GENIALG

GENetic diversity exploitation for Innovative macro-ALGal biorefinery

---

Deliverable D6.4" Baseline genetic diversity and gene flow between farms and populations "

Planned delivery date (as in DoA): June 2021

Actual submission date: **August 2021**
Workpackage: WP6

Workpackage leader: CIIMAR

Deliverable leader: CNRS
Version: 1.0

Project co-funded by the European Commission within the Horizon 2020 Programme (2014 - 2020)	
Dissemination Level	
PU Public	**x**
CI Classified, as referred to Commission Decision 2001/844/EC	
CO Confidential, only for members of the consortium (including the Commission Services)	

Research and Innovation action: GA no. 727892

D6.4 – Baseline genetic diversity and gene flow between farms and populations (48+6) – CNRS –
Lucie Jaugeon, Paolo Ruggieri, Stéphane Mauger, Jérôme Coudret, Mark Cock, Philippe Potin, Christophe Destombe, Myriam Valero

The aims of this deliverable is to characterize wild and farmed strains of the kelp *Saccharina latissima* through the usage of genetic tools (molecular markers) around the European waters. The molecular data are used to quantify the risk of cross breeding (due to existing gene flow) between wild and farmed strains. Particularly, the main outcome achievable from this task is the quantification of the minimum geographic distance required to avoid gene flow between unrelated populations of *S. latissima,* used as seedlings in kelp farms. This deliverable has been completed and a manuscript is in preparation. The draft of the manuscript is given below

INTRODUCTION

The accelerated diversification of the uses of seaweed in the food, pharmaceutical, textile, plastic, biofuel and other industries is accompanied by a major transition in the exploitation of the resource. In order to ensure the sustainability of this resource, the harvesting of natural populations is being replaced by intensive cultivation (Buschmann et al. 2014). New culture methods are currently being developed in seaweed aquaculture (Robinson et al. 2013, Hafting et al. 2015, Garzia-Poza et al. 2020).

The culture of kelp is currently experiencing a considerable boom around the world (Ferdouse et al. 2018). The development of intensive cultivation in open sea can have significant and sometimes irreversible consequences on the environment. In particular, the dispersal of cultivated genotypes and the putative hybridization of these farmed seaweeds with native stocks can result in an impoverishment of local genetic diversity (Hutchings & Fraser 2008, Liu et al. 2012). The introgression of crop genetic material into wild populations which are widely documented in terrestrial agriculture and animal aquaculture (Manchester & Bullock, 2000; Jiang et al., 2012; Ellstrand et al., 2013) remains poorly known in seaweed aquaculture (Goeke

et al. 2020). This phenomenon of genetic pollution is of great concern and can have profound impact on genetic diversity and evolution of natural populations (Loureiro et al. 2015). Cultivated individuals may provide a selective advantage when they spread to the natural environment (Williams and Smith, 2007).

Draft policies in Europe prohibit the use of non-native species and limit the use of non-native genotypes in seaweed aquaculture (Loureiro et al., 2015; Barbier et al., 2019). In kelps, dispersal distances are generally reported to be very limited (Santelices, 1990) and new individuals recruit within a few meters of the parental sporophyte (Sundene, 1962; Anderson & North, 1966; Dayton, 1973). In the cultivated kelp *Saccharina japonica,* Shan and collaborators (2019) demonstrated a very limited gene flow from farms to the adjacent spontaneous populations, but in the contrary direction gene flow was more important. These authors suggest that differences would be essentially linked to the different cultural practices exercised by the local farmers.

In the North Atlantic, the sugar kelp, *Saccharina latissima* is considered a good candidate for the development of intensive aquaculture. Until now, cultivation of *S. latissima* has been tested on a small scale at various locations in Europe from the Iberian Peninsula (Freitas et al. 2015; Peteiro & Freire 2013) to Norway (Handa et al. 2013, Fobord et al. 2020), via Sweden (Vish et al. 2020), Faroe Islands (Bak 2019), Denmark (Boderskov et al. 2021) Scotland (Cappuzo & McKie, 2016), Ireland (Mac Monagail & Morisson, 2020) and Brittany (Mesnildrey et al. 2012). In these different small farms, the crop lines were seeded from a mixture of broodstock sporophytic individuals taken from the local indigenous populations (Araujo et al. 2021).

In Europe, S. latissima's range extends from Spitsbergen in the north to the northern part of Portugal in the south. Recent investigations on genetic diversity reported strong genetic structuring for *S. latissima* (Evankow et al. 2019, Guzinski et al. 2016, 2020, Mao et al. 2020), suggesting that long range exchanges of migrants are an unlikely occurrence (Guzinski et al. 2020). However, at small scale (less than a few 10s of km), gene flow appears to be quite extensive, with evidence of connectivity between sites due to local currents (Mooney et al. 2018).

The objective of this study was to estimate the genetic diversity of a few cultivated populations distributed along the European coasts and to compare it with that of local wild

populations using a combination of microsatellite markers. The goal was to be able to assign cultivated populations to wild populations.

MATERIALS AND METHODS

Sample collection

Sporophytic individuals of *Saccharina latissima* were collected in 27 localities (24 natural populations N = 677 and 3 farms N = 101) across the North-Eastern Atlantic between 2017 and 2019 (see figure 1 and table 1). For the farms, sporophytic individuals were harvested randomly along the aquaculture long lines. For each farms, the population of origin of the spropophytes that were used to seed the farms were known, ie: natural population 9 (Saint-Brieuc) for the farm 25 (CWEED); natural population 24 (Frøya) for the farm 27 (SES) and the natural population that was used to seed the farm 26 (SAMS) was not sampled but close to population 16 (Atlantic bridge).In natural populations, sampling was carried out on individuals attached to the substrate in the lower intertidal zone except for the two localities of Piriac-sur-Mer and Noirmoutier (figure 1) where drifting individuals were sampled along the beach. For the farms, random sampling was performed along the aquaculture long lines. A small disc of tissue (2–4 cm^2) was cut at the base of the blade from each of the sampled individuals. Dried tissue samples were preserved in individual zip-locked plastic bags containing silica gel until DNA extraction.

