BMC Genomics

RESEARCH

Open Access

Characterization of fine geographic scale population genetics in sugar kelp (*Saccharina latissima*) using genome-wide markers

Signe Bråtelund[1*], Tom Ruttink[2,3], Franz Goecke[1], Ole Jacob Broch[4], Gunnar Klemetsdal[5], Jørgen Ødegård[5] and Åshild Ergon[1]

Abstract

Background Kelps are not only ecologically important, being primary producers and habitat forming species, they also hold substantial economic potential. Expansion of the kelp cultivation industry raises the interest for genetic improvement of kelp for cultivation, as well as concerns about genetic introgression from cultivated to wild populations. Thus, increased understanding of population genetics in natural kelp populations is crucial. Genotyping-by-sequencing (GBS) is a powerful tool for studying population genetics. Here, using *Saccharina latissima* (sugar kelp) as our study species, we characterize the population genetics at a fine geographic scale, while also investigating the influence of marker type (biallelic SNPs versus multi-allelic short read-backed haplotypes) and minor allele count (MAC) thresholds on estimated population genetic metrics.

Results We examined 150 sporophytes from 10 locations within a small area in Mid-Norway. Employing GBS, we detected 20,710 bi-allelic SNPs and 42,264 haplotype alleles at 20,297 high quality GBS loci. We used both marker types as well as two MAC filtering thresholds (3 and 15) in the analyses. Overall, higher genetic diversity, more outbreeding and stronger substructure was estimated using haplotypes compared to SNPs, and with MAC 15 compared to MAC 3. The population displayed high genetic diversity (H_E ranging from 0.18–0.37) and significant outbreeding ($F_{IS} \leq -0.076$). Construction of a genomic relationship matrix, however, revealed a few close relatives within sampling locations. The connectivity between sampling locations was high ($F_{ST} \leq 0.09$), but subtle, yet significant, genetic substructure was detected, even between sampling locations separated by less than 2 km. Isolation-by-distance was significant and explained 15% of the genetic variation, while incorporation of predicted currents in an "isolation-by-oceanography" model explained a larger proportion (~ 27%).

Conclusion The studied population is diverse, significantly outbred and exhibits high connectivity, partly due to local currents. The use of genome-wide markers combined with permutation testing provides high statistical power to detect subtle population substructure and inbreeding or outbreeding. Short haplotypes extracted from GBS data and removal of rare alleles enhances the resolution. Careful consideration of marker type and filtering thresholds is crucial when comparing independent studies, as they profoundly influence numerical estimates of population genetic metrics.

Keywords Macroalgae, *Laminariales*, SNPs, Short read-backed haplotypes, Population structure, Inbreeding, Isolation by distance, Genetic diversity, Genomic relationship matrix

*Correspondence:
Signe Bråtelund
signe.bratelund@nmbu.no
Full list of author information is available at the end of the article

Background

Over the last decade, kelps (large brown algae in the order Laminariales [1]) have been recognized as a sustainable source of food and feed and a raw material for production of biofuels, pharmaceuticals and other products [2–6]. This recognition has sparked interest in kelp cultivation in Europe and the Americas. Selection and breeding of suitable genetic material of kelp species has been practiced in Asia for a long time [7, 8] and has recently been initiated in the USA [9]. In aquaculture, genetic introgression from farmed to wild populations is a significant issue (e.g. [10]) and one of the main concerns regarding the expansion of the kelp cultivation industry in Europe and the Americas [11–13]. In Asia, crop-to-wild gene flow from farmed *Undaria pinnatifida*, has been observed [14], though the full ecological effects have not been extensively studied. Genetic introgression could potentially lead to a shift in the wild population allele frequency spectrum or a change in the genetic diversity. In turn, this could make natural populations less robust to biotic and abiotic changes in their environment [15], reduce genetic diversity available for future breeding or change the population in other ways that may have unforeseen ecological consequences.

Since kelps play a key ecological role, both as a primary producers, and as a habitat forming species that supports a diverse marine life [16, 17], it is vital to avoid genetic introgression from kelp farms to wild populations. Two primary strategies can be pursued for this purpose. The first approach involves minimizing the risk of gene flow from cultivated crops to wild populations. Certain measures, such as harvesting the kelp before maturity, can contribute to reducing crop-to-wild gene flow. Still, some risk persists unless completely sterile varieties become available for cultivation. The second strategy entails exclusively using locally sourced kelp for cultivation. With this approach, it is important to determine the degree of local genetic specificity necessary to prevent significant alterations in the genetic composition of wild populations by genetic introgression. To do this it is crucial to deepen the understanding of genetic diversity, population substructure, and gene flow between and within kelp populations at relatively short geographical distances. Such knowledge is also fundamental to protect and restore kelp forests amidst the contemporary challenges of increasing sea temperatures and excessive grazing by sea urchins [18–20].

Our study species, *Saccharina latissima* L. (sugar kelp) forms dense kelp forests in relatively sheltered areas with rocky sediments at depths of less than 30 m [21] along the coast of Europe and the Atlantic coast of North America [22–24]. It is also currently one of the most cultivated species in Europe [7]. Previous studies of *S. latissima* have consistently revealed significant substructure between populations separated by moderate [25, 26] to long [23, 27, 28] distances. The observed genetic substructure between geographically distant populations could be the result of strong selection pressure for local adaptation, which has been suggested in studies of other sessile marine organisms including the kelp species *Macrocystis pyrifera* and *Laminaria digitata* [29–32], or it could be an indicator of limited gene flow over long distances and genetic drift, as most of these studies have observed significant isolation-by-distance (IBD) [23, 25–27, 33–35]. Fewer studies have investigated genetic substructure across short distances, and those that have, have not found significant IBD or significant genetic substructure across distances of less than 10 km [26, 36, 37]. These results are somewhat surprising, considering that kelp spores have been found to typically have high sinking rates and mostly settle within 500 m from their release site [38–40]. It has been proposed that local currents drive connectivity in kelp and other macroalgal species [37, 41], which could explain the absence of substructure between nearby populations. On the other hand, the studies that have investigated genetic substructure over short distances are based on few single sequence repeats (SSRs) as markers, and while they have been able to reveal significant genetic substructure between populations over longer distances, they might lack the resolution that is needed to observe genetic substructure on a very fine geographic scale.

In order to apply statistical methods with sufficient power to study genetic substructure over short geographic distances, it is crucial to understand the impact that multiallelic markers and rare alleles have on estimates of population genetic metrics like genetic diversity, substructure and inbreeding or outbreeding. Next-Generation Sequencing (NGS) techniques have revolutionized the field of population genetics, making genome-wide markers available for non-model organisms at a relatively low cost. One powerful and cost-effective method is Genotyping-by-Sequencing (GBS) [42]. This method involves digestion of the genomic DNA by sequence specific restriction enzymes, ligation of adapters to the resulting fragments, amplification, and sequencing of short regions (100–300 bp) between two neighboring restriction sites, resulting in a reduced representation of the genome. Since GBS does not require a pseudo-chromosome level reference genome assembly and results in tens of thousands of genome-wide markers, mostly SNPs, it is an excellent genotyping method for non-model organisms such as *S. latissima*. An alternative to utilizing SNPs directly in genetic analyses is to combine neighboring SNPs located within each GBS locus into haplotypes, which can increase the statistical

power [43]. These short read-backed haplotypes can be complemented with read mapping polymorphisms as a novel source of genetic diversity information [44]. The resulting short read-backed haplotypes are a genome-wide multiallelic marker type that contains more genetic information than separate SNPs. Despite their potential, such haplotypes are notably underutilized compared to SNPs. In order to remove SNP and haplotype variants that result from sequencing errors, GBS raw data sets are filtered based on Minor Allele Frequency (MAF) or Minor Allele Count (MAC). However, in establishing MAF or MAC thresholds for a study, there is a trade-off between eliminating sequencing errors and sacrificing genetic information by discarding real alleles. Still, it remains common practice to set MAF or MAC thresholds without thorough discussion of the potential unintended effects of such thresholds, despite several studies reporting biases associated with MAC or MAF filtering [45–48].

In this study we have utilized both SNPs and short read-backed haplotypes with two different MAC thresholds in a fine geographic scale population genetic study of the *S. latissima* population around Inntian island in Mid-Norway. We had the following objectives:

1. Characterize the fine geographic scale population genetics of the *S. latissima* population around Inntian in terms of genetic diversity, inbreeding and population substructure using genome-wide markers and computer-intensive significance tests.
2. Examine the effect of (i) short haplotypes versus SNP markers, and (ii) exclusion or inclusion of rare alleles (different MAC filtering thresholds), on commonly used measures of diversity, inbreeding and population substructure.
3. Explore population substructure on the individual level by implementing a genomic relationship matrix (GRM) and adapting it to multi-allelic haplotype markers.
4. Investigate if spore dispersal simulations, based on predicted currents, better explain genetic substructure between sampling locations than physical distance alone.

Results

Polymorphic markers and genetic diversity

A total of 159 sporophytes were sampled across 10 locations close to Inntian island (Fig. 1) and subjected to double-digest GBS fingerprinting, of which 150 high quality samples (creating 10 sampling locations with 14–16 individuals each) were retained for further analysis. SMAP *delineate* [44] was used to identify 20,297 high quality (HQ) GBS loci after extensive filtering (length range 100–275 bp, read depth > 10, and > 95% completeness across the sample set). The HQ GBS loci covered a total of 3.79 Mb of genomic DNA, which corresponds to between 0.49% and 0.64% of the genome, based on estimates for the genome size of *S. latissima* genome ranging between 588 and 774 Mb [49]. SNP calling and filtering identified 20,710 bi-allelic SNPs at a minor allele count (MAC) threshold of 3 (SNP_{MAC3}) and 11,391 SNPs at MAC15 (SNP_{MAC15}) within the HQ GBS loci. Subsequent haplotype calling with SMAP *haplotype-sites* [44] identified 12,012 and 8556 HQ GBS loci with at least two haplotype alleles after filtering haplotypes at MAC3 and MAC15, respectively ($Haplotype_{MAC3}$ and $Haplotype_{MAC15}$). So, 59% and 42% of the HQ GBS loci were polymorphic at MAC3 and MAC15 filtering, respectively. Taken together, this shows that a substantial fraction of the SNPs (45.0%) and haplotype alleles (41.0%) occurred at low frequency (1–5%) across the entire population. These were retained by MAC3 filtering and removed by MAC15 filtering (Fig. 2A-D).

The various sampling locations displayed different responses to MAC filtering, but removal of rare alleles decreased the estimated allelic richness (*AR*) in most of the sampling locations (Fig. 2E). *AR* is known to vary greatly with sample size [51] and most of the significant differences in *AR* ($P < 0.05$) in our study were detected between pairs of sampling locations with different sample size, but notably, location 7 had a significantly lower *AR* compared to all other sampling locations, including locations with the same sample size, at MAC3 (Additional file 1, Table A1). When the MAC threshold was increased to 15 the variance in *AR* between sampling locations decreased, indicating that most of the variation in *AR* between sampling locations was found due to alleles that were rare in the total population.

Two commonly used measures of genetic variation are expected (H_E) and observed heterozygosity (H_O). H_E is calculated based on allelic frequencies and can give an indication of a population's effective size. A low H_E could be caused by, *e.g.*, genetic drift or by inbreeding over time. H_O, on the other hand, is calculated based on genotype frequencies and is more sensitive to contemporary non-random mating and survival. Both the average H_E (Fig. 3A) and the average H_O (Fig. 3B) in our study were higher with the haplotype datasets than with the SNP datasets (at the same MAC threshold) and increased when rare alleles were excluded in MAC15 filtering. The increase in H_E and H_O with exclusion of rare alleles was greater in SNPs than in haplotypes. There were few pairs of sampling locations that differed significantly in average H_E, but the average H_E in the total population was significantly higher than in

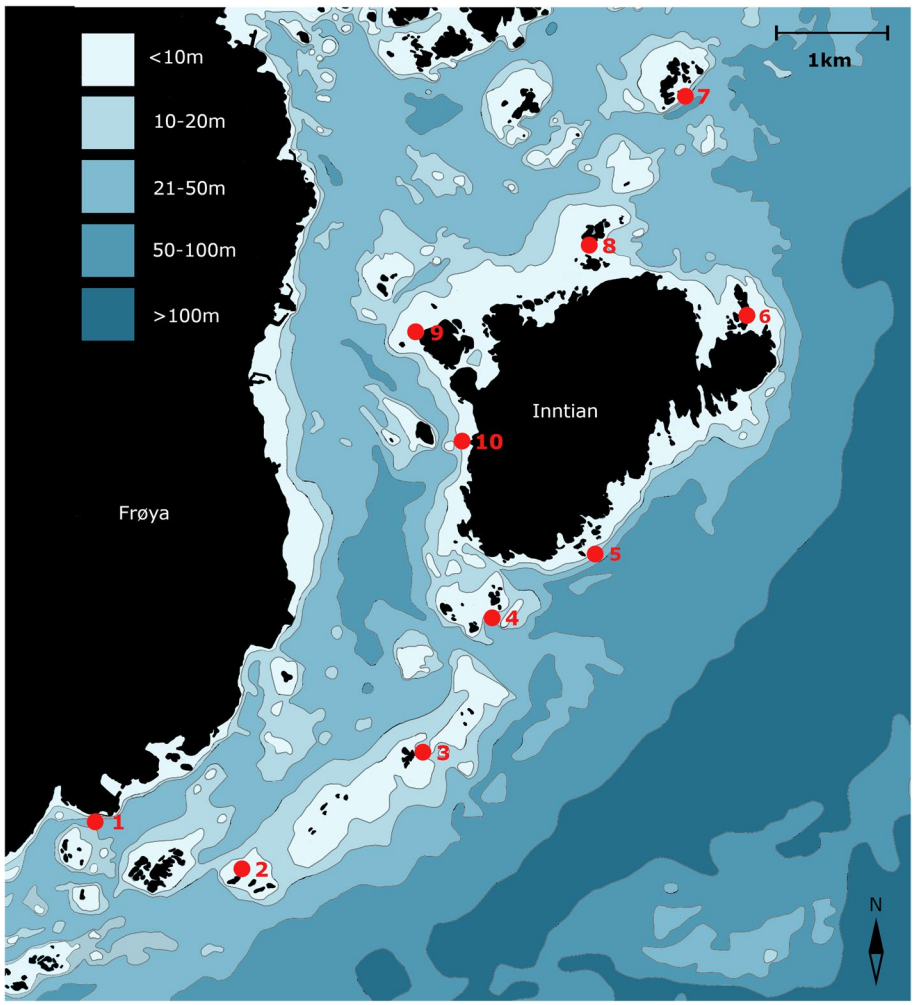


Fig. 1 Map of sampling locations. Red dots mark sampling locations. Sea depths are indicated by various shades of blue [50]

(See figure on next page.)

