GENIALG

GENetic diversity exploitation for Innovative macro-ALGal biorefinery

Deliverable D2.2" Biobanking of novel Ulva *and* S. latissima *strains with characters of interest for biomass production"*

Planned delivery date (as in DoA): June 2021

Actual submission date: August 2021
Workpackage: WP2

Workpackage leader: CNRS

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 727892 (GENIALG). This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Deliverable D2.2 Part A: *Saccharina latissima*

Biobanking of novel *S. latissima* strains with characters of interest for biomass production

Paulo Ruggeri[1], Lucie Jaugeon[2], Julien Dhinaut[2], Jaro Guzinski[2], Antoine Fort[3], Zofia Nehr[1], Bertrand Jacquemin[2], Aurelien Baud[2], Ousseini Issaka-Salia[2], Marie-Mathilde Perrineau[4], Marion Ballenghien[2], Martina Strittmatter[4], Cecilia Rad-Menéndez[4], Callum O'Connell[4], Carla Ruiz-Gonzalez[4], QianYi Zhang[4], Emilie Gouhier[2], Akira F. Peters[5], Jean-Francois Arbonna[6], Inka Bartsch[7], Klaus Valentin[7], Jerome Coudret[2], Delphine Scornet[1], Stéphane Mauger[2], Stéphane Loisel[2], Komlan Avia[8], Agnieszka Lipinska[1], Olivier Godfroy[1], Christophe Destombe[2], Philippe Potin[1], Ronan Sulpice[3], Claire Gachon[4], Myriam Valero[2], J. Mark Cock[1]

Affiliations:

[1]CNRS, Sorbonne Université, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff, France.

[2]IRL EBEA 3614, Evolutionary Biology and Ecology of Algae, CNRS, Sorbonne Université, UC, UACH, Station Biologique de Roscoff, CS 90074, Place Georges Teissier, 29688 Roscoff cedex

[3]National University of Ireland - Galway, Plant Systems Biology Lab, Ryan Institute & MaREI Centre for Marine, Climate and Energy, School of Natural Sciences, Galway, H91 TK33, Ireland.

[4]Scottish Association for Marine Science, Scottish Marine Institute, Oban, United Kingdom.

[5]Behzin Rosko, Santec, France

[6]C-Weed, 15 rue de l' Herminette 35350 Saint-Méloir-des-Ondes

[7]Alfred Wegener Institute, Am Handelshafen 12, 27570, Bremerhaven, Germany.

[8]Université de Strasbourg, INRAE, SVQV UMR-A 1131, F-68000 Colmar, France

TO BE CITED AS: Ruggeri P, Jaugeon L, DhinautJ, GuzinskiJ, Fort A, Nehr Z, Jacquemin B, Baud A, Issaka-Salia O, Perrinea M-M, Ballenghien M, Strittmatter M, Rad-Menéndez C, O'Connell C, Ruiz-Gonzalez C, Zhang QY, Gouhier E, Peters AF, Arbonna J-F, Bartsch I, Valentin K, Coudret J, Scornet D, Mauger S, Loisel S, Avia K, Lipinska A, Godfroy O, Destombe C, Potin P, Sulpice R, Gachon CGM, Valero M, Cock JM (2021). Biobanking of novel S. latissima strains with characters of interest for biomass production. Project GenialG - GENetic diversity exploitation for Innovative macro-ALGal biorefinery, EC Grant agreement no: 727892 (J. M Cock, Ed), Available from https://genialgproject.eu/results/deliverables/.

Executive summary

This deliverable Part A describes work carried out as part of work package 2 of the Genialg project aimed at developing methodologies to identify strains of *Saccharina latissima* of interest for biomass production and to make those strains available through a germplasm biobank. The work carried out focused on two main aspects, establishment of a germplasm collection exploitable for strain selection and the development and application of methodologies to identify strains of interest. The first aspect involved surveying the genetic structure of *S. latissima* populations along the European coastline, planning and implementation of a strategy to sample these populations, biobanking of germplasm in the form of gametophyte cultures and evaluation of genetic strategies to derive second-generation sporophytes from the gametophyte stocks (including evaluation of heritance and of the effects of inbreeding and outbreeding). The second aspect involved the development and application of phenotyping and genotyping methodologies and approaches to use this information for the selection of genotypes of interest. The main output from this deliverable is a large collection of *S. latissima* germplasm, together with genetic-based tools for the identification of strains of interest for biomass production that allow the selection of genotypes from specific geographical locations. The tools developed can therefore be employed to improve production, whilst at the same time limiting the selection process to local genotypes, avoiding the introduction of foreign genotypes into aquaculture sites.

Part B describes the work carried out to identify strains of *Ulva spp.* of interest for biomass production and to make those strains available through a germplasm biobank. The work carried out focused on two main aspects, 1) establishment through cryopreservation of a germplasm collection of *Ulva* exploitable for strain selection and 2) the development and application of methodologies to identify strains by genetic studies.

****

Introduction

This deliverable is associated with task 2.1 "Applying breeding approaches for strain improvement". The objective of this sub-task was to develop and apply phenotyping approaches to identify novel strains with characteristics of industrial interest. Field-collected sporophytes cannot be maintained in culture. The work carried out for this deliverable was therefore based on the established of gametophyte cultures derived from field-collected sporophytes via sporulation of blade sori. These gametophytes can be used to generate a new (F1) generation of sporophytes by inducing fertility and allowing male and female gametophytes to cross fertilise.

Genetic analysis for this work package was based on a large-scale, Europe-wide collection of *Saccharina latissima* germplasm in the form of gametophyte cultures derived from sporulation of field-collected sporophyte blades fragments. Because exploitation of this resource requires production of sporophytes (i.e. the cultivated generation) from the gametophyte strain stocks essential actions for the deliverable included investigating the heritability of phenotypic characters and evaluation of different crossing strategies (selfing compared with outbreeding).

One question of particular interest was whether characteristics of sporophytes and gametophytes are correlated, for example whether the degree of stress resistance during the gametophyte generation is correlated with the degree of stress resistance during the subsequent sporophyte generation. Correlation between the phenotypes of the two generations would allow screens to be carried out on laboratory gametophyte cultures to select for characteristics of interest in the derived sporophyte generation.

Evaluation of the genetic structure of European *S. latissima* populations

Before carrying out large-scale biobanking of *S. latissima* germplasm, a Europe-wide census of the genetic structure of *S. latissima* populations was carried out in relation to local environmental conditions (Guzinski *et al.*, 2020). The aim of this analysis was to obtain an overview of the genetic structure of European *S. latissima* populations and to link genetic structure to local environmental conditions.

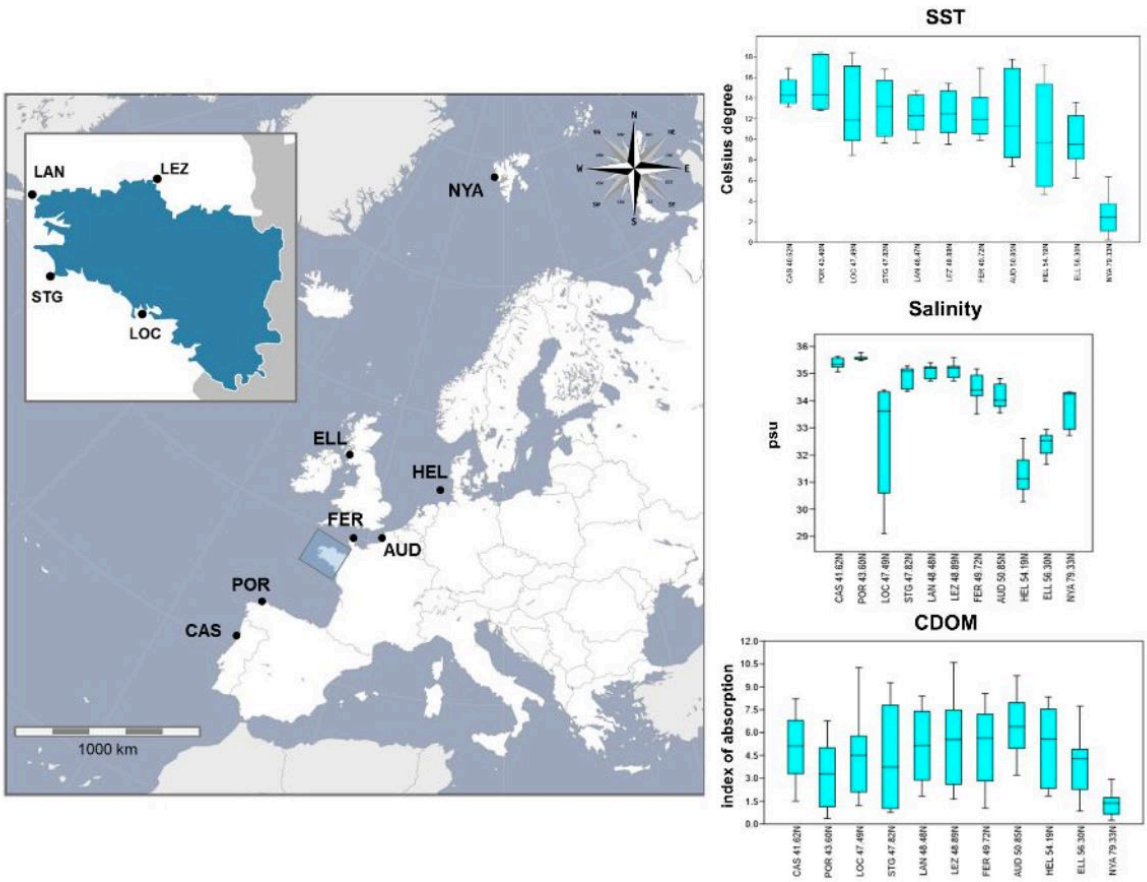


Figure 1. Map showing the 11 sites sampled for *S. latissima*. The box-plots on the right show within-site variation for three environmental variables: sea surface temperature (SST), salinity and chromophobic dissolved organic material (CDOM). The localities are ordered by latitude with the most northern site on the left. Figure reproduced from Guzinski *et al.* (Guzinski *et al.*, 2020)

Double digest restriction site-associated DNA sequencing (ddRAD-seq) derived single nucleotide polymorphisms (SNPs) were used to evaluate the genetic diversity and structure of *S. latissima* from across 11 sites spanning the entire European Atlantic latitudinal range of this species (Figure 1). This study indicated that there was a high degree of genetic differentiation between the sampled populations (Figure 2). Moreover, genetic diversity was significantly higher for the northernmost locality (Spitsbergen) compared to sites at the southern end of the species range in northern Iberia. Analysis of the genetic structure of the *S. latissima* populations in relation to three environmental variables (sea surface temperature, salinity, and water turbidity) identified seven SNPs and 12 microsatellite alleles that were significantly associated with at least one of these environmental variables.

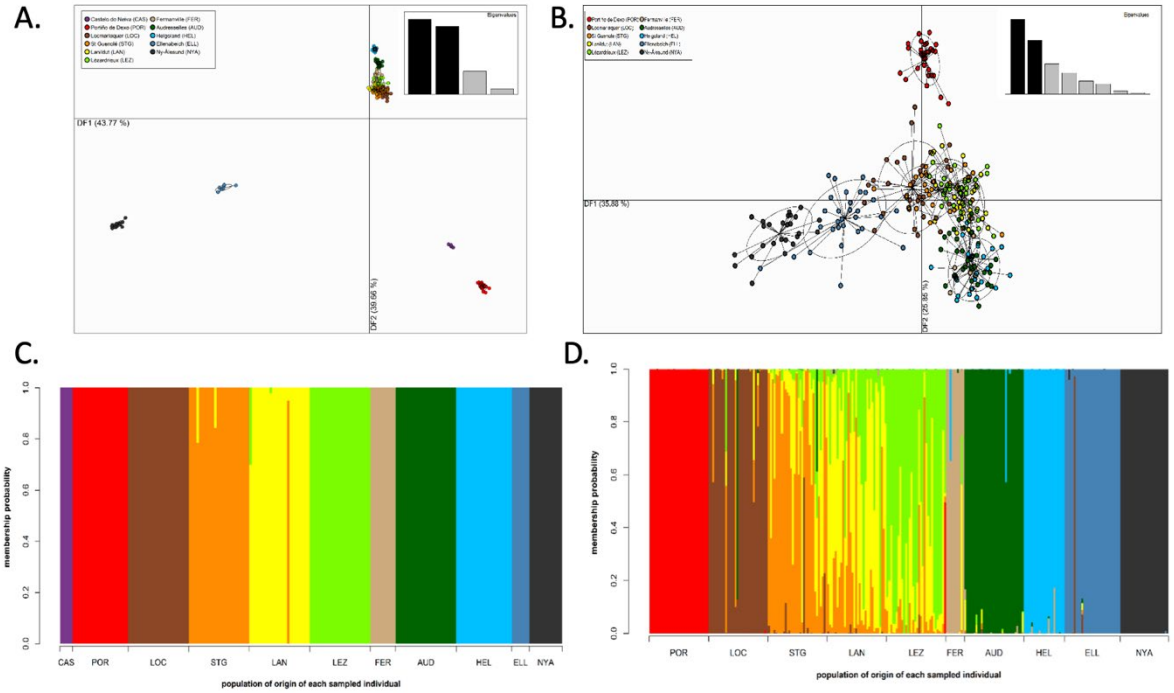


Figure 2. Genetic structure of European *S. latissima* populations. A.,C. Discriminant analysis of principal components (DAPC) scatterplot and compoplot for 199 *S. latissima* sporophytes sampled from 11 localities genotyped using 4,069 SNP markers. B.,D. Discriminant analysis of principal components (DAPC) scatterplot and compoplot for 280 *S. latissima* sporophytes sampled from 10 localities (see key) genotyped at 18 SSR markers. Figure reproduced from Guzinski *et al.* (Guzinski *et al.*, 2020).

This study therefore provided important genetic structure information for the design of the germplasm sampling phase of the project. The study also established a set of more than 4000 single nucleotide polymorphism (SNP) markers and identified genetic loci of potential interest for both conservation and aquaculture strategies targeting *S. latissima*.

Establishment of a Europe-wide collection of *S. latissima* germplasm

To establish a genetically diverse collection of *S. latissima* strains representing population along most of the length of the European Atlantic coastline, parental sporophytes were sampled from 21 different localities extending from Portugal to northern Norway (Figure 3, Table 1). This collection of sporophytes also served as the starting material for the genome-wide association study (GWAS; Deliverable 2.3) and is therefore referred to hereafter as the "GWAS collection".

Below we provide a brief description of the sampling, phenotyping and genotyping of this collection. A more complete description of this process can be found in Deliverable 2.3).

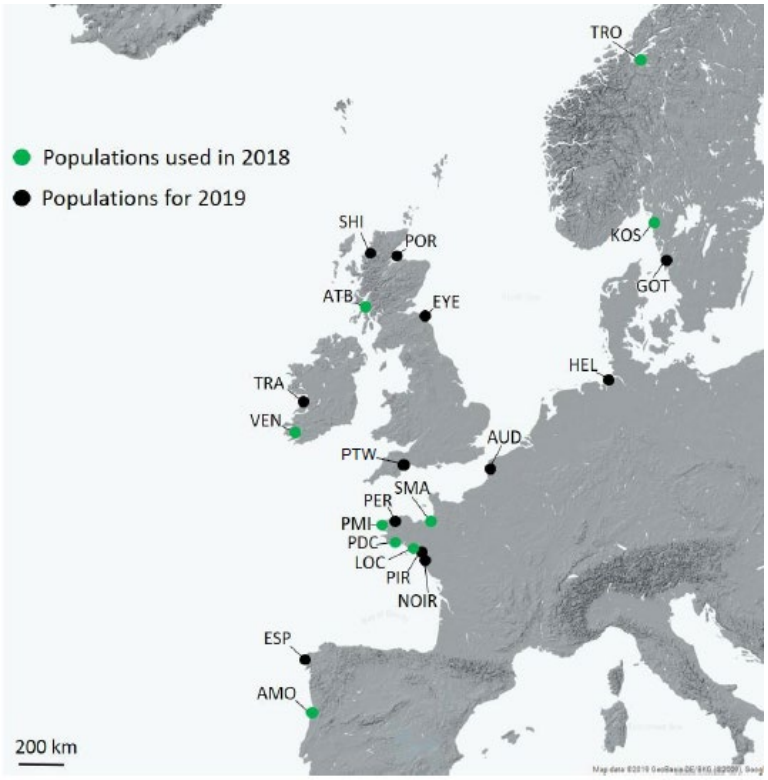


Figure 3. *S. latissima* sampling sites for the GWAS collection.

Table 1. Populations sampled for *Saccharina latissima* sporophytes.