Microsatellites genotyping

Total genomic DNA was extracted from approximately 10-20 mg of silica gel-dried tissue. DNA extraction was performed using the Nucleospin® 96 plant kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's instructions, except for cell lysis that was performed at 65°C for 15 minutes and for the washing steps with buffer PW1 and PW2 that were repeated two times each. These modifications were performed to remove most of the polysaccharides that might inhibit DNA extraction and PCR amplification of microsatellite markers (SSRs).

Amplification and scoring of the SSRs loci were performed as detailed in Guzinski et al. (2016) for 19 expressed sequence tag (EST)-derived SSR loci, and as in Paulino et al. (2016) for 8 genomic SSR loci. Amplification products were separated by electrophoresis on an ABI 3130 XL capillary sequencer (Applied Biosystems, USA). Alleles were sized using the SM594 size standard (Mauger et al., 2012) and scored manually using the software GeneMapper 4.0 (

Table 1: List of samples of *Saccharina latissima*. Site number (#) and name, country, latitude (Lat) and longitude (Lon) in decimal degrees, sample size (N), standardized allelic richness (Â) and standardized number of private alleles (PÂ) for a common sample size of 10 individuals, expected heterozygosity (He), FIS multilocus estimates (* indicates significant deviation from Hardy–Weinberg expectations), selfing rate (s) (¥ indicates a significant heterozygosity disequilibrium) and modified index of association (rbarD) (¤ indicates significant linkage disequilibrium).

#	Name	Country	Lat	Lon	N	Â	PÂ	He	Fis	s	rbarD	Date of collection	Collectors
1	Amorosa	Portugal	41.6418	-8.8233	22	1.9 ± 0.08	1.4 ± 0.55	0.171	0.042	0	-0.002	03/02/2018	I. Azevedo
2	Ría de Muros y Noya	Spain	42.7866	-8.9661	20	3.15 ± 0.14	5.65 ± 1.58	0.319	0.121*	0.209¥	0.021	20/02/2019	I. Azevedo
3	Piriac-sur-Mer	France	47.3818	-2.5386	20	2.6 ± 0.1	1.99 ± 1.19	0.257	0.176*	0.099	-0.005	06/12/2018	A. Peters
4	Noirmoutier	France	46.9912	-2.3027	19	2.57 ± 0.11	2.76 ± 1.24	0.239	0.122*	0.084	-0.005	22/01/2019	A. Peters, A. Berthou, K. Saless
5	Locmariaquer	France	47.5570	-2.9256	39	2.76 ± 0.14	0.54 ± 0.65	0.246	0.029	0.015	-0.008	04/12/2017 & 21/04/2019	C. Destombes, J. Dhinaut, M. Valero
6	Concarneau	France	47.8597	-3.9163	35	3.09 ± 0.13	1.13 ± 0.93	0.309	0.056	0.140¥	-0.002	20/12/2017	L. Lévêque, Y. Fontana
7	Parc Marin d'Iroise	France	48.4676	-4.7723	30	3.25 ± 0.14	3.03 ± 1.42	0.311	0.087*	0.042	0.011	18/12/2017	L. Lévêque, W. Thomas, Y. Fontana
8	Roscoff	France	48.7308	-4.0045	49	3.14 ± 0.17	1.56 ± 1.11	0.297	0.032	0.077	0.003	05/12/2017 & 21/02/2019	Y. Fontana, A. Baud, L. Jaugeon , P. Ruggeri
9	Saint-Brieuc	France	48.9095	-3.0664	24	2.92 ± 0.15	1.01 ± 0.79	0.309	0.056	0.069	-0.003	08/11/2018	JF. Arbona
10	Saint-Malo	France	48.6524	-2.0380	19	2.5 ± 0.1	0.29 ± 0.46	0.297	0.040	0.138¥	0.019	05/12/2017	JF. Arbona
11	Porthallow	England	50.0691	-5.0811	10	2.38 ± 0	1.05 ± 0.21	0.264	0.072	0.144	0.063¤	20/03/2019	T. Van Berkel, C. Warwick-Evans
12	Audresselles	France	50.8326	1.5860	23	2.18 ± 0.09	1.2 ± 0.86	0.204	0.006	0	0.007	24/01/2019	J. Coudret, L. Jaugeon , P. Ruggeri
13	Helgoland	Germany	54.1817	7.8915	33	1.8 ± 0.08	0.8 ± 0.76	0.155	0.069	0.030	-0.013	21/03/2019	A. Wagner
14	Ventry Bay	Ireland	52.1157	-10.3675	23	4.21 ± 0.18	7.7 ± 2.12	0.461	0.007	0.038	0.017	20/02/2018	A. Fort
15	New Quay	Ireland	53.1562	-9.0731	21	3.81 ± 0.17	5.53 ± 1.34	0.453	0.071	0	0.012	20/02/2019	A. Fort

