A working document on the genetic Stock Identification of horse mackerel, *Trachurus trachurus* for the ICES Benchmark workshop on horse mackerel and boarfish stocks (WKBHMB)
Version 4.0, 25th January 2024

Edward D. Farrell[1], Angela P. Fuentes-Pardo[2], David S. Murray[3], Leif Andersson[2]

1. Summary

This working document presents a thorough review of the history and origin of the stock identification of horse mackerel, which highlights the significant uncertainties in the current delineation of the stocks. A new genetic based stock identification approach is also presented, which provides a robust method for defining the biological units (referred to as "populations" from this point) that occur within the stock areas and enables the assignment of individuals to population of origin.

The historical classification of three horse mackerel stocks was initially based on the recognition of three potential spawning areas and the assumption that these may represent spawning grounds for discrete populations. On review of the early working group reports and scientific literature it is clear that there was little empirical evidence to support the delineation of the three stocks. Regardless, the initial stock delineation was largely retained and shaped the subsequent direction of the data collection and stock assessments of the three stocks until it was challenged by the results of the HOMSIR project. Whilst some changes were made to the stock areas based on the HOMSIR results, the project noted that the population structure in the western European waters could be more complicated than the results suggested and that more research was needed to clarify the migration patterns within the Northeast Atlantic. This was particularly relevant to the potential mixing areas between the three stocks, however little further work was conducted and the provisional stock boundaries suggested by HOMSIR have largely remained in place to the present day.

The current study presents the most comprehensive investigation of horse mackerel stock structure in the northeast Atlantic area to date using the most advanced methods available. The western horse mackerel population appears to have the widest distribution and ranges from division 4.a in the north, division 3.a in the east and south into division 9.a. Based on the samples analysed this population spawns to the west and southwest of Ireland, in the Bay of Biscay along the Northern Spanish Shelf and in Portuguese waters. It also occurs in divisions 7.e and 7.d in significant numbers and may also be present in divisions 4.b at certain times of the year.

The horse mackerel that spawn in the southern North Sea are a locally adapted biological unit. Based on the samples analysed this population has a limited distribution and occurs primarily in divisions 4.b and 4.c. It also occurs in division 7.d where it mixes with the western population. It was also recorded as the minority component in samples from division 7.e and in very small numbers in samples from division 4.a.

The southern population was the least well sampled in the current study. A very small number of spawning individuals (<10) were collected in the south of division 9.a. These individuals were characteristic of the southern population, which is more closely related to the north African population than to the western population.

An assignment model was developed that can be used to distinguish individuals from the western and North Sea populations with greater than 90% accuracy. Whilst it was not possible to develop an assignment model to distinguish the southern and western populations it was possible to conclude that there was mixing of non-spawning individuals between the western and southern populations along the Portuguese coast but the majority of the southern individuals were caught south of Lisbon. Widescale application of the genetic approach has indicated that the current delineation of the three horse mackerel stocks is not appropriate for the purposes of data collection for stock assessment.

Table of Contents

1. Summary *1*

2. Background *3*

2.1 Biology and Assessment 3

2.2 Initial stock identification 4

2.3 Advances in stock identification 8

2.4 Post HOMSIR stock identification 13

2.5 Genetic assignment and the aims of the current study 17

3. Materials and Methods *20*

3.1 Sampling, DNA extraction, and SNP genotyping 20

3.2 Genotype data quality control 20

3.3 Preliminary analyses and population genetic structure 21

3.4 Informative marker identification and selection 21

3.4.1 Approach 1 21

3.4.2 Approach 2 22

3.5 Assignment model development 22

3.6 Assignment of mixed samples 23

4. Results and discussion *24*

4.1 Sampling, DNA extraction, and SNP genotyping 24

4.2 Genotype Data quality control 27

4.3 Preliminary analyses and population genetic structure 28

4.4 Informative marker identification and selection 33

4.4.1 Approach 1 33

4.4.2 Approach 2 34

4.5 Assignment model development 37

4.6 Assignment of mixed samples 40

5. Conclusions and summary *43*

6. Recommendations *44*

7. Acknowledgements *45*

8. References *45*

9. Annex 1 – Supplementary Tables *50*

10. Annex 2 – Supplementary Figures *60*

11. Annex 3 – Genotyping Quality Control *75*

12. Annex 4 – Alternative Assignment Model *77*

13. Annex 5 – Additional samples analysed *79*

14. Annex 6 – Combined assignment plots *88*

2. Background

The following section provides a short review of the basis for the current stock delineation of horse mackerel and recent genetic studies which have raised doubts as to the validity of the current stock delineations for data collation and assessment purposes. This provides useful context for the most recent genetic studies, which have shown that the current delineation of the stocks and division of data for assessment purposes is not appropriate.

2.1 Biology and Assessment

The Atlantic horse mackerel, *Trachurus trachurus* (Linnaeus, 1758), is a species of jack mackerel from the Carangidae family that is distributed in the East Atlantic from Norway to southern Africa (FAO Major Fishing areas 27, 34 and 47), and in the Mediterranean Sea (FAO Major Fishing area 37). It is a pelagic shoaling species found on the continental shelf and is one of the most widely distributed species in shelf waters in the northeast Atlantic, where it is a commercially important species.

Horse mackerel are considered to be asynchronous batch spawners with indeterminate fecundity (Gordo et al., 2008; Ndjaula et al., 2009; van Damme et al., 2014) and are known to undertake annual migrations between spawning, feeding and over-wintering areas (Abaunza et al., 2003). In the northeast Atlantic area horse mackerel have a prolonged spawning season, with spawning noted as occurring in shelf waters in the western stock area from at least March to July (ICES WGMEGS, 2023) in the North Sea stock area along the coastal regions of Belgium, the Netherlands, Germany, and Denmark from May to late July/early August (Eltink, 1990; 1992; Iversen et al., 1989; Macer, 1974) and in the southern stock area in shelf waters throughout the year from November to September (Costa, 2022). The temporal distribution of peak spawning varies between geographic regions; west of Ireland in June and July (ICES WGMEGS, 2023), in the North Sea in May and June (Macer, 1974) and in Portuguese waters in February and March (Costa, 2022).

In the Northeast Atlantic horse mackerel are assessed and managed as three separate stocks, The Southern stock (ICES Division 9.a), the North Sea stock (ICES divisions 3.a, 4.b–c, and 7.d), and the Western stock (ICES Subarea 8 and divisions 2.a, 4.a, 5.b, 6.a, 7.a–c, and 7.e–k), which are believed to broadly align with population structure. However, the discreteness of the stocks and the location and levels of mixing between them are unknown, which leads to uncertainty in the input data for the stock assessments. Compared to northeast Atlantic mackerel, very little is known about the seasonal or life-time migratory behaviour of horse mackerel. This situation is exacerbated by the current stock delineation and resulting ICES assessment process, where the western and North Sea stocks are assessed by the ICES Working Group on Widely Distributed Stocks (WGWIDE) and the southern stock is assessed by the ICES Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA). As a result the catch, survey and biological data for each stock are analysed in isolation without developing a cohesive understanding of the distribution or biology of the species across the three stock areas or in the northeast Atlantic area in general. This lack of integration and the different assessment approaches applied to each stock have led to a flawed basis for the development of catch advice.

The spawning stock biomass (SSB) of the southern stock has, according to the stock assessment, increased significantly over the past ten years from an average annual SSB (1992-2010) of c.350k tonnes to over 1.2m tonnes in 2023 (ICES, 2023a). Annual catches remain at the unrestrictive level of approximately 20,000 – 30,000t per year despite advised catch increasing exponentially since 2014, up to a high of ≤ 173,873t in 2024. Conversely the SSB of the western stock has decreased from an estimated peak of over 5m tonnes in 1988 to c. 707,811 tonnes in 2023 (ICES, 2023b), with the advised catches in the western stock having decreased from c. 200,000t in 2011 to zero catch in 2023, despite catch levels being at or below the advised level in almost all years over the past decade. The assessment of the North Sea stock is more uncertain and though the current stock assessment does

not provide estimates of SSB, the North Sea stock is considered to be significantly smaller than either the Western or Southern stocks and it has also declined significantly in recent years (ICES, 2023b).

2.2 Initial stock identification

Prior to 1984 the Advisory Commission on Fishery Management (ACFM) did not make any recommendations about catch levels for horse mackerel because of a lack of biological information about the stocks and doubts about catch statistics (ICES, 1984). However, it was noted that the fisheries in the southern areas (8.c and 9.a) were concentrated on juvenile fish. In 1985 horse mackerel were included in the Working Group on the Appraisal of Sardine Stocks in Divisions VIIIc and IXa where available data from ICES areas 4, 6, 7, 8 and 9 were compiled for the first time (ICES, 1986a). As there was an absence of any information concerning stock identity, the data from each ICES area were considered separately and it was not possible to assess the spawning stock biomass. At the 4th Annual Meeting of NEAFC (26-28 November 1985), the following request to ICES was made: "*Describe the distribution of the horse mackerel in the NEAFC area and if possible assess the state of these stocks*". It was not possible to conduct an assessment in 1986 (ICES, 1986b) but length frequencies of commercial and survey catches in different areas, where available, were presented.

The first attempt to delineate stocks of horse mackerel and to conduct an assessment was in the 1987 Working Group on the Assessment of Pelagic stocks in Divisions VIIIc and IX and Horse Mackerel (ICES, 1987a). The working group first reviewed a horse mackerel ageing workshop that had taken place in 1987 (ICES, 1987b). At the time there were significant differences is the ageing methods used for horse mackerel, which led to significant variability in the age determination between different countries and readers (see Eltink, 1985). It was confirmed at the workshop that this was still a significant issue, though it was noted that by studying age compositions of Dutch samples from 1981-1986 from divisions 6.a and 7 that the 1982 year-class was abundant. This year-class was also abundant in Portuguese samples from 1984-1986 from division 9.a. This was further explored by the working group (ICES, 1987a), who followed a strong 1979 year-class in Dutch catches from ages 3-7 during the year 1982-1986 and showed that disagreement in ages started from age 4 onwards. The English data disagreed and showed a strong 1980 year-class that was also seen in the Portuguese samples from 8.c and 9.a, but this was assumed to be an incorrectly aged 1979 year-class due to the Dutch data. The very strong 1982 year-class was easier to follow in both the western and southern stocks and based on the plots provided in the report it appeared to largely disappear from the southern data from age three onwards. The same pattern was seen with an apparently strong 1985 year-class. This was not discussed in the report, but it suggests that the fish were moving out of the data collection area. It is not clear if this was a move to deeper water away from the target fisheries or a move northward into the western stock area. It also highlights the fact that the 1982 year-class was not a stand-alone year class but during this period there were multiple strong year-classes. Once these reached age 4+ it was difficult to distinguish them and as such it is possible they may have been incorrectly assumed to belong to the same year-class, thus inflating the perception of the 1982-year class.

The 1987 working group also reviewed mackerel egg survey data from the 1977, 1980, 1983 and 1986 surveys and decide that "*the borders of this main spawning area to other spawning areas are the English Channel and the southern part of the surveyed area in Bay of Biscay*". Based on this the group concluded that there were three stocks: the "*Western Horse Mackerel*" (6.a, 7.a-c,e.k, 8.a-b,d-e), the "*North Sea Horse Mackerel*" (2.a, 3.a, 4.a-c, 7.d) and the "*Southern Horse Mackerel"* (8.c and 9.a). The delineation of the southern stock was noted as being provisional ("*for the time being*") because it was not based on biological information on spawning areas. The North Sea stock was noted to spawn mainly in the southern North Sea and it was suggested to "*probably*" overwinter in the English Channel where it would "*mix to some extent*" with the western stock, which highlighted the uncertainties in the stock delineation.

In the 1988 working group the ageing issues were again discussed and the group concluded that the 1-ring-per-year interpretation of the otoliths followed by the Netherlands was "*convincing*" even though it was previously only accepted by the minority. Therefore, the working group advised the age compositions produced by other interpretations should be revised and everyone should follow the Dutch method. At the time only Dutch age data were included in the western stock assessment and only Portuguese age data in the southern stock assessment. It also noted that "*this will be the last Working Group report which contains extensive tables of length composition for horse mackerel. Now that agreement has been reached on criteria for age determination, length-based stock assessment methods need not be considered.*" This was an unfortunate decision as there remained the issue of the difficulty of accurately ageing horse mackerel greater than four years old and continuing to present length data would have enabled a better understanding of the structure of the different stocks. Future data would potentially be compounded by ageing errors and inter reader variability, which was demonstrated to be a significant issue (ICES, 1987b).

Despite the conclusions of the 1987 working group regarding stock structure, the 1988 working group stated that they lacked "*sufficient knowledge to determine the stock structure of horse mackerel*" and that the information on egg and larval distribution was "*not sufficient evidence to infer independent stocks, as adult horse mackerel are highly mobile, and these areas may represent no more than three separate areas where spawning environments are favoured by the fish.*" The presence of the 1982 year-class in all three stock areas, as defined in 1987, indicated a degree of connectivity among the stock areas as did the absence of larger fish in the southern stock and a surplus of larger fish in the northern North Sea. The working group assessed the stocks under two hypotheses: three independent stocks and a single mixed stock and concluded that until the issue stock identification is settled, the working group would continue to produce basic stock assessments on the basis of three separate stocks. Regardless an exploratory combined assessment of the western and southern stock was attempted but the North Sea was excluded as there was insufficient data available. Strong historical recruitment was noted in 1968 and 1969 that contributed to the high level of catches in the 1970's by the USSR. The combined assessment was difficult to perform though given the lack of agreement in the age data between the western and southern samples and also the fact that the southern fishery was primarily based on juveniles and the western fishery on adult fish. A tagging programme was also proposed to investigate the migratory behaviour of fish, with the emphasis being placed on tagging juvenile fish.

The 1989 working group report reiterated the uncertainties in stock delineation in subsequent meetings (ICES, 1989) where it stated "*Egg and larval distributions suggest the existence of separate spawning areas corresponding to the three geographic stocks (Southern, Western, and North Sea) of horse mackerel. This is not sufficient evidence to infer that these are independent stocks as adult horse mackerel are highly mobile and these areas may represent no more than three separate areas where spawning environments are favourable."* The working group also reported that due to increasing catches by Norway in divisions 2.a and 4.a in 1987 and 1988, "*Questions were raised at the ACFM meeting in May 1988 on the distribution of the North Sea horse mackerel "stock" in Divisions 4.a and 2.a, and it was suggested that it could be part of the Western "stock" which includes Division 6.a."* Therefore, investigations were carried out to determine the stock of origin of these catches, which were caught during the mackerel fishery in quarters 3 and 4. They were compared to quarter 4 bottom trawl survey data from the western area and it was concluded that the migration pattern of the western horse mackerel seemed to be very similar to that of the western mackerel. It was also concluded that there was a distinction between these fish and those surveyed off the Danish coast during acoustic surveys in August 1985-1988 because there was a gap between the survey area and the high concentrations of fish observed there and the area where the Norwegian catches were taken. No analyses were undertaken and this was a subjective decision. On this basis the working group decided to expand the area of the western horse mackerel stock to include divisions 2.a and 4.a. It was assumed that mixing between the western and North Sea stocks northeast of the UK was low

because the North Sea spawning occurred in the coastal areas in the southern North Sea and German Bight and that the feeding area of the North Sea stock was likely in the eastern central North Sea. Mixing to the south of the UK in the English Channel was stated as being known to "*occur mainly in division 7.e, when the North Sea horse mackerel overwinters in the English Channel*". However, no evidence for this assumption was presented in 1989 and there was no existing evidence from previous working group reports. As such the origin of this assumption is unclear. Based on these assumptions and the results of egg surveys it was noted that the western stock was approximately seven times larger than the North Sea stock and as such the horse mackerel in 7.e were probably western horse mackerel, so catches in 7.e should be allocated to the western stock. One additional piece of relevant information from the 1989 report was the table of sampling effort in the western stock area (Table 1). It is useful to see that despite overall increasing catches in the western stock area from 1982-1988, the level of sampling had reduced significantly over the period from 66% coverage in 1982 to 29% coverage. This further highlights the uncertainties that were likely to have been present in the biological data used in the assessment.

Table 1. The catch sampling effort by the Netherlands in the Western stock area (ICES, 1989).

Category	1982	1983	1984	1985	1986	1987	1988
Dutch catch (t)	27,500	36,200	54,700	57,500	51,700	75,150	49,140
International catch (t)	41,588	64,862	73,625	80,551	105,765	156,247	170,900
% covered with age sampling	66%	56%	74%	71%	49%	48%	29%

In 1990 the working group was renamed the Working Group on the Assessment of the Stocks of Sardine, Horse Mackerel and Anchovy (ICES, 1990). Due to improvements in data collection the working group was able to divide 90% of the total catch by quarter and area and when viewed in conjunction with survey data led to indications of the seasonal migratory pathways (Figure 1). The North Sea stock was believed to have a smaller migration route than the western stock, mainly within the southern and eastern North Sea. In October 1989 the fishery was noted as having moved from the southeastern North Sea to the west and southwards through the English Channel, though there was no evidence to determine whether the fishery was targeting North Sea or western horse mackerel in this area. After spawning off the southwest of Ireland the western stock was believed to migrate north to the southern part of division 2.a, before moving south into 4.a in August and finally moving back west of Scotland and Ireland in November. The working group concluded that the western and North Sea stocks were probably mixing in division 4.b in quarters 3 and 4 and in parts of 7.e in quarter 4 but further work including tagging, parasite analyses, fatty acid profiles and genetic variation analyses should be undertaken. There appeared to be no hiatus in the distribution of eggs or spawning from the southwest of Ireland to the northern Spanish shelf (division 8.c), though spawning in this area was more intensive in the eastern part of the northern Spanish shelf than in the area around Vigo to the west. There was no mention of Portuguese waters yet it was concluded to be part of the southern stock apparently on the basis of the perception of the horse mackerel on the northern Spanish shelf. Of interest though were the results of larvae surveys carried out from October 1986 to January 1989, which indicated that in Northern Portuguese waters the peak occurrence of larvae was in April, whilst off the south of Portugal larvae were present throughout the year.

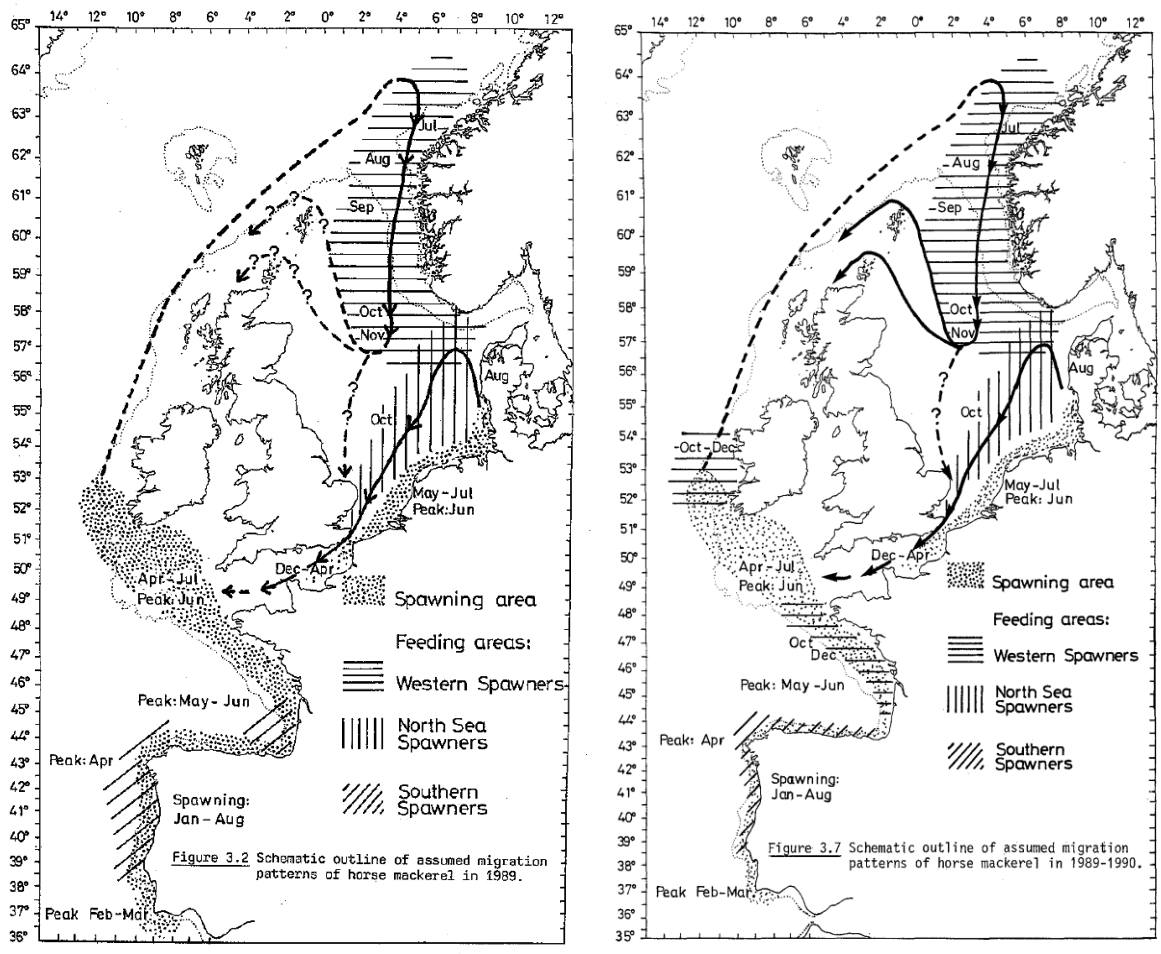


Figure 1. Schematic outline of the assumed migration patterns of horse mackerel in (left panel) 1989 (ICES, 1990) and (right panel) 1990 (ICES, 1991).

Initial results from a parasite analysis study presented at the 1991 working group supported the separation of the horse mackerel that spawned in the southern North Sea from those that spawned in the western area (ICES, 1991). Based on the distribution of the fishery the report also concluded that that not all horse mackerel for the west and southwest of Ireland migrated to the northern North Sea and it was predominately the larger fish that did, which then returned via division 6.a rather than through the English Channel (Figure 1). Analyses of the distribution of 0-group, 1-group and 2+-group horse mackerel caught during the 1990 Q4 Scottish, English, Dutch, French, Spanish and Portuguese bottom trawl surveys highlighted the prevalence of juvenile fish in the southern stock areas and also the Bay of Biscay and English Channel (Figure 2). Much of the focus was again on the migration patterns of the western and North Sea stocks with little emphasis on the potential migration pathways between the western and southern areas. This is surprising as the report noted that ACFM commented in relation to last year's assessment that "*There are basic data problems for this stock. The stock definition is not clear and the Working Group has identified the need for further research into it*". The report also noted that another horse mackerel ageing workshop had taken place in 1990 at which the results of an otolith exchange programme were to be presented. However, the samples from the southern stock area were not analysed in time for the workshop and despite the decision taken at the 1987 ageing workshop to follow the Dutch method of ageing and reanalyse all samples collected to date with the new method, it was reported than only the 1990 southern samples were aged with this method and the older age data still required revision. It was therefore not possible to conduct a full assessment on the southern stock.

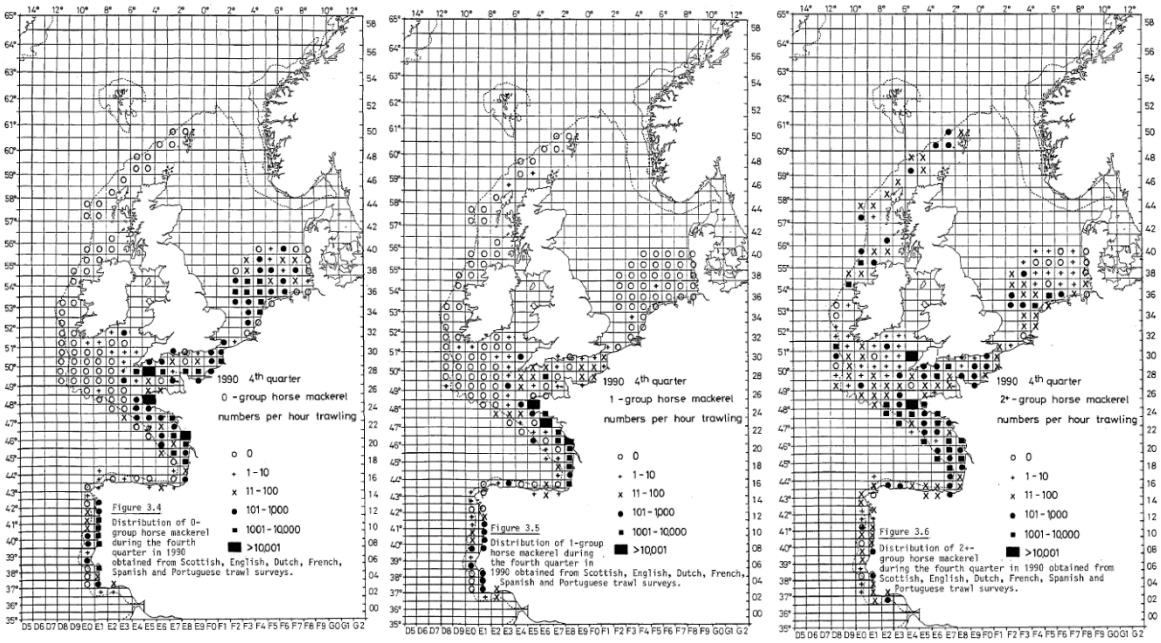


Figure 2. The distribution of (left) 0-group (middle) 1-group (right) 2+-group horse mackerel during the 1990 Q4 Scottish, English, Dutch, French, Spanish and Portuguese bottom trawl surveys (from ICES, 1991).

In 1992 the Study Group on Stock Identify of Mackerel and Horse Mackerel was held in Vigo with the terms of reference to focus specifically on the stock identification issues in sub-areas 8 and 9. After reviewing available data the study group concluded that "*The evidence currently available is not adequate to determine whether two separate stocks or one single stock occupies the Western and Southern areas.*"

From 1993 to 1997 the working group noted each year at the start of the horse mackerel chapter that there was "*there is no well established biological basis*" for separating horse mackerel into the three stocks but that there was "*no new information on which to base a change in the stock-separation used previously*". Therefore, the horse mackerel continued to be assessed as three stocks despite there being little evidence to support this. In 1997 the working group report (ICES, 1998) noted that the "*Southern and Western horse mackerel are thought to have similar migration patterns to the mackerel from the same areas. As for mackerel the egg surveys have demonstrated that it is difficult to determine a realistic border between a western and southern spawning area*". Two new studies based on allozyme differentiation and morphometric characteristics were noted to have not provided any basis for changing the existing stock separation and tagging studies in Portuguese and Spanish waters had failed to recover any tagged fish.

It is evident from the early working group reports that the decision to split horse mackerel into three stocks for assessment purposes was based on a largely ad hoc approach and very little scientific evidence. Whilst there may have been some biological basis for the assumption that the southern North Sea and Western stock were different populations, there was, as noted in multiple reports, little basis for the delineation between the western and southern stock areas. Regardless most of the focus of subsequent working groups was on the development of increasingly complex stock assessment approaches with less and less emphasis on further stock identification at the working group level.

2.3 Advances in stock identification

The EU-funded HOMSIR project (2000-2003) attempted to address the issues with stock identification by employing a multidisciplinary approach including various genetic approaches (allozymes, mtDNA, microsatellites), body morphometrics, otolith shape analyses, parasites as biological tags and the

comparative study of life history traits (growth, reproduction and distribution) (Abaunza et al., 2008a). Samples were collected over two years across the three stock areas, in north African waters and in different parts of the Mediterranean Sea, with each sample being analysed with the suite of different methods.

Despite multiple genetic approaches being used in the HOMSIR project (Cimmaruta et al., 2008; Comesaña et al., 2008; Kasapidis and Magoulas, 2008), no genetic population structure among the samples (Figure 3) was identified. It should be noted that the genetic methods applied were considered state-of-the-art at the time of the study though it is now widely recognised that the approaches used did not have sufficient power to robustly identify population structure in a pelagic species such as horse mackerel (see Andersson et al., 2024).

Body morphometric analyses apparently indicated significant differentiation between the Atlantic and Mediterranean samples and a latitudinal gradient within the Atlantic samples. It was concluded that there was a clear boundary between three clusters corresponding to the North Sea, north of the Galician coast through the west of the Ireland and Britain, and to the Portuguese coast and Gulf of Cádiz (Murta et al., 2008a). However, when only pre-spawning and spawning individuals were considered one of the western samples clustered with the Mediterranean samples and the Portuguese samples clustered as a sub-branch of the North Sea sample. Further, whilst the cross validation of the discriminant analyses indicated reasonable classification to either the Atlantic or Mediterranean samples, the classification with the Atlantic samples was much less clear and there was significant misclassification between samples from the three stock areas. Interestingly, samples from the Algarve on the south coast of Portugal had the highest self-classification rate (76%) for pre-spawning and spawning fish. When all fish were included regardless of maturity stage there also appeared to be a latitudinal pattern to the self-classification with the sample from northwest Portuguese waters displaying a higher percentage of misclassification to the more northerly samples than the sample from Southwest Portugal, which had a higher rate of misclassification to the Algarve and Mauritanian samples. This indicates that these two samples from the west coast of Portugal may be from a mixing zone where more northerly and southerly populations mix. Whilst overall the results indicated that there was likely a level of differentiation between and among the Atlantic samples that loosely corresponded to the existing ideas of stock structure, the results were not conclusive enough to define static stock boundaries.

The results of otolith shape analyses of the same samples were broadly similar in that they did not indicate a separation of the western horse mackerel stock from the North Sea stock but did resolve three clusters of areas: a northern, an Ibero-Mauritanian and an eastern Mediterranean group (Stransky et al., 2008). It was proposed that the high misclassification rate within the northern areas, including Galicia, as well as within the areas along the Portuguese coast supported the separation of these groups from each other and pointed to high within-group similarities. Though similar to the body morphometrics, the sample from the Algarve area had the highest self-classification rate (67%) of fish from the northeast Atlantic area and as with the body morphometric analysis there was a latitudinal gradient evident in the misclassifications of the northwest and southwest Portuguese samples, which suggests western Portuguese waters may be mixing zone between northern and southern populations.

The parasite analyses support this theory as the results indicated evidence of considerable mixing between western, southern and Mauritanian populations (MacKenzie et al., 2008). Interestingly it was also noted that "*the occasional occurrence in some of our northern samples of parasites known to be more common in Trachurus spp. populations off West Africa indicates migration of T. trachurus from West Africa as far north as south-west Norway"*. The distinction between the North Sea and western stocks was supported by the parasite analyses.

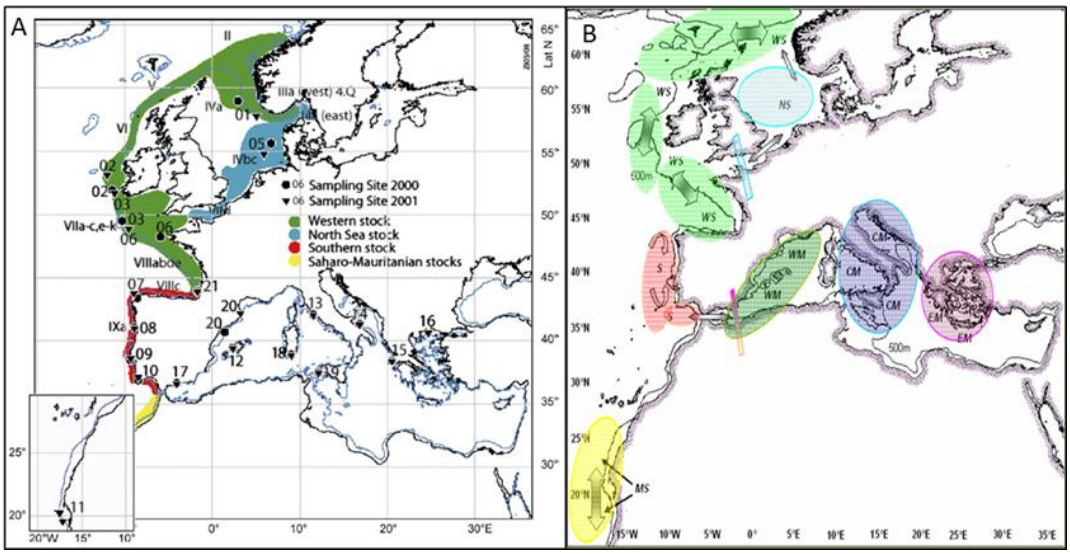


Figure 3. (A)The suggested stocks of horse mackerel prior to the HOMSIR project. The sampling sites in the HOMSIR project in 2000 (circles) and 2001 (triangles). (B) Proposed horse mackerel stocks according to the results of the HOMSIR project. The arrows indicate possible migratory movements. WS: western stock; NS: North Sea stock; S: southern stock; MS: Saharo-Mauritanian stock; WM: western Mediterranean stock; CM: central Mediterranean stock; EM: eastern Mediterranean stock. From Abaunza et al. (2008).

Analyses of life-history data and length-at-age data appeared to indicate an increasing trend in median length-at-age with latitude in two different areas (midway along the Portuguese coast and along the west coast of Ireland), which was reported to suggest the possibility of a particular length-dependent migration pattern during the spawning season in these areas (Abaunza et al., 2008b). It should be noted that not all ages were present in sufficient numbers in the samples from each sampling location for inclusion in the analyses and there was also significant variability in length-at-age within sampling area, which make the interpretation of the results difficult. In the Atlantic, both the length and age at first maturity seemed to increase with increasing latitude and in the northern most sample areas including the Northern North Sea samples (division 4.a) the samples were dominated by larger older fish relative to all other areas. Also not highlighted but clearly discernible from the presented data was the lack of older fish in the more southerly Atlantic samples, including Portuguese waters and in the Mediterranean Sea (Figure 3 in Abaunza et al., 2008b). Abaunza et al. (2008b) also noted that "*there are growth differences in areas belonging to the same stock. This is the case for the so called "western stock" in the northeast Atlantic (from south of the Bay of Biscay until Norway, except the North Sea) and the so-called "southern stock" in the western Atlantic waters of the Iberian Peninsula*", though this was not elaborated on further apart from acknowledging that due to "*the difficulties in obtaining all size groups in each area, the resulting von Bertalanffy growth function parameter estimates often represent large extrapolations beyond the range of the sampling data, and are therefore considered unreliable*". Of interest was the observation that the distribution of commercial landings of horse mackerel through the year, suggested the possibility of migratory movements between the spawning areas (from south-western Ireland to the Iberian Peninsula coasts) and the feeding and wintering areas (Norwegian coasts, northern North Sea, The English Channel). One might also hypothesise that the lack of older fish in the southern stock area may be related to an ontogenic migration of larger older fish to more northerly areas with more southern areas acting as a nursery areas and juvenile areas as is seen in other pelagic species including mackerel (*Scomber scombrus*) and boarfish (*Capros aper*).

Murta et al. (2008) explored such ontogenic migrations along the Iberian Atlantic coast, based on bottom-trawl survey data covering the period 1985–2003 and suggested that the fish in division 9.a appeared to primarily migrate within the area whilst there were indications of fish migrating into division 8.c from either the north or the south. The authors did however acknowledge that the

available data were limited to the same season in each of the years and as such were temporally limited. As a result seasonal migration patterns may not have been identified in the study. In the context of understanding the biology of the species across the northeast Atlantic area it would be both interesting and valuable to undertake further analyses, similar to those of Murta et al. (2008), but including commercial and survey data from across all three stock areas without segregating by stock area.

The HOMSIR project was a groundbreaking study and was in its time perhaps the most comprehensive and detailed stock identification study undertaken in the northeast Atlantic area. The resulting stock structure was broadly similar to that previously described, with the most significant suggested change being the realignment of the stock boundary between the western and southern stocks from the southern Bay of Biscay to Cape Finisterre. This hard boundary was conveniently located on the boundary between divisions 8.c and 9.a, which as noted would make segregation of existing data and collation on new data easier. However given the small number of temporally limited samples collected either side of this boundary and the aforementioned indications of mixing within Portuguese waters, it is difficult to accept that this boundary has been robustly tested. The HOMSIR project did note though that the population structure in the western European waters could be more complicated and that more research was needed to clarify the migration patterns within the Northeast Atlantic. This also was especially relevant to the mixing areas between the North Sea stock and the Western stock (northern North Sea, ICES Division 4.a and English Channel, ICES Division 7.d), where the sampling was relatively sparse in comparison to the southern regions, including the Mediterranean Sea. Despite these uncertainties the HOMSIR based stock delineation has persisted in both the assessment of the species in the northeast Atlantic and has rarely been challenged or reconsidered.

Whilst considering the results of the HOMSIR project it is important to note several key considerations:

1) The sample sizes were small relative to the size of the populations and limited to two consecutive years.
2) Samples were collected opportunistically from surveys and from commercial catches and comprised a mixture of maturity stages and both adult and juvenile specimens.
3) Though the methods applied were considered state of the art at the time this is no longer the case (e.g. genetics) and results should be viewed with a more informed perspective.
4) Whilst a number the analysed approaches indicated mixing in Portuguese waters this was largely ignored when it came to realigning the stock areas.
5) The sampling for the project took place in 2000 and 2001 (23-24 years ago).
6) The assessed abundance of the western stock has declined significantly since this period.
7) The assessed abundance of the southern stock has increased almost exponentially since c. 2010.
8) Populations and environmental conditions are not stable and have changed considerably since the HOMSIR time.

The results of the HOMSIR project were presented to the working group in 2003 but the changes in the delineation of the southern stock were not incorporated in that year (ICES, 2004), though it was recommended to undertake this in time for the following year. The working group noted that that it seemed strange that only catches from the western part of Division IIIa were allocated to the western stock but explained that the "*reason for this was that the catches in the western part of this Division taken in the fourth quarter usually are taken in neighbouring area of catches of western fish in Division IVa*". The group further noted that it was not sure if catches in 3.a and 4.a in quarter 1 and quarter 2 were of western or North Sea origin but that catches were usually low and as such would continue to be allocated to the western stock.

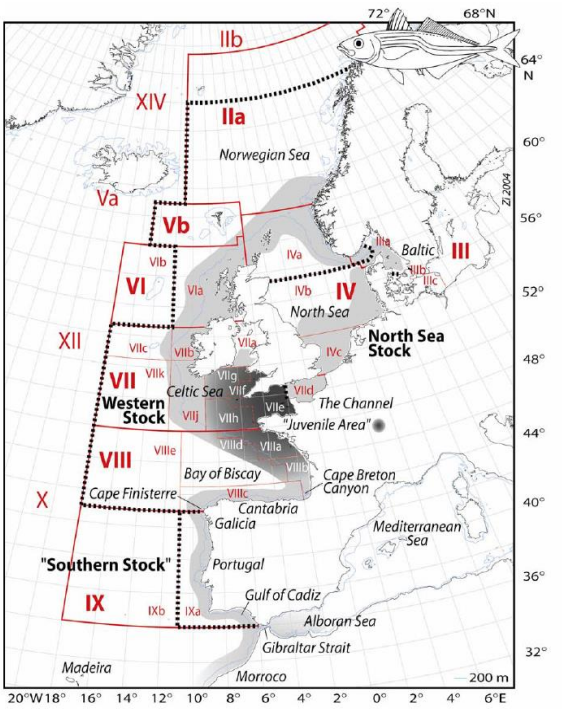


Figure 4. The distribution of horse mackerel in the northeast Atlantic with the new stock definitions according to the HOMSIR project (ICES, 2004).

In 2004 the working group realigned the southern and western stock areas based on the results of the HOMSIR project (Figure 4) and reallocated the catches from division 8.c from the southern to the western stock, as had been decided in 2003 (ICES, 2005a). Surprisingly and without precedent or supporting evidence the working group also reallocated the catches from quarters 1 and 2 in division 4.a to the North Sea stock instead of the western stock. This seemingly arbitrary decision was contrary to what was done in 2003 and no justification was provided despite there being no change in the pattern of catches. In 2005 the working group changed this decision again and allocated the catches in division 4.a to the western stock and those in division 3.a to the North Sea stock, again with no explanation for the change (ICES, 2006). This was repeated at the 2006 working group as the quarters 1 and 2 catches were again small (ICES, 2006b). However, at the 2007 working group catches in 4.a in quarters 1 and 2 were deemed to be significant and the decision was reversed again with these catches being allocated to the North Sea stock (ICES, 2007). In 2008 the new Working Group on Widely Distributed Stocks (WGWIDE) was established, and the three horse mackerel stock assessments were included in its remit (ICES, 2008). The 2008 working group continued to allocate the quarter 1 and 2 catches in 4.a and 3.a to the North Sea and this appeared to become the standard approach in subsequent working groups, despite there being no evidence to support it.

After the HOMSIR project the stock definitions remained largely unchallenged within the working groups assessing horse mackerel, despite their being significant uncertainty in their delineation as summarised above from forty years of working groups. This led to the three stock assessments being developed in isolation from each other and the implementation of different sampling strategies within the different stock areas. These differences likely propagated the perceived differences in the horse mackerel in the different stock areas and prevented a cohesive understanding of the species across the three stock areas being developed. This issue was exacerbated in 2011 when the southern horse mackerel assessment was moved from WGWIDE to the Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA). As a result, the data and assessments for the western and southern stock areas were entirely isolated from each other within ICES and there were no official attempts to resolve the stock structure further.

2.4 Post HOMSIR stock identification

A number of academic and industry funded projects have attempted to address these issues and have continued to develop the knowledge base for stock identification in horse mackerel. More recent genetic studies based on a small number of putatively neutral microsatellite loci (four from *T. trachurus* developed during the HOMSIR project and eight cross-amplified from the Chilean Jack Mackerel *Trachurus murphyi* Nichols, 1920 (Canales-Aguirre et al., 2010) indicated evidence of potential population structure between samples caught in Norwegian waters (ICES Division 4.a) and those caught west of Ireland (ICES Divisions 6.a, 7.c and 7.j). Though this pattern of structure was temporally unstable across the multiannual samples, the authors suggested that it potentially indicated the presence of local populations in ICES Division 4.a in addition to the migratory western stock (Sala-Bozano et al., 2015). No significant genetic structure was detected between the samples from the central North Sea (ICES Division 4.b) and those sampled west of Ireland (Sala-Bozano et al., 2015). It should be noted that spawning samples were not available for analysis during the study and as such analysed samples may also represent mixed stock/population samples.

Healy et al. (2019) utilised the same panel of microsatellites (minus two of the *T. murphyi* microsatellites) to investigate population structure of *T. trachurus* samples from the southern part of the species distribution, including southern Portugal, the Alboran Sea, north- and central- African waters. The analyses indicated no significant structure between the southern Portuguese, Alboran Sea and north African samples (Healy et al., 2019). The other results of the project were not relevant to the stocks in the northeast Atlantic area.