Fig. 2 Allele frequency distributions, allele count distribution and allelic richness. **A** Minor allele frequency distribution of SNPs. All bars combined represent SNPs included in the SNP_{MAC3} dataset. Light gray bars represent the SNPs that are also included in the SNP_{MAC15} dataset; **B** Allele frequency distribution of haplotype alleles. All bars combined represent alleles included in the $Haplotype_{MAC3}$ dataset. Light gray bars represent the alleles that are also included in the $Haplotype_{MAC15}$ dataset; **C** Distribution of number of SNPs per GBS locus in SNP_{MAC3} and SNP_{MAC15}. The numbers of loci are indicated on the bars; **D** Distribution of number of haplotype alleles per GBS locus in $Haplotype_{MAC3}$ and $Haplotype_{MAC15}$. The numbers of polymorphic haplotype loci are indicated on the bars; **E** Allelic richness in each sampling location measured in datasets $Haplotype_{MAC3}$ and $Haplotype_{MAC15}$. Whiskers represent 95% confidence intervals of the *AR*, acquired by bootstrapping over loci (20,000 bootstraps). The number of individuals sampled from each location (*n*) are indicated with gray bars

any of the sampling locations ($P < 0.05$, bootstrap t-test) in all the datasets, which indicates that there was subdivision in the total population. When it comes to average H_O, sampling location 7 had a significantly lower average compared to most of the other locations and to the total population ($P < 0.05$, bootstrap t-test), in all four marker datasets (for more detail, see Additional file 1, Tables A2-A6).

Population structure

The fixation index, F_{ST}, is a measure of population structure within populations that ranges from 0, indicating that there is no population substructure, to a theoretical

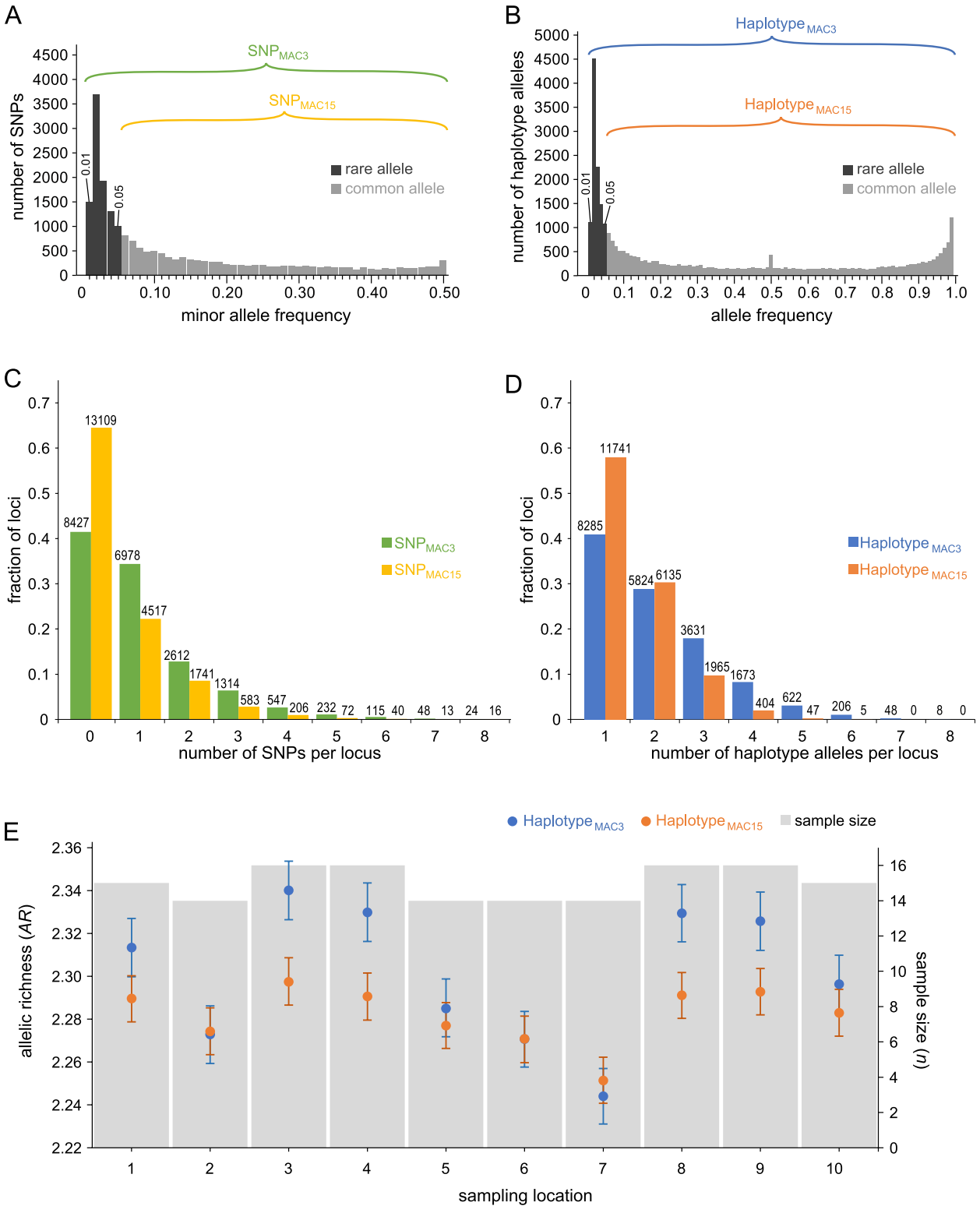


Fig. 2 (See legend on previous page.)

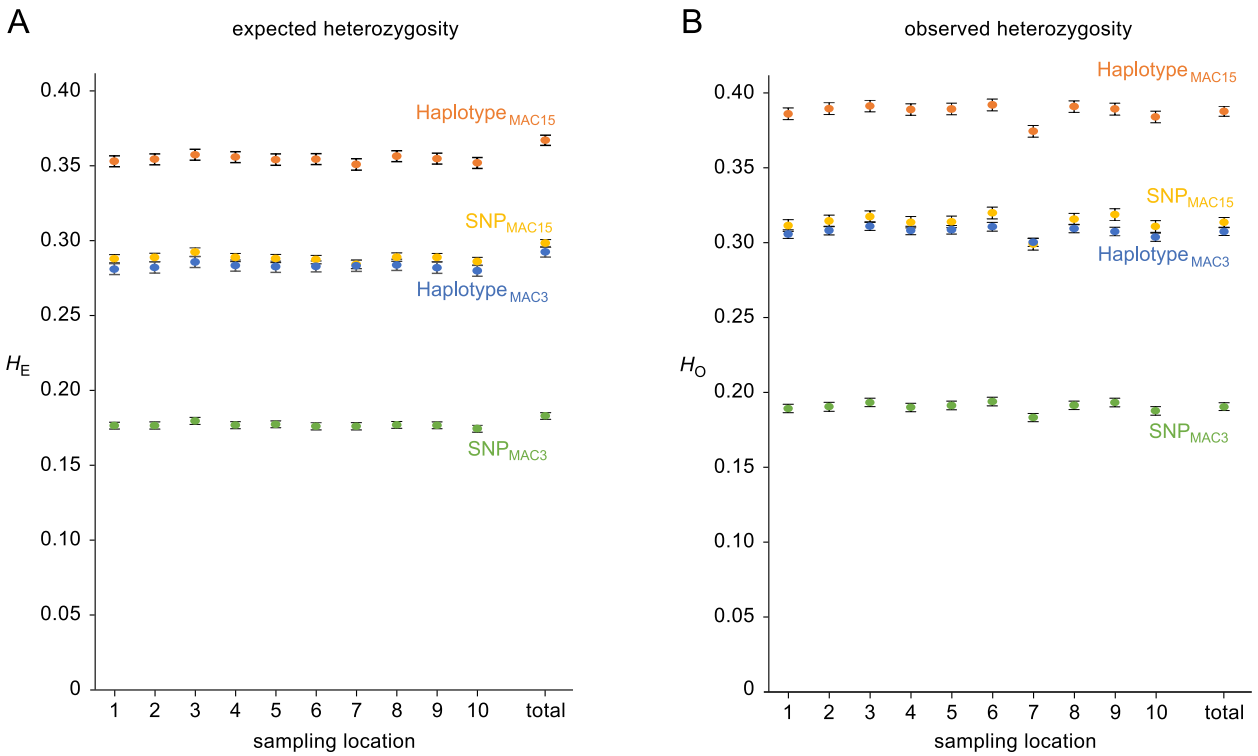


Fig. 3 Expected and observed heterozygosity. **A** Average expected heterozygosity (H_E) and **B** observed heterozygosity (H_O) across all loci, measured for all sampling locations and in the total population based on four data sets: green – SNP_{MAC3}; yellow – SNP_{MAC15}; blue – $Haplotype_{MAC3}$; orange – $Haplotype_{MAC15}$. Whiskers represent 95% confidence intervals, acquired by bootstrapping over loci (20,000 bootstraps)

maximum of 1 which indicates that there is complete fixation in the subpopulations. In this study we estimated F_{ST} in the total population (global F_{ST}) and between pairs of sampling locations (pwF_{ST}) (Fig. 4). To make the F_{ST} estimates based on SNPs and haplotypes as comparable as possible, we applied an F_{ST} estimate based on Hedrick's standardized G_{ST} with Nei's [52] bias correction for small numbers of subpopulations (G''_{ST}) [53]. The global F_{ST} was estimated to be 0.055, 0.058, 0.064 and 0.067 in the datasets SNP_{MAC3}, SNP_{MAC15}, $Haplotype_{MAC3}$ and $Haplotype_{MAC15}$, respectively, and was significantly larger than zero ($P < 10^{-4}$) in all datasets, suggesting that there was significant subdivision between the 10 sampling locations (Additional file 1, Table A7). Pairwise F_{ST} estimates (pwG''_{ST}) between all the pairs of sampling locations in the four datasets revealed that more genetic substructure was observed with haplotypes than with SNPs and with exclusion of rare alleles (*i.e.*, increasing MAC threshold). Despite this, the overall pattern was consistent, showing that sampling location 7 was more genetically different from the rest of the locations than those were from each other. The haplotype datasets show that sampling locations 3, 4, 5, 6 and 7 were significantly different from each other and from the rest of the sampling locations ($P < 0.01$), while sampling locations 1, 2, 8, 9 and 10 were more genetically alike. In contrast, sampling location 4 was not significantly different from most of the other sampling locations based on the two SNP datasets.

