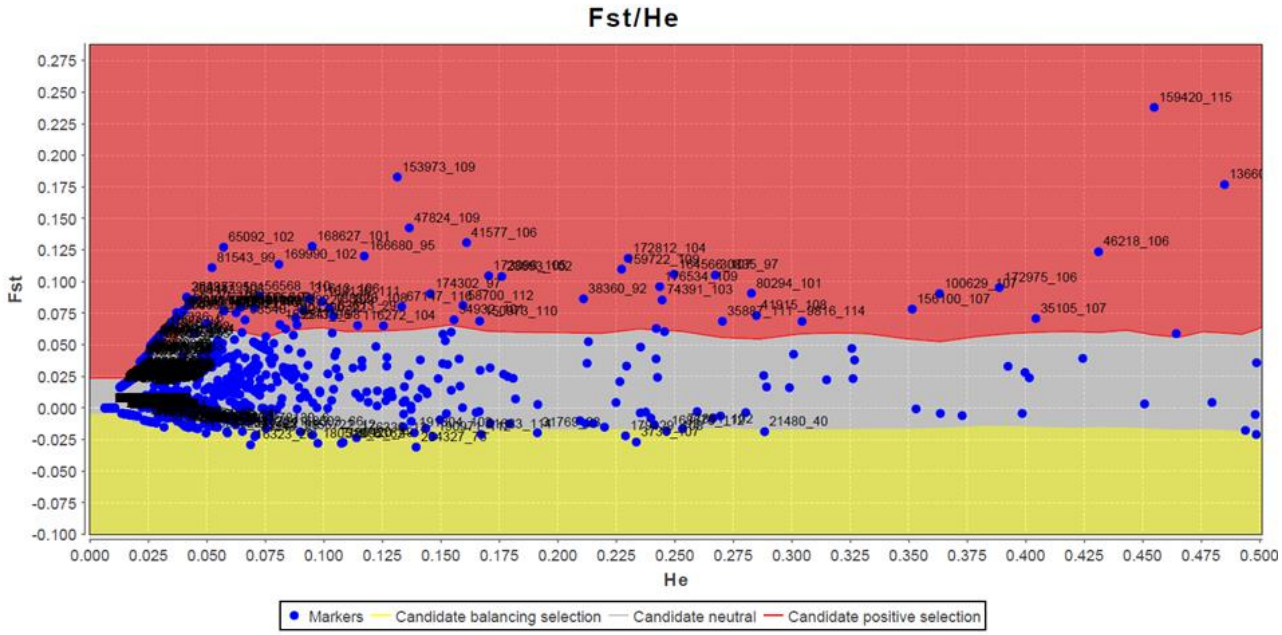
Figure 2- FCA analysis with the putative hybrids removed, and considering two different factor comparisons. a) VIa and VIIgh are relatively distinct; b) IVa, IVb and VIIj are distinguished.





The ICES management areas could therefore be differentiated based on these 1,977 SNP loci, however this is a large number for screening unless a SNP chip were to be developed. We therefore used methods to identify outlier loci, putatively under natural selection, to identify a smaller set of informative loci. From the 1,977 loci applying the four approaches cited in the methods, only one method identified enough loci to potentially provide sufficient resolution. This was the Fdist method implemented in the program Lositan, which identified 166 outlier loci (**Error! Reference source not found.**). The method compares diversity among putative populations against diversity within.

Figure 3 - Identification of outlier loci for the five-population analysis (1,977 SNPs in total) based on neutral expectations, using the Fdist method as implemented in Lositan. F_{ST} = Wright's population level inbreeding coefficient (a measure of differentiation), and He = within population diversity.



When these 166 loci were used to differentiated regions in an FCA analysis, it was possible to identify some distinct clustering, but there was also overlap (**Error! Reference source not found.**). To explore the clustering potential when there was a larger number of SNPs, a second analysis was run including just the four putative populations with high quality read depth data (areas IVa, N=24, IVb, N=29, VIIj, N=25 and VIIg&h, N=31). This run generated 8,742 SNPs. The full 8,742 SNP dataset was initially analysed using the Bayesian clustering method based on minor allele frequency, and this is illustrated in **Error! Reference source not found.**. This method had not provided a clear pattern of clustering when the 1,977 SNP data set was used (data not shown), but for 8,742 SNPs most individuals were assigned to the correct management area, with 2-3 exceptions in each area.

Figure 4 - FCA analysis based on 166 outlier loci (putatively under positive natural selection) identified by the Lositan analysis, comparing factors 1 and 2 (a) and 1 and 3 (b).