In 2015 the Pelagic Freezer Trawler Association (PFA) contracted the Wageningen UR, Institute for Marine Resources and Ecosystem Studies, IJmuiden (IMARES) to undertake a study on North Sea Horse Mackerel (Brunel et al., 2016). The primary aim of the study was to improve the data quality used for an analytical stock assessment model of North Sea horse mackerel. The management boundary between the western and North Sea stocks in the English Channel (corresponding to the separation between divisions 7.e, western Channel and 7.d, eastern Channel) does not correspond to a real biological boundary, as mixing of the two stocks is known to occur in division 7.d in autumn and winter (Brunel et al., 2016). The catches taken in 7.d are officially considered as being North Sea horse mackerel and represent c.80% of the catches from this stock. An unknown proportion of this catch is likely from the western stock, which interferes with the cohort signal in the catch at age matrix, hampering the development of an age-structured assessment model for the North Sea stock. Developing methods to separate catches from the western stock from catches from the North Sea stock in division 7.d was deemed necessary to improve the quality of the catch information for the North Sea stock. Within the project, two pilot studies, based on chemical fingerprint and genetics, were conducted to investigate new methods to determine stock structure and to develop techniques to identify the stock origin of the catches taken in the eastern English Channel.

The chemical fingerprint analysis was carried out by IMARES using two-dimensional gas chromatography (GCxGC-MS), in order to establish a full chemical fingerprint of the horse mackerel samples from both the western and North Sea stocks. Results were inconclusive but suggested that the chemical fingerprint approach was a potential tool to determine stock of origin, with a moderate risk of misclassification. However, more insight on the sources of variation of compound concentrations (seasonal changes, influence of sex, length, age, reproducibility of the results from year to year) would be required before this method can be further developed.

The genetic analyses were contracted to University College Dublin (UCD) to undertake a pilot study to develop a method of genetic stock identification for discriminating North Sea and Western Horse mackerel (Brunel et al., 2016). The aims of the pilot study were to firstly develop and validate at least 24 polymorphic microsatellites markers in horse mackerel and secondly to screen spawning fish collected in 2015 from the Western and North Sea stocks (same samples as the chemical fingerprint

analyses) to establish a genetic baseline of the spawning stocks and test the presence of population structure. Next Generation Sequencing (NGS) and Genotyping by Sequencing (GBS) based approaches, which were developed on cod (*Gadus morhua* Linnaeus, 1758), boarfish (*Capros aper* Lacépède, 1802) and herring (*Clupea harengus*) were used for marker development and screening of spawning samples (Farrell et al., 2016; Vartia et al., 2014 & 2016). The pilot study successfully identified a large number of novel microsatellites, however initial data analyses were confounded by a poor-quality sequencing run and as such the discrimination power between the western and North Sea sample was low. This resulted in the pilot study being unable to separate the two stocks conclusively and unequivocally.

In an effort to develop the analyses further the Northern Pelagic Working Group of the European Association of Fish Producers Organisations (EAPO) funded further analyses to be undertaken. Farrell & Carlsson (2018) used a combination of an established shotgun sequencing approach (see Farrell et al., 2016) and mapping to long read sequences generated using the Oxford Nanopore MinION platform to identify novel horse mackerel microsatellites, which were subsequently genotyped in population samples using a Genotyping-By-Sequencing (GBS) approach. Sampling was conducted over three consecutive years and three spawning seasons and covered a large area of the distribution of the species including the Western, North Sea and Southern stock areas and also northwest African waters. In total 33 samples, comprising 2,295 individual fish were collected from 2015 to 2017 (Figure 5) and were genotyped with thirty novel microsatellites, three *T. trachurus* loci from the HOMSIR project (Kasapidis and Magoulas, 2008), three from *T. murphyi* (Canales-Aguirre et al., 2010) and one locus from the Japanese horse mackerel *Trachurus japonicas* (Temminck & Schlegel, 1844). Significant and temporally stable population structure was identified, for the first time, between the southern North Sea samples and all other areas. Exploratory assignment testing and mixed stock analysis indicated that a large component of the fish caught outside spawning time in the northern North Sea and western English Channel belonged to the Western stock. However, there was insufficient power to perform robust assignments of North Sea and western samples (Farrell & Carlsson, 2018). No significant genetic differentiation was found between samples from the Western or Southern stocks or from North African waters, however, it was suggested that this lack of differentiation was most likely due to a lack of power in the marker panel rather than a true case of panmixia.

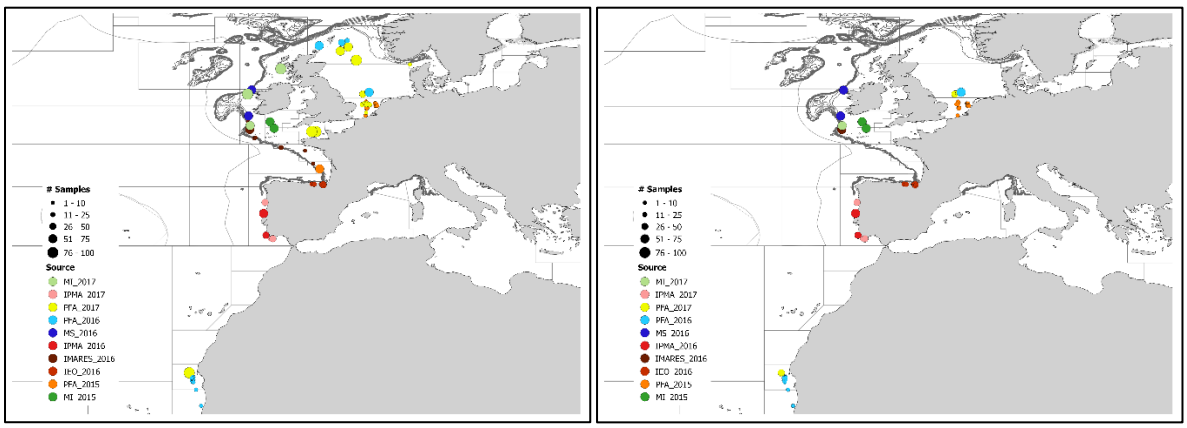


Figure 5. (Left Panel) The horse mackerel samples collected from 2015 to 2017 and (right panel) those included in the baseline dataset (From Farrell & Carlsson, 2018).

In order to improve the ability to identify informative genetic markers for horse mackerel it was deemed necessary to employ the Whole Genome Sequencing (WGS) approaches developed for herring (Han et al., 2020) and described by Andersson et al. (2024). In short, the approach involved developing an annotated draft genome for horse mackerel (Genner & Collins, 2002) and then undertaking population level pooled whole genome sequencing (Pool-Seq) of representative population samples, which were then aligned to the draft genome for analysis (Fuentes-Pardo et al., 2020; 2023). The primary aims of the study were to:

1) Identify population structure underlying the current stock divisions.
2) Estimate the extent of genetic differentiation between populations based on WGS.
3) Identify the evolutionary processes, genetic basis and environmental drivers of local adaptation.
4) Design a genetic marker panel that can be used for population studies and stock identification.

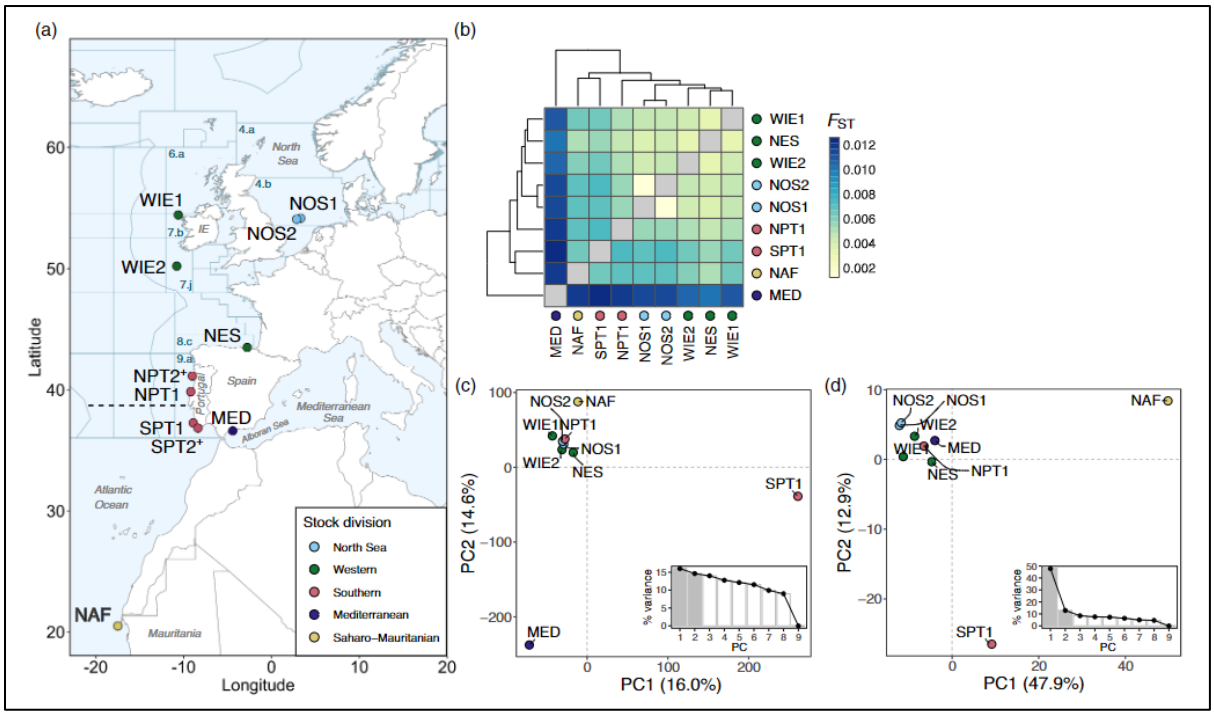


Figure 6. (a) Sampling sites of the Atlantic horse mackerel included in the Pool-Seq analyses. The approximate location of a biogeographical transition zone in central Portugal, near Lisbon, is denoted with a horizontal dashed line. In all plots, each dot represents a sampling location and its colour indicates the corresponding ICES stock. (b) Heatmap plot representing pairwise pool-*FST* values based on ~12.8 million SNPs. (c, d) Principal component analysis (PCA) plot based on (c) undifferentiated (61,543 SNPs) and (d) highly differentiated (818 SNPs) markers. The first two axes are shown.

The samples included in the WGS analyses were a subset of the samples analysed in Farrell & Carlsson (2018), with one additional sample from the Alboran Sea in the western Mediterranean Sea (Figure 6). A total of ~12.8 million polymorphic biallelic Single Nucleotide Polymorphisms (SNPs), distributed across the twenty-four chromosomes were identified, passed all the quality filters and were used in the population analyses. Overall, there were low levels of genetic differentiation among samples, with genetic differences constituting less than 1.5% of the entire genome, though three subtle population structure patterns were statistically significant. First, the largest genomic differences existed between the western Mediterranean Sea and all Atlantic samples. Second, Atlantic samples were genetically differentiated following a latitudinal pattern with a break near mid-Portugal, where samples north of this break (North Sea, west of Ireland, northern Spanish Shelf, northern Portuguese waters) were genetically more similar to each other than to the samples south of this break (Southern Portuguese waters and North Africa). Third, the samples from the southern North Sea were genetically differentiated from all other samples and represented a "*genetically distinct population*" (Fuentes-Pardo et al., 2020; 2023).

Fuentes-Pardo et al. (2023) further examined whether population structure patterns were driven primarily by neutral or selective processes by separately performing analyses on two subsets of SNPs comprising either neutral or adaptive markers. The results indicated that the separation between the westernmost part of the Mediterranean Sea and Atlantic populations may have been driven by neutral

processes, while the latitudinal pattern and separation of North Sea samples was more likely the result of selective processes. Most of the population structure patterns were driven by a few highly differentiated putatively adaptive loci. Seven loci distinguish the North Sea, two the Mediterranean Sea, and a large putative inversion on chromosome 21 underlined the north-south divide and distinguished the North African samples.

To validate Pool-Seq results and to identify a panel of highly informative SNPs for genetic stock identification, a subset of individuals (n=20 + 4 replicates) from each of the pooled samples were individually genotyped with a reduced panel of 63 SNPs, which was reduced to 17 SNPs (9 adaptive and 8 neutral) following further analyses. Genotyping was undertaken by a commercial provider; IdentiGEN, Dublin, Ireland, using their proprietary IdentiSNP genotyping assay chemistry, which utilises target specific primers and universal hydrolysis probes. Following an end-point PCR reaction, different genotypes were detected using a fluorescence reader. The analyses agreed with the Pool-Seq results and indicated that individuals clustered in four main groups: (i) the North Sea; (ii) west of Ireland, northern Spanish shelf, and northern Portugal; (iii) southern Portugal; and (iv) north Africa representing the identified populations (Figure 7). In all groups, some individuals showed admixed ancestry, suggesting that they were probably F1-hybrids or backcrosses between local and migrant individuals. Overall, the results indicates that gene flow occurred more often between neighbouring geographic areas and may explain why neutral markers are ineffective for stock identification in species such as horse mackerel (see Andersson et al., 2024).

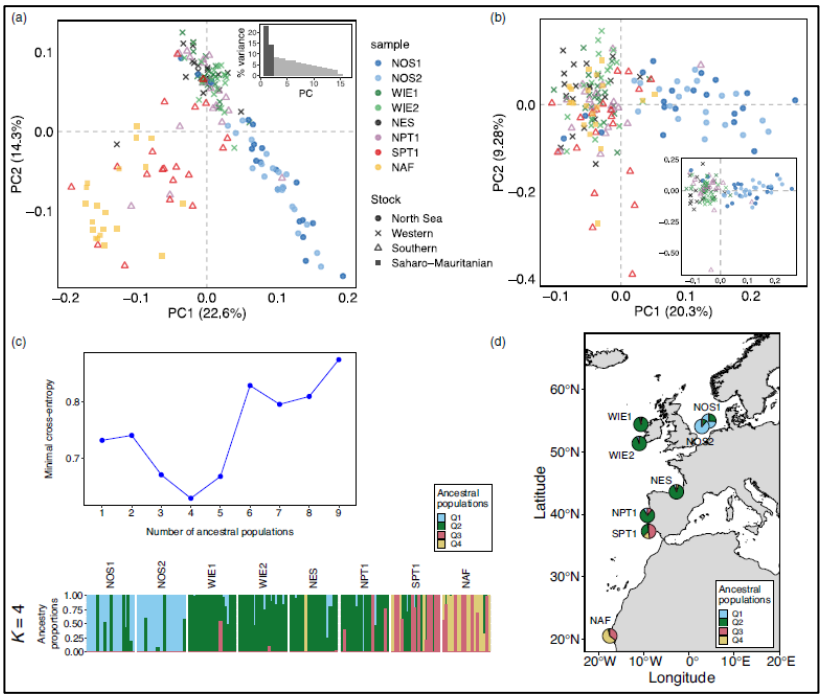


Figure 7. Population structure based on individual genotypes of the 17-SNP panel (from Fuentes-Pardo et al., 2023). (A-B) Principal components analysis (PCA) plot based on (a) all 17 markers (b) PCA with all samples but excluding markers from the chr21 inversion (n=2) (c) Analysis of admixture, (top) cross-entropy criterion plot to identify the most likely number of ancestral populations (K), (bottom) admixture bar plot for K=4 (d) map showing the mean ancestry proportion per location for K = 4.

In addition to investigating population structure Fuentes-Pardo et al. (2023) also undertook a Genome-Environment Association (GEA) analysis in order to identify which environmental variables were related to adaptive genetic variation and local adaptation in the populations identified. The GEA analysis indicated that there was a strong association between outlier SNP characteristics of the

southern North Sea samples and variation in temperature range or correlated environmental parameters such as iron content and primary productivity. The North Sea corresponds to the northern limit of the reproductive range of the species and exhibits a combination of environmental factors that makes this area unique. The southern North Sea is characterized by colder mean temperatures and a higher temperature range (colder winters and warmer summers) than other locations included in the study as well as higher oxygen content, iron content, and primary productivity. The particular environmental conditions in this area and the number of genomic regions that appeared to be under selection suggested a polygenic response to diverse selection pressures driving local adaptation. In simple terms the horse mackerel spawning in the southern North Sea are locally adapted to the specific environmental conditions there and are unlikely to successfully spawn in areas which do not meet these requirements. This is relevant to fisheries stock assessment as future recruitment of the North Sea population may be negatively impacted by changing environmental conditions and given the high level of local adaptation of the population they may not be able to successfully utilise alternative spawning and nursery areas. Such information may be useful in the context of developing recruitment forecasts for this population.

Whilst the aforementioned studies have established, using a variety of approaches, that there are at least three horse mackerel populations with the northeast Atlantic area, none have fully explored or uncovered the spatial and temporal population structure of the species. The current delineation of the stocks is based on a combination of assumptions made by the early working groups, the uncertain results of the HOMSIR project and the arbitrary decisions of the working group in later years (e.g. decisions regarding 4.a in Q1&2 and Q3&4). The most recent genetic analyses (Fuentes-Pardo et al. (2020; 2023) have indicated that some of these assumptions are likely unfounded and require further analyses as the outcome may have a significant impact on the three horse mackerel stock assessments. For example, the southern border of the western population may be further south that currently defined and within division 9.a, catches in division 4.a may need to be allocated to the western stock in all quarters of the year and there may be significant mixing of the western and North Sea stocks in the English Channel (divisions 7.d and 7.e). It is essential to both test the current assumptions and develop a method for continued monitoring.

2.5 Genetic assignment and the aims of the current study

Genetic assignment methods compare genetic data from individuals to genetic profiles of reference samples from potential source populations to determine population of origin, if any, for a given individual (Manel et al., 2005). The methods can also be used to assess the amount of overlap or separation between the reference populations (McMillan and Fewster, 2017). Assignment methods that attempt to solve classification problems rely on computing a discriminant function based on samples from potential source populations and then classify unknown individuals to the group with the highest discriminant score (Manel et al., 2005). Genetic assignment methods have traditionally relied on using the genotypic frequency distribution under the assumption of Hardy–Weinberg equilibrium (HWE) and linkage equilibrium in each source population as their discriminant function (Manel et al., 2005). These genetic assignment methods can be broadly divided into Bayesian (Rannala and Mountain, 1997), frequency (Paetkau et al., 1995) and distance (Cornuet et al, 1999) based methods (Hauser et al., 2006). The underlying assumptions of the methods are quite similar although the distance based methods may be less sensitive to violations of population genetic expectations such as HWE and linkage equilibrium (Cornuet et al., 1999). These methods are commonly implemented in the software GeneClass2 (Piry et al., 2004). In the absence of baseline data to guide classification, Bayesian clustering methods may be used to delineate clusters of individuals based on

their multi-locus genotypes and assign individuals to their individual clusters (Manel et al., 2005). However, these Bayesian clustering analyses such as that implemented in the software Structure (Pritchard et al., 2000) are also constrained by the underlying assumptions of HWE and linkage equilibrium. Multivariate analysis has several advantages over other classical approaches used in population genetics and genetic assignment, the foremost of which is that they do not require the assumptions of HWE or linkage equilibrium (Jombart et al., 2009). Multivariate approaches are particularly suited to solving classification problems when used in the form of supervised machine learning (SML) approaches. SML is concerned with predicting the value of a response label/category on the basis of the input variables/features (Schrider and Kern, 2018). When empirical data are available, SML trains an algorithm based on a training set of the labelled data, which can then be used to predict the category of unknown data. Support Vector Machines (SVM) are a set of SML methods that can be used for classification problems. The objective of SVM algorithms is to find a hyperplane in an N-dimensional space (N - the number of features) that distinctly classifies the data point (see James et al., 2013). SVM models can also be used to classify non-linear data through use of non-linear kernels (James et al., 2013) and can be optimised by adjusting parameters, including cost and gamma, which control the stringency of the boundary and the influence of single training datapoints, respectively. A lower cost means a softer boundary between the classes and means more individual points on the wrong side of the division will be allowed. A low value for gamma means that each data point will have a wider influence than if the gamma was high. SVM models do not directly provide probability estimates, these are calculated using logistic regression on the SVM scores, fit by an additional cross-validation on the training data. The output probabilities can be converted to odds in order to make the values more understandable The R package assignPOP (Chen et al., 2018) has recently made the use of SVM models for assignment more accessible and also allows for the integration of genetic and non-genetic data within the same model, which is an advantage in many stock identification studies which also collect morphometric data.

Such genetic assignment methods are now being employed as part of regular data collection protocols for herring in divisions 6.a, 7.b-c and also in the eastern North Sea and western Baltic areas (Bekkevold et al., 2023; Farrell et al., 2022). The methods are applied to survey and commercial catches and have facilitated the construction of separate survey indices for separate populations and have also led to an increased understanding of the migration and mixing of herring outside of the spawning seasons, when the populations are known to segregate. The most recent genetic analyses of horse mackerel (Fuentes-Pardo et al., 2023) have provided the basis to develop a set of tools with which to develop the capacity to similarly assign individual horse mackerel back to their population of origin. This a key consideration if a complete understanding of the spatial and temporal structure is to be developed. Like most marine fish this is likely to be dynamic, and largely driven by environmental factors. As such it requires ongoing monitoring, particularly if mixing zones between populations are identified. There may be temporal or spatial changes in the levels of mixing which will be reflected in the catch and survey data from those areas and which should be identified and accounted for in the sampling, stock assessment and management.

Whilst the small panel of SNPs tested by Fuentes-Pardo et al. (2023), on a subset of individuals, did prove to be informative and capable of discriminating between the samples tested, there may be other SNPs in the of ~12.8 million SNPs initially identified that had a higher power of discrimination between the populations or were more suitable for assay design. In order to test a wider selection of SNPs on a large number of representative individuals to ensure that final marker selection for a stock assignment panel is as robust as possible, it is necessary to use alternative genotyping methods than those used in Fuentes-Pardo et al. (2023). The most robust and cost-effective method for genotyping thousands of SNPs in thousands of samples is SNP-chip analysis. This approach is also more

reproducible and provides higher quality data than GBS based approaches, which is why it has been used in many human genetic studies, often targeting a million SNPs. SNP-chip analysis also provides a means to establish baseline reference populations by enhancing the comparability of findings among institutions and facilitating the merger of data from different studies (Andersson et al., 2024).

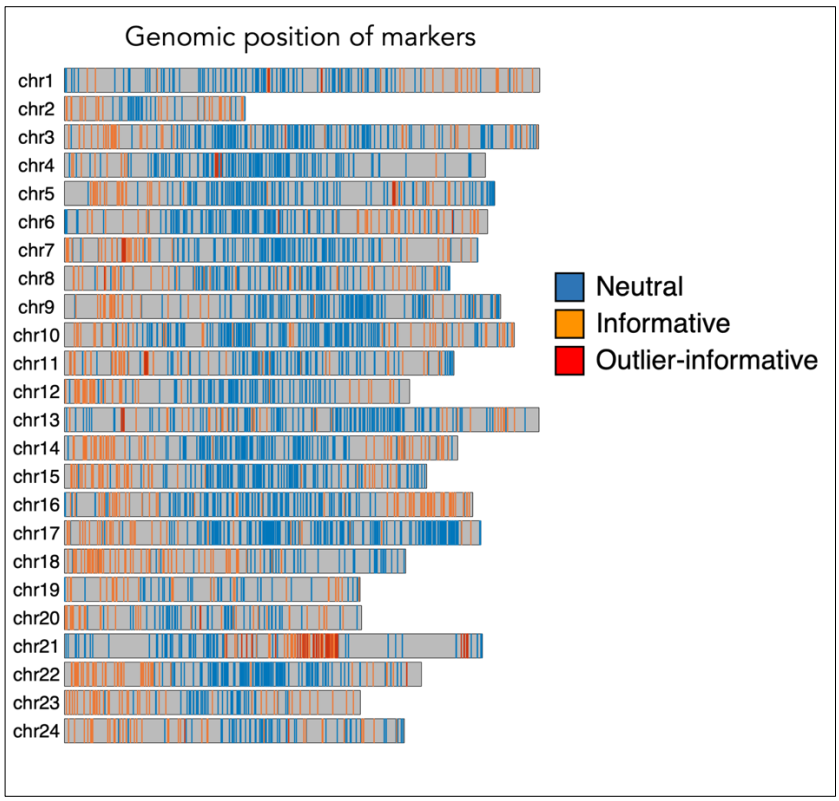


Figure 8. Genomic position of genetic markers included in the Axiom SNP array for the Atlantic horse mackerel.

To this end Uppsala University and Identigen Ltd. (Dublin, Ireland) developed a multispecies Axiom® SNP genotyping array, containing c.3,000-4,000 SNPs each for seven commercial fish species: Atlantic herring (*Clupea harengus*), European sprat (*Sprattus sprattus*), Atlantic horse mackerel (*T. trachurus*), brown trout (*Salmo trutta*), Atlantic salmon (*Salmo salar*), European perch (*Perca fluviatilis*), and Atlantic cod (*G. morhua*). The horse mackerel component of the array comprised 4,242 SNPs of which 1,578 were outlier SNPs and 2,664 putatively neutral SNPs. The workflow followed for marker selection for the SNP array is summarized in Figure S1 in Annex 2 and was not part of the current study. In order to the select the makers for inclusion from the ~12.8 million available, a pre-selection of outlier SNPs was first performed based on their high variance between pools in the existing horse mackerel Pool-Seq dataset (Fuentes-Pardo et al., 2023). The resulting marker set, which covered all 24 chromosomes (Figure 8), was complemented with the 50 most differentiated SNPs from each of the divergent genomic regions identified through genome scans and with previously tested SNPs in the IdentiGEN platform in Fuentes-Pardo et al. (2023).

The aims of the current study were to;

1. Genotype a large set of baseline and potentially mixed horse mackerel samples collected over multiple years from across the three stock areas in the northeast Atlantic.
2. Confirm the population structure identified in the Pool-Seq analyses.
3. Develop an assignment model to assign non-spawning individuals to their population of origin.
4. Use the assignment model to investigate mixing of populations across stock boundaries.

3. Materials and Methods

3.1 Sampling, DNA extraction, and SNP genotyping

Samples were collected opportunistically between 2015 and 2022 through exiting scientific surveys and from both target and non-target fisheries. Specific effort was directed at identifying spawning areas and times in order to collect baseline spawning samples. Effort was also focussed on sampling potential mixing zones identified in the review of historical data. Each fish was measured for total length (TL) to the 0.5 cm below, total body weight (TW) to the nearest 1.0 g and sex and maturity were assessed. As maturity stages were recorded using a number of different maturity keys depending on the country of collection, these were standardised to the six-point international horse mackerel maturity scale (ICES, 2015).

For the majority of samples collected from 2015 to 2021 a c.0.5 cm^3 piece of tissue was excised from the dorsal musculature of each specimen and stored at -20°C in absolute ethanol. After 2021 most samples were collected using the LVL Technologies Genetic Sampling Tool (GST) system, comprising barcoded racks and tubes with an integrated sample collection tool. The samples collected along the Norwegian coast by IMR consisted of fin tissue stored in absolute ethanol.

2,304 samples were sent to Identigen Ltd. (Dublin, Ireland) for magnetic bead based DNA extraction and genotyping with the multispecies Axiom® SNP genotyping array (FSHSTK1D). The samples collected using the GST system were sent directly without the need for subsampling. Samples collected using standard methods were subsampled and 30mg of tissue or fin from each individual transferred to a deep-well plate with ethanol and heat sealed prior to shipping.

Two *Trachurus trecae* samples from the Mauritanian samples, previously identified through COI barcoding in Farrell & Carlsson (2018), were also included in order to test the ability of the genotyping array to distinguish closely related species that may occur in the sampling area. The Mediterranean samples included in Fuentes-Pardo et al. (2020;2023) were not available for analysis in the current study.

Table 2. Conversion of maturity scales to the 6-pt international horse mackerel maturity scale.

International Stage	Name	Walsh (1990)	Walsh (IEO)	Walsh (IPMA)	MSS	MI	IMR
1	Immature	1	1	1	1	1-2	1-2
2	Developing	2-3	2-3	2-3	2	3-4	3-4
3	Spawning	4-5	4-5	4-5	3-4-5	5-6-7	5-6
4	Regressing	6	6	6	6	8-9	7-8
5	Omitted spawning						
6	Abnormal						

3.2 Genotype data quality control

Initial quality control of the genotype data was performed by IdentiGEN following the manufacturers recommended standard protocols (Anon, 2023: Annex 3).

Raw genotype data were downloaded from the IdentiGEN Info Centre as a Variant Call Format (VCF) file. Data were filtered by “Conversion Type” and scrutinised for patterns indicative of errors. After filtering, data were extracted and also converted to Genepop format for some of the further analyses.

Further quality control of the data was performed at a finer scale once the panels of markers had been refined. At a minimum a baseline individual sample had to be genotyped at greater than 90% of the

available SNPs to be retained in the dataset and a marker had to be genotyped in greater than 90% of the individual samples to be retained.

3.3 Preliminary analyses and population genetic structure

The primary aim of the data exploration and preliminary analyses of population structure in the current study were to confirm the results of Fuentes-Pardo et al. (2020; 2023) and to identify a panel of informative markers for population assignment. The analytical approaches followed were tailored to this specific task. Many of the SNPs used in the current study were high-graded to maximise the power of discrimination between the populations identified by Fuentes-Pardo et al. (2020; 2023). Therefore, the dataset may not be suitable for conventional population genetic analyses and as such some of the analyses presented (e.g. estimation of fixation indices) were for exploratory purposes only. More detailed analyses of population structure with a view to more academic topics e.g. evolutionary considerations and Environmental Association analyses are ongoing and are beyond the scope of the current working document.

The data exploration and preliminary analyses were initially conducted on all samples, including the two *Trachurus trecae* samples from Mauritania. In subsequent analyses the *T. trecae* samples were removed from the dataset. Exploratory analyses were first undertaken at the individual fish level through Principal Component Analysis (PCA) using the function *snpgdsPCA* in the *R* package *SNPRelate* (Zheng et al., 2012). Analyses were conducted using the full SNP dataset and also subsets with either the putatively neutral or outlier SNPs as defined in Fuentes-Pardo et al. (2020; 2023).

Subsequent analyses were undertaken at the sample group level and included estimation of multi-locus pairwise F_{ST} and visualisation of the results, to enable exploration of the relationships between different samples, through Principal Coordinate Analysis (PCoA) using the covariance standardised method conducted in GenAlEx 6.51b2 (Peakall and Smouse, 2012). Discriminant analysis of Principal Components (DAPC), from the *R* package *adegenet* (Jombart, 2008; Jombart et al., 2010), is a multivariate approach that transforms multi-locus genotype data using PCA to derive a set of uncorrelated variables, which serve as input for discriminant analysis (DA). The DA aims to maximize among-group variation and minimize within-group variation. DAPC does not make assumptions of underlying population genetic processes (e.g. neutrality, linkage equilibrium, Hardy–Weinberg equilibrium), therefore it was appropriate to use this approach with the data in the current study. DAPC was conducted at a number of different levels on the datasets based on refined panels of markers. DAPC cross-validation was conducted with the *xvalDapc* function in *adegenet*.

3.4 Informative marker identification and selection

Based on the SNP array data, the most informative and cost-effective panels of SNPs for baseline delineation of horse mackerel were selected following two approaches.

3.4.1 Approach 1

It was first determined whether neutral and outlier markers offered equivalent power to identify population structure. Patterns of population structure of baseline (maturity stage 3) and potential baseline (any maturity) samples were compared by performing separate PCAs using only using neutral or outlier markers. Having identified which marker type provided higher population resolution, various quality control filters were applied to retain SNPs with the highest power of discrimination between populations and also highest genotyping rate. Markers with low information content, such as those that were monomorphic or had a minor allele frequency less than 5%, and markers with a missing genotyping rate greater than 10% were excluded. Of the resulting markers, the ten most differentiated in each of the diagnostic genomic regions identified through genome scans comparing allele frequencies between groups of populations in Fuentes-Pardo et al. (2023) (based on their delta allele frequency, dAF, and required dAF > 0.37) were selected. The genotypes of the informative SNPs

in each population, compared across samples, were examined and removed if they deviated from expectations (i.e. if they did not discriminate the population(s) of interest). The aim was to keep five markers per diagnostic genomic region and to ensure a spread across the genomic region by removing SNPs that were too close to each other (< 1 kbp), as they provided equivalent information. A level of redundancy was retained in the SNP panel to ensure that all diagnostic genomic regions were represented even if some probes failed due to poor DNA quality or technical artifacts. A non-redundant minimal SNP panel was also generated by applying linkage disequilibrium (LD) pruning using a r^2 threshold of 0.2.

3.4.2 Approach 2

An alternative approach was also employed to identify the most informative SNPs for including in the assignment model SNP panel for the western versus North Sea model. In this approach DAPC was first performed and then the contribution of individual variables analysed to determine the primary variables responsible for the structure observed. The *loadingplot* function in *adgenet* was used to visualise the contribution of the variables to the differentiation of the two populations. The top 5% of variables were extracted from the DAPC results and up to 3 of the top ranked SNPs were selected from each unlinked genomic region, ensuring that there was a minimum of 1kbp between them. A second iteration of the SNP panel was also generated were only the top SNP in each region was retained in order to avoid including linked SNPs in the panel.

3.5 Assignment model development

The *R* package *assignPOP* (Chen *et al*., 2018), which performs population assignment using a machine-learning framework, was used to develop the assignment model. *assignPOP* uses Monte-Carlo cross-validation (*assign.MC*) to divide the baseline data into a training dataset and test dataset. The assignment model is developed with the training dataset and subsequently tested with the independent test dataset, which avoids introducing 'high-grading bias' (see Anderson, 2010). As the Monte-Carlo procedure samples random individuals each time, it does not guarantee that every individual is sampled. Therefore, *assignPOP* can perform an additional method of *K*-fold cross-validation (*assign.kfold*), which involves randomly dividing the individuals from each population into *K* groups and then using one group from each population as test individuals and the remaining *K-1* groups as the training individuals. Assignment tests are performed until every group and hence individual is tested, resulting in *K* tests. *assignPOP* has a number of classification model options including the *SVM* model from the *R* package *e1071* (Meyer *et al*., 2015). In order to avoid overfitting the model and to objectively determine the optimum number of PCs to be used in the *SVM* tuning in *assignPOP*, *DAPC* cross-validation was conducted with the *xvalDapc* function in *adegenet*. Exploratory analyses in *assignPOP* determined that the optimum model and kernel for the assignment model were the *SVM* model and the radial basis function (RBF) kernel. Exploratory analyses and the *tune*, *tune.control* and *best.svm* functions in *R* package *e1071* (Meyer et al., 2015) were used to perform a grid search for the optimum values for cost and gamma. These parameters were used for testing the rate of self-assignment using both Monte-Carlo cross-validation (*assign.MC*) and also *K*-fold cross-validation (*assign.kfold*) to estimate membership probability. Both Monte-Carlo and *K*-fold cross-validation were performed using 25%, 50%, 75% of the highest F_{ST} loci.

One important consideration when developing the baseline is to determine how many genetic markers are required for accurate assignment. This will enable the threshold for missing data of the unknown samples to be set with a robust basis without compromising the integrity of the assignments. In order to do this the Monte-Carlo cross validation analyses were run again with random sampling of loci (*loci.sample="random"*) rather than highest F_{ST} loci (*loci.sample="fst"*). In this instance 25%, 50% 75% and all the loci were randomly tested to determine the rate of accuracy of self-assignment.

3.6 Assignment of mixed samples

Following the validation of the assignment model and marker panels the potentially mixed samples from the Western and North Sea stock areas (Table S2) were assigned to population of origin using the *svm* model in *assignPOP* with the model parameters derived from the assignment model development in Sections 3.5 and 4.5. Mixed samples were compiled into 6 groups based on temporal and spatial relatedness and relevance to stock identification questions (Tables 3 & S2). One of the potential baseline samples (sample #2/PTB_3), collected in the central North Sea in 2016 and previously analysed in Fuentes-Pardo et al. (2023), that was not included in the baselines was included with the mixed samples for assignment.

Table 3. The assignment groups for the mixed samples assigned to population of origin.

Assignment group	Location	ICES Area	Quarters analysed	Current assumed Stock
1	West of Ireland	7.b	3	Western
2	Northern North Sea	4.a	1, 2	North Sea
3	Northern North Sea	4.a	3, 4	Western
4	Central/Southern North Sea	4.b-c	3	North Sea
5	Eastern Channel	7.d	4	North Sea
6	Western Channel	7.e	3	Western

Individuals that were submitted for genotyping but failed to yield amplifiable DNA or pass initial QC by IdentiGEN were marked as Fail (*F*). Based on the results of the assignment model testing (Section 4.5) the QC genotyping threshold of the mixed samples was set at a conservative level of 75% i.e. individuals had to be successfully genotyped at ≥ 75% of the SNPs in the panel of markers in order to be retained in the dataset for assignment. Individuals below this threshold we marked as Not Assigned (*NA*).

A further assignment threshold was set for the assignment probability at 0.67. This level indicated a situation where one outcome was twice as likely as the alternate outcome (Table 4) and was deemed appropriate. This is the same probability used for the assignment of herring in ICES division 6.a (Farrell et al., 2021; 2022). Individuals with an assignment probability less than 0.67 were marked as Below Threshold (*BT*) and were not considered to be assigned to either the Western (*WS*) or the North Sea (*NS*) populations.

Table 4. Conversion of probabilities to odds. The proposed threshold of 0.67 is indicted with the red line.

Probability	Odds [p/(1-p)]
0.1	0.11
0.2	0.25
0.3	0.43
0.4	0.67
0.5	1
0.6	1.5
0.67	2.03
0.7	2.33
0.75	3
0.8	4
0.85	5.67
0.9	9
0.91	10.11
0.92	11.5
0.93	13.29
0.94	15.67
0.95	19
0.96	24
0.97	32.33
0.98	49
0.99	99
0.999	999

4. Results and discussion

4.1 Sampling, DNA extraction, and SNP genotyping

In total 102 samples comprising 2,304 individuals were collected from across the three ICES stock areas and outgroup samples were also collected in north African waters (Figures 9 and S2-4 and Tables 2 and 3).

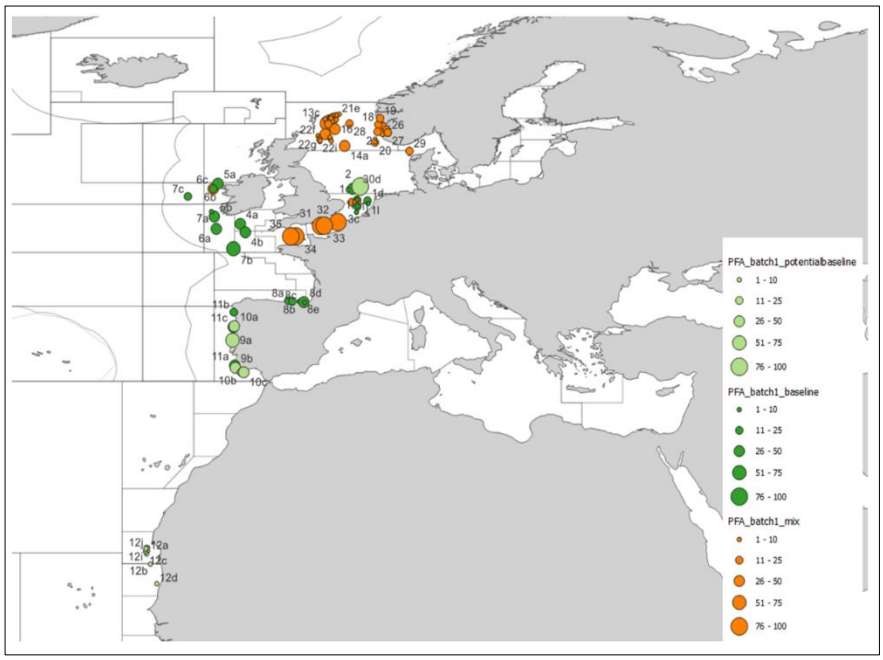


Figure 9. Overview of sampling locations and samples included in the current study. Dark green represents spawning baseline samples, Light green represents potential baseline samples and orange represents mixed samples for assignment.

Baseline spawning samples were collected from the western stock area from four years, from the North Sea from two years and from the southern stock area in a single year (Figures 9 & S2-4 and Tables S1). In the western area historical ichthyoplankton surveys reported in Ellis et al. (2012) suggest that there may have historically been a low level of spawning in the Irish Sea, and a higher degree of spawning in the outer Bristol Channel and eastern Channel. However during the course of the current study no spawning samples were collected in these areas and no spawning activity was reported by institutes or industry collaborators.

Samples that did not meet the strict definition of baseline spawning samples (i.e., mature fish collected at spawning time on the spawning grounds) were also collected and were classified as potential baseline samples. Some of these samples (e.g. the samples from the southern stock area collected in 2016 and 2017) were included in the analyses in Fuentes-Pardo et al. (2020; 2023) as no other baseline spawning samples were available at the time. Similarly, the North Sea potential baseline sample collected in 2016 was included in the analyses in Fuentes-Pardo et al. (2020; 2023) as the 2015 North Sea sample was not of high enough quality for Whole Genome Sequencing but in the current project was of sufficient quality for genotyping on the Axiom Array.

Some of the individual baseline and potential baseline samples were caught as bycatch, contained a small number of individual fish and were caught in close spatial and temporal proximity to other small samples by the same vessel. These samples were grouped where appropriate in order to ensure that each sample had a sufficient number of individuals to enable robust statistical analyses to be undertaken. This resulted in the 49 samples being grouped into 22 grouped samples (Table S1).

Analyses of the length frequency of the western and North Sea baseline and potential baseline samples by groups indicated that different cohorts were sampled in different years (Figure 10 and Table S3). For example, the 2015 North Sea baseline samples had a larger modal length than the 2017 North Sea baseline samples and thus were considered to comprise different cohorts. However, the 2016 North Sea potential baseline sample had a modal length smaller than the 2017 sample (Figure 10) and comprised juvenile fish (Table S3) as such it is not a true baseline sample and may also not be a true temporal replicate but simply the same year class sampled in consecutive years (Figure 10). Temporal replicates of spawning baseline samples were available in the Western and North Sea stock areas. In the southern stock area only a single year of baseline spawning samples was available (2019). It should be noted though that the southern spawning samples collected in 2019 displayed a bimodal length distribution with smaller fish collected in division 9.a south of Lisbon and larger fish collected in northern Portuguese waters. No spawning fish were collected off the southern Portuguese coast where juvenile fish were collected in 2017 (Figure S4 and Tables S3 & S4).

In addition to the baseline and potential baseline samples, potential mixed samples were also collected (Table S2). These were primarily collected outside of the spawning seasons in potential areas of mixing and comprised adult and juvenile horse mackerel (Figure 11 and Table S6). The main purpose of these samples was to assign the individual horse mackerel within them to population of origin with the assignment model that was developed on the baseline samples. The samples collected in division 4.a had a significantly larger modal length than those collected in divisions 4.b-c and 7.d-e. It should be noted that one sample collected in division 4.a in July 2016 comprised primarily stage 3 individuals. This was not considered a baseline spawning sample as it was not collected in a recognised spawning ground.

