

**SNPs reveal geographical population structure of *Corallina officinalis*
(Corallinaceae, Rhodophyta)**

Chris Yesson¹, Amy Jackson², Steve Russell², Christopher J. Williamson^{2,3} and Juliet Brodie²

¹ Institute of Zoology, Zoological Society of London, London, UK

² Natural History Museum, Department of Life Sciences, London, UK

³ Schools of Biological and Geographical Sciences, University of Bristol, Bristol, UK

CONTACT: Chris Yesson. Email: chris.yesson@ioz.ac.uk

Abstract

We present the first population genetics study of the calcifying coralline alga and ecosystem engineer *Corallina officinalis*. Eleven novel SNP markers were developed and tested using Kompetitive Allele Specific PCR (KASP) genotyping to assess the population structure based on five sites around the NE Atlantic (Iceland, three UK sites and Spain), spanning a wide latitudinal range of the species' distribution. We examined population genetic patterns over the region using discriminate analysis of principal components (DAPC). All populations showed significant genetic differentiation, with a marginally insignificant pattern of isolation by distance (IBD) identified. The Icelandic population was most isolated, but still had genotypes in common with the population in Spain. The SNP markers presented here provide useful tools to assess the population connectivity of *C. officinalis*. This study is amongst the first to use SNPs on macroalgae and represents a significant step towards understanding the population structure of a widespread, habitat forming coralline alga in the NE Atlantic.

KEYWORDS Marine red alga; Population genetics; Calcifying macroalga; Corallinales; SNPs; Corallina

Introduction

Corallina officinalis is a calcified geniculate (i.e. articulated) coralline alga that is widespread on rocky shores in the North Atlantic (Guiry & Guiry, 2017; Brodie *et al.*, 2013; Williamson *et al.*, 2016). In the NE Atlantic the species is distributed from southern Greenland, Iceland and northern Norway, in the north, to northern Spain and the Azores in the south (Williamson *et al.*, 2015; Pardo *et al.*, 2015). *C. officinalis* is recognized as an important ecological component of rocky shores: it can create dense turfs, which are habitats for many small invertebrates, providing shelter in highly dynamic intertidal habitats, and a substratum for the settlement of macro- and microalgae (Nelson, 2009). Moreover, *C. officinalis*, like other calcifying algae, can contribute to carbon dioxide fluxes within the ocean through production and dissolution of calcium carbonate, and have an important role in the carbon cycle of coastal marine ecosystems (van der Heijden & Kamenos, 2015).

Despite their ecological significance, calcified seaweeds are at risk from a combination of both local, e.g. sedimentation, eutrophication, change in freshwater flows, and global perturbations, e.g. climate change and ocean acidification. Rising sea temperatures driven by climate change are projected to result in significant range shifts of macroalgal species, with extinctions at lower latitudes and colonization of higher latitudes (Brodie *et al.*, 2014), which may in turn affect inter-specific competition (Kroeker *et al.*, 2010). Ocean acidification, i.e. decreasing ocean pH and carbonate saturation, will have a substantial impact on calcifying organisms, including habitat-forming macroalgae (Koch *et al.*, 2013). Less alkaline waters will lead to corrosion of calcium carbonate skeletons and increase the metabolic costs of calcification (Nelson, 2009). The

North Atlantic is predicted to see a significant reduction of calcifying algae by 2100 under such conditions, which could have dramatic consequences for local ecosystem functioning (Brodie *et al.*, 2014).

The fate of organisms in our rapidly changing marine ecosystems will depend on their genetic diversity and population connectivity, and whether gene flow is great enough to counteract the possibility of local adaptation (Valero *et al.*, 2001). Information on genetic connectivity in marine macroalgae remains sparse (Li *et al.*, 2016a), yet identifying and conserving hotspots of tolerant genotypes will assist in the management of these important habitats in light of future climatic changes. Population genetic tools are vital for the identification of locally adapted genotypes, and an understanding of the connectivity of populations is key for conservation planning (Pauls *et al.*, 2013).

Population genetic studies in the red algae have focussed on a limited number of species, using predominantly DNA sequence regions rather than traditional population genetic markers (Li *et al.*, 2016a). The use of highly variable markers that are the foundation for traditional population genetic studies is rare for red algae, and published studies have focussed on microsatellites, i.e. short sequence repeat markers (Hu *et al.*, 2010; Couceiro *et al.*, 2011; Kostamo *et al.*, 2012; Song *et al.*, 2013; Wang *et al.*, 2013). *Chondrus crispus*, for example, has been studied using microsatellite markers (Krueger-Hadfield *et al.*, 2011) and single nucleotide polymorphisms (SNPs) (Provan *et al.*, 2013), and SNP data were generated for *Furcellaria lumbricalis* from different locations and salinity conditions to see whether the information was congruent with microsatellites (Olsson & Korpelainen, 2013).

Currently, SNPs have emerged as the marker of choice for population genetic studies because, amongst many other properties, they have codominant inheritance and

there are potentially thousands of loci available for analysis (Seeb *et al.*, 2011; Provan *et al.*, 2013). Application of SNPs to macroalgae is rare, and predominantly still in development, for example, the construction of a high density SNP linkage map for the kelp *Saccharina japonica* (Zhang *et al.*, 2015), and the report of SNPs found by examining the plastomes of three species of the red algal genus *Membranoptera* (Hughes *et al.*, 2017). However, SNP markers have been developed, tested and compared to microsatellite alternatives for the red alga *Chondrus crispus* (Provan *et al.*, 2013). The six SNPs used to genotype this species from the UK and Ireland proved effective for analysis of population patterns.

To date, there have been few genetic studies of *Corallina officinalis*, these have examined: the utility of cox 1 region for DNA barcoding (Robba *et al.*, 2006); phylogenetic analysis of *C. officinalis* and related species (Hind & Saunders, 2013; Walker *et al.*, 2009; Williamson *et al.*, 2015); taxonomy (Brodie *et al.*, 2013; Hind *et al.*, 2014); and mitogenomics (Williamson *et al.*, 2016); but none have analysed population genetic patterns. Indeed, the first microsatellite marker for any coralline alga was only recently reported for the maerl-forming crustose species *Phymatolithon calcareum* (Pardo *et al.*, 2014). The development of next generation sequencing technologies has led to an increased focus on high throughput sequencing and publication of genome-level datasets for a number of macroalgal species (DePriest *et al.*, 2014; Kim *et al.*, 2015; Bi *et al.*, 2016; Williamson *et al.*, 2016). Such data have the potential to address many questions relating to taxonomy, phylogeny and evolutionary history in algal genetics (Kim *et al.*, 2014). The recent study documenting a mitochondrial genome for *Corallina officinalis* (Williamson *et al.*, 2016) involved the acquisition of whole genome shotgun sequence data for specimens from Iceland, UK

and Spain. This has created the potential to use these sequence data for the identification of highly variable markers suitable for population genetic analysis.