16	Atlantic Bridge	Scotland	56.3162	-5.5835	35	3.51 ± 0.14	1.52 ± 1.03	0.404	0.089*	0	0.001	15/12/2017	C. Gachon, MM. Perrineau, V. Montalescot, A. Garvetto, N. Bruderer
17	Eyemouth	Scotland	55.8399	-2.0619	26	2.95 ± 0.1	0.83 ± 0.77	0.405	0.151*	0	0.003	23/01/2019	L. Brunner, S. Reed
18	Shieldaig	Scotland	57.5258	-5.65	23	3.34 ± 0.14	2.35 ± 1.16	0.449	0.118*	0.114¥	0.008	06/12/2018	MM. Perrineau, L. Brunner, G. Misol
19	Portmahomack	Scotland	57.8329	-3.8435	25	3.24 ± 0.11	1.3 ± 1	0.444	0.194*	0.100	-0.005	24/01/2019	L. Brunner, S. Reed
20	Timret	Sweden	58.0636	11.4757	25	2.57 ± 0.08	0.01 ± 0.11	0.374	0.077	0.128¥	-0.006	07/02/2019	W. Visch
21	Kosterhavets Nationalpark	Sweden	58.8355	10.9905	28	2.68 ± 0.11	0.54 ± 0.62	0.359	0.137*	0	0.01	27/01/2018	W. Visch
22	Runde	Norway	62.3987	5.6614	12	2.64 ± 0.06	0.29 ± 0.46	0.347	0.378*	0.182	-0.014	24/06/2018	N. Diehl
23	Trondheim	Norway	63.4526	10.4611	11	3.17 ± 0.06	1.02 ± 0.75	0.405	0.154*	0.038	0.038	03/02/2018	T. Helen Johnsplass
24	Frøya	Norway	63.7032	8.8718	27	3.38 ± 0.14	1.66 ± 1.02	0.435	0.212*	0.061	-0.001	September 2019	Seaweed Energy Solutions
25	C-WEED Farm	France	48.5850	-1.9879	35	2.36 ± 0.15	0.3 ± 0.46	0.247	0.035	0.361¥	0.064¤	29/07/2019	C-WEED Aquaculture
26	SAMS Farm	Scotland	56.4901	-5.4700	25	2.26 ± 0.16	0.47 ± 0.54	0.371	0.146*	0.405¥	0.066¤	End of january to may 2018	Scottish Association for Marine Science
27	SES Farm	Norway	63.7032	8.8718	30	1.94 ± 0.11	0.12 ± 0.34	0.229	0.118*	0.215¥	0.035¤	12/05/2019	Seaweed Energy Solutions

Applied Biosystems, Foster City, USA). Two filters have been applied to the raw multilocus genotypes data-set using the poppr R package (Kamvar et al., 2014). First, only loci and individuals with less than 30 % of missing data were retained (4 loci and 89 individuals were removed) and, second, loci with global minor allele frequency (MAF) of less than 0.01 were discarded (2 other loci were excluded). The final dataset included 599 wild and 90 farmedsporophytes characterized with 21 microsatellite loci (all 18 loci used in Guzinski et al., 2020 except Sacl-32 and Sacl-75, and 5 other ones: Sacl-11/37/54/65/95).
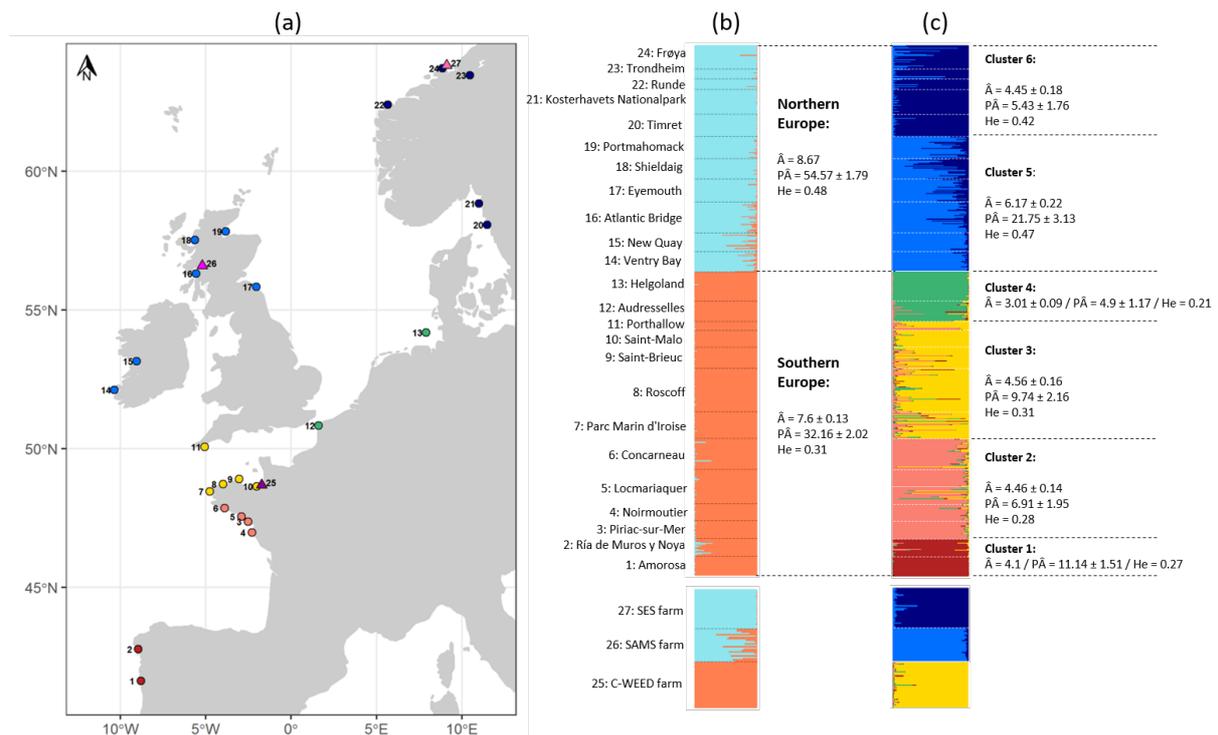
Figure 1: Sampling of *Saccharina latissima* and genetic structure inferred from multi-locus microsatellite genotypes. (a) Sampling locations colored according to genetic structure obtained in STRUCTURE. Circles correspond to natural populations and triangles correspond to farms. (b) Uppermost level of genetic structure, standardized allelic richness (Â), standardized number of private alleles (PÂ) and gene diversity (He) per cluster for a common sample of 256 individuals. (c) Second hierarchical level of structure, Â, PÂ and He per cluster for a common sample of 42 individuals.

