



HAL
open science

Genomic signatures of clonality in the deep water kelp *Laminaria rodriguezii* Genomic signatures of clonality

Lauric Reynes, Thierry Thibaut, Stéphane Mauger, Aurelie Blanfuné, Florian Holon, Corinne Cruaud, Arnaud Couloux, Myriam Valero, Didier Aurelle

► To cite this version:

Lauric Reynes, Thierry Thibaut, Stéphane Mauger, Aurelie Blanfuné, Florian Holon, et al.. Genomic signatures of clonality in the deep water kelp *Laminaria rodriguezii* Genomic signatures of clonality. *Molecular Ecology*, 2021, 10.1111/mec.15860 . hal-03159657

HAL Id: hal-03159657

<https://hal.science/hal-03159657>

Submitted on 4 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Genomic signatures of clonality in the deep water kelp *Laminaria***
2 ***rodriguezii***

3 **Genomic signatures of clonality**

4 Lauric Reynes¹, Thierry Thibaut¹, Stéphane Mauger², Aurélie Blanfuné¹, Florian
5 Holon³, Corinne Cruaud⁴, Arnaud Couloux⁵, Myriam Valero², Didier Aurelle^{1,6}

6 *lauric.reynes@mio.osupytheas.fr

7 (1) Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France

8 (2) IRL 3614, Evolutionary Biology and Ecology of Algae, CNRS, Sorbonne Université, UC, UACH, CS 90074,
9 Place Georges Teissier, 29688 Roscoff, France

10 (3) Andromède Océanologie, Montpellier, France

11 (4) Genoscope, Institut de Biologie François-Jacob, Commissariat à l'Energie Atomique (CEA),
12 Université Paris-Saclay, F-91057 Evry, France.

13 (5) Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry,
14 Université Paris-Saclay, 91057 Evry, France

15 (6) Institut de Systématique Évolution Biodiversité (ISYEB, UMR 7205), Muséum National d'Histoire
16 Naturelle, CNRS, Sorbonne Université, EPHE, 57 rue Cuvier, CP 51, 75005 Paris, France

17

18

19

20

21

22

23

24

25

26 **ABSTRACT**

27 **Partial clonality, mode of reproduction, heterozygote excess, population genomics, kelp**

28 The development of population genomic approaches in non-model species allows for renewed
29 studies of the impact of reproductive systems and genetic drift on population diversity. Here,
30 we investigate the genomic signatures of partial clonality in the deep water kelp *Laminaria*
31 *rodriguezii*, known to reproduce by both sexual and asexual means. We compared these
32 results with the species *Laminaria digitata*, a closely related species that differs by different
33 traits, in particular its reproductive mode (no clonal reproduction). We analysed genome-wide
34 variation with dd-RAD sequencing using 4077 SNPs in *L. rodriguezii* and 7364 SNPs in *L.*
35 *digitata*. As predicted for partially clonal populations, we show that the distribution of F_{IS}
36 within populations of *L. rodriguezii* is shifted toward negative values, with a high number of
37 loci showing heterozygote excess. This finding is the opposite of what we observed within
38 sexual populations of *L. digitata*, characterized by a generalized deficit in heterozygotes.
39 Furthermore, we observed distinct distributions of F_{IS} among populations of *L. rodriguezii*,
40 which is congruent with the predictions of theoretical models for different levels of clonality
41 and genetic drift. These findings highlight that the empirical distribution of F_{IS} is a promising
42 feature for the genomic study of asexuality in natural populations. Our results also show that
43 the populations of *L. rodriguezii* analysed here are genetically differentiated and probably
44 isolated. Our study provides a conceptual framework to investigate partial clonality on the
45 basis of RAD-sequencing SNPs. These results could be obtained without any reference
46 genome, and are therefore of interest for various non-model species.