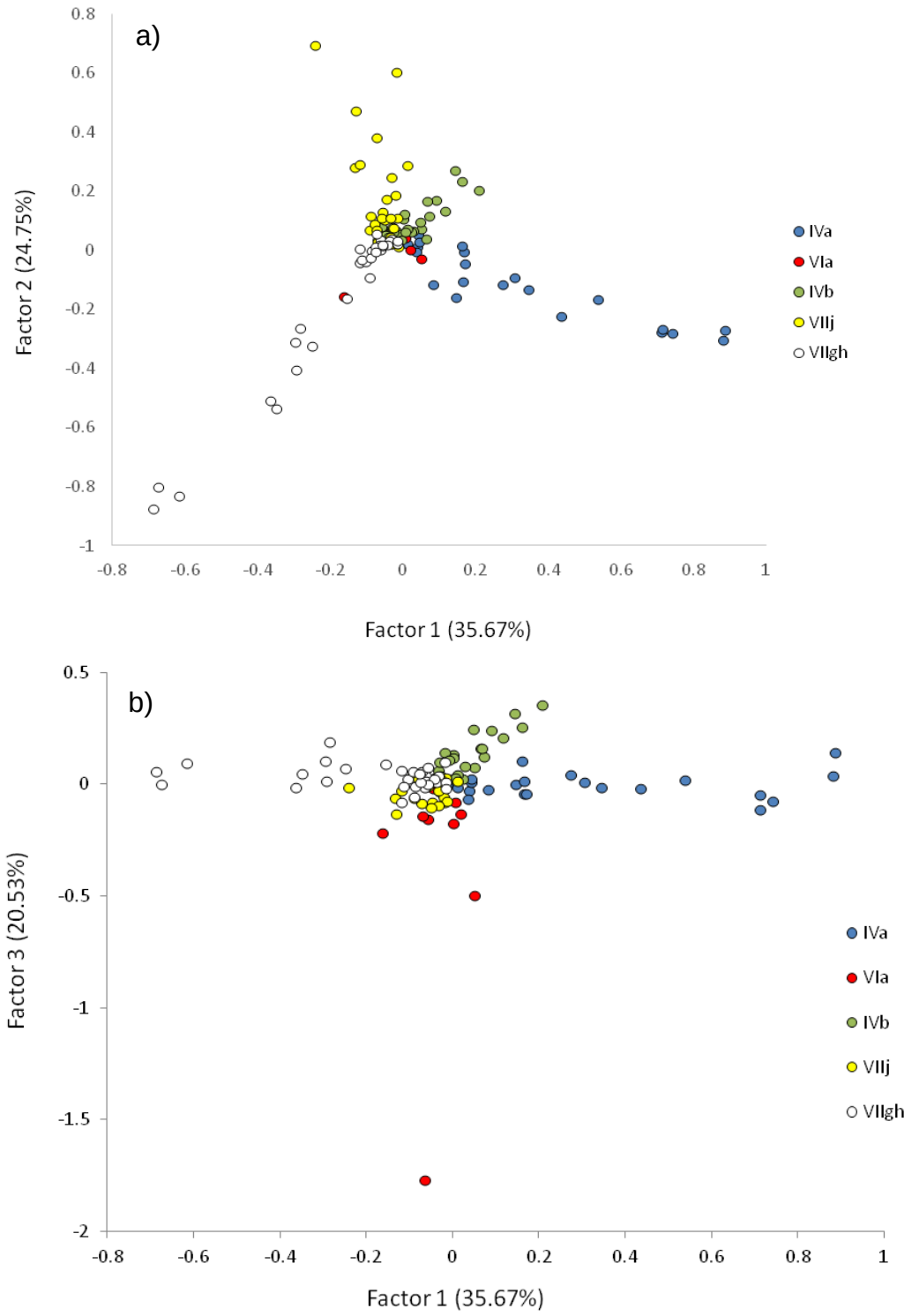
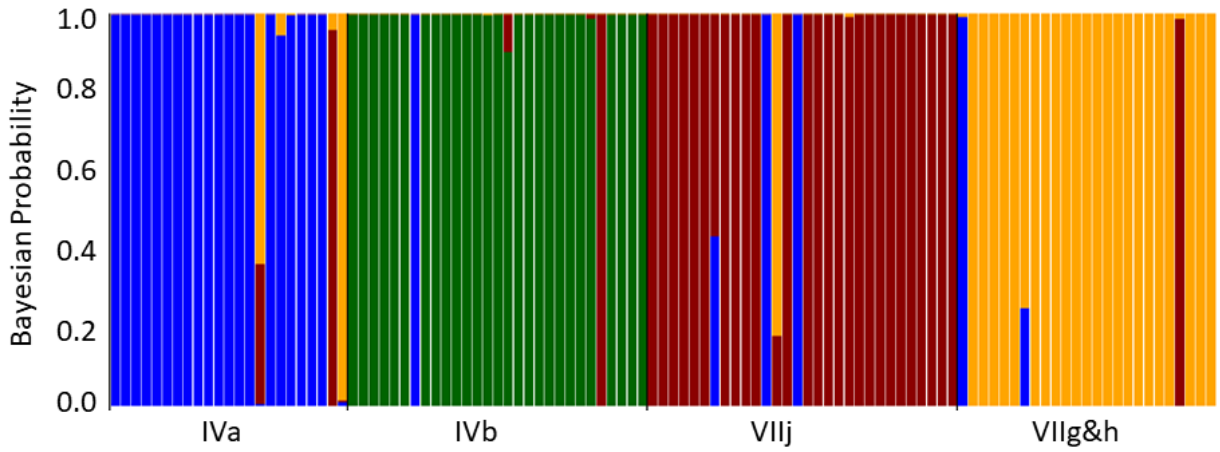
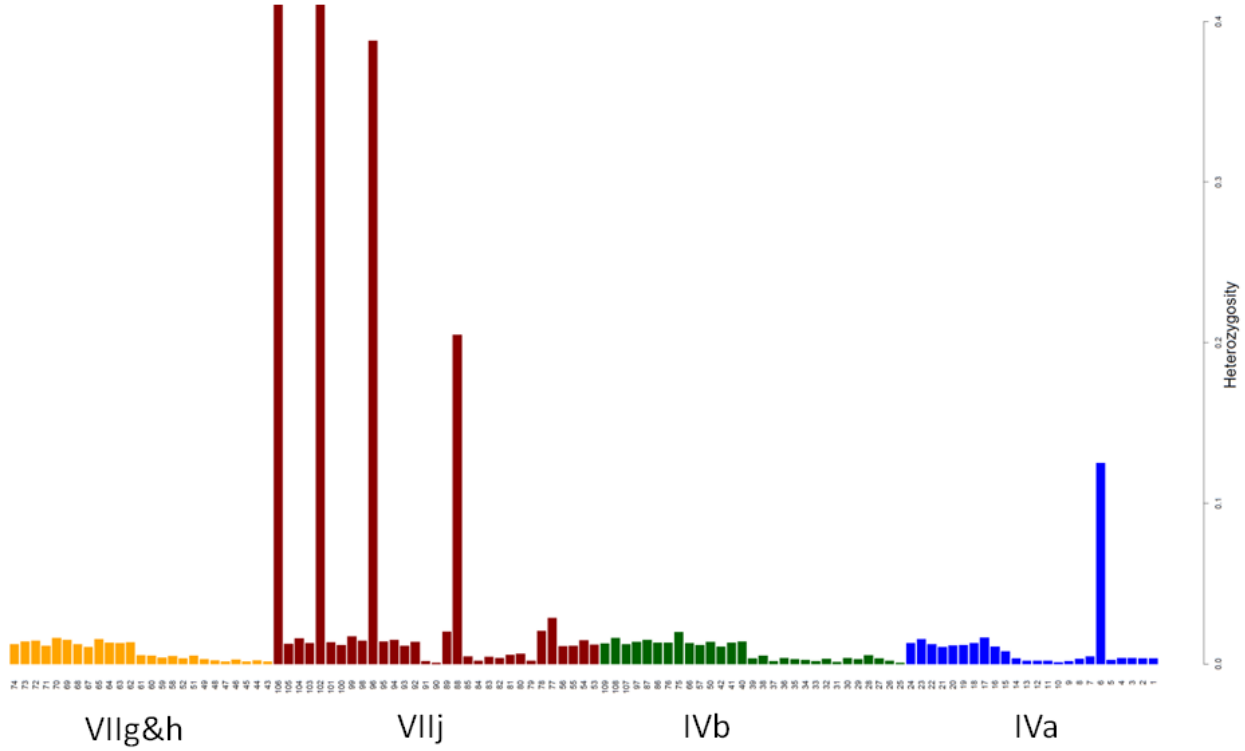