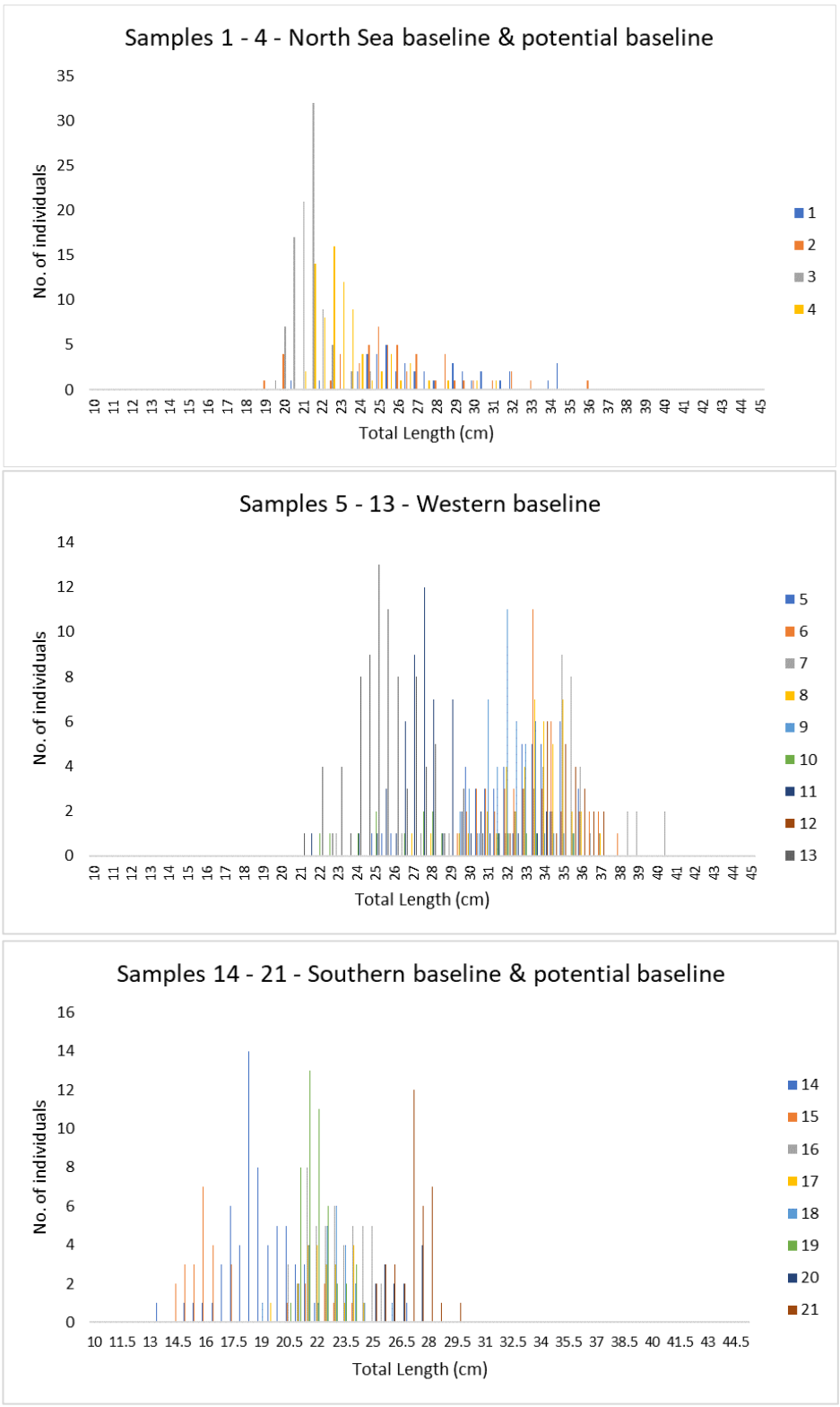


Figure 10. The length frequency of the baseline spawning and potential baseline samples by group in the current study. The sample details are provided in Table S1.

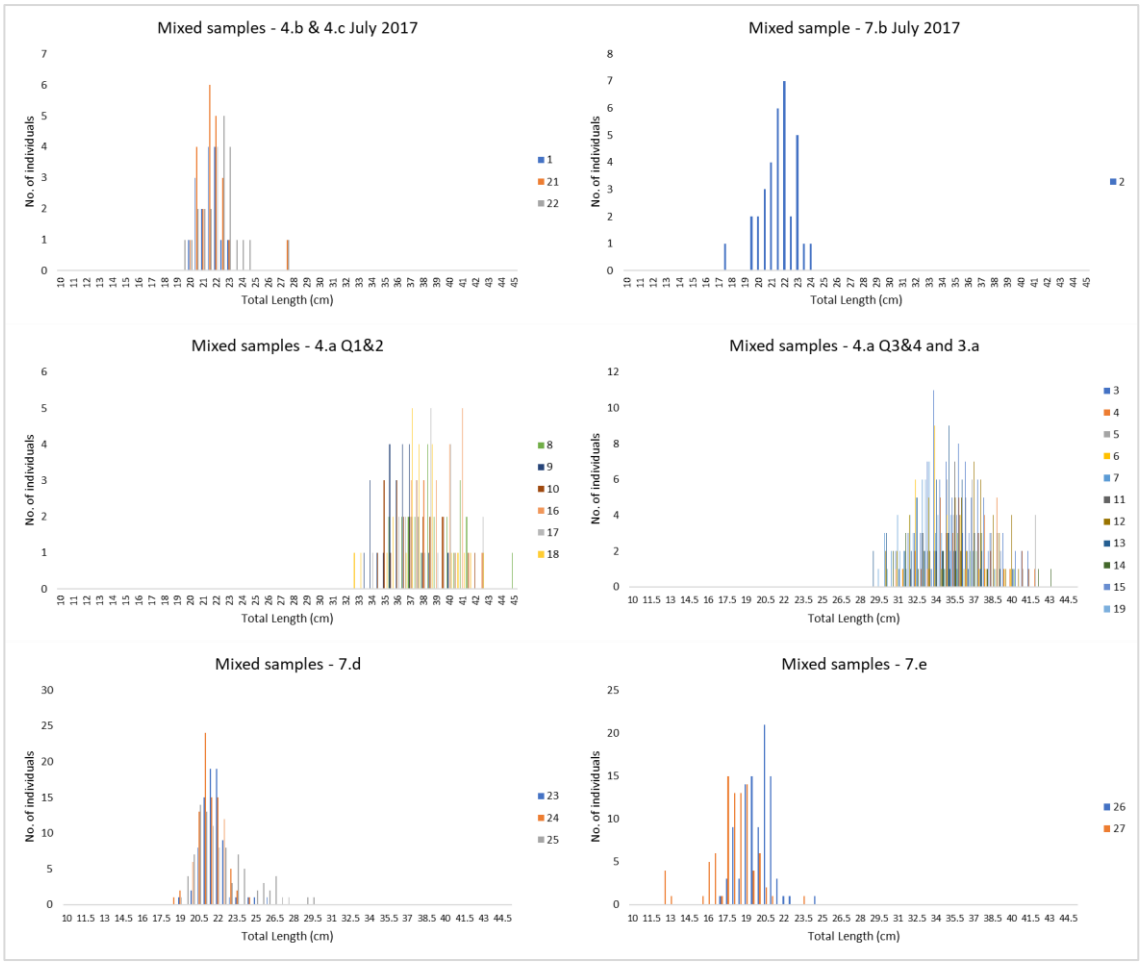


Figure 11. The length frequency of the mixed samples by group in the current study. The sample details are provided in Table S2.

4.2 Genotype Data quality control

2,287 out of 2,304 (99%) individual samples passed DNA quality control following DNA extraction and were genotyped with the Axiom Array. 2,166 of the 2,287 individual samples (95%) passed initial QC performed by IdentiGEN on the raw genotype data, of which 31% (n = 673) corresponded to baseline samples, 17% (n = 363) to potential baselines, 52% to potential mixed samples (n = 1128). Of the 4,103 horse mackerel SNPs on the array, 3,031 (74%) passed initial QC. Of the 3,301 SNPs, 2,550 (77%) were classified as “PolyHighResolution” and 32 as “CallRateBelowThreshold” and were retained in the dataset for further analyses (Figure S5 & Table S7). SNPs classified as “MonoHighResolution”, “NoMinorHom” or “OTV” were removed from the dataset and excluded from all further analyses.

Additional quality controls were applied at the marker panel level following the exploratory analyses and are detailed in the relevant sections below.

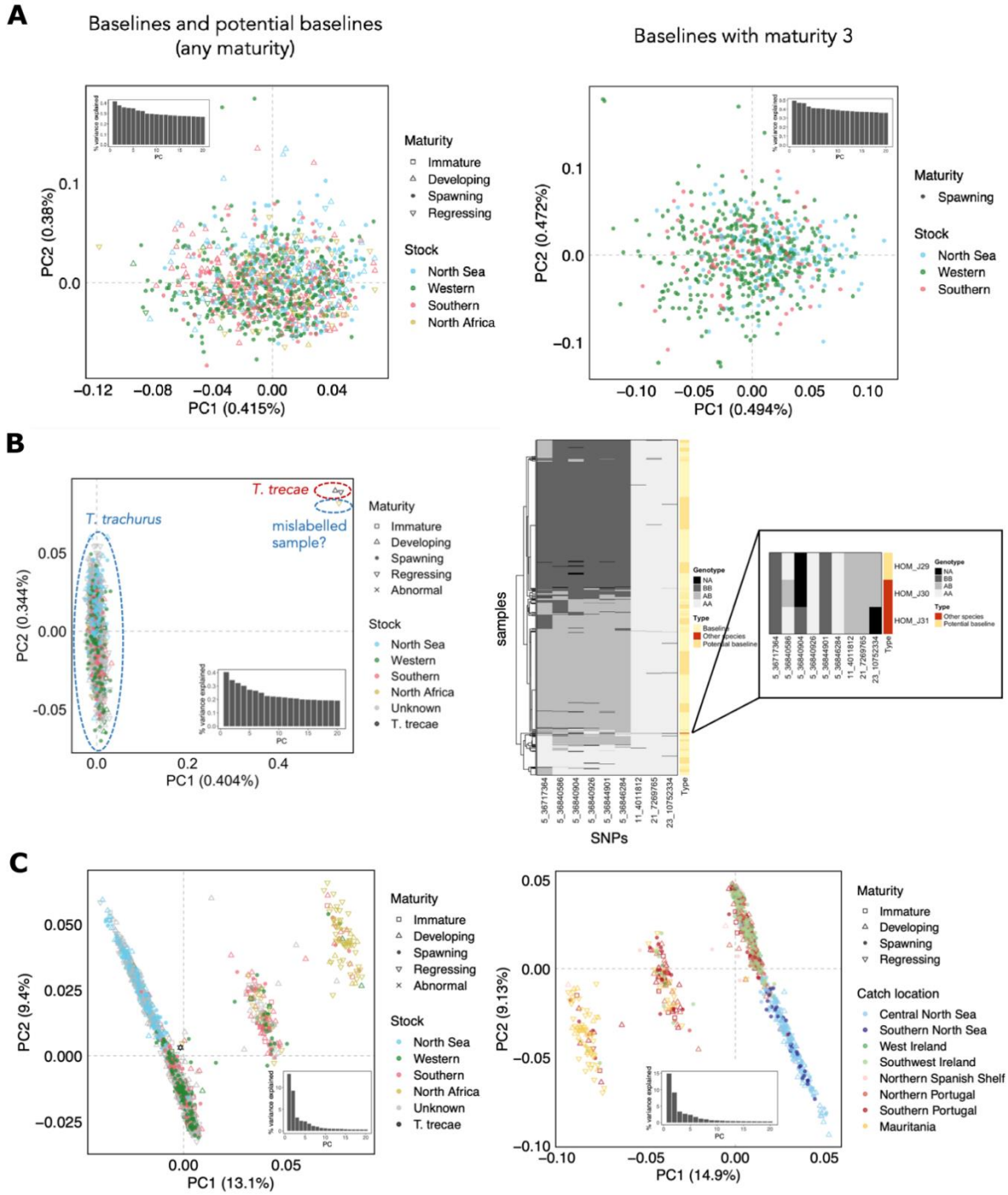


Figure 12. Individual clustering explored with PCA. (A-B) based on neutral SNPs. PCA plot for (A) only for *T. trachurus*, (left) baseline and potential baseline samples all maturity stages (right) only stage 3 spawning samples (B) all samples, including individuals from a sister species, *T. trecae*, (left) baseline samples with maturity stage 3, or (right) baseline and potential baseline samples. (C) based on outlier markers. Each dot corresponds to a single individual. (left) colored by stock of origin, or (right) sampling location.

4.3 Preliminary analyses and population genetic structure

Exploratory clustering analysis of baselines and potential baseline samples at the individual fish level with PCA revealed no genetic differentiation among *T. trachurus* samples using solely neutral genetic markers (Figure 12A). The only differentiation observed with neutral markers occurred between *T.*

trachurus and the related species *T. trecae*, which was driven by three SNPs (Figure 12B). The two known *T. trecae* samples were clearly differentiated and a third individual from the Mauritanian samples was also clustered with them. This highlighted the ability of the SNP array to distinguish different *Trachurus* species. These three individuals were removed from any further analyses.

The PCA based on outlier markers indicated a clear separation of the North Sea individuals from the other individuals from the Western and Southern stock areas and North Africa (Figure 12C), which is in agreement with the Pool-Seq and individual genotyping results of Fuentes-Pardo et al. (2020; 2023). Moreover, a similar 3-cluster pattern was again observed and was driven by markers of the putative inversion on chromosome 21 which distinguished most of the individuals from North Africa. As the putatively neutral markers provided no power to discriminate the populations no further analysis was undertaken with them. The selection of markers for baseline delineation was focussed on the set of outlier markers (see Section 4.4).

Additional analysis of F_{ST} of the baseline and potential baseline samples at the sample group level was performed in order to estimate the differentiation between samples (Table S8). PCoA of the F_{ST} results indicated a distinct pattern of four clusters (Figure 13). It is recognised that the dataset comprised a significant number of linked adaptive markers and as such the underlying assumptions of neutrality and independent markers were violated however the analysis does provide informative outputs regarding the relationship between the different sampling areas. The most differentiated cluster along the primary axis comprised the four North Sea samples, which were clustered close together. The North African sample was the most differentiated sample on the secondary axis. As expected, based on the results of Fuentes-Pardo et al. (2020; 2023) the west/southwest of Ireland (Western) and Northern Spanish Shelf (NSS) samples clustered together. Of note was the fact that all of the northern Portuguese (NPT) samples collected north of Lisbon also clustered with these samples as did the 2019 spawning baseline samples collected south of Lisbon. Three of the southern Portuguese (SPT) samples, which comprised mainly juvenile and developing fish (Figure 10 and Tables S3 & S4) clustered together between the Western/NSS/NPT samples and the NAF sample. Exploratory DAPC yielded similar results, though when the NAF was included it was difficult to see the distinction between the Western/NSS/NPT and 2016 and 2017 SPT samples (Figure 14). This was resolved by removing the NAF samples and rerunning the analyses (Figure 14). DAPC was also performed using the *find.clusters* function to estimate the most likely number of clusters in the combined baseline and potential baseline dataset. Three clusters (k=3) was the optimum number to resolve the data into geographically meaningful clusters (Figure S6). The output of the DAPC was plotted using the *compoplot* function in *adegenet* (Figure 15).

Overall the analyses agreed with the results of Fuentes-Pardo et al. (2020; 2023) in that the spawning horse mackerel samples from the southern North Sea were significantly different from the spawning horse mackerel sampled in the other stock areas i.e. they are a locally adapted biological unit, which for the purposes of this study are referred to as a population. The spawning horse mackerel sampled to the west and southwest of Ireland were not differentiated from the spawning horse mackerel samples along the Northern Spanish Shelf or in Portuguese waters (north or south of Lisbon). They were also not differentiated from the juvenile and developing horse mackerel sampled in northern Portuguese waters in 2016 and 2017. They were however differentiated from the southernmost samples collected to the southwest and south of Portugal (Figure S4). Therefore, as concluded in Fuentes-Pardo et al. (2020; 2023) the horse mackerel in division 9.a appears to comprise two populations; the first being the western population and the second being a more southerly population that is closer related to the North African samples. The *compoplot* (Figure 15) also indicates that there

is likely a degree of mixing between populations with increasing level of mixing towards the south of Portugal.

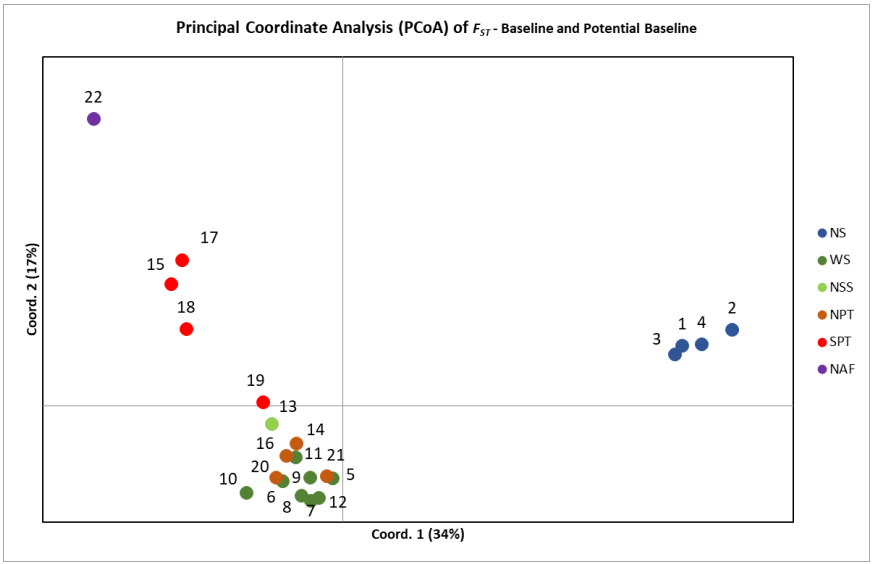


Figure 13. Principal Coordinate Analysis (PCoA) of F_{ST} of baseline and potential baseline samples. NS = North Sea, WS = Western, NSS = Northern Spanish Shelf, NPT = Northern Portugal, SPT = Southern Portugal, NAF = North Africa. The sample details are provided in Table S1.

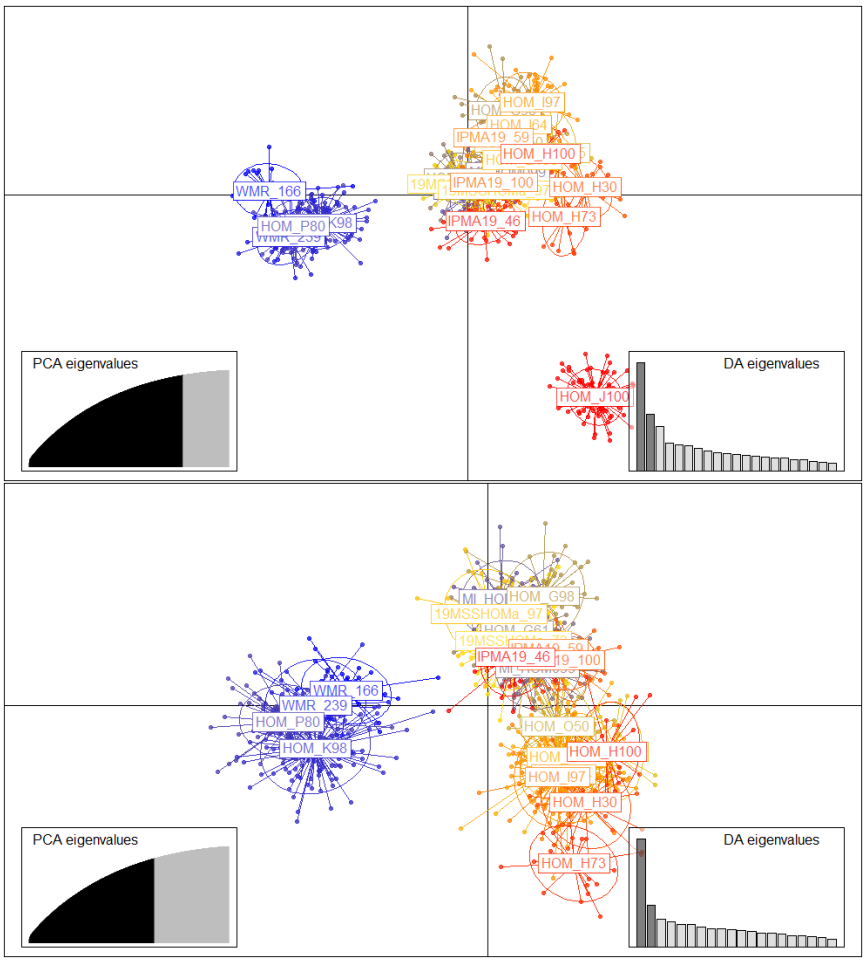


Figure 14. (Top) DAPC of baseline and potential baseline samples (Bottom) DAPC of baseline and potential baseline samples excluding the North African sample.

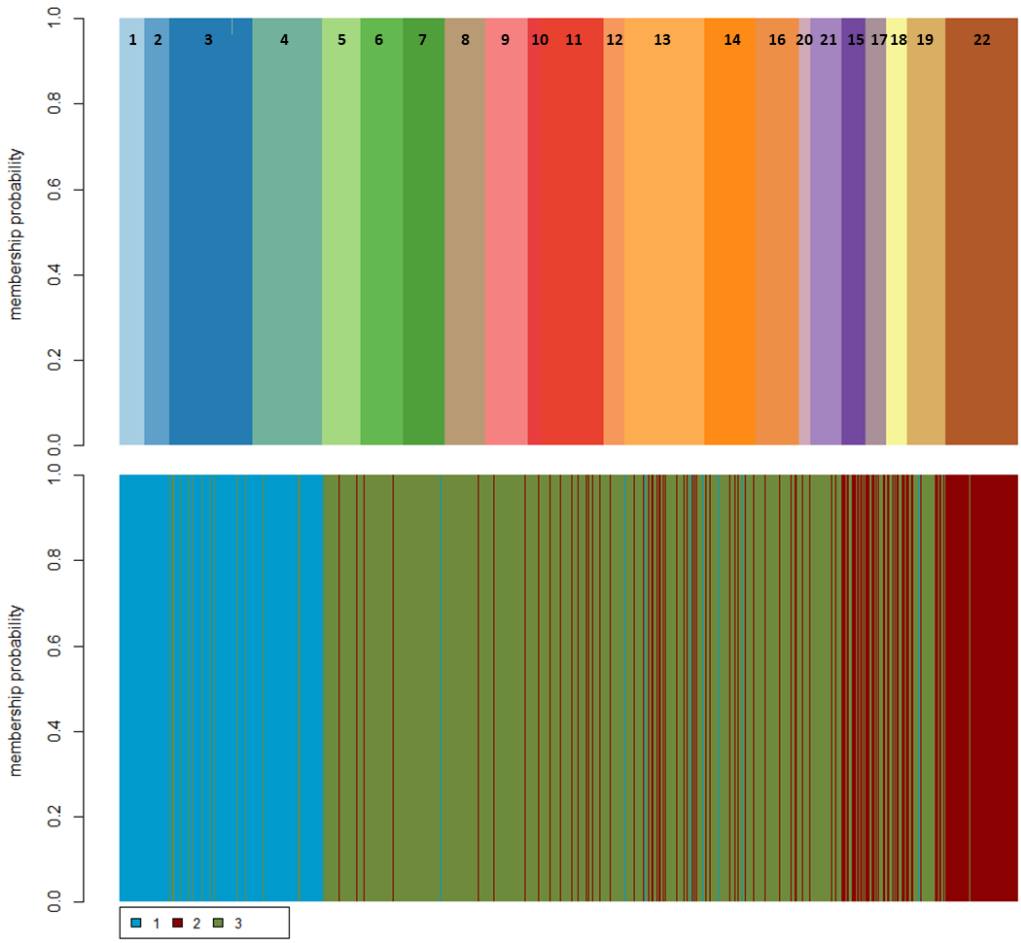


Figure 15. The output of the *find.clusters* informed *DAPC* of the baseline and potential baseline sample plotted using the *compoplot* function in *adegenet*. The top panel indicates the 22 sample groups with each individual represented by a column and the bottom panel indicates the group each individual was assigned to. The sample details are provided in Table S1.

In order to refine the dataset in preparation for developing an assignment model it was necessary to exclude the potential baseline samples and any individuals in the baseline samples that were not classified as maturity stage 3 as these did not meet the strict criteria for spawning baseline samples. The quality control criteria were implemented on the revised dataset and only SNPs genotyped in ≥ 95% of individuals were retained and only individuals that were genotyped with ≥ 95% SNPs were retained in the dataset. The resulting data set comprised 2,421 SNPs and 596 individual fish (Table 5). The 2016 North Sea, 2016 and 2017 Northern Portuguese and the North African samples did not contain any maturity stage 3 individuals and as such were excluded from the analyses. The 2016 and 2017 southern Portuguese samples only contained 2 and 6 stage 3 individuals, respectively. These were combined into a single sample for the initial exploratory analyses. The F_{ST} of the revised dataset was analysed at the sample group level and the results again visualised through PCoA. The pattern of differentiation of the samples was the same as previously observed with three clusters; southern North Sea, West and southwest of Ireland/Northern Spanish Shelf/Portuguese spawners north and south of Lisbon, southern Portugal (Figure 16).

There was an insufficient number of southern Portugal individuals (n=8) with which to build a baseline for an assignment model. And the remaining spawning individuals from Portuguese waters were genetically indistinguishable from the spawning individuals from the western spawning population. Therefore, for the purposes of developing an assignment model it was decided to only consider the

Western (west and southwest of Ireland and Northern Spanish Shelf) and the southern North Sea populations and to exclude the samples collected in division 9.a from further analyses. The resulting dataset comprised 109 and 383 individuals from the North Sea and the Western populations, respectively, and was carried forward to the informative marker identification and selection stage (Sections 3.4 and 4.4).

Table 5. The number of individual maturity stage 3 horse mackerel in each of the sample groups.

Sample Group	Catch Location	ICES Area	Exploratory Mat 3 genotyped
1	Central North Sea	4.b	20
2	Southern North Sea	4.c	19
3	Central North Sea	4.b	0
4	Central North Sea	4.b	70
5	Southwest Ireland	7.g	44
6	Southwest Ireland	7.h	48
7	West Ireland	7.b	39
8	Southwest Ireland	7.j	44
9	Southwest Ireland	7.j	47
10	West Ireland	7.b	16
11	Southwest Ireland	7.j	68
12	West Ireland	7.c	22
13	Northern Spanish Shelf	8.c	55
14	Northern Portugal	9.a	0
15	Southern Portugal	9.a	2
16	Northern Portugal	9.a	0
17	Southern Portugal	9.a	2
18	Southern Portugal	9.a	6
19	Southern Portugal	9.a	45
20	Northern Portugal	9.a	13
21	Northern Portugal	9.a	36
22	Mauritania	NA	0

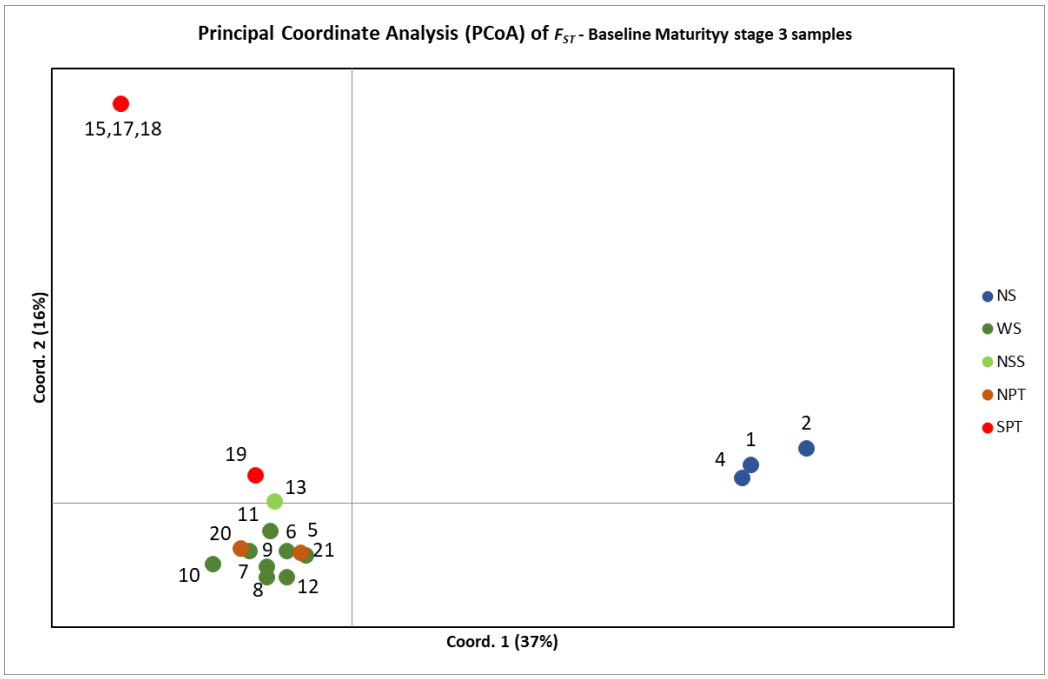


Figure 16. Principal Coordinate Analysis (PCoA) of F_{ST} of baseline and potential baseline samples. NS = North Sea, WS = Western, NSS = Northern Spanish Shelf, NPT = Northern Portugal, SPT = Southern Portugal. The sample details are provided in Table S1.

4.4 Informative marker identification and selection

4.4.1 Approach 1

Approach 1 was based at the individual fish level again. The highest ranked SNPs per informative genomic region in the pairwise contrasts were selected and the pattern of genotypes across the baseline and potential baseline samples compared visually and also compared with the dAF from the existing Pool-Seq data (Figure 17; Fuentes-Pardo et al. 2023). It was evident that there were technical issues with some of the SNPs which had not been filtered out in the preceding QC steps. Some regions also appeared not to be particularly informative and were excluded. As the outlier markers diagnostic of the North Sea population on chromosomes 4, 7, and 11 showed inconsistent patterns in both, the Pool-Seq and the individual genotype datasets, all of them were excluded. Given that no Mediterranean samples were included in the current analysis, the outlier markers from chromosome 5 were also removed. Therefore, the redundant marker set for population differentiation consisted of a total of 25 outlier SNPs. If required three neutral SNPs capable of distinguishing *T. trachurus* from the sister species *T. trecae* could also be added to the panel if north African samples were to be assigned with the resulting assignment model. In the current study these were excluded. An 8 SNP marker panel was also derived by removing the redundancy and only retaining the highest ranked SNP per genomic region. In summary two marker panels were derived from Approach 1: *25_SNP* and *8_SNP* (Table S9).

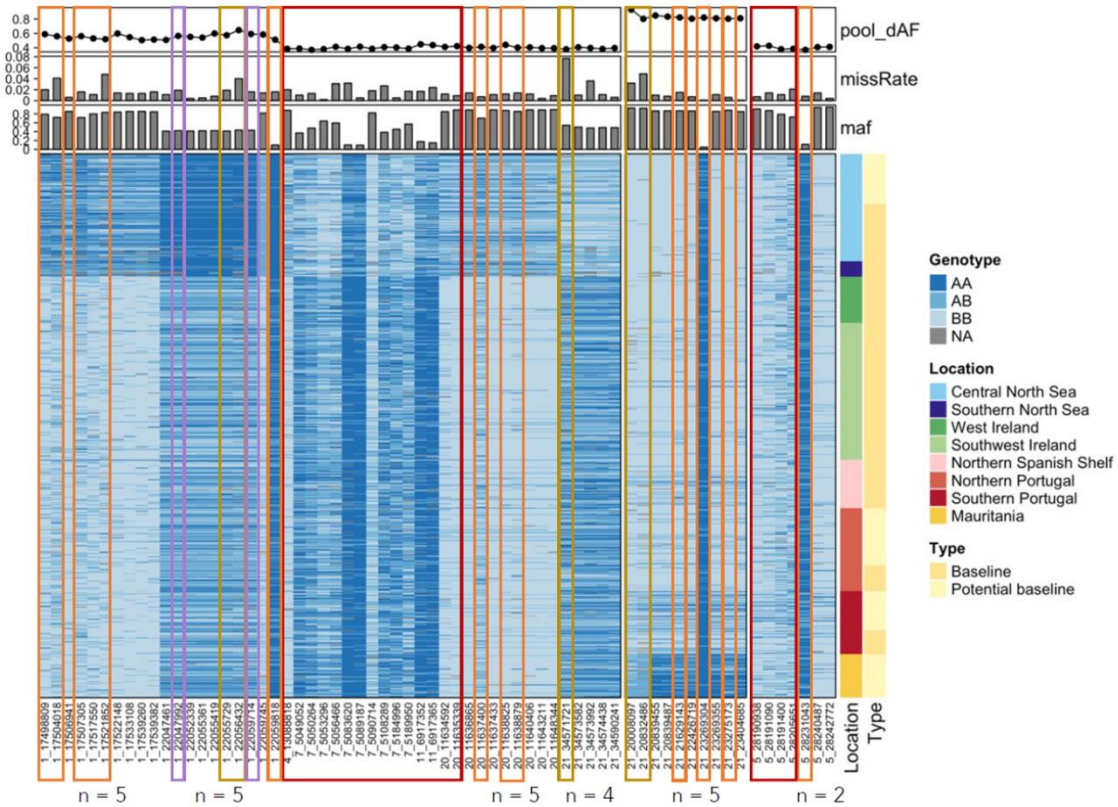


Figure 17. The visual analysis of the highest ranked SNPs per informative genomic region in the pairwise contrasts and the pattern of genotypes across the baseline and potential baseline samples. The dAF from the existing Pool-Seq data is also included at the top for comparison.

4.4.2 Approach 2

In Approach 2 the samples within the Western-North Sea baseline dataset were pooled by population and analysed through *DAPC* and *PCA* again. *DAPC* cross-validation indicated 200 PCs was the optimum. *DAPC* was run again with the sample groups and also with the sample groups pooled by population (Figure 18). PCA was also performed on the dataset with samples pooled by population (Figure 18). The *loadingplot* function in *adgenet* was used to visualise the contribution of the variables to the differentiation of the two populations. The top 5% of variables were extracted from the DAPC results and up to 3 SNPs were selected from each unlinked genomic region, ensuring that there was a minimum of 1kbp between them. This resulted in the *36_SNP* marker panel and the related *17_SNP* marker panel when only the top SNP per region was selected to avoid linkage (Table S9).

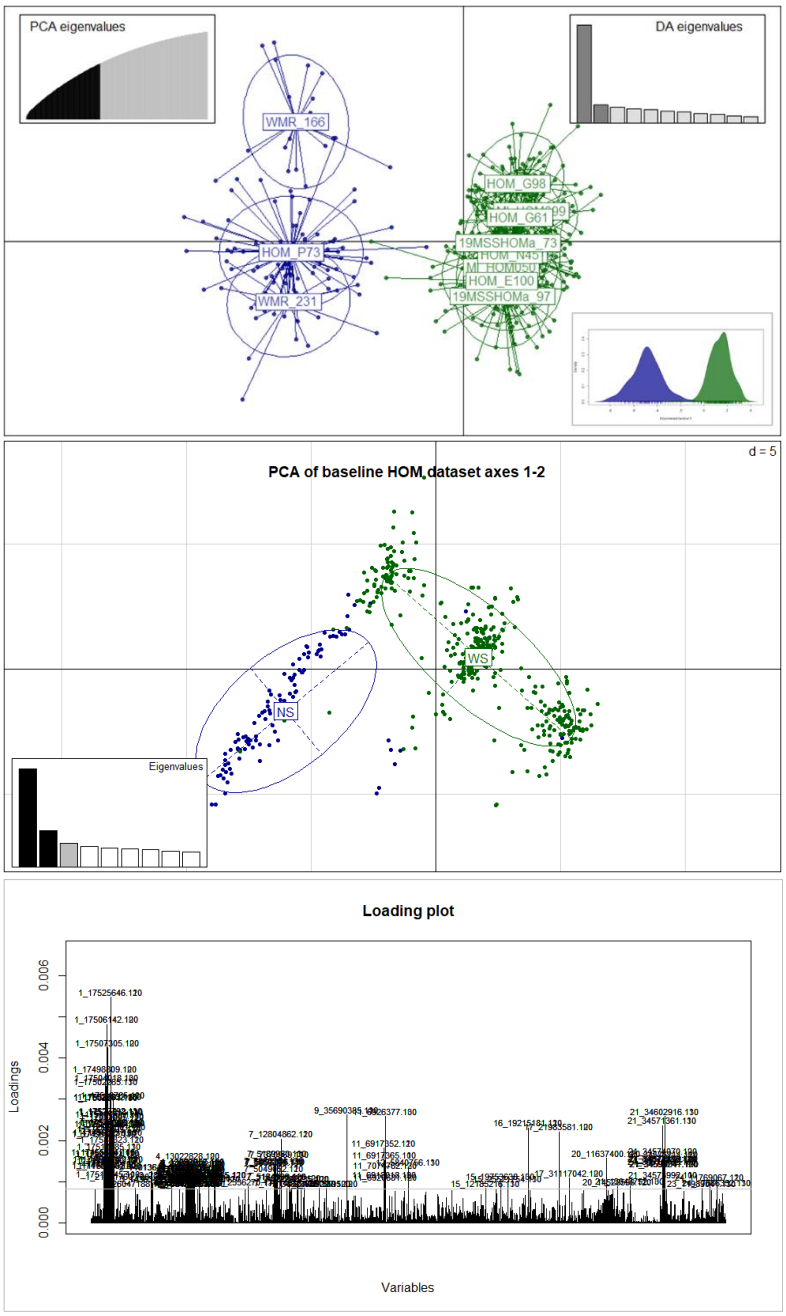


Figure 18. (top) *DAPC* and (middle) *PCA* of the maturity stage 3 North Sea and Western baseline samples. The inset in top panel is samples pooled by population. (bottom) loading plot of the variable contribution in the *DAPC* pooled by population.

Table 6. The number of individual fish in each of the sample groups in each of the baseline assignment datasets.

Sample Group	Catch Location	ICES Area	2421_SNP	25_SNP	8_SNP	36_SNP	17_SNP
1	Central North Sea	4.b	20	22	18	20	20
2	Southern North Sea	4.c	19	21	19	19	19
4	Central North Sea	4.b	70	70	79	70	70
5	Southwest Ireland	7.g	44	44	43	44	44
6	Southwest Ireland	7.h	48	49	45	47	48
7	West Ireland	7.b	39	44	46	39	39
8	Southwest Ireland	7.j	44	45	44	44	44
9	Southwest Ireland	7.j	47	47	48	47	47
10	West Ireland	7.b	16	16	15	16	16
11	Southwest Ireland	7.j	68	70	72	68	68
12	West Ireland	7.c	22	23	23	22	22
13	Northern Spanish Shelf	8.c	55	61	85	55	55

The four datasets were subjected to the 90% quality control threshold where only individuals genotyped at ≥ 90% of the markers and markers genotyped at ≥ 90% of the individuals were retained in the datasets (Table 6). The exploratory F_{ST}/*PCoA*, *DAPC* and *PCA* analyses were conducted again to estimate the potential of the different marker panels, to discriminate the populations, relative to the complete *2421_SNP* dataset (Figures 19, 20, 21). The F_{ST}/*PCoA* indicated that the primary axes, representing the differentiation between the NS versus WS samples, accounted for greater than 80% variation in the analyses with the 25_SNP and 8_SNP (*Approach 1*) panels and greater than 70% variation with the 36_SNP and 17_SNP (*Approach 2*) panels (Figure 19).

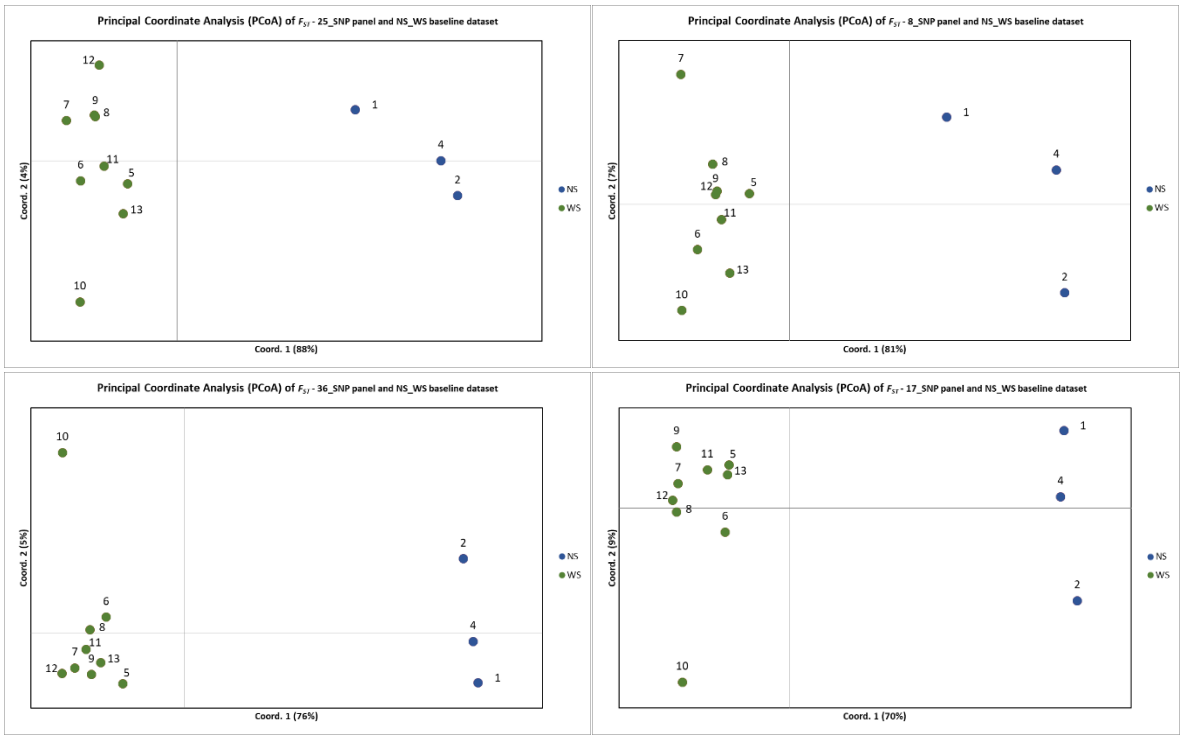


Figure 19. Principal Coordinate Analysis (*PCoA*) of F_{ST} of North Sea and Western baseline samples of the (top left) *Approach 1 - 25_SNP* panel (top right) *Approach 1 – 8_SNP* panel (bottom left) *Approach 2 – 36_SNP* panel (bottom right) *17_SNP* panel.

The secondary axes in all four cases explained less than 10% of the variation. Sample group #10 appeared to have some degree of differentiation associated with it, particularly in the *Approach 2*

analyses. The reason for this is unclear at present and will be investigated further. It had no bearing on the development of assignment models in the current study.