47

48

49

50

51 INTRODUCTION

52 Partial clonality, defined as a mode of reproduction where species reproduce both sexually
53 and asexually, occurs in a wide variety of organisms across the tree of life, and is particularly
54 widespread in fungi, plants, rotifers, cladocerans and insects (Halkett et al., 2005). Clonal
55 reproduction can occur through different mechanisms such as fragmentation, vegetative
56 reproduction, mitotic parthenogenesis, or some forms of meiotic parthenogenesis (De Meeûs
57 et al., 2007; Jaron et al., 2020). Partial clonality has attracted considerable attention among
58 evolutionary biologists due to the potential effects of asexuality on the evolutionary trajectory
59 of species since reproductive modes have a major impact on the level of genome-wide genetic
60 diversity and its distribution within and among populations (Halkett et al., 2005; Duminil et
61 al., 2007). Moreover, reproductive modes profoundly affect other biological traits that play
62 important roles in colonization processes and the spread of populations, such as population
63 growth and dispersal (Kettenring & Mock, 2012; González de León et al., 2016), and in
64 buffering and resilience after catastrophic events (Becheler et al., 2020).

65 Theoretical models predict that when clonality predominates, it will increase heterozygosity
66 (Marshall & Weir, 1979; Balloux et al., 2003; De Meeûs et al., 2006), which may ultimately lead
67 to higher levels of observed heterozygosity (H_o) than those expected (H_e) under panmixia. This
68 deviation has been observed empirically through negative values of fixation index F_{IS} , reported
69 in various partially clonal populations (in animals Adjeroud et al., 2014; Halkett et al., 2005;
70 angiosperms Alberto et al., 2002; Alberto et al., 2005; Arnaud-Haond et al., 2007, and red and
71 brown macroalgae Guillemin et al., 2008; Ardehed et al., 2015; Coleman & Wernberg, 2018,
72 Pardo et al., 2019). Without sex, alleles do not segregate independently, which implies that
73 heterozygosity can be preserved rather than reduced over generations (Judson & Normark,
74 1996). Increasing clonal reproduction can increase effective population size at the allelic scale
75 but it decreases the genotypic effective population size: as a consequence asexual populations
76 can maintain higher allele diversity but lower genotype diversity compared to sexual ones
77 (Balloux et al., 2003). Asexual reproduction also leads to specific patterns of the distribution
78 of F_{IS} over the genome, characterized by a high occurrence of extreme, mainly negative, values
79 (Stoeckel & Masson, 2014; Reichel et al., 2016). These last theoretical studies reported, among
80 other points, that high rates of asexuality increased the variance of the distribution of F_{IS}
81 compared to that obtained from fully sexual populations (Stoeckel & Masson, 2014) but also

82 as compared to fully clonal populations (De Meeûs et al. 2006). Some variation in F_{IS} among
83 loci can also indicate the occurrence of rare recombination events as in the automictic
84 parthenogenetic species *Artemia parthenogenetica* (Nougué et al., 2015). In addition, if
85 asexuality is stable over generations, the accumulation of mutations in different alleles can
86 increase sequence differences among alleles, which corresponds to the "Meselson effect"
87 (Birky 1996; Judson & Normark, 1996; Balloux et al., 2003; De Meeûs et al., 2007). This effect
88 has been empirically demonstrated in ancient asexual *Timema* stick-insects (Schwander et al.,
89 2011) and in the human pathogen *Trypanosoma brucei gambiense* (Koffi et al., 2009; Weir et
90 al., 2016) and *Candida albicans* (Nébavi et al., 2006). Conversely the long term persistence
91 without sex of Bdelloid rotifers, referred as asexual scandals, has been challenged by the
92 suggestion of non-canonical sex (Schwander, 2016).