Figure 5 - Bayesian clustering method based on minor allele frequencies (see methods). Each line represents an individual fish, and the colours represent indentified clusters.



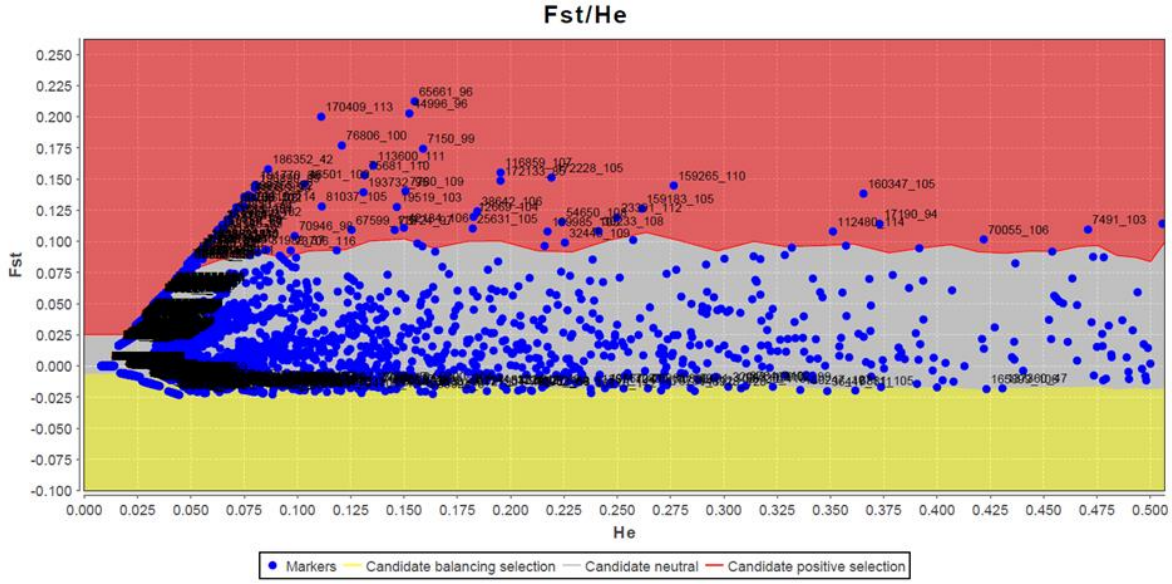
This 4-population dataset was used for more intensive assessments of the diversity and demographic history of putative populations, taking advantage of the higher resolution provided by 8,742 SNPs. The average heterozygosity (low heterozygosity equates to little genetic variability) was 0.03230 for area IVa, 0.04007 for area IVb, 0.03162 for area VIIgh, and 0.07983 for area VIIj. Levels were similar among most individuals apart from a few exceptions, especially in area VIIj (**Error! Reference source not found.**). One possible explanation for the exceptional individuals seen in VIIj would be the presence of F1 hybrids (see discussion).