Data analysis

Prior to genetic diversity and population structure analysis, loci were tested for stuttering, large allele dropout and null alleles using the software MicroChecker v2.2.3 (Van Oosterhout et al., 2004). The frequency of null alleles per locus and per population was estimated according to the EM algorithm (Dempster et al. 1977) using FreeNA software (Chapuis and Estoup, 2007).

Genetic diversity of natural and farmed populations

Genetic diversity was estimated per population as Nei's gene diversity (expected heterozygosity, He), allelic richness (Â) and number of private alleles (PÂ). These were standardized for the smallest sample sizes in terms of individuals within sites (N = 10), within regions (N = 42) and within clades (N = 256) defined thereafter, using 1000 randomizations (Assis et al. 2016).

Estimation of fixation index (F_{IS}) per population was computed using Genetix 4.05 (Belkhir et al., 2004) and alleles were randomized 10000 times among individuals within each sample to test for deficit in heterozygotes due to potential non-random mating. Threshold for significant tests was set at p-values < 0.01.

Heterozygote deficiencies are often occurring in natural populations even in case of limited selfing, due to technical artefacts such as null alleles (Pompanon et al., 2005). To account for technical biases, population selfing rate "s", derived from "g2", was calculated using RMES software (robust multilocus estimate of selfing) (David et al., 2007). This indice g2 measures the extent to which heterozygosities are correlated across loci. Under no inbreeding, statistical independency of the heterozygosities at different loci is expected. To test the null hypothesis of no heterozygosity disequilibrium (g2=s=0), 1000 permutations of the genetic data were performed.

Linkage among loci is very low or not expected at all for SSRs when populations are sexually reproducing because alleles recombine freely into new genotypes during the process of sexual reproduction. To test this phenomenon, linkage disequilibrium was assessed per population, using a single multilocus measurement of LD, rbarD, a modified index of association which accounts for unequal sample size (Brown et al. 1980 modified by Agapow and Burt, 2001). rbarD was computed using poppr R package (Kamvar et al., 2015). 1000 permutations of the genetic data were realized to test the null hypothesis of no linkage among the different loci.

Genetic structure of natural populations & assignment of farmed sporophytes

The potential effect of null alleles on genetic differentiation was checked by estimating pairwise F_{st} of Weir (1996) both using and without using the ENA correction implemented in the software FreeNA and described in Chapuis and Estoup (2007). The uncorrected and corrected pairwise F_{st} were then compared using a paired t-test.

The dataset was used to perform a series of statistical analyses aimed to detect, the number of genetic clusters composed by the different populations of *S. latissima* sampled in our study.

The information concerning the genetic structure of *S. latissima* across Europe was then used to perform an assignment test to detect unexpected gene flows between populations and between wild populations and farms. Genetic structure was inferred using STRUCTURE 2.3.4 (Pritchard et al., 2000) with admixture and a correlated allele frequency model, without any prior population assignments. STRUCTURE was run in a hierarchal way, evaluating first all populations, and then within the two main clusters South and North Europe separately. A range of assumed populations (K, set sequentially from 1 to 16 for all populations, from 1 to 13 for South Europe, and from 1 to 11 for North Europe) was run 10 times using a burn-in of 5×10^5 iterations and a run length of 1×10^6 iterations. The number of clusters was estimated using the DeltaK criterion of Evanno et al. (2005). Pophelper R package (Francis, 2017) was used to summarize assignment results across independent runs and graphically visualize the results. Genetic structure was also inferred by factorial correspondence analysis (FCA) of population multiscores using Genetix software.

A hierarchical analysis of molecular variance (AMOVA, Excoffier et al., 1992), as implemented in the poppr R package with the function poppr.amova(), was performed using all loci. The different hierarchical levels were defined as followed: the uppermost level of genetic structure corresponds to the two main clades N and S Europe, the second level corresponds to the sub-clusters in each clade, called here regions, the lowest level corresponds to the sampling locations, called here populations. Differences between hierarchical levels were tested by 999 permutations of the data with the function randtest() (ade4 R package, Thioulouse et al. 2018).

Structure analyses were complemented with a discriminant analysis of principal components (DAPC) implemented in the R package adegenet 2.1 (Jombart et al., 2008 & 2010). This approach is free of assumptions about Hardy-Weinberg equilibrium or linkage disequilibrium and allows to identify and describe clusters of genetically related individuals. DAPC analyses were realized twice. In the first analysis, the find.clusters() function was used to determine the number of groups (K) de novo. Up to 24 groups (K) were tested, which corresponds to the number of sampling sites. Optimal K was selected as that with the lowest BIC value. In the second analysis, in order to investigate more precisely the potential source populations of farmed individuals, the dataset was split in two parts: one containing the 599 wild individuals, used as "training dataset", and one containing the 90 farmed individuals, used as supplementary individuals. Sampling locations were used as a priori groups. The assignment

of supplementary individuals to groups using the predict.dapc() function was then examined. The optimal number of PCs to use in the DAPC was determined using the cross-validation method (xvalDapc() command) and corresponded to 80 (number of PCs with lowest RMSE). Finally, individual assignment tests were performed to assign farmed sporophytes to the population they have the highest probability of belonging to, using the software GeneClass2 (Piry et al. 2004). The "assign/exclude population as origin of individuals" option was used with natural populations of Saint-Brieuc (France), Atlantic Bridge (Scotland) and Frøya (Norway) as reference populations. These populations corresponded to the populations where the parents of the farmed sporophytes might come from. The Bayesian method of Rannala and Mountain (1997) was choose to compute the likelihood of a multilocus genotype occurring in a population. The relative assignment scores of the individuals were examined at the first and second ranks. For an individual i in a population l, the score was defined as follow:

$score_{i,l} = \frac{L_{i,l}}{\sum_{i=1}^{k} L_{i,j}}$ with $L_{i,l}$ the likelihood value of the individual i in the population l.