93 In this paper, we investigated how population genomic studies of partially clonal species allow
94 analysis of both the rate of clonality and its consequences for the distribution of F_{IS} . With the
95 development of thousands of Single Nucleotide Polymorphism (SNP), we will test whether the
96 observed empirical distribution of F_{IS} follows the aforementioned predicted variations in F_{IS}
97 among loci. The development of reduced representation sequencing methods (see for review
98 Davey et al., 2011) now allows such investigations in non-model species. These approaches
99 are also useful to more precisely delineate clonal lineages, where microsatellite loci have
100 failed to detect any genetic variation, as in the case of the selfing species *Sargassum muticum*
101 in the introduction range (Le Cam et al., 2019). Among the diversity of reduced representation
102 sequencing methods, Restriction Site Associated DNA sequencing (RAD sequencing; Baird et
103 al., 2008) is useful to efficiently develop enough SNPs to study the genomic distribution of
104 population genetic statistics.

105 We used the kelp *Laminaria rodriguezii* as a model species for the study of partial clonality.
106 *Laminaria rodriguezii* is a deep water kelp endemic to the Mediterranean Sea, mainly reported
107 from depths below 70 m, rarely found in shallower waters on seamounts or in upwelling
108 systems (Ballesteros, 2006), and with a maximum depth recorded at 260 m in the Adriatic Sea
109 (Ercegović, 1960). The rare populations of this species are exposed to various levels of
110 disturbances, such as direct destruction by fishing gear and a decline in water transparency,
111 which has led to *L. rodriguezii* being classified as endangered (Zuljevic et al., 2016). Besides
112 sexual reproduction, *L. rodriguezii* can reproduce asexually by stoloniferous or rhizoidal
113 growth. This mode of asexual reproduction involves somatic cells and should lead to clonal

114 lineages. Vegetative reproduction is only known in two *Laminaria* species: *L. rodriguezii* (Huvé,
115 1955) and the northeast Pacific *L. sinclairii* (Demes & Graham, 2011). *Laminaria rodriguezii*
116 can also reproduce sexually with a haplo-diplobiontic life cycle involving the alternation of
117 microscopic dioicous unisexual haploid gametophytes with large diploid sporophytes typically
118 found in kelps (Bringloe et al., 2020). Even if we do not know the importance of self
119 incompatibility in this species, self fertilisation (i.e. fertilisation between male and female
120 gametophytes originating from a single sporophyte parent), was reported to be possible
121 experimentally in most studied species of Laminariales (see the recent review of Goecke et
122 al., 2020). *Laminaria rodriguezii* then presents a mixed reproductive system, for which the
123 genetic consequences have not been studied. This deep water species is also characterized by
124 small and highly fragmented populations (Araújo et al., 2016). Therefore, the populations of
125 *L. rodriguezii* should be impacted by a reduction in genetic polymorphism through drift, a
126 reduction that can be counterbalanced by the aforementioned effects of asexual reproduction
127 on allele diversity (Balloux et al., 2003).

128 We specifically studied here whether the levels of asexual reproduction have discernable
129 effects on the empirical genomic distribution of diversity and F_{IS} among loci. We included
130 populations of the fully sexual congeneric species, *L. digitata*, with different levels of
131 microsatellite genetic diversity (Liesner et al., 2020) to take into account the consequence of
132 the mode reproduction with the confounding effect of variation in the strength of genetic drift
133 in our interpretations. *Laminaria digitata* is analysed only for comparative purposes, and we
134 did not develop the study of the genetic structure of this species in this article. We also
135 interpret our results in the light of previous theoretical studies, especially considering the full
136 distribution of F_{IS} among loci (Stoeckel & Masson, 2014). Among other factors, we tested
137 whether the clonal rates in populations of *L. rodriguezii* lead to a shift toward highly negative
138 F_{IS} values among loci, as expected for the highest levels of clonality (Stoeckel & Masson, 2014).
139 We used a large set of SNPs in order to access the most complete distribution of F_{IS} values over
140 populations of genomes. We also evaluated the robustness of our results by considering
141 different cutoffs in the filtering of SNPs. The implications of the inferred genetic diversity and
142 genetic structure for the management and conservation of this rare deep water kelp *L.*
143 *rodriguezii* are discussed.