Figure 6 - Mean heterozygosity for each sample, colour-coded by management area.



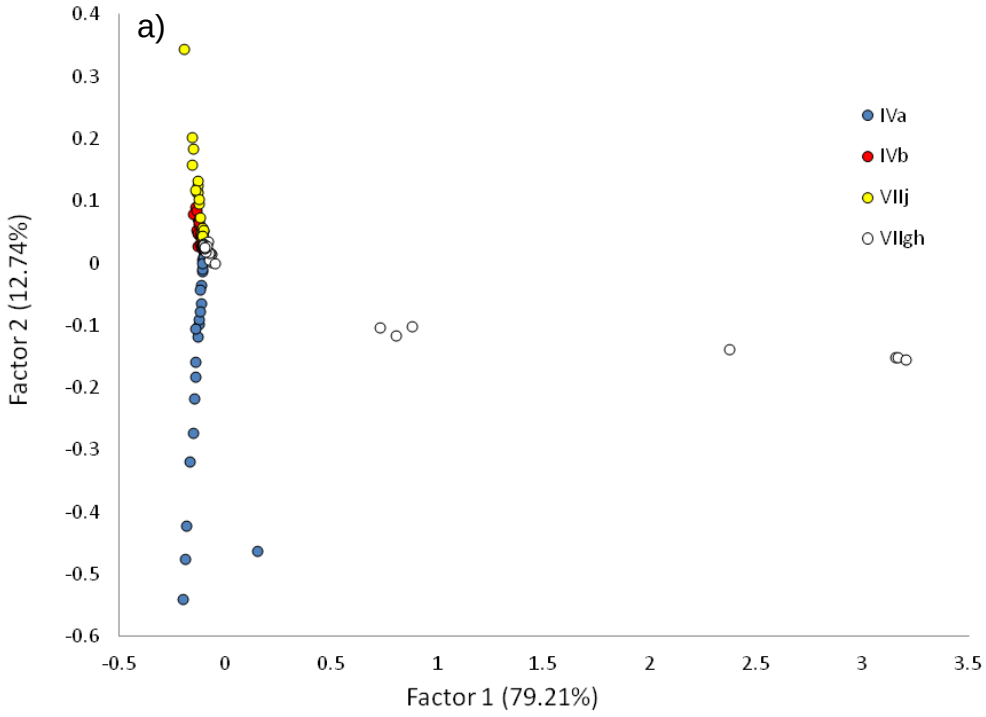
Using the Fdist method we then identified loci putatively under natural selection in the 8,742 SNP dataset, generating 2,459 outlier SNPs (**Error! Reference source not found.**).

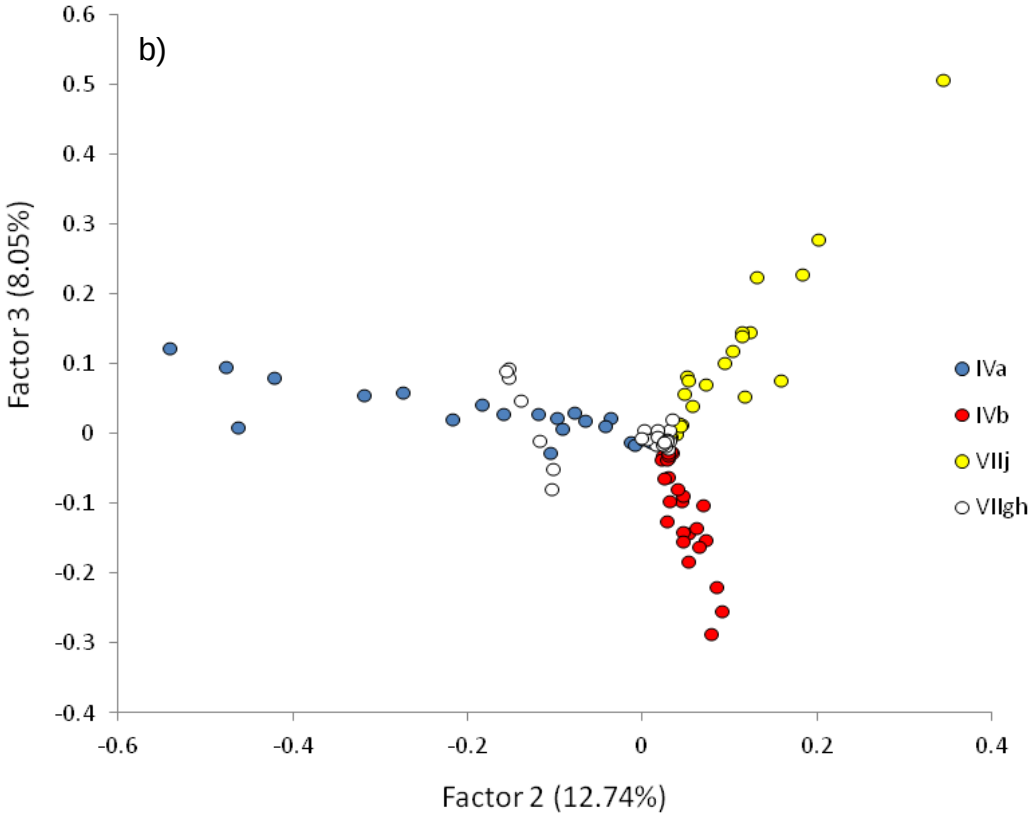
Figure 7 - Identification of outlier loci for the 4-population analysis (8,742 SNPs in total) based on neutral expectations, using the Fdist method as implemented in Lositan.