Population subdivision on the individual level was studied using a genomic relationship matrix (GRM) in which the entries represent genomic covariances between individuals (G_{ij}). GRMs were constructed based on the four datasets and the results were consistent across datasets, showing that most individuals in the entire sample set were somewhat less related to individuals from sampling location 7 than to individuals from other sampling locations (Fig. 5A). Within sampling location 7 there were two groups of genetically similar individuals (Fig. 5B), particularly, individuals 97 and 101 displayed a genomic covariance of 0.29, which is in the range expected for half-sibs (~0.25) and individual 98 displayed a genomic covariance of 0.47–0.48 with individuals 97 and 101, which is in the range expected for full-sibs or parents and offspring (~0.50) [54]. There were several pairs of related individuals within other sampling locations as well, but no close relationships ($G_{ij} > 0.1$) were observed for any pair of individuals from different sampling locations. All the observations mentioned above were consistent across all the SNP and haplotype datasets (Additional file 2). A cluster analysis confirmed the observations, since the

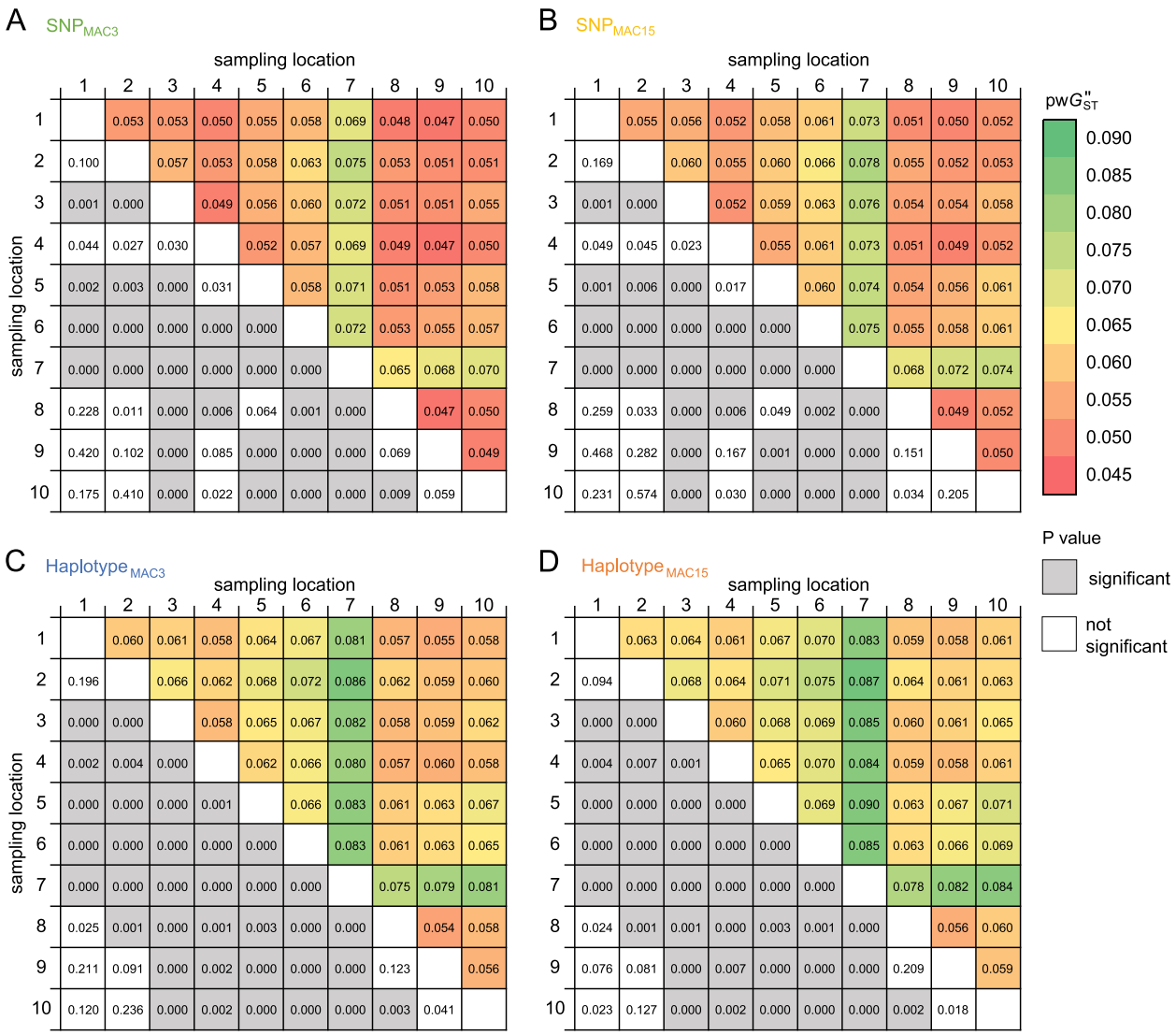


Fig. 4 Pairwise F_{ST} across all loci. Pairwise F_{ST} estimated based on four datasets: **A** SNP$_{MAC3}$; **B** SNP$_{MAC15}$; **C** Haplotype$_{MAC3}$; **D** Haplotype$_{MAC15}$. Pairwise F_{ST} are shown above the diagonals with lower values colored red and higher values colored green. *P*-values, acquired by Monte Carlo permutation testing (10,000 permutations), are shown below the diagonal with significant values ($P < 0.01$) marked in gray

most prominent clusters represented groups of related individuals from sampling location 7, in all datasets, while there was strong admixture between the rest of the sampling locations (Additional file 1, Figure A1).

To investigate if water currents could explain the observed genetic substructure between the 10 sampling locations, a spore dispersal simulation using a 3-dimentional ocean model was performed. The results indicated variable dispersal from the 10 sampling locations to other locations across the study area (Fig. 6A). A relatively large number of particles released from sampling locations 4, 8, 9 and 10 traveled across the area west/north-west of Inntian. A considerable number of simulated particles from sampling locations 3 and 5 also dispersed to a part of the same area, but most of the particles from sampling locations 3 and 5 were concentrated relatively close to their release sites. The particles released from sampling locations 1 and 2 were also somewhat concentrated around the release sites, but some particles dispersed in the area south-west of Inntian. Particles from sampling location 6 were dispersed along the northern coast of Inntian, including to sampling location 8. Particles from sampling location 7 were concentrated around the release site, with some dispersal in the area north of Inntian. The currents at sampling location 7 were spread in all directions and the scalar speed was relatively low here, corroborated by

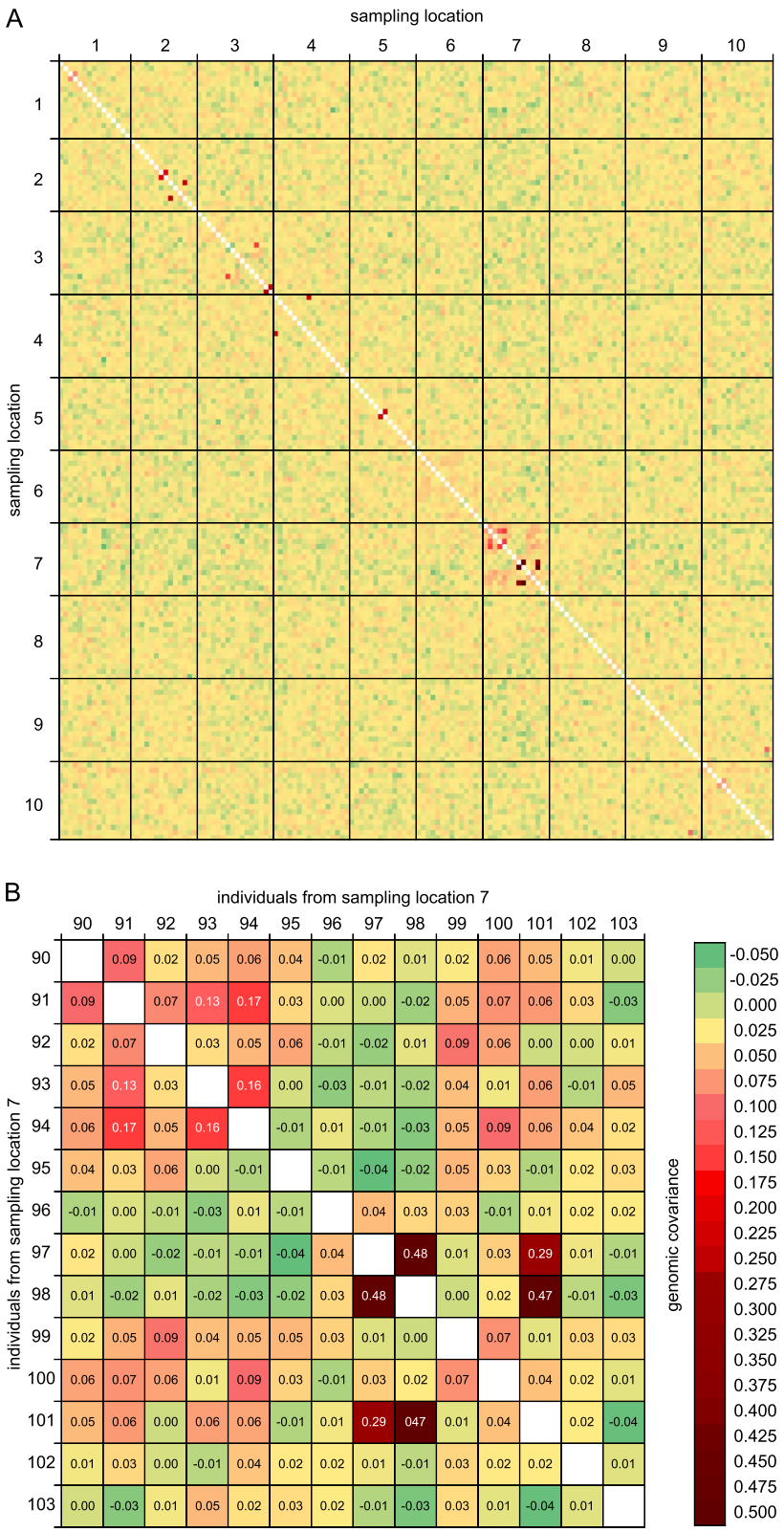


Fig. 5 Heatmaps of individual genomic covariances in the total population and sampling location 7. Heatmaps of genomic covariances (GRM entries) between individuals calculated from the $\text{Haplotype}_{\text{MAC3}}$ dataset. **A** Genomic covariances between all 150 individuals, split by sampling location; **B** Genomic covariances between individuals in sampling location 7. Higher genomic covariances are marked in red and lower genomic covariances are marked in green. The color scale is identical in panes A and B

the almost circular region of high particle concentrations around the site. Overall, the results show that most particles do not disperse far from their release site and the dispersal is not equal in all directions.

The simulated directional connectivity between the sampling locations, measured as the fraction of particles released from one site that reached another site (Fig. 6B) indicated limited flow of particles from sampling location 7 to the remaining sampling locations during the simulated period of time, except for sampling location 8, which received a moderate number of particles from sampling location 7. Sampling location 6 received the least number of particles from other sampling locations; sampling location 8 was the only location that provided a relatively high number of particles to sampling location 6. No pair of sampling locations had zero connectivity in the simulations. Taken together, the dispersal simulation results indicate that genetic exchange is possible between all sampling locations within the sampling area but that there is a spatial pattern of particle displacement and directional connectivity.

To further substantiate substructure between sampling locations, we searched for private alleles per location. Private alleles could only be identified when rare alleles were included, *i.e.*, in the SNP_{MAC3} and $Haplotype_{MAC3}$ datasets. This was expected since the population size per sampling location is in the range of 14 to 16 individuals, making the chances of an allele with at least 15 copies (MAC15) in the total population only being present in one sampling location small. Strikingly, sampling location 7 contained 55% (SNP_{MAC3}) or 58% ($Haplotype_{MAC3}$) of all the private alleles in the datasets (Fig. 6C).

To assess the effects of distance and currents on genetic substructure between the sampling locations, two models of isolation were applied – an isolation-by-distance (IBD) model and an "isolation-by-oceanography" (IBO) model. In the IBD model, linearized pairwise F_{ST} estimates ($pwG''_{ST}/(1-pwG''_{ST})$) were plotted against physical distance with a small, yet significant ($P<0.01$) correlation, indicating that physical distance alone explained around 15% of the genetic substructure between sampling locations (Fig. 7A). In the IBO model, currents were considered in addition to physical distance by using the inversed simulated connectivity from the spore dispersal predictions as a measure of distance (Fig. 7B). The IBO model explained a considerably larger proportion of the genetic substructure between the sampling locations (approx. 26–28%) than the IBD model.

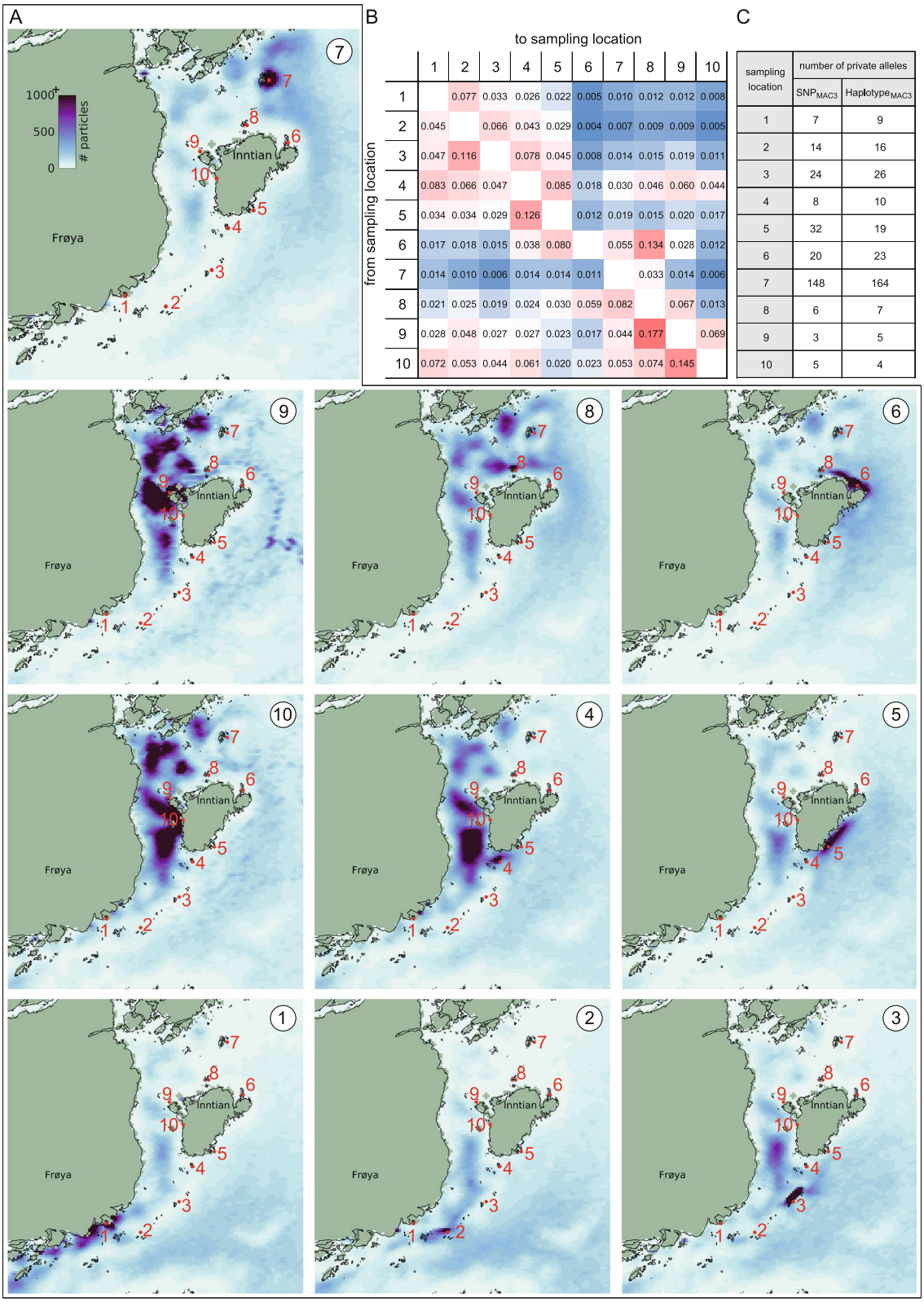
Outbreeding

While F_{IS} measures inbreeding as the excess of homozygotes on the subpopulation level [55], F_{GRM} is expected to estimate the fraction of alleles that are likely to be identical by decent in an individual, emphasizing loci that are homozygous for an allele that is rare in the total population [54]. Both measures were significantly negative in the total population, irrespective of dataset, indicating that the population was somewhat outbred (Table 1). All sampling locations had significantly negative F_{IS} and F_{GRM} values in all the datasets (averages are given in Table 1, for details see Additional file 1, Table A8), except sampling location 7, which had a F_{IS} that was not significantly negative in the SNP_{MAC3} dataset and mean F_{GRM} values that were not significantly different from 0 across all datasets. Overall, F_{IS} values became more negative when rare alleles were excluded and in haplotypes compared to SNPs. F_{GRM} also became more negative with exclusion of rare alleles in the SNP datasets, but in the haplotype datasets, the absolute F_{GRM} values decreased when rare alleles were excluded. The variation in F_{IS} across loci in the total population was between 0.032, and 0.049 and was higher when using haplotypes than SNPs and when applying a higher MAC threshold (more details are given in Additional file 1, Table A9).