144

145 MATERIAL AND METHODS

146 Sampling, DNA extraction

147 Individuals of *L. rodriguezii* were sampled from four sites in the Mediterranean, three in
148 Eastern Provence (Banc Magaud_1, Banc Magaud_2, Cap Camarat) and one in Southern
149 Corsica (Bonifacio), between 65 to 76 meters depth (Figure 1). Sampling was done by scuba
150 divers between June and August 2018. Sampling for DNA collection was performed by
151 collecting a small piece of tissue from the blade of sporophytes. More precisely, a total of 47
152 samples were collected from sporophytes not connected to each other by a stolon, and with
153 a minimum distance of two meters between sampled blades. This strategy aimed to promote
154 the sampling of distinct genets. For one of the two Banc du Magaud sites, namely Banc
155 Magaud_2, we used a different sampling design: the spatial distance between the sampled
156 sporophytes was recorded along a regular transect of 20 meters to study the spatial genetic
157 structure inside the population. Regarding *L. digitata*, we used samples from five northeastern
158 Atlantic populations of *L. digitata* collected in 2018 (Table S1, Supporting Information).
159 Genomic DNA was extracted using the Nucleospin® 96 plant kit (Macherey-Nagel, Düren,
160 Germany), according to the manufacturer's protocol.

161 ddRAD-sequencing

162 Two double-digest RAD-sequencing libraries (ddRAD-seq) with 47 individuals of *L. rodriguezii*
163 and 125 individuals of *L. digitata* were prepared according to Peterson et al., (2012). To
164 minimize bias in library preparation, all DNA samples were randomized across libraries, and
165 two and five replicates of *L. rodriguezii* and *L. digitata*, respectively, were included. DNA
166 concentrations were estimated with PicoGreen (Invitrogen, Carlsbad, CA, USA), and 100 ng of
167 genomic DNA was double digested using PstI and HhaI (NEB) for 12 hours at 37°C. The
168 purification of digested DNA was conducted by the NucleoMag NGS clean-up and size select
169 kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. When these
170 steps were completed, the ligation of sequencing adapters to restriction digested DNA was
171 performed as follows: P1 adapters (overhang PstI), including unique barcodes (6-12 bp) and
172 P2 universal adaptor (overhang HhaI) were ligated to DNA using 0.5 µl (400 U/µL) of T4 DNA
173 ligase (NEB), 6 µl of 10X T4 ligase buffer (NEB) and incubated for 12 hours at room

174 temperature. Before PCR amplification, an additional purification step was done with the
175 NucleoMag NGS clean-up and size select kit. Amplification by PCR was restricted to 15 cycles
176 and performed with the Q5 hot Start High-Fidelity DNA polymerase kit (NEB). Then, samples
177 were multiplexed, and automated size selected for a range between 400 and 900 bp using a
178 Pippin Prep (Sage Science, Beverly, MA, USA). Finally, the libraries were sequenced with the
179 paired-end method (2×150 bp) on an Illumina Hiseq 4000 platform (G enome Qu ebec
180 Innovation Centre, McGill Univ., Montreal, Canada).

181 **Genotyping and SNP filtering**

182 Raw quality reads were checked using FastQC v0.11.7 (Andrews, 2010) and trimmed to 137
183 bp after the removal of adapters sequencing by Trimmomatic (Bolger et al., 2014). Then,
184 paired-reads were demultiplexed by both index and barcode using Stacks's process_radtags
185 (Catchen et al., 2011). Thanks to the phylogenetic proximity between *L. rodriguezii* and *L.*
186 *digitata* (Žuljevi c et al., 2016), we were able to use the draft genome of *L. digitata*
187 (unpublished data, Mark Cock) to align our reads with the BWA-mem algorithm in BWA
188 v0.7.17 (Li & Durbin, 2009). Aligned reads were then assembled into loci and finally, we called
189 single nucleotide polymorphisms (SNPs) with the Stacks v2.4 pipeline (Rochette et al., 2019).
190 Aligned reads of mapping quality below 30 (which means a probability of incorrect alignment
191 below 0.001) were excluded to minimize errors resulting from the erroneous assembly of
192 paralogous loci as a single locus. To avoid a strong linkage between SNPs, we kept only the
193 first SNP of each RAD locus. Further filtering was performed on the basis of minor allele
194 frequency, missing data rates for both individuals (i.e the percentage of missing genotypes per
195 individual) and loci (the percentage of non-genotyped individuals per locus), using VCFtools
196 v0.1.17 (Danecek et al., 2011) (Table 1). We aimed to identify the signatures of clonality from
197 putatively neutral loci. To this end, we excluded outlier loci identified as involved under
198 divergent selection with *pcadapt* v3.5 (Luu et al., 2017). *Pcadapt* starts by analysing population
199 structure with a PCA. Then candidate outliers are identified as those which are excessively
200 correlated with population structure (Luu et al., 2017). The outlier identified with *pcadapt*
201 should then correspond to inflated signals of genetic differentiation.