Although there are clusters identified, overlap is not greatly reduced compared to the alternatives based on the five-population analyses (**Error! Reference source not found.**).

Figure 8 - FCA based on four populations and 2,459 Lositan outlier loci comparing factors 1 and 2 (a) and 2 with 3 (b).





Much of the separation is, however, retained at a similar level when just the strongest 288 or 96 outliers are applied (**Error! Reference source not found.**). The 4-population analysis provided enough power to generate good estimates of historical demography for each population using the stairway plot method (see methods). Although neither the average mutation rate nor the generation time are known for this species, a mutation rate of 2.5×10^{-8}/ site/ generation can be estimated from a review of studies (Lynch 2010), and based on the age of maturation of female anglerfish (Duarte et al. 2001) we estimated the generation time at 14 years. Applying those parameters gives the results illustrated in **Error! Reference source not found.**.

Figure 9 - FCA analysis comparing factors 2 and 3 for 288 SNPs (a) and 96 SNPs (b), in each case the strongest outliers after Lositan analysis.

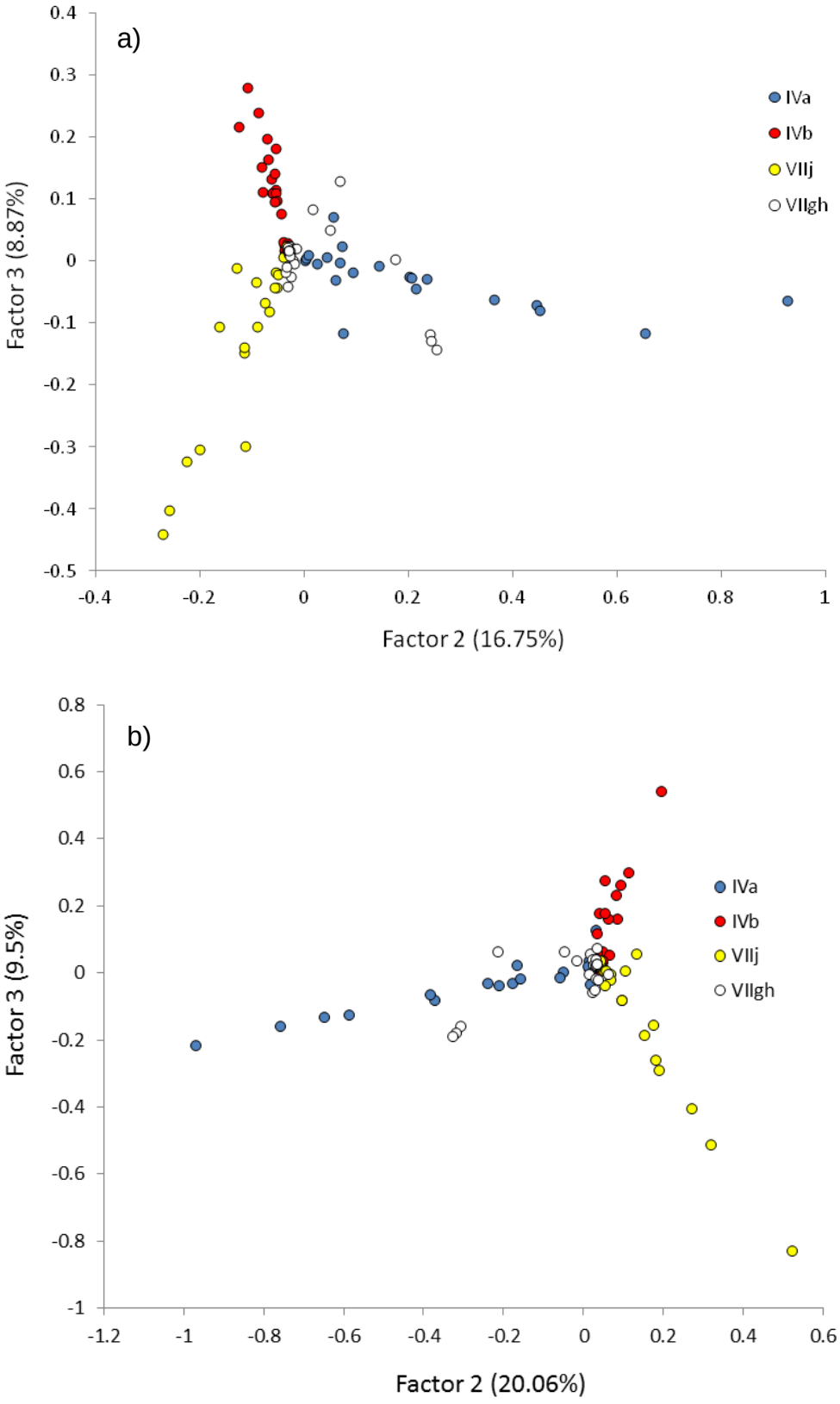
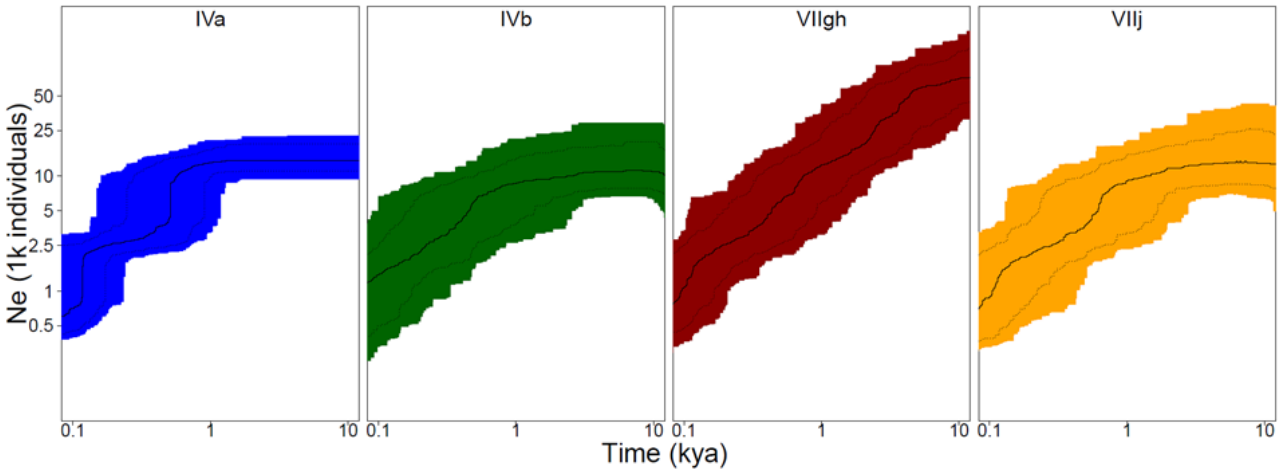