The individual inbreeding coefficients, F_{GRM}, were negative in most of the individuals across populations (Additional file 2). In the dataset SNP_{MAC3} 24 individuals had positive F_{GRM}, of which 9 were from sampling location 7; in SNP_{MAC15} 12 individuals had positive F_{GRM}, of which 6 were from location 7; $Haplotype_{MAC3}$ had 7 individuals with positive F_{GRM}, all from location 7; and $Haplotype_{MAC15}$ had 7 individuals with positive F_{GRM}, of which 5 were from location 7. Individual 98 had an F_{GRM} between 0.307 and 0.390 (0.250 is expected for an offspring of two siblings) in all datasets, which was notably higher than the second most inbred individual in each dataset (individual 94), which had an F_{GRM} ranging from 0.079 to 0.104.

(See figure on next page.)

Fig. 6 Simulated spore dispersal, connectivity and private alleles. **A** Simulated spore dispersal from the 10 sampling locations. All sampling locations are indicated on the map and the release sampling location is indicated in the top-right corner of each map. The colors indicate the number of simulated particles (out of 2882) that have traveled through each pixel of the grid; **B** Heatmap showing the simulated connectivity between the sampling locations measured as the proportion of particles released from the FROM sampling location that reaches the TO sampling location. High connectivity is marked in red and low connectivity is marked in blue; **C** Table showing the number of private alleles in each sampling location in the SNP_{MAC3} and $Haplotype_{MAC3}$ datasets



	1	2	3	4	5	6	7	8	9	10
1		0.077	0.033	0.026	0.022	0.005	0.010	0.012	0.012	0.008
2	0.045		0.066	0.043	0.029	0.004	0.007	0.009	0.009	0.005
3	0.047	0.116		0.078	0.045	0.008	0.014	0.015	0.019	0.011
4	0.083	0.066	0.047		0.085	0.018	0.030	0.046	0.060	0.044
5	0.034	0.034	0.029	0.126		0.012	0.019	0.015	0.020	0.017
6	0.017	0.018	0.015	0.038	0.080		0.055	0.134	0.028	0.012
7	0.014	0.010	0.006	0.014	0.014	0.011		0.033	0.014	0.006
8	0.021	0.025	0.019	0.024	0.030	0.059	0.082		0.067	0.013
9	0.028	0.048	0.027	0.027	0.023	0.017	0.044	0.177		0.069
10	0.072	0.053	0.044	0.061	0.020	0.023	0.053	0.074	0.145	

sampling location	number of private alleles	
	SNP_{MAC3}	$Haplotype_{MAC3}$
1	7	9
2	14	16
3	24	26
4	8	10
5	32	19
6	20	23
7	148	164
8	6	7
9	3	5
10	5	4

Fig. 6 (See legend on previous page.)

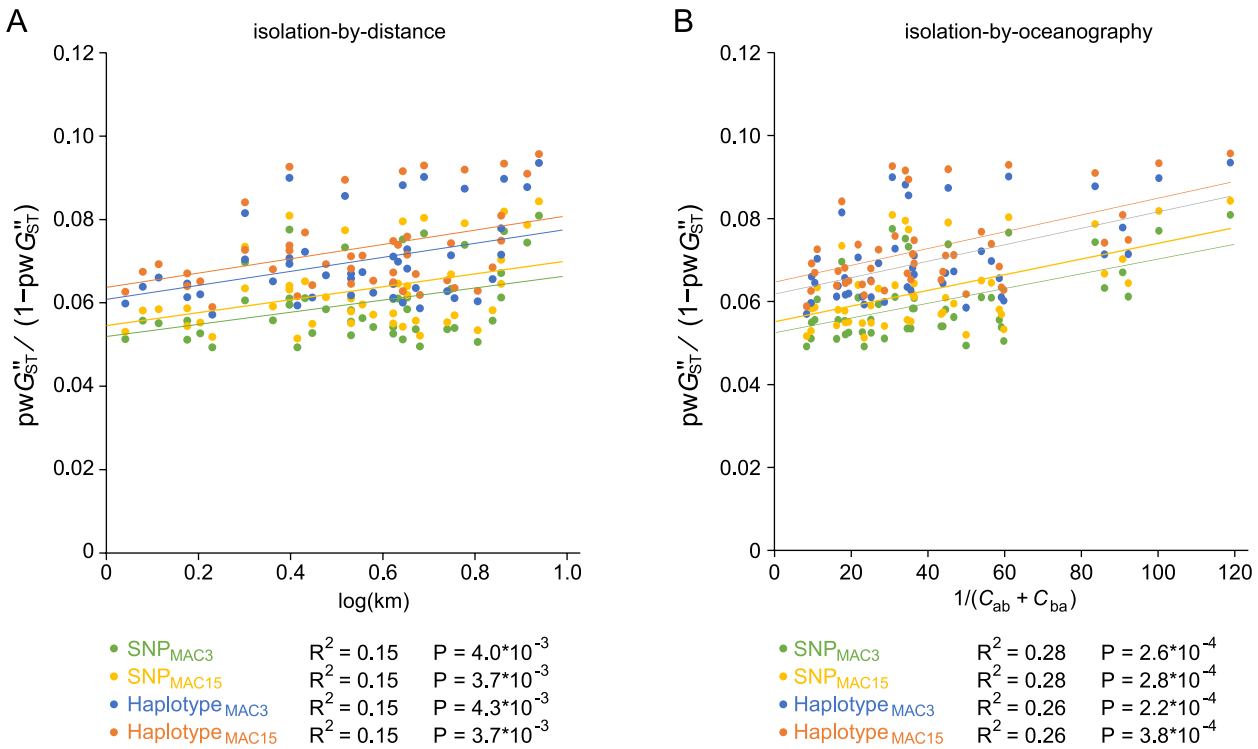


Fig. 7 Isolation-by-distance and "isolation-by-oceanography". **A** Isolation-by-distance (IBD) model with squared correlation between genetic distance, estimated as linearized pairwise F_{ST} (pwG''_{ST}/(1 − pwG''_{ST})), and physical distance, measured as log(km) between each pair of sampling locations; **B** "Isolation-by-oceanography (IBO)" model, also with squared correlation between genetic distance (pwG''_{ST}/(1 − pwG''_{ST})) and oceanographic distance measured as the inverse connectivity between two sampling locations based on spore dispersal simulations. Regression lines are shown. *P*-values are obtained from a simple Mantel test (50,000 permutations)

Discussion

Marker types and genetic information content

In this study we have compared two different types of genome-wide markers – biallelic SNPs and bi- or multi-allelic short read-backed haplotypes extracted from GBS read data. We have selected methods that are considered comparable and analogous between SNPs and haplotypes. Still, the population genetic metrics we have estimated tended to vary between marker types. H_E and H_O were notably higher when based on haplotype datasets than on SNP datasets. This is expected since the number of alleles per locus increases when SNPs are phased together into multi-allelic short read-backed haplotypes. The observed excess of heterozygotes was higher with haplotypes than with SNPs, leading to more negative F_{IS} values and better resolution for detection of inbreeding and outbreeding.

F_{ST} estimates are known to vary between marker types, with the number of alleles per locus, and with the amount of genetic diversity [56, 57]. To make the F_{ST} estimates based on SNPs and haplotypes as comparable as possible in this study, we have applied an F_{ST} estimate based on Hedrick's standardized G_{ST} with Nei's [52] bias correction for small numbers of subpopulations (G''_{ST}) [53]. Still, our haplotype-based F_{ST} estimates are not just numerically higher than their SNP-based equivalents, they are also more significant. Thus, haplotypes provide a better resolution than SNPs in detecting population substructure. It has been shown in simulation studies that

Table 1 Mean F_{IS} and mean F_{GRM} the total population, based on four marker datasets. Mean F_{IS} the unweighted mean of F_{IS} across all the sampling locations. Mean F_{GRM} is the mean individual F_{GRM} (obtained from the GRM) across all individuals. The significance of the F_{IS} values was tested using Monte Carlo permutation testing with simulation of random mating (10,000 permutations) and the significance of the F_{GRM} values was tested using a bootstrap t-distribution of the mean (50,000 bootstraps)

	Dataset			
	SNP_{MAC3}	SNP_{MAC15}	$Haplotype_{MAC3}$	$Haplotype_{MAC15}$
mean F_{IS}	−0.076[a]	−0.085[a]	−0.084[a]	−0.091[a]
mean F_{GRM}	−0.031[b]	−0.041[b]	−0.036[b]	−0.033[b]

[a] Negative mean F_{IS} significantly different from zero ($P < 0.05$)

[b] Negative mean F_{GRM} significantly different from zero ($P < 0.05$)

the statistical power for detecting genetic substructure can increase when SNPs on the same loci are combined into haplotypes [43]. Moreover, short read-backed haplotypes called with SMAP contain read mapping polymorphisms (*i.e.,* a form of genetic polymorphism indirectly derived from indels, and read-reference mismatches leading to soft and hard clipping) additional to SNPs, further increasing the genetic information content per GBS read [44].

Effect of filtering out rare alleles on genetic diversity measures

In this study, we have performed all the genetic analyses both including and excluding rare alleles by applying two different minor allele count (MAC) thresholds: MAC3 (corresponding to a minor allele frequency (MAF) of 1% in these data) and MAC15 (MAF = 5%). If rare alleles that we excluded from the MAC15 datasets relative to the MAC3 datasets were mostly sequencing errors, they would be expected to be evenly distributed across the sampling locations. However, the decrease of *AR* upon exclusion of rare alleles from the haplotype datasets varied between sampling locations (Fig. 2E) and more than half of the private alleles, which were exclusively rare alleles, were found in individuals from sampling location 7 (Fig. 6C). This indicates that some true rare alleles were lost when the MAC threshold was increased from 3 to 15.

Exclusion of rare alleles affected SNP datasets in a different way than haplotype datasets. As expected, removal of rare SNP alleles exclusively increased the mean H_E as SNP loci with low H_E were removed. Similarly, removal of bi-allelic haplotype loci with one rare allele will increase the average H_E. However, when removing rare alleles from multiallelic loci, the loci will be retained if there are at least two non-rare alleles left. Thus, removal of rare alleles from such multiallelic loci will reduce both *AR* and H_E. This is most likely the reason why the increase in H_E with increasing MAC threshold was lower in the haplotype datasets than in the SNP datasets. Similarly to H_E, observed heterozygosity (H_O) increased when rare alleles were excluded, which is expected theoretically and in coherence with a previous study of *S. latissima* [58]. When it comes to F_{IS} estimates, removing rare alleles increased the absolute F_{IS} values, both in the SNP and haplotype datasets, indicating that exclusion of rare alleles (whether real or false) increased the genetic resolution when it comes to detection of heterozygote excess or deficiency. The increase in resolution can be a consequence of the larger maximum possible deviation from Hardy–Weinberg equilibrium (HWE) is possible in loci with H_E closer to 0.5. The F_{ST} estimates also increased when rare alleles were excluded, both in the SNP and haplotype datasets, which is expected since the *major* alleles of loci with low MAC make most subpopulations more genetically similar. Thus, the change in F_{ST} caused by removing rare alleles can be a combination of increased resolution due to exclusion of sequencing errors, and of bias caused by exclusion of real low frequency alleles. In a population genetic study of *S. latissima*, Thomson [58] showed that pairwise F_{ST} estimates increased by 20–30% in two populations when the minor allele frequency threshold was increased from 1 to 5% (corresponding to MAC3 and MAC15 in the current study). Taken together, these results show that it is important to keep the filtering thresholds in mind when comparing estimates of heterozygosity, F_{IS} and F_{ST} across independent studies.