202 **The impact of SNP filtering on population genetic summary statistics**

203 RAD-seq is subject to different sources of genotyping errors (Mastretta-Yanes et al., 2015) that
204 may generate spurious inferences in surveys of genetic variation. However, the noise of
205 artifactual SNPs can be efficiently overcome by applying a rigorous filtering strategy of raw
206 RAD-seq datasets (e.g. McKinney et al., 2017; O'Leary et al., 2018) and measuring its effects
207 on population genetic summary statistics (Shafer et al., 2017). Care was taken in this study to
208 minimize the impact of artifactual SNPs on the true biological signal of the datasets (see
209 Supporting Information "Materials S1"). To achieve this, we studied the impact of different
210 filtering cutoffs on several parameters of genetic diversity and departure from panmixia: i) the
211 mean F_{IS} and standard deviation (SD) among loci, ii) the overall F_{ST} and SD among loci, iii) the
212 SNP error rate, defined as the proportion of SNP mismatches between replicate pairs
213 (Mastretta-Yanes et al., 2015). We have specifically explored the effect of the maximum rate
214 of missing data allowed per SNP (max-missing), the minor allele frequency (MAF), and the
215 mean read depth per locus across all individuals (meanDP), or by considering only loci for
216 which genotypes were strictly identical between replicates. Further details are given in
217 Supporting Information "Materials S1".

218 **Clonality in *Laminaria rodriguezii***

219 Because of somatic mutations and sequencing errors, we do not expect to get two strictly
220 identical genotypes over thousands of loci even if they belong to the same clonal lineage.
221 Consequently, we considered that genotypes with very low divergence corresponded to the
222 same clonal lineage. We chose the maximum distance between replicates ($d = 0.022$ for *L.*
223 *rodriguezii*, $d = 0.020$ for *L. digitata*, see Results) as a threshold to identify repeated genotypes.
224 Distances were computed as the ratio of the number of allelic differences between two
225 individuals (the Hamming distance, implemented in the function *diss.dist* of the Poppr v2.8.2
226 R package). Samples of *L. rodriguezii* separated by a distance lower than the aforementioned
227 threshold were clustered into clonal lineages (referred here as MLLs Multi-Locus Lineages, see
228 Arnaud-Haond et al., 2007) using the UPGMA clustering algorithm (implemented in Poppr
229 v2.8.2 R packages: Kamvar et al., 2014; Kamvar et al., 2015). The matrix of genetic distances
230 was additionally used for the construction of a network with the neighbor-net method
231 implemented in SplitsTree, which was used to visualize genetic proximity of samples from the
232 same clonal lineages, and also the differences among populations (Huson, 1998). We

233 computed the genotypic richness (R) index within populations (Dorken & Eckert, 2001), as
234 follows: $R = \frac{G-1}{N-1}$, where G is the number of distinct genotypes detected in the population and
235 N the total number of individuals sampled. We thus considered each cluster of clonal
236 individuals as a single genotype to calculate G. The distribution of clonal membership was
237 analysed with the Pareto β index as described in Arnaud-Haond et al. (2007) using the Poppr
238 v2.8.2 R package. Finally, the average values of genetic diversity and population differentiation
239 of *L. rodriguezii* were estimated in two complementary ways: firstly, including all genotyped
240 individuals, secondly by keeping only one representative of each clonal lineage.