Figure 10 - Stairway plots for each of the putative populations from the 4-population analysis showing the estimated demographic histories and effective population sizes. Kya = thousands of years ago.



4. Conclusions

The 1M digest restriction associated DNA sequencing method provided nearly 9000 SNPs for an analysis based on the best quality samples, and nearly 2000 SNPs when poorer quality samples were retained (to permit representation of a fifth population, management area VIa). Population structure was found, separating all five included populations (based on the quota ICES areas) by ordination analysis, though the clusters were not widely separated in Euclidian space (see distribution of individual samples as represented by points in Figure 1), and F_{ST} values among putative populations were low, suggesting ongoing or very recent gene flow.

The best resolution was achieved when full datasets were analysed (see Figure 1, **Error! Reference source not found.**& **Error! Reference source not found.**), however in each case there was evidence for potentially recent migrants and admixture meaning that the assignment of new individuals would need to be in that context. Apparent migrants were relatively infrequent (~9% estimated from the analysis presented in **Error! Reference source not found.**), and so the analysis of multiple individuals from a given catch (haul) should greatly reduce the chance of misidentification of the catch location (e.g. if the chance of finding one migrant is 9%, all of a sample of 10 would be found to be migrants with a probability of 3.5×10^{-11}, or 5 of the 10 at a probability of 5.9×10^{-6}).

A SNP chip could be developed to screen individuals for the larger numbers of SNPs, and from our analyses the 1,977 SNPs from the 5-population analysis would be sufficient for this, potentially assigning individual fish with high accuracy. However, it is preferable and often cheaper to screen a smaller number of SNPs. Although there are various methods available for this including Fluidigm arrays[2] and Taqman qPCR (Shen et al. 2009), a relatively efficient and affordable method is

[2] https://www.fluidigm.com/ products/biomark-hd-system

‘genotyping in thousands’ (GT-seq; Campbell et al. 2015). This is because the equipment required for screening is widely available (standard PCR machine and Illumina MiSeq), and the method is very cost effective. Briefly, more than 100 loci can be multiplex amplified in a PCR reaction with independent barcoding (to identify individual fish), and then sequenced relatively inexpensively on the Illumina MiSeq sequencer.

To facilitate this we assessed outlier loci putatively under natural selection, using the method that provided the largest number of outliers, the Fdist method (implemented in Lositan – see Methods). This method errs on the side of also including outliers that may be due to strong genetic drift, but in the context of population assignment, these may also be informative. In an earlier study with a deep sea fish species – Orange roughy (*Hoplostethus atlanticus*) we found that putative natural selection outliers identified a particular population, highly differentiated from the rest (Goncalves da Silva et al 2019). In another study on species of deep sea sharks, most comparisons showed exaggerated differentiation when natural selection outliers were analysed (Keggin 2017). However, every case is different.

Here we find that the proportional degree of differentiation among putative populations did not significantly change for markers putatively under natural selection compared to the full dataset, nor was there a very strong effect of increasing the level of resolution. However, the increase in resolution was sufficient to allow for a smaller number of loci to be applied for screening, though the resolution remained highest with the full datasets.