RESULTS

A total of 221 alleles were detected in the dataset sampled throughout the 27 sites (Figure 1a, Table1) resulting from 688 unique multilocus genotypes. The mean number of alleles per locus reaches 10.52 ± 7.93 (ranging from 3 to 33). There was 3.82 % missing data across the final data set. Microchecker showed no signs of stuttering error or large allele dropout but suggested null alleles for 14 out 21 loci. For 9 of them (Sacl-11, Sacl-21, Sacl-37, Sacl-41, Sacl-54, Sacl-78, SLN32, SLN320, SLN510), null alleles were not consistent across populations, with no more than two populations concerned. For the 5 remaining loci (Sacl-56, SLN34, SLN35, SLN36, SLN54), from 3 to 7 populations were concerned but the average frequency of null alleles remained low (< 0.20, Dakin & Avise, 2004). Averaged frequency of null alleles per locus ranged from 0.003 ± 0.013 for locus Sacl-21 to 0.101 ± 0.134 for locus SLN34.

The genotypic analysis using STRUCTURE and FCA revealed two main clusters (clades): (1) Southern Europe and (2) Northern Europe (Figure 1b, Figure 2). The subsequent hierarchical level of genetic structure divided Southern Europe cluster in four sub-clusters and Northern Europe in two sub-clusters. Six sub-clusters (regions) were thus defined: Portuguese and Spanish populations (sub-cluster 1), southern Brittany populations (sub-cluster 2), northern Brittany and English Channel populations (sub-cluster 3) Audresselles and Helgoland
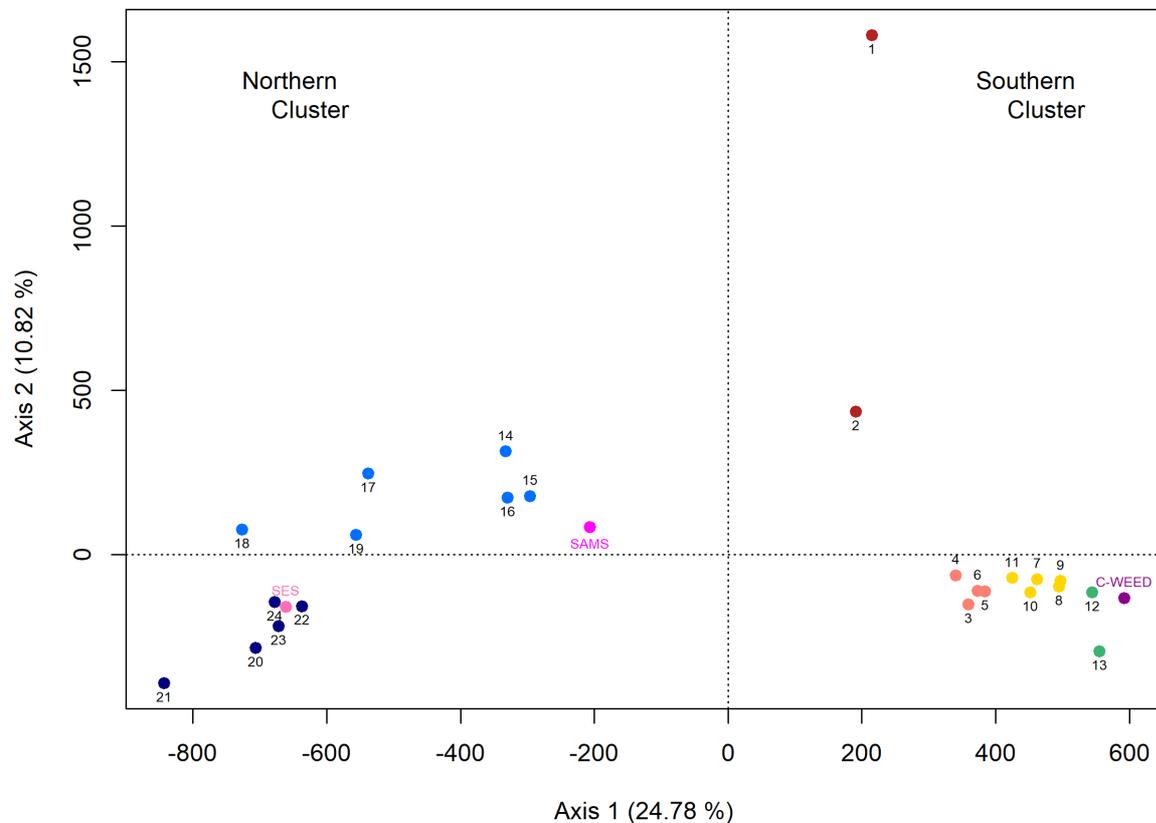
Figure 2: Genetic differentiation of *Saccharina latissima* inferred by factorial correspondence analysis of population multiscores. Numbers refer to sites as listed in Figure 1 and Table 1.

populations (sub-cluster 4), Irish and Scottish populations (sub-cluster 5), Swedish and Norwegian populations (sub-cluster 6) (Figure 1c and supplementary Figure 1 for DeltaK plots). The separation in two main clusters is also supported by the first DAPC analysis using the find.clusters() function, but subsequent hierarchical levels were not well supported. BIC values showed a clear decrease for K=2, but continued to decrease until K=8 and remained more or less stable after (supplementary Figure 2). AMOVA results indicated significant population differentiation at all levels: between clades, between region within clades, between populations within regions and within populations (all p values < 0.001, Table 2). 53 % of the total variance was associated with differences within populations, and 27 % with differences between clades, whereas differences between populations within regions and between regions within clades contributed to 10 % of the total variance each.