241 **Genetic diversity**

242 The within-population genetic diversity was assessed with the datasets of *L. rodriguezii* and *L.*
243 *digitata*, after the removal of potentially selected loci. Given that the percentage of
244 polymorphic loci (%P) may be affected by the variations in sample sizes, we used a random
245 sampling approach implemented in a custom bash script (available at
246 <https://github.com/rebecca-cj/revegetation>). The percentage of polymorphic loci was
247 calculated from 100 random resamplings (with replacement between samples) for each
248 sample size at the population level. In addition, estimates of observed heterozygosity (H_o),
249 expected heterozygosity (H_e), and F_{IS} (Weir & Cockerham, 1984) were calculated with the
250 GENEPOP v1.0.5 R package (Rousset, 2008). The same data sets were also analyzed with the
251 HierFstat R package (Goudet, 2005) to compute the distribution of F_{IS} values across loci. We
252 used a two-sample Kolmogorov-Smirnov (KS) test to test for differences in the distribution of
253 F_{IS} between populations. We used the KS D statistic as the maximum distance between
254 cumulative distributions to assess the magnitude of these differences. The P-value of the D
255 statistic was computed using a bootstrapping method (10^6 bootstraps), with the Matching
256 v4.9.7 R package (Sekhon, 2008). The distributions were divided into ten discrete classes of F_{IS}
257 for which the frequency of occurrence in populations was shown as histograms. The Global
258 Hardy-Weinberg test [Score (U) test] (Rousset & Raymond, 1995), implemented in GENEPOP,
259 was used to test departure from panmixia across all loci for the different populations. The test
260 was performed twice: firstly by considering the alternative hypothesis of heterozygote excess
261 and secondly by considering heterozygote deficiency. The P-value of each test was
262 approximated by Markov chain (MCMC) algorithm with the following settings;

263 dememorization: 10 000, batch: 100, iterations per batch: 10 000. Then, a multisample score
264 test (Rousset & Raymond, 1995) was performed to obtain a global P-value per population.
265 Finally, the two-sided P-value of the Global Hardy-Weinberg test [Score (U) test] was obtained
266 by computing the following formula: $P_{\min} + (1 - P_{\max})$ where P_{\min} and P_{\max} are the smaller
267 and the higher probabilities of unilateral tests, respectively (De Meeûs, 2012). Linkage
268 disequilibrium (LD) among loci was analyzed with the standardized index of multilocus
269 association $\overline{r_d}$ (Agapow & Burt, 2001) in Poppr v2.8.2. The $\overline{r_d}$ index is a summary statistic of
270 linkage disequilibrium based on the variance of pairwise distances among individuals (see
271 details in Agapow & Burt, 2001). This index is particularly useful for datasets with a high
272 number of loci for which pairwise approaches are difficult to apply. To avoid the consequences
273 of a Wahlund effect, we computed the $\overline{r_d}$ index separately for each sampling site. Genotypes
274 were permuted 1000 times among individuals inside populations to test for linkage
275 disequilibrium.

276 **Population differentiation in *Laminaria rodriguezii***

277 The overall level of genetic differentiation (overall F_{ST}) was estimated in Genepop v4.2.2. We
278 performed a G-based test of genotypic differentiation across loci (Goudet et al., 1996) with
279 the same software. The P-value for each locus was computed using the Markov chain method
280 with default parameters (500000 iterations). The significance of the test was evaluated by the
281 method of the harmonic mean P-values (HMP) (Wilson, 2019), implemented in the
282 `harmonicmeanp` v3 R package. The test compares P_R , the HMP of a set of P-values R , to α_L ,
283 the significance threshold for a false positive rate weighted by ω_R , the total number of
284 individual P-values (see Table 1 in Wilson, 2019). The HMP P-value is significant when $P_R \leq$
285 $\alpha_L \omega_R$. We used a false positive rate of 0.05 and an equal weight for each P-value, considered
286 as similar to Bonferroni correction.