Population code	Ocean or sea	Country	Region	Location	Site	Latitude	Longitude	Date of collection	N° indiv. sampled
VEN	Celtic Sea	Ireland	County Kerry	Dingle Peninsula	Ventry bay	52.11575	-10.36756	20/02/2018	30
TRO	Norwegian Sea	Norway	Sør-Trøndelag	Trondheim	Ringvebukta	63.45261	10.46119	04/02/2018	30
KOS	North Sea	Suède	Västra Götaland	Tjärnö*	Kosterhavet national park	58.83552	10.99055	27/01/2018	30
PDC > CON	NE Atlantic	France	Brittany	Concarneau	Pointe du Cabellou	47.85972	-3.91638	20/12/2017	30
LOC	NE Atlantic	France	Brittany	Locmariaquer	Pointe de Kerpenhir	47.55705	-2.92567	04/12/2017	30
LOC	NE Atlantic	France	Bretagne	Locmariaquer	Pointe de Kerpenhir	47.55705	-2.92567	21/04/2019	9

PER	NE Atlantic	France	Brittany	Roscoff	Perharidy	48.73087	-4.00453	05/12/2017	30
PER	NE Atlantic	France	Brittany	Roscoff	Perharidy	48.73087	-4.00453	21/02/2019	30
SMA	NE Atlantic	France	Brittany	Saint-Malo	Saint-Malo	48.65245	-2.03801	05/12/2017	30
PMI	NE Atlantic	France	Brittany	Lanildut	Parc Marin d'Iroise	-4.79261	48.47902	22/01/2017	30
ATB	NE Atlantic	Scotland	Argyll and Bute	Oban	Atlantic Bridge	56.31620	-5.58350	15/12/2017	30
AMO	Atlantic Ocean	Portugal	Norte e Porto	Viana do castelo	Amorosa	41.64187	-8.82340	03/02/2018	30
ESP	Atlantic Ocean	Spain	Uia	Ría de Muros y Noya	Punta Polveira	42.78667	-8.96611	20/02/2019	19
TRA	Celtic Sea	Ireland	County Clare	New Quay	Tra Li	53.15625	-9.07312	20/02/2019	21
PIR	NE Atlantic	France	Pays de la Loire	Vendee	Piriac-sur-mer	47.38189	-2.53866	06/12/2018	20
PTW	NE Atlantic	England	Southwest	Cornwall	Porthallow	-50.06919	-5.08119	20/03/2019	10
POR	North Sea	Scotland	Highlands	Tarbat Peninsula	Portmahomack	57.83290	-3.84350	24/01/2019	25
AUD	NE Atlantic	France	Pas-de-Calais	Cape de Gris Nez	Audreselles	50.83269	1.58602	21/02/2019	22
SHI	NE Atlantic	Scotland	Highlands	Shieldaig	Shieldaig	57.52580	-5.65000	06/12/2018	23
EYE	North Sea	Scotland	Berwickshire	Eyemouth	Eyemouth	55.83990	-2.06190	23/01/2019	26
GOT	North Sea	Sweden	Bohuslän	Mollösund	Mollösund	58.06367	11.47574	07/02/2019	25
NOR	NE Atlantic	France	Loire-Atlantique	Noirmoutier	Pointe du Devin	46.99128	-2.30273	22/01/2019	17
HEL	North Sea	Germany	Helgoland	Helgoland	Helgoland Kurpromenade beach	54.18180	7.89155	21/03/2019	32

The following information was noted for each sampling session: name of the site, date and hour, tide coefficient, corrected depth if subtidal, GPS coordinates of the site, water temperature, names of the operators.

Each individual was identified by an ID code that included "SL" for *Saccharina latissima*, a three-letter corresponding to the sampling site (e.g. "ELL" for Ellenabeich) and a number. For example, the third sporophyte sampled at Ellenabeich was "SLELL3".

At the time of sampling, five morphometric measurements were carried out for each sampled sporophyte (Figure 4).

ID	Width Stipe	Length Stipe	Width Blade	Length Blade	Length Thallus
SLELL1	0.61	11.0	24.0	60.0	71.0
SLELL2	0.82	41.0	27.0	102.0	143.0
SLELL3	0.64	18.0	21.5	84.0	102.0

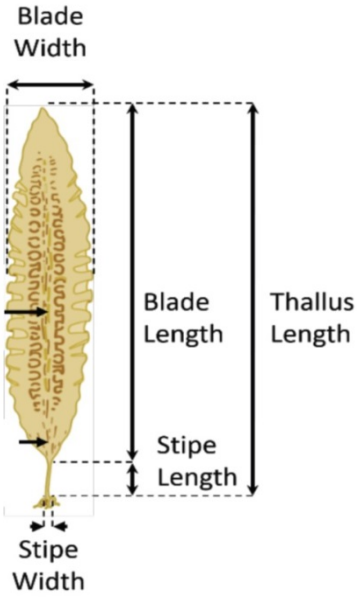


Figure 4. Morphometric parameters measured for sampled sporophytes. The top panel shows an example of morphometric data recorded for three individuals. The bottom panel indicates how the measurements were made.

Sampling involved simultaneous collection of vegetative sporophyte blade tissue for DNA extraction (2 cm^2 blade fragment from near the meristem area and the edge of the blade placed in a sachet with silica gel after rinsing with freshwater and drying with paper) and fertile blade tissue (sorus) for spore release to produce gametophyte cultures (four 1 cm^2 blade sorus fragments rinsed with freshwater, dried with a paper towel and placed in a 50 ml Falcon tube containing 45ml of sterile filtered seawater and two microscope slides to collect settled spores). The samples were then sent to SAMS (Oban, Scotland) for treatment.

Sporulation and generation of F1 sporophytes by selfing of bulk gametophytes

The protocols used to 1) induce sporulation of the sporophyte sorus samples, 2) carry out self crosses using bulks of gametophytes each derived from the same parent sporophyte and 3) grow the derived F1 sporophytes under controlled, common garden conditions in 10 L carboys (Figure 5) are described in Deliverable 2.3 (see the section describing the GWAS experiment).

Figure 5. Culture of F1 sporophytes derived by selfing (i.e. bulk gametophyte crosses) from the Europe-wide collection of field sporophyte sorus samples. On the right, Petri dish with young sporophytes derived by self-fertilisation of the bulked gametophytes from a single sampled parental sporophyte. The inset shows one young sporophyte. In the middle, carboy culture system. On the right, harvested sporophytes.

Phenotyping and genotyping of the F1 sporophytes

Following growth of the F1 sporophytes in 10 L carboys under common garden conditions, the individuals were individually photographed and the following morphometric measurements recorded:

- Stipe length (cm)
- Total length (cm)
- Blade length (cm)
- Blade width (at widest point) (cm)
- Blade surface area (cm^2)

In addition, sporophyte blade tissue samples that had been taken after the young sporophytes had been photographed and rapidly frozen in liquid nitrogen (no more than 30 min after removal from the carboys), were ground to a fine powder under liquid nitrogen and about 5 mg of powder for each sample was used for biochemical assays. For the assays, all 353 samples were

randomized across four different 96 well plates and assays were carried out for nitrite, nitrate, ammonium, amino acids, proteins and four different soluble sugars (see Deliverable 2.3 for details).

The F1 sporophytes were genotyped using the ddRAD-seq method. DNA was extracted from samples of 494 sporophyte individuals and two barcoded ddRAD libraries constructed for 125 bp paired-end Illumina sequencing.

To summarise, a total of 579 sporophyte samples were collected from 21 sites across Europe. Phenotype information was collected both for the harvested field sporophytes and for F1 progeny derived by selfing the field progeny and ddRAD-based genotypes were determined for the F1 sporophyte progeny.

Heritance analysis

To determine whether growth-related (morphometric) traits are inherited genetically, progeny of five individuals from the GWAS collection (SLPER276, SLESP24, SLGOT27, SLHEL39 and SLSHI15) were grown as described above but with the progeny of each individual in a different carboy so that the parent-of-origin could be easily followed. The progeny of each individual (at least 20 progenies were analysed for each parental sporophyte, 147 individuals in all) were then phenotyped as described above. If the measured morphometric traits were inherited genetically, then variance within the progeny of a single individual was expected to have been lower than for randomly selected progeny from the five parent individuals. The progenies were put into culture for the same amount of time using the same 10 L carboy culture system as described above except that progeny of different parents were not mixed. Photographs were taken after the culture period and the following measurements were made: 1) stipe length, 2) total length (blade plus stipe), 3) blade length, 4) blade width 5) blade surface area. Standard deviations were calculated for each of the five groups of progeny and also for 1000 randomised datasets and a permutation test was applied to calculate the significance of the comparison between the observed standard deviations from each sample and the randomized standard deviations.

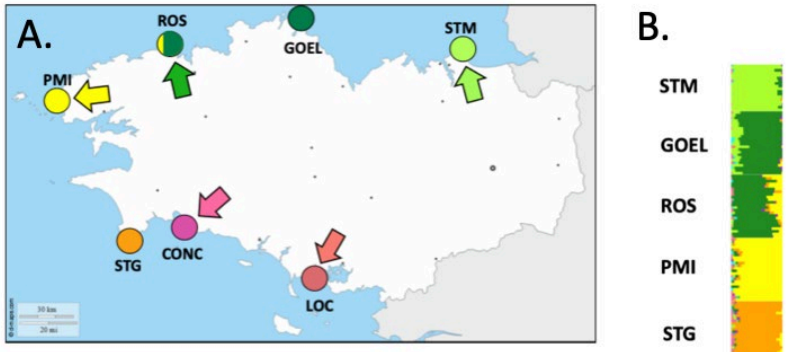
In general the expectations were fully met. The standard deviations for the progeny pools were significantly lower than for the randomised pools. The only exception was for SLESP24 but it

was noted that several of the progenies of this individual were deformed so that may have affected the analysis. Overall, however, this analysis confirmed that growth characteristics are inherited, which was an important observation, making it relevant to select progeny based on phenotypic information about their progenitors.

Evaluation of the effects of inbreeding and outcrossing

The *S. latissima* strains have been biobanked in the form of gametophyte cultures and it is therefore necessary to carry out crosses in order to recover sporophytes (i.e. the generation exploited for aquaculture). The simplest approach to obtain sporophytes derived from a specific, sampled sporophyte is to cross male and female gametophytes that are both derived from the same sporophyte of interest (i.e. via selfing). However, this approach will tend to lead to inbreeding of the strain of interest, with a possible risk of inbreeding depression. Experiments were therefore carried out to evaluate the phenotypic effects of inbreeding and outcrossing to determine whether there were any detectible effects and, thereby, to provide important knowledge for the planning of future strategies for the exploitation of the *S. latissima* germplasm collection.

Crossing experiments were carried out using strains from five different sites (Figure 6A) located around the coast of Brittany (France): Saint Malo (STM), Perharidy (ROS), Parc Marin d'Iroise (PMI), Concarneau (CONC) and Locmariaquer (LOC). About 30 sporophytes were sampled from each of these five locations plus individuals from Saint Guenolé (STG) and Goëlo (GOEL), 168 individuals in total. ddRAD-seq genoptyping (4,104 SNPs) of these individuals (Figure 6B) indicated that there was a clear genetic distinction between populations from northern and southern Brittany but a genetic continuum between populations (PMI, ROS, GOEL) in northwest Brittany (corresponding to a "cold water pocket"; Gallon *et al.*, 2014).



C.

	LOC	CONC	PMI	ROS	STM
LOC	6S+3W				
CONC	2 B	6S+3W			
PMI	2 B	2 B	6S+3W		
ROS	2 B	2 B	2 B	6S+3W	
STM	2 B	2 B	2 B	2 B	6S+3W

Figure 6. Experiments carried out to evaluate the effects of inbreeding and outcrossing. A. Sampled populations of Breton *S. latissima*. B. Representation of population structure based on SNP data for the seven Breton *S. latiss*ima populations. C. Summary of the crosses carried out (S, selfing; W, within population; B, between populations; see text for details). STM, Saint Malo; ROS, Perharidy; PMI, Parc Marin d'Iroise; CONC, Concarneau; LOC, Locmariaquer.

The collected sporophytes were sporulated in the laboratory to produce gametophytes. Crosses were then carried out by mixing pools of male and female gametophytes derived from individual sporophytes ("bulk" crosses). Three types of cross were carried out (Figure 6C):

- S ("self") crosses: selfing of gametophytes all derived from the same individual. Self crosses of six individuals from each site (30 S crosses in total).
- W ("within") crosses: within-population crosses carried out by mixing gametophytes from 30 individuals all from the same population. Three replicate within-population crosses were carried out for each population (15 W crosses in total).
- B ("between") crosses: between-population crosses carried out by mixing gametophytes from 15 individuals all from one population with gametophytes from 15 individuals of a second population. Two replicate between-population crosses were carried out for each combination of two different populations (10 different combinations, 20 B crosses in total).

The sporophytes derived from these crosses were allowed to attach to seeding strings (Figure 7) and the strings then attached to ropes for open sea culture for a period of six months from April to October 2018.

Figure 7. Sporophytes derived from a laboratory cross developing attached to seeding string.

Blade width was measured every month as a proxy for growth performance. The data shown below correspond to measurements made in July after three months of growth.

Comparison of blade widths for progeny issued from selfing with measurements of progeny issued from within-population crosses indicated evidence of inbreeding depression (Figure 8A). Analysis of individual populations indicated that blade widths were particularly low for populations from northern Brittany (Figure 8B,C). Comparison of progeny issued from selfing or within-population crosses for each locality indicated that the strength of inbreeding depression depended on the population, being non significant in the southern population near the range limit of the species (Figure 8D).

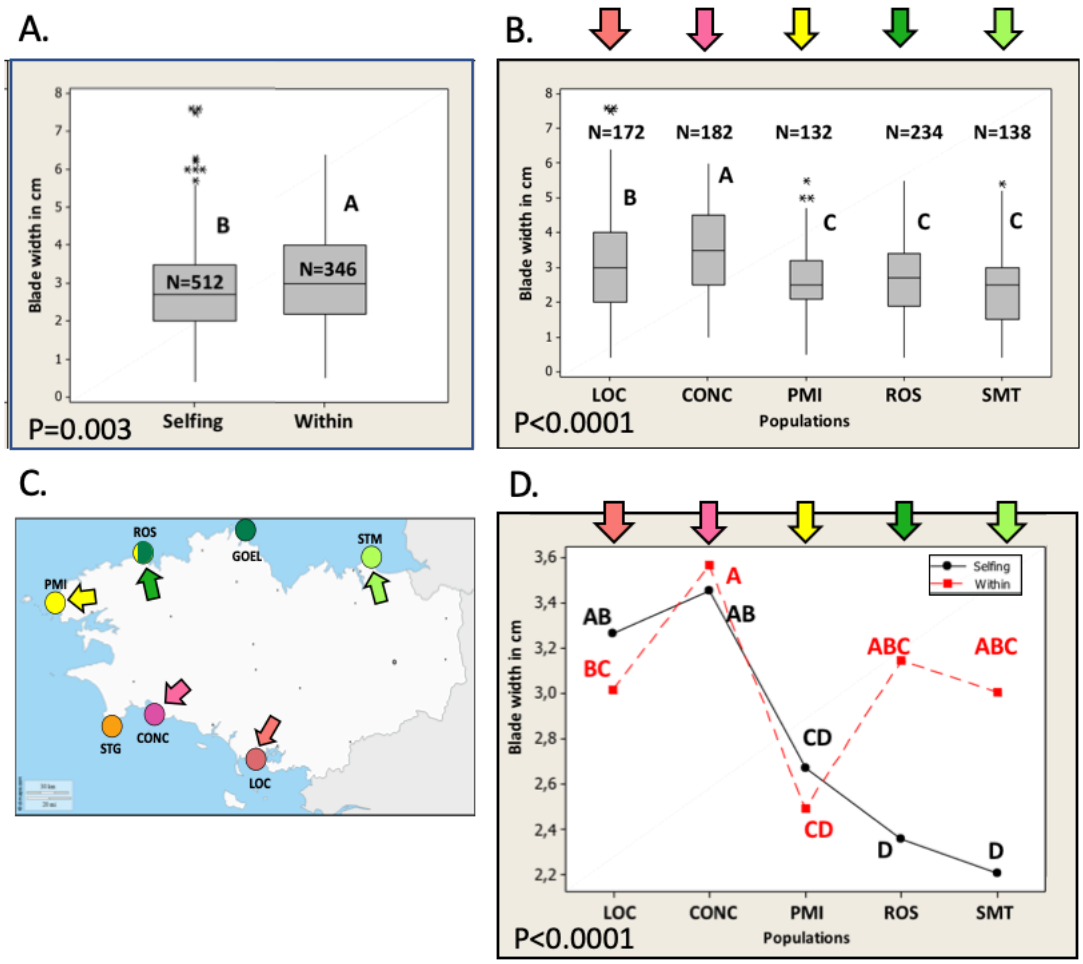


Figure 8. Blade width measurements for progeny issued from selfing or within-population crosses after 3 months of culture. A. Comparison of progeny issued from selfing with progeny issued from within-population crosses. B. Comparison of progeny issued from within-population crosses for each of the five localities. C. Sampling locations for the five populations. D. Comparison of progeny issued from selfing with progeny issued from within-population crosses for each of the five localities. Letters indicate statistically different datasets. STM, Saint Malo; ROS, Perharidy; PMI, Parc Marin d'Iroise; CONC, Concarneau; LOC, Locmariaquer.

For each combination of between-population crosses, blade width measurements were compared with data for within-population crosses for each of the two relevant localities (Figure 9). The results of this analysis were variable, depending on the cross that had been carried out. Evidence of outbreeding depression was observed for three crosses, including two crosses between populations from north and south Brittany (ROS x CONC, STM x CONC). This would be consistent with outbreeding depression being more marked between more distantly related

individuals. For two other combinations (plus possibly a third combination that was not statistically significant), evidence of heterosis was detected. Both of these crosses were between populations from north and south Brittany (PMI x CONC, PMI x LOC).