Here we report on the development of the first SNP markers for a calcifying red alga, and their use in examining the underlying population structure of *Corallina officinalis* based on five populations from a wide latitudinal range in the NE Atlantic.

Materials and Methods

Sampling and DNA extraction

Samples of *Corallina officinalis* were collected from three sites in the UK, one in northern Spain and one in southern Iceland (Table 1). 31-48 samples were collected from each site. Each sample was given a unique identifier and was split into two parts, one of which was dried in silica beads and the other deposited in the BM herbarium. DNA was extracted from approximately 0.5 cm² of dried material using a modified CTAB microextraction protocol (Robba *et al.*, 2006). This method differs from the Doyle & Doyle (1990) protocol: following the CTAB digestion and chloroform/isoamyl centrifugation step, the aqueous supernatant phase is purified and concentrated using the column-based, Illustra GFX PCR DNA and Gel Band Purification kit (GE Healthcare UK Ltd) according to the manufacturer's instructions and eluted from the column in a final volume of 50 µl of Tris buffer (0.1 M Tris-HCl, pH 8.0). DNA concentration was assessed using a Nanodrop 8000 spectrophotometer (ThermoFisher) to ensure sufficient yield of genomic DNA for genotyping (minimum required yield 150 ng).

SNPs

Potential SNP targets were developed by examining data from shotgun sequence reads from the Illumina Miseq platform (<https://www.illumina.com/systems/sequencing-platforms/miseq.html>). Four samples were independently sequenced on the Miseq: one specimen from Iceland, one from Spain, and two from the UK (one each from North Devon and South Devon; Williamson *et al.*, 2016). Miseq reads were cleaned by trimming the start and end of the reads between 66-236 bp. A minimum of 8.6 million

and maximum of 14.5 paired-end reads were obtained for each sample. Sequence quality was assessed using FastQC (v 0.9; Babraham Bioinformatics, Cambridge, UK), and those with mean quality scores below 20 were removed, as were those with trimmed length < 100 bp. Reads matching the mitochondrial (Williamson *et al.*, 2016) and draft chloroplast genomes were removed, leaving only nuclear reads for examination.

The SISRS (Site Identification from Short Read Sequence) package was used to identify potential SNP loci with variable sites from the sequence reads. The SISRS process generates a 'composite genome' (essentially a rapidly assembled and highly fragmented draft genome, based on a subset of reads selected to obtain 10x coverage), which is used as a reference for alignment of shotgun reads for identification of informative variable regions (Schwartz *et al.*, 2015). We assumed a genome size of 105,000,000 bp based on the published genome of the red alga *Chondrus crispus* (Collén *et al.*, 2013), to allow the SISRS process to estimate 10x coverage for the draft genome. The SISRS process identifies putative SNPs on the fragmented draft genome. Potential SNPs were filtered based on a number of control criteria: the genome fragment must have at least 50 bp either side of the variable locus (to allow primer development); a minimum of 5x read coverage, including reads from samples sourced from each country; and the SNP is not near other potential SNPs (either no other SNP was present on the genome fragment, or at least 1000 bp between potential SNPs). Putative SNPs passing these filters were Blast matched to the NCBI nucleotide database (<https://blast.ncbi.nlm.nih.gov/>) and any producing positive matches to other taxa were discarded, on the assumption that these are contaminants. A final subset of 12 SNPs was selected for genotyping by selecting those with the highest read coverage but one

SNP failed amplification for the majority of samples, leaving 11 SNPs for analysis.

Genotyping

Genotyping was performed using a KASP (Kompetitive Allele Specific PCR) assay (Semagn *et al.*, 2014). DNA samples were eluted in 10 mM Tris buffer and placed into two 96-well plates with two control wells, at concentrations of over 5 ng μl^{-1} at 40 μl volumes and sent to LGC genomics (www.lgcgenomics.com/genotyping/kasp-genotyping-chemistry) for genotyping. Results were delivered as a bi-allelic scoring for each of the 11 SNPs.

Statistical analyses

All statistical analyses were performed in the statistics software R version 3.1.3 (<https://cran.r-project.org/>). An assessment of Linkage Disequilibrium (LD) between SNP markers was performed using the LD function of the R package “genetics” (Warnes & Leisch, 2013), using Bonferroni correction of *p*-values to assess significance. Summary statistics of observed and expected heterozygosity (H_o and H_e) were calculated using the function `HWE.test` (R package genetics).

Structure between populations was assessed using a pairwise *F_{st}* (Weir & Cockerham, 1984), with significance measured based on 10,000 permutations using the `boot.pfst` function in the R package `heirfstat` (Goudet & Jombart, 2015). Isolation By Distance (IBD) was tested using a Mantel test, based on genetic distances measured by *F_{st}* and geographic distance measured as distance between sites over water (approximated with the distance tool in the GIS package Quantum GIS). The Mantel test was performed using the `mantel.randtest` function in the R package “ade4” (Dray & Dufour, 2007) using 10,000 replicates.

An assessment of genetic patterns in the data was performed using a Discriminant Analyses of Principal Component (DAPC) using the R package “adegenet” (Jombart, 2008). The number of genetic clusters was investigated using the `find.cluster` function of `adegenet`, which runs successive K-means for clustering and assesses the optimal K by reference to the Bayesian information criterion. The automatic cluster selection procedure “diffNgroup” was used with `n.iter` set to 10^7 and `n.start` set to 10^4 , and principal components to retain for analysis were selected based on a percentage of variance explained threshold (95%). The ordination (DAPC) analysis was performed using the `dapc` function. The optimal number of principal components to retain was assessed using the `optim.a.score` function. An assessment of sample assignment to genetic cluster was performed using the `compoplot` function.

A power analysis was performed to test the ability of similar-sized datasets to produce significant results. An estimate of effective population size (N_e) was performed using `NeEstimator V2` (Do *et al.*, 2014). The `PowSim` program (Ryman & Palm, 2006) was run using the estimate of population size ($N_e=20$), 100 simulations and 25 generations of drift for both the complete dataset (11 SNP markers) and the unlinked marker set (5 SNPs). `PowSim` tests for significant F_{st} results from simulated datasets of a given size to determine whether a dataset of that size is sufficient to consistently produce significant results (Ryman & Palm, 2006).

Results

265 samples from five populations (Table 1) were collected and genotyped. The 11 SNP markers analysed are presented in Table 2, along with observed and expected heterozygosity these markers. All except one of the 11 markers (Coff4) showed significant deviation from Hardy Weinberg Equilibrium (HWE). Significant linkage (LD) was found between 19 out of 55 pairs of markers. The largest subset of unlinked markers contains 5 SNPs, with 3 combinations of 5 unlinked markers, set 1: 1, 3, 4, 5, 6; set 2: 1, 3, 6, 7, 10; set 3: 3, 7, 8, 10, 11.

