**Archimer**
https://archimer.ifremer.fr

# From seascape ecology to population genomics and back. Spatial and ecological differentiation among cryptic species of the red algae *Lithophyllum stictiforme/L. cabiochiae,* main bioconstructors of coralligenous habitats

De Jode Aurélien [1, *], David Romain [1], Haguenauer Anne [1, 2], Cahill Abigail E. [1, 3], Erga Zinovia [1], Guillemain Dorian [1, 4], Sartoretto Stephane [5], Rocher Caroline [1], Selva Marjorie [1], Le Gall Line [6], Féral Jean-Pierre [1], Chenuil Anne [1]

[1] Aix Marseille Univ, Univ Avignon, CNRS, IRD, IMBE, Marseille, France
[2] EPHE PSL Research University, USR 3278 CRIOBE CNRS-UPVD, BP 1013, Moorea, 98729, French Polynesia
[3] Biology Department, Albion College, Michigan 49224, USA
[4] Institut Pythéas: Observatoire des Sciences de l'Univers
[5] IFREMER, Zone Portuaire de Brégaillon, 83500 La Seyne-sur-mer, France
[6] Institut Systématique Evolution Biodiversité (ISYEB), Muséum national d'Histoire naturelle, CNRS, Sorbonne Université, EPHE, 57 rue Cuvier, CP 39, 75005 Paris, France

* Corresponding author : Aurélien De Jode, email address : aurelien.dejode@imbe.fr

**Abstract :**

Ecosystem engineering species alter the physical structure of their environment and can create or modify habitats, having a massive impact on local biodiversity. Coralligenous reefs are highly diverse habitats endemic to the Mediterranean Sea built by calcareous benthic organisms among which Crustose Coralline Algae are the main engineering species. We analyzed the diversity of Lithophyllum stictiforme or L. cabiochiae in coralligenous habitats combining a multiple barcode and a population genomics approach with seascape features. Population genomics allowed disentangling pure spatial effects from environmental effects. We found that these taxa form a complex of eight highly divergent cryptic species that are easily identifiable using classic barcode markers (psbA, LSU, COI). Three factors have a significant effect on the relative abundances of these cryptic species: the location along the French Mediterranean coast, depth and Photosynthetic Active Radiation (PAR). The analysis of around 5000 SNPs for the most abundant species revealed genetic differentiation among localities in the Bay of Marseille but no differentiation between depths within locality. Thus, the effect of depth and PAR on cryptic species communities is not a consequence of restricted connectivity but rather due to differential settlement or survival among cryptic species. This differential is more likely driven by irradiance levels rather than by pressure or temperature. Both the genetic and species diversity patterns are congruent with the main patterns of currents in the Bay. Ecological differentiation among these engineering cryptic species, sensitive to ocean warming and acidification, could have important consequences on the diversity and structure of the coralligenous communities.

**Graphical abstract**



**Highlights**

► We found 8 cryptic species in Lithophyllum stictiforme / L. cabiochiae. ► Cryptic species relative abundances vary along the French Mediterranean coast. ► Cryptic species relative abundances vary among depth on the same locality. ► No genetic differentiation among depths within locality in the most abundant species. ► Cryptic species composition is influenced by locality and depth.

**Keywords** : engineering species, cryptic species, ecological niche, coralligenous habitats, ecological differentiation, Crustose Coralline Algae

**Introduction**

Ecosystem engineers are organisms that alter their abiotic environment in such a way that they create or modify habitats, thereby having large effects on the associated species community (Jones *et al.* 1994; Crain & Bertness 2006). Thus, phenotypic variation among ecosystem engineering organisms potentially have important consequences on the species community and on the ecosystem (Whitham *et al.* 2003). The phenotypic variation can arise at the intra specific level by plasticity or genetic differentiation, as well as interspecifically when different engineering species have different ecological traits (Badano & Cavieres 2006; Lamit *et al.* 2011). In the marine realm, animal organisms acting as ecosystem engineers promote biodiversity of the associated communities (Romero *et al.* 2015). Algal engineer species also have tremendous impacts on marine biodiversity. Within seascapes, kelp forests are the most conspicuous three-dimensional habitats hosting a high diversity of species (Teagle *et al.* 2017). Crustose Coralline Algae (hereafter CCA) are also major engineering organisms and contribute to the three-dimensional structure of several habitats such as coral reefs, maërl beds and coralligenous habitats.

In the Mediterranean Sea, coralligenous habitats are emblematic calcareous biogenic constructions built-up in dim light conditions mainly by calcareous algae (Corallinacea and Peyssonneliacea) and reinforced by calcareous invertebrates (e.g. bryozoans, serpulid polychaetes, scleractinians) (Ballesteros 2006). The resulting framework is complex and harbors various micro-habitats that shelter at least 1600 species (Ballesteros 2006), making coralligenous habitats an important biodiversity hot-spot in the Mediterranean Sea (Boudouresque 2004). These habitats provide various ecosystem services (e.g. food provisioning, recreational diving, research material) (Paoli *et al.* 2016; Thierry de Ville

d'Avray *et al.*, 2019), yet they are threatened by global ocean warming and acidification (Martin & Gattuso 2009; Lombardi *et al.* 2011; Martin *et al.* 2013; Linares *et al.* 2015; Rodríguez-Prieto 2016) and local human activities (e.g. fishing, anchoring or sewage outfalls) (Hong 1980; Ballesteros 2006; Balata *et al.* 2005; Balata *et al.* 2007). The lack of knowledge regarding the biodiversity of these habitats impedes our understanding of their ecological functioning and our capacity to protect them efficiently (SPA/RAC 2017).

In the coralligenous habitats, CCA are considered to be the major engineering group (Laborel 1961; Laubier 1966; Sartoretto *et al.* 1996); however, the phylogenetic affinities of these CCA among other Corallinales has not been tested yet with molecular systematic tools even though these tools have strongly modified the perception of coralline diversity (Bittner *et al.* 2011, Pardo *et al.* 2014, Peña *et al.* 2015, Rösler *et al.* 2017). In the genus *Lithophyllum*, *L. stictiforme* (Areschoug) Hauck (1877) *and L. cabiochiae* (Boudouresque & Verlaque) Athanasiadis (1999) are considered the main coralligenous builders below 20 m depth (Sartoretto *et al.* 1996). However, identification of these two nominal species based on macro-morphological characteristics or anatomical structures is uncertain. Moreover, recent studies using molecular systematic tools (Rindi *et al.* 2017, Pezzolesi et al. 2019) have revealed the presence of cryptic diversity but did nit support distinction between these two species. Ignoring the presence of cryptic species within a nominal species may have important consequences for biodiversity management. In particular, when cryptic species are ecologically differentiated, environmental changes may result in higher risks of extinction (local or global) than expected for a single generalist species (Chenuil *et al.*, in press). Furthermore, recent studies showed that "*L. stictiforme*" survival and reproduction were affected by irradiance levels and temperature (Rodríguez-Prieto 2016) and that "*L. cabiochiae*" photosynthesis was reduced under elevated $pCO_2$ (Martin *et al.* 2013). This

highlights the need for studies of biodiversity at both inter- and intraspecific levels to evaluate the potential of adaptation to global change of these ecosystem engineering species.

In this study, we combined barcoding and exon capture sequencing to reveal cryptic species among 438 individuals initially collected as *L. cabiochiae/stictiforme* found in sympatry along the French Mediterranean coast, providing the opportunity to study the ecological determinants of their co-distribution. We used a fine scale, ecologically contextualized design in order to distinguish spatial effects (resulting from migration and connectivity) from ecological effects (fitness differences among cryptic species in distinct environments). Light is the most important environmental factor shaping coralligenous communities (Ballesteros 2006), thus we recorded environmental variables affecting the irradiance levels received by the community on each site. We analyzed community composition within the species complex (i.e. the relative abundances of the distinct cryptic species) in relation to ecological conditions (location, depth, orientation, slope) to determine if the different species were found in different niches. Population genetic analyses were then carried out in the most abundant of the cryptic species to determine the connectivity matrix among studied locations in order to disentangle pure spatial effects from environmental effects of depth.

# Materials and methods

## 2.1 Field sampling and DNA extraction

Samples were collected by scuba diving in 13 different localities along the continental French Mediterranean coast and in Corsica (Figure 2 & 3, Supplementary Material I). In the Marseille and Toulon area, horizontal transect lines were divided into 5 m segments. For each segment, up to 4 fragments of *Lithophyllum* spp. of about 3 cm large were collected. To avoid sampling clones of the same specimen, we left a minimum of 1m between collected specimens. An average of 16 samples was collected at each sampling site. Moreover, 3 physical parameters were recorded in situ: slope, orientation, and rugosity. Slope was divided in 4 levels: (i) Flat, when the angle formed by the substrate with the horizontal line was between 0° and 18.45°, (ii) Inclined, for angles between 18.45° and 71.69°, (iii) vertical, for angles between 71.69° and 90° and (iv) overhanging, when there is an overhang at least as large as a person above the observer and covers most of the segment. Orientation was measured with a compass handled by the diver and directed perpendicularly to the substrate wall (in the horizontal plane; it could not be defined for horizontal slopes). We considered 8 modalities: North (N), South (S), East (E), West (W) and the four intermediate orientations (Northeast (NE), Northwest (NW), Southeast (SE), Southwest (SW)). Note that a North orientation in our dataset corresponds to a South exposition of the substrate. Rugosity characterizes the size of crevices, holes and faults observed in the segment and was characterized as follows: (i) "Tiny", when holes were smaller than a fist (about 10 cm wide); (ii) "Small", when holes were larger than a fist but smaller than a head; (iii) "Medium" (M): holes and crevices that were approximately head-sized (about 30 cm wide); (iv) "Large" (L):

crevices, holes and faults can contain at least the upper body (about 1 m wide with air tanks).

Most transects were carried out at a depth of 24-31 m (depth category D1), or 37-46 m (depth

category D2). In Banyuls-sur-mer, Villefranche-sur-mer and Corsica transects were not

segmented and all individuals were sampled at the same depth category. All CCA were dried,

and preserved in silica gel at room temperature in a dark place or in ethanol 96% at 4°C until

DNA extraction. A piece of algal tissue was excised and cleaned of epiphytes by scraping the

surface with a razor blade. The excised sample was disrupted using TissueLyser II system

(Qiagen) with a 3 mm stainless steel bead. DNA was extracted using Chelex 100 chelating

resin (Walsh *et al.* 1991). Around 20 mg of tissue along with 500 µl of Chelex 10% and 3µl of

Proteinase K (20 mg/mL) were incubated at 60 °C for 90 min. Then sample was heated at 100

°C for about 10 min to deactivate the Proteinase K.