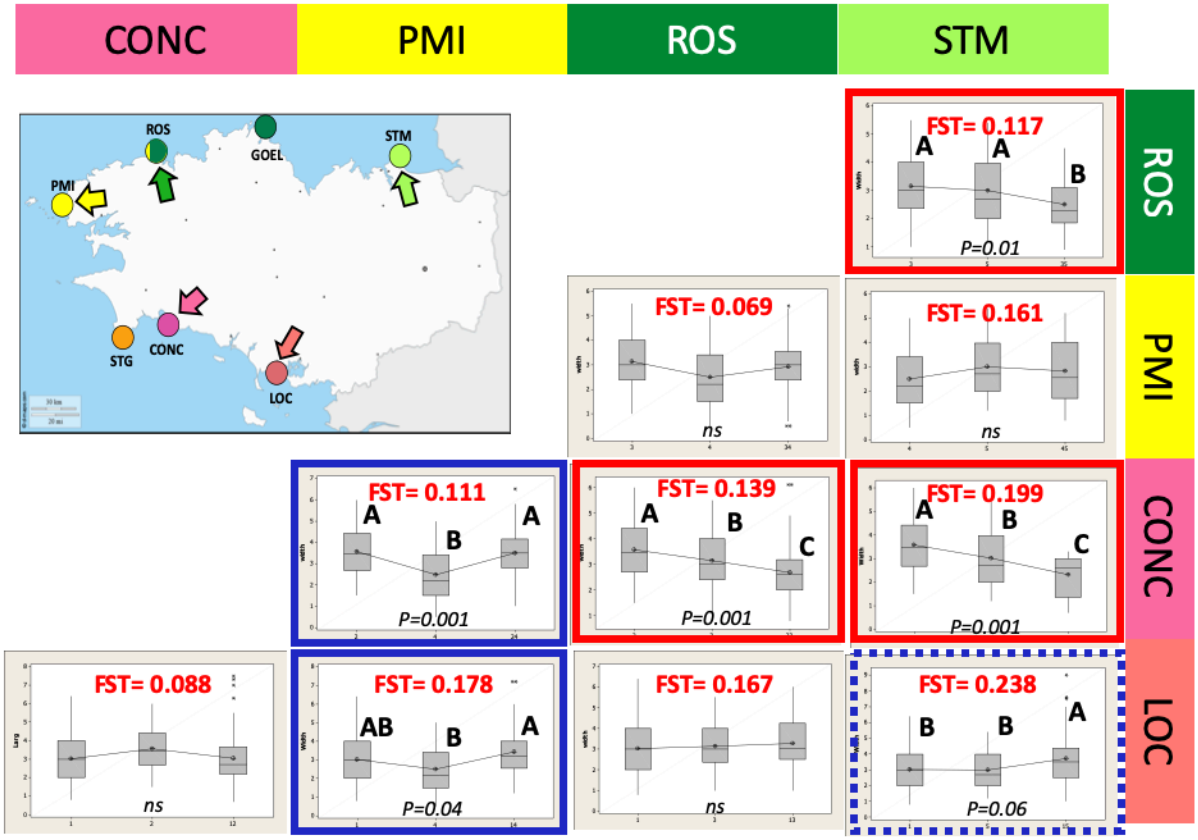


Figure 9. Blade width measurements for progeny issued from within-population crosses and from between-population crosses. Each graph shows a box plots for the progeny derived from between-population crosses on the right together with the box plots for within-population crosses for the two relevant populations on the left. Letters indicate statistically different datasets. Red boxes indicate evidence of outbreeding depression. Blue boxes indicate evidence of heterosis, with a marginal case indicated by a dotted blue box. Fixation index (FST) values provide an estimation of the genetic distance between each pair of populations (high F_{ST} values indicate high genetic distance). STM, Saint Malo; ROS, Perharidy; PMI, Parc Marin d'Iroise; CONC, Concarneau; LOC, Locmariaquer.

To summarise, therefore, these experiments detected evidence of inbreeding depression, outbreeding depression and heterosis but all of these effects were dependent on the study populations that were being used for each cross. The results of this study have provided a useful

base for the design of future breeding strategies and they indicate that inbreeding depression, outbreeding depression and heterosis should be taken into account when designing future breeding strategies. However, as in most cases the observed effects were not very strong, they did not argue strongly against the use of selfing as the method to derive sporophytes from stored gametophyte germplasm. Given that selfing simplifies tracking of genetic effects across generations we would therefore propose that it be used as the preferred method to generate sporophytes from gametophyte stock cultures.

Development of large-scale phenotyping approaches

Phenotyping of sporophytes derived from the collection of biobanked *S. latissima* gametophytes (the GWAS collection) has so far been limited to simple morphometric measurements at the time of harvest and assays of primary metabolites (nitrite, nitrate, ammonium, amino acids, proteins and four different soluble sugars). In an effort to develop additional, more precise phenotyping methodologies, the growth monitoring system initially developed for *Ulva* phenotyping, was tested with young *S. latissima* sporophytes to determine if it could be employed to precisely measure growth rate (as a proxy for biomass production capacity). These experiments, which were carried out in Galway (NUIG) using material generated in Roscoff (CNRS), were co-financed by the exchange program Partenariat Hubert Curien (PHC) Ulysses.

Young, developing *S. latissima* sporophytes were gently flattened against the bottom of an aquarium containing circulated Provasoli-enriched natural seawater using 3D printed circles and nylon mesh and grown at 15°C with a photoperiod of 12h:12h day:night and a light intensity of 100 $\mu mol.m^{-2}s^{-1}$ (Figure 10A). The cultures were photographed every 2 min and averaged images calculated for each period of 10 min (Figure 10B,C). Images taken over a 7 day growth period were analysed with ImageJ and growth rates calculated in R.

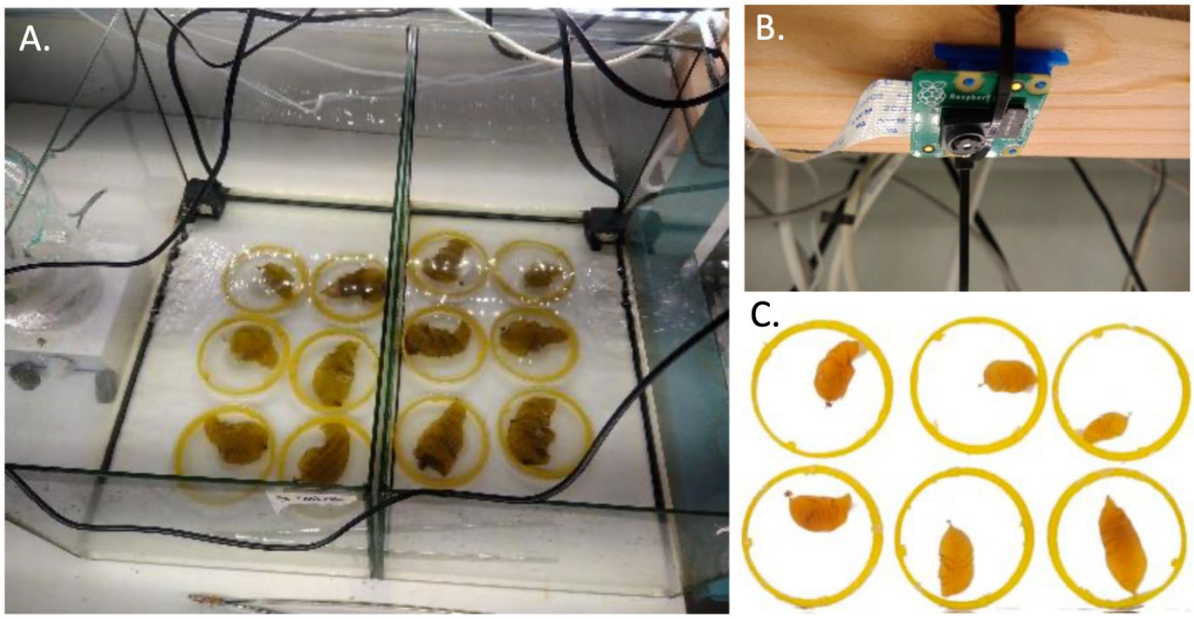


Figure 10. Measuring growth rates of young sporophytes using the Galway phenotyping system. A. View of the aquarium with young sporophytes in the growth system. B. Camera. C. Example of an image taken to obtain growth measurements.

Figure 11 shows an example of data obtained for four individuals (SLSHI-15, SLSHI-23, SLPER-272 et SLHEL-26). Differences in growth rates could be detected using the system but error bars were large for the test experiment, which only used 6 replicates. Growth was normalised based on the initial size of the sporophyte at the start of the experiment but, not unexpectedly, initial size had an influence on growth rate in that individuals with a larger start size (e.g. SLSHI-15 in Figure 11, 3-4 cm) grew more slowly than individuals with a smaller start size (SLSHI-23, SLPER-272 et SLHEL-26, <1 cm).

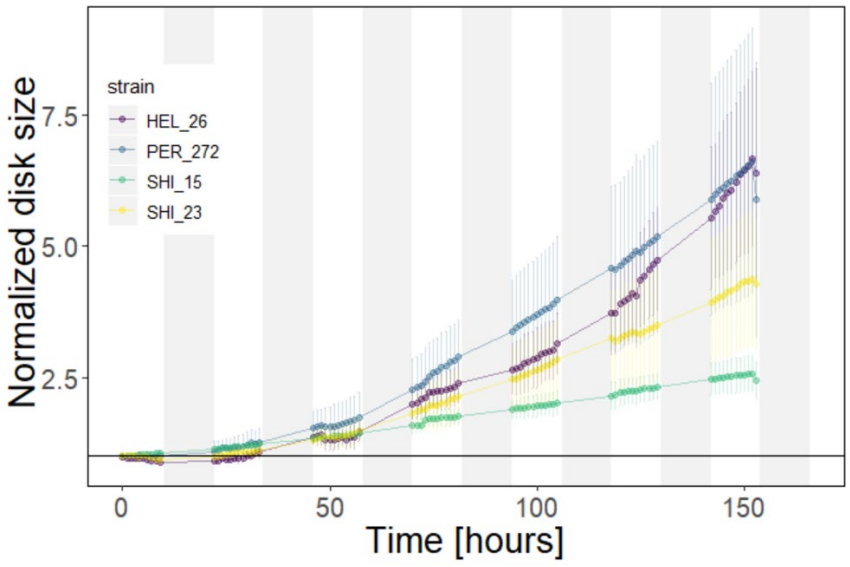


Figure 11. Normalised growth rates for sporophytes of four strains of *S. latissima* measured on the Galway phenotyping system.

These experiments demonstrated that the phenotyping system could be used for measurement of growth for young sporophytes with a starting size of between 0.5 and 4.0 cm but future experimental designs should include more replicates (at least 12) and the methodology will need to take into account differences in the sizes of the sporophytes at the start of an experiment.

Biobanking of *S. latissima* strains with characters of interest for biomass production

The actions presented above have established a large collection of biobanked gametophyte germplasm derived for Europe-wide populations of *S. latissima*, they have established that phenotypic traits are inherited from one generation to the next and they have evaluated different crossing strategies to take into account potential effects of inbreeding depression, outbreeding depression and heterosis. Large-scale genotyping based on ddRAD-seq has been developed and applied to large collections of *S. latissima* individuals and phenotypic information has been collected for progeny of the entire biobanked collection. Moreover, additional phenotyping approaches are being developed that are specifically adapted to the analysis of large *S. latissima* collections. These important advances now make it possible to exploit the biobanked *S. latissima* germplasm resource to select *S. latissima* strains with characters of interest, for example for biomass production. An important prerequisite for this for the future will be to obtain ddRAD-seq (large-scale SNP) genotype information for the entire gametophyte

collection (using the approach that is already being applied to genotype the large F2 family for the gametophyte-based QTL analysis, see Deliverable 2.3). The strategy to select *S. latissima* strains with characters of interest will then be to exploit the genotype and phenotype information obtained for the F1 sporophytes to identify gametophyte strains carrying alleles that are significantly associated with phenotypes of interest, including biometric proxies linked to biomass production. Note that the structuring of the bioresource into samples from populations corresponding to different geographic localities means that this approach can be applied at the local level by selecting optimal strains from a specific locality. This approach will not only allow selection of strains of interest for aquaculture but will also ensure that selected strains are endemic and avoid problems linked to the introduction of material from other geographical locations.

References

Gallon RK, Robuchon M, Leroy B, Gall LL, Valero M, Feunteun E (2014) Twenty years of observed and predicted changes in subtidal red seaweed assemblages along a biogeographical transition zone: inferring potential causes from environmental data. **J Biogeogr** 41:2293–2306.

Guzinski J, Ruggeri P, Ballenghien M, Mauger S, Jacquemin B, Jollivet C, Coudret J, Jaugeon L, Destombe C, Valero M (2020) Seascape Genomics of the Sugar Kelp Saccharina latissima along the North Eastern Atlantic Latitudinal Gradient. **Genes** 11.

Deliverable D2.2 Part B: *Ulva spp.*

Biobanking of novel *Ulva* strains with characters of interest for biomass production

Contributors; Antoine Fort, Marcus McHale, Kevin Cascella, Philippe Potin, Marie-Mathilde Perrineau, Philip D. Kerrison, Elisabete da Costa, Ricardo Calado, Maria do Rosário Domingues, Isabel Costa Azevedo, Isabel Sousa-Pinto, Claire Gachon, Adrie van der Werf, Willem de Visser, Johanna E. Beniers, Henrice Jansen, Michael D. Guiry, Ronan Sulpice

TO BE CITED AS: Antoine Fort, Marcus McHale, Kevin Cascella, Philippe Potin, Marie-Mathilde Perrineau, Philip D. Kerrison, Elisabete da Costa, Ricardo Calado, Maria do Rosário Domingues, Isabel Costa Azevedo, Isabel Sousa-Pinto, Claire Gachon, Adrie van der Werf, Willem de Visser, Johanna E. Beniers, Henrice Jansen, Michael D. Guiry, Ronan Sulpice (2021). Biobanking of novel Ulva strains with characters of interest for biomass production. Project GenialG - GENetic diversity exploitation for Innovative macro-ALGal biorefinery, EC Grant agreement no: 727892 (R. Sulpice Ed), Available from https://genialgproject.eu/results/deliverables/.

1) Cryobiobanking of novel *Ulva* strains

Biobanking is time consuming, costly and largely unreliable for *Ulva* because strains die over time. We have therefore developed a **cryopreservation protocol** for *Ulva* thalli in order to safeguard the sequenced strains.

Material and Methods

Solutions:

40% glycerol stock in artificial seawater made in 0.1M HEPES pH 8
Artificial seawater made in 0.1M HEPES pH 8
Isopropanol

Equipment:
2 mL eppis
MrFrosty (Fisher Scientific)
-80C freezer
Liquid nitrogen dewar with tube containers

Protocol
Largely adapted from ***Lee and Nam, 2016***

Cryopreservation:

⑩ Incubate *Ulva* thallus in darkness for two days (13C) prior to cryopreservation.

- ↘ *This appears important, with better recovery. Our reasoning is that ulvans, which comprise 9-30% of Ulva biomass, are located in the apoplast. Having such coating of the cells might impede glycerol penetration, and by keeping the sample in the dark prior to cryogenisation we hope to reduce the amount of ulvans in the aploplast.*

- Cut Ulva discs of ~3 cm diameter
  - ↘ *The bigger the size the most likely patches of live tissue will survive.*
- Place disc in 2 mL eppis
- Add 750 uL ASW 0.1M HEPES pH8
  - ↘ *Not sure how important the buffering is, we haven't tried with normal ASW*
- Slowly add 750 uL 40% glycerol stock (50 uL or 100 uL each time up to 750 uL, shaking between additions, typically over 15 minutes total), for a final concentration of 20% glycerol
  - ↘ *DMSO doesn't work in our hands. 20% or 10% glycerol OK.*
- Incubate at room temperature for 45'
  - ↘ *We have tried longer incubations, doesn't seem to affect results*
- Place in MrFrosty box filled with Isopropanol (to achieve cooling rate of -1C/min)
- Place MrFrosty in -80 C freezer for two hours
  - ↘ *We haven't tested different (shorter) incubation times, but it might impact recovery. Effectively the samples are at -80C after the two hours incubation period.*
- After two hours, place eppis in liquid nitrogen and keep in nitrogen container, samples done.
  - ↘ *Liquid nitrogen is the most important factor here, keeping samples in -80C or placing at -80C after liquid nitrogen treatment kills the cells.*
  - ↘ *Only ever tried <5 days incubation, not sure how they will survive long term – something to test.*

Recovery:

- Quickly thaw samples at 40C (water bath or thermoshaker) for ~1 minute until ice is gone
- Replace seawater 4 times to get rid of glycerol
  - ↘ *Pre-cooled ASW/HEPES, not sure if that impacts recovery.*
- Place discs in jars filled with ASW + nutrients (1X Guillard's F/2) in growth room, light intensity ~80 umoles.
  - ↘ *Recent tests indicate that it's better to leave the samples alone for a week. Adding bubbling straight away negatively impacts recovery. I'd recommend adding bubbling for stimulating growth after 1 week of recovery. Similar reasoning regarding light intensity: too high straight away might overstress the cells to the point of inviability (not tested).*
- Dead cells will be visible after ~5 days, with patches of live green tissue. One can manually remove the dead tissue, or add bubbling which will detach the dead bits, and the live tissue should start growing after a week.
- After the first week of recovery, replace media, add bubbling and monitor live tissue growth.

This protocol is currently used by NUIG to maintain phenotyped and genotyped strains (Fort et al., 2021). Our initial trials showed a ~90% survival rate of the strains, either undergoing vegetative growth directly after cryopreservation (Figure 1), or via sporulation and regeneration

of gametophytes or sporophytes post-cryopreservation. A paper is currently being drafted and will be submitted after the end of the Genialg project.

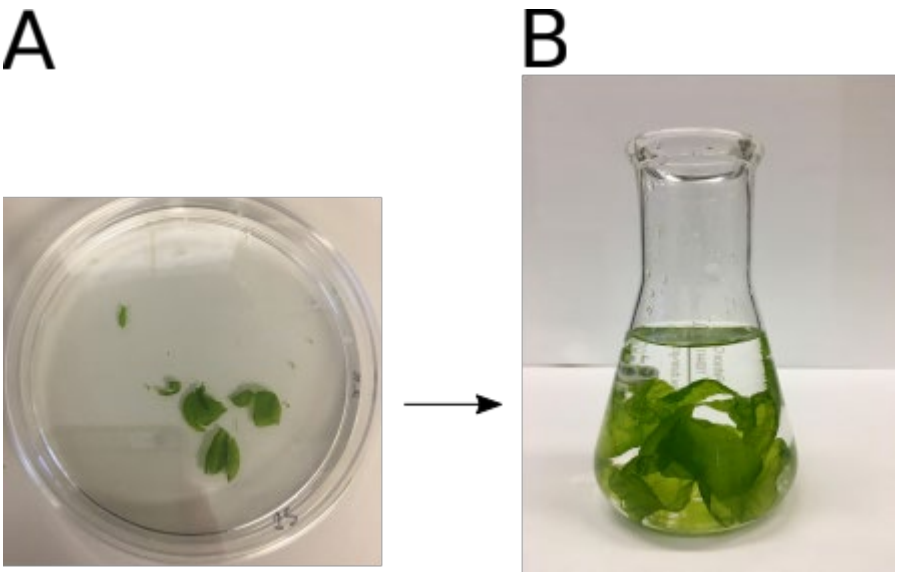


Figure 1. Successful cryopreservation of *Ulva.* A) *Ulva* discs one week after cryopreservation. B) Same *Ulva* discs after one month post-cryopreservation, showing vegetative growth.