The *DAPC* and *PCA* analyses indicated, visually, that there was a higher level of differentiation between the samples from the North Sea and Western populations with the *Approach 2* marker panels (Figures 20 and 21). In order to quantitatively assess the ability of the marker panels in an assignment model the four reduced panels and the complete dataset (Tables 5 & S9) were carried forward for testing of self-assignment rates.

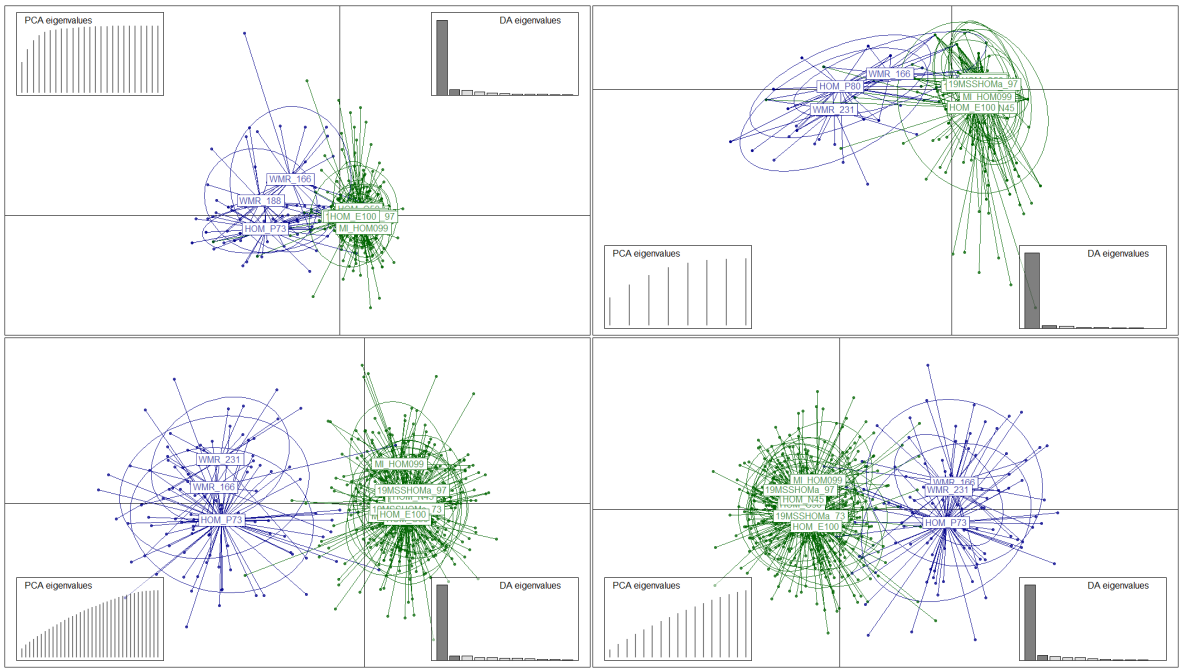


Figure 20. *DAPC* of North Sea and Western baseline samples of the (top left) *Approach 1 - 25_SNP* panel (top right) *Approach 1 – 8_SNP* panel (bottom left) *Approach 2 – 36_SNP* panel (bottom right) *17_SNP* panel.

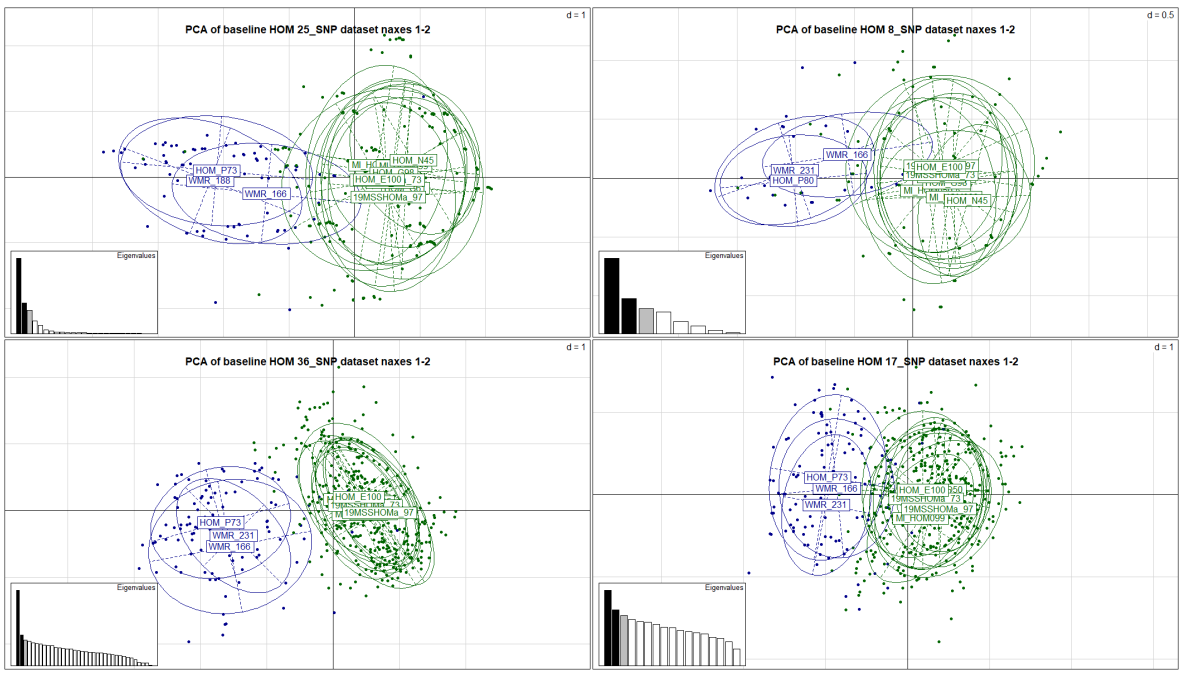


Figure 21. *PCA* of North Sea and Western baseline samples of the (top left) *Approach 1 - 25_SNP* panel (top right) *Approach 1 – 8_SNP* panel (bottom left) *Approach 2 – 36_SNP* panel (bottom right) *17_SNP* panel.

4.5 Assignment model development

In the Monte-Carlo cross-validation 25, 50 and 75 individuals were retained in the baseline training dataset and in the *K*-fold cross validation *K* groups was tested at 3, 5 and 10. The small number of individuals in the training dataset was due to the smaller number of individuals in the North Sea baseline dataset. The optimum model tuning parameters are provided in Table S10.

Monte-Carlo cross-validation and *K*-fold cross-validation of the *2421_SNP* baseline indicated a higher level of self-assignment in the Western population (99%) than the North Sea (93%), however both were at an acceptable level (Figure 22 and Table 7). A small number of outliers were evident when only 25 individuals were used in the training dataset, which is not unusual, and indicated that a larger number of training individuals were required. The most stable assignments were achieved with 75 individuals in the training dataset. It was also clear than not all of the 2421 SNPs were required for an accurate assignment and even 25% of the highest F_{ST} SNPs would likely suffice.

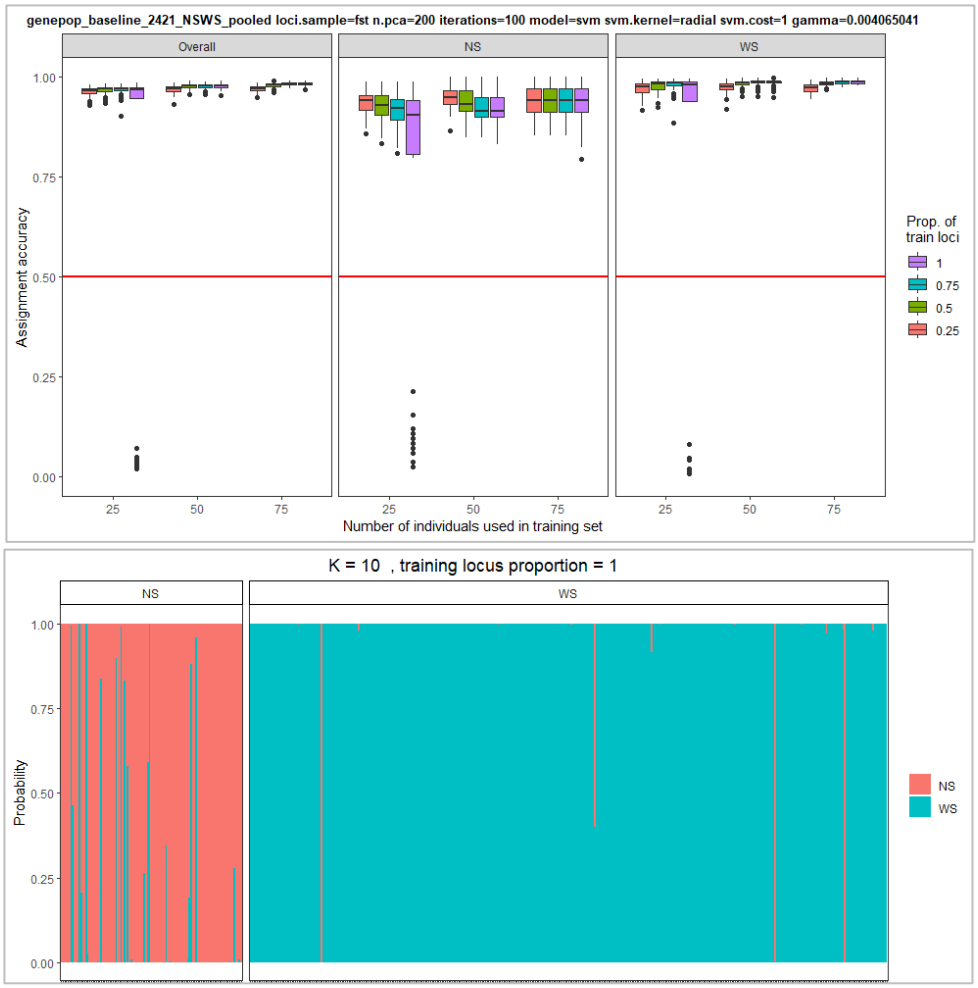


Figure 22. (top) Monte-Carlo cross-validation and (bottom) *K*-fold cross-validation of the *2421_SNP* baseline dataset assignment. NS = North Sea, WS = Western.

The assignments with the *25_SNP* and *8_SNP* panels derived from the *Approach 1* marker selection contrasted significantly with each other. The *25_SNP* panel had a self-assignment rate of 91% and 93% for the NS and WS populations, respectively (Table 7) and the Monte-Carlo and *K*-fold cross-validation plots indicated a relatively stable assignment when greater than 75% of the loci were used (Figures 23 and 24). Conversely the *8_SNP* panel was highly uncertain with only 27% and 67% self-classification rate for the NS and WS, respectively (Figure 23 & Table 7) and as such was considered to be unsuitable for use as with the SVM based assignment model.

Table 7. Assignment matrix for the Monte-Carlo cross-validation and *K*-fold cross-validation of the North Sea (NS) versus Western (WS) baseline dataset with the five different marker panels. The Monte-Carlo outputs are based on a training dataset of 75 individuals and training loci =1. The *K*-fold outputs are based on K=10 and training loci =1. SD = standard deviation.

Panel		Monte-Carlo Cross Validation		
		NS	**WS**	SD
2421_SNP	**NS**	0.93	0.07	± 0.04
	WS	0.01	0.99	± 0.00
25_SNP	**NS**	0.91	0.09	± 0.04
	WS	0.07	0.93	± 0.01
8_SNP	**NS**	0.27	0.73	± 0.27
	WS	0.33	0.67	± 0.29
36_SNP	**NS**	0.94	0.06	± 0.04
	WS	0.04	0.96	± 0.01
17_SNP	**NS**	0.71	0.29	± 0.08
	WS	0.29	0.71	± 0.03

Panel		*K*-fold Cross Validation		
		NS	**WS**	
2421_SNP	**NS**	0.89	0.11	± 0.12
	WS	0.01	0.99	± 0.02
25_SNP	**NS**	0.79	0.21	± 0.13
	WS	0.03	0.98	± 0.03
8_SNP	**NS**	0.71	0.29	± 0.12
	WS	0.04	0.96	± 0.02
36_SNP	**NS**	0.90	0.10	± 0.08
	WS	0.02	0.98	± 0.03
17_SNP	**NS**	0.78	0.22	± 0.14
	WS	0.03	0.98	± 0.04

The assignments with the *36_SNP* and *17_SNP* panels derived from the *Approach 2* marker selection performed better than the *Approach 1* panels. The *36_SNP* panel had a self-assignment rate of 94% and 96% for the NS and WS populations, respectively (Table 7) and the Monte-Carlo and *K*-fold cross-validation plots indicated a relatively stable assignment when greater than 50% of the loci were used (Figures 23 and 24). It also appeared to be more stable and displayed less uncertainty compared to the *25_SNP* panel. The *17_SNP* panel had a self-assignment rate of 71% for both the NS and WS, was also considered to be unsuitable for use as with the SVM based assignment model.

The sensitivity testing of the markers in the baseline panels confirmed that the *2421_SNP* panel had the most redundancy and was not sensitive to the number of randomly chosen SNPs down to at least 25% of the total (Table S11). The *25_SNP* and *36_SNP* panels retained greater than 90% accuracy down to at least 75% of the total SNPs, which also indicated an acceptable level of redundancy. The self-assignment rate of the *8_SNP* panel increased with less than 100% of markers, though it was still not considered reliable enough for use in an assignment model. The reason for the increase was unclear and will be investigated further at a later date. The uncertainty with the *17_SNP* panel increased significantly when using less than 75% of the markers (Table S11).

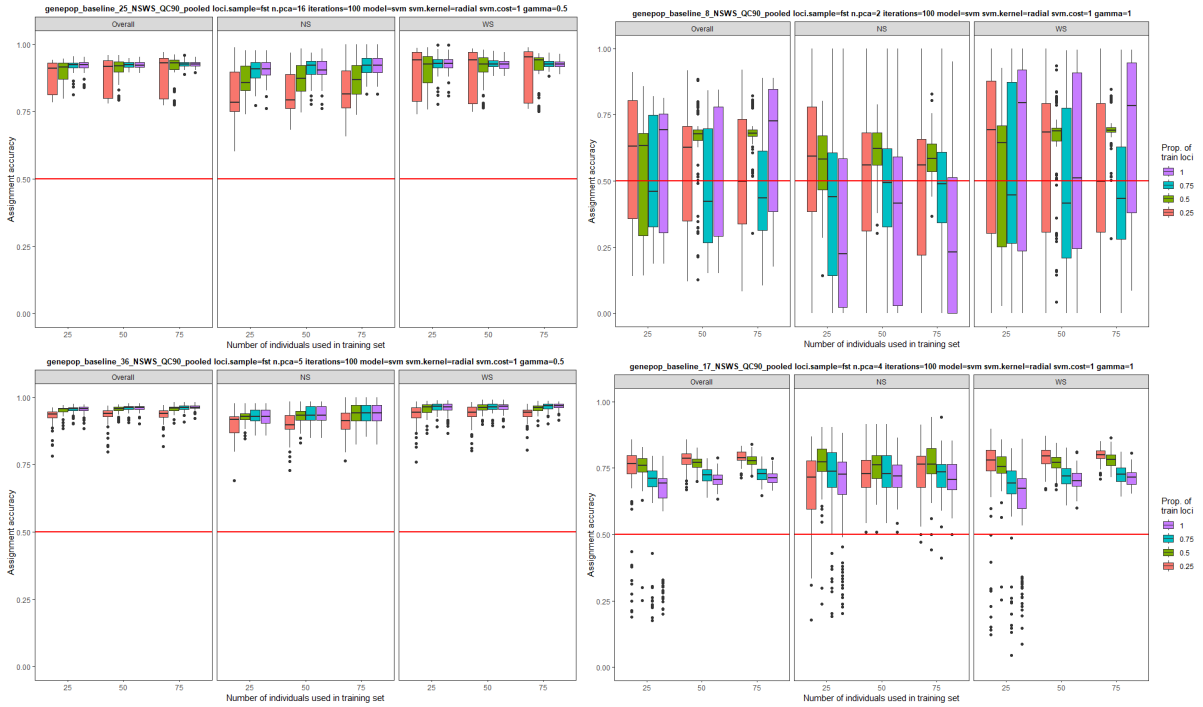


Figure 23. Monte-Carlo cross-validation of the (top left) *25_SNP* (top right) *8_SNP* (bottom left) *36_SNP* (bottom right) *17_SNP* baseline dataset assignments. NS = North Sea, WS = Western.

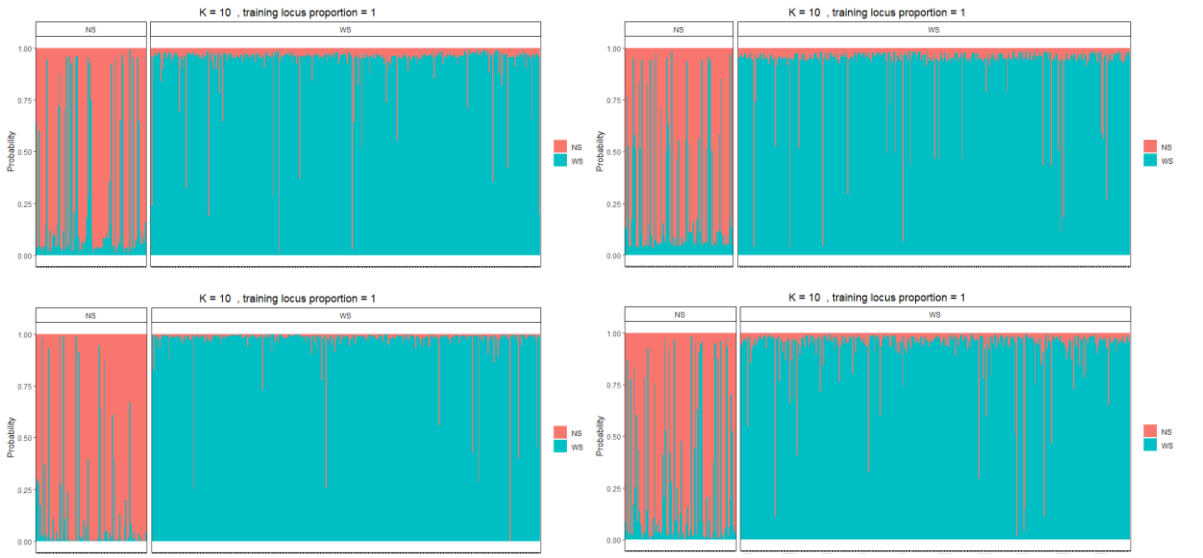


Figure 24. *K*-fold cross-validation of the (top left) *25_SNP* (top right) *8_SNP* (bottom left) *36_SNP* (bottom right) *17_SNP* baseline dataset assignments. NS = North Sea, WS = Western.

In summary three marker panels tested with the multivariate assignment models were considered to be suitable for further exploration (*2421_SNP, 25_SNP and 36_SNP*) and were used to assign the mixed samples in Section 4.6.

4.6 Assignment of mixed samples

The results of the assignments with the three marker panels were quite similar with the main difference being the higher number of *Below Threshold* individuals in the assignments with the *25_SNP* and *36_SNP* panels (Table 8). This is not unexpected given the smaller number of markers in these panels and the lower demonstrated assignment accuracy relative to the *2421_SNP* panel (Section 4.5). The majority of the samples classified as Fail were from the Norwegian coast, which were fin tissue as opposed to muscle tissue, and this was likely a result of sample degradation prior to preservation in ethanol. The majority of the *Below Threshold* individuals were from Assignment Groups 4 and 5 in the central/southern North Sea and Eastern Channel (57% in *2421_SNP*, 81% in *25_SNP* and 55% in *36_SNP*). These were the areas with the highest proportion of North Sea horse mackerel and it is possible that the higher level of *Below Threshold* individuals was due to a higher degree of cross contamination in these samples as they were collected prior to the adoption of the LVL GST protocol. The *Not Assigned* category only comprised a small number of individuals in each case and was not deemed significant.

Table 8. The results of the assignment of the mixed samples with the three marker panels. The number of individuals assigned to each category is indicated. WS = Western, NS = North Sea, BT = Below Threshold, NA = Not Assigned, F= Fail.

		2421_SNP					25_SNP					36_SNP				
Assignment Group	**Sample group**	**WS**	**NS**	**BT**	**NA**	**F**	**WS**	**NS**	**BT**	**NA**	**F**	**WS**	**NS**	**BT**	**NA**	**F**
1	**2**	33	0	0	0	1	33	0	0	0	1	33	0	0	0	1
2	**8**	24	0	0	0	0	24	0	0	0	0	23	0	1	0	0
2	**9**	22	1	0	0	1	20	2	0	1	1	22	1	0	0	1
2	**10**	23	0	0	0	1	23	0	0	0	1	23	0	0	0	1
2	**16**	21	0	0	0	3	21	0	0	0	3	20	0	1	0	3
2	**17**	23	0	0	0	1	23	0	0	0	1	23	0	0	0	1
2	**18**	19	2	1	1	1	19	3	0	1	1	18	4	0	1	1
3	**3**	94	2	0	0	0	93	3	0	0	0	94	2	0	0	0
3	**4**	45	2	0	0	0	45	1	1	0	0	45	2	0	0	0
3	**5**	48	1	0	0	0	48	1	0	0	0	47	2	0	0	0
3	**6**	44	3	1	0	0	43	3	2	0	0	44	4	0	0	0
3	**7**	47	0	0	0	1	47	0	0	0	1	45	0	2	0	1
3	**11**	3	0	0	0	21	2	0	0	1	21	1	0	0	2	21
3	**12**	41	6	0	0	1	42	5	0	0	1	41	4	2	0	1
3	**13**	46	0	0	0	2	45	0	1	0	2	45	1	0	0	2
3	**14**	13	0	0	0	11	12	1	0	0	11	12	0	1	0	11
3	**15**	5	1	0	1	17	6	1	0	0	17	4	1	1	1	17
3	**19**	11	0	0	1	0	12	0	0	0	0	10	0	0	2	0
3	**20**	11	0	0	0	11	11	0	0	0	11	10	0	1	0	11
4	**1**	1	13	1	0	1	1	13	1	0	1	2	11	2	0	1
4	**21**	3	20	0	0	0	4	19	0	0	0	3	20	0	0	0
4	**22**	4	21	0	0	0	4	20	1	0	0	3	22	0	0	0
4	**PTB_3**	12	82	2	0	0	18	76	2	0	0	17	74	5	0	0
5	**23**	53	43	0	0	0	55	38	3	0	0	54	40	2	0	0
5	**24**	21	75	0	0	0	22	64	10	0	0	22	70	4	0	0
5	**25**	26	67	1	0	2	31	58	4	1	2	31	60	3	0	2
6	**26**	82	12	1	0	1	84	11	0	0	1	82	10	3	0	1
6	**27**	79	9	0	0	8	78	9	1	0	8	76	11	1	0	8
Totals		854	360	7	3	84	866	328	26	4	84	850	339	29	6	84

For ease of exploring the results the six assignment groups were plotted as pie charts in a combined map for each marker panel (Figure 25). The results for each assignment group were also plotted separately as pie charts (Figures S7-S11) and as membership plots with associated probabilities (Figures S12-S17).

All individuals in Assignment Group 1 collected west of Ireland were assigned to the Western Stock with a high probability (Figures S7 & S12).

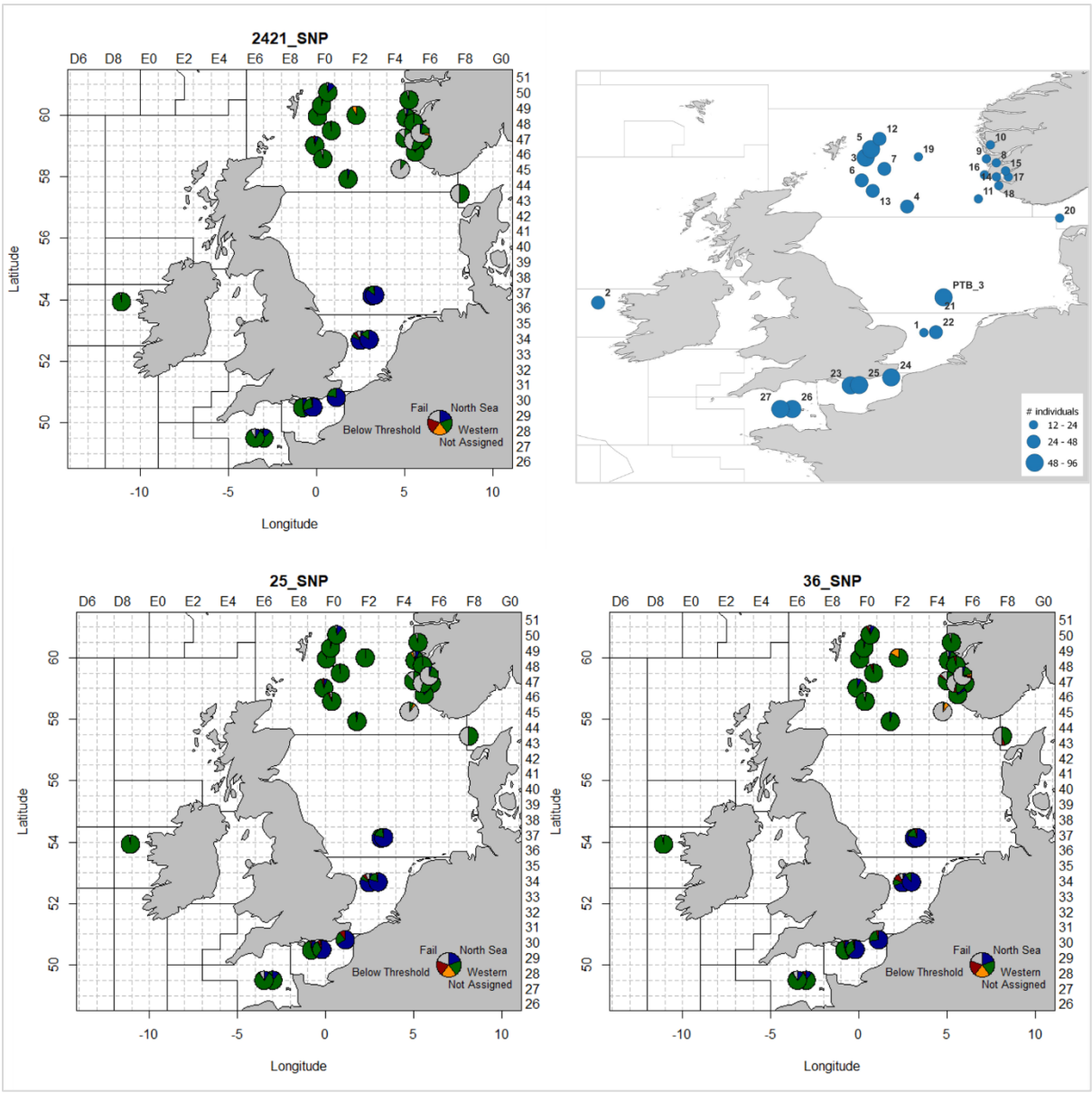


Figure 25. The combined outputs of the assignment of the mixed samples with the three marker panels.

Table 9. The percentage assignments to each assignment group (excluding the *Not Assigned* and *Fail*) categories for each of the marker panels.

	2421_SNP			25_SNP			36_SNP		
Assignment Group	**WS**	**NS**	**BT**	**WS**	**NS**	**BT**	**WS**	**NS**	**BT**
1	100.00	0.00	0.00	100.00	0.00	0.00	100.00	0.00	0.00
2	97.06	2.21	0.74	96.30	3.70	0.00	94.85	3.68	1.47
3	96.23	3.54	0.24	95.53	3.53	0.94	94.54	3.80	1.66
4	12.58	85.53	1.89	16.98	80.50	2.52	15.72	79.87	4.40
5	34.97	64.69	0.35	37.89	56.14	5.96	37.41	59.44	3.15
6	87.98	11.48	0.55	88.52	10.93	0.55	86.34	11.48	2.19

The Assignment Group 2 samples were collected in the northern North Sea (division 4.a) in quarters 1 and 2 in 2019 and 2022 and assumed to consist of North Sea horse mackerel according to the current stock delineation. However, the majority of samples assigned to the Western population (97% in *2421_SNP*, 96% in *25_SNP* and 95% in *36_SNP*) and only 2-4% were assigned to the North Sea population (Figures 25, S8, S13 and Table 9). This result indicated that, based on the samples analysed, there was little support for the assumption that the horse mackerel in division 4.a in quarters 1 and 2 originate from the North Sea population.

The Assignment Group 3 samples were collected in the northern North Sea (division 4.a) in quarters 3 and 4 in 2016, 2017, 2019-2022 (Table S2) and as such were assumed to consist of Western horse mackerel. This was confirmed with the majority of samples assigned to the Western population (96% in *2421_SNP*, 96% in *25_SNP* and 95% in *36_SNP*) and only 4% assigned to the North Sea population (Figure 25, S9, S14 and Table 9).

Conversely the Assignment Group 4 samples collected in the central and southern North Sea in July 2016 and 2017 assigned primarily to the North Sea population (86% in *2421_SNP*, 81% in *25_SNP* and 80% in *36_SNP*), though a significant proportion (13%, 17%, 16%) also assigned to the Western population (Figure 25, S10, S15 and Table 9). This highlights that there may still be a significant degree of mixing in parts of this area and further sampling is required to refine the delineation of the populations. The spatial and temporal gaps in sampling are evident in Figure 25 where there is a large gap between the division 4.a and 4.b samples.

The three samples from the eastern Channel (Assignment Group 5) were collected in the same year (2020) in three different months (September, November and December). Overall the samples assigned primarily to the North Sea population (65% in *2421_SNP*, 56% in *25_SNP* and 59% in *36_SNP*), though a significant proportion (35%, 38%, 37%) also assigned to the Western population (Figure 25, S11, S16 and Table 9). There was also a temporal and possible spatial pattern evident between the months with the September sample, which was the furthest west, comprising more Western than North Sea horse mackerel, whilst the November and December samples had similar proportions of 22-28% Western and 72-78% North Sea horse mackerel (Figures S11 & S16 and Table 9). These results confirm that there is a mixing issue in the eastern Channel and the current assumption that division 7.d is exclusively part of the North Sea stock is incorrect.

The two samples collected in the western Channel were collected on the same day in close proximity to each other and had a similar composition with the majority of fish being of Western origin (88% in *2421_SNP*, 89% in *25_SNP* and 86% in *36_SNP*) and the minority (11%) being of North Sea origin (Figures 25, S11, S17 and Table 9). It should also be noted that the modal length of individuals sampled was less than 20cm in the western channel and c.20cm in the eastern channel, highlighting the fact that these are important juvenile areas.

In summary the assignment of the mixed samples has highlighted that there are significant issues with the current stock delineation between the Western and North Sea stocks and that the stock areas do not align with the spatial distribution of the populations. Any of the three marker panels may be used for the assignment of mixed individuals but at this stage it is prudent to use the *2421_SNP* panel as it had the lowest error rate and the lowest number of individuals classified as *Below Threshold*. When new baseline samples are collected and processed the model will need to be retrained prior to rerunning the assignments.

5. Conclusions and summary

The review of the history and origin of the stock identification of horse mackerel (Section 2) highlighted the significant uncertainties in the current delineation of the stocks. The identification of the three stocks was based purely on the recognition of three potential spawning areas without significant evidence to support their discreteness. It is clear that the initial separation of the southern stock from the western stock was not based on any biological information but merely aligned with the southern boundary of the mackerel egg survey, south of which there was evidence of spawning but no survey coverage and limited data. Similarly, the boundary between the North Sea and the western stocks was based on noted spawning in the southern North Sea and the suggestion that this stock "*probably*" overwintered in the English Channel where it would "*mix to some extent*" with the western stock, which spawned to the west and southwest of Ireland. Further assumptions were made regarding the origin of the non-spawning horse mackerel in divisions 2.a and 4.a, which were assumed to belong to the western stock and likely mixing with the North Sea stock in division 4.b. The lack of evidence for the assumed limits of the three stocks was noted multiple times by the Working Groups and it was clearly stated that the egg and larvae distribution data was "*not sufficient evidence to infer independent stocks, as adult horse mackerel are highly mobile, and these areas may represent no more than three separate areas where spawning environments are favoured by the fish.*"

Regardless of the significant uncertainties and lack of robust evidence, the initial stock delineation was largely retained over the following twenty years and it shaped the subsequent direction of the data collection and stock assessments of the three stocks. It was not until the HOMSIR project (2000-2003) that the existing structure was challenged, and some changes made. However, the project noted that the population structure in the western European waters could be more complicated than the results suggested and that more research was needed to clarify the migration patterns within the Northeast Atlantic. This was particularly relevant to the potential mixing areas between the North Sea stock and the western stock (divisions 4.a and 7.d) and between the re-defined southern and western stocks (divisions 8.c and 9.a) where the sampling was relatively sparse. Despite this key recommendation little further work was conducted and the provisional stock boundaries suggested by HOMSIR have remained in place to the present day except for the arbitrary and unsupported reallocation of catches from quarters 1 and 2 in division 4.a to the North Sea stock instead of the western stock.

The current study has presented the most comprehensive investigation of horse mackerel stock structure in the northeast Atlantic area to date using the most advanced methods available for defining the biological units that underly the stocks. The analyses presented are robust and have been rigorously tested and the assignment model developed can be used to distinguish individuals from the western and North Sea populations. However, further work is required before a complete realignment of the stocks can be undertaken as detailed in Section 6.

The primary conclusions were that the horse mackerel that spawn in the southern North Sea comprise a locally adapted biological unit, which in this study is referred to as a population. Based on the samples analysed to date this population seems to have a limited distribution and occurs primarily in divisions 4.b and 4.c. It also occurs in division 7.d where it mixes with the western population and may be the dominant population at certain times of the year. It was also recorded as the minority component in samples from division 7.e and in very small numbers in samples from division 4.a.

The western horse mackerel population appears to have the widest distribution and ranges from division 4.a in the north, division 3.a in the east south to division 9.a south of Lisbon. Based on the samples analysed this population spawns to the west and southwest of Ireland, in the Bay of Biscay along the Northern Spanish Shelf and in Portuguese waters. It also occurs in divisions 7.e and 7.d in significant numbers and may also be present in divisions 4.b at certain times of the year.

The southern population was the least well sampled in the current study. A very small number of spawning individuals (<10) were collected in the south of division 9.a. These individuals were characteristic of the southern population, which is more closely related to the north African population than to the western population. Whilst it was not possible to develop an assignment model to distinguish the southern and western populations it was possible to conclude that there was mixing of non-spawning individuals between the western and southern populations along the Portuguese coast but the majority of the southern individuals were caught south of Lisbon.

In summary the key results were as follows:

- At least three populations of horse mackerel occur within the northeast Atlantic area.
- Individuals from the North Sea population and western populations may be assigned to population of origin with greater than 90% accuracy.
- Samples from division 4.a assigned primarily to the western population regardless of quarter.
- Samples from 7.d indicated a mix of western and North Sea populations.
- As a result the North Sea assessment is likely over-estimating abundance in the IBTS survey in 7.d and overestimating catch of North Sea individuals in 4.a and 7.d.
- The western stock area does not align with the distribution of the western population, which extends south into 9.a, east into 7.d and is the primary population found in 4.a.
- The indices used in the western assessment are likely underestimating recruitment and biomass as they do not contain the population area.
- The catch of the western population is underestimated as it does not account for the catches taken in 4.a in quarters 1 and 2 or catches in 7.d or 9.a which contain a high proportion of juvenile individuals.
- Horse mackerel from at least two populations occur within the southern stock area.
- Most samples from the southern area were genetically identical to the western population.
- The most southerly samples (Gulf of Cadiz) represent a different 'southern' population.
- The current stock delineation of the three stocks is not appropriate for the assessment of the three stocks.

6. Recommendations

The results in this working document support the assertion that the current stock delineation of the three horse mackerel stocks is not appropriate, and this should be considered in the forthcoming benchmark of the three stocks in 2024. The authors acknowledge that further work is required before a complete realignment of the stocks can be undertaken however at a minimum the current benchmark should account for the results and incorporate them as far as possible into the revised assessments. The genetic resources developed in the current study will be made opensource and are freely available to the benchmark process and to the institutes involved. It is imperative that the countries involved in the horse mackerel fisheries and the relevant institutes involved in the assessments initiate sampling programs and contribute to the further development of this work.

The following recommendations should be adopted to continue and further develop the work for the purposes of improving the assessments beyond the current benchmark process.

1. The assessment methods and models used for the three stock assessments should be aligned to facilitate future revision of the stock boundaries.
2. Within the 2024 benchmark a sensitivity analyses of the three assessments based on the results of the current study should be undertaken to test the impact of future revisions of the stock boundaries.
3. Consideration should be given to allocating all catches in division 4.a to the western stock regardless of quarter as there is no basis or evidence to support the current allocation approach.

4. As part of the benchmark a longer term (3-5 year) plan should be developed where significant changes will be made to the assessments after a complete revision of the stock boundaries of the three stocks based on widescale genetic analyses. This plan should include:
    a. Developing standard genetic sampling of baseline spawning samples of the three populations across entire spawning seasons and multiple years. This is particularly relevant to the North Sea and southern populations.
    b. Identifying and sampling spawning areas not sampled or analysed in the current study (e.g. see Ellis et al. 2012) in order to ensure the baselines are comprehensive and appropriate.
    c. Analysing catch data across three stock areas and developing a sampling plan for genetic splitting of catches into population of origin where necessary.
    d. Identifying the survey indices in the three assessments that will be required to be genetically split and determine the sampling levels required.
    e. A step-wise development of the new assessments over the course of the 3-5 year period prior to the next benchmark.
5. A key benchmark output should be a list of urgent actions that national administrations/institutes are advised to follow in order to implement the plan developed in point 4.
6. All samples collected for genetic analysis should be genotyped using the Axiom® SNP genotyping array (FSHSTK1D) or equivalent to ensure data generated is combinable with the data in the current study.
7. All samples collected for genetic analysis should comprise muscle tissue in preference to fin tissue as the results in the current project have highlighted to increased likelihood of genotyping failure for fin tissue samples.

7. Acknowledgements

The authors would like to thank Martin Pastoors, Maria Manuel Angélico, Finlay Burns, Gersom Costas, Cindy Van Damme, Cristina Nunes, Brendan O'Hea, Ciaran O'Donnell, Michael O'Malley, Niels Hintzen, Claus Reedtz Sparrevohn and all scientists and crew on the survey and commercial vessels involved in sampling. We thank the members of the Northern Pelagic Working Group of the European Association of Fish Producers Organisations for continued funding of the research programme and the Pelagic Advisory Council for funding and ongoing discussions of the implication of the results for the management of the horse mackerel stocks. We also thank the members of the ICES Working Group on Widely Distributed Stocks (WGWIDE) and ICES Working Group on Mackerel and Horse Mackerel Egg Surveys (WGMEGS) for their input. Finally we thank UPPMAX in Uppsala for providing the high-throughput sequencing and computational infrastructure under the NAISS projects 2020/5-36 and 2020/16-14.

8. References

Abaunza, P., Murta, A.G., Campbell, N. et al. 2008a. Stock identity of horse mackerel (*Trachurus trachurus*) in the Northeast Atlantic and Mediterranean Sea: integrating the results from different stock identification approaches. Fisheries Research, 89: 196–209.

Abaunza, P., Gordo, L.S., García Santamaría, M.T., Iversen, S.A., Murta, A.G., Gallo, E. 2008b. Life history parameters as an important basis for the initial recognition of stock management units in horse mackerel (*Trachurus trachurus*). Fisheries Research, 89: 167-180.

Abaunza, P., Gordo, L., Karlou-Riga, C., Murta, A., Eltink, A.T.G.W., Santamaria, M.T.G., Zimmermann, C., Hammer, C., Lucio, P., Iversen, S.A., Molloy, J., and Gallo, E. 2003. Growth and Reproduction of Horse Mackerel, *Trachurus trachurus* (Carangidae). Fish Biology and Fisheries, 13:27–61.

Andersson L, Bekkevold D, Berg F, Farrell ED, Felkel S, Ferreira M, Fuentes-Pardo AP, Goodall J, Pettersson M. 2024. How Fish Population Genomics Can Promote Sustainable Fisheries: a Road Map. Annual Review in Animal Biosciences, 12, 10.1146/annurev-animal-021122-102933.

Anderson, E.C. 2010. Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased. Molecular Ecology Resources, 10: 701–710.

Anon. 2023. Axiom™ Genotyping Solution Data Analysis. User guide. Publication Number MAN0018363. Revision C.0. 166 pp.

Bekkevold D, Berg F, Polte P, Bartolino V, Ojaveer H, et al. 2023. Mixed-stock analysis of Atlantic herring (*Clupea harengus*): a tool for identifying management units and complex migration dynamics. ICES Journal of Marine Science 80: 173-84

Brunel, T., Farrell, E.D., Kotterman, M., Kwadijk, C., Verkempynck, R., Chen, C and Miller, D. 2016. Improving the knowledge basis for advice on North Sea horse mackerel. Developing new methods to get insight on stock boundaries and abundance. Wageningen, IMARES Wageningen UR (University & Research centre), Wageningen Marine Research report C092/16 57 pp.

Canales-Aguirre, C.B., Ferrada, S. and Galleguillos, R. 2010. Isolation and characterization of microsatellite loci for the jack mackerel (*Trachurus murphyi* Nichols, 1920). Journal of Conservation Genetics, 11: 1235-1237.

Chen, K., Marshall, E.A., Sovic, M.G., Fries, A.C., Gibbs, H.L., Ludsin, S.A. 2018. assignPOP: An r package for population assignment using genetic, non-genetic, or integrated data in a machine-learning framework. Methods in Ecology and Evolution, 9: 439-446. (doi.org/10.1111/2041-210X.12897)

Cimmaruta, R., Bondanelli, P., Ruggi, A., Nascetti, G. 2008. Genetic structure and temporal stability in the horse mackerel (*Trachurus trachurus*). Fisheries Research, 89: 114-121.

Comesaña, A.S., Martínez-Areal, M.T., Sanjuan, A. 2008. Genetic variation in the mitochondrial DNA control region among horse mackerel (*Trachurus trachurus*) from the Atlantic and Mediterranean areas. Fisheries Research, 122-131.

Cornuet, J.-M., Piry, S., Luikart, G., Estoup, A. and Solignac, M. 1999. New methods employing multilocus genotypes to select or exclude populations as origins of individuals. Genetics, 15: 1989-2000.

Costa, A.M. 2022. First step for the implementation of visual assessment of the mesenteric fat content of horse mackerel, *Trachurus trachurus* (Linnaeus, 1758) in the Portuguese coast. Relatórios Científicos e Técnicos do IPMA nº35. 37p.

Ellis, J.R., Milligan, S.P., Readdy, L., Taylor, N. and Brown, M.J. 2012. Spawning and nursery grounds of selected fish species in UK waters. Sci. Ser. Tech. Rep., Cefas Lowestoft, 147: 56pp.