Null alleles had some effect on the genetic structure: the ENA method gave slightly, but significantly, lower F_{st} values (average F_{st} with ENA = 0.268 ± 0.132) than those obtained
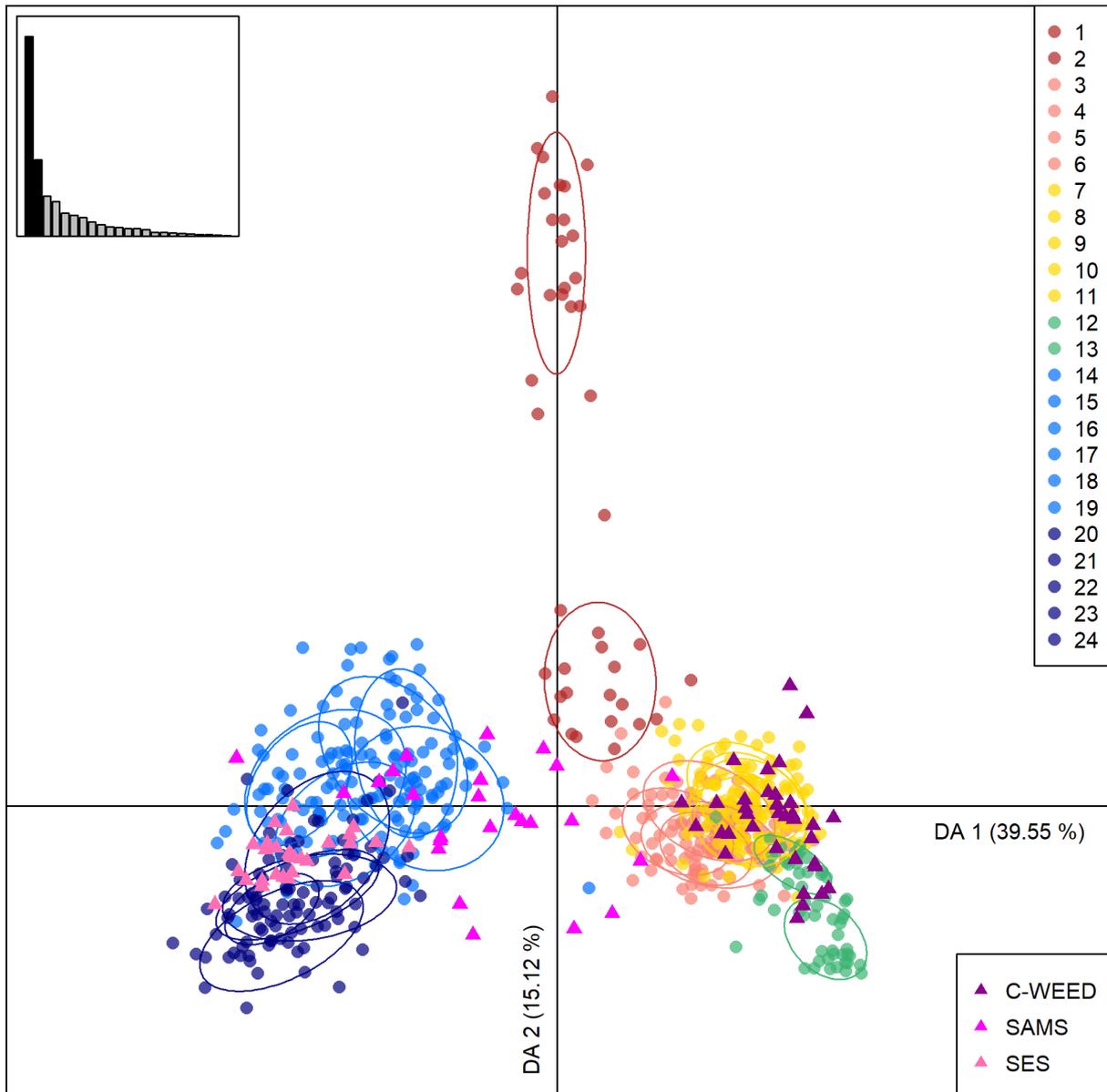
Figure 3: Discriminant Analysis of Principal Components (DAPC) analysis using supplementary individuals. Individuals used in the "training dataset" are represented as dots and the populations as inertia ellipses. Individuals were coloured according to STRUCTURE and FCA results. Eigenvalues of the analysis are displayed in inset. Supplementary individuals from the farms are represented with triangles. Numbers refer to sites as listed in Figure 1 and Table 1.

without ENA correction for the presence of null alleles (average F_{st} without ENA = 0.274 ± 0.133; paired t= 19.253, p < 2.2e-16). However, the most and least differentiated pairwise comparisons and the main pattern of differentiation between northern and southern populations persisted with the ENA correction.