Following the identification, by WUR, of highly productive, high quality strains (see below), these strains needed to be preserved over the years, which is of special relevance for commercial large scale seaweed cultivation. From 2017-2019 WUR bio-banked over 30 *Ulva* spp. populations collected along the coast of the Eastern Scheldt (see report 2019). Irrespective of the outcome, bio-banking of seaweed is laborious way of preserving high quality material, and it is suggested to look for other ways to preserve specific strains. During our second Genialg meeting in York 2019 this was further discussed and it was decided that one specific cryo-preservation protocol designed by NUIG (and largely adapted from Lee and Nam, 2016) would be tested by several laboratories within the Genialg consortium.

In a series of 5 experiments of over 40 days each, WURs first results indicated that they could successfully cryo-preserve collected material, irrespective of the time (up to 14 months) the material was cryo-preserved. In almost all cases sporulation was observed shortly after thawing according to the slightly adapted NUIG protocol, and sporulated material could generally be successfully re-grown into healthy looking thallus. Incidental growth analyses indicated RGR-values varying between 0.1 and 0.3. However, massive re-sporulation also occurred and

depending on the number of " spider-like-thalli" a second re-growth phase occurred (see Figure 2). For now we conclude that cryo-preservation could be a way to store promising strains. However, successful re-growth strategies still awaits further experimentation. We are currently setting-up new experiments to further analyse this.

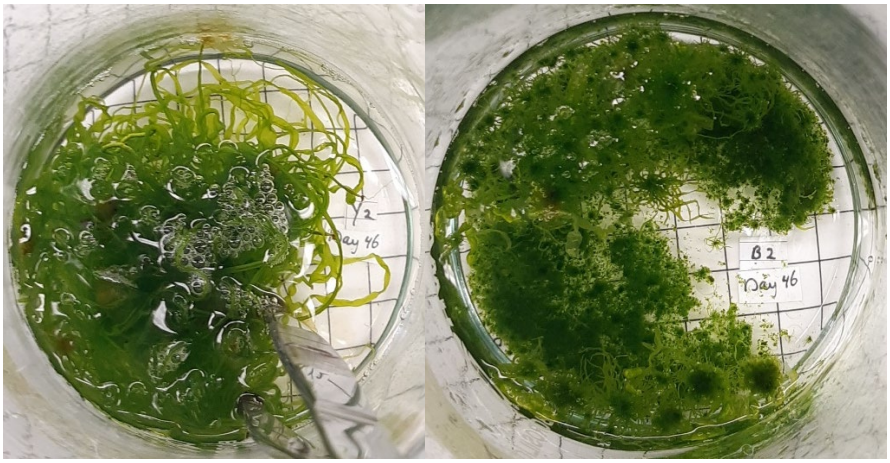


Figure 2. Illustration of " spider-like-thalli"

Both populations sporulated shortly after thawing. Population Y2 grew into healthy looking thalli at day 46. Population B2 partly re-sporulated half-way, and produced mainly "spider-like-thalli" after 46 days.

References

Fort A, McHale M, Cascella K, Potin P, Perrineau M-M, Kerrison PD, Costa E da, Calado R, Domingues M do R, Azevedo IC, Sousa-Pinto I, Gachon C, Werf A van der, Visser W de, Beniers JE, Jansen H, Guiry MD, Sulpice R (2021) Exhaustive reanalysis of barcode sequences from public repositories highlights ongoing misidentifications and impacts taxa diversity and distribution. **Molecular Ecology Resources** n/a Available at: https://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13453 [Accessed July 19, 2021].

Lee YN, Nam KW (2016) Cryopreservation of gametophytic thalli of *Ulva prolifera* (Ulvales, Chlorophyta) from Korea. **J Appl Phycol** 28:1207–1213.

2) **Exhaustive reanalysis of barcode sequences from public repositories highlights ongoing misidentifications and impacts taxa diversity and distribution**

Fort A, McHale M, Cascella K, Potin P, Perrineau M-M, Kerrison PD, Costa E da, Calado R, Domingues M do R, Azevedo IC, Sousa-Pinto I, Gachon C, Werf A van der, Visser W de, Beniers JE, Jansen H, Guiry MD, Sulpice R

Executive summary

Accurate species identification often relies on public repositories to compare the barcode sequences of the investigated individual(s) with taxonomically assigned sequences. However, the accuracy of identifications in public repositories is often questionable, and the names originally given are rarely updated. For instance, species of the Sea Lettuce (*Ulva* spp.; Ulvophyceae, Ulvales, Ulvaceae) are frequently misidentified in public repositories, including herbaria and gene banks, making species identification based on traditional barcoding unreliable. We DNA barcoded 295 individual distromatic foliose strains of *Ulva* from the North-East Atlantic for three loci (*rbc*L, *tuf*A, ITS1). Seven distinct species were found, and we compared our results with all worldwide *Ulva* spp. sequences present in the NCBI database for the three barcodes *rbc*L, *tuf*A and the ITS1. Our results demonstrate a large degree of species misidentification, where we estimate that 24%–32% of the entries pertaining to foliose species are misannotated and provide an exhaustive list of NCBI sequences reannotations. An analysis of the global distribution of registered samples from foliose species also indicates possible geographical isolation for some species, and the absence of *U. lactuca* from Northern Europe. We extended our analytical framework to three other genera, *Fucus*, *Porphyra* and *Pyropia* and also identified erroneously labelled accessions and possibly new synonymies, albeit less than for *Ulva* spp. Altogether, exhaustive taxonomic clarification by aggregation of a library of barcode sequences highlights misannotations and delivers an improved representation of species diversity and distribution.

1 INTRODUCTION

Species identification of biological specimens is paramount for assessing the diversity of ecosystems (Johannesson & Andre, **2006**), identify invasion events (Dunbar et al., **2021**; Estoup & Guillemaud, **2010**), and qualify the distribution of species of interest (Mendez et al., **2010**). While morphological characteristics can be used for species identification (Dugon et

al., 2012), precise species identification often relies on the analysis of "barcode" sequences, which are small standardized genetic loci used for taxonomic identification of the samples (Valentini et al., 2009). Indeed, morphological characters can be a poor indicator of the underlying complexity of the genetic diversity within a genus (Packer et al., 2009).

For example, due to the phenotypic plasticity of the genus *Ulva* —the type genus of the *Ulvophyceae*, *Ulvales* and *Ulvaceae*— in response to environmental factors, and relatively subtle morphological differences between species (Hofmann et al., 2010; Malta et al., 1999), DNA barcoding is necessary to attribute species names to specimens, even for the most common species. DNA barcoding for the purpose of identifying specimens relies on the amplification and sequencing of specific loci in the genome. In plants and algae, it is often through chloroplast markers such as *rbc*L and *tuf*A, but also nuclear markers such as parts of the 45S rRNA repeats (most commonly the Internal Transcribed Spacer 1 [ITS1]) (Coat et al., 1998; Fort et al., 2018, 2019, 2020; Miladi et al., 2018; O'Kelly et al., 2010). The sequences obtained from those barcodes are then compared with sequences associated to species names which are publicly available in repositories, such as the National Center for Biotechnology Information (NCBI).

Typically, NCBI sequences with high percentage identity compared with the query sequence are considered as belonging to the same species and used as reference for phylogenetic trees when no statistical inference of species delimitation is used (Heesch et al., 2009; Saunders & Kucera, 2010; Steinhagen et al., 2019). The risk in such case is that the species attributed to the matching sequences present in the NCBI can be erroneous, leading to the misidentification of the investigated individual. Indeed, the taxonomic information in the NCBI is not always accurate, and often contains "putative" species names (Garg et al., 2019), erroneous classifications (Chowdhary et al., 2019; Nasehi et al., 2019), or nonupdated species names following nomenclature adjustments (Hughey et al., 2021). Therefore, improving the nomenclature and taxonomic classification of sequences of any genus of interest requires a careful exhaustive reanalysis of barcodes sequences, to ensure accurate classification of new specimens, and to provide an updated list of reannotations.

Here, we deployed such an analytical framework to revisit the phylogeny of *Ulva* spp., a genetically diverse group of green macroalgal species ubiquitous in the world's ocean, brackish and even in freshwater environments. Over 400 *Ulva* names have been coined of which about 90 are currently recognised as taxonomically valid (Guiry & Guiry, 2021), many of which are uncommon or rare and only about 25 are frequently reported (Guiry & Guiry, 2021). The morphology of *Ulva* species can be grouped into two general types, one containing foliose

"sheet-like" species (distromatic foliose blades commonly known as "Sea Lettuce"), and another with tubular or partially tubular thalli (monostromatic tubes formerly recognized as the genus *Enteromorpha*). However, the phenotypic plasticity between tubular and foliose morphotypes is not solely genetic, but can be based on both abiotic and biotic factors (Wichard et al., **2015**). We generated DNA barcodes (*rbc*L, *tuf*A, ITS1) on 185 strains of distromatic foliose *Ulva* from the North East Atlantic, and used data and species delimitation from our previous study containing another 110 strains (Fort, McHale, et al., **2021**), as a primer for large-scale phylogenetic analysis of all *Ulva* sequences for the three common barcodes present in the NCBI database. The main aim of this study was to develop an analytical framework allowing the extent of misannotations in the sequences of any taxa of interest to be highlighted, taking as proof of concept the case of distromatic foliose *Ulva* species. We provide a detailed view of the phylogenetic relationships and possible misannotations between all sequences in the NCBI database, and propose readjustment for misannotated NCBI accessions, a list of appropriate reference vouchers for large foliose species, and a nomenclature adjustment between certain *Ulva* species. Finally, we employed the same analytical framework for three other seaweed genera, *Fucus*, *Porphyra* and *Pyropia* and identified clades containing misannotations and potential new synonymies.

2 MATERIALS AND METHODS

2.1 Foliose *Ulva* sample collection and DNA extraction

We collected individual thalli from foliose *Ulva* individuals with a thalli area >1,000 mm^2 in 34 sites in Ireland, Brittany (France), Spain, Portugal, the United Kingdom and the Netherlands between June 2017 and September 2019. The list of strains and associated metadata are available in Table **S1**. A total of 185 strains were collected for this study. On collection, samples were placed in clip-seal bags filled with local seawater and sent to Ireland in cold insulated boxes. On arrival, thalli were thoroughly washed with artificial seawater and a ~50 mm^2 piece of biomass collected and placed in screw caps tubes (Micronic). The tubes were immediately flash-frozen in liquid nitrogen and stored at –80°C. Then, samples were freeze dried, ground to a fine powder using a ball mill (Qiagen TissueLyser II), and ~5 mg of powder used for DNA extraction, using the magnetic-beads protocol described in Fort et al. (**2018**).

2.2 DNA amplification and sanger sequencing

The extracted DNA was amplified using three different primers combinations to obtain partial sequences for the nuclear 45S rRNA repeats (ITS1), as well as the chloroplast *rbc*L and *tuf*A barcodes. The primers used in this study are available in Table **S2**, and originate from (Heesch et al., **2009**) and (Saunders & Kucera, **2010**) for *rbc*L and *tuf*A, respectively. The ITS1 primers

were designed from the data set obtained in Fort, McHale, et al. (2021), and used in Fort, Linderhof, et al. (2021). PCR amplification was performed in 25 µl reaction volume containing 1 µl of undiluted DNA, 0.65 µl of 20 pmol forward and reverse primers, 9.25 µl of miliQ water and 12.5 µl of MyTaq Red mix (Bioline). The PCR protocol used 35 cycles of denaturation at 95°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 30 s. PCR products were precipitated using 2.5 volumes of 100% EtOH and 0.1 volume of 7.5 M ammonium acetate and incubated on ice for 30 min. Pellets were centrifuged at 4,000 *g* for 30 min at 4°C, and washed twice with 75% EtOH. Finally, PCR amplicons were sent to LGC Genomics GmbH for Sanger sequencing using the forward primer for each barcode.

2.3 Data set compilation for phylogenetic analyses

Our phylogenetic analysis aimed to consider all sequences attributed to *Ulva* species (foliose and tubular) in the NCBI database, including tubular and partially tubular species, and detect any evidence of species misannotation therein. We designed an analysis pipeline that could be used in any other taxa of interest, summarised in Figure 1. Command line codes and links to download the software used are available in Appendix S1. We downloaded all available sequences in the NCBI for ITS, *rbc*L and *tuf*A (as of 13 July 2020), in addition to the sequences from our previous study (Fort, McHale, et al., 2021). The search keywords were as follows: "*Ulva* (organism) AND internal transcribed" for ITS sequences, "*Ulva* (organism) AND *rbc*L (gene) AND plastid (filter)" for *rbc*L sequences, and "*Ulva* (organism) AND tufa (gene) AND plastid (filter)" for *tuf*A sequences. This search strategy yielded 1,679 ITS sequences (1,975 in total including this study and Fort, McHale, et al. (2021), 1,432 *rbc*L sequences (1,732 in total) and 1,114 *tuf*A sequences (1,393 sequences in total).

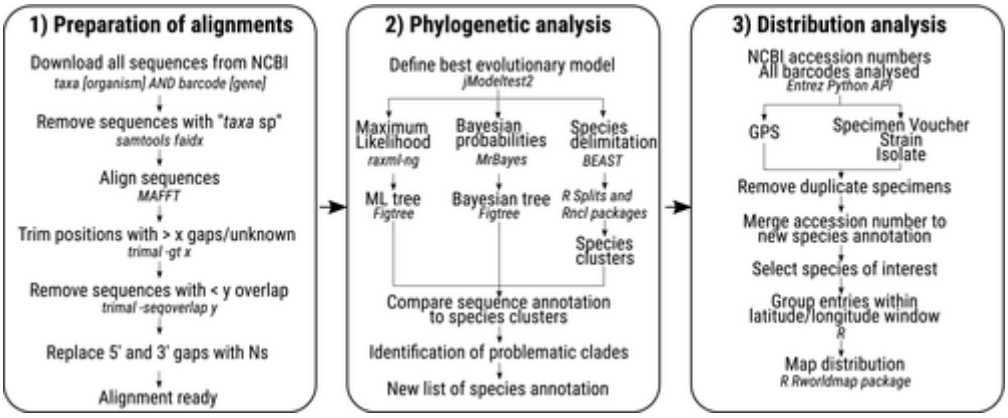


FIGURE 1

Open in figure viewerPowerPoint

Analysis framework used in this study. The list of scripts and software is available in Appendix S1

National Center for Biotechnology Information entries that did not contain species information (containing "*Ulva* sp." as organism) were then removed from the data set, by selecting all sequences not containing "*Ulva* sp." in their title, and using Samtools faidx (Li et al., **2009**) to extract their corresponding sequences. This filtering yielded 1,726, 1,312 and 1,321 sequences for ITS1, *rbc*L and *tuf*A, respectively. Sequences were then aligned using MAFFT (Katoh et al., **2019**) using the default settings for *rbc*L and *tuf*A, and the iterative FFT-NS-i method for the ITS1 alignment, due to the numerous gaps present. Because each study might amplify a slightly different portion of the barcodes due to the use of different primers, we then removed nucleotide positions that were absent in (i) more than 60% of the sequences using Trimal (Capella-Gutiérrez et al., **2009**) -gt 0.4 for *rbc*L and *tuf*A, and (ii) in more than 91% of the sequences for ITS1 (Trimal -gt 0.09). This step effectively trimmed the 5' and 3' ends of the alignment as to retain informative nucleotides, thereby avoiding large missing positions due to the use of different primers in different studies. Sequences containing more than 50% unknown bases in the trimmed alignments were then removed using Trimal (trimal -seqoverlap 50) (for *rbc*L and *tuf*A), and more than 70% unknown bases for the ITS1 alignment (trimal -seqoverlap 70). The use of two different filtering methods between the organellar barcodes (*rbc*L and *tuf*A) and ITS1 was because the ITS1 alignment contains gaps that are biologically relevant (the ITS1 length varies between species), while *rbc*L and *tuf*A coding sequences generally do not vary in length, but only in sequence. The filtering steps yielded final alignments containing 1,245 sequences (270 bp), 1,062 sequences (1,231 bp) and 1,320 sequences (801 bp) for ITS1, *rbc*L and *tuf*A, respectively. The 5′ and 3′ gaps introduced by the presence of missing positions in some of the sequences due to missing data were modified into "n" (i.e., unknown) bases. The missing nucleotides at the beginning and end of the sequences were due to the use of different primers (or sequencing length), and not to genetically relevant differences.

The *Fucus* and *Poyphyra* +Pyropia data sets were generated as above, using the search terms "Fucus (organism) AND (COI [gene] OR COX1[gene])", "Fucus (organism) AND internal transcribed", and "(porphyra [organism] OR pyropia [organism] OR neoporphyra [organism] OR neopyropia [organism]) AND (COX1[gene] OR COI[gene])". The final alignment data sets contained 174 sequences for *Fucus* COI, 452 sequences for Fucus nrRNA-ITS and 1,296 sequences for *Porphyra* + Pyropia COI/COX1. We kept entries with no taxonomically accepted names to encompass all genetic information available for those clades.