The use of a genomic relationship matrix

The genomic relationship matrix (GRM) [54] is extensively used in quantitative genetics, but we here show its value also in population genetics. It provides insight on the individual level, with individual inbreeding coefficients on the diagonal ($1 + F_{GRM}$) and genomic covariances (corresponding to additive genetic relationships) between individuals, on the off-diagonal. Furthermore, it can be used for hierarchical clustering to examine subdivisions in populations [59]. Originally, the GRM was developed for bi-allelic data [54], but in this study, we have shown that GRMs can be constructed for multi-allelic markers with a method that treats each allele as a separate locus, a method that is analogous to VanRaden's first method [54].

Since pairs of individuals sharing rare alleles will exhibit higher genetic covariances (off-diagonal) compared to pairs sharing common alleles [54], the GRM can be used to detect pairs or groups of individuals that share more rare alleles than others and that are therefore likely related. This is because shared rare alleles are more likely to be identical by descent, whereas shared common alleles are more likely to be identical by chance. The emphasis on rare alleles in F_{GRM} makes it a valuable addition to F_{IS} in population genetic studies since it can detect accumulation of rare alleles in an individual in addition to high levels of homozygosity across loci. Consequently, F_{GRM} can capture signals of inbreeding in previous generations, thereby complementing F_{IS}, which primarily detects contemporary inbreeding. Moreover, the genomic covariances estimated by VanRaden's first method have similar expectations to pedigree-based relationships, making them directly interpretable and valuable for identifying pairs of siblings or clones.

Population genetics of the *S. latissima* population around Inntian

Genetic diversity

Previous population genetic studies of *S. latissima* have reported varying levels of genetic diversity (Additional file 1, Table A10). Evankow et al. [27] found that the populations in the South Norwegian Sea ecoregion, which also spans the sampling location of our current study, had the highest H_E of all the studied ecoregions along the whole coast of Norway. They estimated H_S (subpopulation or region specific H_E) to be 0.468 and 0.420 in two sampling locations in the South Norwegian Sea. Another population genetic study of *S. latissima* in Norway by Ribeiro et al. [26], that included a sampling location near Frøya (near our study location), estimated H_E and H_O to be 0.633 and 0.643, respectively. These two studies used 9–12 SSR markers and their heterozygosity estimates were considerably higher than the ones we found in any of our datasets (0.18–0.37; Fig. 3A). This difference can be attributed to the use of different marker types since studies have shown that the use of SSRs can result in dramatically higher estimates of expected heterozygosity than what is estimated using SNPs [23]. We here present the first study of *S. latissima* populations in Norway that utilizes genome-wide markers, but there are a few such studies from other geographical areas (Additional file 1, Table A10). Mao et al. [33] studied the *S. latissima* population in the north-eastern United States and estimated H_S and H_O to be 0.26–0.31 and 0.26–0.32, respectively, which is similar to what we have found in the SNP dataset with the same MAC filtering threshold (SNP_{MAC3} corresponding to MAF1). Thomson [34, 35] reported H_S estimates of 0.112–0.140 along the west coast of Scotland and 0.231–0.249 along the west coast of Sweden based on SNPs filtered at MAF3 (corresponding to MAC9 in this study). Considering the MAF filtering thresholds applied in both studies, the *S. latissima* population around Inntian (estimated H_S of approx. 0.18 with MAC3 and 0.29 with MAC15 in the SNP datasets) appears to be more diverse than the populations on the west coast of Scotland and similar to the populations on the west coast of Sweden.

Outbreeding

The significantly negative F_{IS} and mean F_{GRM} observed in our study suggest that the *S. latissima* population around Inntian displays a low degree of outbreeding (Table 1). Ribeiro et al. [26] also reported slightly negative F_{IS} near our sampling site but did not assess the statistical significance of the F_{IS} values. Previous studies from other locations report both significantly positive and significantly negative F_{IS} values (Additional file 1, Table A10), however, few studies have reported more negative F_{IS} values than what was observed in our current study, the exceptions being one population around Helgoland outside the North Sea coast of Germany [23] and a few populations in Maine [33]. One reason why the population that we have studied displays negative F_{IS} and F_{GRM} values can be inflow of genetic material from nearby locations. The studied population is unlikely to be isolated since there are large areas with similar sea depths as our study area (Fig. 1) around Frøya island [50] and *S. latissima* forests have been observed and predicted [21] at several locations in the archipelago that spans our study. Furthermore, generally high connectivity has been observed along the Norwegian coast [26] and between populations in the South Norwegian Sea ecoregion [27], and the low, although significant F_{ST} values, as well as the particle dispersal predictions in the current study suggest high connectivity on the local geographical scale as well. Another factor contributing to negative F_{IS} values could be partial clonality [60, 61]. Clonal reproduction of diploid sporophytes can occur in Laminariales through some forms of parthenogenesis or apospory (see Goecke et al. 2020 [7]). If vegetative reproduction was dominating, the variance in F_{IS} between loci would be expected to be large [61], but in our case it was relatively low. Additionally, no pairs of clones were identified among the samples, as they would have been easily recognized in the GRM with a genomic covariance close to 1. Taken together, this implies that although contemporary cloning may occur, sexual reproduction appears to be predominant in the population. Finally, the excess of heterozygotes (negative F_{IS}) and the low individual inbreeding coefficients (F_{GRM}) could be an indication of inbreeding depression, which have been shown to affect fertility and survival in the kelp species *Macrocystis pyrifera* [62].

Population substructure

Previous studies using SSRs and considering short distances have not observed significant genetic substructure between subpopulations of *S. latissima* that are less than 10 km apart [26, 36, 37] (see also Additional file 1, Table A10), while in the current study significant pairwise F_{ST} was detected over distances of less than 2 km. The global F_{ST} estimates in the current study (0.055–0.067) were also considerably higher than what has been found within small geographic areas in previous studies [33, 36], and comparable to F_{ST} estimates found across larger areas (within 750 km) [37]. F_{ST} values can depend on marker type; F_{ST} values based on SNPs were about twice as large as those based on SSRs characterized in the same populations [23]. F_{ST} values can also vary depending on estimation methods and on removal or inclusion of rare alleles [63], and one possible reason for the high F_{ST} estimates in our current study, compared to previous

fine geographic scale studies, may be that we have used an F_{ST} estimate based on G_{ST}, while previous studies have used AMOVA (Φ_{ST}) [64] or Weir and Cockerham's [65] unbiased F_{ST} estimate (θ_{ST}). While Φ_{ST} and θ_{ST} compare populations to the most recent common ancestral population, G_{ST} compares populations to the total current population, sometimes resulting in higher F_{ST} estimates [63].

While there was high genetic connectivity between all the sampling locations in this study, some were more connected than others and location 7 clearly stood out from the rest, both on the subpopulation (F_{ST}) and individual (GRM) level. Furthermore, location 7 was less diverse and less outbred, and it had more private alleles than the other sampling locations. The spore dispersal simulations suggest that location 7 is somewhat isolated from the other sampling locations by currents, which could explain some of the substructure. Moreover, the lower degree of outbreeding could also be the result of isolation. The higher number of inbred individuals (with positive F_{GRM}) from sampling location 7 indicated that some degree of inbreeding has occurred in previous generations. One explanation for the accumulation of rare and private alleles in individuals from location 7 could be a reduced gene flow from location 7 to the other sampling locations. The closely related groups of individuals from location 7 could also cause a positive bias in private allele count, since a rare allele in one individual in a group is likely to be found in the other individuals in the same group, raising the allele count over the MAC3 threshold. On the other hand, the presence of a highly related group of individuals in location 7 does not seem coincidental as this location contains several related groups and has more complex substructure on the individual level than any of the other sampling locations (Fig. 5, see also Additional file 1, Figure A1). Overall, location 7 stands out as notably distinct from the other sampling locations, and this distinction cannot be attributed solely to physical distance. However, one would need a larger sample from sampling location 7 as well as samples from outside of the geographic area covered in this study to determine whether the difference is caused by connectivity to populations outside of the study or isolation. Moreover, local adaptation could contribute to making location 7 genetically different from the rest of the studied population, as this is fairly common in sessile ocean organisms, even at distances of a few kilometers [31, 32].

Effects of currents and distance on population substructure

Previous studies have shown that isolation-by-distance (IBD) is an important driving force of genetic substructure between *S. latissima* populations across moderate to large distances [23, 25, 26, 33–35]. At shorter distances, on the other hand, significant correlations between genetic substructure and physical distance have not been observed [26, 37]. In the current study significant IBD was detected, but it only explained around 15% of the genetic variation, indicating that additional causative factors are important. In a study of *S. latissima* in the Northern Irish Sea, combining genotyping and hydrodynamic modelling, Mooney et al. [37] argued that local currents drive connectivity at small distances. This is also supported by early spore dispersal experiments that have revealed that kelp spores do not travel far from their release location under calm conditions [38–40]. In our study, inclusion of predicted currents increased the fraction of genetic substructure explained from ~ 15% in the IBD model to ~ 27% in the "isolation-by-oceanography" (IBO) model, supporting the notion that local currents play an important role in driving genetic connectivity between populations.

The distance measures used in our IBO model were chosen to make it as comparable to the IBD model as possible, but there are several ways to represent both physical and genetic distances that could potentially further increase the correlation coefficient of the model. One interesting aspect of spore dispersal simulations as a measure of distance is that they are directional. Directionality was not yet implemented in our IBO model, but by using directional measures of genetic distance, such as the estimate of directional migration proposed by Sundqvist et al. [66], this additional dimension could be included in future isolation models based on dispersal simulations. In the seagrass species *Zostera marina*, strong correlation between directional migration and simulated dispersal probabilities have been observed [41]. When gene flow is stronger in one direction one would expect populations with limited gene flow *out* to have more private alleles than other populations. This study revealed coherence between the predicted particle flow between sampling locations and the number of private alleles. That is, sampling location 7, with the least predicted particle flow *out*, contained more than half of all private alleles observed in all 10 locations together.

Furthermore, connectivity and thus genetic exchange and population structure are not only driven by particle transport but also by (a)biotic environmental factors and biological features of the particles themselves. The SINMOD system that we have used reproduces circulation dynamics realistically in the studied area [67], and therefore physical dispersal patterns are expected to be realistic. However, the particle tracking module does not currently account for water temperature or biological factors such as the biphasic life cycle of Laminariales, where dispersal can occur at the spore stage, gametophyte stage and in the form of sporophyte fragments with

reproductive tissue [68, 69]. Other biological factors are the size, weight, number, and lifespan of reproductive particles as well as interactions with other species, *e.g.*, sea urchins grazing on, and potentially translocating, gametophytes [70]. The spore dispersal simulation results in this study should therefore be interpreted as relative potentials for transportation rather than an estimate of actual gene flow. With increasing knowledge of kelp dispersal and reproductive biology, more detailed dispersal models could be constructed, potentially improving gene flow predictions further.

Conclusions and outlook

We have conducted a population genetic analysis of a natural *S. latissima* population at a finer geographical scale than what has been done before, using both SNPs and short read-backed haplotypes, and both including and excluding rare alleles. We demonstrate that the variability in numerical estimates of population genetic metrics achieved using different marker types and rare allele filtering thresholds should be considered when performing or comparing population genetic studies. Notably, higher MAC thresholds and the use of haplotypes instead of SNPs generally yielded increased absolute numerical values of the estimated population genetic measures (H_E, H_O, F_{IS}, F_{GRM}, and F_{ST}). Our findings also show that haplotypes yield comparable results to SNPs and exhibit enhanced statistical power in detecting genetic substructure. Filtering out rare alleles enhanced resolution in detecting outbreeding and subdivision within the population, albeit at the expense of losing real genetic information such as private alleles.

The *S. latissima* population near Inntian island in Mid-Norway is genetically diverse and slightly outbred. Low F_{ST} estimates indicate that the level of outcrossing between sampling locations is large enough to keep the subpopulations far from fixation. This is supported by the particle dispersal model that predicts particle flow between all sampling sites within one generation. Despite the high connectivity in the studied area, we were able to detect subtle, but significant genetic subdivision between most pairs of sampling locations, particularly between sampling location 7 and the remaining sampling locations. The use of a GRM additional to more widely used population genetic methods revealed inbred individuals and groups of related individuals from sampling location 7, indicating a slight isolation of that location over more than one generation. The observed genetic subdivision between locations in the studied area was partly explained by physical distance, but a larger part of the genetic differences was explained when predicted currents were taken into consideration, suggesting that local currents play a role in driving genetic subdivision on a fine geographic scale.

Read-backed haplotype calling and permutation testing of population genetic metrics can facilitate monitoring of subtle genetic changes in wild populations near cultivation sites, which has been recommended for sustainable management of genetic resources in the development of the kelp cultivation industry [71], as well as monitoring of genetic changes in response to various stressors, including increasing ocean temperatures and sea urchin grazing. These methods, combined with predictions of particle dispersal, can be further refined and implemented in development of models predicting genetic introgression of cultivated kelp into natural populations, necessary for knowledge-based regulations and policies regarding the extent and localization of kelp cultivation as well as the sourcing of genetic material for cultivation. The low degree of substructure in the population around Inntian indicates that individuals from any of the sampling locations can be used as genetic material for kelp cultivation within the studied area without risking considerably changing the genetic composition of the local population as a whole. A similar study spanning a larger area, and including additional metrics used in conservation genetics (e.g., Jost et al. 2017 [72]) would be required to determine how far one could go.

Methods

Sample collection, DNA extraction and sequencing

In September 2019, 16–17 randomly chosen mature sporophytes were collected from each of ten locations within an area of approximately 50 km^2 at the subtidal zone around the Inntian island, Frøya in Trøndelag county, Norway (Fig. 1). Tissue was sampled from the basal blade (meristematic region) of individual sporophytes and frozen at –20˚C prior to freeze-drying.