Eltink, A. 1992. Horse mackerel egg production and spawning stock size in the North Sea in 1991. ICES C.M. 1992/H:21

Eltink, A. 1990. Horse mackerel egg production and spawning stock size in the North Sea in 1989. ICES C.M. 1990/H:20

Eltink, A. 1985. Results of the Horse Mackerel (*Trachurus trachurus* L.) Otolith Exchange Program. ICES CM 1985/H:40.

Farrell, E.D., Andersson, L., Bekkevold, D., Campbell, N., Carlsson, J., et al. 2022. A baseline for the genetic stock identification of Atlantic herring, *Clupea harengus*, in ICES Divisions 6.a, 7.b-c. R Soc Open Sci 9: 220453

Farrell, E.D., Campbell, N., Carlsson, J., Egan, A., Gras, M., Lusseau, S.M., Nolan, C., O'Connell, S., O'Malley, M., and White, E. 2021 Herring in Divisions 6.a, 7.b and 7.c: Scientific Assessment of the Identity of the Southern and Northern Stocks through Genetic and Morphometric Analysis. Service Contract EASME/EMFF/2017/1.3.2.1/SI2.767459. DOI: 10.2826/208498

Farrell, E. D., & Carlsson, J. 2018. Genetic stock Identification of Northeast Atlantic Horse mackerel, *Trachurus trachurus*. A report prepared for the members of the Northern Pelagic Working Group. 40pp.

Farrell, E. D., Carlsson, J. E. L., & Carlsson, J. 2016. Next Gen Pop Gen: implementing a high-throughput approach to population genetics in boarfish (*Capros aper*). Royal Society Open Science, 3(12), 160651.

Fuentes-Pardo, A.P., Farrell, E.D., Pettersson, M., Sprehn, C.G., Andersson, L. 2023. The genomic basis and environmental correlates of local adaptation in the Atlantic horse mackerel (*Trachurus trachurus*). Evolutionary Applications, 16:1201-1219.

Fuentes-Pardo, A.P., Pettersson, M., Sprehn, C.G., Andersson, L. & Farrell, E.D. 2020. Population structure of the Atlantic horse mackerel (*Trachurus trachurus*) revealed by whole-genome sequencing. A report prepared for the members of the Northern Pelagic Working Group and the Pelagic Advisory Council. 38 pp.

Genner, M. & Collins, R. 2022. The genome sequence of the Atlantic horse mackerel, *Trachurus trachurus* (Linnaeus 1758). Wellcome Open Research, 7, 118.

Gordo, L.S., Costa, A., Abaunza, P., Lucio, P., Eltink, A.T.G.W. and Figueiredo, I. 2008. Determinate versus indeterminate fecundity in horse mackerel. Fisheries Research, 89: 181–185.

Han, F., Jamsandekar, M., Pettersson, M.E., Su, L., Fuentes-Pardo, A., Davis, B., Bekkevold, D., Berg, F., Cassini, M., Dahle, G., Farrell, E.D., Folkvord, A. and Andersson, L. 2020. Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. eLife, 9: e61076.

Healey, A.J.E., Farthing, M.W., Nunoo, F.K.E., Potts, W.M., Sauer, W.H.H., Skujina, I., King, N., Becquevort, S., Shaw, P.W., & McKeown, N.J. 2020. Genetic analysis provides insights into species distribution and population structure in East Atlantic horse mackerel (*Trachurus trachurus* and *T. capensis*). Journal of Fish Biology, 96: 795–805.

ICES. 2023a. Working Group on Southern Horse Mackerel, Anchovy and Sardine (WGHANSA). Draft Report. ICES Scientific Reports. 5:67. 396 pp.

ICES. 2023b. Working Group on Widely Distributed Stocks (WGWIDE). ICES Scientific Reports. 5:82. 980 pp.

ICES. 2023. Working Group on Mackerel and Horse Mackerel Egg Surveys (WGMEGS). ICES Scientific Reports. 5:81. 118 pp. https://doi.org/10.17895/ices.pub.23790201

ICES. 2015. Report of the Workshop on Maturity Staging of Mackerel and Horse Mackerel (WKMSMAC2), 28 September–2 October 2015, Lisbon, Portugal. ICES CM 2015/SSGIEOM:17. 93 pp.

ICES. 2010. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 28 August - 3 September 2010, Vigo, Spain. ICES CM 2010/ACOM:15: 612 pp.

ICES. 2008. Report of the Working Group on Widely Distributed Stocks (WGWIDE), 21 – 1 September 2008, ICES Headquarters Copenhagen. Diane. 67 pp.

ICES. 2007. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy (WGMHSA). 4-13 September, Copenhagen. ICES CM 2007/ACFM:31

ICES. 2006a. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy (WGMHSA). 6-15 September, Vigo, Spain. ICES CM 2006/ACFM:08

ICES. 2006b. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy (WGMHSA). 5-14 September, Galway, Ireland. ICES CM 2006/ACFM:36

ICES. 2005. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. 7-16 September, Copenhagen. ICES CM 2005/ACFM:08

ICES. 2004. Report of the Working Group on the Assessment of Mackerel, Horse Mackerel, Sardine and Anchovy. 9-18 September, Copenhagen. ICES CM 2004/ACFM:08

ICES. 1998. Report of the Working Group on the assessment of mackerel, horse mackerel, sardine and anchovy. 9-18 September 1997, ICES Headquarters, Copenhagen. ICES CM 1998/Assess:6: Part 1: 176pp.

ICES. 1995. Report of the Working Group on the assessment of mackerel, horse mackerel, sardine and anchovy. 21 June-1 July 1994, ICES Headquarters, Copenhagen. ICES CM 1995/Assess:2: Part 1: 165pp.

ICES. 1992. Report of the Study Group on Stock Identify of Mackerel and Horse Mackerel. 22-24 January, Vigo, Spain. ICES CM 1992:H:4
ICES. 1991. Report of the Working Group on the Assessment of the Stocks of Sardine, Horse Mackerel and Anchovy. 18-27 June 1991, Copenhagen. ICES CM 1990/Assess:22.
ICES. 1990. Report of the Working Group on the Assessment of the Stocks of Sardine, Horse Mackerel and Anchovy. 20-29 June 1990, ICES Headquarters, Copenhagen. ICES CM 1990/Assess:24.
ICES. 1989. Report of the Working Group on the assessment of pelagic stocks in Divisions VIIIc and IXa and horse mackerel. 10-19 May 1989, ICES Headquarters, Copenhagen. ICES CM 1989/Assess:19: 143pp.
ICES. 1988. Report of the Working Group on the assessment of pelagic stocks in Divisions VIIIc and IXa and horse mackerel. 4-11 May 1988, ICES Headquarters, Copenhagen. ICES CM 1988/Assess:22: 160pp.
ICES. 1987a. Report on the Working Group on the Appraisal of Sardine Stocks in Divisions VIIIc and IXa and horse mackerel. ICES CM 1987/Assess 23
ICES. 1987b. Report Horse Mackerel (scad) Age Determination Workshop. Lowestoft, 9-14 April 1987.
ICES. 1986a. Report on the Working Group on the Appraisal of Sardine Stocks in Divisions VIIIc and IXa. ICES CM 1986/Assess 2
ICES. 1986b. Report on the Working Group on the Appraisal of Sardine Stocks in Divisions VIIIc and IXa. ICES CM 1986/Assess 23
ICES. 1984. Extract of Report of the Advisory Committee on Fishery Management, 1984 to the North-East Atlantic Fisheries Commission. Part 1. ICES CM 1984/ ACFM:FC33
Iversen, S.A. , A. Eltink, E. Kirkegaard and D.W. Skagen 1989. The egg production and spawning stock size of the North Sea Mackerel and Horse Mackerel stocks in 1988. ICES C.M. 1989/H:16 22pp
James, G., Witten, D., Hastie, T. and Tibshirani, T. 2013. An introduction to statistical learning, volume 112. Springer, 2013.
Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24: 1403-1405.
Jombart, T., Pontier, D and Dufour, A-B. 2009. Genetic markers in the playground of multivariate analysis. Heredity, 102: 330-341.
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.
Macer, C.T. 1974. The reproductive biology of the horse mackerel *Trachurus trachurus* (L.) in the North Sea and English Channel. Journal of Fish Biology, 6: 415-438.
MacKenzie, K., Campbell, N., Mattiucci, S., Ramos, P., Pinto, A.L., Abaunza, P. 2008. Parasites as biological tags for stock identification of Atlantic horse mackerel *Trachurus trachurus* L. Fisheries Research, 89: 136-145.
Manel, S., Gaggioti, O.E., Waples, R.S. 2005. Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology and Evolution, 20: 136–142.
McMillan, L. and Fewster, R. 2017. Visualizations for genetic assignment analyses using the saddlepoint approximation method. Biometrics, 73: 1029-1041.
Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C.-C., and Lin, C.-C. 2015. E1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071). TU Wien.
Murta, A.G., Pinto, A.L., Abaunza, P. 2008a. Stock identification of horse mackerel (*Trachurus trachurus*) through the analysis of body shape. Fisheries Research, 89: 152-158.
Murta, A.G., Abaunza, P., Cardador, F., S´anchez, F. Ontogenic migrations of horse mackerel (*Trachurus trachurus*) along the Iberian coast: implications for stock identification. Fisheries Research, this issue-a.
Ndjaula, H.O.N., Hansen, T., Kruger-Johnsen, M. and Kjesbu, O. S. 2009. Oocyte development in captive Atlantic horse mackerel *Trachurus trachurus*. ICES Journal of Marine Science, 66: 623–630.

Paetkau, D., Calvert, W., Stirling, I. and Strobeck C. 1995. Microsatellite analysis of population structure in Canadian polar bears. Molecular Ecology, 4: 347-354.

Peakall, R. and Smouse P.E. 2012. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28, 2537-2539.

Piry, S., Alapetite, A., Cornuet, J-M., Paetkau, D., Baudouin, L., Estoup, A. 2004. geneclass2: a software for genetic assignment and first-generation migrant detection. Journal of Heredity, 95: 536– 539.

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

Rannala, B. and J. L. Mountain. 1997. Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. USA 94: 9197-9221.

Sala-Bozano, M., Mariani, S., Barratt, C.D., Sacchi, C., Boufana, B. and Coscia, I. 2015. Spatio-temporal variability in the population structure in North-east Atlantic stocks of horse mackerel (*Trachurus trachurus*). Biology and Environment: Proceedings of the Royal Irish Academy, 115: 211-220.

Schrider, D.R. and Kern, A.D. 2018. Supervised machine learning for population genetics: a new paradigm. Trends in Genetics, 34:301-312.

Stransky, C., Murta, A.G., Schlickeisen, J., Zimmermann, C. 2008. Otolith shape analysis as a tool for stock separation of horse mackerel (*Trachurus trachurus*) in the Northeast Atlantic and Mediterranean. Fisheries Research, 89: 159-166.

Van Damme, C.J.G., Thorsen, A., Fonn, M., Alvarez, P., Garabana, D., O'Hea, B., Perez, J.R. et al. 2014. Fecundity regulation in horse mackerel. ICES Journal of Marine Science, 71: 546–558.

Vartia, S., Collins, P.C., Cross, T.F., FitzGerald, R.D., Gauthier, D.T., McGinnity, P., Mirimin, L. & Carlsson J. 2014. Multiplexing with three-primer PCR for rapid and economical microsatellite validation. Hereditas, 151: 43–54.

Vartia, S., Villanueva, J.L., ., Finarelli, J., Farrell, E.D., Hughes, G., Carlsson, J.E.L., Collins, P.C., Gauthier, D.T., McGinnity, P., Cross, T.F., FitzGerald, R.D., Mirimin, L., Cotter, P. & Carlsson, J. 2016. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. Royal Society Open Science, 3 DOI: 10.1098/rsos.150565.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. 2012. A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28(24), 3326–3328.

9. Annex 1 – Supplementary Tables

Table S1. Baseline and potential baseline* horse mackerel samples analysed in the current study. Samples are sorted by collection date and area.

Sample	Group	Catch Location	ICES Area	Date	Lat	Lon	Collected	Submitted	Genotyped	Previous analysed
1c	1	Central North Sea	4.b	01/07/2015	53.92	2.62	10	10	9	Yes[1,2]
1d	1	Central North Sea	4.b	08/07/2015	52.82	4.02	10	11	8	Yes[1,2]
1e	1	Central North Sea	4.b	01/07/2015	54.03	2.90	6	4	3	Yes[1,2]
1f	1	Central North Sea	4.b	29/06/2015	54.00	2.77	10	10	5	Yes[1,2]
1g	1	Central North Sea	4.b	28/06/2015	53.93	2.57	7	7	4	Yes[1,2]
1a	2	Southern North Sea	4.c	06/08/2015	52.28	3.03	15	12	12	Yes[1,2]
1b	2	Southern North Sea	4.c	09/07/2015	52.85	3.99	10	8	8	Yes[1,2]
1h	2	Southern North Sea	4.c	29/06/2015	52.30	3.05	10	10	2	Yes[1,2]
1i	2	Southern North Sea	4.c	30/06/2015	53.05	2.93	10	9	4	Yes[1,2]
1j	2	Southern North Sea	4.c	05/07/2015	52.92	3.01	10	5	0	Yes[1,2]
1k	2	Southern North Sea	4.c	02/07/2015	53.97	2.82	10	9	3	Yes[1,2]
1l	2	Southern North Sea	4.c	06/07/2015	52.55	4.09	1	1	0	Yes[1,2]
2*	3	Central North Sea	4.b	08/09/2016	54.15	3.30	100	96	96	Yes[2,3]
3a	4	Central North Sea	4.b	05/07/2017	54.07	2.85	18	18	18	Yes[2]
3b	4	Central North Sea	4.b	04/07/2017	54.03	2.90	21	21	21	Yes[2]
3c	4	Central North Sea	4.b	06/07/2017	51.74	2.94	6	6	6	Yes[2]
3d	4	Central North Sea	4.b	06/07/2017	53.93	2.55	35	35	35	Yes[2]
4a	5	Southwest Ireland	7.g	18/07/2015	50.66	-8.43	50	47	44	Yes[1,2]
4b	6	Southwest Ireland	7.h	20/07/2015	49.88	-7.94	50	49	49	Yes[1,2]
5a	7	West Ireland	7.b	30/06/2016	54.42	-10.62	51	49	49	Yes[2,3]
5b	8	Southwest Ireland	7.j	04/07/2016	51.35	-10.98	49	47	46	Yes[2,3]
6a	9	Southwest Ireland	7.j	17/06/2017	50.20	-10.79	50	49	49	Yes[2,3]
6b	10	West Ireland	7.b	07/07/2017	53.93	-11.09	55	16	16	Yes[2,3]
7a	11	Southwest Ireland	7.j	05/07/2019	51.80	-11.26	11	10	10	Yes[2]
7b	11	Southwest Ireland	7.j	09/07/2019	48.35	-9.11	62	62	62	Yes[2]
7c	12	West Ireland	7.c	16/07/2019	53.22	-13.57	24	24	24	Yes[2]
8a	13	Northern Spanish Shelf	8.c	10/04/2016	43.51	-3.74	20	20	20	Yes[2,3]
8b	13	Northern Spanish Shelf	8.c	10/04/2016	43.45	-3.34	23	23	22	Yes[2,3]
8c	13	Northern Spanish Shelf	8.c	12/04/2016	43.45	-2.67	3	3	3	Yes[2,3]
8d	13	Northern Spanish Shelf	8.c	12/04/2016	43.37	-2.22	44	44	43	Yes[2,3]
8e	13	Northern Spanish Shelf	8.c	12/04/2016	43.33	-2.15	4	4	4	Yes[2,3]
9a*	14	Northern Portugal	9.a	18/03/2016	39.83	-9.20	64	62	59	Yes[2,3]
9b*	15	Southern Portugal	9.a	27/04/2016	37.26	-8.92	30	29	28	Yes[2,3]
10a*	16	Northern Portugal	9.a	26/07/2017	41.13	-9.03	50	50	50	Yes[2,3]
10b*	17	Southern Portugal	9.a	23/07/2017	36.84	-8.38	23	23	23	Yes[2,3]
10c*	18	Southern Portugal	9.a	23/07/2017	36.84	-8.10	27	27	24	Yes[2,3]
11a	19	Southern Portugal	9.a	05/02/2019	37.47	-8.98	46	46	45	Yes[2]
11b	20	Northern Portugal	9.a	24/02/2019	42.44	-9.08	13	13	13	Yes[2]
11c	21	Northern Portugal	9.a	26/02/2019	41.06	-9.13	41	37	36	Yes[2]
12a*	22	Mauritania	NA	17/06/2016	20.17	-17.52	9	9	9	Yes[2,3]
12b*	22	Mauritania	NA	10/06/2016	18.96	-17.24	9	9	9	Yes[2,3]
12c*	22	Mauritania	NA	13/06/2016	19.91	-17.61	9	9	9	Yes[2,3]
12d*	22	Mauritania	NA	29/05/2016	17.14	-16.62	3	3	3	Yes[2,3]
12e*	22	Mauritania	NA	09/06/2016	20.09	-17.71	9	9	9	Yes[2,3]
12f*	22	Mauritania	NA	10/06/2016	20.40	-17.67	9	8	8	Yes[2,3]
12g*	22	Mauritania	NA	28/05/2016	20.49	-17.50	9	9	9	Yes[2,3]
12h*	22	Mauritania	NA	13/06/2016	20.52	-17.64	9	9	9	Yes[2,3]
12i*	22	Mauritania	NA	13/06/2016	20.28	-17.73	8	8	8	Yes[2,3]
12j*	22	Mauritania	NA	12/06/2016	20.42	-17.68	10	10	10	Yes[2,3]

[1]Brunel et al., 2016; Farrell et al., 2016. [2]Farrell & Carlsson 2018. [3]Fuentes-Pardo et al., 2020; 2023.

Table S2. Potentially mixed population samples for assignment

Sample	Group	Assignment group	Catch Location	Assumed stock	ICES Area	Date	Lat	Lon	Collected	Submitted	Genotyped
3e	1	4	Southern North Sea	North Sea	4.c	17/07/2017	52.68	2.48	20	16	15
6c	2	1	West Ireland	Western	7.b	07/07/2017	53.93	-11.09	45	34	33
13a	3	3	Northern North Sea	Western	4.a	20/07/2016	59.97	0.05	20	20	20
13b	3	3	Northern North Sea	Western	4.a	20/07/2016	60.27	0.70	12	12	12
13c	3	3	Northern North Sea	Western	4.a	20/07/2016	60.02	0.05	64	64	64
14a	4	3	Northern North Sea	Western	4.a	01/07/2017	57.93	1.78	47	47	47
14b	5	3	Northern North Sea	Western	4.a	20/07/2017	60.32	0.28	53	49	49
15	6	3	Northern North Sea	Western	4.a	28/07/2017	59.02	-0.11	73	48	48
16	7	3	Northern North Sea	Western	4.a	09/08/2017	59.50	0.83	66	48	47
17	8	2	Norwegian coast	North Sea	4.a	08/03/2019	59.75	5.50	30	24	24
18	9	2	Norwegian coast	North Sea	4.a	10/04/2019	59.92	5.08	30	24	23
19	10	2	Norwegian coast	North Sea	4.a	23/05/2019	60.50	5.25	30	24	23
20	11	3	Norwegian coast	Western	4.a	11/09/2019	58.25	4.75	28	24	3
21a	12	3	Northern North Sea	Western	4.a	18/07/2020	60.43	-0.13	1	1	1
21b	12	3	Northern North Sea	Western	4.a	20/07/2020	60.70	0.87	1	1	1
21c	12	3	Northern North Sea	Western	4.a	21/07/2020	60.75	0.63	1	1	1
21d	12	3	Northern North Sea	Western	4.a	21/07/2020	60.78	1.03	1	1	1
21e	12	3	Northern North Sea	Western	4.a	21/07/2020	60.88	1.25	3	3	3
21f	12	3	Northern North Sea	Western	4.a	22/07/2020	60.82	0.93	2	2	2
21g	12	3	Northern North Sea	Western	4.a	22/07/2020	60.80	0.97	2	2	2
21h	12	3	Northern North Sea	Western	4.a	23/07/2020	60.77	0.63	10	10	9
21i	12	3	Northern North Sea	Western	4.a	23/07/2020	60.57	0.32	2	2	2
21j	12	3	Northern North Sea	Western	4.a	24/07/2020	60.63	0.38	2	2	2
21k	12	3	Northern North Sea	Western	4.a	24/07/2020	60.67	0.47	6	6	6
21l	12	3	Northern North Sea	Western	4.a	24/07/2020	60.48	0.33	6	6	6
21m	12	3	Northern North Sea	Western	4.a	24/07/2020	60.52	0.43	6	6	6
21n	12	3	Northern North Sea	Western	4.a	25/07/2020	60.38	0.97	7	5	5
22a	13	3	Northern North Sea	Western	4.a	17/08/2020	59.97	0.18	11	11	11
22b	13	3	Northern North Sea	Western	4.a	18/08/2020	58.58	0.35	5	5	4
22c	13	3	Northern North Sea	Western	4.a	18/08/2020	58.60	0.37	4	4	4
22d	13	3	Northern North Sea	Western	4.a	18/08/2020	58.75	0.35	5	5	5
22e	13	3	Northern North Sea	Western	4.a	19/08/2020	58.38	0.48	5	5	5
22f	13	3	Northern North Sea	Western	4.a	20/08/2020	58.87	-0.83	7	7	6
22g	13	3	Northern North Sea	Western	4.a	21/08/2020	58.68	-0.70	5	5	5
22h	13	3	Northern North Sea	Western	4.a	21/08/2020	58.52	-0.75	2	2	2
22i	13	3	Northern North Sea	Western	4.a	22/08/2020	58.35	-0.65	3	3	3
22j	13	3	Northern North Sea	Western	4.a	22/08/2020	58.42	-0.62	8	1	1
23	14	3	Norwegian coast	Western	4.a	05/11/2021	59.17	5.50	30	24	13
24	15	3	Norwegian coast	Western	4.a	16/11/2021	59.42	5.88	30	24	7
25	16	2	Norwegian coast	North Sea	4.a	27/04/2022	59.25	5.00	30	24	21
26	17	2	Norwegian coast	North Sea	4.a	02/05/2022	59.17	6.00	30	24	23
27	18	2	Norwegian coast	North Sea	4.a	24/05/2022	59.17	6.00	30	24	23
28	19	3	Norwegian coast	Western	4.a	07/09/2022	60.00	2.25	30	24	12
29	20	3	Skagerrak	Western	3.a	30/08/2017	57.45	8.13	11	11	11
30a	21	4	Central North Sea	North Sea	4.b	11/07/2017	54.15	3.22	12	12	12
30d	21	4	Central North Sea	North Sea	4.b	18/07/2017	54.13	3.15	16	11	11
30b	22	4	Southern North Sea	North Sea	4.c	12/07/2017	52.88	3.02	11	11	14
30c	22	4	Southern North Sea	North Sea	4.c	13/07/2017	52.70	2.98	14	14	11
31	23	5	Eastern Channel	North Sea	7.d	21/09/2020	50.48	-0.57	100	96	96
32	24	5	Eastern Channel	North Sea	7.d	13/11/2020	50.82	1.12	100	96	96
33	25	5	Eastern Channel	North Sea	7.d	07/12/2020	50.50	-0.22	100	96	94
34	26	6	Western Channel	Western	7.e	25/05/2017	49.50	-3.00	100	96	95
35	27	6	Western Channel	Western	7.e	25/05/2017	49.50	-3.50	100	96	88

Table S3. The length frequency of baseline and potential baseline horse mackerel samples included in the current study.

	Baseline and Potential Baseline Sample Group																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
13																						
13.5														1								
14																						
14.5															2							
15														1	3							
15.5														1	3							
16														1	7							
16.5														1	4							
17														3								
17.5														6	3							
18														4								
18.5														14								
19		1												8				1				
19.5			1											4			1					
20		4	7											5								
20.5	1		17											5	1	3			1			
21			21	2									1	3		2	2	2	8			
21.5			32	14							1			3	2	8	4	4	13			2
22	1		9	8						1			4	1		5	4	1	11			
22.5		1	5	16						1			1		2	5	3	5	6			5
23		4		12			1						4		1	6	3	6	2			3
23.5			2	9									1			4	1	4	2			12
24	2	3		4						1	1		8		1	5	4	2	3			7
24.5	4	5	2	1									9			5	1	1				9
25	4	7		2	1					2	1		13			5				2	2	6
25.5	5	5		4	1						3		11			2				3	3	5
26	2	5		1	1						1		8					1		2	3	7
26.5	3	2		3			1			1	6		3							2	2	3
27	2	4						1			9		8	1							12	1
27.5	2			1			1			2	12		4							4	6	2
28	1	1						1		2	7		5								7	1
28.5		4		1						1	1		1								1	1
29	3	1					1				7											4
29.5	2	1				1		1	2		2		3								1	3
30	1	1		1	4	2		1	3		1											2
30.5	2				3	3	1		1		2		1									3
31		1		1	3	3		2	7		1											
31.5	1				3	2		1	4	1	1											1
32	2	2			4	3	1	4	11	1		1										1
32.5					1	3	2	2	6		1											1
33		1			5	3	3	4	5	1												1
33.5					5	11	3	7	6	1	1											
34	1				5	3	4	6	1		2	6										
34.5	3				2	6	2	5	1			1										
35					6	2	9	7	1		3	5										
35.5							8	2	1	1		4										
36		1			3	2	4	2				3										
36.5						2	1					2										
37						2	1	1				2										1
37.5																						
38						1																
38.5							2															
39							2															1
39.5																						
40																						
40.5							2															1
41																						
41.5																						
42																						

Table S4. The maturity stages (6-pt scale ICES, 2015) of baseline and potential baseline horse mackerel samples included in the current study.

	Maturity stage (6-pt)					
	1	**2**	**3**	**4**	**5**	**6**
1			42			
2			54			
3		89		7		
4			70	10		
5			47			
6			49			
7		4	44	1		
8		1	46			
9			47	2		
10			16			
11		2	70			
12		1	23			
13		29	64	1		
14		62				
15	21	5	3			
16		49	1			
17		18	2	3		
18		21	6			
19			46			
20			13			
21			37			
22		17	2	64		

Table S5. The length frequency of mixed horse mackerel samples by group included in the current study.

	Potential Mixed Samples																										
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
10																											
10.5																											
11																											
11.5																											
12																											
12.5																											4
13																											1
13.5																											
14																											
14.5																											
15																											
15.5																											1
16																											5
16.5																											6
17																										1	1
17.5		1																								3	15
18																										9	13
18.5																								1		3	13
19																							1	2	1	14	14
19.5		2																				1			4	15	4
20	1	2																			1	1	2	6	7	9	6
20.5	3	3																			4	2	8	13	14	21	2
21	2	4																			2	2	15	24	13	15	1
21.5	4	6																			6	2	19	15	11	3	
22	4	7																			5	4	19	15	8	1	
22.5	1	2																			3	5	9	12	8	1	
23	1	5																			1	4	1	5	3		
23.5		1																				1	1	2	7		1
24		1																				1			5		
24.5																						1	1	1		1	
25																							1		2		
25.5																									3		
26																							1		2		
26.5																									4		
27																									1		
27.5																					1	1			1		
28																											
28.5																											
29													2												1		
29.5							1																		1		
30						2	3					2	3	1					1								
30.5							2							2													
31			2	1		2	4						1						2								
31.5				1		1	2						2	3					1								
32			3			4	4				1	2	2						3								
32.5			3	2		6	5						5	1				1									
33			3	1	3	2	6				3			2	2			1	6								
33.5			7	1	2	5	7		1		1	2		1	1				1								
34			11	2	2	9	3		3		2	1	6		1		1		4								
34.5			6	5	1	1	3		1	1	2	1	2	2					1	1							
35			7	3	6	4	3		1	3	3	3	9	1	2			1	4								
35.5			5	3	4	2	2	2	4	1	7	5	2	1	4			2	1	3							
36			8	5	4	4	1		3	3	2	5	6	3	3		2			1							
36.5			7	1	5	2		2	4	2	2		1	1	3		2	1		2							
37			5	2	6		2	2	4	2		7	3			3	2	5		2							
37.5			6	3	3	1		2		3	1	6	1	3	2		2	4									
38			5	4	1	1		1	1	2		1		1	2	3	1										
38.5			3	2	3			4	1	2		4			1	2	5	4		1							
39			5	5	3			2				3	1		2	3	1										
39.5			3	1		1				2		1				2	2										
40				1		1		2	1			4	1		1	4	4	1		1							
40.5			2					1		1		1						1									
41			2	2	1			3								5		1									
41.5			2	1	1			2		1						1		1									
42			1	1	4					1				1													
42.5																1	2	1									
43														1													
43.5																											
44																											
44.5																											
45								1																			

Table S6. The maturity stages (6-pt scale ICES, 2015) of mixed horse mackerel samples included in the current study.

	Maturity stage (6-pt)						
	1	2	3	4	5	6	NA
3e				16			
6c	34						
3		1	52	43			
4				47			
5				49			
6				48			
7				48			
8		19		5			
9		24					
10		24					
11				24			
12							48
13							48
14		4		20			
15		12		11		1	
16		22		2			
17		23		1			
18		24					
19				24			
20				11			
21				23			
22				25			
23		96					
24		94		2			
25		93	3				
26	94	2					
27	96						

Table S7. Quality call data of the complete raw dataset prior to filtering.

	Quality Call					
Chr	**CallRateBelowThreshold**	**MonoHighResolution**	**NoMinorHom**	**OTV**	**PolyHighResolution**	**Total**
1	2	2	12	10	183	209
2		1	3	4	24	32
3	1	3	8	6	140	158
4	2	2	3	6	146	159
5	2	1	10	3	124	140
6	1		7	4	107	119
7	1	2	6	15	125	149
8	2		7	8	68	85
9	1		1	13	115	130
10	1		10	10	131	152
11	1	3	13	10	104	131
12	2	1	5	8	76	92
13		1	4	11	127	143
14	3		4	11	98	116
15	3	2	12	7	110	134
16	2		7	15	120	144
17	2	3	14	9	181	209
18	1	3	18	16	32	70
19	2	1	4	6	40	53
20	1	2	6	6	83	98
21	1	2	11	13	178	205
22		2	16	7	134	159
23			16	3	39	58
24		1	10	6	61	78
988					2	2
9116					1	1
9125					1	1
9146			1			1
9156			1			1
9157	1					1
9249				1		1
Total	32	32	209	208	2550	3031

Table S8. F_{ST} of the baseline and potential baseline samples by sample group.

Sample Group		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
NS_15a	1	0.000																					
NS_15b	2	0.011	0.000																				
NS_16	3	0.008	0.007	0.000																			
NS_17	4	0.008	0.007	0.003	0.000																		
WS_15a	5	0.020	0.022	0.016	0.019	0.000																	
WS_15b	6	0.022	0.025	0.019	0.022	0.007	0.000																
WS_16a	7	0.021	0.024	0.018	0.021	0.006	0.006	0.000															
WS_16b	8	0.022	0.025	0.019	0.021	0.007	0.006	0.006	0.000														
WS_17a	9	0.021	0.024	0.017	0.019	0.006	0.006	0.006	0.006	0.000													
WS_17b	10	0.030	0.033	0.027	0.029	0.013	0.012	0.012	0.013	0.012	0.000												
WS_19a	11	0.020	0.023	0.017	0.019	0.005	0.005	0.005	0.006	0.005	0.011	0.000											
WS_19b	12	0.023	0.026	0.020	0.022	0.009	0.009	0.009	0.009	0.009	0.016	0.008	0.000										
NSS_16	13	0.021	0.023	0.017	0.019	0.006	0.005	0.006	0.006	0.005	0.011	0.004	0.009	0.000									
NPT_16	14	0.020	0.023	0.017	0.019	0.006	0.006	0.006	0.006	0.005	0.011	0.004	0.009	0.003	0.000								
NPT_17a	16	0.021	0.024	0.018	0.020	0.007	0.006	0.007	0.007	0.006	0.011	0.005	0.010	0.004	0.005	0.000							
NPT_19a	20	0.028	0.031	0.026	0.028	0.013	0.013	0.014	0.014	0.013	0.019	0.012	0.016	0.012	0.012	0.013	0.000						
NPT_19b	21	0.020	0.023	0.017	0.019	0.007	0.007	0.007	0.007	0.007	0.014	0.006	0.009	0.007	0.006	0.007	0.014	0.000					
SPT_16	15	0.029	0.033	0.027	0.029	0.015	0.013	0.015	0.015	0.013	0.018	0.011	0.018	0.009	0.010	0.011	0.019	0.015	0.000				
SPT_17a	17	0.030	0.033	0.027	0.029	0.016	0.015	0.017	0.017	0.015	0.020	0.013	0.020	0.011	0.012	0.013	0.021	0.017	0.010	0.000			
SPT_17b	18	0.029	0.032	0.026	0.029	0.014	0.012	0.014	0.013	0.013	0.017	0.011	0.016	0.009	0.010	0.011	0.019	0.014	0.010	0.011	0.000		
SPT_19	19	0.022	0.025	0.019	0.021	0.007	0.007	0.008	0.008	0.007	0.013	0.005	0.010	0.005	0.006	0.007	0.013	0.008	0.010	0.011	0.010	0.000	
NAF_16	22	0.042	0.046	0.040	0.042	0.028	0.026	0.028	0.028	0.027	0.031	0.024	0.030	0.021	0.024	0.025	0.031	0.028	0.014	0.015	0.019	0.020	0.000

Table S9. The SNP panels derived from the two approaches used to identify informative markers for the North Sea versus Western baseline dataset.

Approach 1		Approach	
Panel_1	**Panel_2**	**Panel_3**	**Panel_4**
25_SNP	**8_SNP**	**36_SNP**	**17_SNP**
1_17506941	1_17539260	1_17525646	1_17525646
1_17522148	1_22047461	1_17506142	1_21711760
1_17533108	5_28240487	1_17507305	3_14013644
1_17539260	20_11640406	1_21711760	4_13022828
1_17539382	21_20839487	1_26047188	6_2336270
1_22047461	21_22426719	3_14013644	7_12804862
1_22052339	21_23404685	3_28964757	8_24769297
1_22055361	21_34573992	3_14363455	9_35690385
1_22055419		4_13022828	11_6926377
5_28240487		4_13092907	12_5840756
5_28242772		4_13088818	15_19753638
20_11636865		6_2336270	16_19215181
20_11637433		7_12804862	17_21983581
20_11640406		7_5189950	20_11637400
20_11643211		7_7370969	21_34602916
20_11648344		8_24769297	23_11989041
21_20839455		9_35690385	24_11769067
21_20839487		9_18815195	
21_22426719		11_6926377	
21_23269355		11_6917352	
21_23404685		11_7074782	
21_34573582		12_5840756	
21_34573992		15_19753638	
21_34574438		15_25521754	
21_34590241		15_12185216	
		16_19215181	
		17_21983581	
		17_31117042	
		20_11637400	
		20_14528596	
		21_34602916	
		21_34571361	
		21_34574070	
		23_11989041	
		24_11769067	
		24_17686352	

Table S10. The tuning parameters for the assignment models in *assignPOP*.

	Complete	**Approach 1**		**Approach 2**	
	2421_SNP	**25_SNP**	**8_SNP**	**36_SNP**	**17_SNP**
# individuals (NS/WS)	109/383	113/399	116/421	109/382	109/383
PCs - highest mean success	200	16	2	5	4
Iterations	100	100	100	100	100
Cost	1	1	1	1	1
Gama	0.004065	0.5	1	0.5	1

Table S11. Assignment matrix for the Monte-Carlo cross-validation of the North Sea (NS) versus Western (WS) baseline dataset with the five different marker panels with random loci selection. SD = standard deviation.

Panel	%loci	Monte-Carlo Cross Validation			
			NS	WS	SD
2421_SNP	100	NS	0.92	0.08	± 0.04
		WS	0.01	0.99	± 0.00
	75	NS	0.92	0.08	± 0.04
		WS	0.01	0.99	± 0.01
	50	NS	0.93	0.07	± 0.04
		WS	0.02	0.98	± 0.01
	25	NS	0.92	0.08	± 0.05
		WS	0.03	0.97	± 0.02
25_SNP	100	NS	0.91	0.09	± 0.04
		WS	0.07	0.93	± 0.02
	75	NS	0.90	0.10	± 0.05
		WS	0.08	0.92	± 0.03
	50	NS	0.86	0.14	± 0.15
		WS	0.10	0.90	± 0.05
	25	NS	0.65	0.35	± 0.18
		WS	0.32	0.68	± 0.26
8_SNP	100	NS	0.32	0.68	± 0.28
		WS	0.38	0.62	± 0.30
	75	NS	0.54	0.46	± 0.25
		WS	0.36	0.64	± 0.21
	50	NS	0.63	0.37	± 0.23
		WS	0.33	0.67	± 0.21
	25	NS	0.68	0.32	± 0.30
		WS	0.30	0.70	± 0.28
36_SNP	100	NS	0.93	0.07	± 0.04
		WS	0.03	0.97	± 0.01
	75	NS	0.90	0.10	± 0.07
		WS	0.07	0.93	± 0.10
	50	NS	0.73	0.27	± 0.16
		WS	0.26	0.74	± 0.18
	25	NS	0.58	0.42	± 0.22
		WS	0.37	0.63	± 0.21
17_SNP	100	NS	0.72	0.28	± 0.08
		WS	0.29	0.71	± 0.03
	75	NS	0.70	0.30	± 0.09
		WS	0.31	0.69	± 0.10
	50	NS	0.63	0.37	± 0.19
		WS	0.36	0.64	± 0.21
	25	NS	0.63	0.37	± 0.20
		WS	0.42	0.58	± 0.25

10. Annex 2 – Supplementary Figures

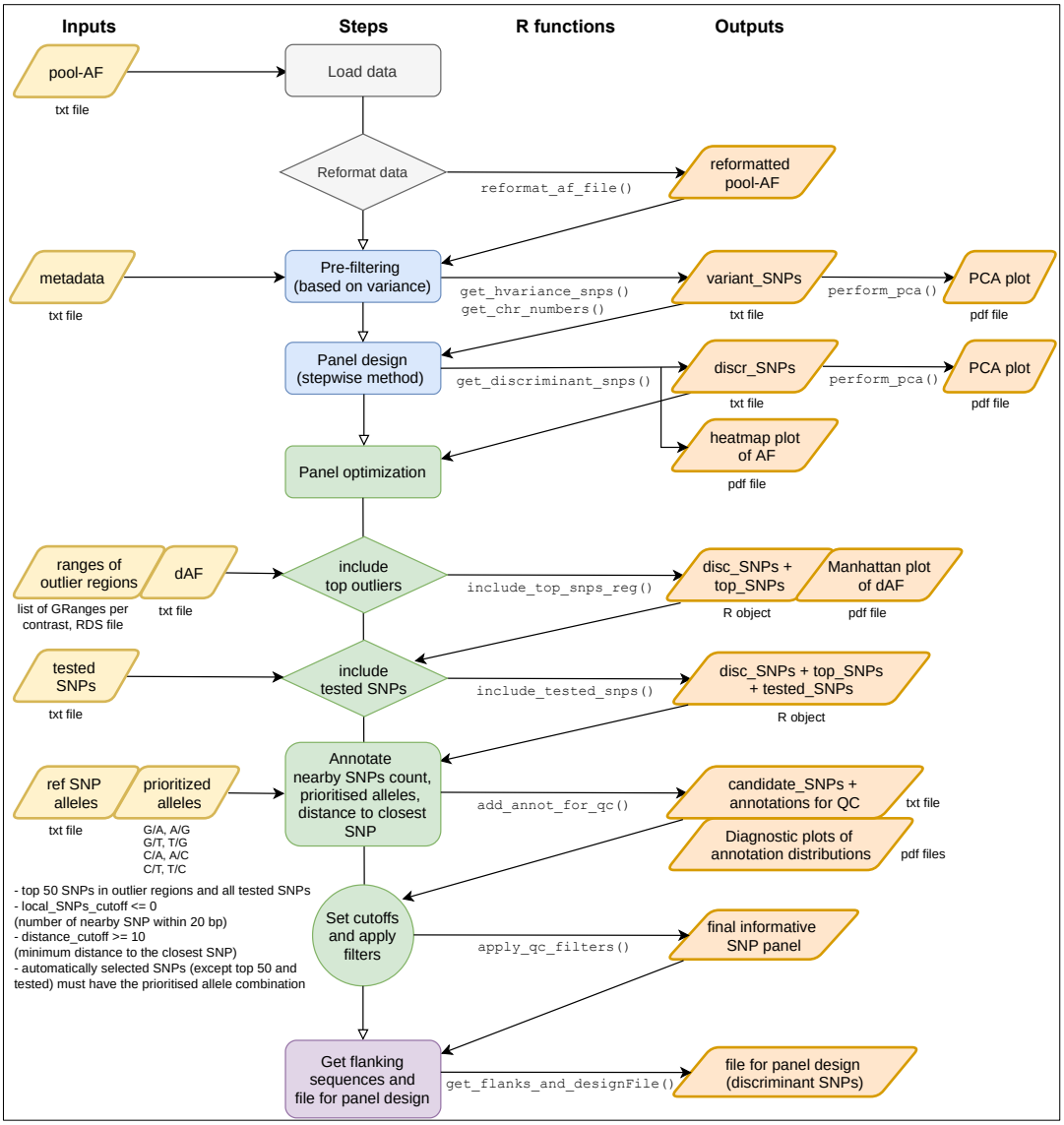


Figure S1. Workflow to develop a SNP chip for the Atlantic horse mackerel.