2.4 Phylogenetic analyses

We used both maximum likelihood and Bayesian MCMC phylogenetic analyses for the ITS1, *rbc*L and *tuf*A data sets to create maximum likelihood and Bayesian trees for each barcode. First, the best evolutionary model for each of the three alignments was determined based on their Akaike information criterion (AIC) score using jModeltest 2 (Darriba et al., **2012**; Posada & Buckley, **2004**). For all three alignments, general time reversible +gamma distribution +proportion of invariants sites (GTR + G + I) was deemed the most appropriate. Maximum likelihood trees were obtained using RAxML-NG (Kozlov et al., **2019**) using the "--all" option (20 maximum likelihood inferences, then bootstrap trees). Bootstrapping was stopped automatically using a MRE-based Bootstopping Test (Pattengale et al., **2010**) once reaching convergence values below 0.03. Bootstrap values were computed using the "--bs-metric tbe" option, representing transfer bootstrap expectation (TBE) values, expected to produce higher support for large trees with hundreds of sequences (Lemoine et al., **2018**), compared with classical Felsenstein bootstrap proportions (FBP). Bayesian MCMC analyses were performed using MrBayes with MPI support (Ronquist et al., **2012**), with a varying number of generations between the three data sets, until the average standard deviation of split frequencies reached a maximum of 0.05, and estimated sample sizes (ESSs) were higher than 200 for all parameters.

For species delimitation, we used the same method as per Fort et al. (**2019**) and Fort, McHale, et al. (**2021**), with a general mixed yule coalescent model (Fujisawa & Barraclough, **2013**; Pons et al., **2006**) in BEAST, and 50 million Markov Chain Monte Carlo (MCMC), using the BEAGLE library for decreasing computational time (Suchard & Rambaut, **2009**). Convergence was confirmed in Tracer (Rambaut et al., **2018**), with an ESS score >200 for all relevant parameters. Species delimitation was performed using the Rncl and Splits packages in R (Fujisawa & Barraclough, **2013**). All trees were visualised using Figtree (**http://tree.bio.ed.ac.uk/software/figtree/**), and annotated in Inkscape (**https://inkscape.org/**).

For detecting putative species disagreement within clades, all species names of the accessions present within GMYC clusters were compared and a percentage agreement metric per cluster was generated. For each cluster, the maximum number of accessions with the same species names was divided by the total number of accessions within the clade. This ratio indicates how divergent species names are within the GMYC clade, and all clades below 100% agreement can indicate a possible misannotation or new synonymies. The R script to generate the species delimitation and this ratio is available in Appendix **S2**.

2.5 Taxonomic assignment of sequence names

Regarding foliose *Ulva* species, since several species names have been found to be synonymous, and we used the species names listed in Table **1** as our reference. Where holotype or lectotype reference sequences were available, we attributed the species names of the reference to all sequences within the same GMYC clade. Where such type sequences were not available, we based our species attribution with comparisons from sequences from the literature and the GMYC clustering, with the caveat that indeed the nomenclature of the GMYC clade could change once holotype sequences become available. The rationale behind the selection of reference sequences is detailed in Appendix **S1**.

TABLE 1. Names and synonyms used in this study

Species	Synonymous name	Reference
Ulva lactuca Linnaeus	*Ulva fasciata* Delile	Hughey et al. (2019)
Ulva australis Areschoug	*Ulva pertusa* Kjellman, *Ulva laetevirens* Areschoug	Kraft et al. (2010), Hughey et al. (2021)
Ulva compressa Linnaeus	*Ulva mutabilis* Föyn	Steinhagen et al. (2019)
Ulva expansa (Setchell) Setchell & N.L. Gardner	*Ulva lobata* (Kützing) Harvey	Hughey et al. (2019)
Ulva lacinulata (Kützing) Wittrock	*Ulva scandinavica* Bliding, *Ulva armoricana* Dion, Reviers & Coat, 1998	Hughey et al. (2021), this study

2.6 Species distribution of distromatic foliose *Ulva* species

The country of origin, GPS coordinates, specimen name and publication name of all of the NCBI entries in the three data sets were recovered using custom python scripts (Appendix **S3** and **S4**), restricted to vouchers assigned in our analysis as belonging to the 11 main distromatic foliose *Ulva* species (namely, *U. australis* Areschoug, *U. fenestrata* Postels & Ruprecht, *U. lactuca* Linnaeus, *U. gigantea* [Kützing] Bliding, *U. lacinulata* [Kützing] Wittrock, *U.ohnoi* M. Hiraoka & S. Shimada, *U. rigida* C. Agardh, *U. expansa* [Setchell] Setchell & N. L. Gardner, *U. arasakii* Chihara and *U. ohiohilulu* H. L. Spalding & A. R. Sherwood), and *Ulva* sp. *A*. Publications associated with NCBI entries missing GPS coordinates and/or location of origin were manually searched to retrieve GPS coordinates where

available. Accessions whose area of origin were uncertain were removed from the analysis. Duplicated specimens (i.e., specimens with more than one barcode sequenced in the NCBI) were removed and only one entry was kept. The complete list of vouchers, specimen, name, publication, GPS coordinates and proposed species attribution is available in Table S3. The world map and pie-chart distribution of *Ulva* species was created in R using the package Rworldmap (South, 2011).

3 RESULTS

Using the analysis pipeline we created, recovered and analysed all *Ulva* sequences in the NCBI, as well as 185 additional strains from the North-East Atlantic sequenced in this study, for the three most common barcodes used in *Ulva* phylogeny, namely *rbc*L, *tuf*A and ITS1.

3.1 Analysis of all *Ulva* spp. *rbc*L sequences from public repositories

We used the *rbc*L data set generated in this study, that from Fort, McHale, et al. (2021), as well as all available *rbc*L sequences from *Ulva* entries in the NCBI (see Materials and Methods). From the *rbc*L alignment, we generated a maximum likelihood phylogenetic tree containing 1,062 sequences. GMYC analysis revealed the presence of 24 clades containing more than two sequences (confidence interval 19–28) (Figure 2). Of these, 10 belong to obligatory distromatic foliose species, namely *U. arasakii*, *U.* sp. *A*, *U. expansa*, *U. fenestrata*, *U. australis*, *U. gigantea*, *U. ohnoi*, *U. lactuca*, *U. rigida* and *U. lacinulata* (Table 2). The GMYC species delimitation, however, failed to discriminate between five species. *U. lacinulata* and *U.* sp *A* were found to be conspecific, despite previous evidence to the contrary (Fort, McHale, et al., 2021; Heesch et al., 2009), as well as a single clade containing both *U. lactuca* and *U. ohnoi*, and another clade containing *U. rigida* and *U. adhaerens*. The full maximum likelihood tree (including bootstrap support), the Bayesian MCMC analysis tree (including probabilities), and entries species names for *rbc*L can be found in Figure S1, and Table S3).

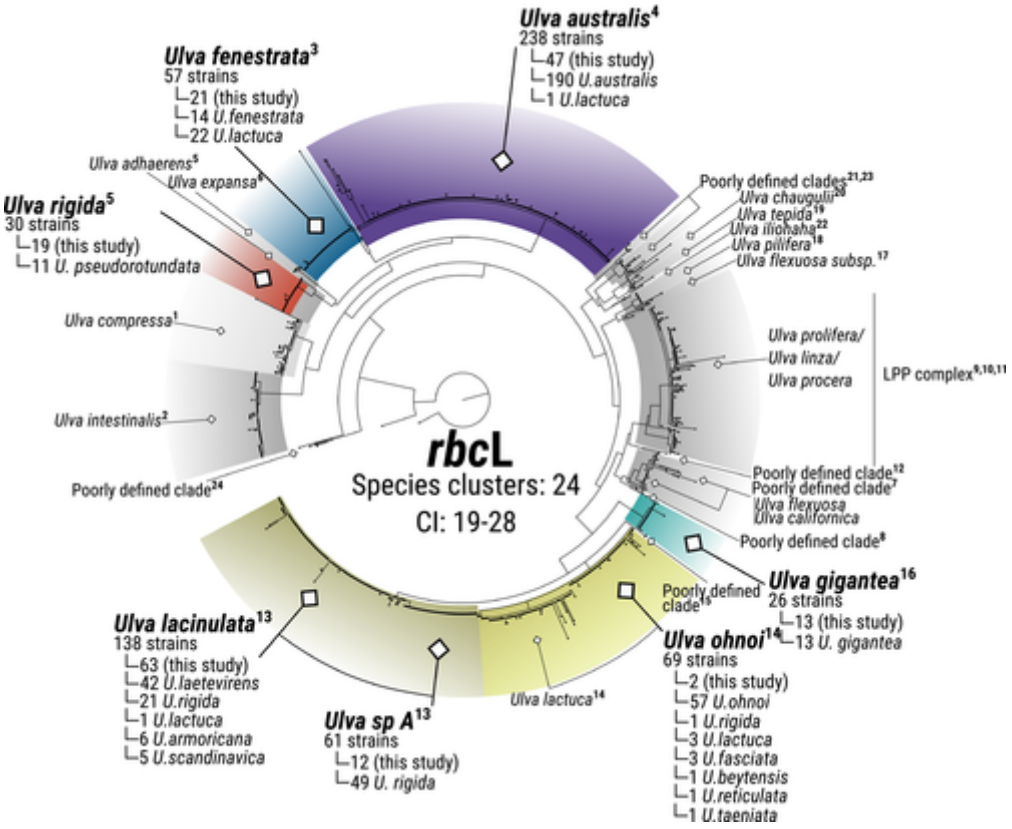


FIGURE 2

[Open in figure viewerPowerPoint]

Maximum likelihood tree of the *rbc*L alignment, rooted on *Umbraulva* sequences. Coloured clades represent distromatic foliose species found in this study. Shaded clades represent tubular or partially tubular species and/or species with no representative in this study. Numbers, shaded and/or coloured clades represent species clusters determined using GMYC. Full trees including bootstrap values and bayesian posterior probabilities are available in Figure S1

TABLE 2. Summary of sequences and annotations available for foliose *Ulva* species

Species clade	ITS1			*rbc*L			*tuf*A		
	This study	Total	% misannotated	This study	Total	% misannotated	This study	Total	% misann
U. arasakii	0	11	0	0	1	0	0	10	0
U. lacinulata	62	134	49.3	63	138	46.4	59	140	57.9
U. australis	48	90	2.2	47	238	0.4	43	175	0
U. expansa	0	4	0	0	6	0	0	32	0

Species clade	ITS1			*rbc*L			*tuf*A		
	This study	Total	% misannotated	This study	Total	% misannotated	This study	Total	% misann
U. fenestrata	21	53	37.7	21	57	38.6	21	225	38.2
U. gigantea	15	25	0	13	26	0	14	32	0
U. lactuca	0	26	38.5	0	58	17.2	0	16	37.5
U. ohiohilulu	0	0	0	0	0	0	0	9	0
U. ohnoi	3	35	11.4	2	69	14.5	3	92	4.3
U. rigida	18	26	28	20	30	37	18	24	25
U. sp. A	15	50	70	12	61	80	11	40	73

The 177 *rbc*L sequences from this study originating from the North East Atlantic belong to seven distinct clades, with 19 samples identified as *U. rigida*, 21 samples as *U. fenestrata*, 47 as *U. australis*, 13 as *U. gigantea*, two as *U. ohnoi*, 12 as *U.* sp. *A* and 63 as *U. lacinulata*.

The clades containing *U. australis* and *U. gigantea* are the most consistent, with minimal discrepancies between species names within the clades. The other clades appear more problematic, with significant species names discrepancies in the *U. fenestrata*, *U. ohnoi*, *U. lacinulata* and *U. rigida* clades (Figure S2).

We found 69 strains belonging to the *U. ohnoi* clade, two in this study, 57 *U. ohnoi* vouchers from the NCBI database (described in Hiraoka et al., 2004; Krupnik et al., 2018; Melton et al., 2016, including the type), as well as several likely misannotated entries, including one *U. rigida*, three *U. lactuca*, three *U. fasciata*, one *U. beytensis* Thivy & Sharma, one *U. reticulata* Forsskål and one *U. taeniata* (Setchell) Setchell & N. L. Gardner. Most entries originated from the same unpublished population set (number 452119310). Next, the *U.* sp. *A* clade contained 12 strains from this study, as well as 49 *U. rigida* entries from the NCBI, described in Heesch et al. (2009); Rautenberger et al. (2015) and Loughnane et al. (2008). Finally, the *U. lacinulata* clade containing 138 strains appeared to contain several cases of likely species misidentification. This clade contained 63 individuals from this study, 38 individuals from Fort, McHale, et al. (2021) (which are now renamed *U. lacinulata* following nomenclatural reassignment [Hughey et al., 2021]) and four *U. laetevirens* entries [two from

(Kraft et al., 2010), and two from China (Du et al., 2014)]. However, 21 entries in the *U. lacinulata* clade were assigned as *U. rigida*. The presence of a large number of *U. laetevirens* individuals in this clade stems from the recent sequencing of the *U. laetevirens* holotype (Hughey et al., 2021), which was found to belong to *U. australis*. Thus, the sequences formerly known as *U. laetevirens* should now be reclassified as *U. lacinulata*, whose sequenced holotype belong to the same clade (Hughey et al., 2021). Interestingly, all five *U. scandinavica* entries also cluster within the *U. lacinulata* clade, with two out of five *U. scandinavica* entries being indistinguishable from *U. lacinulata* ones, and the other three possessing a single polymorphic site. Altogether, *U. scandinavica* are likely to be synonymous with *U. lacinulata*. Finally, following nomenclatural adjustment via sequencing of the lectotype (Hughey et al., 2021), the *U. rigida* clade contains *U. pseudorotundata* sequences.

Of the large foliose species not represented in our data set, *U. arasakii* is represented by a single individual, and the *U. expansa* clade contains six NCBI entries, four *U. expansa* and two *U. lobata*, which have been shown to be synonymous (Hughey et al., 2019), Table 1. Concerning other species, *U. compressa* Linnaeus and *U. intestinalis* Linnaeus are well defined, with no misidentification of *U. intestinalis*, and only three likely misannotated sequences in the *U. compressa* clade: one *U. intestinalis* and two *U. pseudocurvata* entries. The other species are more problematic, with several poorly defined clades containing a mixture of *U. prolifera*, *U. linza*, *U. flexuosa*, *U. californica* and *U. tanneri*.

Altogether, we found a relatively low agreement between the species names assigned to the NCBI vouchers and the GMYC clusters for *rbc*L, with only seven out of 24 GMYC clusters containing 100% of sequences with the same species name annotation (Figure S2). Disagreements between GMYC clades and species names within them do not necessarily indicate misannotations, due to poor detection of species boundaries by the GMYC analysis using this barcode. Nonetheless, the results show that *rbc*L sequences are probably poor at defining *Ulva* species, and that each clade should be investigated in detail, as significant naming discrepancies are present.

3.2 Analysis of all *tuf*A sequences from public repositories

We performed the same analysis using the *tuf*A barcode (Figure 3, Figure S3 and Table S3). We found significantly more species clusters than for the *rbc*L barcode (40 species clusters, confidence interval 37–46).

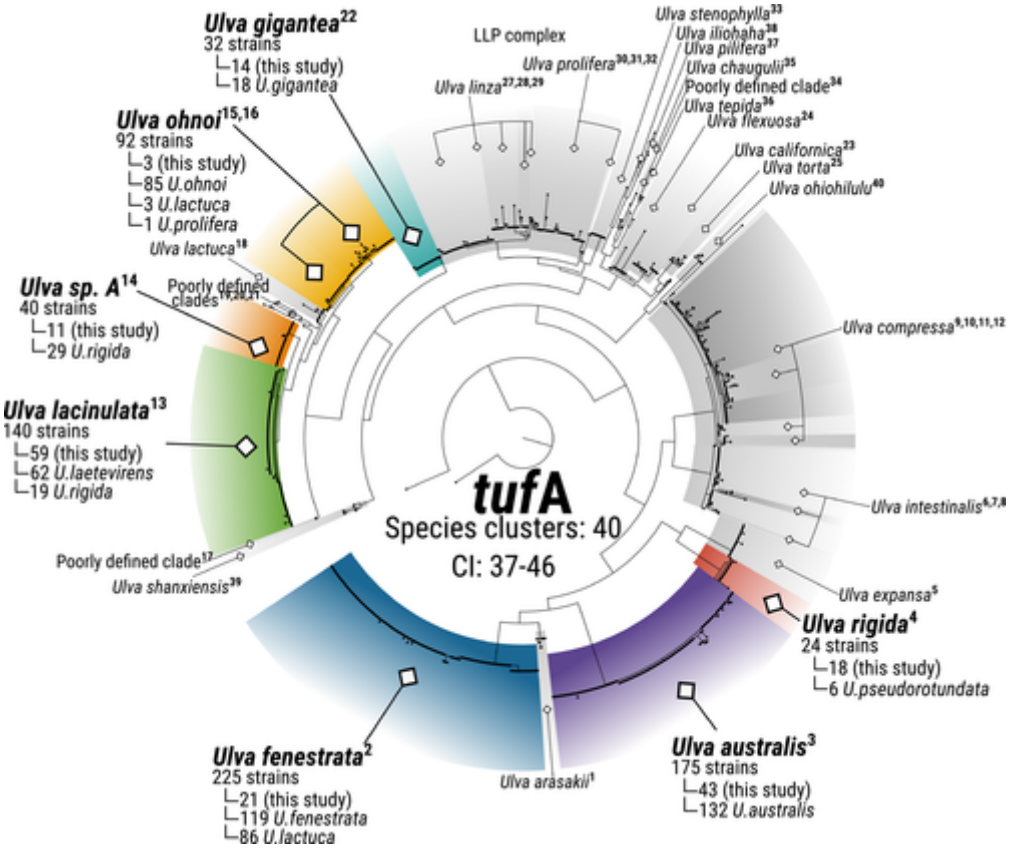


FIGURE 3

Open in figure viewerPowerPoint

Maximum Likelihood phylogenetic tree of 1,320 *Ulva* spp. *tuf*A sequences, and description of the entries belonging to the main distromatic foliose *Ulva* species. Maximum likelihood tree of the *tuf*A alignment, rooted on *Umbraulva* sequences. Coloured clades represent distromatic foliose species found in this study. Shaded clades represent tubular or partially tubular species and/or species with no representative in this study. Numbers, shaded and/or coloured clades represent species clusters determined using GMYC. Full trees including bootstrap values and bayesian posterior probabilities are available in Figure S3

For foliose species (Table 2), as expected, the *U. fenestrata* clade shows the same name misapplication with *U. lactuca*, with 225 individuals, 21 in this study, 119 *U. fenestrata* entries and 86 *U. lactuca* entries. *U. australis*, *U. gigantea tuf*A clades appear well defined, with no name misapplication, similar to the *rbc*L results. *U. ohnoi* is also generally well circumscribed. The *U. lacinulata* and *U.* sp. *A* clades are separated by the GMYC analysis using *tuf*A, and 19 *U. rigida* sequences are clustering within the *U. lacinulata* clade. Less common foliose species, such as *U. expansa*, *U. arasakii* and *U. ohiohilulu* are represented with more than two entries, each with their separate clades.