Analyses were conducted on the complete dataset of all SNPs for samples with at least 9 non-null SNPs (N=208, minimum population sample: 31). Additionally, a parallel analysis was performed on a subset of these data including only the 5 unlinked SNPs with the fewest null alleles (set 3 above). The power analysis indicated that datasets of both 5 and 11 SNPs are sufficient to produce significant *Fst* results (100/100 of simulations for both datasets produced significant results, $p < 0.05$). There was significant genetic differentiation between all populations (Table 3, supplementary table S1 for unlinked data). The greatest genetic distance was observed between the Iceland and UK (Kent) populations (for both the complete dataset and unlinked data), although the greatest geographic distance was between the Iceland and Spain sites (Iceland to Spain, 2,500 km; Iceland to UK, Kent, 2,000 km). A marginally insignificant pattern of isolation by distance was shown in the data set (complete dataset: 0.68, $p = 0.075$), reflecting a weak pattern that geographically distant sites showed greater genetic divergence. Analysing just the five unlinked markers produced a similar result (unlinked data: 0.77, $p = 0.084$ – see supplementary figure 1 for a comparison with the complete dataset).

Cluster analysis selected four genetic clusters which appeared to have a degree of site-specificity (Table 3). All Iceland samples were placed in a single genetic cluster (cluster 1) and 81% of the Spanish population were placed in Cluster 3. However, no genetic cluster was entirely site-specific (e.g. a single sample from Spain was in the 'Icelandic cluster', see Table 3, Fig. 1). The greatest genetic similarity was observed between the UK populations. There were two predominantly UK-based genetic clusters, an eastern cluster (4), in which the majority of Kent samples occurred, and a western cluster (2), which predominated in the Devon sites and was not observed in Kent. There was a broadly similar pattern of clustering based on just unlinked markers, with a predominantly Icelandic cluster, an East/West split in the UK but less isolation observed for the Spanish population (Supplementary Table 1).

The DAPC used three principal components and 2 discriminant functions, and the proportion of conserved variance was 55% (Figure 2). The scatter plot (Fig. 2) shows the relative isolation of the Iceland population (cluster 1), and the closer affinity of the Thanet (Kent) and Combe Martin (North Devon) populations which both contained a high proportion of samples from cluster 4. Cluster assignment probabilities were typically high (163/208 higher than 90%, Fig. 3). There were several samples from both Iceland and Spain with affinities to both the Icelandic and Spanish genetic clusters, indicating that even these distant locations showed some connection.

Discussion

In this study, SNPs were developed for first time for a coralline red alga. The application of these markers revealed significant genetic structuring within *C. officinalis* populations in the NE Atlantic. These markers are a useful tool to analyse genetic connectivity of this important habitat-forming coralline red algae. The marginally insignificant isolation by distance result is somewhat inconclusive, and may be resolved by greater sampling. Findings for other (non-invasive) red algal species, have reported significant isolation by distance over a variety of spatial scales, such as *Gelidium canariense* in the Canary Isles (Bouza *et al.*, 2006), *Chondrus crispus* around the UK and Ireland (Provan *et al.*, 2013), and *Ahnfeltiopsis pusilla* in N Spain (Couceiro *et al.*, 2011).

There is genetic connectivity between even the most geographically isolated populations of *C. officinalis*, with the Icelandic population containing an individual from the “Spanish” genetic cluster. In addition, each of the reported genetic clusters is present in at least two sites. The relatively low diversity seen in SW Iceland (a single genetic cluster present), agrees with findings for the red algae *Palmaria palmata* and *Chondrus crispus* of relatively low genetic diversity in SW Iceland in comparison to other areas of the NE Atlantic (Li *et al.*, 2015; Li *et al.*, 2016a). However, these examples show greater connectivity of Icelandic populations than we found for *C. officinalis*, with Icelandic *P. palmata* sharing genotypes with the West Ireland but not SW England, while for *C. crispus* the Icelandic genotype was found in a variety of locations in southern England from the Cornish coast to the North Sea (Li *et al.*, 2016a). However, it should be noted that these studies are based on mitochondrial, plastid and nuclear

sequence regions, which will be less variable than SNP markers.

In contrast to the findings reported here, in a study of the large brown intertidal macroalga *Fucus serratus*, North Spain had the most isolated of NE Atlantic intertidal populations, but this study did not include populations from the UK or Iceland (Coyer *et al.*, 2003). Although 80% of our Spanish *C. officinalis* samples were found to be from a single genetic cluster, the Spanish site was the only location where all genetic clusters were found. Therefore the Spanish site could be viewed as a repository of genetic diversity, which supports findings that North Spain has relatively high macroalgal diversity and may have been a refugium during the last glacial maximum (Li *et al.*, 2016a). However, in *Chondrus crispus*, Iberian populations in NW Portugal are highly genetically isolated from northern European populations (Hu *et al.*, 2010).

There is significant genetic differentiation between every population (pairwise $F_{st} > 0$), even between the UK sites. This fits the expected pattern, given that genetic differentiation is reported at scales of 1-10 km for many seaweeds (Krueger-Hadfield *et al.*, 2011), and even the nearest UK sites sampled here are 400 km distant. Although there is a moderate east/west divide in the UK *C. officinalis* populations, the most distant UK populations in North Devon (Combe Martin) and Thanet, Kent share a majority of samples fitting the Eastern UK genetic cluster (4), while the intermediate site in South Devon has a majority of samples fitting the Western UK genetic cluster (2). The Kent population has relatively low diversity (e.g. 29/31 samples from Kent belong to the same genetic cluster), with only the Iceland population showing a lower diversity. Although there is no directly comparable study of red algae for these locations, the North Sea population of *Chondrus crispus* (the more geographically isolated Helgoland) showed relatively low diversity (Hu *et al.*, 2010). Furthermore,

there appears to be a closer relationship between North Devon and the North Sea for *Mastocarpus stellatus* than for south coast UK populations (Li *et al.*, 2016b).

There is significant linkage evidenced in the 11 markers tested in this study. The largest set of markers showing no linkage contained 5 SNPs. Linkage disequilibrium can be a problem for parental and sibling analysis (Huang *et al.*, 2004). However, tests comparing results from a subset of data including just unlinked markers produced similar results to analysis of the full dataset (see supplementary figure and table), so we have chosen to present the results based on the complete data. Our analysis also used the DAPC method which does not depend on a population genetics model and carries no assumptions of linkage equilibrium (Jombart *et al.*, 2010).