Highest genetic diversity was found in the northern clade with standardized allelic richness (Â) = 8.67, standardized private allelic richness (PÂ) = 54.57 ± 1.79 and expected heterozygosity (He) = 0.48 (Figure 1). Within natural populations, standardized allelic richness (Â) ranged from

1.8 ± 0.08 in Helgoland (Germany) to 4.21 ± 0.18 in Ventry Bay (Ireland) (Table 1). Within farms, Â was comprised between 1.94 ± 0.11 (SES farm, Norway) and 2.36 ± 0.15 (C-WEED farm, France). Standardized private allelic richness (PÂ) varied from 0.01 ± 0.11 in Timret (Sweden) to 7.7 ± 2.12 in Ventry Bay (Ireland) within natural populations and was comprised between 0.12 ± 0.34 (SES farm, Norway) and 0.47 ± 0.54 (SAMS farm, Scotland) for cultivated populations. Expected heterozygosity (He) ranged from 0.155 in Helgoland (Germany) to 0.461 in Ventry Bay (Ireland) within natural populations and from 0.229 in SES farm (Norway) to 0.371 in SAMS farm (Scotland) for cultivated populations. Significant positive F_{is} values were obtained in 12 out of 24 natural populations and in 2 out of 3 farms. Population selfing rate (s), varied from 0 to 0.209 within natural populations and from 0.215 to 0.405 within farms. All the three farms and five natural populations showed significant excess in heterozygotes (Ría de Muros y Noya in Spain, Concarneau in Southern Brittany France and Saint-Malo in Northern Brittany, France, Shieldaig in Scotland and Timret in Sweden). Significant linkage disequilibria were only found within natural population of Porthallow (England) and within all the three farms.

All the three methods used to assign cultivated sporophytes to the natural populations (STRUCTURE, DAPC and GenClass2) suggested that no genetic separation existed between farmed and wild populations. Bayesian approach used in STRUCTURE program revealed that individuals from C-WEED farm were assigned to the southern clade/sub-cluster 3, while individuals from SAMS or SES farms were assigned to the northern clade/sub-cluster 5 and 6 respectively (Figure 1). Second analysis realized with DAPC, using sporophytes from the natural populations as "training dataset" and the cultivated sporophytes as supplementary individuals revealed that the percentage of correctly assigned sporophytes to their parental population is higher for C-WEED farm (91.43 %), than for SAMS farm (76 %) or SES farm (66.67 %) (Figure 3 and 4). Results of assignment tests realized with GeneClass2 showed the same pattern, all the individuals from C-WEED and SAMS farms were correctly assigned to their populations of origin, with really high assignment scores (mean assignment score of 100 % ± 0 for C-WEED farm and 99.9 % ± 0.04 for SAMS farm). In contrast, only 63.3 % of the individuals from SES farm were correctly assigned to their population of origin with an average assignment score of 64.5 % Figure 5).
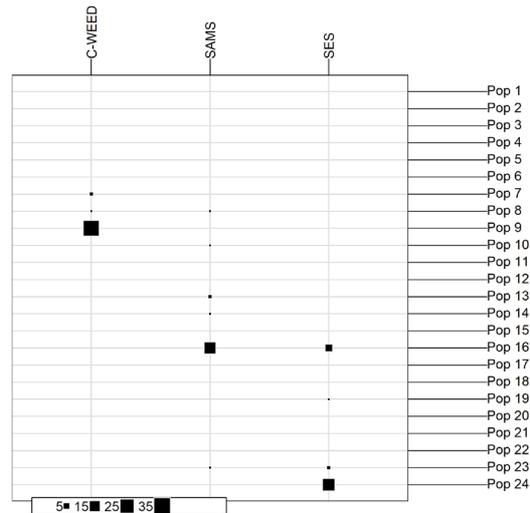
Figure 4: Assignment of the cultivated sporophytes (supplementary individuals) to the natural populations using DAPC. Columns correspond to the farms and rows to the natural populations. Numbers refer to sites as listed in Figure 1 and Table 1.
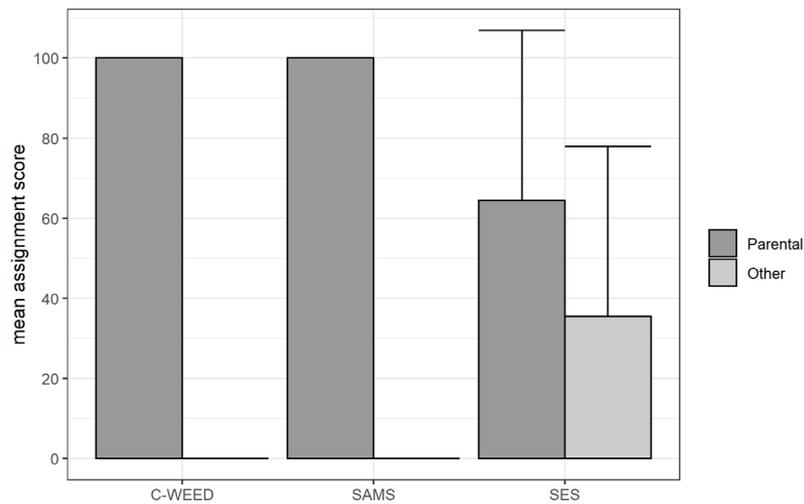
Figure 5: Mean assignment score of the farmed sporophytes from C-WEED, SAMS and SES farms to the parental population.

CONCLUSION

The 27 microsatellite locus developped in this species can be used to quantify the level of gene flow between farms and natural populations at least between the six genetically differentiated European regions. These molecular tools can then be useful to quantify the risk of cross breeding. (2) Farms are not genetically differentiated from their population of origin, (3) Farming have reduced genetic diversity and increased inbreeding compared to the populations of origin.

REFERENCES

Agapow, P. M., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular Ecology Notes*, *1*(1-2), 101-102.

Araújo, R., Vázquez Calderón, F., Sánchez López, J., Azevedo, I. C., Bruhn, A., Fluch, S., ... & Ullmann, J. (2021). Current status of the algae production industry in Europe: an emerging sector of the Blue Bioeconomy. *Frontiers in Marine Science*, *7*, 1247.

Assis, J., Coelho, N. C., Lamy, T., Valero, M., Alberto, F., & Serrão, E. Á. (2016). Deep reefs are climatic refugia for genetic diversity of marine forests. *Journal of biogeography*, *43*(4), 833-844.