For other species, *tuf*A appears more appropriate than *rbc*L for species delimitation, with a clear separation between *U. linza* and *U. prolifera* clades, as well as between *U. californica* and *U. flexuosa*, without apparent misidentifications apart from one *U. mediterranea* Alongi, Cormaci & G. Furnari and one *U. prolifera* vouchers, both displaying 100% identity with *U. flexuosa*. *Ulva compressa* and *U. intestinalis* are similarly well defined in the *tuf*A data set.

Consequently, the percentage agreement of species names within GMYC clusters in the *tuf*A data set is significantly higher than with *rbc*L, with 23/40 GMYC clusters showing complete agreement (Figure S2).

3.3 Analysis of all ITS1 sequences from public repositories

Finally, the analysis was repeated on the ITS1 barcode data set (Figure 4, Figure S4 and Table S3). Once again, the results are in general agreement with the previous barcodes, particularly with *tuf*A. Indeed, species delimitation predicts 42 species clusters (compared with 40 with *tuf*A), with a confidence interval of 34–59.

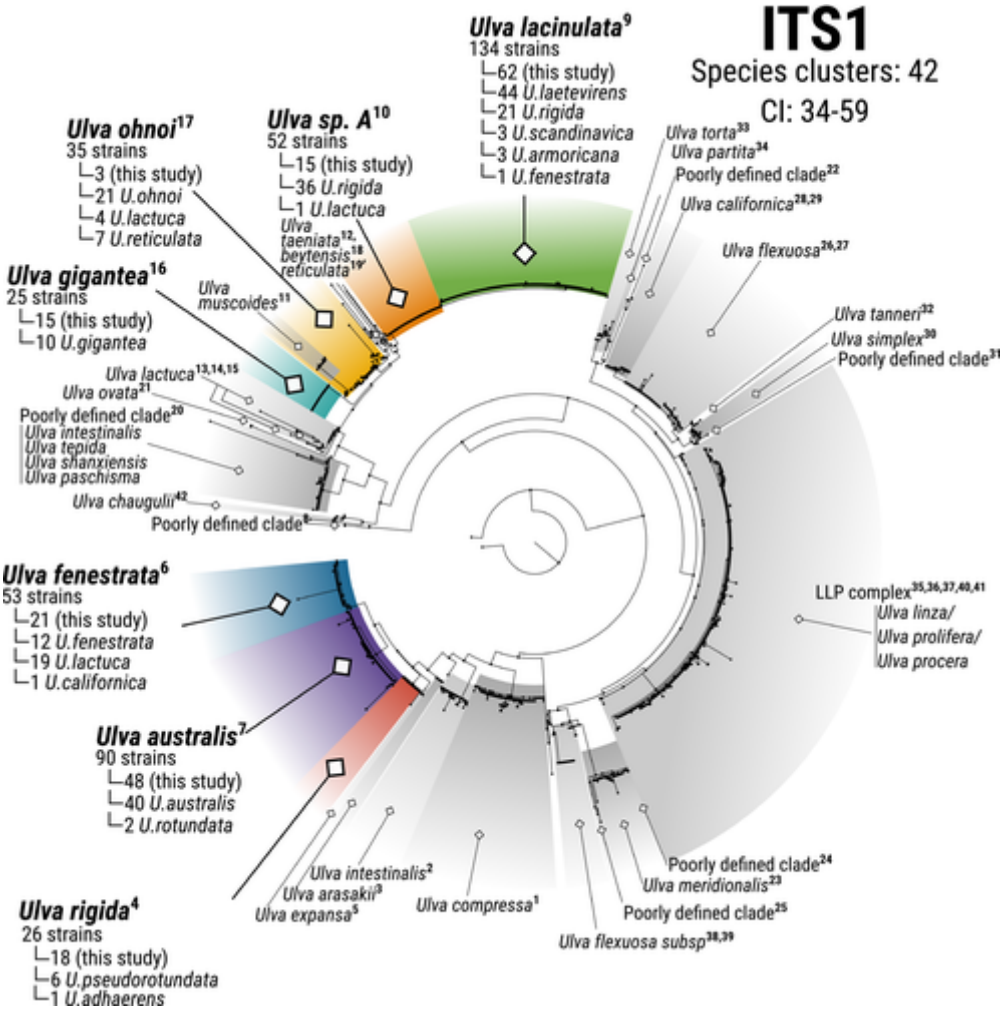


FIGURE 4

Open in figure viewerPowerPoint

Maximum likelihood phylogenetic tree of 1,245 *Ulva* spp. ITS1 sequences, and description of the entries belonging to the main distromatic foliose *Ulva* species. Maximum likelihood tree of the ITS1 alignment, rooted on *Umbraulva* sequences. Coloured clades represent distromatic foliose species found in this study. Shaded clades represent tubular or partially tubular species and/or species with no representative in this study. Numbers, shaded and/or colored clades represent species clusters determined using GMYC. Full trees including bootstrap values and Bayesian posterior probabilities are available in Figure S4

The *U. australis*, *U. gigantea* and *U. ohnoi* clades are well conserved, with only minor discrepancies (Table **2**). The *U. fenestrata* clade however contains 33 *U. fenestrata* accessions and 19 erroneous *U. lactuca* accessions. The *U. lacinulata* clade contains 134 sequences with 62 from this study, the holotype of *U. armoricana* (NCBI accession **MT078962**; Coat et al., **1998**), and 44 *U. laetevirens*. As for the *rbc*L results, we found *U. scandinavica* within the *U. lacinulata* clade, all of which show 100% identity with most other *U. lacinulata* sequences.

With regard to narrow-tubular species, the "Linza-Procera-Prolifera" (LPP) complex is poorly delimited, with NCBI entries of all three species intertwined within five clades. Outside of the LPP complex, other narrow-tubular *Ulva* species appear well delimited, with two exceptions. The *U. meridionalis* R. Horimoto & S. Shimada (Horimoto et al., **2011**) clade contains 12 probably misannotated *U. prolifera* vouchers. Similarly, the clade containing *U. tepida* Y. Masakiyo & S. Shimada contains several entries annotated as *U. intestinalis*, *U. shanxiensis* L. Chen, J. Feng & S. -L. Xie *and U. paschisma* F. Bast.

Out of 42 GMYC clusters, only 14 show complete agreement in species names (Figure **S2**). This shows that a significant number of misannotations are probably present in the ITS sequences of the *Ulva* genera.

3.4 Impact of NCBI accession reanalysis on species distribution

After reassigning species name for each NCBI entry, we generated a world map of the distribution of the eleven large foliose *Ulva* species from which there is genetic evidence (Figure **S5**). Strikingly, no *U. lactuca* individuals are present in the North Atlantic and the Baltic Sea, outside of a specimen recovered from an aquarium and misannotated as *U. laetevirens* (Vranken et al., **2018**), and a single specimen in Massachusetts, USA. As shown above, the reports of *U. lactuca* in many regions are all referable to *U. fenestrata*. Importantly, while the number of misannotations in the NCBI is significant, the problem is even higher in other databases that do not rely on DNA sequencing for reporting species records. For

instance, the Ocean Biodiversity Information System (OBIS) contains >4,700 records for *U. lactuca*, most of which located in the North Atlantic, in seeming contradiction with our results (Figure 5). Hence, reanalysis of barcode sequences can drastically change species distribution.

FIGURE 5

Open in figure viewerPowerPoint

Comparison of *U. lactuca* species distribution based on different databases. Each dot represents a single record

3.5 **Extension of the analytical framework with *Fucus, Porphyra*** and ***Pyropia* spp**.

We used the same analytical pipeline to detect possible misannotations or new synonymies in three other genera of economically and ecologically important macroalgae: *Fucus* spp. (*Phaeophyceae*, *Fucaceae*), and two Bangiales genera, *Porphyra* and *Pyropia* spp.

For *Fucus* spp., we used all publicly available sequences for the *COX1* and nrRNA-ITS barcodes, and generated a maximum likelihood tree and species delimitation as for the *Ulva* data sets. The GMYC analysis predicts eight and nine species for *COX1* and nrRNA-ITS sequences, respectively (Figure 6), with the *Fucus distichus* clade being split into six different predicted species by the GMYC analysis of *COX1* sequences. For the nrRNA-ITS data set, the clades containing *F. serratus* and *F. vesiculosus* species names are separated into two and four predicted clades, respectively. Overall, the species names within the GMYC clusters are well conserved, with 5/8 and 7/9 clusters displaying 100% agreement (Figure S2). However, one clade in both barcode data sets appears problematic. *F. vesiculosus* and *F. spiralis* sequences are intertwined in both data sets. This indicates that the two species are frequently misannotated. Indeed, sequences with both names are in some cases

indistinguishable, with 100% identity. The full maximum likelihood trees are available in Figure **S6**.

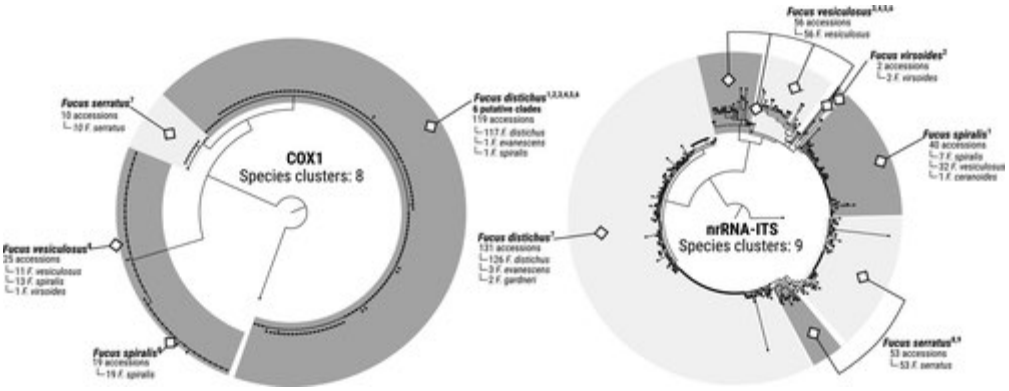


FIGURE 6

Open in figure viewerPowerPoint

Maximum likelihood phylogenetic tree of *Fucus* spp. COX1 and nrRNA-ITS sequences. Numbers and shaded clades represent species clusters determined using GMYC. Full ML trees are available in Figure S6

The *Porphyra* and *Pyropia* data set contains 1,296 *COX1* sequences, separated into 62 GMYC clusters (Figure **7**, full tree available in Figure **S7**). Unlike *Ulva*, the species names within GMYC clusters appear remarkably consistent in this data set, with only twelve out of 62 GMYC clusters containing sequences with different species names (Figure **S2**). Furthermore, most of those relate to clusters containing vouchers with undetermined species names, hence do not represent misannotations *per se*. Only one clade is potentially problematic, with sequences named either *Porphyra linearis* or *Porphyra umbilicalis*, despite being identical in sequence.

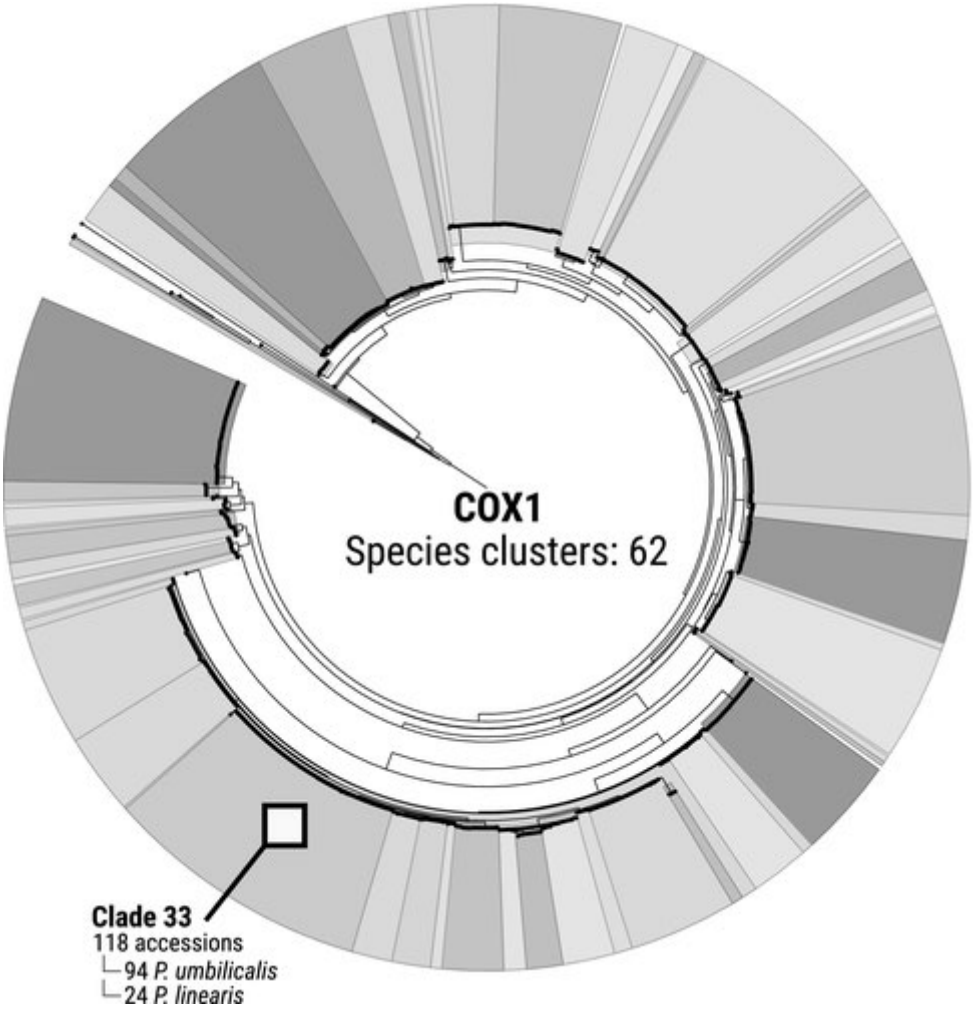


FIGURE 7

Open in figure viewerPowerPoint

Maximum Likelihood phylogenetic tree of *Porphyra* + *Pyropia* COX1 sequences. Shaded clades represent species clusters determined using GMYC. Full ML tree is available in Figure S7

Altogether, the three additional data sets show a lower extent of potential misannotations than the *Ulva* data sets, even when using a species-rich family such as the Bangiaceae. We generated a histogram of the percentage of agreement in the species names of all GMYC clusters between the three groups of species investigated here (Figure **8**), which shows a significant number of GMYC clusters below 100% agreement in *Ulva*, compared to *Fucus*, *Porphyra*, and *Pyropia* data sets.