### 2.2 Sanger Sequencing

Three independent molecular markers were used to identify the species of *Lithophyllum*: the

plastid marker *psb*A, (Yoon *et al.* 2002), a fragment of the nuclear 28S (or large subunit)

rDNA marker (Harper & Saunders 2001) and the mitochondrial marker COI (Saunders 2005).

PCR reaction mixes were the same for the three markers, and PCR programs were identical

for *psb*A and 28S (Supplementary Material II). PCR products were verified by electrophoresis

migration on 1.5 % agarose gel TBEx1 and then sent to Eurofins Genomics for Sanger

sequencing using primers Gaz-R1, *psb*A-F1 and T04. Fragment sizes were approximately 700

bp for COI and 1000 bp for *psb*A and 28S. Sequences were checked and aligned using

BioEdit software (Hall 1999) before further analyses.

## 2.3 Miseq Sequencing

In a second step, to determine the lineages of additional individuals with lower sequencing costs, we designed shorter fragments within the *psb*A and 28S markers to allow high throughput sequencing (Illumina, Miseq paired end 2x 250 bp) (Supplementary Material II). PCR cycles and reactions mixes were the same for both markers (Supplementary Material II). Amplicons were sequenced, and Miseq Reads were processed as described in Cahill *et al.* (2017) with a few exceptions. For the plastid marker, *psb*A, sorting was done as in Cahill *et al.* (2017) for mitochondrial loci with slight modifications: sequences were retained if the total number of reads was $\geq 20$ (i.e. 20X coverage) and the count ratio $\leq 0.14$ (i.e. the most abundant read was at least 7 times more abundant than the second most abundant read). For the 28S nuclear locus, the total number of reads was sufficient for all individuals (at least 66 reads for each). For both markers, forward and reverse reads did not overlap and were attached end to end.

## 2.4 Haplotype networks

For each marker, sequences from Sanger sequencing and Miseq sequencing were aligned visually with Bioedit (Hall 1999). Sequence positions found in both Sanger and Miseq sequencing were kept to build an alignment with all sequences of the same size. For each marker, we built haplotype networks from the longer sequence alignment obtained by Sanger sequencing and also, for *psb*A and 28S, from the shorter alignments including Miseq sequences. Haplotype networks were generated using the median-joining algorithm of the Network software, v.5.0.0.1 (Bandelt *et al.* 1999). The average proportion of differences and average Kimura distance (K2P, Kimura 1980) between haplogroups were computed from the long alignments (i.e. using MEGA v.4 (Tamura *et al.* 2007)).

## 2.5 Transcriptomics

Individuals were collected in 2015 at the CAS, MEJ, LPD, COU localities by scuba diving, and preserved in seawater during transportation to the lab. Around 2 $cm^2$ of each individual was immediately cleaned of epiphytes with a razor blade, placed in 1 ml QIAzol lysis reagent (Qiagen) and disrupted using TissueLyserII instrument (Qiagen). The rest was preserved in 96% ethanol at 4°C and used to determine the haplogroup of each individual following the protocol for Sanger sequencing described above in this paper.

Total RNA isolation was performed according to the manufacturer's instructions, except for overnight precipitation at −20 °C (1/10 volume sodium acetate 3M pH 5.2, 2 volumes ethanol). Contaminants were eliminated by further cleaning using an RNeasy plus mini kit (Qiagen). RNA integrity was assessed using Agilent 2100 *Bioanalyzer* system, and concentration and purity using Nanodrop instrument (Thermofisher). Residual DNA was digested using TurboDNAse (Ambion) following the manufacturer's instructions.

RNA-Seq libraries were generated using the TruSeq Stranded mRNA Illumina kit according to the manufacturer's protocol. During this preparation, libraries were individually tagged to allow their pooling before sequencing. The size distributions of libraries' RNA fragments were controlled with a Fragment Analyzer™ Automated CE System from Advanced Analytical Technologies, Inc. (AATI).

The eight libraries were quantified by qPCR following the manufacturer's protocol. Libraries were pooled before sequencing on one lane on the Illumina HiSeq3000 (Illumina Inc., San Diego, CA) as paired-end reads of length 150 bp. Library preparation and sequencing were performed at the Genotoul platform (http://get.genotoul.fr/).

Three samples were highly contaminated with 30 %, 10 % and 7% of their reads mapping against the *E. coli* genome. Those reads were removed before the assembly step. Assembly of

each library was performed using the RunDrap Pipeline with default parameters as described by Cabau *et al.* (2017).

To build a reference transcriptome, one meta-assembly of the two individuals of the most abundant species (the C1 species) transcriptome was conducted, using the Run Meta Pipeline described in Cabau *et al.* (2017).

Sequences from each assembly and meta-assembly were blasted using the program blastn version 2.2.26 (Altschul *et al.* 1990; Camacho *et al.* 2009) against different databases available on the Genotoul Cluster: Bacteria (ftp://ftp.ncbi.nih.gov/genomes/archive/old_refseq/Bacteria), H_sapiens (ftp://ftp.ncbi.nih.gov/genomes/H_sapiens), Drosophila_melanogaster (ftp://ftp.ncbi.nih.gov/genomes/archive/old_refseq/Drosophila_melanogaster), yeast (ftp://ftp.ncbi.nih.gov/blast/db/FASTA/ ) , M_musculus (ftp://ftp.ncbi.nih.gov/genomes/M_musculus) and contigs with a hit (e-value threshold 0.001) against any of them were removed from further analyses. These species are commonly used in genomic studies and are potential sources of contamination on the genomic platform. Also, their genomes are available allowing to compare our sequences with these potential contaminants.

The meta-assembly of the two individuals of the C1 haplogroup was used as a reference in the following steps and contig names were modified using a custom script. For all individuals, reverse reads were renamed and pooled with forward reads in one file. All reads were mapped on the reference with Bwa mem (Li 2013) with default parameters. SNP calling was conducted using the Reads2snp v2.0 script from PopPhyl project (Tsagkogeorga *et al.* 2012; Gayral *et al.* 2013). Open reading frames (ORFs) were detected using Transdecoder 3.0.0

(Haas *et al.* 2013) and the ORF output file was converted to get.ORF format using a custom script.

Biotinylated RNA probes were designed by and ordered from MYcroarray (Ann Arbor, MI, USA; now Arbor Biosciences). Based on the ORF sequences, 18757 candidate sequences (total size 18.3Mb) were soft-masked for simple repeats and low-complexity DNA using RepeatMasker (Smit *et al.* 2013). Any strings of Ns between 1-10bp were replaced with Ts to facilitate probe design; larger strings of Ns were left alone. 120 bp probes with 2x flexible tiling density were designed for all sequences, and dG, GC%, and % soft-masked were tabulated. Only probes with (i) dG greater than -9, (ii) 30-50% GC, (iii) 0 soft-masked bases were kept and divided in two subsets: the first one contained probes with 1 to 7 SNPs with the reference and the second one the probes with no SNP with the reference.

Among the first probe subset we randomly sampled one probe per ORF for a total of 14403 probes. For each randomly selected probe we extracted the closest and the furthest probe when possible. Finally, we took 2000 probes in the closest, 2000 in the furthest (from the first probe selected in the ORF) and 1617 in the non-polymorphic probes for a total of 20020 probes.

### 2.5 Exon-Capture genotyping

For capture sequencing, total genomic DNA was extracted according a protocol derived from Sambrook *et al.* (1989), followed by one or two purifications using NucleoSpin® gDNA Clean-up (Macherey-Nagel) (see Supplementary Material III for a detailed protocol).

DNA was fragmented using The Bioruptor® Pico (diagenode) to obtain fragment size of around 250 pb. Dual-indexed NGS Libraries were made using NEBNext® Ultra™ DNA Library Prep Kit for Illumina® kit following the manufacturer's protocol. Two conditions of

in-solution target enrichment were performed according to manufacturer's recommendations following the MYbaits v3 protocol (http://www.mycroarray.com/mybaits/manuals.html). The first condition was 32 libraries in 30 μL reactional volume, repeated 3 times for a total of 128 enriched libraries, whereas the second condition was conducted in 15μL reactional volume with 16 libraries. Post-capture, libraries were amplified following Mybaits protocol recommendations (Mix KAPA HiFi, PCR at 60°C, 14 cycles), post PCR purification using Ampure XP, and 1.6pM DNA was provided for sequencing on one MID flowcell of NextSeq Illumina System, Paired-end sequencing (2x150bp).

First, raw reads were split according to their sequencing lane using a custom python script and fastq files were converted to SAM files. We followed the GATK good practices (Van der Auwera *et al.* 2013) to call SNPs except for the BQSR and VQSR steps because of the lack of reference SNPs data sets. We used ORFs of the C1 haplogroup meta assembly as a reference in all the above steps and SNPs calling was restricted to an area starting 400 bp before the first base covered by the probe to 400 bp after the last base covered by the probe.

To study the inter haplogroup divergence the jointGenotyping step was conducted with all individuals from the Bay of Marseilles. To study the intra C1 haplogroup diversity the jointGenotyping step was conducted with all individuals from the C1 haplogroup. In both cases, obtained SNPs were filtered based on GATK recommended parameters (QD < 2.0, FS > 60.0, MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0), then SNPs with minor allele frequency lower than 0.01 were removed. For the inter haplogroup study, SNPs with more than 10% missing value were removed from the dataset. For the intra C1 haplogroup dataset, SNPs with more than 25% missing values or failing HWE in one or more of the 7 populations were removed using vcftools (Danecek *et al.* 2011), with a p-value threshold of 0.01. Finally, for both datasets one SNP per ORF was randomly extracted using a custom

python script and individuals with more than 20% missing genotypes were excluded from the dataset.

## 2.6 Population Genomics Analyses

First, a neighbor joining tree on multilocus genotypes was built with individuals of the inter clade dataset using the APE package (Paradis *et al.* 2004) in R (R Core team 2017). Individuals were colored according to their haplogroup determined by one of the three barcoding markers. Individuals with undetermined haplogroups were assigned to a species based on their positions in the tree, using the phytools (Revell 2012) package in R (R Core team 2017). Principal Component analyses (PCA) were conducted on individuals assigned to the C1 species using package the adegenet R package (Jombart 2008; Jombart & Ahmed 2011). Calculation of F statistics, diversity indices, and tests of genetic differentiation were carried out with the GENEPOP R package (Rousset 2008).