This study is a significant first step towards understanding the genetic structure of *C. officinalis*. However, it is noted that 11 SNPs studied here is a relatively small number of markers in comparison to some other SNP studies where thousands have been investigated (Seeb *et al.*, 2011). However, 11 SNPs with binary variation could potentially yield $2^{11} = 2048$ genotypes, which is an order of magnitude larger than the number of samples analysed. Additionally, this study shows that significant structure can be determined from a small number of markers, as has been found for *Chondrus crispus*, where just 6 SNPs were used to detect significant genetic structure around the UK (Provan *et al.*, 2013). This compares favourably to traditional sequencing studies, which might sequence thousands of bases but end up analysing a handful of variable sites (e.g. Li *et al.*, 2016b sequenced 541 bp of the ITS region for 329 individuals and found 14 variable sites). Similarly, microsatellite studies have shown similar numbers of discriminating characters when examining population genetics of red algae from the NE Atlantic (Kostamo *et al.*, 2012; Provan *et al.*, 2013).

These are the first SNP markers described for any coralline alga and only the third for the red algae. These markers are useful tools to assess the population connectivity of the important ecosystem engineer *C. officinalis*, showing that the most geographically isolated population from SW Iceland has the most genetic isolation, although even the most distant sites show some genetic connectivity. Calcifying algae are at risk from ocean acidification and climate change, and this study should form the foundation of a wider analysis, incorporating more samples from across the Northeast Atlantic to identify the most isolated and potentially at risk populations.

Acknowledgements

We would like to thank Ian Tittley, César Peteiro, Noemi Sanchez and Karl Gunnarsson for assistance with specimen collection. We also thank Chris Maggs, Cristina Pardo and one anonymous reviewer for their constructive suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was partly funded by the Natural History Museum Departmental Investment Fund.

ORCID

Chris Yesson: orcid.org/0000-0002-6731-4229, Juliet Brodie orcid.org/0000-0001-7622-2564

Author contributions

CY, JB, CW conceived the study; AJ, SR performed lab work; AJ, CY performed

analysis; All contributed to manuscript writing.

References

- Bi, G., Liu, G., Zhao, E. & Du, Q. (2016). Complete mitochondrial genome of a red calcified alga *Calliarthron tuberculosum* (Corallinales). *Mitochondrial DNA Part A*, **27**: 2554–2556.
- Bouza, N., Caujapé-Castells, J., González-Pérez, M.Á., & Sosa, P.A. (2006). Genetic structure of natural populations in the red algae *Gelidium canariense* (Gelidiales, Rhodophyta) investigated by random amplified polymorphic DNA (RAPD) markers. *Journal of Phycology*, **42**: 304-311.
- Brodie, J., Walker, R., Williamson, C. & Irvine, L.M. (2013) Epitypification and redescription of *Corallina officinalis* L., the type of the genus, and *C. elongata* Ellis & Solander (Corallinales, Rhodophyta). *Cryptogamie, Algologie*, **34**: 49-56.
- Brodie, J., Williamson, C.J., Smale, D.A., Kamenos, N.A., Mieszkowska, N., Santos, R., Cunliffe, M., *et al.* (2014). The future of the northeast Atlantic benthic flora in a high CO₂ world. *Ecology and Evolution*, **4**: 2787–2798.
- Collén, J. *et al.* (2013). Genome structure and metabolic features in the red seaweed *Chondrus crispus* shed light on evolution of the Archaeplastida. *Proceedings of the National Academy of Sciences*, **110**: 5247-5252.
- Couceiro, L., Maneiro, I., Ruiz, J.M. & Barreiro, R. (2011). Multiscale genetic structure of an endangered seaweed *Ahnfeltiopsis pusilla* (Rhodophyta): implications for its conservation. *Journal of Phycology*, **47**: 259–268.
- Coyer, J.A., Peters, A.F., Stam, W.T. & Olsen, J.L. (2003). Post-ice age recolonization

and differentiation of *Fucus serratus* L. (Phaeophyceae; Fucales) populations in Northern Europe. *Molecular Ecology*, **12**: 1817–1829.

DePriest, M.S., Bhattacharya, D. & López-Bautista, J.M. (2014). The mitochondrial genome of *Grateloupia taiwanensis* (Halymeniaceae, Rhodophyta) and comparative mitochondrial genomics of red algae. *The Biological Bulletin*, **227**: 191–200.

Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., & Ovenden, J.R. (2014). NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (N_e) from genetic data. *Molecular Ecology Resources*, **14**: 209-214.

Dray, S. & Dufour, A.B. (2007). The ade4 package: implementing the duality diagram for ecologists. *Journal of Statistical Software*, **22**: 1–20.

Goudet, J. & Jombart, T. (2015). *hierfstat: Estimation and Tests of Hierarchical F-Statistics*. <http://cran.r-project.org/package=hierfstat>.

Guiry, M.D. & Guiry, G.M. (2017). *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. <http://www.algaebase.org>; searched on 08 August 2017.

Hind, K.R., Gabrielson, P.W., Lindstrom, S.C., & Martone, P.T. (2014). Misleading morphologies and the importance of sequencing type specimens for resolving coralline taxonomy (Corallinales, Rhodophyta): *Pachyarthron cretaceum* is *Corallina officinalis*. *Journal of Phycology*, **50**: 760-764.

Hind, K.R., & Saunders, G.W. (2013). A molecular phylogenetic study of the tribe Corallineae (Corallinales, Rhodophyta) with an assessment of genus level

- taxonomic features and descriptions of novel genera. *Journal of Phycology*, **49**: 103-114.
- Hu, Z., Guiry, M.D., Critchley, A.T. & Duan, D. (2010). Phylogeographic patterns indicate transatlantic migration from Europe to North America in the red seaweed *Chondrus crispus* (Gigartinales, Rhodophyta). *Journal of Phycology*, **46**: 889–900.
- Huang, Q., Shete, S. & Amos, C.I. (2004). Ignoring linkage disequilibrium among tightly linked markers induces false-positive evidence of linkage for affected sib pair analysis. *The American Journal of Human Genetics*, **75**: 1106–1112.
- Hughey, J. R., Hommersand, M. H., Gabrielson, P. W., Miller, K. A. & Fuller, T. (2017). Analysis of the complete plastomes of three species of *Membranoptera* (Ceramiales, Rhodophyta) from Pacific North America. *Journal of Phycology*, **53**: 32–43.
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**: 1403–1405.
- Jombart, T., Devillard, S. & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics*, **11**: 94.
- Jueterbock, A., Kollias, S., Smolina, I., Fernandes, J.M.O., Coyer, J.A., Olsen, J.L. & Hoarau, G. (2014). Thermal stress resistance of the brown alga *Fucus serratus* along the North-Atlantic coast: Acclimatization potential to climate change. *Marine Genomics*, **13**: 27–36.
- Kim, K.M., Park, J.-H., Bhattacharya, D. & Yoon, H.S. (2014). Applications of next-

generation sequencing to unravelling the evolutionary history of algae.