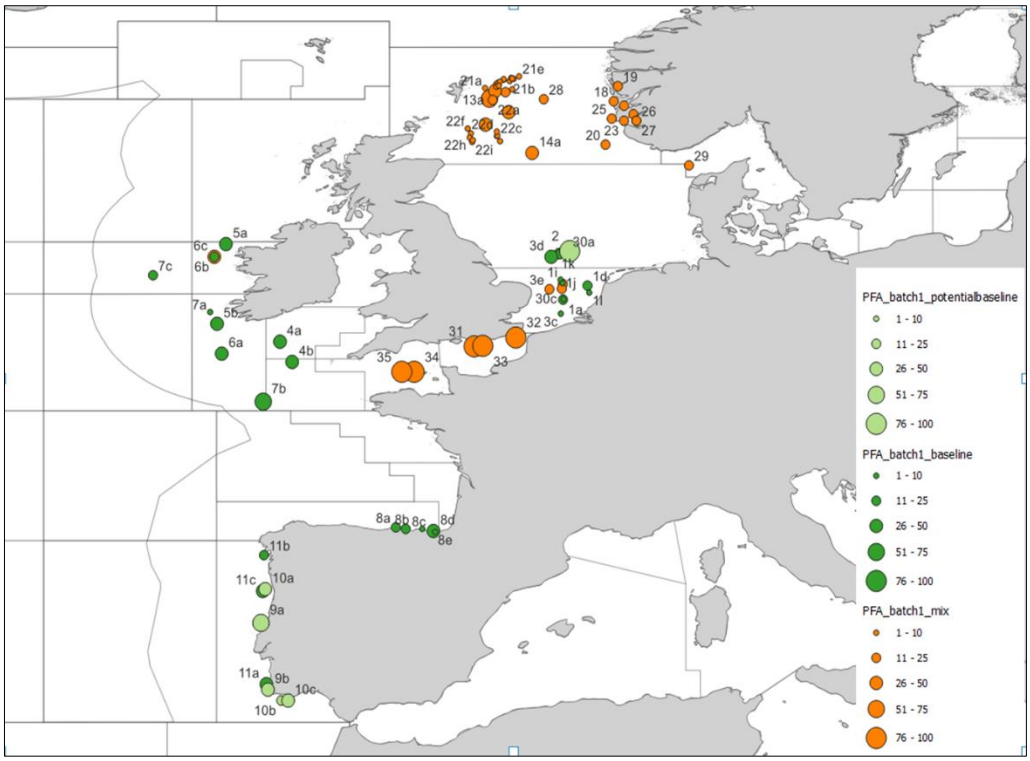


Figure S2. Overview of sampling locations and samples in the northeast Atlantic area included in the current study. Dark green represents spawning baseline samples, Light green represents potential baseline samples and orange represents mixed samples for assignment.

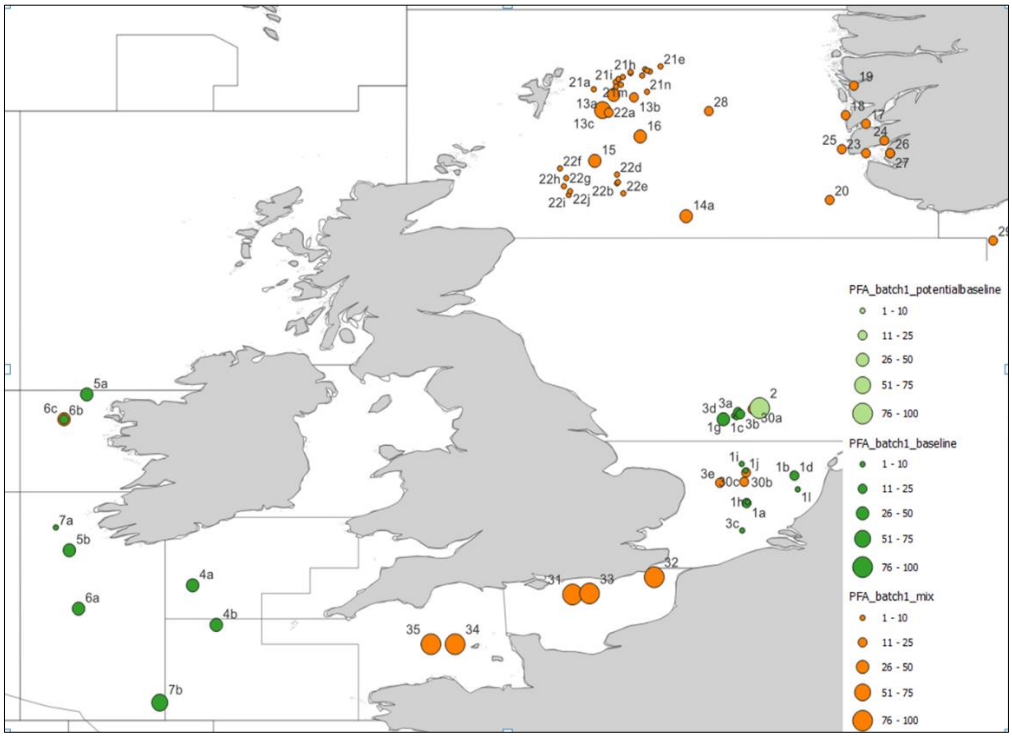


Figure S3. Overview of sampling locations and samples in northwestern waters and the North Sea included in the current study. Dark green represents spawning baseline samples, Light green represents potential baseline samples and orange represents mixed samples for assignment.

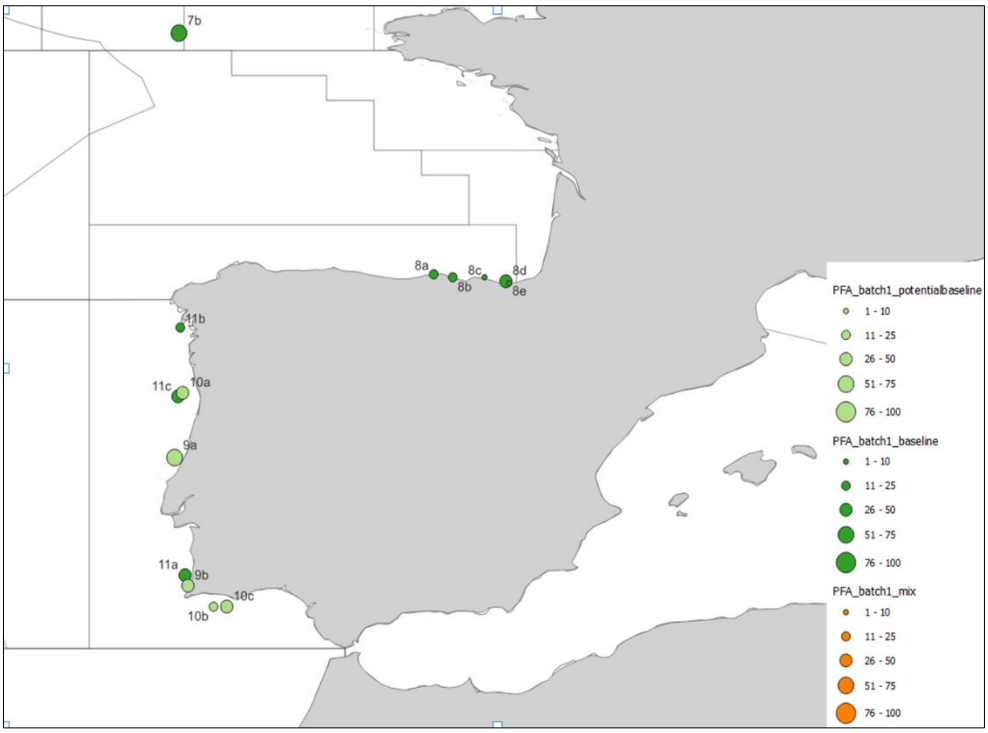


Figure S4. Overview of sampling locations and samples in Spanish and Portuguese waters included in the current study. Dark green represents spawning baseline samples, Light green represents potential baseline samples and orange represents mixed samples for assignment.

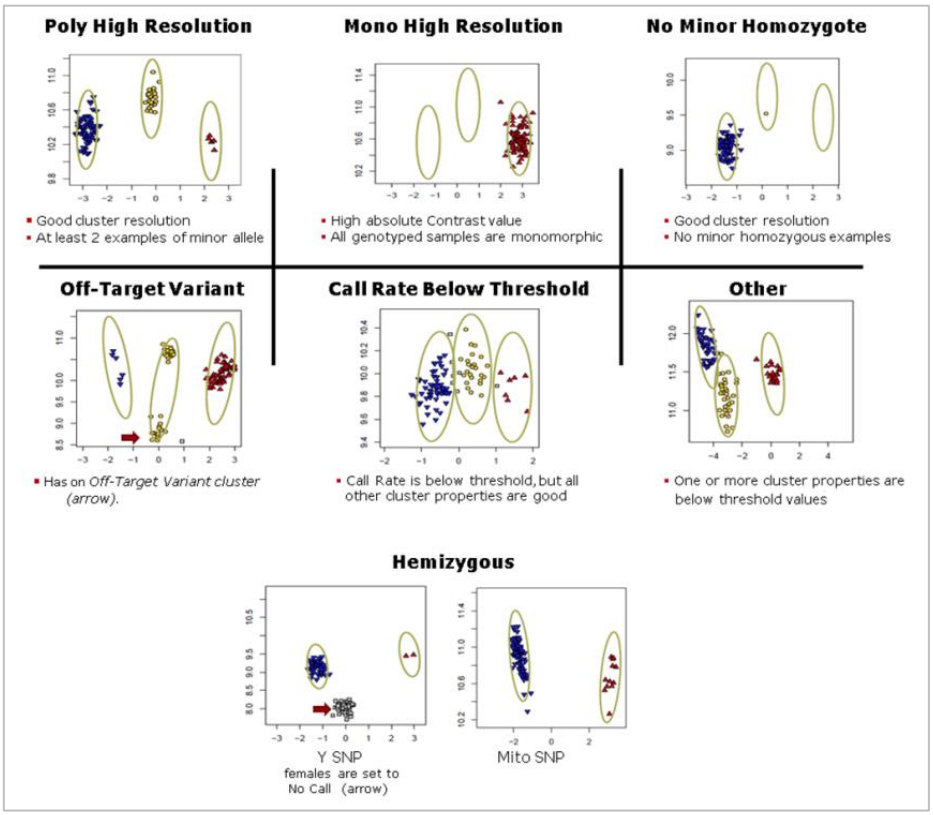


Figure S5. Cluster Plot examples and descriptions of Axiom Array SNP classification categories.

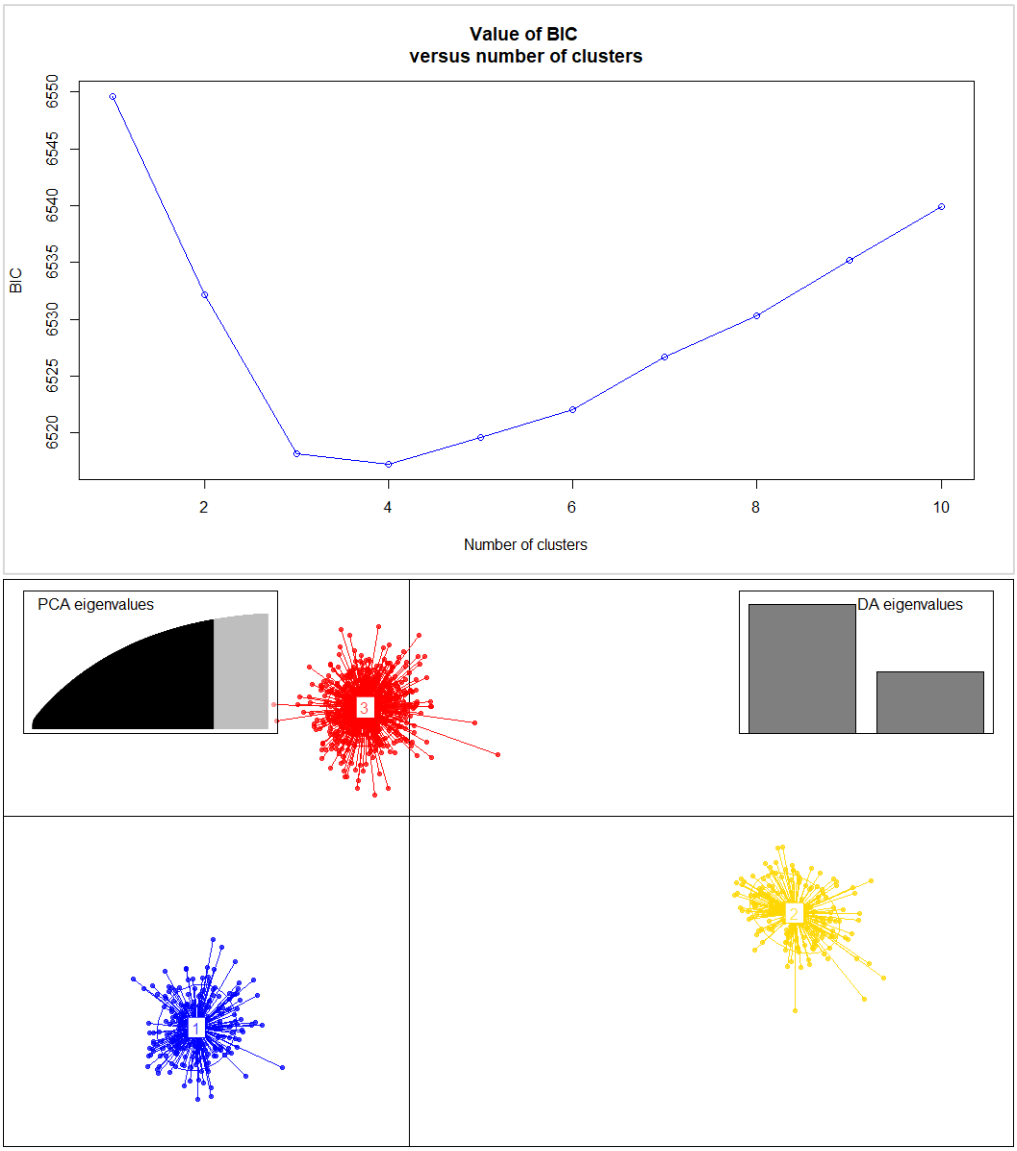


Figure S6. Outputs of the *find.clusters* analyses of the baseline and potential baseline sample dataset.

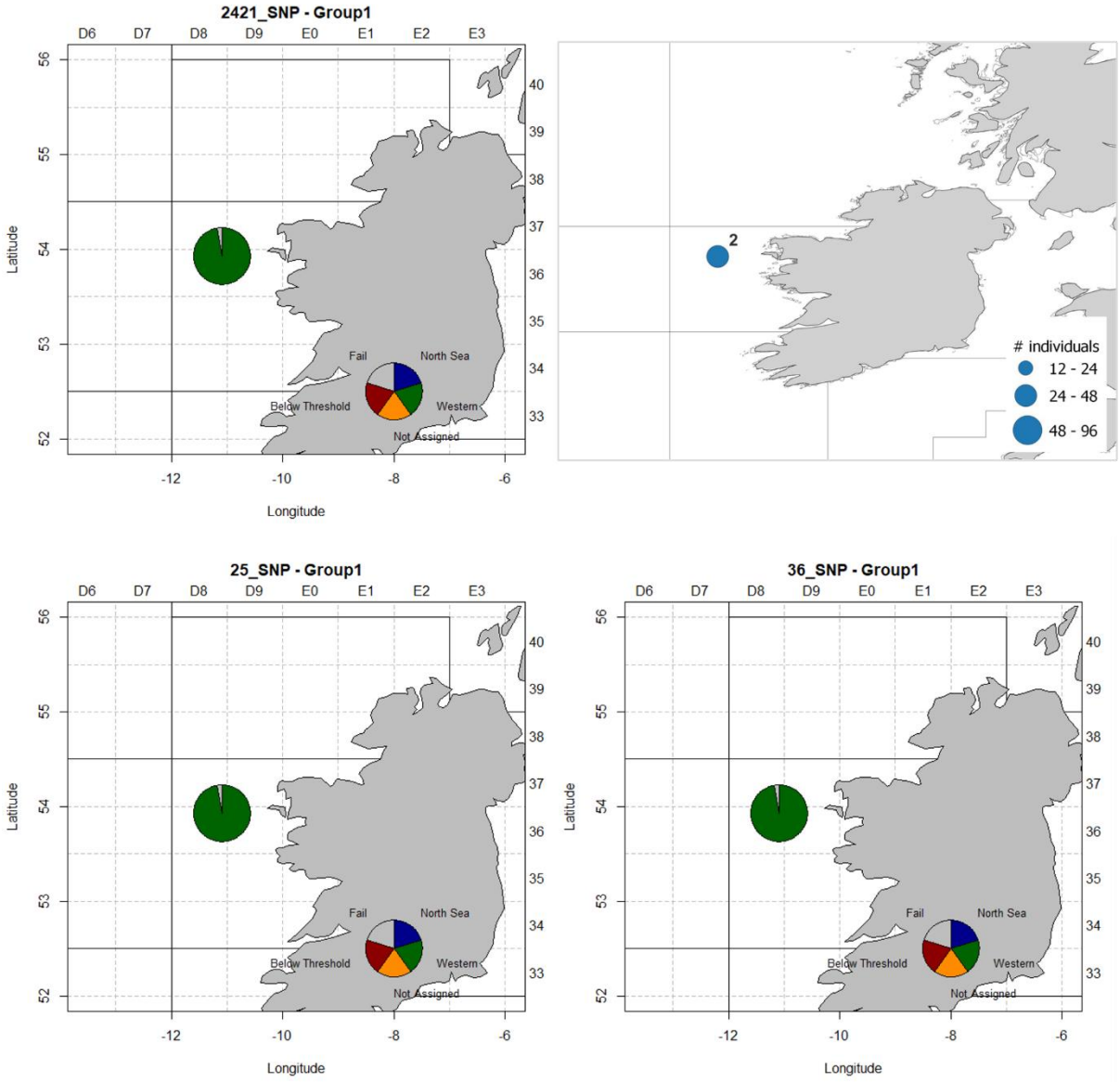


Figure S7. The outputs of the Group 1 assignment of the mixed samples with the three marker panels.

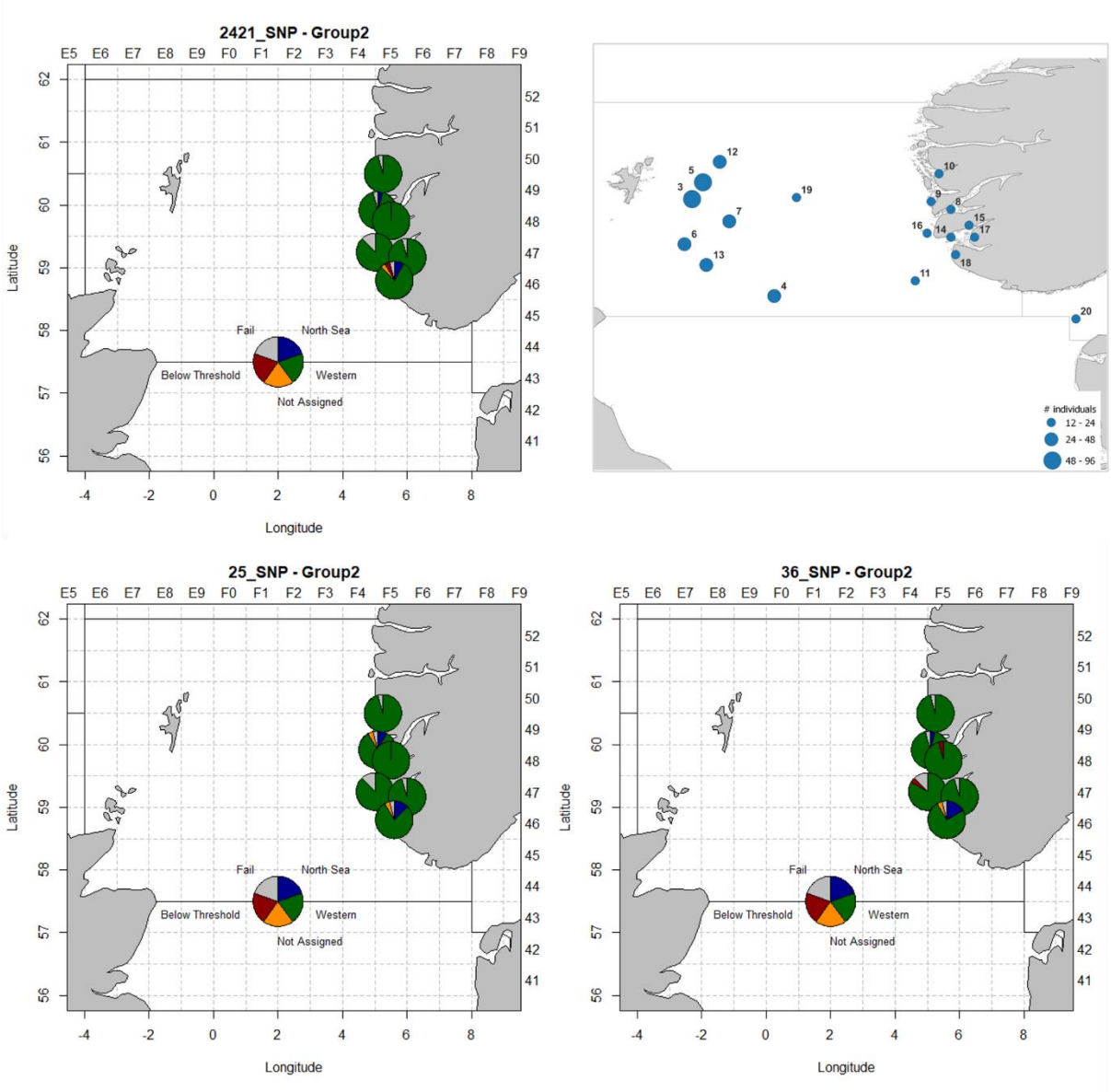


Figure S8. The outputs of the Group 2 assignment of the mixed samples with the three marker panels.

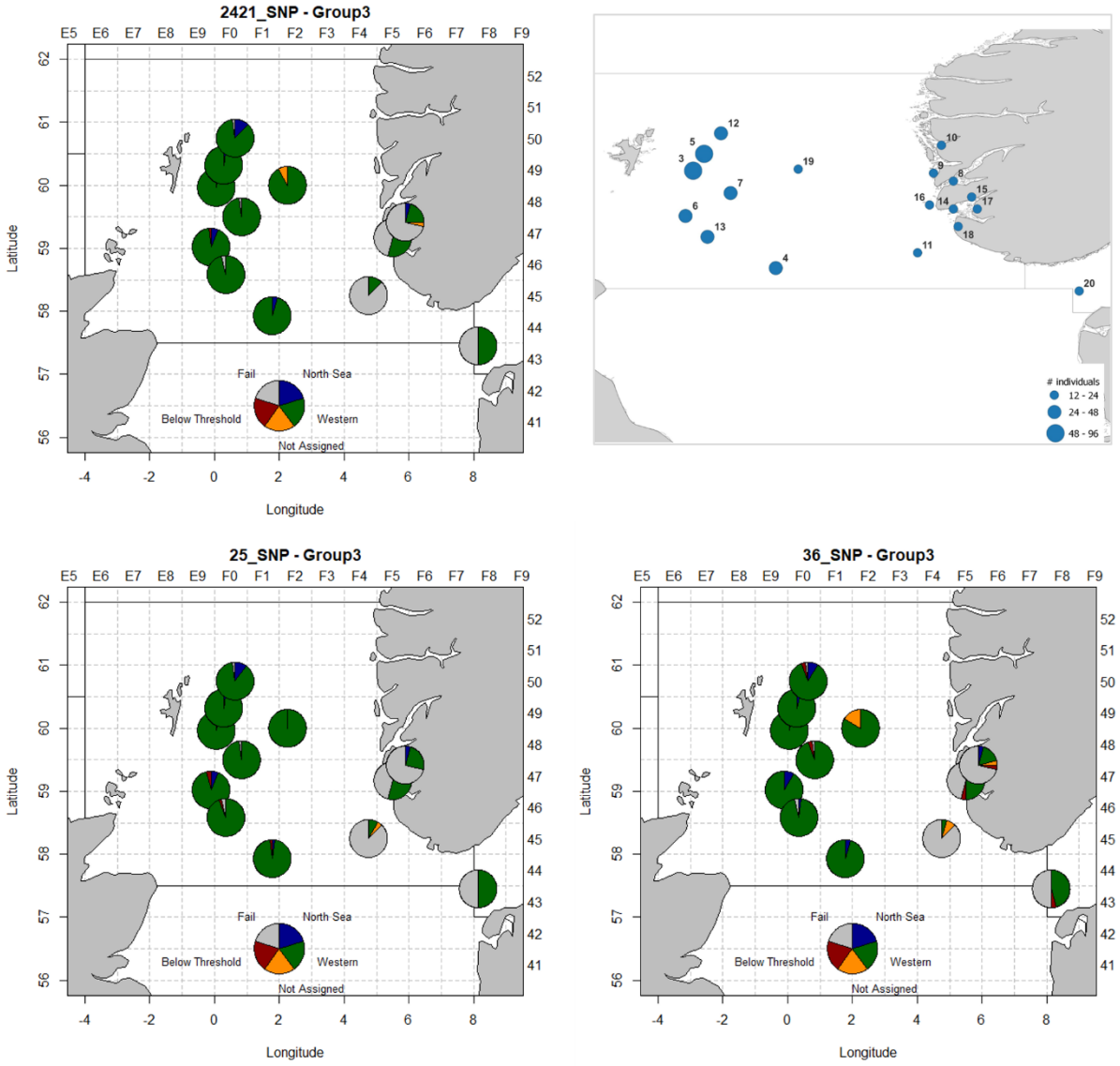


Figure S9. The outputs of the Group 3 assignment of the mixed samples with the three marker panels.

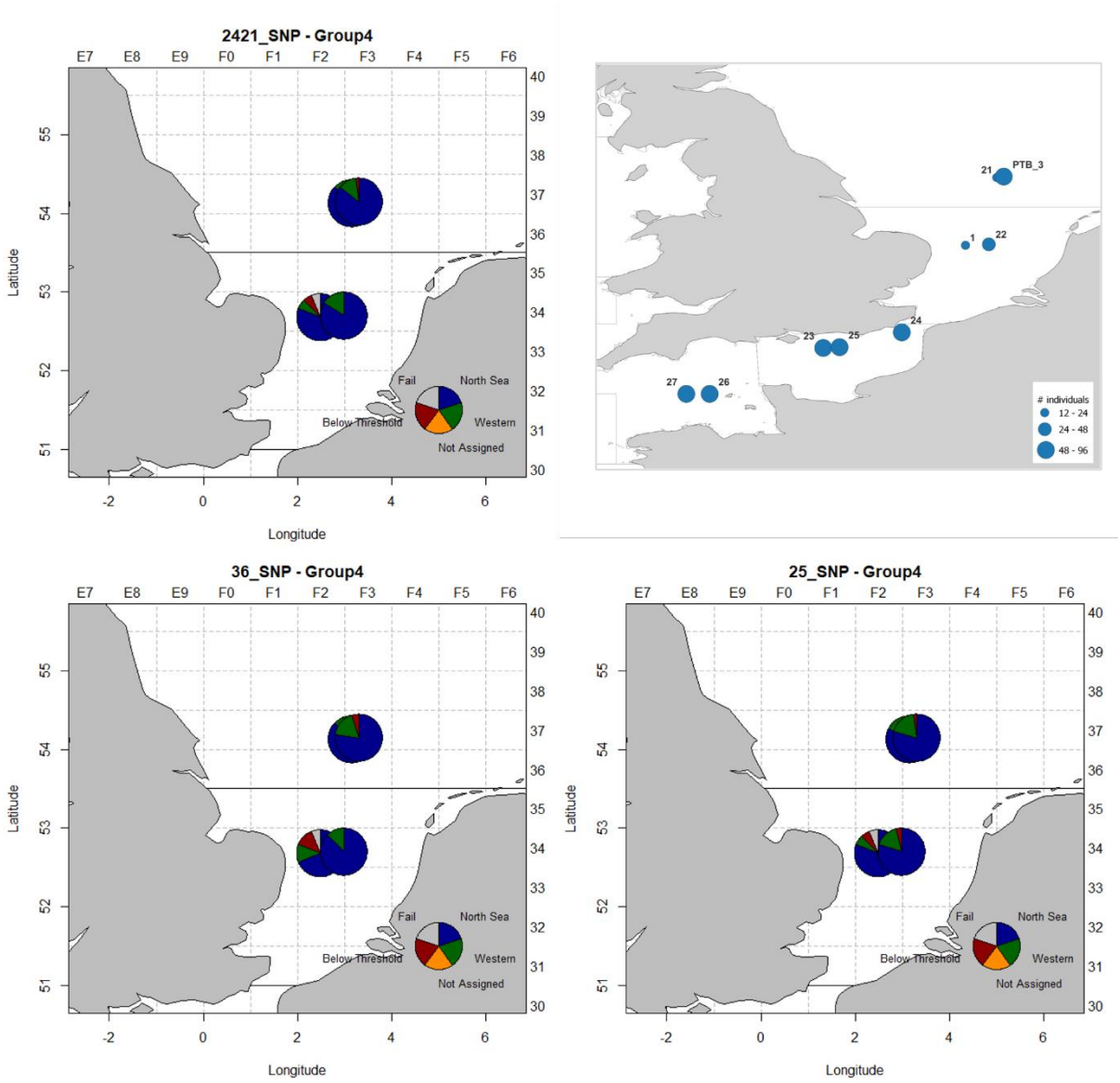


Figure S10. The outputs of the Group 4 assignment of the mixed samples with the three marker panels.

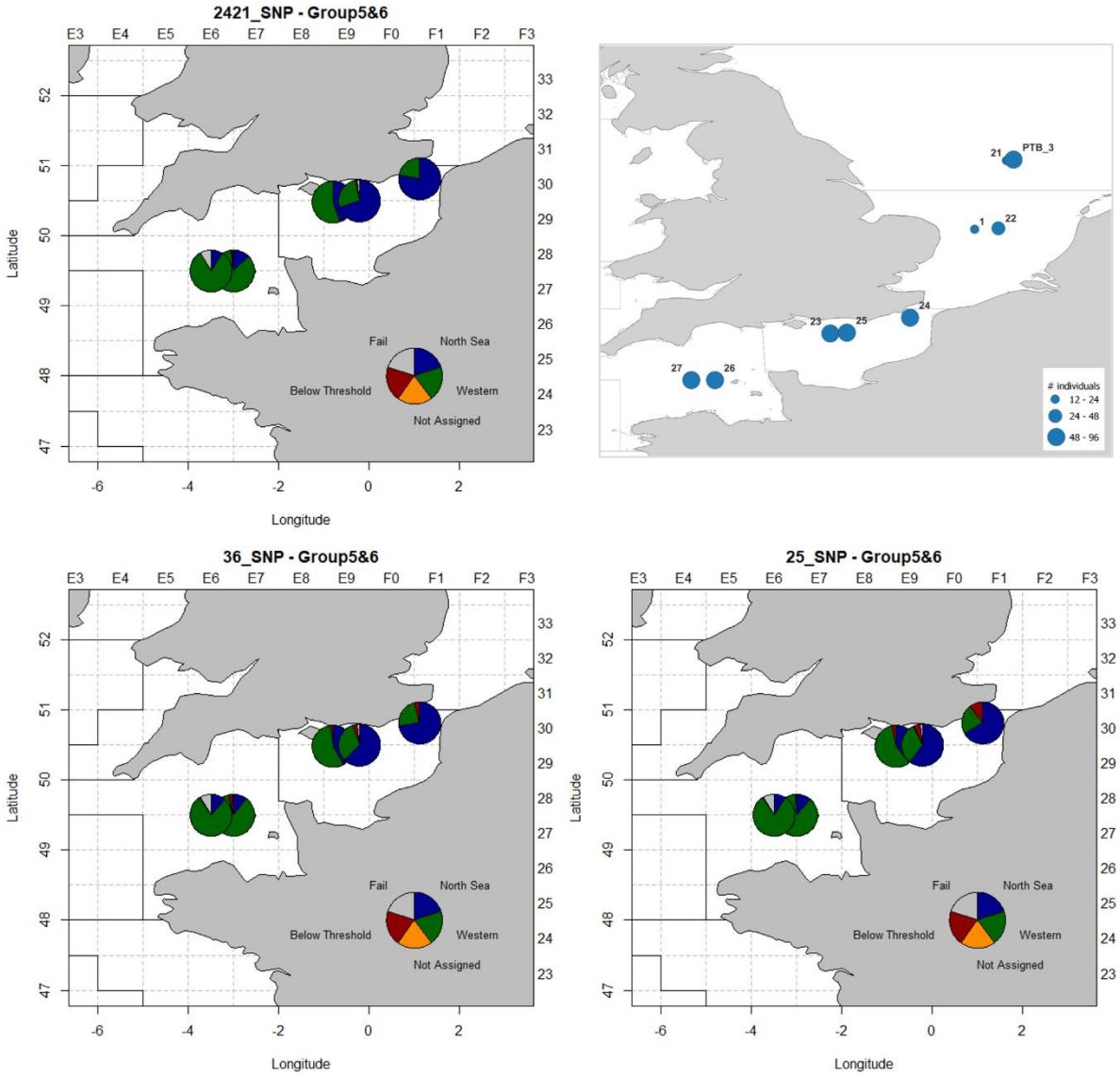


Figure S11. The outputs of the Group 5 & 6 assignment of the mixed samples with the three marker panels.

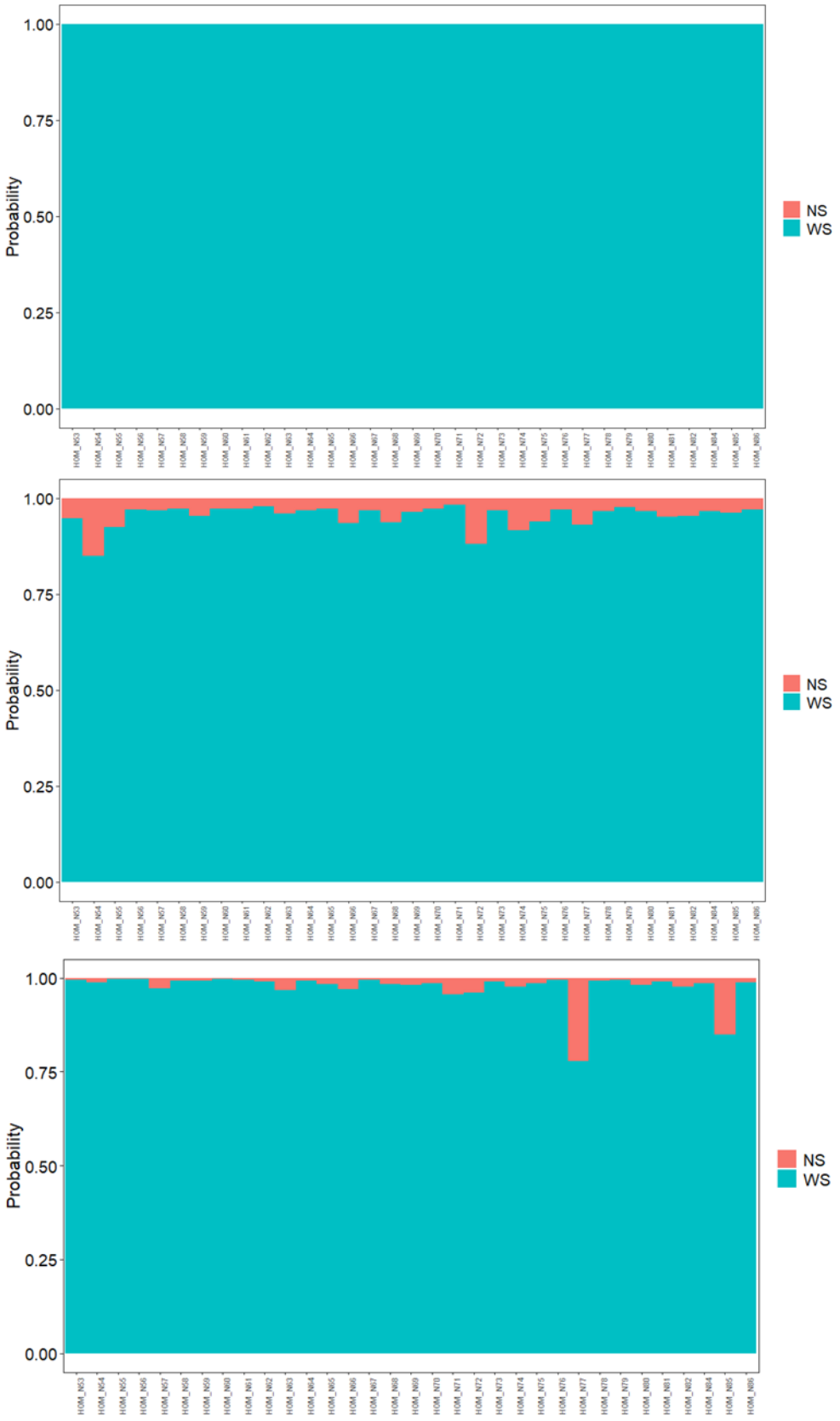


Figure S12. The outputs of the Group 1 assignment with the three marker panels (top) *2421_SNP* (middle) *25_SNP* (bottom) *36_SNP*. Each column represents an individual with the associated assignment probability.

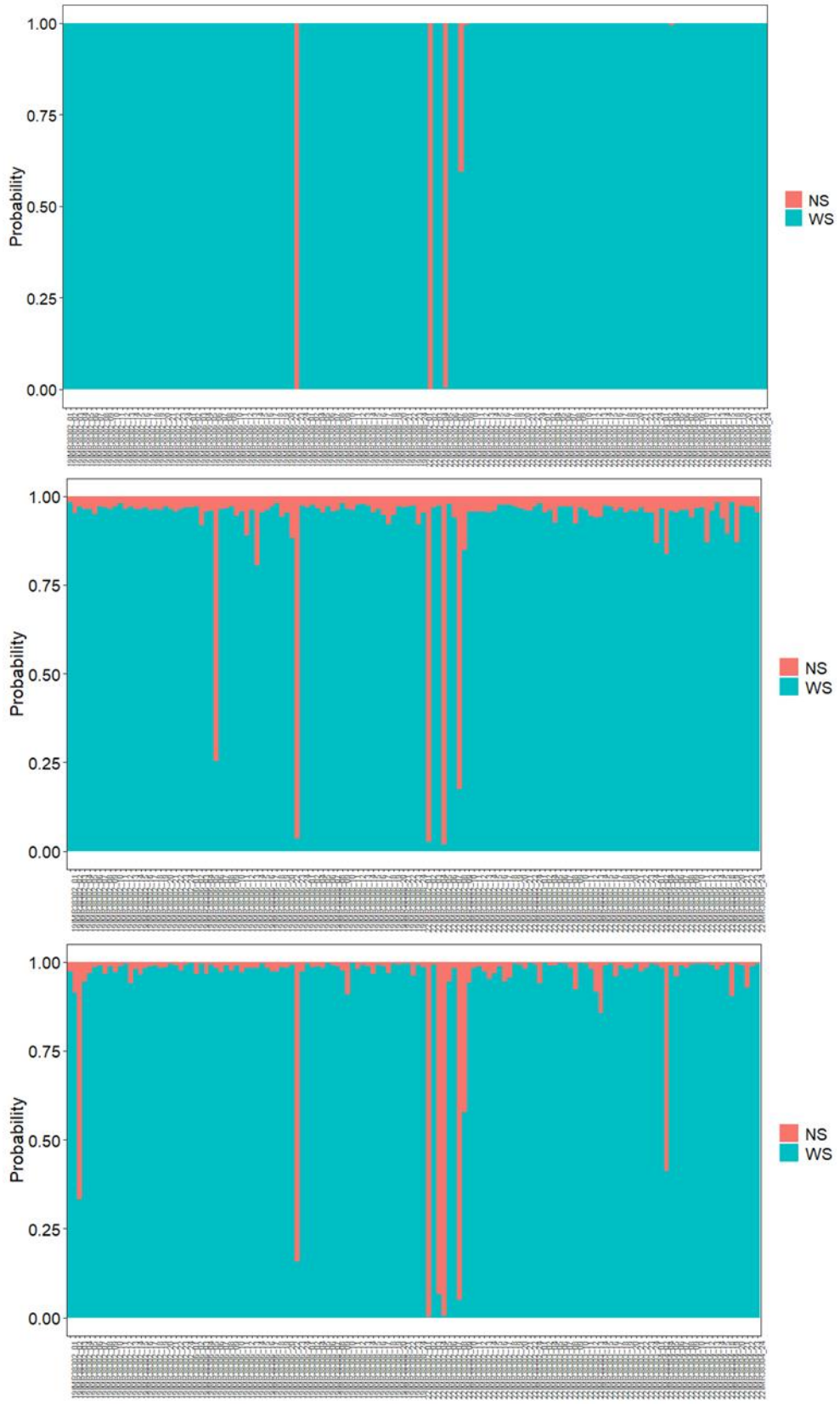


Figure S13. The outputs of the Group 2 assignment with the three marker panels (top) *2421_SNP* (middle) *25_SNP* (bottom) *36_SNP*. Each column represents an individual with the associated assignment probability.

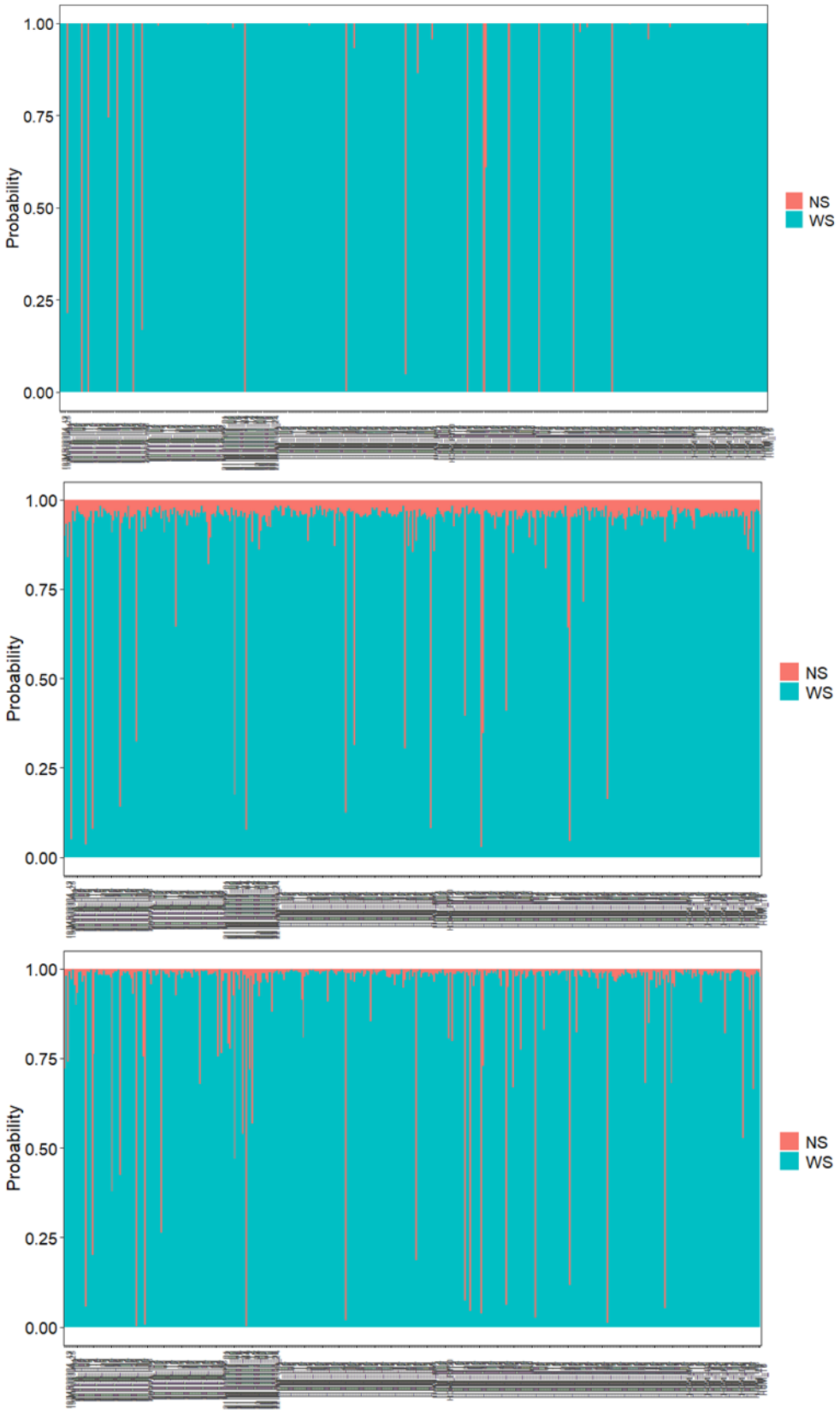


Figure S14. The outputs of the Group 3 assignment with the three marker panels (top) *2421_SNP* (middle) *25_SNP* (bottom) *36_SNP*. Each column represents an individual with the associated assignment probability.