## 2.7 Clonality

To assess the importance of clonal reproduction in the C1 species we used the functions genet_dist and genet_dist_sim from the Rclone package (Bailleul *et al.* 2016). The genet_dist function was used to compute a matrix of pairwise number of alleles differences between our multilocus genotypes (MLG) in each population. The genet_dist_sim function allowed for simulation of a sexual reproduction event between pairs of unique MLGs (outcrossing) or pairs of MLGs (partial selfing) of all our populations and computed two matrices of pairwise genetic distances within the resulting population after 1000 sexual events (one generation each) . The first matrix was obtained by simulating sexual reproduction including outcrossing and selfing and the second was obtained by simulating only outcrossing events (no selfing). The three distributions (empirical, simulated with selfing and simulated without selfing) were

compared to assess if the distribution of empirical distances can be obtained by sexual reproduction alone. We performed the analysis on 2176 biallelic SNPs with  missing values.

## 2.8 Community analyses

All the community statistical analyses were conducted using the PRIMER software version7 (Clarke *et al.* 2014 & 2015). Due to the different sampling procedure, the Banyuls-sur-mer, Villefranche-sur-mer and Corsica localities were not used in statistical analyses linking community composition to environmental factors (PERMANOVAs and PERMANCOVAs). However, all the localities were used in the non-metric Multi-Dimensional Scaling (nMDS) and for the estimation of diversity indices. The final community matrix used for the Marseille area contained 187 rows, corresponding to each segment of marine substrate, and for each row, the number of individuals of each of the 7 species present in this area (corresponding to 7 columns). The final environmental matrix had the same row number and names as the final community matrix, and for each row, the level of each of the environmental variables (depth category, locality, orientation, slope, rugosity). Community data were standardized by the total number of individuals and square-root transformed before computing pairwise Bray-Curtis dissimilarity indices. To calculate diversity metrics for each locality, we summed the abundances of each segment (Table 1). The nMDS was produced using Bray-Curtis similarity indices on the table of abundances per locality.

Due to the lack of replication, the PERMANOVA designs only included 2 factors: locality as a random factor, one other environmental factor as a fixed factor and the interaction term between the two. The PERMANCOVAs designs included the locality factor and the depth (as a numeric variable) or the Photosynthetic Active Radiation (PAR) as covariable. The PAR was calculated for each depth by averaging the values obtained from the SOMLIT recorder in the Frioul locality between 2007 and 2017. Thus, PAR was not directly measured at each location

but, according to its depth, each sampling segment was assigned an estimated PAR value. Since the effect of the 'location' factor was considered in the analysis using the PAR covariable, the variable PAR represents the effect that depth has on irradiance attenuation.

## Results

### 3.1 Haplotype networks

Sanger (i.e. longer) sequence alignments were respectively 594 bp, 965 bp and 802 bp for the mitochondrial (COI), plastid (*psb*A) and the nuclear markers (28S rDNA), after primers and lower quality terminal regions were removed. Shorter alignments including Miseq sequences for the *psb*A and 28S markers were respectively 365 bp and 425bp. Figures in Supplementary Material IV display the networks built from subsets of individuals which had been sequenced for several markers (Fig. networks: 3 sanger, 2 Miseq). Seven haplogroups were identified in the haplotype network build from long *psb*A sequences: C1, C2, C3, C4, C5, C6 and C7. The individual compositions of these seven haplogroups perfectly matched with those of seven haplogroups formed by the other markers (with the exception of a single individual, likely a contamination or a labeling error), for both long and short alignments (273 individuals were sequenced for both the 28S rDNA and *psb*A markers). The number of substitutions separating the distinct haplogroups varied (in long alignments) from 4 to 22 for the most conserved marker which is the 28S rDNA, from 14 to 48 for *psb*A and from 40 to 77 for COI, the most rapidly evolving marker (Supplementary Material IV). The minimum and maximum Kimura distances were respectively 0.003 (28S rDNA, C6-C7) and 0.105 (COI, C1-C6) (Supplementary Material IV). Since the haplogroups were congruent among markers and

found in sympatry, their genetic isolation and their status of cryptic species was established (cf discussion) so in the following sections, they were considered as such.

## 3.2 Capture sequencing

For the inter species dataset, a total of 7068 SNPs were obtained for 122 individuals. Among these individuals, 69 were already classified in one of the 7 species using at least one of the three barcode markers. For all individuals from the Marseille area, species determination based on barcode marker or the multilocus genotype distances between individuals (in number of different alleles) gave the same results (Figure 1). Among the 53 remaining individuals, 3 clustered together forming the C8 species, and the others were assigned to one of the 7 species according to their position on the neighbor joining tree. The mean Euclidean distances based on genotypes between individuals of the different species ranged from 52.76 between C7 and C8 to 115.42 between C2 and C5 (Supplementary Material VI). The C4 species had the highest intra species mean distances: 24.10. Three clusters were distinguished on the tree: C1, C2 to C4, and all the C5 to C8 species (Figure 1).

## 3.3 Population genomics for the C1 species

For the individuals of the C1 species, a total of 4744 SNPs were obtained for 75 individuals. The expected heterozygosity by population ranged from 0.1090 in FTF_D1 to 0.1603 in RMO. Significant $F_{IS}$ values were found in 4 populations: FTF_D1, FTF_D2, CAS_D1, CAS_D2 ranging from -0.0096 in CAS_D2 to 0.1260 FTF_D1. The global $F_{ST}$ value was 0.0464 and $F_{ST}$ values between pairs of populations ranged from 0.0077 between CAS_D1 and CAS_D2 to 0.0911 between RMO and LPD. The genetic differentiation for the two pairs

displaying contrasted depths was not significant. On the PCA plot (Figure 4), the individuals were clustered according to their localities (but the PCA did not suggest clonality because all individuals were separated on at least one combination of axes). Individuals were spread in three different clusters from left to right on the first axis (5.71 % of the total variability). The second axis represented 3.06% of the total variability and no clear cluster was formed along this axis. The third axis represented 2.67% of variability and separated LPD population from all the others (Figure 4 B). In clonality analyses, all multilocus genotypes (MLG) were clearly distinct and empirical genetic distances distributions match with simulated genetic distance distributions obtained under the hypothesis of outcrossing in all populations (Supplementary Material VIII).

### 3.4 Cryptic species diversity and distribution

The species richness among localities ranged from 2 in the COU, CTF and PLN localities to 5 in the CIR and CAS localities. Simpson diversity indices in localities ranged from 0.1498 at the FTF locality to 0.781 at the VPR locality (Table 1). *Lithophyllum* cryptic species communities were differentiated between localities at different spatial scales (Figure 2, 3 & Table 2 & 3). At large spatial scale, the BPT locality was very distant from all other localities (Figure 2B) and was the only one where species C1 was absent and where species C5 was present. Then, the VPR locality was isolated from localities of the Marseille area on the nMDS (Figure 2B) and harbored a high proportion of the C6 and C7 cryptic species as well as the highest species diversity (Table 1). The Corsica community was grouped with communities of the Marseille area (Figure 2B). At small spatial scale (from the COU locality to the LPD locality), species were not randomly distributed across localities (2 factors PERMANOVA, p(perm)=0.0001, Table 2 & Figure 3): the C1 species was ubiquitous and the

C2 species was missing in a single locality (COU). Species C5, C7 and C8 were only found in the CAS locality and the C6 species only in the RMO locality. The C4 species was mainly found at the Côte Bleue (COU, MEJ, CSC) with few exceptions (Figure 3A). Based on the species relative abundances and the nMDS four groups of localities were discernible (Figure 3): (i) the Côte Bleue (COU, CSC, MEJ) and LPD, (ii) Marseille (FTF, RMO), (iii) the PSO and CTF localities and (iv) the CAS locality was highly distant from all the others.

### 3.5 Cryptic species ecology

In the two-way PERMANOVAs, the random locality factor was always significant, but the fixed environmental factor was never significant (Table 2). However, the interaction between locality and sampling depth category had a significant effect on the *Lithophyllum* cryptic species community composition (PERMANOVA, p(perm)=0.014). The community differed between the two depth categories in the RMO locality (PERMANOVA, p(mc) =0.0439) and the CAS locality (PERMANOVA, p(perm)=0.0067). In the RMO locality the C2 and C6 species were found at depth 28 meters and the most abundant C1 species was found at all depths. In the CAS locality, the C5 species was only found in the D1 depth category, the C8 species was only found in the D2 depth category (Figure 3A). In both PERMANCOVAs, the random locality factor and the numeric covariable (i.e. depth, in meters, and PAR) were significant (Table 3).

**Discussion**

## 4.1 The *L. stictiforme/cabiochiae* species complex encompasses at least eight cryptic species identifiable by barcoding in the French Mediterranean coast.

Mitochondrial, plastid and nuclear markers, which belong to different genomes, gave consistent results to reveal distinct haplogroups. In case of asexual reproduction mitochondrial, plastid and nuclear genomes are linked (Dudgeon *et al.* 2017) thus giving the same information. Inside the C1 species, clonality analyses revealed the absence of duplicated multilocus genotypes. Moreover, the distributions of genetic distances among multilocus genotypes match the distributions obtained with only sexual reproduction (Supplementary Material VIII) suggesting that there are no clones in our dataset. Therefore the consistency of the results obtained with the mitochondrial, plastid and nuclear markers is not likely due to presence of clonal lineage. Those haplogroups were highly divergent for the markers COI and 28S (above 10 % of divergence for COI) a level of variation which is generally recognized as interspecific variation for red algae (e.g. Saunders 2005; Le Gall & Saunders 2007) and more specifically for coralline algae (e.g. Pardo et al. 2017).In addition, the analyses based on thousands of SNPs confirmed that these seven haplogroups are genetically isolated and highly differentiated even when found in sympatry and even within a single 5 m segment. There is thus no doubt that these haplogroups are reproductively isolated. One may argue that strong inbreeding may create such a pattern (when outcrossing events are rare), and positive $F_{IS}$ values have been reported in other red algae such as *C. crispus* (Kruger-Hadfield *et al.* 2011 & 2015) and may result from low dispersal capacities leading to inbreeding. Nevertheless, the relatively moderate intra populations $F_{IS}$ values in C1 (Table 4) rule out this hypothesis. We thus established that these haplogroups are separate biological species, i.e. cryptic species of the *Lithophyllum stictiforme/cabiochiae* complex.