International journal of systematic and evolutionary microbiology, **64**: 333–345.

- Kim, K.M., Yang, E.C., Kim, J.H., Nelson, W.A. & Yoon, H.S. (2015). Complete mitochondrial genome of a rhodolith, *Sporolithon durum* (Sporolithales, Rhodophyta). *Mitochondrial DNA*, **26**: 155–156.
- Kostamo, K., Korpelainen, H. & Olsson, S. (2012). Comparative study on the population genetics of the red algae *Furcellaria lumbricalis* occupying different salinity conditions. *Marine Biology*, **159**: 561–571.
- Kroeker, K.J., Kordas, R.L., Crim, R.N. & Singh, G.G. (2010). Meta-analysis reveals negative yet variable effects of ocean acidification on marine organisms. *Ecology Letters*, **13**: 1419–1434.
- Krueger-Hadfield, S.A., Collén, J., Daguin-Thiébaud, C. & Valero, M. (2011). Genetic Population Structure and Mating System in *Chondrus crispus* (Rhodophyta). *Journal of Phycology*, **47**: 440–450.
- Li, J.J., Hu, Z.M., & Duan, D.L. (2015). Genetic data from the red alga *Palmaria palmata* reveal a mid-Pleistocene deep genetic split in the North Atlantic. *Journal of Biogeography*, **42**: 902–913.
- Li, J.-J., Hu, Z.-M. & Duan, D.-L. (2016a). Survival in glacial refugia versus postglacial dispersal in the North Atlantic: The cases of red seaweeds. In *Seaweed Phylogeography* (Hu, Z.-M. & Fraser, C., editors), 309–330. Springer Netherlands, Dordrecht.
- Li, J.-J., Hu, Z.-M., Liu, R.-Y., Zhang, J., Liu, S.-L. & Duan, D.-L. (2016b). Phylogeographic surveys and apomictic genetic connectivity in the North Atlantic

red seaweed *Mastocarpus stellatus*. *Molecular Phylogenetics and Evolution*, **94**: 463–472.

Nelson, W.A. (2009). Calcified macroalgae - critical to coastal ecosystems and vulnerable to change: a review. *Marine and Freshwater Research*, **60**: 787-801.

Olsson, S., & Korpelainen, H. (2013). Single nucleotide polymorphisms found in the red alga *Furcellaria lumbricalis* (Gigartinales): new markers for population and conservation genetic analyses. *Aquatic Conservation: Marine and Freshwater Ecosystems*, **23**: 460-467.

Pardo, C., Peña, V., Bárbara, I., Valero, M. & Barreiro, R. (2014). Development and multiplexing of the first microsatellite markers in a coralline red alga (*Phymatolithon calcareum*, Rhodophyta). *Phycologia*, **53**: 474–479.

Pardo, C., Peña, V., Barreiro, R., & Bárbara, I. (2015). A molecular and morphological study of *Corallina* sensu lato (Corallinales, Rhodophyta) in the Atlantic Iberian Peninsula. *Cryptogamie, Algologie*, **36**: 31-54.

Pauls, S.U., Nowak, C., Bálint, M. & Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. *Molecular Ecology*, **22**: 925–946.

Provan, J., Glendinning, K., Kelly, R. & Maggs, C. A. (2012). Levels and patterns of population genetic diversity in the red seaweed *Chondrus crispus* (Florideophyceae): a direct comparison of single nucleotide polymorphisms and microsatellites. *Biological Journal of the Linnean Society*, **108**: 251–262.

Robba, L., Russell, S.J., Barker, G.L. & Brodie, J. (2006). Assessing the use of the mitochondrial cox1 marker for use in DNA barcoding of red algae (Rhodophyta).

American Journal of Botany, **93**: 1101–1108.

Ryman, N., & Palm, S. (2006). POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Resources*, **6**: 600-602.

Schwartz, R.S., Harkins, K.M., Stone, A.C. & Cartwright, R.A. (2015). A composite genome approach to identify phylogenetically informative data from next-generation sequencing. *BMC Bioinformatics*, **16**: 193.

Seeb, J.E., Carvalho, G., Hauser, L., Naish, K., Roberts, S. & Seeb, L.W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, **11**: 1–8.

Semagn, K., Babu, R., Hearne, S. & Olsen, M. (2014). Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding*, **33**: 1–14.

Song, S.-L., Lim, P.-E., Phang, S.-M., Lee, W.-W., Lewmanomont, K., Largo, D.B. & Han, N.A. (2013). Microsatellite markers from expressed sequence tags (ESTs) of seaweeds in differentiating various *Gracilaria* species. *Journal of Applied Phycology*, **25**: 839–846.

Valero, M., Engel, C., Billot, C., Kloareg, B., & Destombe, C. (2001). Concept and issues of population genetics in seaweeds. *Cahiers de Biologie Marine*, **42**: 53-62.

van der Heijden, L.H. & Kamenos, N.A. (2015). Reviews and syntheses: Calculating the global contribution of coralline algae to total carbon burial. *Biogeosciences*, **12**: 6429–6441.

- Wang, J., Peng, C., Liu, Z., Tang, Z. & Yang, G. (2013). Isolation and characterization of microsatellites of *Grateloupia filicina*. *Conservation Genetics Resources*, **5**: 763–766.
- Warnes, G. & Leisch, F. (2013). *Genetics: Population Genetics*. <http://cran.uvigo.es/web/packages/genetics/> (Accessed 28 March 2017).
- Weir, B. & Cockerham, C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, **38**: 1358-1370.
- Williamson, C.J., Brodie, J., Goss, B., Yallop, M., Lee, S. & Perkins, R. (2014). *Corallina* and *Ellisolandia* (Corallinales, Rhodophyta) photophysiology over daylight tidal emersion: interactions with irradiance, temperature and carbonate chemistry. *Marine Biology*, **161**: 2051–2068.
- Williamson, C.J., Walker, R.H., Robba, L., Yesson, C., Russell, S., Irvine, L.M. & Brodie, J. (2015). Toward resolution of species diversity and distribution in the calcified red algal genera *Corallina* and *Ellisolandia* (Corallinales, Rhodophyta). *Phycologia*, **54**: 2–11.
- Williamson, C.J., Yesson, C., Briscoe, A.G. & Brodie, J. (2016). Complete mitochondrial genome of the geniculate calcified red alga, *Corallina officinalis* (Corallinales, Rhodophyta). *Mitochondrial DNA Part B*, **1**: 326–327.
- Zhang, N., Zhang, L., Tao, Y., Guo, L., Sun, J., Li, X., Zhao, N., *et al.* (2015). Construction of a high density SNP linkage map of kelp (*Saccharina japonica*) by sequencing Taq I site associated DNA and mapping of a sex determining locus. *BMC Genomics*, **16**: 189.