The technology readiness level (TRL) (as defined by the European Research Council (2014)) based on the work done so far likely increased from TRL1 (a theoretical idea) to TRL6 (Technology demonstrated in relevant environment) or TRL 7 (System prototype demonstration in operational environment). Further understanding of the potential for other markers (such as otolith chemistry and stable isotopes) to assign individual fish to population of origin would help determine the degree of further resolution required from the genetic screening. The most cost-effective approach would be to multiplex ~96 outlier loci (see **Error! Reference source not found.**) and screen individuals by amplicon sequencing. The cost per sample, however, would depend on the number of samples screened. Given that some individuals from a given location overlap with individuals from other locations in ordination plots, multiple samples from a given catch would greatly increase the power of assignment.

If samples from 8 fish were taken from each catch to be analysed, then 12 catch analyses would fit into a 96-well plate, and two plates could be included in a single-end 150bp run on a MiSeq lane. The cost of the lane would be ~£900, so the sequencing cost would be £37.50 per catch. Additional costs to cover PCR amplification and associated consumables would bring the cost per catch to £60. This would require initial investment in PCR primers that would then last for a large number of subsequent analyses. Further funding would be required to cover technician time, but the cost per unit would depend on the volume of work.

Our analyses based on the best quality samples among four putative populations (generating 8,742 SNP loci) provided two further observations of relevance to effective management. First, in a region where the ranges of *L. piscatorius* and *L.*

budegassa overlap most extensively (see Charrier et al. 2006), there was evidence for the presence of F1 hybrids (see **Error! Reference source not found.**). However, this would need to be confirmed by genotyping *L. budegassa* at the same loci and comparing confirmed pure line *L. piscatorius* and *L. budegassa* with the putative hybrids. Hybrid zones among congeneric fish species are not uncommon though, even with hybridisation at similar frequencies as suggested here, though the F1 hybrids are often sterile (e.g. Crespin et al. 2002). The pattern we see for anglerfish would be most readily explained by a hybrid zone and sterile hybrids, since there is no strong differentiation between the affected population and other regions after the putative hybrids are removed from the analysis (see Table 2).

The second observation is about the demographic histories of these putative populations, as illustrated in **Error! Reference source not found.**. Early Holocene (starting 11,700 years ago) values of effective population size (Ne) were around 10,000 for three of the four regions, which is consistent with some other deep sea species (e.g. White et al. 2011) and the general observation that the apparent ratio between Ne and the census population size (Nc) is small for these species (meaning that diversity will be lost faster than expected based on the census population size; see Hare et al. 2011). The exception was area VIIgh which appears from this analysis to have had a larger historical population size. The effective population size reflects the level of diversity available for evolutionary change, and is associated with the rate at which diversity would be lost over time (see Hare et al. 2011). For each population there is a more recent population decline, suggested to be starting about 1000 years ago (and earlier for area VIIgh), but the timing will be imprecise due to our limited knowledge of the true mutation rate and generation time. The analysis suggests a decline starting ~40 generations ago in most regions, and the drop is nearly an order of magnitude. Given the timing, environmental factors may be important, though recent declines could have been amplified by fisheries.

5. Future work & recommendations

This study illustrates that the objective of screening anglerfish for the identification of natal population (place of birth) and provenance is feasible at the resolution of the ICES areas resolved here. Implementation should ideally be founded on larger sample sizes to more fully establish the reference database and refine the screening protocol. Available samples should be more inclusively extracted for DNA (there are 100s of samples already available), and a more severe quality screen be applied for the inclusion of samples in the analysis. High molecular weight samples should be included and further samples acquired if necessary so that a broader geographic range could be represented by 30-50 samples per region. However, if funding is limited the existing data presented here may be sufficient to establish the screening protocol, though more reference samples (broader geographic representation) would be desirable. A protocol would need to be established for all samples to be collected for future screening against the database, such that the time between death of the fish and sample collection is minimized (down to no more than a matter of minutes if possible), and the samples are immediately immersed in an appropriate preservative (either 85% ethanol or 20% DMSO (Dimethyl sulfoxide) saturated with NaCl (sodium chloride)). It appears that 100-200 of the SNP loci identified here will be sufficient, especially if used in conjunction with other markers such as stable isotopes. At the same time, further investment in development may increase the resolution. Routine

screening will be possible using methods such as GT-seq or Fluidigm allowing high throughput at relatively low costs. Although there are various alternatives, we recommend the following steps to move the TRL of the approach towards an operational level for MMO use:

- Aim to sample 8-12 fish from a given catch, and ensure that the samples are collected into DNA preservative (e.g. 85% ethanol) as a fin clip from living fish, or collected immediately post-mortem.

- Start with the 96 loci illustrated in **Error! Reference source not found.**(b) for the initial screening, and use the GT-seq method described in the discussion.

- If possible, undertake some further development using a broader geographic range of samples, only high molecular weight DNA, and the same protocol as described here (which will allow existing data to be incorporated into the new analyses).

- Use ordination methods, such as the FCA analyses undertaken in this study, to assign individual fish to clusters. Accept assignment when a majority of the 8-12 fish obtained from the trial catch assign to a given population.

- Screen samples of *L. budegassa* at the same loci to further explore the extent and range of hybridisation between *L. budegassa* and *L. piscatorius*.