Approximately 10 mg of dried biomass per individual was crushed to a fine powder using 2 ml Eppendorf tubes and 3 mm tungsten beads in a TissueLyser homogenizer. Tubes were shaken for 4×30 s at 25 Hz. Then, 700 µL of lysis buffer (100 mM Tris–HCl pH 8, 3% hexadecyltrimethylammonium bromide (CTAB), 1.4 M NaCl, 20 mM EDTA pH 7.5 – 8, 2.5% polyvinylpyrolidone (PVP-40) [73], and 10 µL of RNaseA was added to each tube followed by shaking for 2×30 s at 25 Hz. After incubation for 2 h at room temperature, 700 µL cold chloroform was added, and the tubes were vortexed for 10 s. The tubes were then centrifuged for 15 min at 20,000 g and 4˚C, and the upper (aquatic) phase was transferred to new tubes. Isopropanol (450 µL) was added and the contents were mixed by turning each tube 10 times prior to incubation at room temperature for 1 h. Subsequently, the tubes were centrifuged at 20,000 g and 4˚C for 20 min,

and all the supernatant was discarded. The pellets were washed twice by adding 1 mL 70% ethanol, vortexing for 10 s, centrifuging at 20,000 g and 4˚C for 5 min and discarding the liquid. The tubes were left open for 20 min at room temperature to dry the pellets and finally, the pellets were resuspended by adding 100 µL sterile filtered 10 mM Tris–HCl pH 8.5, incubating at 65˚C for 10 min, vortexing and incubating for another 10 min at 65˚C.

The samples were stored at -20˚C until shipment to LGC Genomics (Berlin, Germany) for preparation of genotyping-by-sequencing (GBS) libraries and sequencing. This was done according to LGC Genomics procedure for normalized GBS (nGBS) [74] and included double enzyme digestion with PstI and MseI, adapter ligation, PCR amplification, molecular normalization, size selection, and 150 bp paired-end (PE-150) Illumina sequencing on a NovaSeq instrument.

SNP and haplotype calling

GBS locus identification and SNP and haplotype calling were conducted using the workflow shown in Fig. 8. First, the raw GBS data were preprocessed using GBprocesS v3.0.4 [75] with default settings. The preprocessing involved demultiplexing, adapter trimming, barcode and spacer sequences, merging corresponding forward and reverse reads using PEAR v0.9.8 [76], and quality filtering. Additionally, reads with internal PstI and/or MseI restriction sites were removed because they could result from partial digests or be chimeras created by genomic fragments that ligated to each other instead of to the adapters during GBS library preparation. The trimmed and merged sequencing data is available in NCBI SRA (BioProject PRJNA1037756). A draft de novo reference genome sequence was created by WGS shotgun Illumina sequencing (PE-150) of genomic DNA (BioProject PRJNA1039286) extracted from one of the individuals in the sample set and DeBruyn graph assembly was performed using CLCbio Genomics Workbench (v20.0.2) with default settings were used. Then, the GBS reads were mapped to the de novo assembly using BWA-MEM in BWA 0.7.17 [77] with default settings. Only unambiguously mapped reads (mapping quality of at least Q30) were retained.

Then, SMAP *delineate* [44] was used to identify 10 individuals with low coverage (less than 60% coverage before removing low coverage loci). These individuals were removed, leaving 150 individuals in the dataset. Moreover, SMAP *delineate* was used to select high-quality GBS loci according to the following criteria: locus length of 100–275 bp, minimum read depth of 10 per locus per individual and sample completeness of at least 95% (from here on called "HQ GBS loci"). SNPs in HQ GBS loci were called using GATK UnifiedGenotyper [78]. Then, SNPs were filtered using VCFtools 0.1.16 [79] with its options for sequencing quality (Q/GQ), read depth (DP) and minor allele count (MAC) set as: *min-meanDP* 30, *mac* 3(15) and *minQ* 20 and then *minDP* 10 and *minGQ* 30. SNPs were filtered for minor allele count (MAC) at two different levels: MAC 3 and MAC 15, which in these data corresponded to minor allele frequencies of approximately 1% and 5%, respectively. Indels and multi-allelic SNPs were then removed

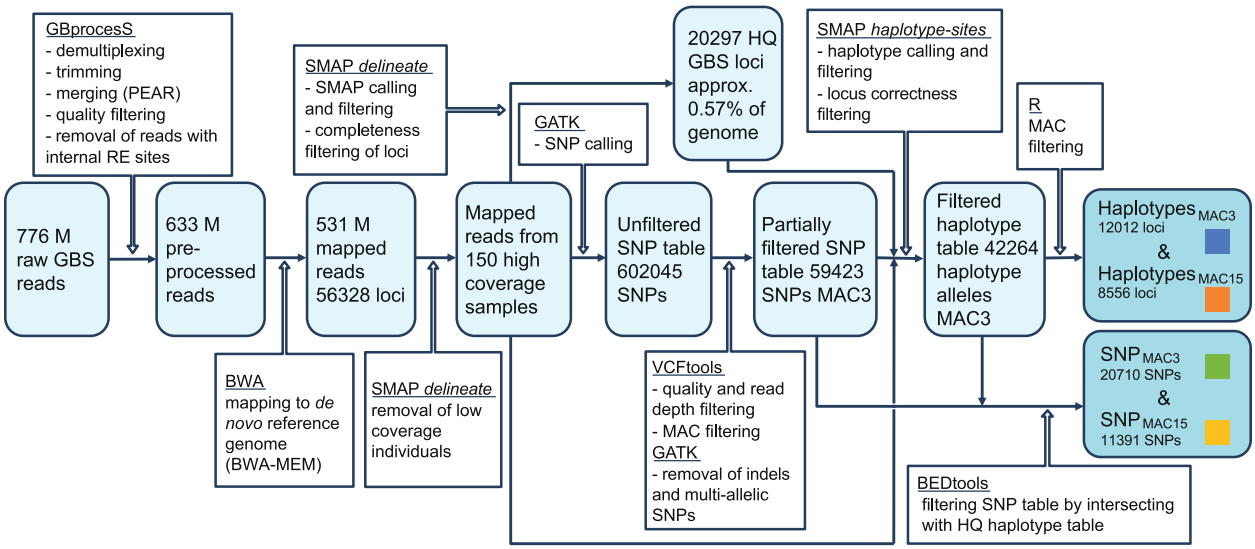


Fig. 8 Bioinformatics workflow from raw GBS reads to fully filtered SNP and Haplotype datasets. White boxes show processing steps, light blue boxes show intermediate files with number of reads, samples, loci, SNPs or haplotype alleles indicated, and dark blue boxes show the final files that were used for genetic analyses

using GATK. Short read-backed haplotypes were called using the SMAP *haplotype-sites* module [44] with recommended settings and mapped reads (excluding low coverage individuals), HQ GBS loci and the SNP table filtered at MAC3 as input. Short read-backed haplotypes were defined by joining the polymorphisms across the length of a given GBS locus (100–275 bp), as defined by SMAP *delineate,* and included stack mapping anchor points (SMAPs [44]) and the selected SNPs. These haplotypes can include combinations of several polymorphisms each, and therefore be either bi-allelic or multi-allelic. For each individual and locus, haplotype frequencies were transformed to discrete dosage calls using default settings for diploids (see [80]), generating a genotype call per haplotype per individual with one of the following states: absent (0), heterozygous (1) or homozygous (2). During the haplotype calling, only loci with high completeness (>95% across the sample set), and with high genotype call quality were retained (*i.e.,* loci with a total allele dosage of 2 (as expected for a diploid) in more than 95% of the samples (>95% correctness)). This filtering process also provides a means to select high quality SNP sites (additional to GATK and VCF filtering). We therefore only retained SNPs with positional overlap with high-quality (HQ) haplotype loci using BEDtools intersect. This also ensured that the SNP and haplotype data sets were compared on the same set of genomic loci. The haplotypes were filtered for MAC in R v.4.2.3 [81] at the same thresholds as the SNPs (MAC3 and MAC15). Additionally, markers that were not absent, i.e., markers that were not sequenced or filtered out due to low quality, poor read count etc., in at least 10 individuals from any sampling location were removed from the SNP and haplotype datasets using R. All these filtering steps resulted in four datasets: SNP_{MAC3}, SNP_{MAC15}, $Hapotype_{MAC3}$ and $Haplotype_{MAC15}$.

Numerical spore dispersal modelling

Spore dispersal modelling was performed using the 3-dimensional biophysical ocean model system SINMOD [82]. SINMOD solves the primitive Navier–Stokes equations using a finite difference scheme on an Arakawa C-grid in *z*-coordinates with a hydrostatic assumption. A model domain of 160 m horizontal resolution, with vertical resolution (layer thickness) ranging from 1 to 5 m for the upper 40 m of the water column to 25 m deeper down, was set up for the region of Central Norway. Application of atmospheric forcing, diffuse and riverine freshwater outflow, and nesting from coarser model setups are all detailed in Broch et al. [83]. The model system has been shown to reproduce circulation dynamics on the Norwegian shelf [84] and in complex coastal regions in higher resolutions [67] in a realistic manner. A Lagrangian particle tracking module using a 4th order Runge–Kutta numerical scheme was used to simulate spore dispersal. The model was run online with the hydrodynamic model, i.e., updated with the basic simulation time step of 30 s. Numerical particles were released from each of the 10 sampling locations in the model grid cell closest to the bottom every 0.5 h and were tracked for the time period of November–December 2018, the expected period of spore release for the parents of the studied population. A total of 2882 particles were released from each site. The particles were assumed to follow the water currents passively and in particular were not assumed to have any biological traits such as swimming speed or lifetime.

A particle was assumed to hit a sampling location if it entered a circle of radius 2 grid cells (=320 m, the hit radius) around that location and was counted only once. A quantification of water connectivity between sampling locations was obtained by considering the fraction C of particles from site a reaching site b of the total number particles released from site a. The greater the fraction, the closer the sites are. This relation quantity is not symmetric, i.e., the distance from a to b may be different from the distance form b to a.

Analysis of genetic diversity

To assess the genetic diversity of the population we calculated the SNP density, the fraction of polymorphic GBS loci, the allelic richness (AR), expected heterozygosity (H_E), and observed heterozygosity (H_O). SNP density and the fraction of polymorphic GBS loci were calculated using the HQ GBS loci (see above). The SNP density was calculated as the number of polymorphic SNP sites (in SNP_{MAC3} and SNP_{MAC15} separately) divided by the total number of nucleotides covered by reads in HQ GBS loci. The fraction of polymorphic GBS loci was calculated as the number of HQ GBS loci with at least one SNP or at least two distinct haplotype alleles, divided by the total number of HQ GBS loci.

The overall distribution of the number of haplotype alleles per locus and location-specific allele richness (AR), defined as the average number of alleles per locus, were calculated based on the datasets $Haplotype_{MAC3}$ and $Haplotype_{MAC15}$. Confidence intervals (95%) for AR were obtained by bootstrapping over loci. That is, the R (4.2.3) function boot() [81] was used to make 20,000 bootstrap samples by selecting m random GBS loci (m is the number of loci in the dataset) with replacement from the haplotype dataset and calculating AR for these samples. Then, a 95% confidence interval was obtained from the 2.5 and 97.5 percentiles of the bootstrap AR values. The confidence intervals were used to assess the significance of pairwise differences in AR. The distribution of

SNPs per GBS locus was calculated based on the datasets SNP_{MAC3} and SNP_{MAC15} using R [81].

Expected heterozygosity (H_E), on the basis of SNPs and haplotypes, was calculated according to the HWE distribution [85] and averaged across all SNPs or haplotype loci (Additional file 1, Equation A1).

Observed heterozygosity (H_O) was averaged across all loci (or SNPs) for each sampling location and for the total population. Confidence intervals for H_E and H_O were calculated in the same way as for *AR* and the significance of pairwise differences in H_E and H_O between sampling locations was assessed using the confidence intervals.

Analysis of population structure

The fixation index, F_{ST}, was used to measure the level of genetic substructure between sampling locations and pairwise F_{ST} was used as a measure of substructure between pairs of sampling locations. To ensure comparability between SNPs and haplotypes, global and pairwise F_{ST} were estimated using Hedrick's [57] standardized estimate of F_{ST}, G'_{ST}, which is a standardized version of Nei's (1973) G_{ST} [86] that allows for comparison between markers with different numbers of alleles. Because the number of subpopulations (sampling locations) in our study was small, and because we compared pairs, the bias correction suggested by Nei [52] was applied to calculate G''_{ST} [53] (Additional file 1, Equations A2-A3). F_{ST} estimates were averaged over all loci (or SNPs) by averaging the numerator and denominator before division, as recommended by Bhatia et al. [63]. The significance of the F_{ST} estimates was tested by Monte Carlo permutation testing as follows: All individuals in the total population were randomly drawn without replacement into 10 new subpopulations of equal sample size as the real subpopulations 10,000 times. For each permutation, F_{ST} estimates were calculated for the total population and on the pairwise level to make a set of null distributions. The F_{ST} estimates were compared to their corresponding null distributions to determine the level of significance.