Tables

Table 1 – Geographic locations of *Corallina officinalis* samples collected for this study.

Country	Location	Code	Latitude	Longitude	N
Iceland	Þorlákshöfn, Ölfus	I	63° 50' 54.5856" N	21° 21' 41.5512" W	45
UK	Thanet, Kent	K	51° 17' 20.9904" N	01° 22' 46.0956" E	31
	Combe Martin, Devon	C	51° 13' 00.2856" N	04° 01' 32.5812" W	42
	Wembury Point, Devon	W	50° 18' 44.8632" N	04° 04' 50.6424" W	42
Spain	Comillas, Cantabria	S	43° 23' 31.2216" N	04° 17' 27.6684" W	48

Table 2. SNP markers and genetic estimates for *Corallina officinalis*. Null signifies proportion of null alleles (N=265). Var is the possible bases for that SNP. Ho/He observed/expected heterozygosity. D/D'/r² are the Hardy Weinberg disequilibrium statistics (D is the raw difference in frequency between observed and expected heterozygotes, D' is D rescaled to the range -1,1, r is the correlation coefficient between two alleles) and the p value tests deviation from Hardy Weinberg equilibrium (test whether D=0).

Name	Sequence	Var.	Null	Ho	He	D	D'	r ²	p-value
Coff1	TATTGGAATTTAAAA[A/T]TTGTGACTCTGAAA	A/T	19.6%	0.859	0.772	-0.044	-2.532	0.147	1.75E-06
Coff2	ACCAAGGGCCCTGCT[G/A]CCGCCGACAATGCG	G/A	21.1%	0.761	0.500	-0.130	-0.547	0.272	1.87E-14
Coff3	GACAGTGTATAGGAG[C/T]TGTGCCGTATTGAA	C/T	17.0%	0.918	0.835	-0.042	-5.050	0.255	1.22E-08
Coff4	CAATTGACAGACTAA[A/T]GTACAAATCTAACG	A/T	24.5%	0.545	0.500	-0.022	-0.094	0.008	2.05E-01
Coff5	TGTGTAAGGTGATGA[C/T]CATCGTCGTCGAAC	C/T	21.1%	0.617	0.525	-0.046	-0.306	0.038	5.58E-03
Coff6	GCTGGCTACAAGACC[G/C]AGACAAAACAACGC	G/C	24.9%	0.749	0.619	-0.065	-0.989	0.116	3.90E-06
Coff7	ATCCCATCTAGGTCC[A/C]GTTTTTCATGTACAG	A/C	29.1%	0.691	0.558	-0.067	-0.614	0.091	5.69E-05
Coff8	ATGAAGAACGAAGTA[T/A]GTCACTATGCGTCT	T/A	3.8%	0.686	0.511	-0.088	-0.481	0.129	1.16E-08
Coff9	TTGATTAAGATAGA[G/A]TTTTTATTATTTA	G/A	14.3%	0.678	0.601	-0.039	-0.511	0.038	4.42E-03
Coff10	TTATTCTATGGAATG[C/T]CAAGGGCAATATCA	C/T	23.8%	0.292	0.513	0.111	0.455	0.207	7.41E-11
Coff11	AAGAGACACCGAAAC[A/T]TTCAAATTCGGAAG	A/T	16.6%	0.783	0.501	-0.141	-0.613	0.319	8.25E-18

Table 3. Regional genetic and geographic distances and assignment of samples to genetic clusters (all specimens are assigned to one of 4 genetic clusters). Lower triangle shows pairwise *Fst* values (all $p < 0.001$ except the WP/CM comparison for which $p < 0.01$). Upper triangle shows approximate over-water distances in thousands of kilometres. N = number of samples, 1-4 are genetic clusters.

Region	Code	Genetic cluster									
		I	K	C	W	S	N	1	2	3	4
þorlákshöfn, Ölfus	I	-	2	1.8	2	2.5	45	45	0	0	0
Thanet, Kent	K	0.49	-	0.8	0.4	1.1	31	0	0	2	29
Combe Martin, Devon	C	0.36	0.09	-	0.4	0.9	42	0	10	5	27
Wembury Point, Devon	W	0.36	0.24	0.09	-	0.8	42	0	34	3	5
Comillas, Cantabria	S	0.26	0.34	0.21	0.18	-	48	1	4	42	1

Figure Captions

Fig. 1. Location of sampling and spread of genetic clusters.