287 **RESULTS**

288 **Sequencing and SNPs filtering**

289 Approximately 1.2 billion (1 154 828 457) raw reads were produced from 49 samples of *L.*
290 *rodriguezii* (i.e. 47 individuals and two replicates) and 130 samples of *L. digitata* (i.e. 125

291 individuals and five replicates). After quality filtering and read trimming, an average of 2.7
292 million (± 1.5 million SE) and 2.3 million (± 2.1 million SE) high-quality reads per individual were
293 retained for *L. rodriguezii* and *L. digitata*, respectively. The percentage of reads mapped to the
294 *L. digitata* genome, ranged from an average of 85.4% (min: 53.9%; max: 92.2%) for *L.*
295 *rodriguezii*, to 94.2% (min: 67.9%; max: 97.6%) for *L. digitata*. The initial filtering of loci, which
296 retained those shared by a high proportion of individuals, resulted in catalogs of 44 716 loci
297 and 60 942 loci for *L. rodriguezii* and *L. digitata*, respectively. At this stage, the number of SNPs
298 was more than eight times higher for *L. digitata* than for *L. rodriguezii* (Table 1). Additional
299 quality filtering steps followed by outlier exclusion resulted in the selection of 4077 putatively
300 neutral SNPs among 43 individuals of *L. rodriguezii* and 7364 putatively neutral SNPs among
301 116 individuals of *L. digitata* (Table 1). Specifically, 289 SNPs (6.62%) and 174 SNPs (2.31%)
302 were identified as outliers for *L. rodriguezii* and *L. digitata*, respectively. The rate of missing
303 data was slightly higher in *L. digitata* (i.e. 3.36%) than in *L. rodriguezii* (i.e. 2.24%). The mean
304 SNP error rate across replicates was estimated to be 1.06% in *L. digitata* and 3.89% in *L.*
305 *rodriguezii* datasets.

306 **Clonal structure of *Laminaria rodriguezii***

307 In *L. rodriguezii*, the distribution of pairwise genetic distances was discontinuous with a gap
308 from the threshold $d = 0.022$ to the lowest distance reported above this threshold ($d = 0.042$)
309 (Figure 2A). This threshold corresponded to the maximum distance observed among
310 replicates. The use of this threshold allowed us to identify four different clonal lineages with
311 several members: namely MLL-A, MLL-B, MLL-C and MLL-D, composed of three, three, seven,
312 two samples respectively (see the network, Figure 3). By selecting only one sample for each
313 clonal lineage, we finally retained 32 different samples in our data analyses from the 43
314 samples initially present in the whole dataset. At Bonifacio (Southern Corsica), we reported a
315 relatively strong occurrence of clonal lineages ($R = 0.20$), with the dominance of a few large
316 clones ($\beta = 0.56$), which contrasted with the low occurrence of clonal lineages ($R = 0.90$) of
317 small size ($\beta = 2.66$) in Eastern Provence (Table S2, Supporting Information). For instance, the
318 maximum clonal size reported was seven at Bonifacio against three in Eastern Provence. The
319 analysis of the spatial genetic structure along the transect at Banc Magaud_2 (Figure S1,
320 Supporting Information) revealed that: i) individuals belonging to the same clonal lineage
321 were relatively close to one another (less than one meter) and ii) the genetic distance between

322 individuals increased with spatial distance (Mantel test, $R = 0.40$, $p < 0.001$). For the other
323 sites, the sampling protocol (with a minimum spacing of two meters), did not allow a similar
324 investigation of the spatial genetic structure. Contrary to *L. rodriguezii*, no pairwise distance
325 among distinct individuals of *L. digitata* was below the distance threshold defined by
326 replicates (Figure 2B), indicating that there was no cluster of genotypes grouped in clonal
327 lineages. The lowest divergence among individuals in *L. digitata* was around $d = 0.035$, and
328 corresponded to intra-population differences within the Helgoland population.