*Lithophyllum* is the most species diverse genus among the Corallinales with 130 species currently recognized (Guiry & Guiry 2018); however, the use of molecular systematics to clarify relationships among species of the genus *Lithophyllum* highlighted our lack of knowledge on the diversity of this genus (Basso *et al.* 2015, Hernandez-Kantun *et al.* 2015 & 2016, Peña *et al.* 2018). Previous studies (e.g. Bittner *et al.* 2011; Hernandez-Kantun *et al.* 2015; Rosler *et al*. 2016) already underlined the necessity of a detailed molecular study of the *Lithophyllum* genus to unravel potential cryptic diversity. Recently, Pezzolesi *et al.* (2019) unravelled cryptic diversity in the *L. stictiforme* complex using three barcode markers. Their study showed the presence of at least 13 species at the Mediterranean scale. Our results underline the usefulness of molecular tools to delineate species in this genus, whereas determination of the species *in situ* or by observing classical macro morphological characters is seldom possible.

## 4.2 Community composition and genetic structure cannot wholly be explained by spatial distances and current patterns.

At the global scale of this study, the high difference in composition between the Banyuls-sur-mer (westernmost) cryptic species community and all other communities (Figure 2) is noteworthy. There are three non-exclusive explanations. (1) The scarcity of suitable habitats for these cryptic species (i.e. rocky substratum found in dim light condition) between the Banyuls-sur-mer (BPT) and the Couronne (COU) localities (Martin *et al.* 2014) may impede the stepwise colonization across these areas, even considering several generations, and the high geographic distance separating Banyuls-sur-mer from the other study sites may impede the colonization in a single generation of propagules (mainly spores in CCA). (2) The Rhône

flow at the west of the COU locality and the presence of vortex structures in the Lion Gulf may constitute barriers to dispersal. (3) The different environmental conditions found in Banyuls-sur-mer may select for different cryptic species: salinity and water temperature are highly variable and turbidity is higher in Banyuls-sur-mer compared to any other sampling site (SOMLIT data: http://somlit.epoc.u-bordeaux1.fr/fr/). At the opposite Eastern end of the study, the Villefranche-sur-mer (VPR) community position on the nMDS could be explained both by its eastern geographical origin and by the fact that it was sampled at shallower depth than the other localities (between 15 and 20 m).

Finally, all other localities were clustered on the nMDS (Figure 2B) in a group gathering the Marseille Area (except the CAS locality), the CIR and the LDM localities, despite the important geographic distances between them. All together this pattern suggests that geographical distance alone does not provide a sufficient explanation of the distribution of these eight cryptic species at the regional scale. Thus, their abundances may be influenced by complicated current patterns and/or by changing environmental factors across the different sampling sites.


At a closer scale around the bay of Marseille, we know the current patterns in more detail and we can benefit from comparative population genetics studies to investigate whether currents and distance can explain species composition (and genetic structure within species). Globally, the *Lithophyllum stictiforme / L. cabiochiae* display a good concordance between cryptic species community and genetic structures in the Marseille Area. The frequencies of the eight cryptic species were highly variable among localities within the Bay of Marseille but three clusters were distinguishable based on community similarities (Figure 3): (i) the Côte Bleue (COU, CSC, MEJ) and LPD; (ii) Marseille (FTF & RMO), (iii) the PSO and CTF localities.

The CAS locality community composition was highly distinct from all the others. The genetic diversity structure within cryptic species C1 presented similarities with the community structure pattern with the same 3 spatial clusters: the CAS population was highly differentiated from all the other populations (Figure 4 & Supplementary Material VII), the LPD and COU populations on one hand, the FTF and RMO populations on the other hand clustered together on the PCA (Figure 4A).

Homogenization of species (or allele) frequencies among localities requires both (i) migration of the viable propagules among localities and (ii) successful settlement and growth in new localities (depending among other things on the availability of suitable habitats).

The first condition is mainly determined by the propagules' ability to disperse and the hydrodynamics of water masses in the area (Cowen & Sponaugle 2009, Weersing & Toonen 2009). In CCA, propagules have low dispersal capacities and settle closely to their source of emission (Norton 1992, Opazo & Otaíza 2007) which can explain the differentiation among localities observed in this study at both the interspecific and the intraspecific levels.

Comparing $F_{ST}$ value in this study ($F_{ST}$ global 0.0464) with those obtained by Cahill *et al.* (2017) using data from several invertebrate species sampled in the same geographic area, the C1 species would correspond to the lecithotroph "larval type" species in terms of dispersal capacities. Both population genomics and cryptic species composition patterns showed a differentiation between the Côte Bleue, and the east part of the Bay of Marseille. This pattern has been identified in previous studies encompassing 9 animal taxa (Cahill *et al.* 2017), and a brown alga (Thibaut *et al.* 2016) and can be attributed to the main currents in the Bay of Marseille that prevent connectivity between these two areas (Pradal & Millet 2013) at least in sessile organisms with low dispersal abilities (Cahill *et al.* 2017). The CAS locality presents the highest level of *L. stictiforme/cabiochiae* cryptic species diversity (Table 1) and is also

very distinct from the other localities (Figure 3 B) in species composition. Moreover, within the C1 species, the CAS population is highly differentiated from the others (Figure 4 & Supplementary Material VII). This sampling locality is east of the head of the "Cassidaigne" marine canyon and presents very different characteristics from other sampled sites, both in terms of currents and biogeochemical parameters. In particular, this locality is not subjected to upwellings as strong as in the area west of the canyon (such as the RMO locality), is more influenced by the North current and often experiences deep eastward currents (Albérola & Millot 2003, Pairaud *et al.* 2011).

### 4.3 Cryptic species community composition is influenced by environmental factors.

Indeed, our combined results in community composition and population genomics established that environmental factors are influential in the composition in cryptic species. The cryptic species community in the RMO and CAS localities were different among depth categories. This was especially true in the CAS locality, where the C5 species was totally absent at deepest sites ((the D2 depth category) and replaced by the C8 species (Figure 2). At a higher taxonomic level, Sartoretto *et al.* (1996) also observed a shift in the frequencies of the main algal builders in corralligenous habitats in the Marseille area across depths: *L. cabiochiae* was reported as the dominant one in deep waters whereas *Mesophyllum alternans* was more restricted to shallower waters. Importantly, no significant genetic differentiation was found for the C1 species between the populations from the D1 and D2 depth categories in the CAS locality (Figure 4 & Supplementary Material VII), suggesting that gene flow between the two depths is not restricted and propagules can travel between the two depth categories. Thus, the differentiation of the communities between depths (found in the PERMANCOVA) should be

explained by environmental factors varying across depths such as light or temperature. Light is known to have an influence on coralligenous assemblages (Ballesteros 2006), because CCA only develop at specific values of irradiance (Ballesteros 1992). Our study shows an influence of the PAR on the community composition of the cryptic species. It could therefore be interesting to experimentally compare physiological parameters of the distinct cryptic species such as photosynthesis, growth rate, and carbonate precipitation under different level of irradiances. Differences in temperature (particularly in temperature variability) may also have a role but our depth categories (30 m and 40-45m) are not very contrasted in relation to temperature, both being below the summer thermocline threshold (around 16-20 m) (Harmelin 2004, Haguenauer *et al.* 2013). Finally, experimental studies showed that the interaction between light and temperature impacted survival in *L. stictiforme* (Rodríguez-Prieto 2016).

## 4.4 Ecological consequences and conservation implications of the cryptic diversity for coralligenous habitats

Different engineering species harboring different ecological traits (Badano & Cavieres 2006; Lamit *et al.* 2011) or phenotypic variation among individuals of the same species (Whitham *et al.* 2003) influence the diversity and structure of the associated communities. Since the cryptic species of the *L. stictiforme/cabiochiae* complex are, together with other CCA of the genus *Mesophyllum,* the main engineers of the coralligenous habitats, the distribution of these cryptic engineering species as well as their intraspecific genetic structure potentially have important consequences on the composition of the benthic assemblages found on

coralligenous habitats. CCA are also known to influence the settlement of other invertebrates by producing chemical cues inducing the recruitment of larvae in several habitats (e.g. coral reefs and vermetid reefs) (Diaz-Pulido *et al.* 2010; Spotorno-Oliveira *et al.* 2015, Quéré & Nugues 2015; Elmer *et al.* 2018). To our knowledge, these kinds of interactions between L. *stictiforme/cabiochiae* and invertebrate larvae (e.g. Anthozoa) have not been studied in coralligenous habitats.

Coralligenous habitats are a major marine biodiversity hotspot of the Mediterranean Sea, yet their protection is still pending mainly because there is still a large gap of knowledge about biodiversity and ecological functioning of these habitats (SPA/RAC 2017).

Engineer species are priority targets for conservation programs because their protection has a large impact in retaining community and ecosystems integrity and functions (Crain and Bertness 2006). The high structure found both at the species and the genetic diversity levels make these cryptic species particularly vulnerable to local threats such as water pollution or mechanical degradation. Protecting coralligenous habitats (or just evaluating their vulnerability) requires taking into account the geographic distribution of the eight cryptic species along the French Mediterranean coast and at smaller spatial scales. Due to the high *L. stictiforme/cabiochiae* community composition differentiation between the biogeographic regions of Banyuls-sur-mer, Marseille and Villefranche-sur-mer, it is necessary to consider each of these areas as a unique protection unit. In the Marseille area, the CAS locality harbored the highest level of cryptic species diversity and is the only locality where the C5, C7 and C8 species are found; the C4 species is mainly found on Côte Bleue (Figure 3 and Table 1). Consequently, to have all the cryptic species under protection requires protecting at least the CAS locality and one locality on the Côte Bleue. Our study strongly suggests that these eight cryptic species have different biotope preferences potentially reflecting contrasted

physiological abilities. The ocean acidification and warming components of the global change due to human activity are two of the major threats on coralligenous habitats (Ballesteros 2006). Recent studies by Rodríguez-Prieto (2016) and Martin *et al.* (2013) showed that the metabolism, reproduction and survival of the *L. stictiforme/cabiochiae* species complex are affected by irradiance levels, temperature and pCO2. However, in these studies species were identified ignoring the presence of cryptic species and thus missing potential differences of responses of the cryptic species to warming and acidification. It emphasizes the need for more investigations to determine if these different species have different capacities to cope with global change.