Bak, U. G. (2019). Seaweed cultivation in the Faroe Islands: An investigation of the biochemical composition of selected macroalgal species, optimised seeding technics, and open-ocean cultivation methods from a commercial perspective.

Belkhir K., Borsa P., Chikhi L., Raufaste N. & Bonhomme F. 1996-2004 GENETIX 4.05, logiciel sous Windows TM pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5171, Université de Montpellier II, Montpellier (France).

Boderskov, T., Nielsen, M. M., Rasmussen, M. B., Balsby, T. J. S., Macleod, A., Holdt, S. L., ... & Bruhn, A. (2021). Effects of seeding method, timing and site selection on the production and quality of sugar kelp, *Saccharina latissima*: A Danish case study. *Algal Research*, *53*, 102160.

Brown, A. H. D., Feldman, M. W., & Nevo, E. (1980). Multilocus structure of natural populations of Hordeum spontaneum. *Genetics*, *96*(2), 523-536.

Capuzzo, E.; McKie, T. (2016) Seaweed in the UK and Abroad–Status, Products, Limitations, Gaps and CefasRole. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/546679/FC002I__Cefas_Seaweed_industry_report_2016_Capuzzo_and_McKie.

Chapuis, M. P., & Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular biology and evolution*, *24*(3), 621-631.

Dakin, E. E., & Avise, J. C. (2004). Microsatellite null alleles in parentage analysis. *Heredity*, *93*(5), 504-509.

David, P., Pujol, B., Viard, F., Castella, V., & Goudet, J. (2007). Reliable selfing rate estimates from imperfect population genetic data. Molecular ecology, 16(12), 2474-2487.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, *39*(1), 1-22.

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

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479-491.

Francis, R. M. (2017). pophelper: an R package and web app to analyse and visualize population structure. *Molecular ecology resources*, *17*(1), 27-32.

García-Poza, S., Leandro, A., Cotas, C., Cotas, J., Marques, J. C., Pereira, L., & Gonçalves, A. M. (2020). The Evolution Road of Seaweed Aquaculture: Cultivation Technologies and the Industry 4.0. *International Journal of Environmental Research and Public Health*, *17*(18), 6528

Goudet, J. F. S. T. A. T. (1995). FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of heredity, 86(6), 485-486.

Goudet J (2005) HIERFSTAT, a package for R to compute and test hierarchical F-statistics. Mol Ecol Notes 5: 184–186.

Guzinski, J., Mauger, S., Cock, J. M., & Valero, M. (2016). Characterization of newly developed expressed sequence tag-derived microsatellite markers revealed low genetic diversity within and low connectivity between European Saccharina latissima populations. Journal of applied phycology, 28(5), 3057-3070.

Guzinski, J., Ruggeri, P., Ballenghien, M., Mauger, S., Jacquemin, B., Jollivet, C., ... & Valero, M. (2020). Seascape Genomics of the Sugar Kelp Saccharina latissima Along the North Eastern Atlantic Latitudinal Gradient. *Genes*, *11*(12), 1503.

Jombart, T., Lyon, D. & Biome, L. De. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics,24: 1403–1405.

Jombart, T., Devillard, S., & Balloux, F. (2010). Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, *11*(1), 1-15.

Forbord, S., Steinhovden, K. B., Solvang, T., Handå, A., & Skjermo, J. (2020). Effect of seeding methods and hatchery periods on sea cultivation of *Saccharina latissima* (Phaeophyceae): a Norwegian case study. *Journal of Applied Phycology*, *32*(4), 2201-2212.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ, 2, e281.

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in genetics, 6, 208.

Liu, F., Yao, J., Wang, X., Repnikova, A., Galanin, D. A., & Duan, D. (2012). Genetic diversity and structure within and between wild and cultivated *Saccharina japonica* (Laminariales, Phaeophyta) revealed by SSR markers. *Aquaculture*, *358*, 139-145.

Mauger, S., Couceiro, L., & Valero, M. (2012). A simple and cost-effective method to synthesize an internal size standard amenable to use with a 5-dye system. Prime Research on Biotechnology, 2(3), 40-46.

Mac Monagail, M., & Morrison, L. (2020). The seaweed resources of Ireland: a twenty-first century perspective. *Journal of Applied Phycology*, 1-14.

Mesnildrey, L., Jacob, C., Frangoudes, K., Reunavot, M., & Lesueur, M. (2012). Seaweed industry in France. Report Interreg program NETALGAE

Paulino, C., Neiva, J., Coelho, N. C., Aires, T., Marbà, N., Krause-Jensen, D., & Serrão, E. A. (2016). Characterization of 12 polymorphic microsatellite markers in the sugar kelp Saccharina latissima. Journal of applied phycology, 28(5), 3071-3074.

Piry, S., Alapetite, A., Cornuet, J. M., Paetkau, D., Baudouin, L., & Estoup, A. (2004). GENECLASS2: a software for genetic assignment and first-generation migrant detection. Journal of heredity, 95(6), 536-539. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000; 155: 945–959.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nature Reviews Genetics*, *6*(11), 847-859.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics, 155(2), 945-959.

Rannala, B., & Mountain, J. L. (1997). Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences, 94(17), 9197-9201.

Thioulouse, J., Dray, S., Dufour, A. B., Siberchicot, A., Jombart, T., & Pavoine, S. (2018). Multivariate analysis of ecological data with ade4. New York, NY: Springer.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, *4*(3), 535-538.

Visch, W., Kononets, M., Hall, P. O., Nylund, G. M., & Pavia, H. (2020). Environmental impact of kelp (*Saccharina latissima*) aquaculture. *Marine Pollution Bulletin*, *155*, 110962.

Weir B.S. (1996) Genetic data analysis II. Sunderland (MA): Sinauer Associates.