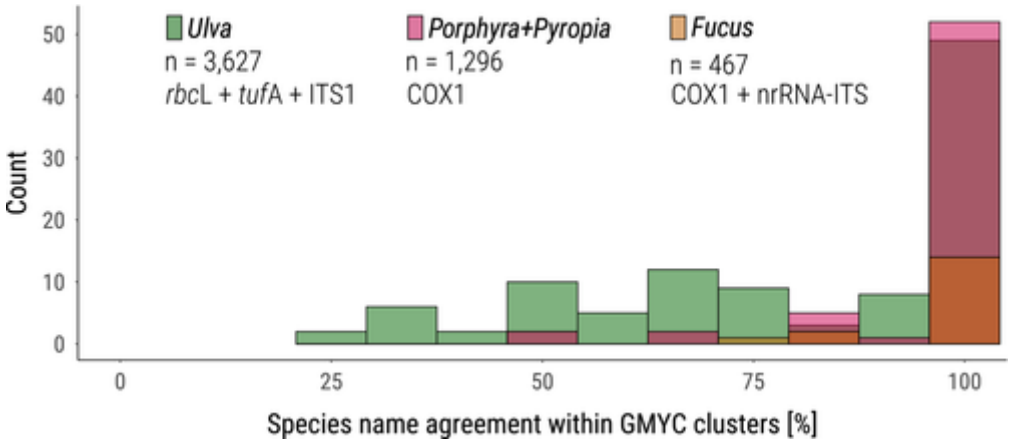


FIGURE 8

Open in figure viewerPowerPoint

Distribution of species names agreement per GMYC cluster between *Ulva*, *Fucus* and *Porphyra* + Pyropia data sets

4 DISCUSSION

4.1 Limitations of species delimitation using single barcodes

In this study, we endeavoured exhaustively to assess the genetic information available for our taxa of interest. We used all publicly available sequences from the NCBI for three common barcodes. Notably, species delimitation using such a large number of sequences yields relatively large species clusters confidence intervals. For instance, using *rbc*L alone did not allow to separate certain taxa that were previously shown to be separate species (Fort, McHale, et al., 2021; Hiraoka et al., 2004; Hughey et al., 2019), such as *U.* sp. *A* and *U. lacinulata* or *U. ohnoi* and *U. lactuca*. This could be due to the use of a smaller length of alignment for *rbc*L in this study, as opposed to concatenated *rbc*L + *tuf*A sequences in Fort, McHale, et al. (2021) for the *U.* sp. *A*/*U. armoricana* separation. In addition, such a discrepancy is inherent to large-scale species delimitation analyses when using limited genetic information (Leliaert et al., 2014; Tang et al., 2014). Indeed, the presence of possibly spurious sequences in the entire data set can skew the speciation threshold of the GMYC analysis, especially when a single barcode containing a limited number of SNPs between species is used. This probably explains the relatively large confidence intervals we observed for *rbc*L. In contrast, using *tuf*A alone we were able to separate *U. lacinulata* and *U.* sp. *A*, which is in agreement with previous studies (Fort, McHale, et al., 2021; Hayden & Waaland, 2002; Heesch et al., 2009; Tan et al., 1999). *tuf*A displays more SNPs than *rbc*L when comparing those two species (nine vs. two, respectively), allowing for a species delimitation between the two clades. The ITS1 barcode similarly allowed for the separation of

those two species. However, while we are able to separate *U. lactuca* and *U. ohnoi* using *tuf*A, *U. ohnoi* is separated into two different clades. Similarly, *U. linza*, *U. compressa*, *U. intestinalis* and *U. prolifera* clades are separated into several clades. Finally, while seven *U. reticulata* vouchers originating from (Monotilla et al., **2018**), are included in the *U. ohnoi* clade using the ITS1 barcode, these probably do not represent erroneous annotation, since in their study, Monotilla et al. (**2018**) showed that *U. ohnoi* and *U. reticulata* are sexually isolated, despite having little to no sequence divergence in this barcode sequence.

Thus, appropriate species delimitation analysis should ideally be performed on a larger amount of genetic information, such as full organellar genomes, or concatenated sequences from the same specimens. Additionally, other species delimitation algorithms are available, such as Poisson tree processes (PTP) or the automatic barcode gap discovery for primary species delimitation (ABGD) (Puillandre et al., **2012**; Zhang et al., **2013**). It is likely that using different methodologies for species delimitation will yield a different number of species clades in the same data set, and a combination of approaches could be used to precisely delimitate all *Ulva* species. Regardless of precise species delimitation however, the methodology described here allows to quickly test putative clades and their associated sequence names for possible misannotations or new synonymies. Notably, the use of "agreement of species names within clade" (Figure **S2**, Figure **8**) from the GMYC output helps to identify potentially problematic clades and species names. It provides a visual representation of the diversity within the data set and serves as a steppingstone for in-depth reassessment of the taxonomy and diversity of genera of interest.

Regarding our findings with *Ulva*, the number of "species names" in the entries from the NCBI data set is 56, nine of which are classified as synonyms. Of the 47 unique species names remaining, this analysis, despite its limitations, found ~40 species clusters containing more than two sequences, thus broadly agreeing with the present number of species described in NCBI. These numbers are significantly lower than that of the number of currently accepted species taxonomically (84 according to (Guiry & Guiry, **2021**)). This apparent discrepancy could be explained by the presence of numerous species entities described morphologically in past studies from which there is no genetic evidence. These specimens should be sequenced if they are available, or their type locality resampled, as the NCBI database probably only contains a subset of all *Ulva* species.

4.2 Nomenclature, taxonomy and species misidentifications in public repositories

The main issue with the use of public repositories to assign species name to sequences is the underlying quality of the species annotation within the repository. Two issues can be present, a nomenclatural issue, where the naming of the taxa is erroneous, or taxonomical issues, where the relationships between taxa is at fault (de Queiroz, 2006). The analytical framework described here allows us to identify clades that contain sequences with different species names, which could represent new synonymies for nomenclatural adjustments, and/or detect problematic taxonomic relationships when sequences of the same species name are present in different clades. Importantly, both of those points do not require prior knowledge of the nomenclature or taxonomy of the genus. For example, the presence of a significant amount of *U. lactuca* sequences intertwined with *U. fenestrata* accessions in one clade highlights misannotation of many specimens of *U. lactuca*, while multiple clades containing only one species name could represent undescribed new taxa.

However, to resolve the nomenclatural issues highlighted requires the systematic sequencing of all available types or the designation of epitypes. This work in *Ulva* is currently underway by Hughey and colleagues, leading to nomenclatural adjustments of several species names (Hughey et al., ,2019, 2021; Hughey et al., 2021). For example, the clade described here as *U. lacinulata* was previously referred to as *U. laetevirens* and *U. armoricana* (Fort, McHale, et al., 2021; Kirkendale et al., 2013; Miladi et al., 2018). Following sequencing of the *U. laetevirens* lectotype (Hughey et al., 2021), the name *U. laetevirens* was found to be synonymous with *U. australis*. Recently, the sequencing of *U. lacinulata* lectotype revealed that it was the oldest valid and available name for this clade (Hughey et al., 2021). We therefore renamed our accessions as *U. lacinulata*. Furthermore, sequencing of the *U. rigida* lectotype revealed that it belongs to the clade previously known as *U. pseudorotundata*, for which the oldest available name is *U. rigida* (Hughey et al., 2021). Finally, given that the sequences previously assigned as *U. rigida* by us do not currently have an available name with a sequenced type, these sequences are provisionally referred to as *Ulva* sp. *A*. This highlights that nomenclature adjustments are likely to continue until all available types sequences become available, a huge task made more difficult by missing types and prohibitions on sampling of types by herbaria. Nonetheless, taxonomically, such adjustments do not impact the clustering of sequences into species clades and the analytical framework described here, which aims to provide an exhaustive view of sequences names, agreements, and species clusters for a genus of interest.

For instance, it was recently reported by Hughey et al. (2019) that several misidentifications were found within the *U. fenestrata* clade. Here, using all sequences available, we found that

this misidentification is indeed significant. Some 40% of sequences belonging to *U. fenestrata* are misannotated (127/334). Hence, caution should be exercised when comparing *U. fenestrata* sequences using BLAST since some of the best matches will erroneously be referred to "*U. lactuca*." We naturally support the use of *U. fenestrata* type as described by Hughey et al. (**2019**) as the baseline for this species (Table **3**). This significant amount of species misannotation lead to a drastic change in the species distribution of *U. lactuca* (Figure **5**) and should not be overlooked. Only *Ulva* products labelled as containing "*Ulva lactuca*" are officially authorized for food consumption in Europe outside of France (Barbier et al., **2019**). Furthermore, accurate description of the species used in the literature is essential for natural products biodiscovery, nutritional profile and traceability (Leal et al., **2016**). This highlights the need to both improve the identification of *Ulva* species and to change the European food regulation by inclusion of the *Ulva* species which are effectively consumed at present under the name of "*Ulva lactuca*" or to treat "Ulva lactuca" as a commercial name encompassing all foliose *Ulva* species.

TABLE 3. Proposed reference sequences for foliose *Ulva* species. Rationale available in Appendix S1

Species	NCBI ITS accession	NCBI *rbcL* accession	NCBI *tufA* accession	Reference
Ulva australis	MT894708	MT160564	MT160674	Fort, McHale, et al. (2021)
Ulva lacinulata	MW544060 [b]	MW543061 [b]	MT160697	Hughey et al. (2021); Fort, McHale, et al. (2021)
Ulva sp. A	MT894534	MT160573	MT160683	Fort, McHale, et al. (2021)
Ulva ohnoi	AB116031 [b]	AB116037 [b]	MT894753	Hiraoka et al. (2004); This study
Ulva rigida	MW544059 [b]	MW543060 [b]	MT160722	Hughey et al. (2021); Fort,

Species	NCBI ITS accession	NCBI *rbcL* accession	NCBI *tufA* accession	Reference
				McHale, et al. (2021)
Ulva gigantea	MT894480	MT160606	MT160716	Fort, McHale, et al. (2021); this study
Ulva lactuca	AY260561 [a]	MK456395 [b]	MF172082 [a]	Hayden et al. (2004); Hughey et al. (2019), Miladi et al. (2018)
Ulva fenestrata	MT894725	MK456393 [b]	MT160728	Fort, McHale, et al. (2021); Hughey et al. (2019)
Ulva arasakii	AB097650	AB097621	MK992126	Shimada et al. (2003); Kang et al. (2014)
Ulva expansa	MH730161 [b]	MH730975 [b]	MH731007 [b]	Hughey et al. (2018)
Ulva ohiohilulu	KT881224 [b]	KT932996 [b]	KT932977 [b]	Spalding et al. (2016)

- [a] Annotated as *U. fasciata*.
- [b] Holotype/lectotype sequence.

Finally, our study shows that *U. "rigida"* (now *U.* sp. *A*) and *U. lacinulata* are also commonly misannotated in public repositories, which was hinted by Miladi et al. (**2018**). It perhaps is not surprising since both species sequences are relatively close, with only a handful of discriminating SNPs contained within those three barcodes, and the viability of interspecific hybrids (Fort, Linderhof, et al., **2021**; Fort, McHale, et al., **2021**). However, previous species delimitation analysis on *rbc*L + *tuf*A using different methodologies (GMYC and bPTP), and the sequence identity differences between the organellar genomes of the two clades indicates

that they are probably two separate species (Fort, McHale, et al., **2021**), and not the single taxon as postulated by Hughey et al. (**2021**). While we consider that the *U. lacinulata* clade is fully resolved due to the presence of *U. lacinulata* type within the clade (Hughey et al., **2021**), the sequence of the *U*. sp. *A* type specimen is not currently available in public repositories. Hence, sequences of the *U*. sp. *A* clade will need to be renamed when a suitable type is found.

Overall, the analysis of large foliose *Ulva* species showed ~26% of misannotated entries in the NCBI database, a percentage probably much higher when tubular or partially tubular species are considered. A significant amount of the misannotations originates from recent nomenclature changes, which renders the work presented in this study particularly important, as we provide in Table **S3** all of the NCBI accession numbers of the foliose species highlighted here, as well as the updated species attribution. We encourage the *Ulva* scientific community to use the trees described here as potential "accession quality check" for species annotation based on BLAST results. In Figures **S1**, **S3** and **S4** we provide the trees of all three barcodes in order to allow researchers to use the search function of PDF viewers for searching specific NCBI accessions and identifying to which clade they belong. However, we encourage the use of exhaustive trees for phylogenetic analyses (i.e., including all available NCBI sequences), instead of trees containing a subset of "selected" NCBI entries. For example, a BLAST result of NCBI accession HQ610342.1 shows 11 matches with 100% identity, 10 of which are classified as *U. lactuca*. Therefore, if a tree was generated using the first five NCBI hits as reference, the sequence will probably be classified as *U. lactuca*. Conversely, using the entire NCBI data set highlights that all of those *U. lactuca* sequences are misannotated *U. fenestrata*. Including all sequences leads to a significant increase in computational time, but with the use of multithreading by raxml-NG and MrBayes, and the BEAGLE library for BEAST, we found that generating trees and GMYC analyses with >1,000 sequences takes ~48 h on eight CPU cores, decreasing further to ~10 h with 64 CPU cores.

Nevertheless, in Table **3** we suggest a list of reference NCBI accessions for all three barcodes of the 11 large foliose *Ulva* species. The rationale for this list is available in Appendix **S1**. As it is simple to update the information associated to NCBI sequences (see **https://www.ncbi.nlm.nih.gov/genbank/update/**), we encourage authors that have deposited sequences on the NCBI to update, if incorrect, the "organism" information of their accession numbers, thus avoiding the amplification and recurrence of misannotated *Ulva* species, such as *U. lactuca*, and to update taxonomic assignments due to nomenclatural adjustments.

Concerning tubular and or partially tubular species, the major hurdle found here lies within the separation of *U. linza*, *U. procera* and *U. prolifera* individuals. This appears to be an ongoing issue with the delimitation of the species within the Linza-Procera-Prolifera (LPP) complex (Cui et al., **2018**; Kang et al., **2014**; Leliaert et al., **2009**), and will require further reanalysis of the NCBI entries after organelle sequencing of holotype specimens. The precise species delimitation of those clusters is outside the scope of this study but indicates that caution should also be taken when analysing the sequences of those species, as misidentifications are likely to be present.

The taxonomic groups described here could also be used to study possible introduction event(s) of non-native species. Notably, Sauriau et al. (**2021**) recently questioned the introduction of *U. australis* in Europe by using all available NCBI sequences of *U. australis* to infer introduction events. Indeed, the separation of sequences from a given species into haplotypes allows for a more granular analysis of species diversity and the detection of the introduction of new genotypes into the environment (Zhao et al., **2021**). The use of haplotype network tools such as POPART (Leigh & Bryant, **2015**), together with the output of the analytical framework presented here, could allow to quickly revisit introduction events of any taxa of interest.

4.3 *Ulva* spp., a particularly problematic genus compared to *Fucus* and *Porphyra*/*Pyropia* genera

Altogether, the potential for misidentifications in public repositories should not be overlooked, and in case of *Ulva* is significant. Comparing the results obtained from *Ulva* with those from *Fucus* and *Porphyra*/*Pyropia* demonstrated that *Ulva* is a particularly problematic genus (Figure **8**). In the case of *Fucus* spp., we only found a single clade that seems problematic, with apparent misannotations between *F. spiralis* and *F. vesiculosus*. With the *Porphyra*/*Pyropia* data set, which contains some 62 GMYC clades, one clade contained a mixture of *Porphyra linearis* Greville and *Porphyra umbilicalis* Kützing. Given that this clade is the only one containing either species' names, it is likely that those two species are synonymous. One species, *Neoporphyra haitanensis* (T.J. Chang & B.F. Zheng) J. Brodie & L.-E. Yang, whose genome has been released (Cao et al., **2020**), appears to be frequently misannotated, given that sequences with this species name are present in multiple clades containing other species names.

The striking consistency in the Bangiales data set over the *Ulva* one (Figure **8**) is probably due to the efforts of the Bangiales scientific community, that have collaboratively reassessed the Bangiales taxonomy and nomenclature over the last 20 years (Sutherland et al., **2011**; Yang et al., **2020**). Perhaps the ubiquitous distribution of *Ulva*, its phenotypical plasticity, and the slow

release of holotype/lectotype specimen sequences, contribute to the considerable discrepancies in *Ulva* taxonomy. We believe that a similar approach to that of the Bangiales order is needed to appropriately revisit *Ulva* nomenclature and taxonomy, and the analytical framework described here could be used as the first step towards that goal.

5 CONCLUSIONS

Due to the increasingly large number of sequences being deposited in public repositories, it is becoming important regularly to reassess the genetic information of taxa of interest, to highlight ongoing species identification issues, update NCBI accessions with new nomenclatures, and potentially reassign names to previously uncharacterised synonymous species. Here, we investigated all *Ulva*, *Fucus* and *Porphyra*/*Pyropia* sequences in the NCBI public repository for common barcodes, as a contribution to clarify the species composition and annotation of these four genera. This data set can be used for future species identification, accession validation and classification purposes, to ensure accurate representation of the species names and taxa within the databases. The analytical framework described here in detail could be transferred to any other taxa of interest, particularly those that show subtle morphological differences between taxa and contain large amount of sequences and suspected misannotations.

ACKNOWLEDGEMENTS

The authors would like to thank Ricardo Bermejo (NUI Galway), Lars Brunner and Sarah Reed (Scottish Association for Marine Science), Dan Smale and Cat Wilding (Marine Biological Organisation), Tim van Berkel and Caroline Warwick-Evans (The Cornish Seaweed Company) and Wave Crookes (SeaGrown), Helena Abreu (Alga +), for providing some of the strains used in this study. This work was funded by the European Union Horizon 2020 programme (project ID 727892, GenialG - GENetic diversity exploitation for Innovative Macro-ALGal biorefinery, **http://genialgproject.eu/**), SFI Frontiers for the Future (Project Pristine Coasts, grant number 19/FFP/6841) and the European Union Northern Periphery and Arctic Programme (project number 366, SW-GROW - Innovations for Seaweed Producers in the Northern Periphery Area project; **http://sw-grow.interreg-npa.eu/**). The authors would like to thank the reviewers and MER editors for their constructive criticism of the manuscript and helpful comments. Open access funding provided by IReL.

AUTHOR CONTRIBUTION

Antoine Fort and Ronan Sulpice designed the experiments; all authors provided biological material; Antoine Fort performed the experiments; Antoine Fort, Marcus McHale, Michael D Guiry and Ronan Sulpice analysed the results; Kevin Cascella and Philippe Potin provided

administrative and technical support; Antoine Fort, Michael D Guiry, Marcus McHale and Ronan Sulpice wrote the manuscript. All authors reviewed the manuscript.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the NCBI at **https://www.ncbi.nlm.nih.gov/**, reference numbers **MT894471**- **MT895108**. Scripts and pipeline are available in GitHub: **https://github.com/FortAnt/BarcodeAnalysis**.