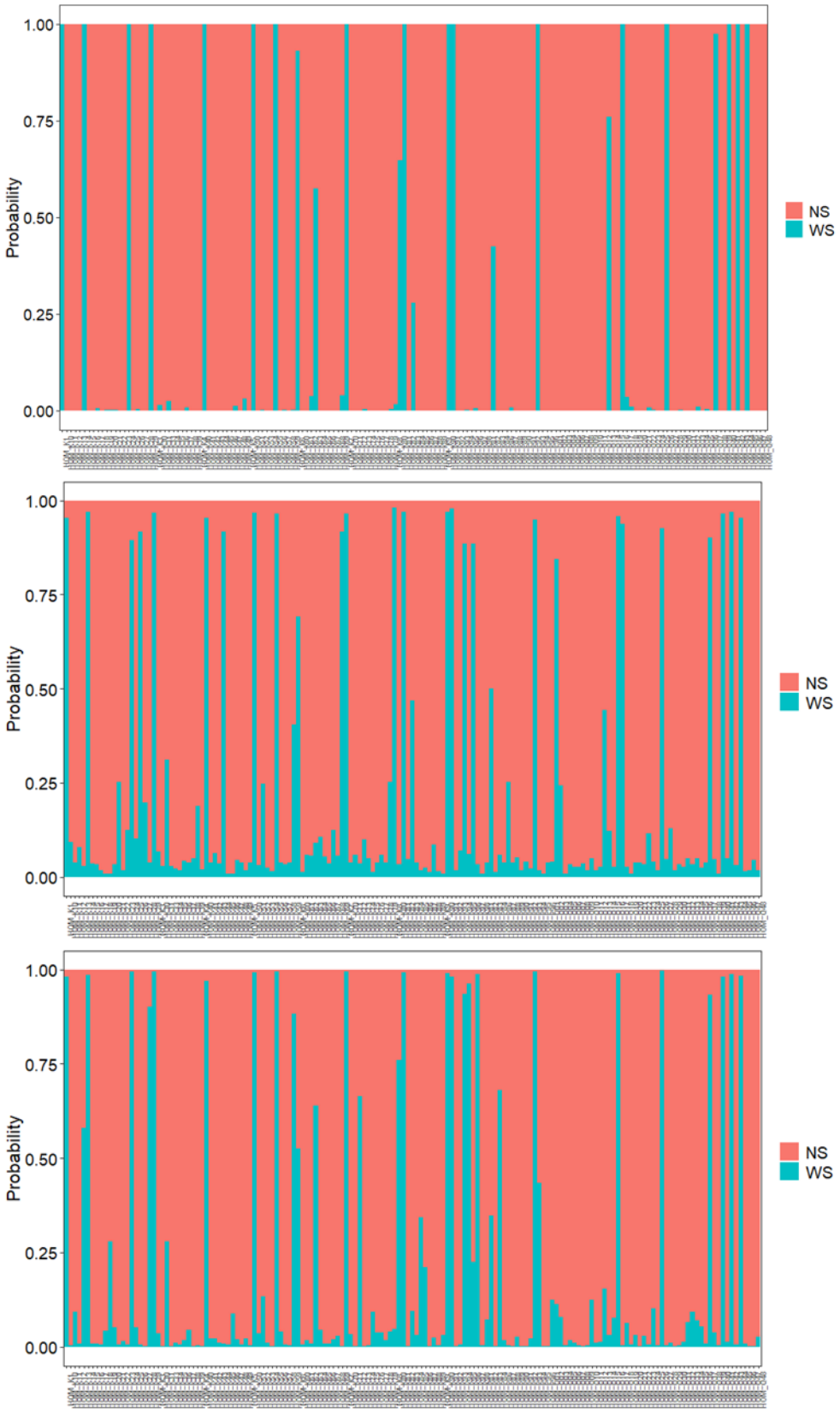


Figure S15. The outputs of the Group 4 assignment with the three marker panels (top) *2421_SNP* (middle) *25_SNP* (bottom) *36_SNP*. Each column represents an individual with the associated assignment probability.

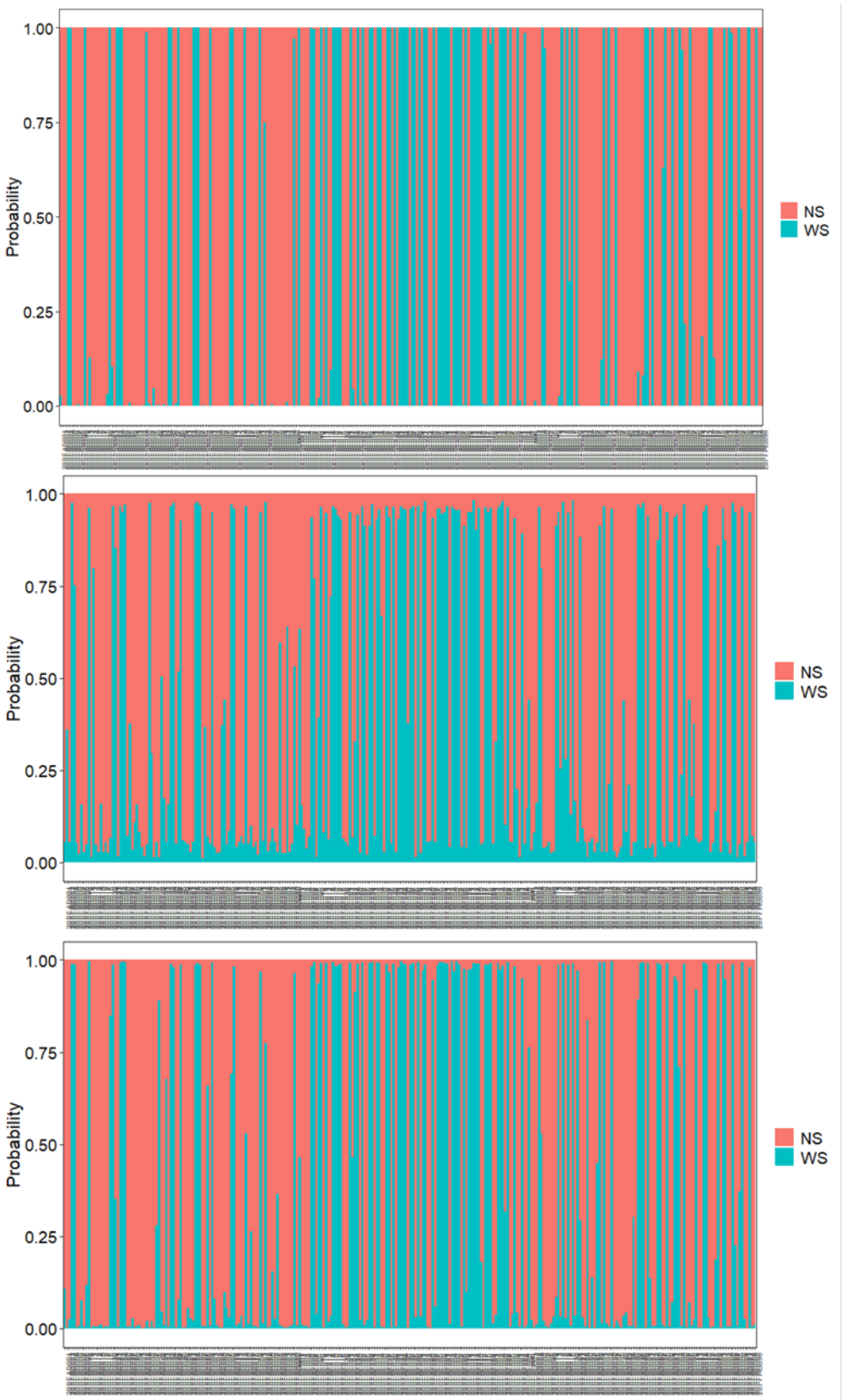


Figure S16. The outputs of the Group 5 assignment with the three marker panels (top) *2421_SNP* (middle) *25_SNP* (bottom) *36_SNP*. Each column represents an individual with the associated assignment probability.

Figure S17. The outputs of the Group 5 assignment with the three marker panels (top) *2421_SNP* (middle) *25_SNP* (bottom) *36_SNP*. Each column represents an individual with the associated assignment probability.

11. Annex 3 – Genotyping Quality Control

Analysis Summary

- **Batch Name:** PFA_HOM_Batch_1
- **Array Package Name:** Axiom_FSHSTK1D.r1
- **Array Type Name:** Axiom_FSHSTK1D
- **Array Display Name:** Axiom_FSHSTK1D.r1 (Horse Mackerel)
- **Workflow Type:** Best Practices Workflow
- **Date Created:** 19/01/2023 18:57:26

Sample Summary

- Number of input samples: 2287
- Samples passing DQC: 2245 out of 2287
- Samples passing DQC and QC CR: 2166 out of 2287
- Samples passing DQC, QC CR and Plate QC: 2166 out of 2287 (94.709%)
- Number of failing samples: 121
- Number of input samples without QC information: 0
- Number of Samples Genotyped: 2166
- Average QC CR for the passing samples: 98.522
- Inbred Penalty Applied: no

Plate QC Summary

Plate Barcode	Result	Number of files in a batch	Number of files failing dish QC	Number of files failing QC Call rate	Number of samples that passed	Percent of passing samples	Average call rate for passing samples	Filtered Call Rate
5514634456004030824977	PASSED	373	10	12	351	94.102	98.041	96.976
5514634456004030824984	PASSED	382	5	3	374	97.906	98.138	97.21
5514634456004030824988	PASSED	382	17	29	336	87.958	98.175	96.521
5514634456005030724308	PASSED	382	0	3	379	99.215	99.022	98.45
5514634470868113024410	PASSED	288	1	1	286	99.306	98.878	98.398
5514634470868113024411	PASSED	384	1	1	382	99.479	99.145	98.703
5514634470868113024416	PASSED	96	8	30	58	60.417	96.782	93.055

ProbeSet Metrics Summary

- Number of ProbeSets: 4242

ConversionType	Count	Percentage
PolyHighResolution	2627	61.928
Other	1122	26.45
OTV	215	5.068
NoMinorHom	214	5.045
MonoHighResolution	32	0.754
CallRateBelowThreshold	32	0.754

Marker Metrics Summary

- Number of Markers: 4103
- Number of BestandRecommended: 3031
- Percent BestandRecommended: 73.873

ConversionType	Count	Percentage
PolyHighResolution	2550	62.15
Other	1072	26.127
NoMinorHom	209	5.094
OTV	208	5.069
MonoHighResolution	32	0.78
CallRateBelowThreshold	32	0.78

Sample QC Thresholds

- DQC: ≥ 0.82
- QC call_rate: ≥ 90
- Average call rate for passing samples: ≥ 90
- Percent of passing samples: ≥ 95

SNP QC Thresholds

- species-type: Diploid
- cr-cutoff: ≥ 90
- fld-cutoff: ≥ 3.6
- het-so-cutoff: ≥ -0.1
- het-so-XChr-cutoff: ≥ -0.1
- het-so-otv-cutoff: ≥ -0.3
- hom-ro-1-cutoff: ≥ 0.6
- hom-ro-2-cutoff: ≥ 0.3
- hom-ro-3-cutoff: ≥ -0.9
- hom-ro: false
- num-minor-allele-cutoff: ≥ 2
- hom-ro-hap-1-XChr-cutoff: ≥ 0.1
- hom-ro-hap-1-MTChr-cutoff: ≥ 0.4
- hom-ro-hap-2-XChr-cutoff: ≥ 0.05
- hom-ro-hap-2-MTChr-cutoff: ≥ 0.2
- aaf-XChr-cut: < 0.36
- fld-XChr-cut: ≥ 4
- homfld-XChr-cut: ≥ 6.5
- homfld-YChr-cut: ≥ 6.5
- min-YChr-samples-cut: ≥ 5
- priority-order: PolyHighResolution, NoMinorHom, MonoHighResolution, OTV, UnexpectedGenotypeFreq, CallRateBelowThreshold, Other, OtherMA
- recommended: PolyHighResolution, NoMinorHom, OTV, MonoHighResolution, CallRateBelowThreshold
- y-restrict: ≤ 0.2
- min-genotype-freq-samples: ≥ 20
- genotype-p-value-cutoff: ≥ 1E-06

Multi-Allelic SNP QC Thresholds

- HomMMA-cutoff: > 10
- FLD-MA-cutoff: > 5.2
- FLD-MA-2-cutoff: > 5.2

- Min-FLD-MA-cutoff: > 0
- Min-FLD-MA-2-cutoff: > 0
- HetSO-MA-2-cutoff: > -0.1
- HomRO-MA-cutoff: > 0.5
- HomRO-MA-2-cutoff: > 0.5
- HomRO-MA-1-cutoff: > 1
- priority-order-MA: PolyHighResolution, NoMinorHom, MonoHighResolution, Hemizygous, UnexpectedGenotypeFreq, CallRateBelowThreshold, Other, OtherMA
- Best-CR-MA-cutoff: > 90

12. Annex 4 – Alternative Assignment Model

The authors believe the assignment models developed in Sections 3.5 and 4.5, based on multivariate analyses and a machine learning approach, were the most appropriate given the type of genetic markers used to discriminate the populations. However, for completeness and to enable comparison alternative assignment methods were also performed. The alternative assignments were conducted in *Geneclass2* (Piry et al., 2004) using two different assignment methods: a Bayesian based method (Rannala and Mountain, 1997) and an allele frequency based method (Paetkau et al., 1995). Both methods were initially tested with the five marker panels developed in the current study (Sections 3.5 and 4.5); *2421_SNP*, *25_SNP*, *8_SNP*, *36_SNP*, *17_SNP* and with default settings in Geneclass2.

Annex 4 Table 1. The results of the self-assignment analyses in Geneclass2 with the five marker panels developed in the current study.

	Rannala & Mountain, 1997		Paetkau et al., 1995	
Dataset	Quality Index	Correctly assigned	Quality Index	Correctly assigned
2421_SNP	96.91	97.00	96.72	96.70
25_SNP	93.43	93.60	93.44	93.60
8_SNP	89.91	93.50	89.96	93.50
36_SNP	95.98	96.10	95.99	96.10
17_SNP	88.30	92.70	88.36	92.70

The rates of correct assignment for the five marker panels were very high for all five marker panels (Annex 4 Table 1). It should be noted however that by default the methods in Geneclass2 use jack-knifing (leave-one-out) to test the self-assignment rate of the baselines, which can upwardly bias assignment accuracy because the training data used in the validations are nearly identical. Therefore, these should not be directly compared to the self-assignment rates of the *assignPOP* based models as those were more rigorously tested. Regardless the results indicated a high level of self-assignment for all of the marker panels. Following the approach in the *assignPOP* analyses only the *2421_SNP*, *25_SNP* and *36_SNP* were carried forward for the assignment of mixed samples.

Annex 4 Table 2. The number of fish in each assignment group assigned to the Western (WS) or North Sea (NS) populations based on the analyses of the three different datasets (*2421_SNP, 25_SNP and 36_SNP*) in performed *assignPOP* and in *Geneclass2* using the Rannala & Mountain (R&M) and Paetkau (PTK) methods.

Marker Panel	Assignment Group	assignPOP		Geneclass2 (R& M)		Geneclass2 (PTK)	
		WS	NS	WS	NS	WS	NS
2421_SNP	1	33	0	33	0	33	0
	2	134	3	133	4	133	4
	3	411	15	402	24	402	24
	4	22	137	19	140	19	140
	5	101	185	94	192	94	192
	6	161	22	158	25	159	24
25_SNP	1	33	0	32	1	32	1
	2	132	5	132	5	132	5
	3	409	17	400	26	400	26
	4	28	131	24	135	24	135
	5	114	172	103	183	102	184
	6	163	20	160	23	160	23
36_SNP	1	33	0	33	0	33	0
	2	130	7	128	9	128	9
	3	407	19	396	30	396	30
	4	29	130	22	137	22	137
	5	110	176	98	188	98	188
	6	160	23	157	26	157	26

The assignment of the mixed samples was performed using the same genepop files for each of the assignment groups in Sections 3.6 and 4.6. Whilst *Geneclass2* does output loglikelihood and probability values, these are different to the probabilities generated by *assignPOP* and as such cannot be directly compared. In order to be able to undertake a direct comparison of the outputs of the different assignment models the most likely source population in each assignment was taken as the final assignment regardless of probability i.e. no assignment probability threshold was applied. Therefore these assignments are for comparative purposes only.

The results of the R&M and PTK assignments in *Geneclass2* were almost identical within marker panel (Annex 4 Tables 2 & 3) and differed by one individual assignment in only two of the comparisons (*2421_SNP Assignment Group 6* and *25_SNP Assignment Group 5*). The results of the Geneclass2 based assignments were also very similar to the *assignPOP* assignments, differing by at most 4% in the *25_SNP Assignment Group 5* comparison (Annex 4 Tables 2 & 3). Examination of the individual assignments revealed that a high proportion of the disagreements concerned individuals with a lower assignment probability, soe of which were classified as *Below Threshold* (raw data available in supporting files).

Overall the results of the alternative assignment models were not considered to be significantly different to the *assignPOP* results however using adaptive markers which are known to be in linkage disequilibrium violates the basic theoretical assumptions of these model. Therefore, the *svm* based machine learning models are more appropriate for use with the marker panels in the current study and the alternative models were not developed further.

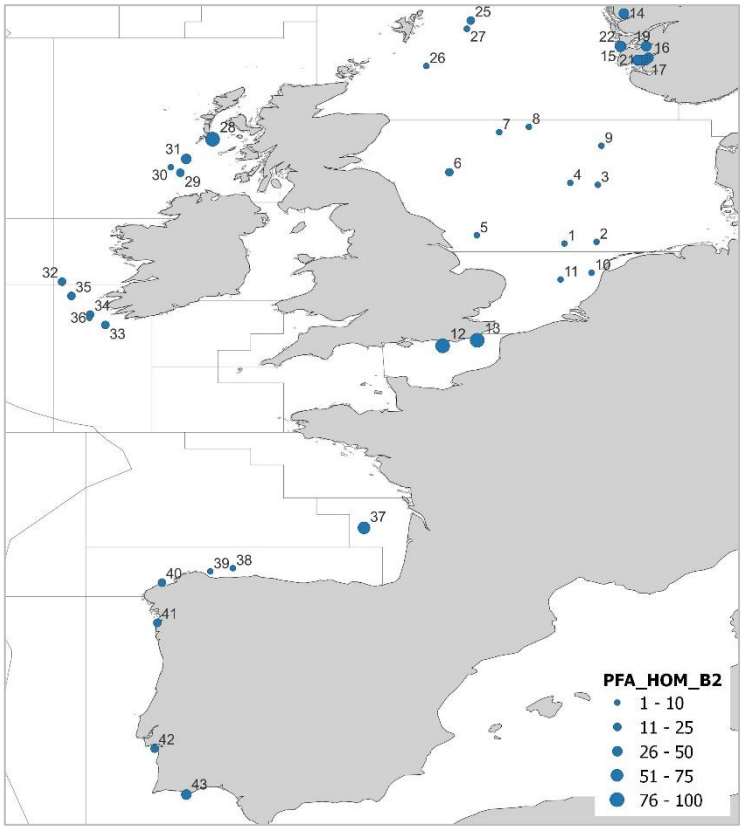
Annex 4 Table 3. The percentage of fish in each assignment group assigned to the Western (WS) or North Sea (NS) populations based on the analyses of the three different datasets (*2421_SNP, 25_SNP and 36_SNP*) in performed *assignPOP* and in *Geneclass2* using the Rannala & Mountain (R&M) and Paetkau (PTK) methods.

Marker Panel	Assignment Group	assignPOP		Geneclass2 (R& M)		Geneclass2 (PTK)	
		WS	NS	WS	NS	WS	NS
2421_SNP	1	100	0	100	0	100	0
	2	98	2	97	3	97	3
	3	96	4	94	6	94	6
	4	14	86	12	88	12	88
	5	35	65	33	67	33	67
	6	88	12	86	14	87	13
25_SNP	1	100	0	97	3	97	3
	2	96	4	96	4	96	4
	3	96	4	94	6	94	6
	4	18	82	15	85	15	85
	5	40	60	36	64	36	64
	6	89	11	87	13	87	13
36_SNP	1	100	0	100	0	100	0
	2	95	5	93	7	93	7
	3	96	4	93	7	93	7
	4	18	82	14	86	14	86
	5	38	62	34	66	34	66
	6	87	13	86	14	86	14

13. Annex 5 – Additional samples analysed

Subsequent to the completion of the analyses in the current document it was apparent that there were additional samples collected, by a number of organisations, that would contribute to the further understanding of horse mackerel populations. Therefore the NPWG of EAPO decided to fund the analyses of these samples as a priority before the 2024 benchmark.

The samples were collected from across the three stock areas and filled in some of the previous gaps in sampling, particularly in divisions 4.b and 6.a (Annex 5 Figure 1 and Table 1). Similar to the samples in the main document the most northerly samples collected in division 4.a contained comparatively larger individuals than the more southerly samples (Annex 5 Table 1, 2a, 2b). The samples from divisions 4.b and 4.c contained comparatively smaller individuals though it should be noted that these samples were collected during the North Sea IBTS in August and each sample comprised only a small number of individuals (Annex 5 Table 1). Comparison of the contents of these samples and their respective hauls of origin on the North Sea IBTS (data not shown) indicated that in some cases there were more individuals captured per haul than were sampled for genetic analysis and that there appears to have been a selection applied. Therefore the length-frequency of all of the samples in divisions 4.b. and 4.c may not be representative of the horse mackerel caught in these areas.



Annex 5 Figure 1. Additional horse mackerel samples included in the Annex 5 analyses.

The samples in the eastern channel had a similar length frequency to those in the main document and comprised maturing individuals with a modal length of 21-22cm (Annex 5 Tables 1 and 2a). The samples in division 6.a off the northwest of Ireland comprised comparatively larger fish than those off the southwest of Ireland (Annex 5 Tables 1 and 2b), though larger fish were present further south in division 8.b. The smallest fish were sampled in division 9.a, north and south of Lisbon, though the fish sampled on the southern Portuguese coast were larger than those samples off the southwest of Portugal (Annex 5 Tables 1 and 2b).

Annex 5 Table 1. Details of the additional samples analysed subsequent to the completion of the main analysis in the current document.

Collector	Sample	Catch Location	ICES stock	ICES Areas	Date	Lat	Lon	Collected	Submitted	Genotyped
CEFAS	1	Central North Sea	North Sea	4.b	10/08/2023	53.74	3.53	1	1	1
CEFAS	2	Central North Sea	North Sea	4.b	10/08/2023	53.80	4.50	2	2	2
CEFAS	3	Central North Sea	North Sea	4.b	13/08/2023	55.53	4.55	1	1	1
CEFAS	4	Central North Sea	North Sea	4.b	13/08/2023	55.59	3.71	3	3	3
CEFAS	5	Central North Sea	North Sea	4.b	14/08/2023	54.00	0.87	7	7	7
CEFAS	6	Central North Sea	North Sea	4.b	16/08/2023	55.92	0.03	13	13	13
CEFAS	7	Central North Sea	North Sea	4.b	21/08/2023	57.13	1.55	10	10	10
CEFAS	8	Central North Sea	North Sea	4.b	21/08/2023	57.30	2.45	10	10	10
CEFAS	9	Central North Sea	North Sea	4.b	22/08/2023	56.72	4.66	10	3	3
CEFAS	10	Southern North Sea	North Sea	4.c	09/08/2023	52.86	4.35	5	5	5
CEFAS	11	Southern North Sea	North Sea	4.c	09/08/2023	52.65	3.41	4	4	3
PFA	12	Eastern Channel	North Sea	7.d	20/10/2020	50.63	-0.17	100	95	94
PFA	13	Eastern Channel	North Sea	7.d	27/11/2020	50.80	0.88	100	49	49
IMR	14	Norwegian Coast	North Sea	4.a	03/02/2022	60.75	5.33	30	24	9
IMR	15	Norwegian Coast	North Sea	4.a	03/01/2023	59.75	5.25	30	24	19
IMR	16	Norwegian Coast	North Sea	4.a	03/01/2023	59.40	6.08	30	24	20
IMR	17	Norwegian Coast	North Sea	4.a	05/01/2023	59.33	6.00	30	24	8
IMR	18	Norwegian Coast	North Sea	4.a	18/01/2023	59.33	5.92	30	24	15
IMR	19	Norwegian Coast	North Sea	4.a	20/01/2023	59.75	6.02	30	24	21
IMR	20	Norwegian Coast	North Sea	4.a	19/01/2023	59.32	5.75	30	24	21
IMR	21	Norwegian Coast	North Sea	4.a	24/01/2023	59.33	5.83	30	24	23
IMR	22	Norwegian Coast	North Sea	4.a	20/01/2023	59.74	5.22	30	24	20
IMR	23	Norwegian Coast	North Sea	4.a	20/01/2023	59.33	5.75	30	24	20
PFA	24	Northern North Sea	Western	4.a	15/07/2022	60.28	0.57	3	3	3
PFA	25	Northern North Sea	Western	4.a	19/07/2022	60.53	0.68	12	12	12
PFA	26	Northern North Sea	Western	4.a	23/07/2022	59.15	-0.67	3	3	3
PFA	27	Northern North Sea	Western	4.a	15/07/2022	60.28	0.57	10	10	10
MI	28	West of Scotland	Western	6.a	19/07/2017	56.92	-7.16	100	48	48
MI/CEFAS	29	Northwest Ireland	Western	6.a	09/07/2023	55.90	-8.14	11	11	10
MI/CEFAS	30	Northwest Ireland	Western	6.a	10/07/2023	56.07	-8.43	1	1	1
MI/CEFAS	31	Northwest Ireland	Western	6.a	11/07/2023	56.32	-7.96	28	28	26
PFA	32	West Ireland	Western	7.b	11/07/2022	52.58	-11.73	18	18	18
PFA	33	Southwest Ireland	Western	7.j	10/07/2022	51.27	-10.42	16	16	16
PFA	34	Southwest Ireland	Western	7.j	10/07/2022	51.58	-10.88	13	13	13
PFA	35	Southwest Ireland	Western	7.j	11/07/2022	52.15	-11.45	12	12	12
PFA	36	Southwest Ireland	Western	7.j	12/07/2022	51.48	-10.90	9	9	9
PFA	37	Bay of Biscay	Western	8.b	08/03/2015	45.09	-2.56	63	48	48
IEO	38	Northern Spanish Shelf	Western	8.c	21/04/2023	43.86	-6.54	6	6	6
IEO	39	Northern Spanish Shelf	Western	8.c	20/04/2023	43.76	-7.23	8	8	8
IEO	40	Northern Spanish Shelf	Western	8.c	17/04/2023	43.42	-8.70	21	14	14
IEO	41	Northwest Spain	Southern	9.a	09/04/2023	42.20	-8.84	20	20	20
IPMA	42	Southwest Portugal	Southern	9.a	25/03/2023	38.37	-8.92	21	21	21
IPMA	43	South Portugal	Southern	9.a	29/03/2023	36.97	-7.96	30	27	25

Only a small number of spawning individuals (maturity stage 3 – 6 point scale) were present in the samples collected. These were primarily from division 8.c and 9.a, which were sampled in April and March, respectively (Annex 5 Tables 1 and 3). There were also a small number of spawning individuals mixed with the samples collected off the northwest and southwest of Ireland and in the western part of division 4.a, all of which were collected in July, which is the end of the horse mackerel spawning season. No maturity information was available from the samples collected in Norwegian waters, though these were collected in January and February and as such can be assumed not to contain spawning fish. The samples collected in divisions 4.b and 4.c were noted to contain a high proportion of suggested stage 4.b individuals on the 9-point maturity scale. This equates to maturity stage 2 on the 6-point maturity scale (Table 2) and as such these are not considered to be spawning individuals. It should be noted that observing a significant proportion of maturity stage 2 individuals in August would be considered unusual as the spawning season usually ends around July, therefore it may be the case that these were actually stage 7-8 (9-pt scale) or stage 4 (6-pt scale) fish. As horse mackerel is an indeterminate spawner it can be difficult to macroscopically distinguish these stages. This does not affect the genetic assignments and analyses presented in the current study.

Annex 5 Table 2a. Length frequency of the additional horse mackerel samples analysed in Annex 5.

	Sample number																						
TL (cm)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
14																							
14.5																							
15																							
15.5																							
16																							
16.5																							
17																							
17.5												2											
18												6											
18.5												4											
19		1								2		7	1										
19.5										2		8											
20										1		9	2										
20.5												18	6										
21		1									1	10	9										
21.5												10	18										
22	1			1							1	9	10										
22.5					1							8	13										
23											1	2	6										
23.5												3	10										
24					2						1	1	3										
24.5					2							1	6										
25													5		1								
25.5					1							1	1										
26					1								3										
26.5												1	3									1	
27													1									1	
27.5													1									2	
28																							
28.5													1									1	
29																							
29.5													1										
30																							
30.5																							
31																							
31.5																							
32						2	1	2															
32.5						3																1	
33						1	6	1									1		1	1		1	1
33.5						2								2	2		2						1
34				1		2	1	3									2	2	1	3	3	2	3
34.5															2	1	1	2	1	1	1	2	
35			1			1		2	3					1	3	1	8	2	2	2	3	2	1
35.5				1										3	4	1	3	3	5	2	4	6	5
36						1	1	2						9	3	2	3	1	1	6	2	1	4
36.5						1								1	1	2	1	6	2		4	2	2
37							1							3	3	4	3	4	3	3	3		1
37.5														4	3	5	1	4	2	4	1	3	4
38															1		2		1	3	2		2
38.5														3	3	2	1	1	3	4	1	2	1
39														1		3	1	1	2		3		3
39.5																3	1		2		1	1	1
40															1	3		2	1		1		
40.5														2	1			2			1		1
41																			1	1			
41.5														1		1			1			1	
42															1	1							
42.5															1	1							
43																			1				
43.5																							
44																							
44.5																							
45																							
45.5																							
46																						1	

Annex 5 Table 2b. Length frequency of the additional horse mackerel samples analysed in Annex 5.

	Sample number																		
TL (cm)	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43
14																				
14.5																		2		
15																		1	1	
15.5																			3	
16																		3	5	
16.5																		2	5	
17																		6	6	
17.5																		3	1	
18																		3		
18.5																				
19																				
19.5																				
20																				
20.5																	1			
21																				
21.5																	3			1
22																	2			3
22.5																1				4
23																	5			7
23.5																1	1			4
24																1				2
24.5																1				5
25						1											1			3
25.5																				
26						1					1	1		3		1				
26.5											1	1	1	2		1				
27																				
27.5											1			1						
28					2	1					1	1		1						
28.5					1					2	1			1		1	1			1
29					1						1	1	2	1						
29.5					1					1	1	3	1	1						
30									2		1		3		1					
30.5					2				1		1	2	2	2		1				
31					8				2		1				1					
31.5					5				4	3		1		1	1					
32	1				1			5		1					1					
32.5					8				3	5	2	2								
33				1	16			5	2	2				1						
33.5			1		11				2					2						
34	1		1		13	4		6	2					1	1					
34.5		1			6									2						
35				2	13	2	1	6		1				2						
35.5		4		3	3					1				10	1					
36		2		1		1		3						5						
36.5		2	1	1	2									10						
37		1		1	3	1		2			1			5						
37.5														3						
38		2			1			1						3						
38.5														2						
39	1				2									1						
39.5				1										2						
40														1						
40.5																				
41					1															
41.5																				
42																				
42.5																				
43																				
43.5																				
44																				
44.5																				
45																				
45.5																				
46																				

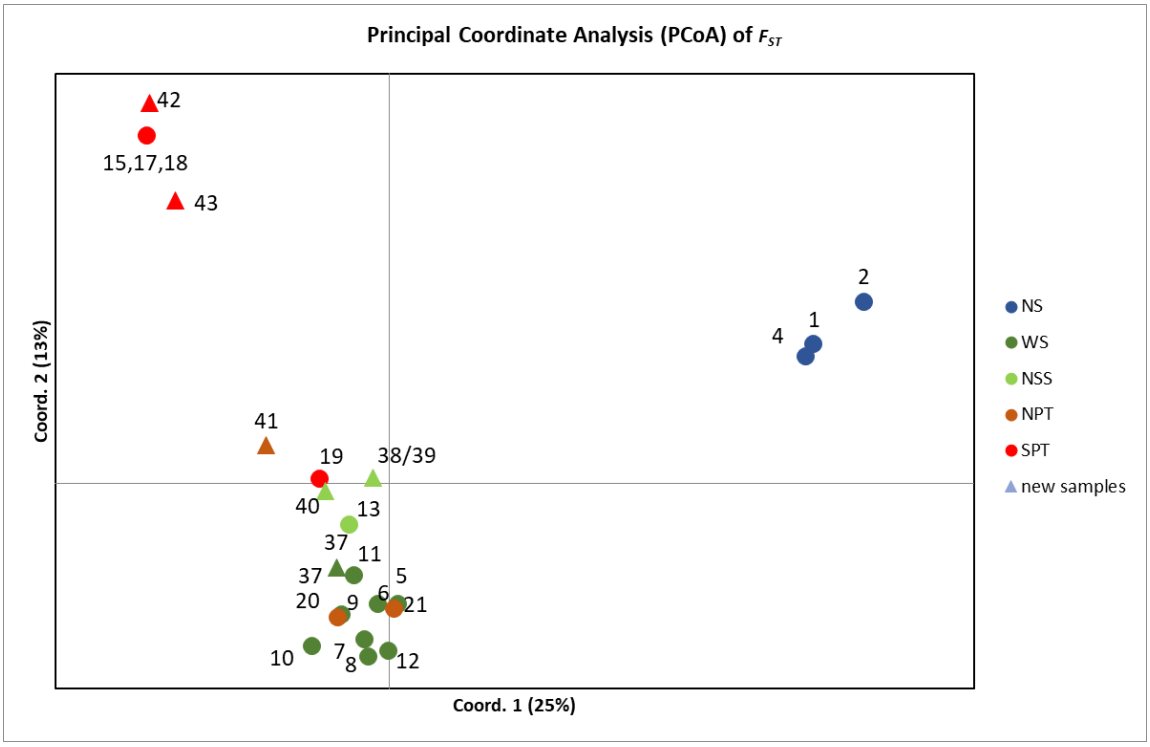
Annex 5 Table 3. Maturity stages (6-pt) of the additional horse mackerel samples analysed in Annex 5.

	Maturity Stage (6-pt)						
Sample number	**1**	**2**	**3**	**4**	**5**	**6**	**NA**
1	1						
2	2						
3		1					
4		3					
5	4	3					
6	2	11					
7		10					
8		10					
9		3					
10	5						
11	1						3
12	10	88		2			
13		100					
14							30
15							30
16							30
17							30
18							30
19							30
20							30
21							30
22							30
23							30
24		1	2				
25		10	2				
26		3					
27		6	4				
28			3	97			
29		10		1			
30		1					
31		20	8				
32		17	1				
33		12	4				
34		13					
35		12					
36		7	2				
37		63					
38			6				
39			8				
40			14				
41	4	5		11			
42	17	4					
43		6	24				

The quality control and genetic analyses of the additional samples followed the methods described in Sections 3 and 4 of the main document and as such they are not repeated here. The analyses in the current annex were divided into two parts, the first concerning the samples collected in the southern part of the stock areas (divisions 8.b, 8.c and 9.a) and the second concerning the samples collected in the North Sea and Western stock areas.

As described in Section 4.3 of the main document it was not possible to develop an assignment model that included the southern population due to the low number of representative spawning baseline samples available for that population. Addition of the new samples did not change this situation and therefore it was decided to instead qualitatively assess the population of origin of these samples (samples #37-43, Annex 5 Figure 1 and Table 1). The number of individuals in some of the samples was low and therefore the results should be viewed with caution. In order to ensure that there was a minimum of 10 individuals per sample for inclusion in the analyses samples #38 and 39 were combined before proceeding.

The samples were analysed as per the exploratory analyses detailed in Sections 3.3 and 4.3 using the *2421_SNP* panel. In the *2421_SNP* dataset fifteen SNPs from eleven chromosomes did yield any usable data in the genotyping however given the high level of redundancy in the SNP panels and the results of the sensitivity analyses detailed in Section 4.5 it was concluded that the missing data did not have an impact on the subsequent analyses. The data for samples #37-43 were added to the dataset used to derive Figure 16 in the main document and the F_{ST} analyses was performed again, with the results again visualised through PCoA. The two southern most samples (#42 from Setubal and #43 from Faro) clustered with the samples representing the southern population (Annex 5 Figures 1 and 2). The sample from the northern part of division 9.a (#41) and those from divisions 8.c (#38/39, 40) and 8.b (#37) clustered with the western population samples. Though the analysis was largely qualitative it supports the previous results that indicates that there is a southern population of horse mackerel and that there is mixing between the western and southern populations within division 9.a. It also further highlights that sampling levels in this area in the current study are low and more intensive spatial and temporal sampling of both baseline spawning samples and potential mixed juvenile and adult samples is required to resolve the pattern of mixing in this area.



Annex 5 Figure 2. Principal Coordinate Analysis (PCoA) of F_{ST} of baseline samples from Section 4.3 in the main document and the additional samples. The new samples are indicated with a triangle icon and are colour coded as per the baseline samples according to sampling location. NS = North Sea, WS = Western, NSS = Northern Spanish Shelf, NPT = Northern Portugal, SPT = Southern Portugal. The sample details are provided in Annex 5 Table 1.

Samples #1-37, which ranged from the Norwegian coast to the Bay of Biscay (Annex 5 Figure 1 and Table 1) were assigned with the assignment models developed in Sections 3.5 and 4.5 of the main document and are directly comparable and combinable with the assignments of mixed samples in Section 4.6. All model parameters were the same and the same QC thresholds were applied pre and post assignment. As noted above, in the *2421_SNP* dataset fifteen SNPs from eleven chromosomes did yield any usable data in the genotyping and in the *36_SNP* dataset one SNP did not yield any usable

data. Given the high level of redundancy in the SNP panels and the results of the sensitivity analyses detailed in Section 4.5 it was concluded that the missing data did not have an impact on the subsequent assignments. It should be noted that the samples from the Norwegian coast, which comprised fin tissue instead of muscle tissue had the highest number of failed samples. This was also the case in the analyses in the main document and highlights the need for standardised sampling of muscle tissue to be introduced in preference to fin tissue.

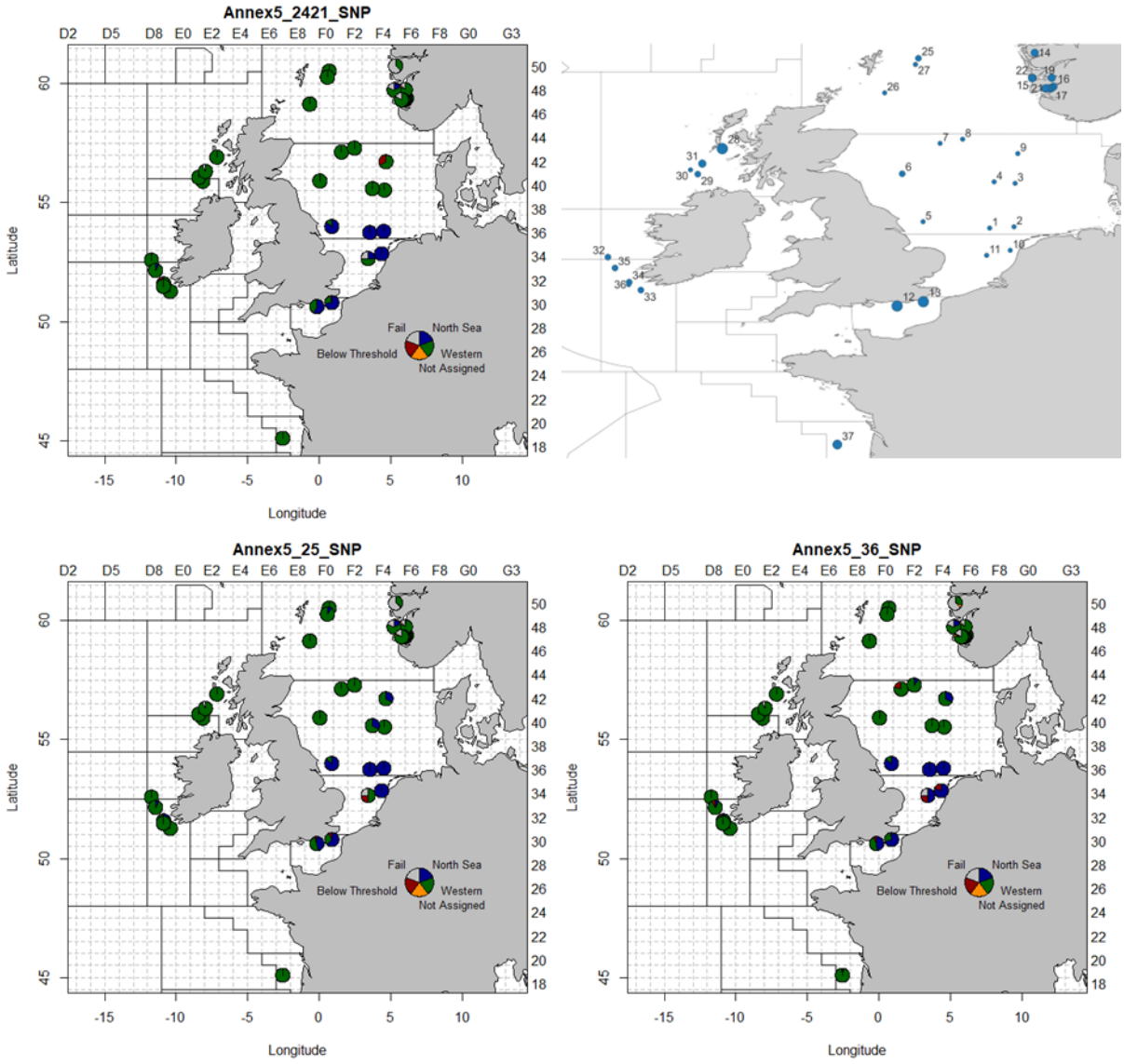
Annex 5 Table 4. The results of the assignment of the additional samples with the three marker panels. The number of individuals assigned to each category is indicated. WS = Western, NS = North Sea, BT = Below Threshold, NA = Not Assigned, F= Fail.