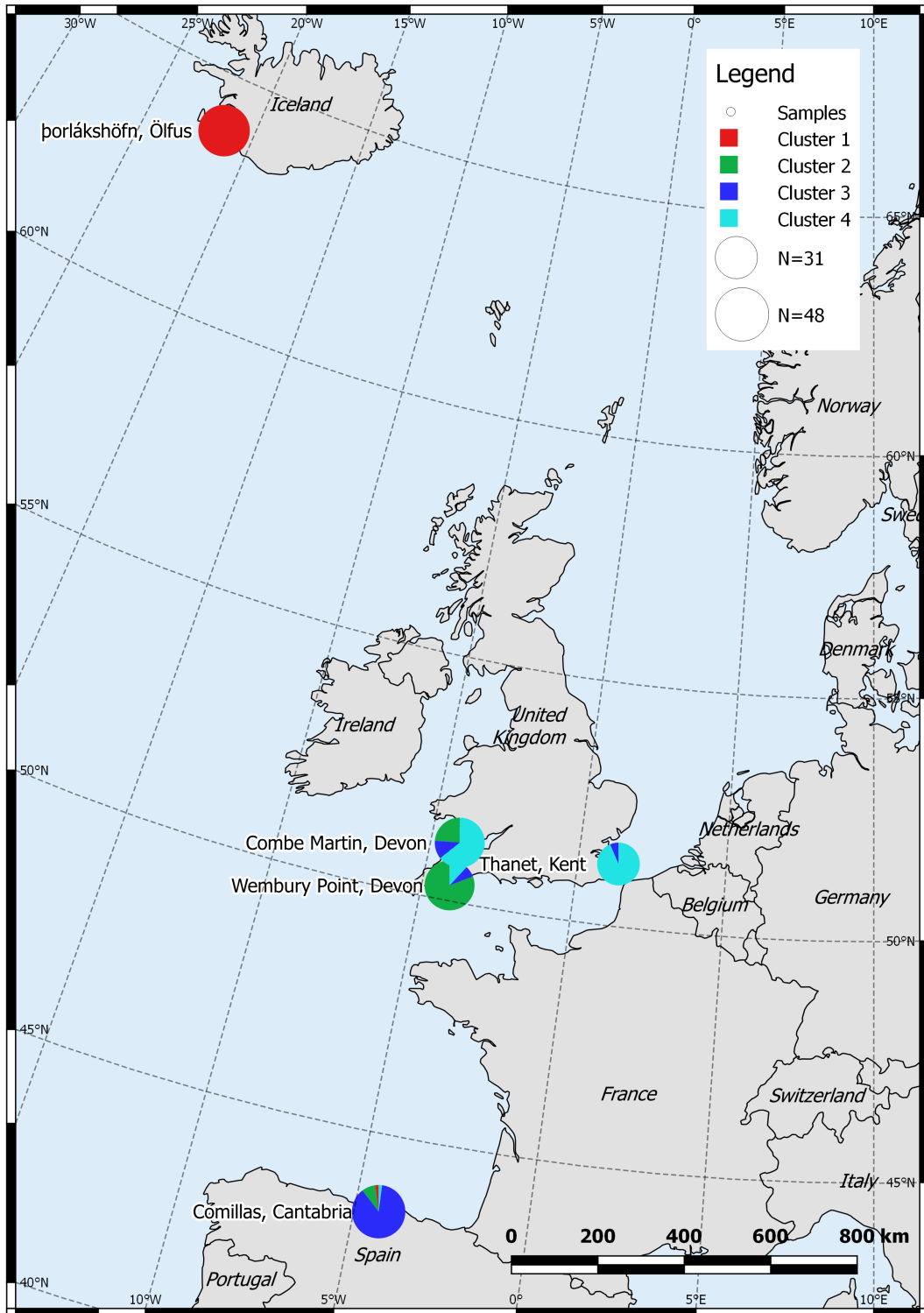


Fig. 2. Scatter plot visualizing the discriminant analysis of principal components (DAPC). Ellipses are centred on each of the four genetic clusters. Letters represent sample locations. I=Iceland, K=Kent, S=Spain, W=Wembury Point, C=Combe Martin. Proportion of conserved variance is 55%, with axis 1 (x) accounting for 41% and axis 2 (y) 13%.