Finally, genetic diversity and structure are both key pieces of information needed to design efficient species protection policy. This study is the first that gives an insight into the genetic diversity and structure at the genomic level for the bioengineer algae of coralligenous habitats (C1 species): genetic differentiation occurred at a very small spatial scale resulting from small dispersal capacities of propagules and particularities of the currents in the Marseille area (Figure 4, Supplementary Material VII). Yet neither the genetic structure at a larger geographical scale nor the impact of selective processes potentially shaping the differentiation between populations living in variable environmental conditions are known. Therefore, investigating genetic diversity and structure in these cryptic species can reveal different capacity (different level of genetic diversity for example) of adaptation to global change which should be considered in conservation policy of coralligenous habitats.

## Acknowledgments

## Funding

# References

Albérola, C., & Millot, C. (2003). Circulation in the French mediterranean coastal zone near Marseilles: the influence of wind and the Northern Current. *Continental Shelf Research*, *23*(6), 587–610. doi:10.1016/S0278-4343(03)00002-5

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Athanasiadis, A. (1999). The taxonomic status of *Lithophyllum stìctaeforme* (Rhodophyta, Corallinales) and its generic position in light of phylogenetic considerations. *Nordic Journal of Botany*, *19*(6), 735–745. doi:10.1111/j.1756-1051.1999.tb00682.x

Badano, E. I., & Cavieres, L. A. (2006). Impacts of ecosystem engineers on community attributes: effects of cushion plants at different elevations of the Chilean Andes. *Diversity and Distributions*, *12*(4), 388–396. doi:10.1111/j.1366-9516.2006.00248.x

Bailleul, D., Stoeckel, S., & Arnaud-Haond, S. (2016). RClone: a package to identify MultiLocus Clonal Lineages and handle clonal data sets in r. *Methods in Ecology and Evolution*, *7*(8), 966–970. doi:10.1111/2041-210X.12550

Balata, D., Piazzi, L., Cecchi, E., & Cinelli, F. (2005). Variability of Mediterranean coralligenous assemblages subject to local variation in sediment deposition. *Marine Environmental Research*, *60*(4), 403–421. doi:10.1016/j.marenvres.2004.12.005

Balata, D., Piazzi, L., & Cinelli, F. (2007). Increase of sedimentation in a subtidal system: Effects on the structure and diversity of macroalgal assemblages. *Journal of Experimental Marine Biology and Ecology*, *351*(1–2), 73–82. doi:10.1016/j.jembe.2007.06.019

Ballesteros, E. (1992). *Els vegetals i la zonació litoral: espècies, comunitats i factors que influeixen en la seva distribució*. Institut d'Estudis Catalans.

Ballesteros, E. (2006). Mediterranean coralligenous assemblages: a synthesis of present knowledge. *Oceanography and Marine Biology: An Annual Review*, *44*, 123–195.

Bandelt, H. J., Forster, P., & Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. doi:10.1093/oxfordjournals.molbev.a026036

Basso, D., Caragnano, A., Gal, L. L., & Rodondi, G. (2015). The genus *Lithophyllum* in the north-western Indian Ocean, with description of L. yemenense sp. nov. , L. socotraense sp. nov. , L. subplicatum comb. et stat. nov. , and the resumed L. affine , L. kaiseri , and L. subreduncum (Rhodophyta, Corallinales). *Phytotaxa*, *208*(3), 183–200. doi:10.11646/phytotaxa.208.3.1

Bassi, D., Braga, J. C., & Iryu, Y. (2009). Palaeobiogeographic patterns of a persistent monophyletic lineage: *Lithophyllum pustulatum* species group (Corallinaceae, Corallinales, Rhodophyta). *Palaeogeography, Palaeoclimatology, Palaeoecology*, *284*(3), 237–245. doi:10.1016/j.palaeo.2009.10.003

Bittner, L., Payri, C. E., Maneveldt, G. W., Couloux, A., Cruaud, C., de Reviers, B., & Le Gall, L. (2011). Evolutionary history of the Corallinales (Corallinophycidae, Rhodophyta) inferred from nuclear, plastidial and mitochondrial genomes. *Molecular Phylogenetics and Evolution*, *61*(3), 697–713. doi:10.1016/j.ympev.2011.07.019

Boudouresque, C.-F. (2004). Marine biodiversity in the mediterranean: status of spicies, populations and communities. *Scientific Reports of the Port-Cros National Park*, *20*, 97–146.

Cabau, C., Escudié, F., Djari, A., Guiguen, Y., Bobe, J., & Klopp, C. (2017). Compacting and correcting Trinity and Oases RNA-Seq de novo assemblies. *PeerJ*, *5*, e2988. doi:10.7717/peerj.2988

Cahill, A. E., De Jode, A., Dubois, S., Bouzaza, Z., Aurelle, D., Boissin, E., … Chenuil, A. (2017). A multispecies approach reveals hot spots and cold spots of diversity and connectivity in invertebrate species with contrasting dispersal modes. *Molecular Ecology*, *26*, 6563–6577. doi:10.1111/mec.14389

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421. doi:10.1186/1471-2105-10-421

Chenuil A, Cahill AE, Délémontey N, Du Salliant Du Luc E, Fanton A. (in press). Problems and Questions Posed by Cryptic Species. A Framework to Guide Future Studies. In E. Casetta et al. (eds.), From Assessing to Conserving Biodiversity , History, Philosophy and Theory of the Life Sciences 24, Springer Nature Switzerland. https://doi.org/10.1007/978-3-030-10991-2_4

Clarke K.R., Gorley R.N., Somerfield P.J., Warwick R.M. (2014). Change in marine communities: an approach to statistical analysis and interpretation, 3rd edition. PRIMER-E, Plymouth, 260pp

Clarke, K.R., Gorley, R.N. (2015). PRIMER v7: User Manual/Tutorial. PRIMER-E, Plymouth, 296pp

Cowen, R. K., & Sponaugle, S. (2009). Larval Dispersal and Marine Population Connectivity. *Annual Review of Marine Science*, *1*(1), 443–466. doi:10.1146/annurev.marine.010908.163757

Crain, C. M., & Bertness, M. D. (2006). Ecosystem Engineering across Environmental Gradients: Implications for Conservation and Management. *BioScience*, *56*(3), 211–218. doi:10.1641/0006-3568(2006)056[0211:EEAEGI]2.0.CO;2

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi:10.1093/bioinformatics/btr330

Diaz-Pulido, G., Harii, S., McCook, L. J., & Hoegh-Guldberg, O. (2010). The impact of benthic algae on the settlement of a reef-building coral. *Coral Reefs*, *29*(1), 203–208. doi:10.1007/s00338-009-0573-x

Dudgeon, S., Kübler, J. E., West, J. A., Kamiya, M., & Krueger-Hadfield, S. A. (2017). Asexuality and the cryptic species problem. *Perspectives in Phycology*, *4*(1), 47–59. doi:10.1127/pip/2017/0070

Elmer, F., Bell, J. J., & Gardner, J. P. A. (2018). Coral larvae change their settlement preference for crustose coralline algae dependent on availability of bare space. *Coral Reefs*, *37*(2), 397–407. doi:10.1007/s00338-018-1665-2

Gayral, P., Melo-Ferreira, J., Glémin, S., Bierne, N., Carneiro, M., Nabholz, B., … Galtier, N. (2013). Reference-Free Population Genomics from Next-Generation Transcriptome Data and the Vertebrate–Invertebrate Gap. *PLoS Genetics*, *9*(4), e1003457. doi:10.1371/journal.pgen.1003457

Haas, B. J. (2013). Transdecoder v.3.0.0 https://github.com/TransDecoder/TransDecoder/releases/tag/v3.0.0

Haguenauer, A., Zuberer, F., Ledoux, J.-B., & Aurelle, D. (2013). Adaptive abilities of the Mediterranean red coral Corallium rubrum in a heterogeneous and changing environment: from population to functional genetics. *Journal of Experimental Marine Biology and Ecology*, *449*, 349–357. doi:10.1016/j.jembe.2013.10.010

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, *41*, 95–98.

Harmelin, J.-G.. (2004). Environnement thermique du benthos côtier de l'île de Port-Cros (parc national, France, Méditerranée nord-occidentale) et implications biogéographiques. *Scientific Reports of the Port-Cros National Park*, *20*, 173–194.

Harper, J. T., & Saunders, G. W. (2001). The application of sequences of the ribosomal cistron to the systematics and classification of the florideophyte red algae (Florideophyceae, Rhodophyta). *Cahiers de Biologie Marine*, *42*(1–2), 25–38.

Hernandez-Kantun, J. J., Rindi, F., Adey, W. H., Heesch, S., Peña, V., Le Gall, L., & Gabrielson, P. W. (2015). Sequencing type material resolves the identity and distribution of the generitype *Lithophyllum incrustans*, and related European species L. hibernicum and L. bathyporum (Corallinales, Rhodophyta). *Journal of Phycology*, *51*(4), 791–807. doi:10.1111/jpy.12319

Guiry, M.D. & Guiry, G.M. 2018. *AlgaeBase*. World-wide electronic publication, National University of Ireland, Galway. http://www.algaebase.org

Hernandez-Kantun, J. J., Gabrielson, P., Hughey, J. R., Pezzolesi, L., Rindi, F., Robinson, N. M., Peña V., Riosmena-Rodiguez, R., Le Gall, L., Adey, W. (2016). Reassessment of branched *Lithophyllum* spp. (Corallinales, Rhodophyta) in the Caribbean Sea with global implications. *Phycologia*, *55*(6), 619–639. doi:10.2216/16-7.1

Hong, J.-S. (1980). *Etude faunistique d'un fond de concrétionnement de type coralligène soumis à un gradient de pollution en Méditerrannée nord-occidentale (Golf de Fos)*. Faculté des Sciences, Aix-Marseille II.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. doi:10.1093/bioinformatics/btn129

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. doi:10.1093/bioinformatics/btr521

Jones, C. G., Lawton, J. H., & Shachak, M. (1994). Organisms as Ecosystem Engineers. In *Ecosystem Management* (pp. 130–147). Springer, New York, NY. doi:10.1007/978-1-4612-4018-1_14

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. doi:10.1007/BF01731581

Krueger-Hadfield, S. A., Roze, D., Correa, J. A., Destombe, C., & Valero, M. (2015). O father where art thou? Paternity analyses in a natural population of the haploid–diploid seaweed *Chondrus crispus*. *Heredity*, *114*(2), 185–194. doi:10.1038/hdy.2014.82

Krueger-Hadfield, Stacy A., Collén, J., Daguin-Thiébaut, C., & Valero, M. (2011). GENETIC POPULATION STRUCTURE AND MATING SYSTEM IN CHONDRUS CRISPUS (RHODOPHYTA)1: MATING SYSTEM IN CHONDRUS CRISPUS. *Journal of Phycology*, *47*(3), 440–450. doi:10.1111/j.1529-8817.2011.00995.x

Laborel, J. 1961. Le concretionnement algal "coralligène" et son importance géomorphologique en Méditerranée. Recueil Travaux Station Marine d'Endoume , 23: 37-60.