	2421_SNP					25_SNP					36_SNP				
Sample	**WS**	**NS**	**BT**	**NA**	**Fail**	**WS**	**NS**	**BT**	**NA**	**Fail**	**WS**	**NS**	**BT**	**NA**	**Fail**
1	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0
2	0	2	0	0	0	0	2	0	0	0	0	2	0	0	0
3	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0
4	3	0	0	0	0	2	1	0	0	0	3	0	0	0	0
5	1	6	0	0	0	1	6	0	0	0	1	6	0	0	0
6	13	0	0	0	0	13	0	0	0	0	13	0	0	0	0
7	10	0	0	0	0	10	0	0	0	0	8	0	2	0	0
8	10	0	0	0	0	10	0	0	0	0	9	1	0	0	0
9	2	0	1	0	0	2	1	0	0	0	2	1	0	0	0
10	0	5	0	0	0	0	5	0	0	0	0	4	1	0	0
11	2	1	0	0	1	2	0	1	0	1	0	2	1	0	1
12	40	54	0	0	1	46	44	4	0	1	42	47	5	0	1
13	11	36	2	0	0	14	31	4	0	0	14	34	1	0	0
14	9	0	0	0	15	8	0	0	1	15	7	0	0	2	15
15	19	0	0	0	5	19	0	0	0	5	18	0	0	1	5
16	19	0	0	1	4	17	0	1	2	4	18	0	0	2	4
17	8	0	0	0	16	8	0	0	0	16	5	1	2	0	16
18	15	0	0	0	9	15	0	0	0	9	15	0	0	0	9
19	20	0	0	1	3	19	1	0	1	3	20	0	0	1	3
20	21	0	0	0	3	20	1	0	0	3	21	0	0	0	3
21	23	0	0	0	1	23	0	0	0	1	22	0	1	0	1
22	16	4	0	0	4	16	4	0	0	4	16	4	0	0	4
23	20	0	0	0	4	19	0	0	1	4	18	0	1	1	4
24	2	1	0	0	0	3	0	0	0	0	2	1	0	0	0
25	12	0	0	0	0	12	0	0	0	0	12	0	0	0	0
26	3	0	0	0	0	3	0	0	0	0	3	0	0	0	0
27	10	0	0	0	0	9	1	0	0	0	10	0	0	0	0
28	48	0	0	0	0	48	0	0	0	0	48	0	0	0	0
29	10	0	0	0	1	10	0	0	0	1	9	0	1	0	1
30	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0
31	26	0	0	0	2	26	0	0	0	2	26	0	0	0	2
32	18	0	0	0	0	18	0	0	0	0	18	0	0	0	0
33	16	0	0	0	0	16	0	0	0	0	16	0	0	0	0
34	11	0	2	0	0	11	2	0	0	0	11	1	1	0	0
35	11	1	0	0	0	11	1	0	0	0	10	1	1	0	0
36	9	0	0	0	0	9	0	0	0	0	9	0	0	0	0
37	48	0	0	0	0	48	0	0	0	0	46	1	1	0	0

As was the case with the results in the main document the assignment of the samples with the three marker panels were consistent with each other and only small differences were observed (Annex 5 Table 4). The pattern of assignment of individuals across the sampling area (Annex 5 Figures 3 and 4) was also consistent with what was observed in the analyses in the main document (Figure 25). The samples from the Norwegian coast (#14-23) collected in Q1 and the samples from the northern North Sea (#24-27) collected in July comprised almost exclusively western horse mackerel. The samples collected in division 6.a in July 2017 (#28) and July 2023 (#29-31), in divisions 7.b (#32) and 7.j (#33-36) in July 2022 and in 8.b in March 2015 (#37) also comprised almost entirely western horse mackerel. The small numbers of fish assigned to the North Sea in these areas was consistent with the error rates

of the assignment models and as such may be error or may be migrants. Further analyses are underway to investigate if it is possible to differentiate these two potential outcomes.

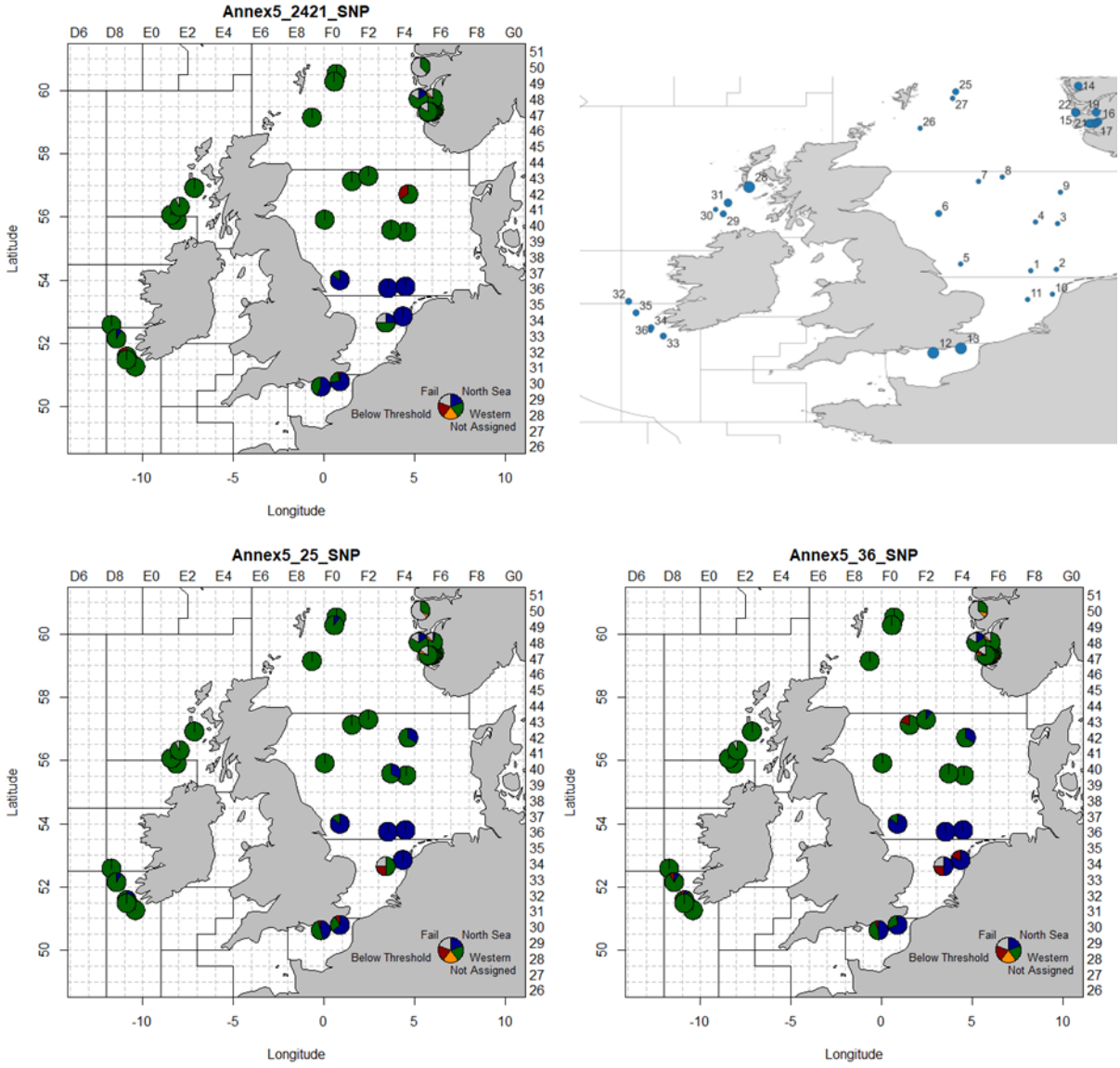
As noted the number of individuals in the samples collected in division 4.b was small (Annex 5 Table 1). There was a spatial pattern to the assignment of the individuals in these samples with those in the northern part of division 4.b (#3, 4, 6-9) being predominantly assigned to the western horse mackerel and those in the southern part (#1, 2, 5) being assigned to the North Sea horse mackerel (Annex 5 Figures 3 and 4). The assignment of samples in division 4.c (#10-11) was consistent with this with a higher proportion of individuals assigned to the North Sea. The assignments of the samples in the eastern channel (#12-13) collected in October and November 2020 were consistent with the assignments of the eastern channel samples in Section 4.6. The October sample had a higher proportion of western horse mackerel than the later sample, which was also seen with the September sample in Section 4.6. It should be noted that these earlier samples were also caught further west than the later samples which may indicate a temporal and spatial pattern to the mixing in the area. Further sampling over multiple years is required to investigate this further.



Annex 5 Figure 3. The outputs of the assignment of the additional samples with the three marker panels.

In summary all of the patterns observed in the analyses of the samples in Annex 5 were consistent with previous results and further add to the understanding of the migrations and mixing of the horse mackerel populations. Further samples can be added to the analyses as they become available in the future, which will help refine the spatial and temporal distribution of the populations.

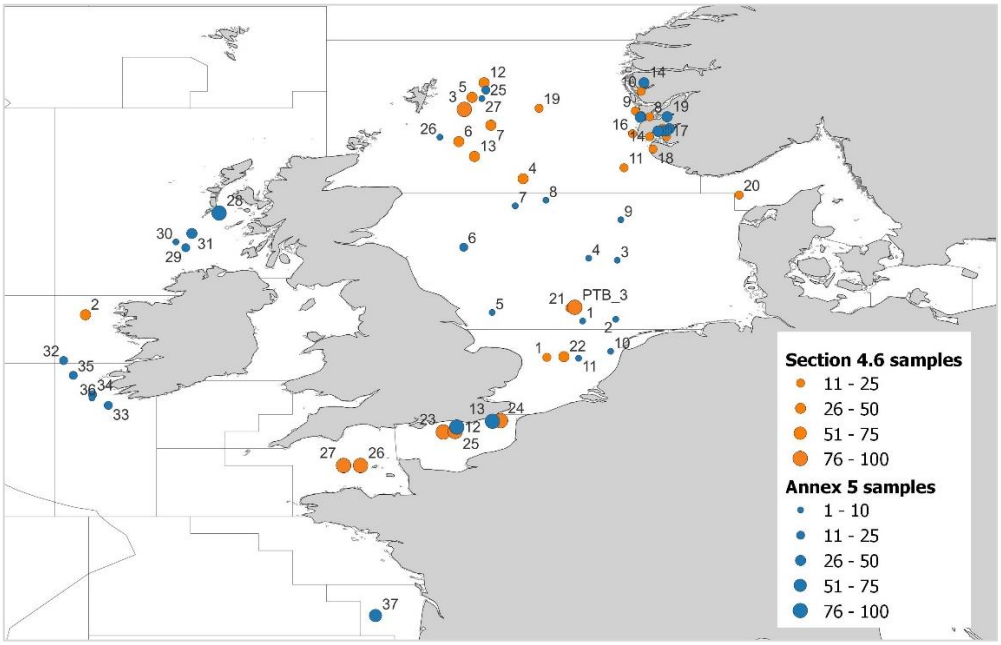
For ease of interpretation the assignments presented in Annex 5 and in Section 4.6 were combined into a single set of plots in Annex 6.



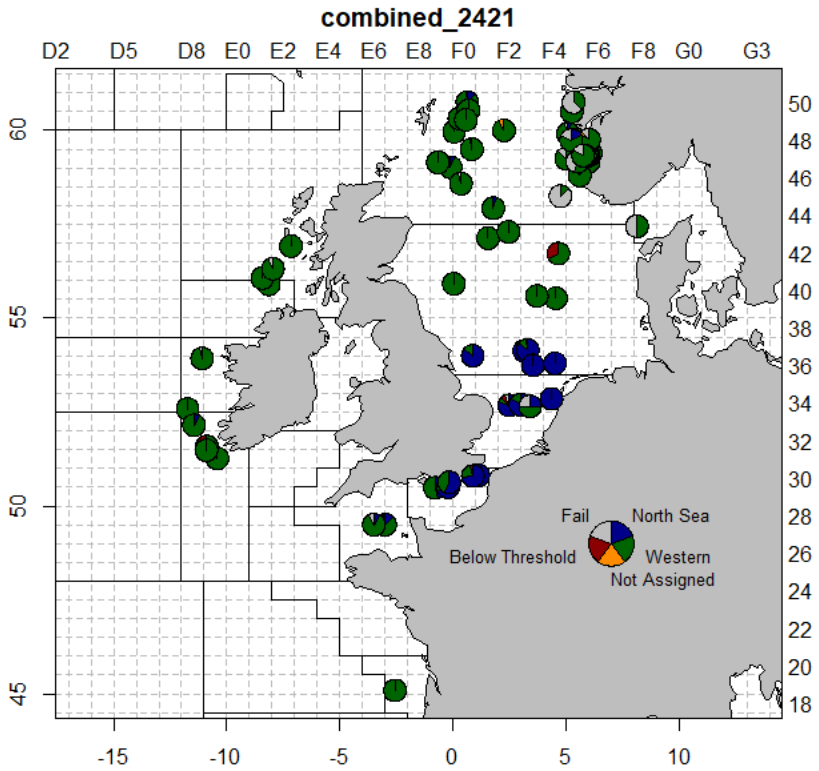
Annex 5 Figure 4. The outputs of the assignment of the additional samples with the three marker panels with the Bay of Biscay sample not shown in order to focus on the more northerly samples.

14. Annex 6 – Combined assignment plots

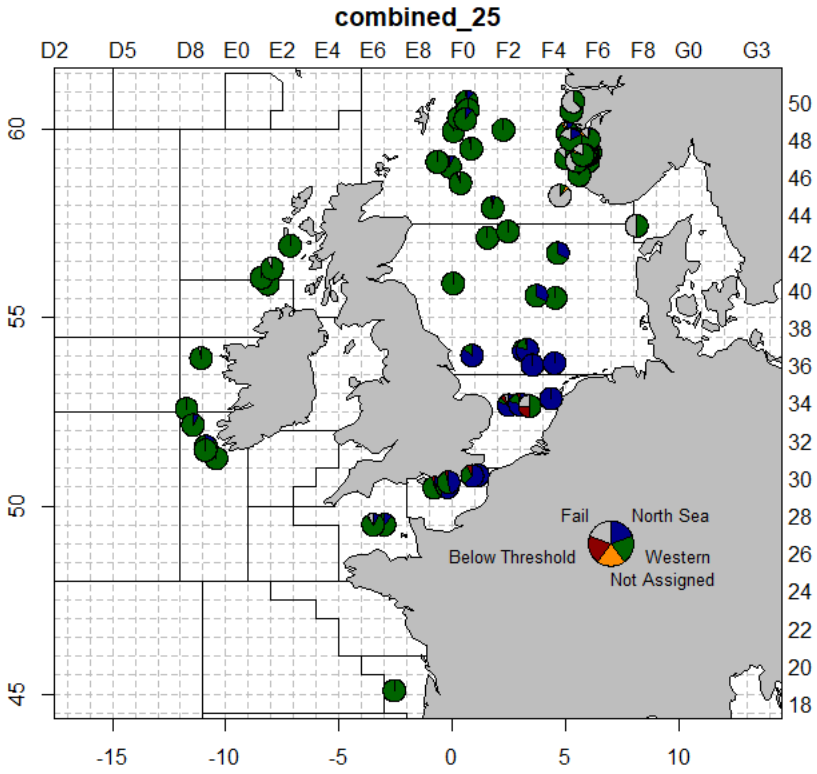
The following Annex contains combined figures for the mixed samples assigned in Section 4.6 of the main document and in Annex 5.



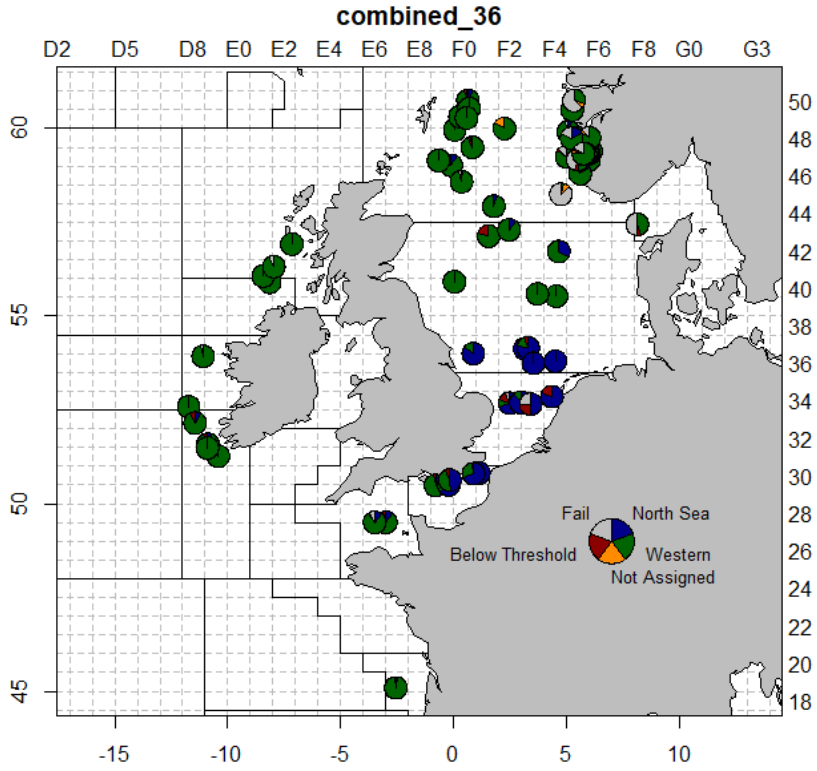
Annex 6 Figure 1. A combined plot of the mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5.



Annex 6 Figure 2. The output of the combined assignments with the *2421_SNP* panel of the mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5.

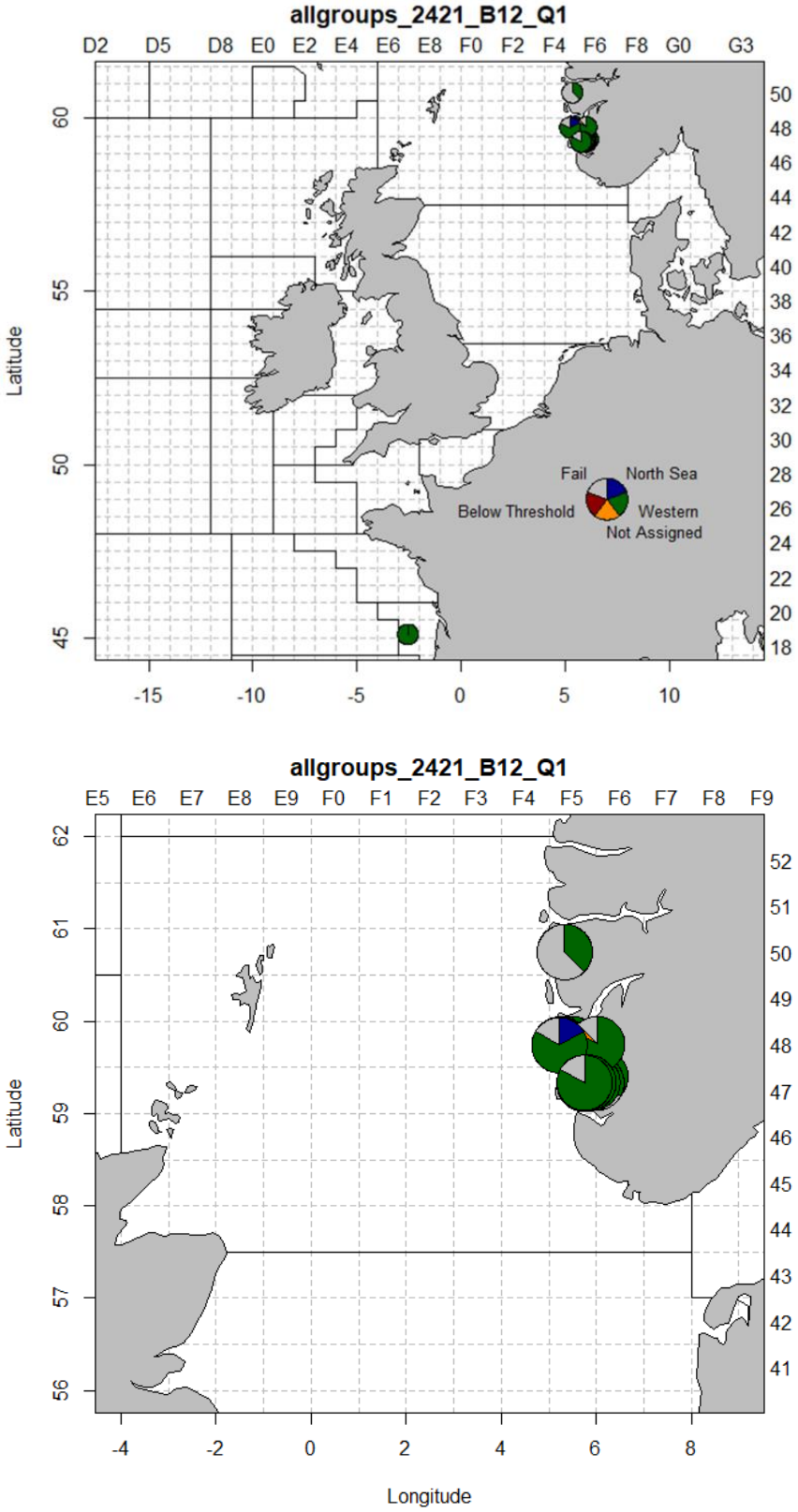


Annex 6 Figure 3. The output of the combined assignments with the *25_SNP* panel of the mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5.

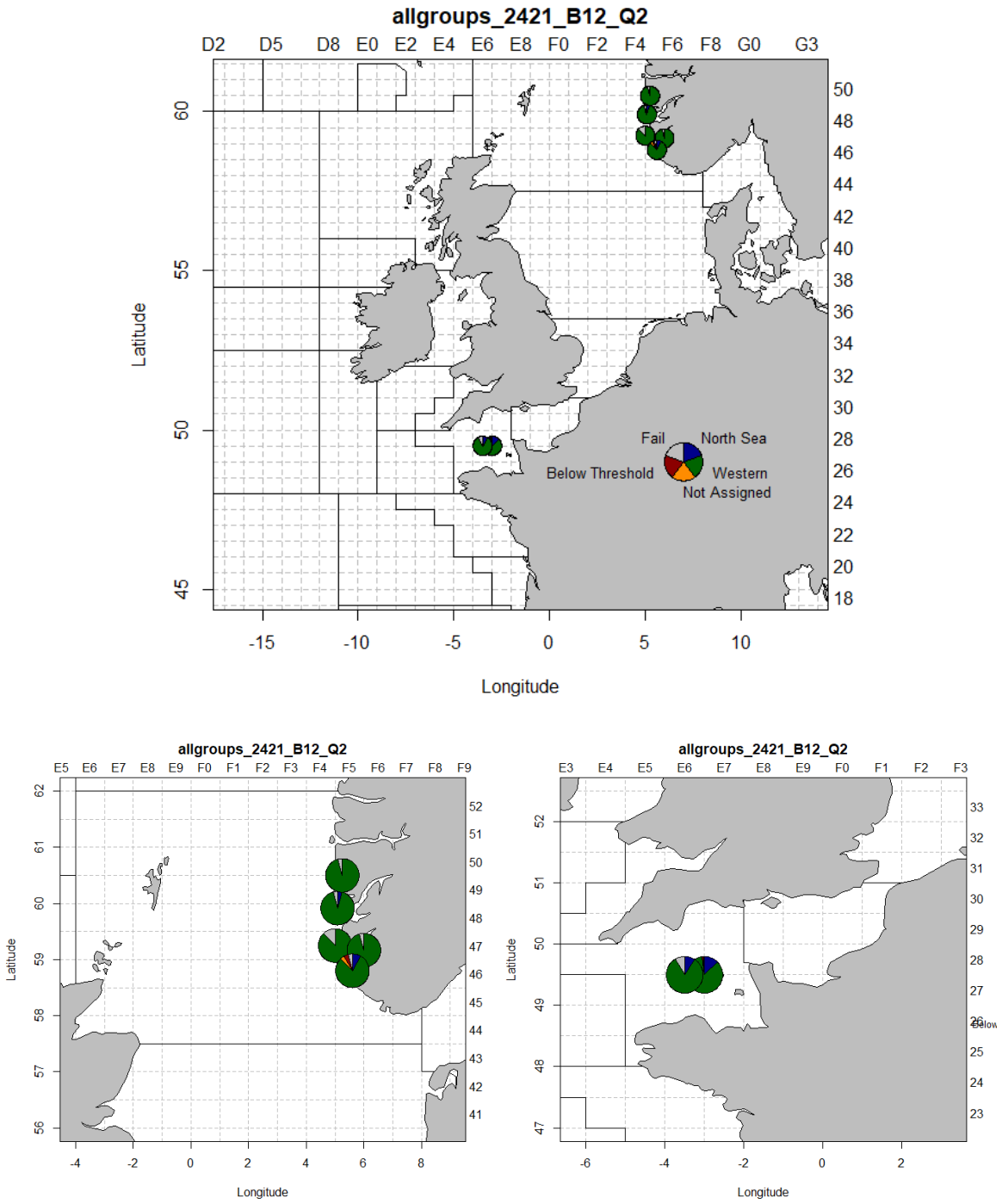


Annex 6 Figure 4. The output of the combined assignments with the *36_SNP* panel of the mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5.

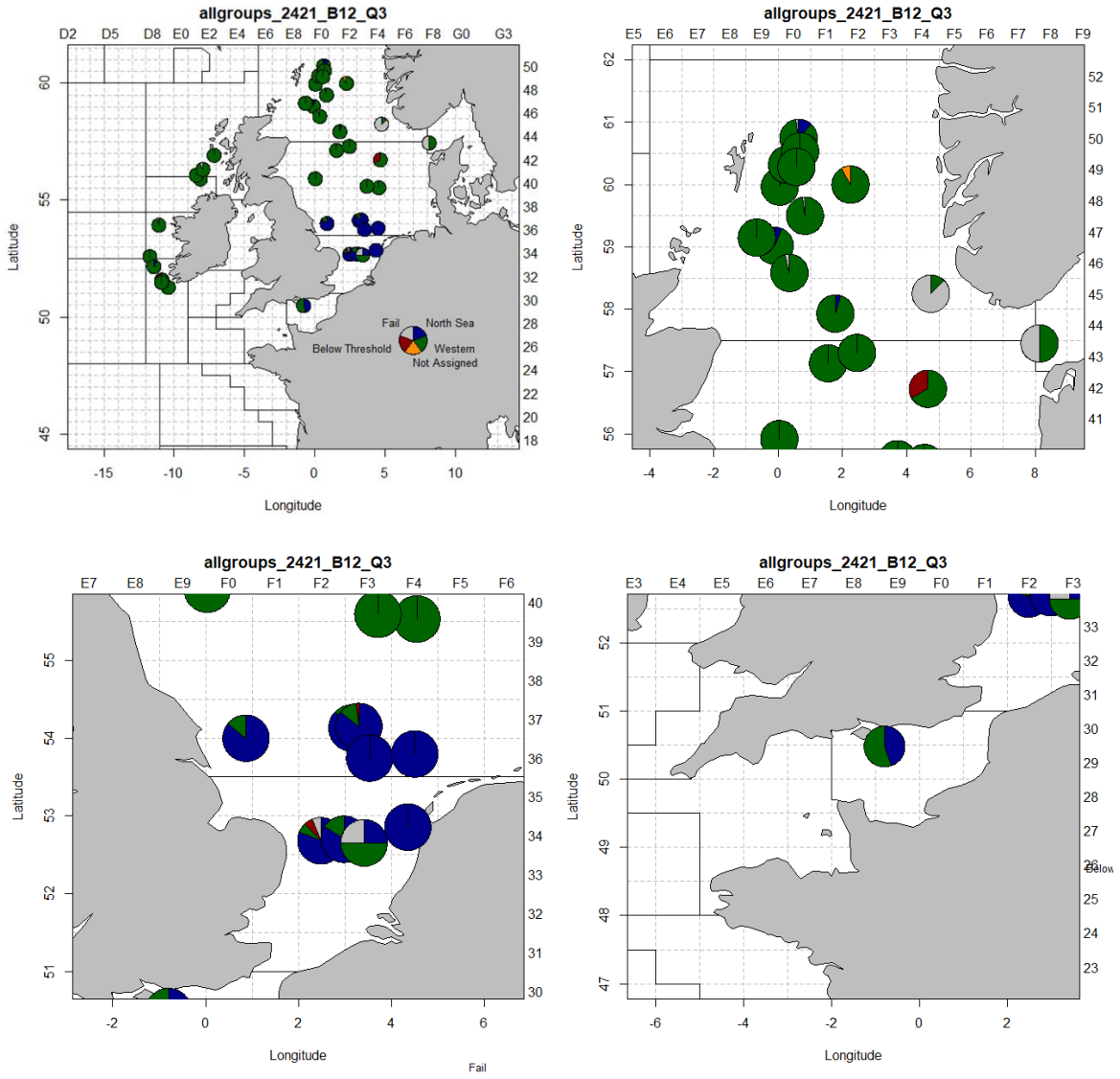
In order to help with the interpretation of the assignment results with the *2421_SNP* panel were also split by quarter.



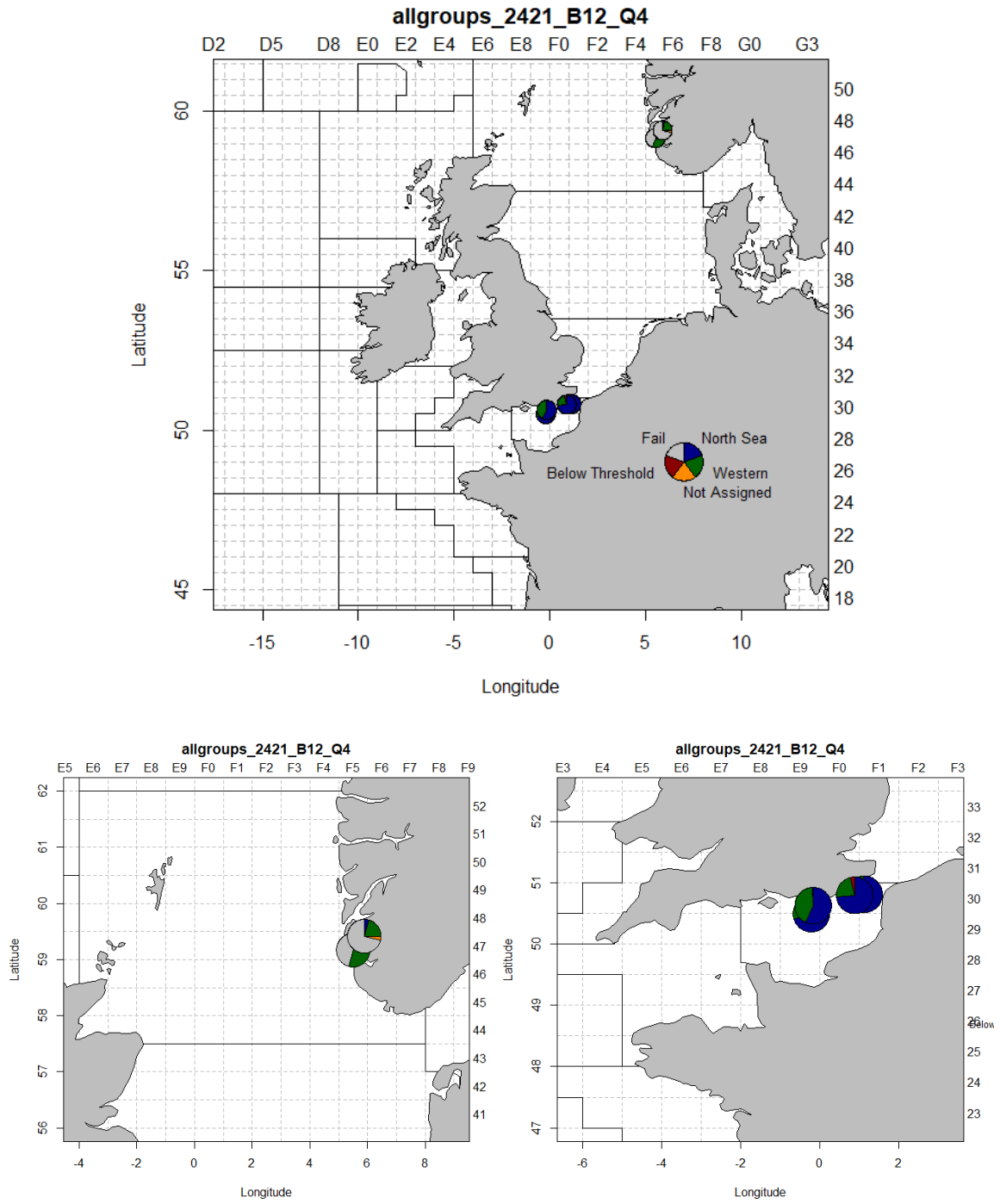
Annex 6 Figure 5. The output of the combined assignments with the *2421_SNP* panel of the quarter 1 mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5. Top = all quarter 1, bottom panel = division 4.a only



Annex 6 Figure 6. The output of the combined assignments with the *2421_SNP* panel of the quarter 2 mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5.



Annex 6 Figure 7. The output of the combined assignments with the *2421_SNP* panel of the quarter 3 mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5.



Annex 6 Figure 7. The output of the combined assignments with the *2421_SNP* panel of the quarter 4 mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5.

Annex 6 Table 1. The output of the combined assignments with the *2421_SNP* panel of the mixed samples assigned in Section 4.6 of the main document and the samples assigned in Annex 5.

Sample number	Date	Quarter	Catch Location	Lat	Lon	ICES Areas	Assigned Western	Assigned North Sea	Below Threshold	Not Assigned
B2_37	08/03/2015	1	Bay of Biscay	45.09	-2.56	8.b	48	0	0	0
17	08/03/2019	1	Norwegian coast	59.75	5.50	4.a	24	0	0	0
B2_14	03/02/2022	1	Norwegian Coast	60.75	5.33	4.a	9	0	0	0
B2_15	03/01/2023	1	Norwegian Coast	59.75	5.25	4.a	19	0	0	0
B2_16	03/01/2023	1	Norwegian Coast	59.40	6.08	4.a	19	0	0	1
B2_17	05/01/2023	1	Norwegian Coast	59.33	6.00	4.a	8	0	0	0
B2_18	18/01/2023	1	Norwegian Coast	59.33	5.92	4.a	15	0	0	0
B2_20	19/01/2023	1	Norwegian Coast	59.32	5.75	4.a	21	0	0	0
B2_19	20/01/2023	1	Norwegian Coast	59.75	6.02	4.a	20	0	0	1
B2_22	20/01/2023	1	Norwegian Coast	59.74	5.22	4.a	16	4	0	0
B2_23	20/01/2023	1	Norwegian Coast	59.33	5.75	4.a	20	0	0	0
B2_21	24/01/2023	1	Norwegian Coast	59.33	5.83	4.a	23	0	0	0
34	25/05/2017	2	Western Channel	49.50	-3.00	7.e	82	12	1	0
35	25/05/2017	2	Western Channel	49.50	-3.50	7.e	79	9	0	0
18	10/04/2019	2	Norwegian coast	59.92	5.08	4.a	22	1	0	0
19	23/05/2019	2	Norwegian coast	60.50	5.25	4.a	23	0	0	0
20	27/04/2022	2	Norwegian coast	59.25	5.00	4.a	3	0	0	0
25	27/04/2022	2	Norwegian coast	59.25	5.00	4.a	21	0	0	0
26	02/05/2022	2	Norwegian coast	59.17	6.00	4.a	23	0	0	0
27	24/05/2022	2	Norwegian coast	59.17	6.00	4.a	19	2	1	1
13a	20/07/2016	3	Northern North Sea	59.97	0.05	4.a	20	0	0	0
13b	20/07/2016	3	Northern North Sea	60.27	0.70	4.a	12	0	0	0
13c	20/07/2016	3	Northern North Sea	60.02	0.05	4.a	62	2	0	0
PTB_3	08/09/2016	3	Central North Sea	54.15	3.30	4.b	12	82	2	0
14a	01/07/2017	3	Northern North Sea	57.93	1.78	4.a	45	2	0	0
6c	07/07/2017	3	West Ireland	53.93	-11.09	7.b	33	0	0	0
30a	11/07/2017	3	Central North Sea	54.15	3.22	4.b	0	12	0	0
30b	12/07/2017	3	Southern North Sea	52.88	3.02	4.c	2	9	0	0
30c	13/07/2017	3	Southern North Sea	52.70	2.98	4.c	2	12	0	0
3e	17/07/2017	3	Southern North Sea	52.68	2.48	4.c	1	13	1	0
30d	18/07/2017	3	Central North Sea	54.13	3.15	4.b	3	8	0	0
B2_28	19/07/2017	3	West of Scotland	56.92	-7.16	6.a	48	0	0	0
14b	20/07/2017	3	Northern North Sea	60.32	0.28	4.a	48	1	0	0
15	28/07/2017	3	Northern North Sea	59.02	-0.11	4.a	44	3	1	0
16	09/08/2017	3	Northern North Sea	59.50	0.83	4.a	47	0	0	0
29	30/08/2017	3	Skagerrak	57.45	8.13	3.a	11	0	0	0
21a	18/07/2020	3	Northern North Sea	60.43	-0.13	4.a	1	0	0	0
21b	20/07/2020	3	Northern North Sea	60.70	0.87	4.a	0	1	0	0
21c	21/07/2020	3	Northern North Sea	60.75	0.63	4.a	1	0	0	0
21d	21/07/2020	3	Northern North Sea	60.78	1.03	4.a	1	0	0	0
21e	21/07/2020	3	Northern North Sea	60.88	1.25	4.a	3	0	0	0
21f	22/07/2020	3	Northern North Sea	60.82	0.93	4.a	2	0	0	0
21g	22/07/2020	3	Northern North Sea	60.80	0.97	4.a	1	1	0	0
21h	23/07/2020	3	Northern North Sea	60.77	0.63	4.a	8	1	0	0
21i	23/07/2020	3	Northern North Sea	60.57	0.32	4.a	2	0	0	0
21j	24/07/2020	3	Northern North Sea	60.63	0.38	4.a	2	0	0	0
21k	24/07/2020	3	Northern North Sea	60.67	0.47	4.a	6	0	0	0
21l	24/07/2020	3	Northern North Sea	60.48	0.33	4.a	5	1	0	0
21m	24/07/2020	3	Northern North Sea	60.52	0.43	4.a	5	1	0	0
21n	25/07/2020	3	Northern North Sea	60.38	0.97	4.a	4	1	0	0
22a	17/08/2020	3	Northern North Sea	59.97	0.18	4.a	11	0	0	0
22b	18/08/2020	3	Northern North Sea	58.58	0.35	4.a	4	0	0	0
22c	18/08/2020	3	Northern North Sea	58.60	0.37	4.a	4	0	0	0
22d	18/08/2020	3	Northern North Sea	58.75	0.35	4.a	5	0	0	0
22e	19/08/2020	3	Northern North Sea	58.38	0.48	4.a	5	0	0	0
22f	20/08/2020	3	Northern North Sea	58.87	-0.83	4.a	6	0	0	0
22g	21/08/2020	3	Northern North Sea	58.68	-0.70	4.a	5	0	0	0
22h	21/08/2020	3	Northern North Sea	58.52	-0.75	4.a	2	0	0	0
22i	22/08/2020	3	Northern North Sea	58.35	-0.65	4.a	3	0	0	0
22j	22/08/2020	3	Northern North Sea	58.42	-0.62	4.a	1	0	0	0
31	21/09/2020	3	Eastern Channel	50.48	-0.57	7.d	53	43	0	0
B2_33	10/07/2022	3	Southwest Ireland	51.27	-10.42	7.j	16	0	0	0
B2_34	10/07/2022	3	Southwest Ireland	51.58	-10.88	7.j	11	0	2	0
B2_32	11/07/2022	3	West Ireland	52.58	-11.73	7.b	18	0	0	0
B2_35	11/07/2022	3	Southwest Ireland	52.15	-11.45	7.j	11	1	0	0
B2_36	12/07/2022	3	Southwest Ireland	51.48	-10.90	7.j	9	0	0	0

B2_24	15/07/2022	3	Northern North Sea	60.28	0.57	4.a	2	1	0	0
B2_27	15/07/2022	3	Northern North Sea	60.28	0.57	4.a	10	0	0	0
B2_25	19/07/2022	3	Northern North Sea	60.53	0.68	4.a	12	0	0	0
B2_26	23/07/2022	3	Northern North Sea	59.15	-0.67	4.a	3	0	0	0
28	07/09/2022	3	Norwegian coast	60.00	2.25	4.a	11	0	0	1
B2_29	09/07/2023	3	Northwest Ireland	55.90	-8.14	6.a	10	0	0	0
B2_30	10/07/2023	3	Northwest Ireland	56.07	-8.43	6.a	1	0	0	0
B2_31	11/07/2023	3	Northwest Ireland	56.32	-7.96	6.a	26	0	0	0
B2_10	09/08/2023	3	Southern North Sea	52.86	4.35	4.c	0	5	0	0
B2_11	09/08/2023	3	Southern North Sea	52.65	3.41	4.c	2	1	0	0
B2_1	10/08/2023	3	Central North Sea	53.74	3.53	4.b	0	1	0	0
B2_2	10/08/2023	3	Central North Sea	53.80	4.50	4.b	0	2	0	0
B2_3	13/08/2023	3	Central North Sea	55.53	4.55	4.b	1	0	0	0
B2_4	13/08/2023	3	Central North Sea	55.59	3.71	4.b	3	0	0	0
B2_5	14/08/2023	3	Central North Sea	54.00	0.87	4.b	1	6	0	0
B2_6	16/08/2023	3	Central North Sea	55.92	0.03	4.b	13	0	0	0
B2_7	21/08/2023	3	Central North Sea	57.13	1.55	4.b	10	0	0	0
B2_8	21/08/2023	3	Central North Sea	57.30	2.45	4.b	10	0	0	0
B2_9	22/08/2023	3	Central North Sea	56.72	4.66	4.b	2	0	1	0
B2_12	20/10/2020	4	Eastern Channel	50.63	-0.17	7.d	40	54	0	0
32	13/11/2020	4	Eastern Channel	50.82	1.12	7.d	21	75	0	0
B2_13	27/11/2020	4	Eastern Channel	50.80	0.88	7.d	11	36	2	0
33	07/12/2020	4	Eastern Channel	50.50	-0.22	7.d	26	67	1	0
23	05/11/2021	4	Norwegian coast	59.17	5.50	4.a	13	0	0	0
24	16/11/2021	4	Norwegian coast	59.42	5.88	4.a	5	1	0	1