6. References

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F-st-outlier method. *BMC Bioinformatics* **9**:323.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Catch F (2002) GENETIX 4.04, software under Windows TM for the genetics of the populations. Laboratory Genome, Populations, Interactions, CNRS UMR 5000, University of Montpellier II, Montpellier (France).

Berg PR, Star B, Pampoulie C, Bradbury IR, Bentzen P, Hutchings JA, Jentoft S, Jakobsen KS (2017) Trans-oceanic genomic divergence of Atlantic cod ecotypes is associated with large inversions. *Heredity*, **119**, 418-428

Campbell NR, Harmon SA, Narum SR (2015) Genotyping-in-Thousands by sequencing (GT-seq): A cost effective SNP genotyping method based on custom amplicon sequencing *Molecular Ecology Resources* **15**, 855-867

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171-182

Charrier G, Chenel T, Durand JD, Girard M, Quiniou L, Larache J (2006) Discrepancies in phylogeographical patterns of two European anglerfishes (*Lophius budegassa* and *Lophius piscatorius*). *Molecular Phylogenetics and Evolution* **38**, 742-754

Crespin L, Berrebi P, Lebreton JD (2002) Spatially varying natural selection in a fish hybrid zone. *Journal of Fish Biology* **61**, 696-711.

Crozier WW (1987) Biochemical genetic variation and population structure in angler-fish *Lophius piscatorius* L. from the Irish Sea and west of Scotland. *Journal of Experimental Marine Biology and Ecology* **106**, 125-136.

Cuveliers EL, Larmuseau MHD, Hellemans B, Verherstraeten SLNA, Volckaert FAM, Maes GE (2012) Multi-marker estimate of genetic connectivity of sole (*Solea solea*) in the North-East Atlantic Ocean. *Marine Biology* **159**,1239–1253

Duarte R, Azevedo M, Landa J, Pereda P (2001) Reproduction of anglerfish (*Lophius budegassa* Spinola and *Lophius piscatorius* Linnaeus) from the Atlantic Iberian coast. *Fisheries Research* **51**, 349-361.

Duforet-Frebourg N, Bazin E, Blum MGB. (2014) Genome scans for detecting footprints of local adaptation using a Bayesian factor model. *Molecular Biology and Evolution* **31**, 2483–2495.

European Research Council (2014). Technology readiness levels (TRL). Available at: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf [last accessed June 2019].

Fishbase (2019) FAO Capture Production of Lophius piscatorius. Available at (http://www.fishbase.org/report/FAO/FAOCatchList.php?scientific=Lophius+piscatorius) [Last accessed December 2019].

Foll M, 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.

Gaither MR, Gkafas GA, de Jong M, Sarigol F, Neat F, Regnier T, Moore D, Gröcke DR, Hall N, Liu X, Kenny J, Lucaci A, Hughes M, Haldenby S, Hoelzel AR (2018) Genomics of habitat choice and adaptive evolution in the deep sea. *Nature Ecology & Evolution* **2**, 680–687

Goncalves da Silva A, et al. (2019) Evolutionary incompatible fish populations for conservation in the deep sea. *Journal of Applied Ecology* In press

Günther T, Coop G. (2013) Robust identification of local adaptation from allele frequencies. *Genetics* **195**, 205–220.

Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples RS, Ruegg K, Palstra F (2011) Understanding and estimating effective population size for practical application in marine species management. *Conservation Biology* **25**,438–449

Hoelzel AR. Ed. (1998) Molecular Genetic Analysis of Populations; a Practical Approach. 2nd Edition. Oxford University Press, Oxford.

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

Keggin T (2017) Population genomics of deep sea sharks species. MSc Thesis, Durham University.

Liu XM, Fu YX. 2015. Exploring population size changes using SNP frequency spectra. *Nature Genetics* **47**, 1099-1099.

Lynch M (2010) Evolution of the mutation rate. *TREE* **26**, 345-352.

Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Thomas Moen, Torild Johansen, Christian Mittelholzer, Geir-Lasse Taranger, Rob Ogden, Gary R Carvalho (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology* **9**:276.

O'Sullivan M, Wright PJ, Verspoor E, Knox D, Piertney S (2006) Absence of spatial and temporal genetic differentiation at microsatellite loci in north east Atlantic anglerfish (*Lophius piscatorius*). *Journal of Fish Biology* **69**, 261

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model Species. *PLoS ONE*, **7**, e37135

Russell, FS., (1976) The eggs and planktonic stages of British marine fishes. Academic Press, London, UK. 524 p.

Shen GQ., Abdullah KG, Wang QK. (2009) The TaqMan Method for SNP Genotyping. In: Komar A. (eds) Single Nucleotide Polymorphisms. Methods in Molecular Biology™ (Methods and Protocols), vol 578. Humana Press, Totowa, NJ

White TA, White TA, Stefanni S, Stefanni S, Stamford J, Stamford J, Hoelzel AR, Hoelzel AR. (2009) Unexpected panmixia in a long-lived, deep-sea fish with well-defined spawning habitat and relatively low fecundity. *Molecular Ecology* **18**:2563–2573.

White TA, Fotherby H, Hoelzel AR (2011) Comparative assessment of population genetics and demographic history of two congeneric deep sea fish species living at different depths. *Marine Ecology Progress Series* **434**, 155-164