Supporting Information

References

- Barbier, M., Charrier, B., Araujo, R., Holdt, S., Jacquemin, B., & Rebours, C. (2019). In M. Barbier & B. Charrier (Eds.), *PEGASUS-PHYCOMORPH European guidelines for a sustainable aquaculture of seaweeds*. COST action FA1406.
- Cao, M., Xu, K., Yu, X., Bi, G., Liu, Y., Kong, F., & Mao, Y. (2020). A chromosome-level genome assembly of Pyropia haitanensis (Bangiales, Rhodophyta). *Molecular Ecology Resources*, 20(1), 216– 227.
- Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972– 1973.
- Chowdhary, A., Singh, A., Singh, P. K., Khurana, A., & Meis, J. F. (2019). Perspectives on misidentification of Trichophyton interdigitale/Trichophyton mentagrophytes using internal transcribed spacer region sequencing: Urgent need to update the sequence database. *Mycoses*, 62(1), 11– 15.
- Coat, G., Dion, P., Noailles, M.-C., De Reviers, B., Fontaine, J.-M., Berger-Perrot, Y., & Loiseaux-De Goër, S. (1998). *Ulva armoricana* (Ulvales, Chlorophyta) from the coasts of Brittany (France). II. Nuclear rDNA ITS sequence analysis. *European Journal of Phycology*, 33(1), 81– 86.
- Cui, J., Monotilla, A. P., Zhu, W., Takano, Y., Shimada, S., Ichihara, K., & Hiraoka, M. (2018). Taxonomic reassessment of Ulva prolifera (Ulvophyceae, Chlorophyta) based on specimens from the type locality and Yellow Sea green tides. *Phycologia*, 57(6), 692– 704.

- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772.
- de Queiroz, K. (2006). The PhyloCode and the distinction between taxonomy and nomenclature. *Systematic Biology*, 55(1), 160– 162.
- Dion, P., De Reviers, B., & Coat, G. (1998). Ulva armoricana sp. nov. (Ulvales, Chlorophyta) from the coasts of Brittany (France). I. Morphological identification. *European Journal of Phycology*, 33(1), 73– 80.
- Du, G., Wu, F., Mao, Y., Guo, S., Xue, H., & Bi, G. (2014). DNA barcoding assessment of green macroalgae in coastal zone around Qingdao, China. *Journal of Ocean University of China*, 13(1), 97– 103.
- Dugon, M. M., Black, A., & Arthur, W. (2012). Variation and specialisation of the forcipular apparatus of centipedes (Arthropoda: Chilopoda): A comparative morphometric and microscopic investigation of an evolutionary novelty. *Arthropod Structure & Development*, 41(3), 231– 243.
- Dunbar, J. P., Vitkauskaite, A., O'Keeffe, D. T., Fort, A., Sulpice, R., & Dugon, M. M. (2021). Bites by the noble false widow spider *Steatoda nobilis* can induce Latrodectus-like symptoms and vector-borne bacterial infections with implications for public health: A case series. *Clinical Toxicology*, 1– 12.
- Estoup, A., & Guillemaud, T. (2010). Reconstructing routes of invasion using genetic data: Why, how and so what? *Molecular Ecology*, 19(19), 4113– 4130.
- Fort, A., Guiry, M. D., & Sulpice, R. (2018). Magnetic beads, a particularly effective novel method for extraction of NGS-ready DNA from macroalgae. *Algal Research*, 32, 308– 313.
- Fort, A., Lebrault, M., Allaire, M., Esteves-Ferreira, A. A., McHale, M., Lopez, F., & Sulpice, R. (2019). Extensive variations in diurnal growth patterns and metabolism among *Ulva spp*. strains. *Plant Physiology*, 180(1), 109– 123.
- Fort, A., Linderhof, C., Coca-Tagarro, I., Inaba, M., McHale, M., Cascella, K., Potin, P. & Sulpice, R. (2021). A sequencing-free assay for foliose Ulva species identification, hybrid detection and bulk biomass characterisation. *Algal Research-Biomass Biofuels and Bioproducts*, 55, 102280.
- Fort, A., Mannion, C., Fariñas-Franco, J. M., & Sulpice, R. (2020). Green tides select for fast expanding *Ulva* strains. *Science of the Total Environment*, 698, 134337.
- Fort, A., McHale, M., Cascella, K., Potin, P., Usadel, B., Guiry, M. D., & Sulpice, R. (2021). Foliose Ulva species show considerable inter-specific genetic diversity, low

intra-specific genetic variation, and the rare occurrence of inter-specific hybrids in the wild. *Journal of Phycology*, 57(1), 219– 233.

- Fujisawa, T., & Barraclough, T. G. (2013). Delimiting species using single-locus data and the generalized mixed Yule Coalescent approach: A revised method and evaluation on simulated data sets. *Systematic Biology*, 62(5), 707– 724.
- Garg, A., Leipe, D., & Uetz, P. (2019). The disconnect between DNA and species names: Lessons from reptile species in the NCBI taxonomy database. *Zootaxa*, 4706(3), 401– 407.
- Guiry, M., & Guiry, G. (2021). *AlgaeBase*. World-wide electronic publication, National University of Ireland. Retrieved from http://www.algaebase.org
- Hayden, H. S., Blomster, J., Maggs, C. A., Silva, P. C., Stanhope, M. J., & Waaland, J. R. (2003). Linnaeus was right all along: Ulva and Enteromorpha are not distinct genera. *European journal of phycology*, 38(3), 277– 294. https://doi.org/10.1080/1364253031000136321
- Hayden, H. S., & Waaland, J. R. (2002). Phylogenetic systematics of the Ulvaceae (Ulvales, Ulvophyceae) using chloroplast and nuclear DNA sequences. *Journal of Phycology*, 38(6), 1200– 1212.
- Heesch, S., Broom, J. E. S., Neill, K. F., Farr, T. J., Dalen, J. L., & Nelson, W. A. (2009). *Ulva*, *Umbraulva* and *Gemina*: Genetic survey of New Zealand taxa reveals diversity and introduced species. *European Journal of Phycology*, 44(2), 143– 154.
- Hiraoka, M., Shimada, S., Uenosono, M., & Masuda, M. (2004). A new green-tide-forming alga, Ulva ohnoi Hiraoka et Shimada sp. nov. (Ulvales, Ulvophyceae) from Japan. *Phycological Research*, 52(1), 17– 29.
- Hofmann, L. C., Nettleton, J. C., Neefus, C. D., & Mathieson, A. C. (2010). Cryptic diversity of Ulva (Ulvales, Chlorophyta) in the Great Bay Estuarine System (Atlantic USA): Introduced and indigenous distromatic species. *European Journal of Phycology*, 45(3), 230– 239.
- Horimoto, R., Masakiyo, Y., & Ichihara, K. (2011). Enteromorpha-like Ulva (Ulvophyceae, Chlorophyta) growing in the Todoroki River, Ishigaki Island, Japan, with special reference to Ulva meridionalis Horimoto et Shimada, sp. nov. *Bulletin of the National Museum of Nature and Science. Series B, Botany*, 37, 155– 167.
- Hughey, J. R., Gabrielson, P. W., Maggs, C. A., & Mineur, F. (2021). Genomic analysis of the lectotype specimens of European Ulva rigida and Ulva lacinulata (Ulvaceae,

Chlorophyta) reveals the ongoing misapplication of names. *European Journal of Phycology*, 1– 11.

Hughey, J. R., Gabrielson, P. W., Maggs, C. A., Mineur, F., & Miller, K. A. (2021). Taxonomic revisions based on genetic analysis of type specimens of Ulva conglobata, U. laetevirens, U. pertusa and U. spathulata (Ulvales, Chlorophyta). *Phycological Research*, 69(2), 148– 153.

- Hughey, J. R., Miller, K. A., & Gabrielson, P. W. (2018). Mitogenome analysis of a green tide forming Ulva from California, USA confirms its identity as Ulva expansa (Ulvaceae, Chlorophyta). *Mitochondrial DNA Part B*, 3(2), 1302– 1303. https://doi.org/10.1080/23802359.2018.1535859
- Hughey, J. R., Maggs, C. A., Mineur, F., Jarvis, C., Miller, K. A., Shabaka, S. H., & Gabrielson, P. W. (2019). Genetic analysis of the Linnaean *Ulva lactuca* (Ulvales, Chlorophyta) holotype and related type specimens reveals name misapplications, unexpected origins, and new synonymies. *Journal of Phycology*, 55(3), 503– 508.
- Johannesson, K., & Andre, C. (2006). Life on the margin: Genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology*, 15(8), 2013– 2029.
- Kang, E. J., Kim, J.-H., Kim, K., Choi, H.-G., & Kim, K. Y. (2014). Re-evaluation of green tide-forming species in the Yellow Sea. *Algae*, 29(4), 267– 277.
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160– 1166.
- Kirkendale, L., Saunders, G. W., & Winberg, P. (2013). A molecular survey of *Ulva* (Chlorophyta) in temperate Australia reveals enhanced levels of cosmopolitanism. *Journal of Phycology*, 49(1), 69– 81.
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453– 4455.
- Kraft, L. G., Kraft, G. T., & Waller, R. F. (2010). Investigations into southern Australian *Ulva* (Ulvophyceae, Chlorophyta) taxonomy and molecular phylogeny indicate both cosmopolitanism and endemic cryptic species. *Journal of Phycology*, 46(6), 1257– 1277.
- Krupnik, N., Paz, G., Douek, J., Lewinsohn, E., Israel, A., Carmel, N., Mineur, F., & Maggs, C. A. (2018). Native, invasive and cryptogenic *Ulva* species from the Israeli

Mediterranean Sea: Risk and potential. *Mediterranean Marine Science*, 19(1), 132– 146.

- Leal, M. C., Hilário, A., Munro, M. H., Blunt, J. W., & Calado, R. (2016). Natural products discovery needs improved taxonomic and geographic information. *Natural Product Reports*, 33(6), 747– 750.
- Leigh, J. W., & Bryant, D. (2015). Popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110– 1116.
- Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J. M., Zuccarello, G. C., & De Clerck, O. (2014). DNA-based species delimitation in algae. *European Journal of Phycology*, 49(2), 179– 196.
- Leliaert, F., Zhang, X., Ye, N., Malta, E.-J., Engelen, A. H., Mineur, F., Verbruggen, H., & De Clerck, O. (2009). Research note: Identity of the Qingdao algal bloom. *Phycological Research*, 57(2), 147– 151.
- Lemoine, F., Domelevo Entfellner, J.-B., Wilkinson, E., Correia, D., Dávila Felipe, M., De Oliveira, T., & Gascuel, O. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, 556(7702), 452– 456.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078– 2079.

  Loughnane, C. J., McIvor, L. M., Rindi, F., Stengel, D. B., & Guiry, M. D. (2008). Morphology, *rbc*L phylogeny and distribution of distromatic *Ulva* (Ulvophyceae, Chlorophyta) in Ireland and southern Britain. *Phycologia*, 47(4), 416– 429.
- Malta, E.-J., Draisma, S., & Kamermans, P. (1999). Free-floating Ulva in the southwest Netherlands: Species or morphotypes? A morphological, molecular and ecological comparison. *European Journal of Phycology*, 34(5), 443– 454.
- Melton, J. T., Collado-Vides, L., & Lopez-Bautista, J. M. (2016). Molecular identification and nutrient analysis of the green tide species Ulva ohnoi M. Hiraoka & S. Shimada (2004) (Ulvophyceae, Chlorophyta), a new report and likely nonnative species in the Gulf of Mexico and Atlantic Florida, USA. *Aquatic Invasions*, 11(3), 225– 237.
- Mendez, M., Rosenbaum, H. C., Subramaniam, A., Yackulic, C., & Bordino, P. (2010). Isolation by environmental distance in mobile marine species: molecular

ecology of franciscana dolphins at their southern range. *Molecular Ecology*, 19(11), 2212– 2228.

- Miladi, R., Manghisi, A., Minicante, S. A., Genovese, G., Abdelkafi, S., & Morabito, M. (2018). *A DNA barcoding survey of Ulva (Chlorophyta) in Tunisia and Italy reveals the presence of the overlooked alien U. ohnoi. Cryptogamie, Algologie*.
- Monotilla, A. P., Nishimura, T., Adachi, M., Tanii, Y., Largo, D. B., & Hiraoka, M. (2018). Examination of prezygotic and postzygotic isolating barriers in tropical Ulva (Ulvophyceae, Chlorophyta): Evidence for ongoing speciation. *Journal of Phycology*, 54(4), 539– 549.
- Nasehi, A., Al-Sadi, A. M., Nasr Esfahani, M., Ostovar, T., Rezaie, M., Atghia, O., Alsultan, W., & Javan-Nikkhah, M. (2019). Molecular re-identification of *Stemphylium lycopersici* and *Stemphylium solani* isolates deposited in NCBI GenBank and morphological characteristics of Malaysian isolates. *European Journal of Plant Pathology*, 153(3), 965– 974.
- O'Kelly, C. J., Kurihara, A., Shipley, T. C., & Sherwood, A. R. (2010). Molecular assessment of *Ulva spp*. (Ulvophyceae, Chlorophyta) in the Hawaiian Islands. *Journal of Phycology*, 46(4), 728– 735.
- Packer, L., Gibbs, J., Sheffield, C., & Hanner, R. (2009). DNA barcoding and the mediocrity of morphology. *Molecular Ecology Resources*, 9, 42– 50.
- Pattengale, N. D., Alipour, M., Bininda-Emonds, O. R. P., Moret, B. M. E., & Stamatakis, A. (2010). How many bootstrap replicates are necessary? *Journal of Computational Biology*, 17(3), 337– 354.
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., Kamoun, S., Sumlin, W. D., & Vogler, A. P. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55(4), 595– 609.
- Posada, D., & Buckley, T. R. (2004). Model selection and model averaging in phylogenetics: Advantages of Akaike information criterion and Bayesian
- Puillandre, N., Lambert, A., Brouillet, S., & Achaz, G. (2012). ABGD, automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, 21(8), 1864– 1877.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Systematic Biology*, 67(5), 901– 904.

- Rautenberger, R., Fernandez, P. A., Strittmatter, M., Heesch, S., Cornwall, C. E., Hurd, C. L., & Roleda, M. Y. (2015). Saturating light and not increased carbon dioxide under ocean acidification drives photosynthesis and growth in Ulva rigida (Chlorophyta). *Ecology and Evolution*, 5(4), 874– 888.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, 61(3), 539– 542.
- Saunders, G. W., & Kucera, H. (2010). An evaluation of *rbc*L, *tuf*A, UPA, LSU and ITS as DNA barcode markers for the marine green macroalgae. *Cryptogamie, Algologie*, 31(4), 487– 528.
- Sauriau, P.-G., Dartois, M., Becquet, V., Aubert, F., Huet, V., Bréret, M., & Pante, E. (2021). Multiple genetic marker analysis challenges the introduction history of Ulva australis (Ulvales, Chlorophyta) on French coasts. *European Journal of Phycology*, 1– 13.
- Shimada, S., Hiraoka, M., Nabata, S., Iima, M., & Masuda, M. (2003). Molecular phylogenetic analyses of the Japanese Ulva and Enteromorpha (Ulvales, Ulvophyceae), with special reference to the free-floating Ulva. *Phycological Research*, 51, 99– 108. https://doi.org/10.1046/j.1440-1835.2003.00296.x
- South, A. (2011). Rworldmap: A new R package for mapping global data. *R Journal*, 3(1), 35.
- Spalding, H. L., Conklin, K. Y., Smith, C. M., O'Kelly, C. J., & Sherwood, A. R. (2016). New Ulvaceae (Ulvophyceae, Chlorophyta) from mesophotic ecosystems across the Hawaiian archipelago. *Journal of Phycology*, 52(1), 40– 53. https://doi.org/10.1111/jpy.12375
- Steinhagen, S., Karez, R., & Weinberger, F. (2019). Cryptic, alien and lost species: Molecular diversity of *Ulva sensu lato* along the German coasts of the North and Baltic Seas. *European Journal of Phycology*, 54(3), 466– 483.
- Suchard, M. A., & Rambaut, A. (2009). Many-core algorithms for statistical phylogenetics. *Bioinformatics*, 25(11), 1370– 1376.
- Sutherland, J. E., Lindstrom, S. C., Nelson, W. A., Brodie, J., Lynch, M. D. J., Hwang, M. S., Choi, H.-G., Miyata, M., Kikuchi, N., Oliveira, M. C., Farr, T., Neefus, C., Mols-Mortensen, A., Milstein, D., & Müller, K. M. (2011). A new look at an ancient order: generic revision of the Bangiales (Rhodophyta) 1. *Journal of Phycology*, 47(5), 1131– 1151.

- Tan, I. H., Blomster, J., Hansen, G., Leskinen, E., Maggs, C. A., Mann, D. G., Sluiman, H. J., & Stanhope, M. J. (1999). Molecular phylogenetic evidence for a reversible morphogenetic switch controlling the gross morphology of two common genera of green seaweeds, *Ulva* and *Enteromorpha*. *Molecular Biology and Evolution*, 16(8), 1011– 1018.
- Tang, C. Q., Humphreys, A. M., Fontaneto, D., & Barraclough, T. G. (2014). Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution*, 5(10), 1086– 1094.
- Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution*, 24(2), 110– 117.
- Vranken, S., Bosch, S., Peña, V., Leliaert, F., Mineur, F., & De Clerck, O. (2018). A risk assessment of aquarium trade introductions of seaweed in European waters. *Biological Invasions*, 20(5), 1171– 1187.
- Wichard, T., Charrier, B., Mineur, F., Bothwell, J. H., Clerck, O. D., & Coates, J. C. (2015). The green seaweed *Ulva*: A model system to study morphogenesis. *Frontiers Plant Science*, 6(72).
- Yang, L.-E., Deng, Y.-Y., Xu, G.-P., Russell, S., Lu, Q.-Q., & Brodie, J. (2020). Redefining Pyropia (Bangiales, Rhodophyta): Four New Genera, Resurrection of Porphyrella and Description of Calidia pseudolobata sp. nov. From China. *Journal of Phycology*, 56(4), 862– 879.
- Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29(22), 2869– 2876.
- Zhao, W., Dong, L., Hong, D. D., Brodie, J., Chen, W.-Z., Tien, D. D., Zhou, W., Lu, Q.-Q., Zhang, M.-R., & Yang, L.-E. (2021). Haplotype networks of Phycocalidia tanegashimensis (Bangiales, Rhodophyta) indicate a probable invasion from the South China Sea to Brazil. *Marine Biodiversity*, 51(2), 1– 11.