Lamit, L. J., Wojtowicz, T., Kovacs, Z., Wooley, S. C., Zinkgraf, M., Whitham, T. G., … Gehring, C. A. (2011). Hybridization among foundation tree species influences the structure of associated understory plant communities. *Botany*, *89*(3), 165–174. doi:10.1139/b11-006

Laubier, L. (1966). *Le coralligène des Albères : monographie biocénotique.* (Vol. 137–316). Monaco: Annales de l'Institut océanographique.

Le Gall, L., & Saunders, G. W. (2007). A nuclear phylogeny of the Florideophyceae (Rhodophyta) inferred from combined EF2, small subunit and large subunit ribosomal DNA:

Establishing the new red algal subclass Corallinophycidae. *Molecular Phylogenetics and Evolution*, *43*(3), 1118–1130. doi:10.1016/j.ympev.2006.11.012

Le Gall, L., Delsuc, F., Hourdez, S., Lecointre, G., & Rasplus, J.-Y. (2017). Toward the DNA Library of Life. *European Journal of Taxonomy*, (266). doi:10.5852/ejt.2017.266

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. Retrieved from http://arxiv.org/abs/1303.3997

Linares, C., Vidal, M., Canals, M., Kersting, D. K., Amblas, D., Aspillaga, E., … Ballesteros, E. (2015). Persistent natural acidification drives major distribution shifts in marine benthic ecosystems. *Proc. R. Soc. B*, *282*(1818), 20150587. doi:10.1098/rspb.2015.0587

Lombardi, C., Gambi, M. C., Vasapollo, C., Taylor, P., & Cocito, S. (2011). Skeletal alterations and polymorphism in a Mediterranean bryozoan at natural $CO_2$ vents. *Zoomorphology*, *130*(2), 135–145. doi:10.1007/s00435-011-0127-y

Martin, C. S., Giannoulaki, M., Leo, F. D., Scardi, M., Salomidi, M., Knittweis, L., … Fraschetti, S. (2014). Coralligenous and maërl habitats: predictive modelling to identify their spatial distributions across the Mediterranean Sea. *Scientific Reports*, *4*, 5073. doi:10.1038/srep05073

Martin, S., & Gattuso, J.-P. (2009). Response of Mediterranean coralline algae to ocean acidification and elevated temperature. *Global Change Biology*, *15*(8), 2089–2100.

Martin, S., Cohu, S., Vignot, C., Zimmerman, G., & Gattuso, J.-P. (2013). One-year experiment on the physiological response of the Mediterranean crustose coralline alga, Lithophyllum cabiochae , to elevated $pCO_2$ and temperature. *Ecology and Evolution*, *3*(3), 676–693. doi:10.1002/ece3.475

Norton, T. A. (1992). Dispersal by macroalgae. *British Phycological Journal*, *27*(3), 293–301. doi:10.1080/00071619200650271

Opazo, L. F., & Otaíza, R. D. (2007). Vertical distribution of spores of blade-forming *Sarcothalia crispata* (Gigartinaceae) and crustose corallines (Corallinaceae) in the water column. *Botanica Marina*, *50*(2). doi:10.1515/BOT.2007.011

Pairaud, I. L., Gatti, J., Bensoussan, N., Verney, R., & Garreau, P. (2011). Hydrology and circulation in a coastal area off Marseille: Validation of a nested 3D model with observations. *Journal of Marine Systems*, *88*(1), 20–33. doi:10.1016/j.jmarsys.2011.02.010

Paoli, C., Montefalcone, M., Morri, C., Vassallo, P., & Bianchi, C. N. (2016). Ecosystem Functions and Services of the Marine Animal Forests. In S. Rossi, L. Bramanti, A. Gori, & C. Orejas Saco del Valle (Eds.), *Marine Animal Forests* (pp. 1–42). Cham: Springer International Publishing. doi:10.1007/978-3-319-17001-5_38-1

Paradis E., Claude J. & Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 20: 289-290.

Pardo C., Lopez L., Peña V., Hernández-Kantún J., Le Gall L., Bárbara I. & Barreiro R. 2014 - A multilocus species delimitation reveals a striking number of species of coralline algae forming maerl in the OSPAR maritime area. PLoS ONE 9 (8) : e104073.

Pardo, C., Bárbara, I., Barreiro, R., & Peña, V. (2017). Insights into species diversity of associated crustose coralline algae ( Corallinophycidae, Rhodophyta ) with Atlantic European maerl beds using DNA barcoding. *Anales Del Jardín Botánico de Madrid*, *74*(2), 059.

Peña V., De Clerck O., Afonso-Carrillo J., Ballesteros E., Bárbara I., Barreiro R. & Le Gall L. 2015 - An integrative systematic approach to species diversity and distribution in the genus Mesophyllum (Corallinales, Rhodophyta) in Atlantic and Mediterranean Europe. European Journal of Phycology 50 (1) : 20–36.

Peña, V., Hernandez-Kantun, J. J., Adey, W. H., & Gall, L. L. (2018). Assessment of Coralline Species Diversity in the European Coasts Supported by Sequencing of Type Material: The

Case Study of *Lithophyllum nitorum* (Corallinales, Rhodophyta). *Cryptogamie, Algologie*, *39*(1), 123–137. doi:10.7872/crya/v39.iss1.2018.123

Pezzolesi, L., Falace, A., Kaleb, S., Hernandez-Kantun, J. J., Cerrano, C., & Rindi, F. (2017). Genetic and morphological variation in an ecosystem engineer, *Lithophyllum byssoides* (Corallinales, Rhodophyta). *Journal of Phycology*, *53*(1), 146–160. doi:10.1111/jpy.12488

Pezzolesi, L., Peña, V., Gall, L. L., Gabrielson, P. W., Kaleb, S., Hughey, J. R., Rindi, F. (2019). Mediterranean Lithophyllum stictiforme (Corallinales, Rhodophyta) is a genetically diverse species complex: implications for species circumscription, biogeography and conservation of coralligenous habitats. *Journal of Phycology*, *0*(ja). doi:10.1111/jpy.12837

Pradal, M.-A., & Millet, B. (2013). Spatial Heterogeneity of Artificial Reefs Functioning according to Wind-Induced Lagrangian Circulation. *ISRN Oceanography*, *2013*, 1–9. doi:10.5402/2013/568487

Quéré, G., & Nugues, M. M. (2015). Coralline algae disease reduces survival and settlement success of coral planulae in laboratory experiments. *Coral Reefs*, *34*(3), 863–870. doi:10.1007/s00338-015-1292-0

Revell, L. J. (2012) phytools: An R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3 217-223. doi:10.1111/j.2041-210X.2011.00169.x

Rindi, F., Peña, V., Le Gall L., Braga J. C., Falace A., Hernandez-Kantun, J. J., Pezzolesi, L., Kaleb S. (2017). Evolutionary history and diversity of Mediterranean coralline algae: how much do we know?
August 2017, Conference: 11th International Phycological Congress. *Phycologia* 56(4) (Suppl.): 158-159

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rodríguez-Prieto, C. (2016). Light and temperature requirements for survival, growth and reproduction of the crustose coralline *Lithophyllum stictaeforme* from the Mediterranean Sea. *Botanica Marina*, *59*(2–3). doi:10.1515/bot-2015-0070

Romero, G. Q., Gonçalves-Souza, T., Vieira, C., & Koricheva, J. (2015). Ecosystem engineering effects on species diversity across ecosystems: a meta-analysis: Ecosystem engineering effects across ecosystems. *Biological Reviews*, *90*(3), 877–890. doi:10.1111/brv.12138

Rösler, A., Perfectti, F., Peña, V., & Braga, J. C. (2016). Phylogenetic relationships of corallinaceae (Corallinales, Rhodophyta): taxonomic implications for reef-building corallines. *Journal of Phycology*, *52*(3), 412–431. doi:10.1111/jpy.12404

Rousset, F. (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103–106. doi:10.1111/j.1471-8286.2007.01931.x

Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989). Molecular cloning: a laboratory manual. *Molecular Cloning: A Laboratory Manual.*, (Ed. 2). Retrieved from https://www.cabdirect.org/cabdirect/abstract/19901616061

Sartoretto, S., Verlaque, M., & Laborel, J. (1996). Age of settlement and accumulation rate of submarine "coralligène"(- 10 to- 60 m) of the northwestern Mediterranean Sea; relation to Holocene rise in sea level. *Marine Geology*, *130*(3–4), 317–331.

Saunders, G. W. (2005). Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *360*(1462), 1879–1888. doi:10.1098/rstb.2005.1719

Saunders, G. W., & Moore, T. E. (2013). Refinements for the amplification and sequencing of red algal DNA barcode and RedToL phylogenetic markers: a summary of current primers, profiles and strategies. *ALGAE*, *28*(1), 31–43. doi:10.4490/algae.2013.28.1.031

Smit, A. F. A., Hubley, R. & Green, P. *RepeatMasker Open-4.0*. 2013-2015 <http://www.repeatmasker.org>. unpublished data. Current Version: open-4.0.6 (RMLib: 20160829 & Dfam: 2.0)

SPA/RAC (2017). Action Plan for the Conservation of the Coralligenous and Other Calcareous Bio-concretions in the Mediterranean Sea. UN Environment/MAP Athens. Greece 2017. Retrieved from http://www.rac-spa.org/sites/default/files/action_plans/pa_coral_en.pdf

Spotorno-Oliveira, P., Figueiredo, M. A. O., & Tâmega, F. T. S. (2015). Coralline algae enhance the settlement of the vermetid gastropod *Dendropoma irregulare* (d'Orbigny, 1842) in the southwestern Atlantic. *Journal of Experimental Marine Biology and Ecology*, *471*, 137–145. doi:10.1016/j.jembe.2015.05.021

Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Molecular Biology and Evolution*, *24*(8), 1596–1599. doi:10.1093/molbev/msm092

Teagle, H., Hawkins, S. J., Moore, P. J., & Smale, D. A. (2017). The role of kelp species as biogenic habitat formers in coastal marine ecosystems. *Journal of Experimental Marine Biology and Ecology*, *492*, 81–98. doi:10.1016/j.jembe.2017.01.017

Thibaut, T., Bottin, L., Aurelle, D., Boudouresque, C.-F., Blanfuné, A., Verlaque, M., … Millet, B. (2016). Connectivity of Populations of the Seaweed Cystoseira amentacea within the Bay of Marseille (Mediterranean Sea): Genetic Structure and Hydrodynamic Connections. *Cryptogamie, Algologie*, *37*(4), 233–255. doi:10.7872/crya/v37.iss4.2016.233

Thierry de Ville D 'Avray, L., Ami, D., Chenuil, A., David, R., Féral, J.-P. (2019). Application of the ecosystem service concept at a small-scale: The cases of coralligenous habitats in the North-western Mediterranean Sea. Marine Pollution Bulletin, 138, 160-170. doi:10.1016/j.marpolbul.2018.10.057

Tsagkogeorga, G., Cahais, V., & Galtier, N. (2012). The Population Genomics of a Fast Evolver: High Levels of Diversity, Functional Constraint, and Molecular Adaptation in the Tunicate Ciona intestinalis. *Genome Biology and Evolution*, *4*(8), 852–861. doi:10.1093/gbe/evs054

Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., … DePristo, M. A. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis et Al.*, *11*(1110), 11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43

Walsh, P. S., Metzger, D. A., & Higuchi, R. (1991). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, *10*(4), 506–513.