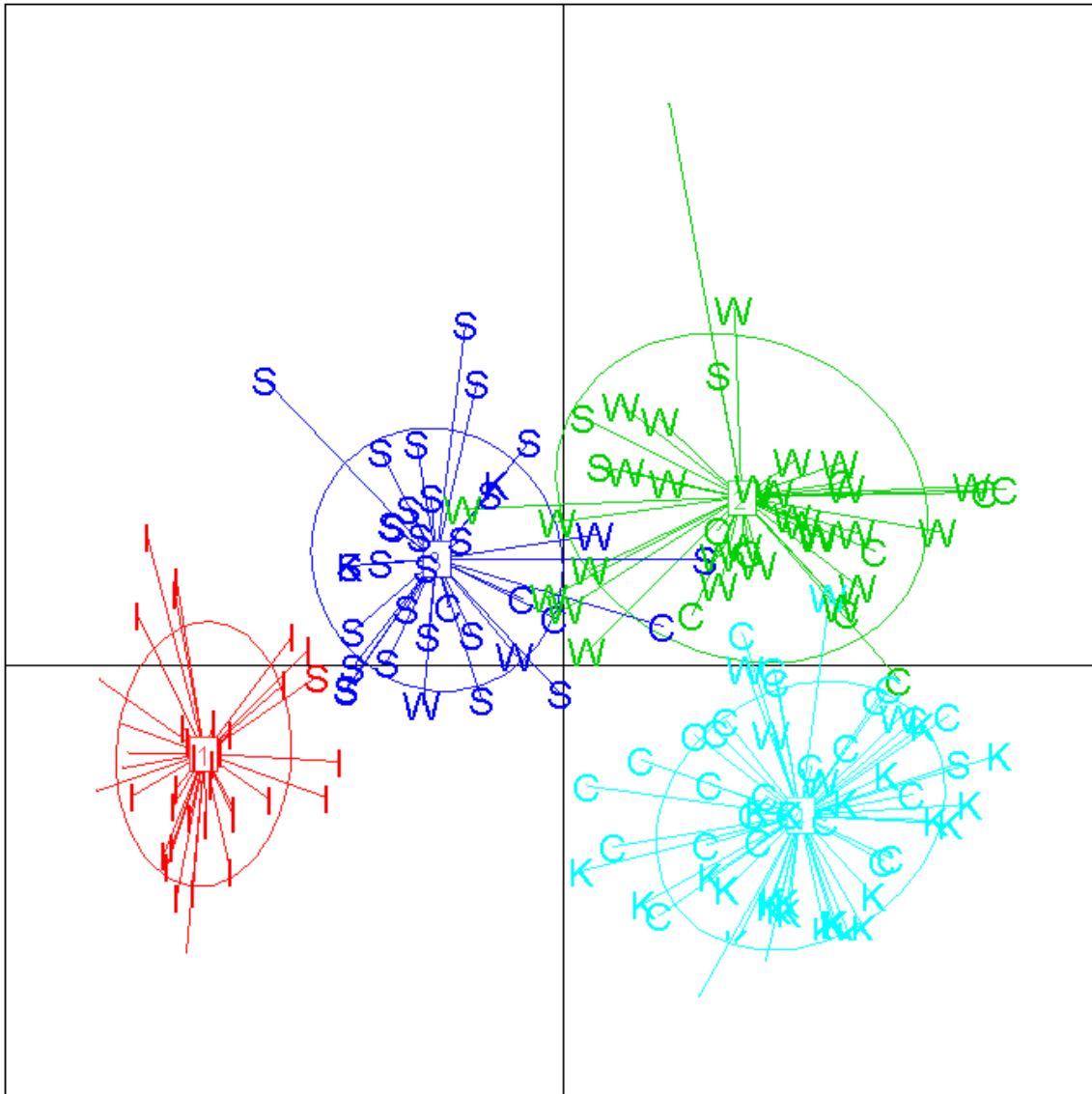
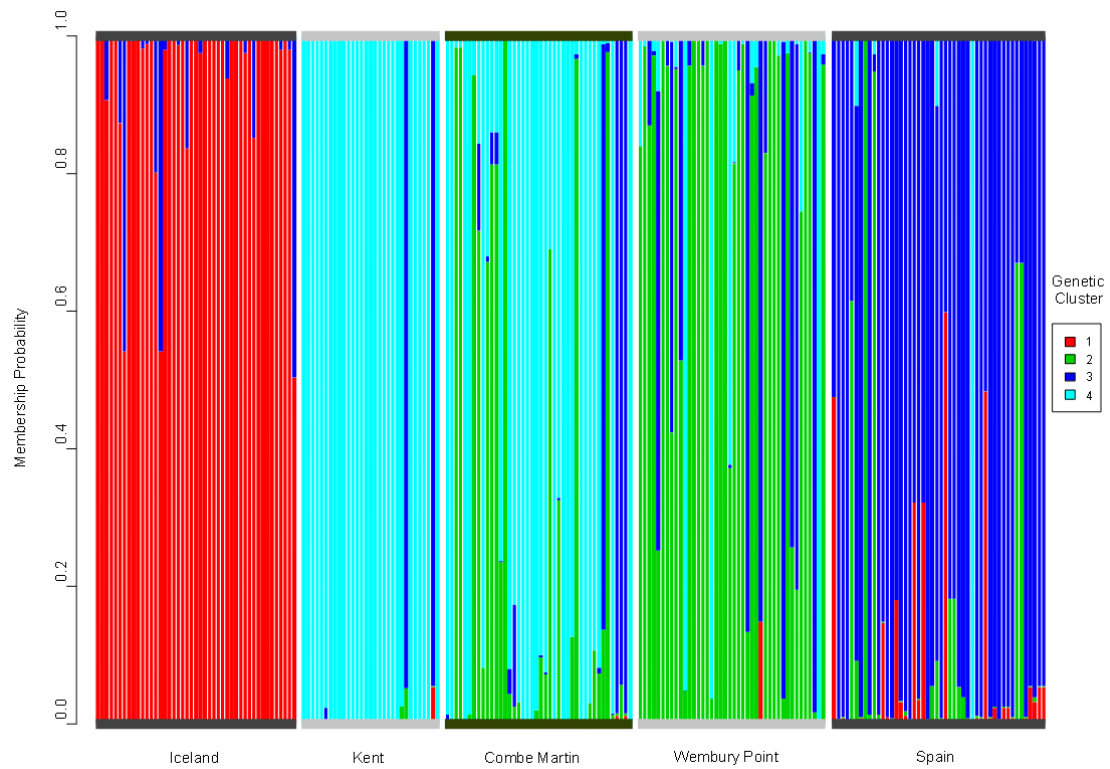
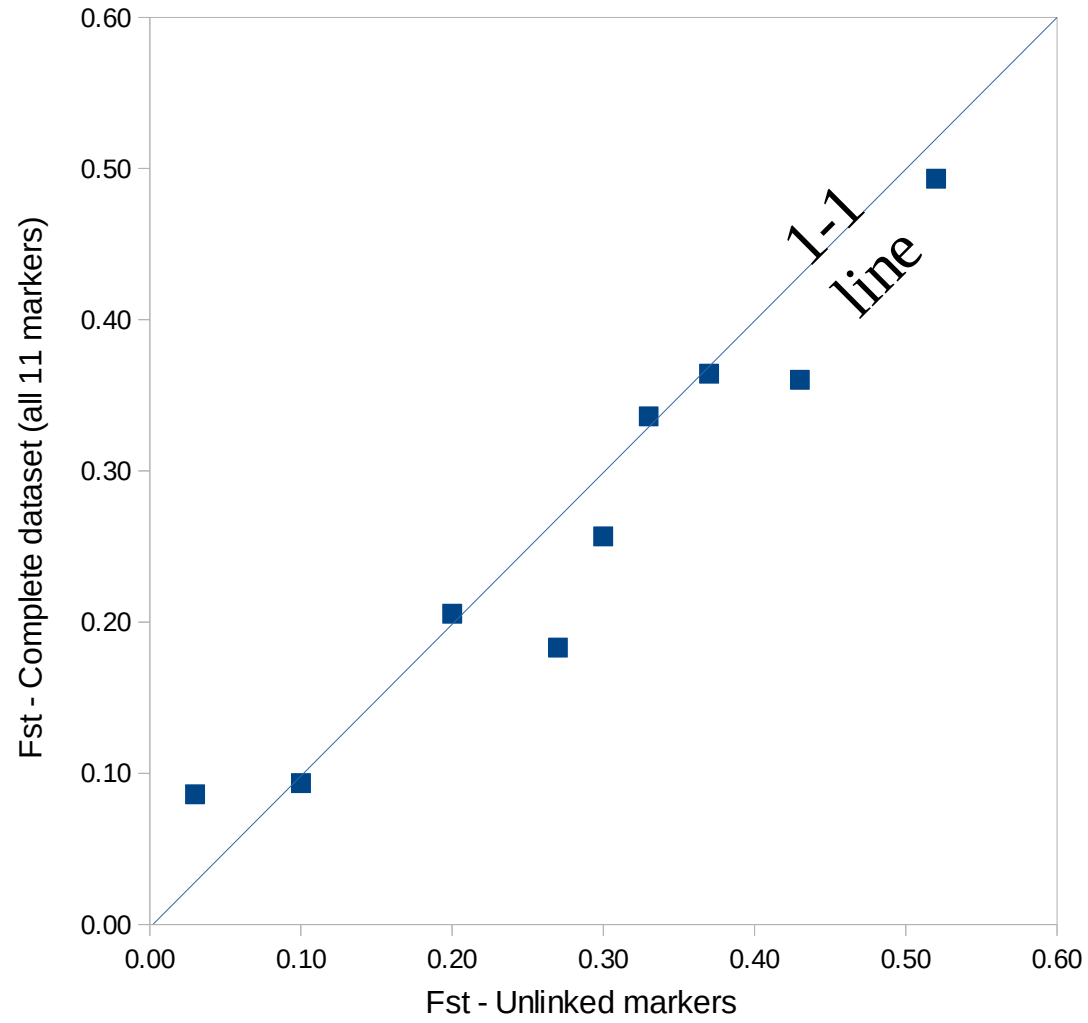


Fig. 3. Assignment probabilities of genetic clusters for all individuals used in this study.



Fst comparison - Unlinked markers and complete dataset



Supplementary Table 1 – Regional genetic and geographic distances and assignment of samples to genetic clusters (based on 5 unlinked markers). Lower triangle shows pairwise F_{st} values (***) indicates $p < 0.001$, ** $p < 0.01$, * $p < 0.05$). Upper triangle shows approximate over-water distances in thousands of kilometres. N = number of samples, 1-4 are genetic clusters.

Region	IC	KE	CM	WP	SP	N	Genetic Cluster			
							1	2	3	4
Þorlákshöfn, Ölfus (IC)	-	2.00	1.80	2.00	2.50	45	0	0	0	45
Thanet, Kent (KE)	0.52***	-	0.80	0.40	1.10	31	13	8	8	2
Combe Martin, Devon (CM)	0.37***	0.03*	-	0.40	0.90	42	12	14	14	2
Wembury Point, Devon (WP)	0.43***	0.15	0.10*	-	0.80	42	8	12	17	5
Comillas, Cantabria (SP)	0.30***	0.33***	0.20***	0.27**	-	48	22	4	22	0