Isolation-by-distance (IBD) was assessed by plotting genetic distance between pairs of sampling locations measured in linearized pairwise F_{ST}, i.e., pwG''_{ST}/(1- pwG''_{ST}), against the logarithm of the physical distance, as proposed by Rousset [87]. Physical distance was measured in kilometer water distance, and the logarithm of the physical distance was used because the sampling locations are spread across a two-dimensional area, not along a line, and the distances are short [87]. Moreover, the linearized F_{ST} estimates were plotted against the oceanographic distance, D_O, and, since the two-dimensionality of the area was already taken into account in the spore dispersal model, the oceanographic distance was implemented directly instead of using logarithmic distance in the model. Oceanographic distance between two locations *a* and *b* was defined as:

$$D_O = \frac{1}{C_{ab} + C_{ba}}, \quad (1)$$

where C_{ab} is the connectivity from *a* to *b*, i.e., the fraction of particles released from location *a* in the spore dispersal model that reached location *b*, and C_{ba} is the connectivity from *b* to *a*. A simple Mantel test [88] (Additional file 1, Equation A4) with 50,000 resamplings was used to assess the significance of IBD and IBO. To further assess population division, the number of private alleles in each sampling location was counted. Private alleles were defined as alleles that were only present in one sampling location and calculated based on the four datasets using R [81].

Genomic relationship matrix

To investigate the genetic substructure in the population on the individual level, a genomic relationship matrix (GRM) was constructed for each of the four datasets. For bi-allelic SNPs, a GRM can be obtained using VanRaden's [54] first method, but since the short read-backed haplotypes were both bi- and multi-allelic, VanRaden's first method could not be applied directly and was slightly modified to accommodate multi-allelic markers. To avoid consequent underestimation of genomic covariances due to unknown genotypes (no-calls) at some loci, the matrix was scaled by the number of loci present in each pair of individuals (Eq. 2).

$$G_{ij} = \frac{\mathbf{X}_i \mathbf{X}'_j}{2\sum c_{il} c_{jl} p_l (1 - p_l)} \quad (2)$$

Where $\mathbf{X}_i$ and $\mathbf{X}_j$ are centered row-vectors of genotypes for all SNP loci (in the SNP datasets) or of haplotype alleles (in the haplotype datasets), i.e., the entries of $\mathbf{X}_i$ is $\alpha_{li} - 2p_l$, where α_{li} is either the copy number of the non-reference allele at SNP loci *l* of individual *i* (in the SNP datasets) or the copy number of the haplotype allele *l* in individual *i* (in the haplotype datasets), and p_l is either the frequency of the non-reference allele at SNP loci *l* (in the SNP datasets) or the frequency of haplotype allele *l* (in the haplotype datasets). No-calls are set to an arbitrary value in $\mathbf{X}_i$ and $\mathbf{X}_j$ and c_{il} and c_{jl} are indicators taking the value 1 if the genotype is observed for SNP locus or haplotype *l* of the individual (*i*/*j*), and 0 otherwise.

Clustering genomic relationships

The GRM was clustered using agglomerative hierarchical clustering. First, the GRMs were converted to distance objects using the built-in R function dist() with euclidean distance [81]. Then, the function hclust() from the R

package dendextend [89] was used to perform hierarchical clustering. Complete linkage was used, meaning that the linkage distance between two subsets was defined by the distance between the two individuals in the subsets that were furthest apart. The number of clusters, k, was determined by maximizing the mean silhouette [90] (Additional file 1, Equation A5). The mean silhouette across all individuals, S_k, was calculated for different numbers of clusters, $k = [2, 15]$. Out of those k that gave informative clusters the k with the highest mean silhouette was chosen for each dataset.

Analysis of inbreeding

The level of inbreeding in each sampling location and in the total population was assessed using two different measures of inbreeding: the inbreeding coefficient (F_{IS}) [55], and the mean of the average individual inbreeding coefficients from the GRM (F_{GRM}). F_{IS} was estimated for each sampling location using Nei's (1987) estimator of F_{IS} [55] (Additional file 1, Equation A6) and averaged across sampling locations. The significance of the F_{IS} values was assessed by adapting the Monte Carlo permutation method suggested by Li et al. [91]. That is, random mating was simulated by splitting all allele pairs into two random gametes for each individual and then drawing random pairs of gametes without replacement until all gametes were drawn. Simulations were repeated 10,000 times, and each time F_{IS} values were calculated. The simulated F_{IS} values made up a null distribution that was used for significance testing. This was done for all sampling locations and for the total population. Since the relative phases of markers on different reads were unknown, a 50% chance of changing the phase between reads was implemented. Pairs of SNPs that were on the same read were assumed to be in close linkage with no recombination, hence the relative phase of SNPs on a single read was not changed.

F_{GRM}, were found by subtracting 1 from the diagonal of the GRMs [54], which corresponds to the genomic variance of an individual. Mean F_{GRM} values were calculated for each sampling location and for the total population. The significance of the mean F_{GRM} was assessed by implementing a one-tailed t-test with 50,000 bootstrap resamplings using the built-in R function t.test() [81] and the R function boot() from the R package "boot" [92, 93].

Abbreviation

AMOVA	Analysis of molecular variance
AR	Allelic richness (here measured as the average number of alleles per locus)
BWA-MEM	Burrows-Wheeler alignment - maximum exact matches
F_{GRM}	Individual inbreeding coefficient from the diagonal of a GRM
F_{IS}	Inbreeding coefficient of a population based on excess of homozygotes
F_{ST}	Wright's fixation index
GBS	Genotyping-by-sequencing
G_{ij}	Genomic covariance calculated based on SNPs
GRM	Genomic relationship matrix
G_{ST}	Nei's (1973) estimate of F_{ST}
G'_{ST}	Hedrick's standardized F_{ST} estimate
G''_{ST}	Hedrick's standardized F_{ST} estimate with Nei's (1987) bias correction
H_E	Expected heterozygosity at HWE
H_O	Observed heterozygosity
H_S	H_E in a subpopulation (sampling location)
H_T	H_E when considering all subpopulations as one
HQ	High-quality
HWE	Hardy-Weinberg equilibrium
IBD	Isolation-by-distance
IBO	Isolation-by-oceanography
MAC	Minor allele count
MAF	Minor allele frequency
nGBS	Normalized GBS
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
PE	Paired-end
PEAR	Paired-end read merger
pwF_{ST}	Pairwise F_{ST}
pwG''_{ST}	Pairwise G''_{ST}
SMAP	Stack mapping anchor point
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeat (microsatellite)
WGS	Whole-genome sequencing
θ_{ST}	Wier and Cockerham's F_{ST} estimate
Φ_{ST}	AMOVA based F_{ST} estimate

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10793-2.

Additional file 1. Supplementary method descriptions, results, figures, and tables.

Additional file 2. GRMs constructed based on all datasets and individual F_{GRM} values.

Acknowledgements

We thank Diogo Raposo and Andreas Lavik at Seaweed Solutions for helping us to collect the sporophytes used in this study.
The authors acknowledge the Orion High Performance Computing Center (OHPCC) at the Norwegian University of Life Sciences (NMBU) for providing computational resources that have contributed to the research results reported within this paper. URL (internal): https://orion.nmbu.no.

Authors' contributions

ÅE coordinated the research; FG and ÅE participated in collection of sporophytes and extracted the DNA; TR performed the SNP and haplotype calling; JØ developed the method for constructing genomic relationship matrices from short read-backed haplotypes; SB performed the genetic analyses with supervision from GK, ÅE and TR; OJB performed the numerical spore dispersal modelling; SB wrote the paper with some input, supervision, and/or revision from JØ, OJB, ÅE, TR, GK and FG. All authors have read and approved the final version.

Funding

This study was financed by Breed4Kelp2Feed (Research Council of Norway Project No. 280534).

Availability of data and materials

The data generated and analyzed during the current study are available in NCBI SRA. Individual GBS data – BioProject PRJNA1037756. De novo reference sequence assembly – Bioproject PRJNA1039286, BioSample SAMN38208215. To ensure reproducibility and transparency, the R scripts used for statistical analyses are publicly available on https://github.com/signebra/Population-genetics/.

Declarations

Ethics approval and consent to participate

The wild sporophytes were collected by hand from the area around Inntian island, Frøya. According to the Norwegian Environment Agency, collection of wild kelp from open country (such as the coast of Frøya and Inntian Islands) is permitted as long as the harvesting is done sustainably, and the harvested kelp is not used for commercial purposes [94]. Hence the kelp was harvested according to the Norwegian Outdoor Recreation Act.

Consent of publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Plant Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway. [2]Plant Sciences Unit, Flanders Research Institute for Agriculture, Fisheries and Food (ILVO), Caritasstraat 39, 9090 Melle, Belgium. [3]Department of Plant Biotechnology and Bioinformatics, Faculty of Sciences, Ghent University, Technologiepark 71, 9052 Ghent, Belgium. [4]Sintef Ocean, P.O. Box 4762 Torgarden, Trondheim 7465, Norway. [5]Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, P.O. Box 5003, N-1432 Ås, Norway.

Received: 31 October 2023 Accepted: 11 September 2024
Published online: 30 September 2024

References

1. Lane CE, Mayes C, Druehl LD, Saunders GW. A multi-gene molecular investigation of the kelp (*Laminariales*, *Phaeophyceae*) supports substantial taxonomic re-organization 1. J Phycol. 2006;42(2):493–512.
2. Purcell-Meyerink D, Packer MA, Wheeler TT, Hayes M. Aquaculture production of the brown seaweeds *Laminaria digitata* and *Macrocystis pyrifera*: Applications in food and pharmaceuticals. Molecules. 2021;26(5):1306.
3. Kumar D, Pugazhendi A, Bajhaiya AK, Gugulothu P. Biofuel production from Macroalgae: present scenario and future scope. Bioengineered. 2021;12(2):9216.
4. Øverland M, Mydland LT, Skrede A. Marine macroalgae as sources of protein and bioactive compounds in feed for monogastric animals. J Sci Food Agric. 2019;99(1):13–24.
5. Hafting JT, Craigie JS, Stengel DB, Loureiro RR, Buschmann AH, Yarish C, Edwards MD, Critchley AT. Prospects and challenges for industrial production of seaweed bioactives. J Phycol. 2015;51(5):821–37.
6. Sugumaran R, Padam BS, Yong WTL, Saallah S, Ahmed K, Yusof NA. A retrospective review of global commercial seaweed production—current challenges, biosecurity and mitigation measures and prospects. Int J Environ Res Public Health. 2022;19(12):7087.
7. Goecke F, Klemetsdal G, Ergon Å. Cultivar Development of Kelps for Commercial Cultivation—Past Lessons and Future Prospects. Front Mar Sci. 2020;7:110.
8. Shan T, Pang S. Breeding in the Economically Important Brown Alga *Undaria pinnatifida*: A Concise Review and Future Prospects. Front Genet. 2021;12:801937.
9. Olsen Y. Resources for fish feed in future mariculture. Aquac Environ Interact. 2011;1(3):187–200.
10. Bolstad GH, Karlsson S, Hagen IJ, Fiske P, Urdal K, Sægrov H, Florø-Larsen B, Sollien VP, Østborg G, Diserud OH. Front Genet. 2021;7(52):eabj3397.
11. FAO. The State of World Fisheries and Aquaculture 2022. Towards Blue Transformation. In: The State of World Fisheries and Aquaculture (SOFIA). Rome: Food and Agriculture Organization of the United Nations; 2022.
12. Norderhaug KM, Hansen PK, Fredriksen S, Grøsvik BE, Naustvoll LJ, Steen H, Moy FE. Miljøpåvirkning fra dyrking av makroalger—Risikovurdering for norske farvann. Rapport fra havforskningen. 2021;24:18–25.
13. Loureiro R, Gachon CMM, Rebours C. Seaweed cultivation: potential and challenges of crop domestication at an unprecedented pace. New Phytol. 2015;206(2):489–92.
14. Shan T, Pang S, Wang X, Li J, Su L. Assessment of the genetic connectivity between farmed and wild populations of *Undaria pinnatifida* (*Phaeophyceae*) in a representative traditional farming region of China by using newly developed microsatellite markers. J Appl Phycol. 2018;30(4):2707–14.
15. Wernberg T, Coleman MA, Bennett S, Thomsen MS, Tuya F, Kelaher BP. Genetic diversity and kelp forest vulnerability to climatic stress. Sci Rep. 2018;8(1):1851.
16. Christie H, Norderhaug KM, Fredriksen S. Macrophytes as habitat for fauna. Mar Ecol Prog Ser. 2009;396:221–33.
17. Teagle H, Hawkins SJ, Moore PJ, Smale DA. The role of kelp species as biogenic habitat formers in coastal marine ecosystems. J Exp Mar Biol Ecol. 2017;492:81–98.
18. Simonson E, Scheibling R, Metaxas A. Kelp in hot water: I.Warming seawater temperature induces weakening and loss of kelp tissue. Mar Ecol Prog Ser. 2015;537:89–104.
19. Moy FE, Christie H. Large-scale shift from sugar kelp (*Saccharina latissima*) to ephemeral algae along the south and west coast of Norway. Mar Biol Res. 2012;8(4):309–21.
20. Filbee-Dexter K, Scheibling RE. Sea urchin barrens as alternative stable states of collapsed kelp ecosystems. Mar Ecol Prog Ser. 2014;495:1–25.
21. Kvile KØ, Andersen GS, Baden SP, Bekkby T, Bruhn A, Geertz-Hansen O, Hancke K, Hansen JL, Krause-Jensen D, Rinde E. Kelp forest distribution in the Nordic region. Front Mar Sci. 2022;9: 850359.
22. Lee JA, Brinkhuis BH. Reproductive Phenology of *Laminaria Saccharina* (L.) Lamour. (*Phaeophyta*) at the Southern Limit of its Distribution in the Northwestern Atlantic Ocean. Journal of Phycology. 1986;22(3):276–85.
23. Guzinski J, Ruggeri P, Ballenghien M, Mauger S, Jacquemin B, Jollivet C, Coudret J, Jaugeon L, Destombe C, Valero M. Seascape Genomics of the Sugar Kelp Saccharina latissima along the North Eastern Atlantic Latitudinal Gradient. Genes. 2020;11(12):1503.
24. Luttikhuizen PC, van den Heuvel FH, Rebours C, Witte HJ, van Bleijswijk JD, Timmermans K. Strong population structure but no equilibrium yet: Genetic connectivity and phylogeography in the kelp *Saccharina latissima* (*Laminariales*, *Phaeophyta*). Ecol Evol. 2018;8(8):4265–77.
25. Nielsen MM, Paulino C, Neiva J, Krause-Jensen D, Bruhn A, Serrão EA. Genetic diversity of *Saccharina latissima* (*Phaeophyceae*) along a salinity gradient in the North Sea-Baltic Sea transition zone. J Phycol. 2016;52(4):523–31.
26. Ribeiro PA, Næss T, Dahle G, Asplin L, Meland K, Fredriksen S, Sjøtun K. Going With the Flow –Population Genetics of the Kelp *Saccharina latissima* (*Phaeophyceae, Laminariales*). Front Mar Sci. 2022;9:876420.
27. Evankow A, Christie H, Hancke K, Brysting AK, Junge C, Fredriksen S, Thaulow J. Genetic heterogeneity of two bioeconomically important kelp species along the Norwegian coast. Conserv Genet. 2019;20(3):615–28.
28. Guzinski J, Mauger S, Cock JM, Valero M. Characterization of newly developed expressed sequence tag-derived microsatellite markers revealed low genetic diversity within and low connectivity between European *Saccharina latissima* populations. J Appl Phycol. 2016;28(5):3057–70.
29. King NG, McKeown NJ, Smale DA, Wilcockson DC, Hoelters L, Groves EA, Stamp T, Moore PJ. Evidence for different thermal ecotypes in range centre and trailing edge kelp populations. J Exp Mar Biol Ecol. 2019;514–515:10–7.
30. Solas M, Correa RA, Barría F, Garcés C, Camus C, Faugeron S. Assessment of local adaptation and outbreeding risks in contrasting thermal environments of the giant kelp, *Macrocystis pyrifera*. J Appl Phycol. 2024;36(1):471–83.
31. Hays CG, Hanley TC, Hughes AR, Truskey SB, Zerebecki RA, Sotka EE. Local Adaptation in Marine Foundation Species at Microgeographic Scales. Biol Bull. 2021;241(1):16–29.
32. Sanford E, Kelly MW. Local Adaptation in Marine Invertebrates. Ann Rev Mar Sci. 2011;3(1):509–35.
33. Mao X, Augyte S, Huang M, Hare MP, Bailey D, Umanzor S, Marty-Rivera M, Robbins KR, Yarish C, Lindell S, Jannink J-L. Population Genetics of Sugar Kelp Throughout the Northeastern United States Using Genome-Wide Markers. Front Mar Sci. 2020;7:00694.
34. Thomson AI. Population Genomics of the Sugar Kelp, *Saccharina latissima*. Chapter IV: Drivers and Spatial Scales of Local Adaptation and