Watt, C. A., & Scrosati, R. A. (2013). Bioengineer effects on understory species richness, diversity, and composition change along an environmental stress gradient: Experimental and mensurative evidence. *Estuarine, Coastal and Shelf Science*, *123*, 10–18. doi:10.1016/j.ecss.2013.02.006

Weersing, K., & Toonen, R. (2009). Population genetics, larval dispersal, and connectivity in marine systems. *Marine Ecology Progress Series*, *393*, 1–12. doi:10.3354/meps08287

Whitham Thomas G., Young William P., Martinsen Gregory D., Gehring Catherine A., Schweitzer Jennifer A., Shuster Stephen M., … Kuske Cheryl R. (2003). Community and

ecosystem genetics: a consequence of the extended phenotype. *Ecology*, *84*(3), 559–573.

doi:10.1890/0012-9658(2003)084[0559:CAEGAC]2.0.CO;2

Yoon, H. S., Hackett, J. D., & Bhattacharya, D. (2002). A single origin of the peridinin-and

fucoxanthin-containing plastids in dinoflagellates through tertiary endosymbiosis.

*Proceedings of the National Academy of Sciences*, *99*(18), 11724–11729.

## Data statement

New sequences data generated for this project will be deposited in Genbank prior to

publications. The raw data files from all the Illumina sequencing runs will be deposited at

DataDryad prior to publication. Bioinformatics code for processing transcriptomics and

capture sequencing data will be available on Github. Bioinformatics scripts for processing

Miseq amplicon Sequencing are available in Github (hhttps://github.com/chaby/dana).

## Author contributions

AC, J-PF, RD and ADJ conceived the study.

ADJ, ZE, RD, AH, AEC, DG, SS, CR and MS produced the data.

ADJ and AC analyzed the data and wrote the paper.

ADJ, AEC, SS, LLG and AC contributed to manuscript editing and revisions.

Tables

*Table 1 Community diversity index for each locality. S: total number of species. N: total number of individuals. D: Margalef species richness index. J': Pielou evenness index. H': Shannon index,1-Lambda': Simpson corrected for small samples. N1: Hill number of order one. N2: Hill number of order 2*

| LOCALITY | S | N | D | J' | H'(LOGE) | 1-LAMBDA' | N1 | N2 |
|----------|---|----|--------|--------|----------|-----------|-------|-------|
| BPT | 3 | 25 | 0.6213 | 0.835 | 0.9174 | 0.57 | 2.503 | 2.208 |
| CAS | 5 | 67 | 0.9513 | 0.7326 | 1.179 | 0.592 | 3.252 | 2.399 |
| CIR | 5 | 69 | 0.9447 | 0.3603 | 0.5799 | 0.2835 | 1.786 | 1.388 |
| COU | 2 | 19 | 0.3396 | 0.9495 | 0.6581 | 0.4912 | 1.931 | 1.87 |
| CSC | 3 | 28 | 0.6002 | 0.6908 | 0.7589 | 0.455 | 2.136 | 1.782 |
| CTF | 2 | 37 | 0.2769 | 0.909 | 0.6301 | 0.4505 | 1.878 | 1.78 |
| FTF | 3 | 51 | 0.5087 | 0.2904 | 0.319 | 0.1498 | 1.376 | 1.172 |
| LDM | 3 | 34 | 0.5672 | 0.4039 | 0.4438 | 0.221 | 1.559 | 1.273 |
| LPD | 3 | 20 | 0.6676 | 0.865 | 0.9503 | 0.5895 | 2.586 | 2.273 |
| MEJ | 3 | 58 | 0.4926 | 0.6568 | 0.7216 | 0.4168 | 2.058 | 1.694 |
| MOY | 3 | 56 | 0.4969 | 0.3145 | 0.3456 | 0.1682 | 1.413 | 1.198 |
| PLN | 2 | 28 | 0.3001 | 0.8113 | 0.5623 | 0.3889 | 1.755 | 1.6 |
| VPR | 4 | 15 | 1.108 | 0.9665 | 1.34 | 0.781 | 3.818 | 3.689 |

*Table 2 Results of the different PERMANOVA designs. Details of the PERMANOVA tables are available in Supplementary Material V*

| | DEPTH CATEGORY | SLOPE | ORIENTATION | RUGOSITY |
|---|---|---|---|---|
| *RANDOM LOCALITY FACTOR* | P(perm)=0.0001 | P(perm)=0.0001 | P(perm)=0.0001 | P(perm)=0.0003 |
| *FIXED ENVIRONMENTAL FACTOR* | P(perm)=0.2844 | P(perm)=0.7645 | P(perm)=0.7176 | P(perm)=0.7969 |
| *INTERACTION* | P(perm)=0.014 | P(perm)=0.3684 | P(perm)=0.3765 | P(perm)=0.3407 |

*Table 3 Results of the different PERMANCOVA designs. Details of the PERMANOVA tables are available in Supplementary Material V*

| | DEPTH | PAR |
|---|---|---|
| *RANDOM LOCALITY FACTOR* | P(perm)=0.0001 | P(perm)=0.001 |
| *NUMERICAL COVARIABLE* | P(perm)=0.0015 | P(perm)=0.0019 |

| *INTERACTION* | P(perm)=0.0582 | P(perm)=0.0773 |
|---|---|---|

Table 4 Genetic diversity (expected heterozygosity) and $F_{IS}$ of the C1 species calculated in each population and each locality (in grey). Populations with a very low individual count (< 9) are colored in red. *Significant results, ** highly significant results

| Pop | Number of individuals | Gene diversity | $F_{IS}$ |
|---|---|---|---|
| CAS_D1 | 16 | 0.1267 | 0.0245** |
| CAS_D2 | 14 | 0.1226 | -0.0096** |
| CAS | 30 | 0.1249 | 0.0159** |
| COU | 3 | 0.1112 | -0.2431 |
| FTF_D1 | 15 | 0.1172 | 0.1260** |
| FTF_D2 | 15 | 0.1204 | 0.0124** |
| FTF | 30 | 0.1188 | 0.0696** |
| LPD | 9 | 0.1090 | -0.0477 |
| RMO | 3 | 0.1603 | -0.1712 |
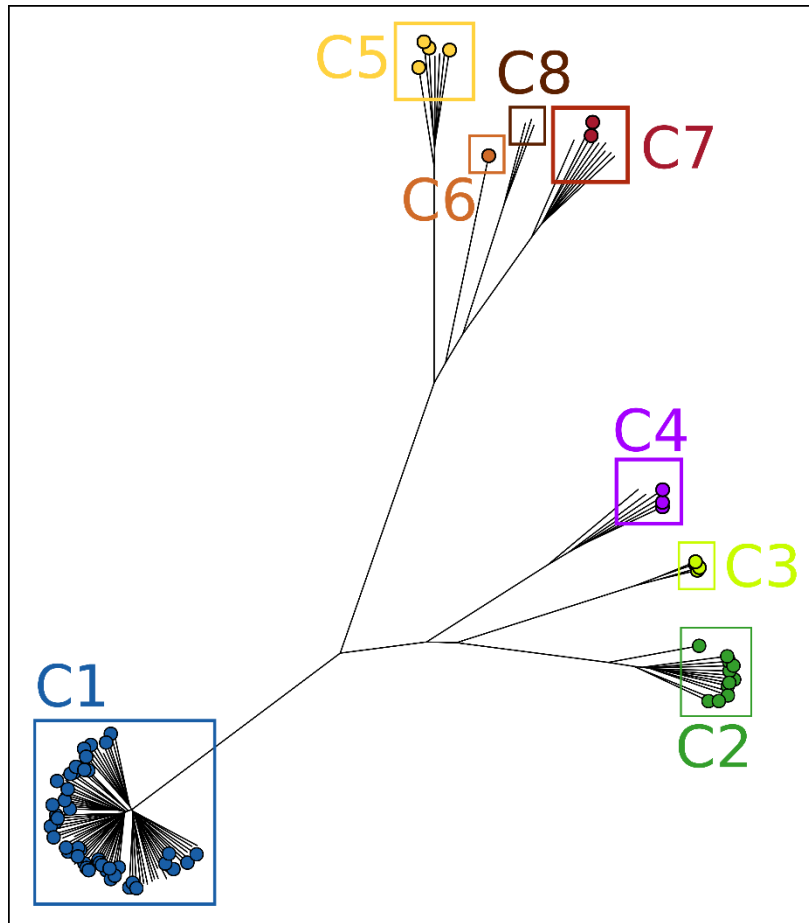
*Figure 1 Neighbour Joinning tree based on euclidean distances between individual multiloci genoptype obtained by capture sequencing (7068 SNPs). Individuals are colored according to their haplogroup determined using the psbA marker. Uncolored tips correspond to individuals with no barcoding sequence.*
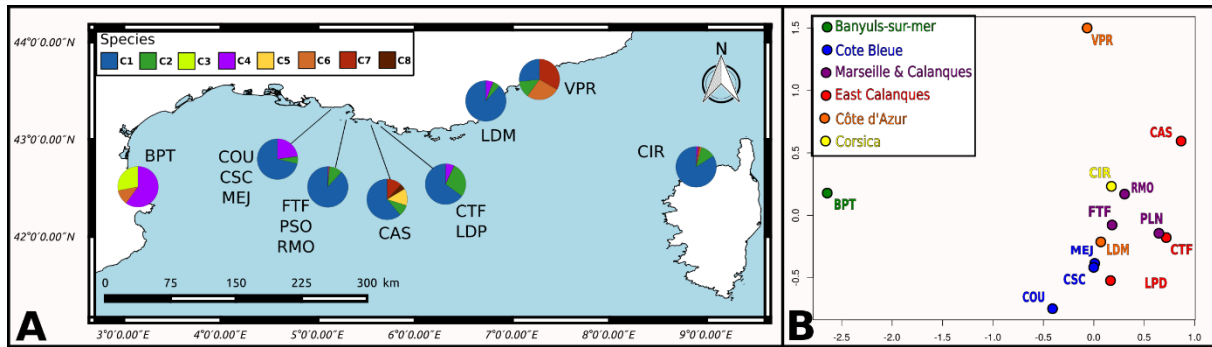
*Figure 2 Distribution of the eight cryptic species along the French Mediterranean coastline. **A:** Map of the study area with pie chart representing the relative abundances of the cryptic species at the different localities. **B:** nMDS on Bray-Curtis distances between localities at large spatial scale: from Banyuls-sur-mer to Villefranche-sur-mer and including Corsica. Stress: 0.04*
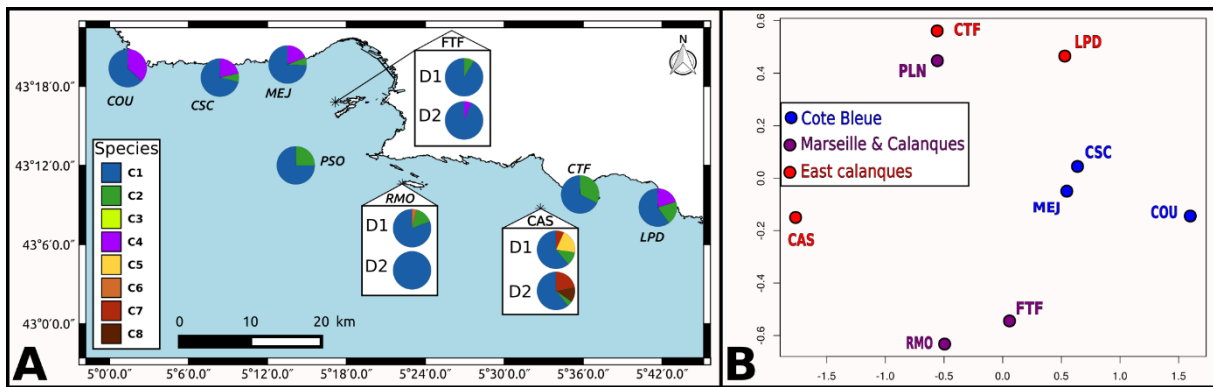
*Figure 3 Distribution of the eight cryptic species around Marseille coastline. **A**: Map of the Marseille area with pie chart representing the relative abundances of the cryptic species at the different localities and different depth. D1: shallow depth category between 24 and 31 meters' depth. D2: deep depth category between 37-46 meters. **B**: nMDS on Bray-Curtis distances between localities at small spatial scale: from the COU locality to the LPD locality. Stress: 0.04*
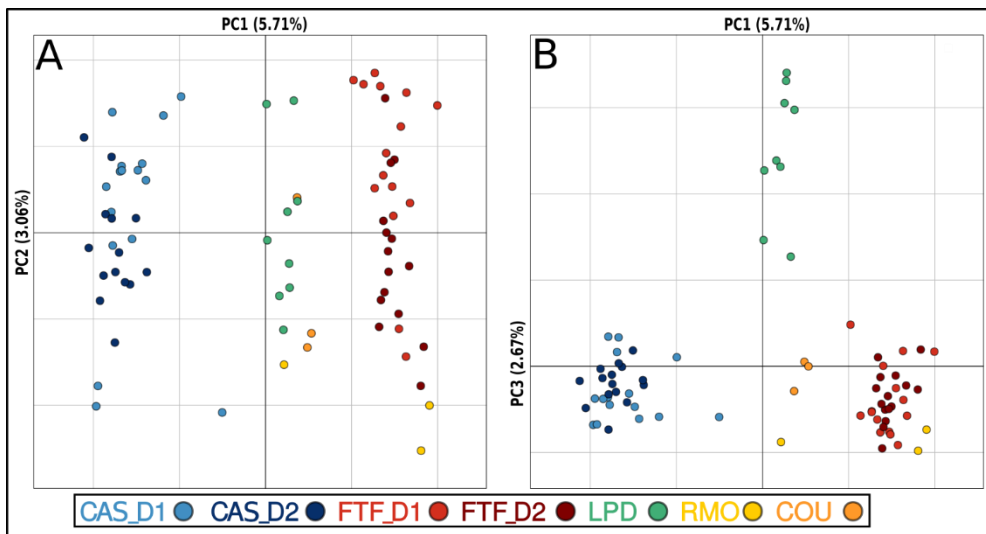


*Figure 4 Principal Component Analysis (PCA) on the multilocus genotype for individuals of the C1 cryptic species. Individuals are colored according to their population, and D1 and D2 in species names correspond to depth categories. **A**: first and second axes are drawn. **B**: first and third axes are drawn.*

Highlights

We found 8 cryptic species in *Lithophyllum stictiforme / L. cabiochiae*

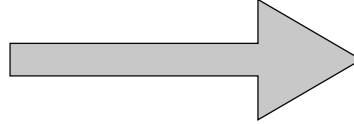Cryptic species relative abundances vary along the French Mediterranean coast

Cryptic species relative abundances vary among depth on the same locality

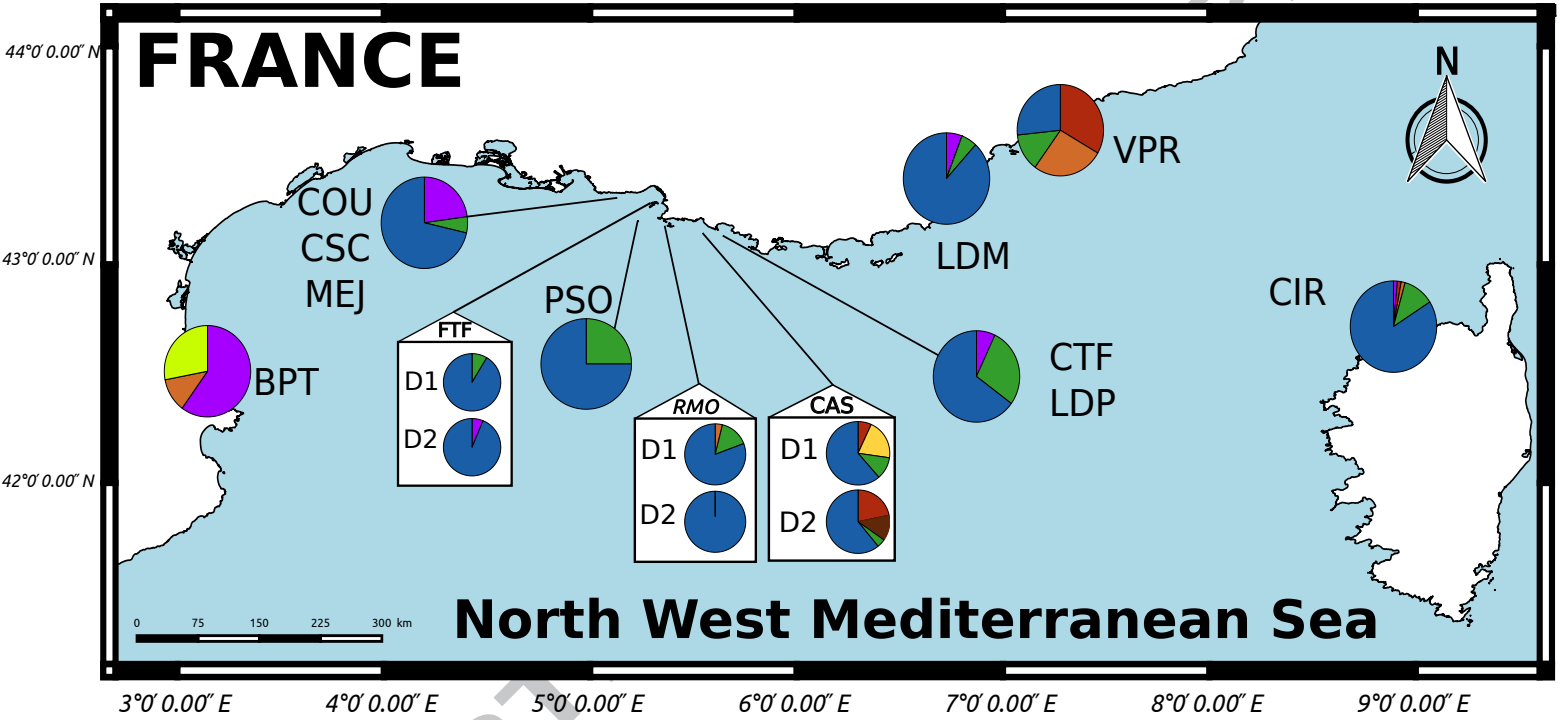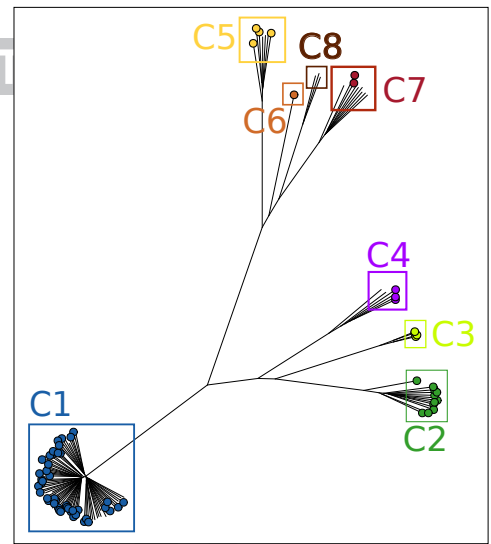No genetic differentiation among depths within locality in the most abundant species

Cryptic species composition is influenced by locality and depth