Connectivity in the Sugar Kelp, *Saccharina latissima*. Inverness: University of the Highlands and Islands; 2021. https://pure.uhi.ac.uk/en/studentTheses/population-genomics-of-the-sugar-kelp-saccharina-latissima.
35. Thomson AI. Population Genomics of the Sugar Kelp, *Saccharina latissima*. Chapter III: Population Genomics Inform the Development of Kelp Cultivation on the West Coast of Scotland. Inverness: University of the Highlands and Islands; 2021. https://pure.uhi.ac.uk/en/studentTheses/populationgenomics-of-the-sugar-kelp-saccharina-latissima.
36. Breton TS, Nettleton JC, O'Connell B, Bertocci M. Fine-scale population genetic structure of sugar kelp, *Saccharina latissima* (*Laminariales*, *Phaeophyceae*), in eastern Maine, USA. Phycologia. 2018;57(1):32–40.
37. Mooney KM, Beatty GE, Elsäßer B, Follis ES, Kregting L, O'Connor NE, Riddell GE, Provan J. Hierarchical structuring of genetic variation at differing geographic scales in the cultivated sugar kelp *Saccharina latissima*. Mar Environ Res. 2018;142:108–15.
38. Gaylord B, Reed DC, Raimondi PT, Washburn L, McLean SR. A Physically Based Model of Macroalgal Spore Dispersal in the Wave and Current-Dominated Nearshore. Ecology. 2002;83(5):1239–51.
39. Gaylord B, Reed DC, Raimondi PT, Washburn L. Macroalgal Spore Dispersal in Coastal Environments: Mechanistic Insights Revealed by Theory and Experiment. Ecol Monogr. 2006;76(4):481–502.
40. Reed DC, Laur DR, Ebeling AW. Variation in Algal Dispersal and Recruitment: The Importance of Episodic Events. Ecol Monogr. 1988;58(4):321–35.
41. Jahnke M, Jonsson PR, Moksnes PO, Loo LO, Nilsson Jacobi M, Olsen JL. Seascape genetics and biophysical connectivity modelling support conservation of the seagrass *Zostera marina* in the Skagerrak-Kattegat region of the eastern North Sea. Evol Appl. 2018;11(5):645–61.
42. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 2011;6(5): e19379.
43. Morin PA, Martien KK, Taylor BL. Assessing statistical power of SNPs for population structure and conservation studies. Mol Ecol Resour. 2009;9(1):66–73.
44. Schaumont D, Veeckman E, Van der Jeugt F, Haegeman A, van Glabeke S, Bawin Y, Lukasiewicz J, Blugeon S, Barre P, Leyva-Pérez MdlO, et al. Stack Mapping Anchor Points (SMAP): a versatile suite of tools for read-backed haplotyping [Internet]. bioRxiv [Preprint]. 2022 [cited 2023 Sep 12]: 9 p. Available from https://doi.org/10.1101/2022.03.10.483555.
45. Rick JA, Brock CD, Lewanski AL, Golcher-Benavides J, Wagner CE. Reference Genome Choice and Filtering Thresholds Jointly Influence Phylogenomic Analyses. Syst Biol. 2023;73(1):76–101.
46. Linck E, Battey CJ. Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. Mol Ecol Resour. 2019;19(3):639–47.
47. Jakobsson M, Edge MD, Rosenberg NA. The Relationship Between FST and the Frequency of the Most Frequent Allele. Genetics. 2013;193(2):515–28.
48. Roesti M, Salzburger W, Berner D. Uninformative polymorphisms bias genome scans for signatures of selection. BMC Evol Biol. 2012;12(1):94.
49. Goecke F, Gómez Garreta A. Martín–Martín R, Rull Lluch J, Skjermo J, Ergon Å: Nuclear DNA Content Variation in Different Life Cycle Stages of Sugar Kelp. Saccharina latissima Marine Biotechnology. 2022;24(4):706–21.
50. ©norgeskart.no: Map of Inntian. Norwegian mapping authority. https://www.norgeskart.no/#!?project=norgeskart&layers=1002&zoom=11&lat=7080976.15&lon=197936.64
51. Leberg PL. Estimating allelic richness: Effects of sample size and bottlenecks. Mol Ecol. 2002;11(11):2445–9.
52. Nei M. Molecular Evolutionary Genetics. New York: Columbia University Press; 1987.
53. Meirmans PG, Hedrick PW. Assessing population structure: FST and related measures. Mol Ecol Resour. 2011;11(1):5–18.
54. VanRaden PM. Efficient Methods to Compute Genomic Predictions. J Dairy Sci. 2008;91(11):4414–23.
55. Nei M. Definition and Estimation of Fixation Indices. Evolution. 1986;40(3):643–5.
56. Li Z, Löytynoja A, Fraimout A, Merilä J. Effects of marker type and filtering criteria on Q (ST)-F (ST) comparisons. R Soc Open Sci. 2019;6(11):190666.
57. Hedrick PW. A Standardized Genetic Differentiation Measure. Evolution. 2005;59(8):1633–8.
58. Thomson AI. Population Genomics of the Sugar Kelp, *Saccharina latissima*. Chapter V: History Matters: Influence and Ascertainment Bias from Divergent Histories in a Population Genomic Comparison of Kelp.: Inverness: University of the Highlands and Islands; 2021. https://pure.uhi.ac.uk/en/studentTheses/population-genomics-of-the-sugar-kelp-saccharina-latissima.
59. Olsen HF, Klemetsdal G. Clustering the relationship matrix as a supportive tool to maintain genetic diversity in the Scandinavian cold-blooded trotter. Acta Agriculturae Scandinavica, Section A — Animal Science. 2020;69(1–2):109–17.
60. Arnaud-Haond S, Stoeckel S, Bailleul D. New insights into the population genetics of partially clonal organisms: When seagrass data meet theoretical expectations. Mol Ecol. 2020;29(17):3248–60.
61. Stoeckel S, Arnaud-Haond S, Krueger-Hadfield SA. The combined effect of haplodiploidy and partial clonality on genotypic and genetic diversity in a finite mutating population. J Hered. 2021;112(1):78–91.
62. Camus C, Solas M, Martínez C, Vargas J, Garcés C, Gil-Kodaka P, Ladah LB, Serrão EA, Faugeron S. Mates matter: gametophyte kinship recognition and inbreeding in the giant kelp, *Macrocystis pyrifera* (*Laminariales*, *Phaeophyceae*). J Phycol. 2021;57(3):711–25.
63. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of rare variants. Genome Res. 2013;23(9):1514–21.
64. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics. 1992;131(2):479–91.
65. Weir BS, Cockerham CC. Estimating F-Statistics for the Analysis of Population Structure. Evolution. 1984;38(6):1358–70.
66. Sundqvist L, Keenan K, Zackrisson M, Prodöhl P, Kleinhans D. Directional genetic differentiation and relative migration. Ecol Evol. 2016;6(11):3461–75.
67. Broch OJ, Klebert P, Michelsen FA, Alver MO. Multiscale modelling of cage effects on the transport of effluents from open aquaculture systems. PLoS ONE. 2020;15(3):e0228502.
68. Norton T. Dispersal by macroalgae. Brit Phycol J. 1992;27(3):293–301.
69. Oppliger LV, Von Dassow P, Bouchemousse S, Robuchon M, Valero M, Correa JA, Mauger S, Destombe C. Alteration of sexual reproduction and genetic diversity in the kelp species at the southern limit of its range. PLoS ONE. 2014;9(7):e102518.
70. Veenhof RJ, Coleman MA, Champion C, Dworjanyn SA. Urchin grazing of kelp gametophytes in warming oceans. J Phycol. 2023;59:838–55.
71. Campbell I, Macleod A, Sahlmann C, Neves L, Funderud J, Øverland M, Hughes AD, Stanley M: The Environmental Risks Associated With the Development of Seaweed Farming in Europe - Prioritizing Key Knowledge Gaps. Frontiers in Marine Science. 2019;6.
72. Jost L, Archer F, Flanagan S, Gaggiotti O, Hoban S, Latch E. Differentiation measures for conservation genetics. Evol Appl. 2018;11:1139–48.
73. Gachon CMM, Strittmatter M, Müller DG, Kleinteich J, Küpper FC. Detection of Differential Host Susceptibility to the Marine Oomycete Pathogen Eurychasma dicksonii by Real-Time PCR: Not All Algae Are Equal. Appl Environ Microbiol. 2009;75(2):322–8.
74. Arvidsson S, Fartmann B, Winkler S, Zimmermann W: Efficient high-throughput SNP discovery and genotyping using normalised Genotyping-by-Sequencing (nGBS). LGC Technical note AN-16110401 2016.
75. GBprocesS [https://gbprocess.readthedocs.io/en/devel/index.html].
76. Zhang J, Kobert K, Flouri T, Stamatakis A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics. 2014;30(5):614–20.
77. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–60.
78. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1110):11.10.11–11.10.33.
79. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156–8.
80. SMAP manual: Haplotyping for different sample types (individual or PoolSeq) [https://ngs-smap.readthedocs.io/en/latest/sites/sites_feature_description/SampleType/index.html].
81. R: A language and environment for statistical computing [https://www.R-project.org/].

82. Slagstad D, McClimans TA. Modeling the ecosystem dynamics of the Barents Sea including the marginal ice zone: I. Physical and chemical oceanography. J Mar Sys. 2005;58(1–2):1–18.
83. Broch OJ, Hancke K, Ellingsen IH. Dispersal and deposition of detritus from kelp cultivation. Front Mar Sci. 2022;9:840531.
84. Skarðhamar J, Svendsen H. Circulation and shelf–ocean interaction off North Norway. Cont Shelf Res. 2005;25(12–13):1541–60.
85. Nei M. Genetic distance between populations. Am Nat. 1972;106(949):283–92.
86. Nei M. Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A. 1973;70(12):3321–3.
87. Rousset F. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. Genetics. 1997;145(4):1219–28.
88. Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Res. 1967;27(2_Part_1):209–20.
89. Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015;31(22):3718–20.
90. Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65.
91. Li R, Wang M, Jin L, He Y. A Monte Carlo Permutation Test for Random Mating Using Genome Sequences. PLoS ONE. 2013;8(8):e71496.
92. Davison AC, Hinkley DV: Bootstrap methods and their application: Cambridge university press; 1997.
93. Canty A, Ripely B. Bootstrap Functions (Originally by Angelo Canty for S). In: CRAN; 2022. Available at: https://cran.r-project.org/web/packages/boot/boot.pdf.
94. Miljødirektoratet (The Norwegian Environment agency). Spørsmål og svar om allemannsretten. 2023. https://www.miljodirektoratet.no/ansvarsomrader/friluftsliv/friluftsliv-og-allemannsretten/ofte-stilte-sporsmal-om-allemannsretten/. Accessed 21 Nov 2023.

Publisher's Note