329 Besides the detection of clonal lineages, we observed in *L. rodriguezii* three individuals, with
330 distinct long branches in the phylogenetic networks (indicated as outliers in Figure 3). These
331 three outlier individuals shared a common characteristic: the highest proportion of
332 heterozygous genotypes among loci per individual, with an average of 33% against 16% in
333 other individuals (Figure S2, Supporting Information). These outlier individuals also had the
334 highest percentage of private alleles among all individuals (average percentage of private
335 alleles at the individual level (Pr) = 20.3% for outlier individuals and $Pr = 1.0\%$ among the other
336 40 genotyped individuals, see Figure S2, Supporting Information). By definition, these private
337 alleles were different among the three individuals.

338 **Genetic diversity**

339 *Laminaria rodriguezii* and *Laminaria digitata* displayed similar expected heterozygosity (H_e),
340 with an average H_e within populations of 0.14 and 0.15, respectively (Table 2). The proportion
341 of polymorphic loci (P) within populations (estimated with the minimum sampling size $n = 10$
342 and 100 random resamples) was higher in *L. rodriguezii* (average $P = 54.42\%$) than in *L. digitata*
343 (average $P = 39.39\%$) (see Table 2, Figure S3, Supporting Information). We observed a
344 departure from panmixia in both species, with an excess of heterozygotes in *L. rodriguezii*, and
345 a deficit of heterozygotes in *L. digitata* (see Table 3 for *L. rodriguezii* and Table S1, Supporting
346 Information for *L. digitata*). The heterozygotes excess was statistically significant (Score [U]
347 test, $P < 0.001$) within populations for *L. rodriguezii*, except for Banc Magaud_1. The
348 heterozygotes deficit was significant (Score [U] test, $P < 0.001$) within populations of *L.*
349 *digitata*.

350 The distributions of F_{IS} were significantly different for all intra and inter-specific comparisons
351 of *L. rodriguezii* and *L. digitata* populations (KS tests, $p < 0.001$, D statistic: 0.07-0.53). At the
352 intraspecific level, the most striking difference in the distributions of F_{IS} (KS tests, D statistic:

353 0.26-0.27) was detected for *L. rodriguezii* between Eastern Provence (Banc Magaud_1,
354 Banc_Magaud_2 and Cap Camarat) and Bonifacio (Figure 4A). The distributions of F_{IS} in
355 Eastern Provence were moderately skewed to negative values with a long tail to positive
356 values (skewness = 0.69; kurtosis = 4.07). The distribution of F_{IS} in Bonifacio was highly skewed
357 toward negative values (skewness = -0.95; kurtosis = 3.14) and showed a much higher
358 proportion of extreme negative values as above to -0.8 (13.6%) in comparison to those
359 measured in Eastern Provence (average, 0.6%). Similarly, the variance in F_{IS} among loci was
360 higher at Bonifacio (var = 0.13) than in Eastern Provence (var = 0.08, 0.07 and 0.06 for Banc
361 Magaud_1, Banc Magaud_2, and Cap Camarat, respectively).

362 In contrast to these results, F_{IS} values were mainly positive in *L. digitata*, whatever the
363 population considered even if the distribution mode of F_{IS} is always observed between -0.1
364 and 0 (Figure 4B). Extreme negative F_{IS} (above -0.8) and even moderate to high negative values
365 (those between -0.5 and -0.3) were almost absent in *L. digitata* with frequencies of 0.03% and
366 0.06%, respectively. The distribution mode of F_{IS} was observed for slightly negative values
367 between -0.1 and 0 whatever the population considered (Figure 4B). The asymmetry of the
368 distributions, as revealed by the positive skewness (skewness = 0.98; kurtosis = 3.32) is
369 indicative of a long tail toward high and extreme positive values (0.5 to 1). This class of
370 extreme positive F_{IS} corresponded to 28% of all loci in Helgoland. At the intra-specific level,
371 the second major difference in the shape of the distribution of F_{IS} was reported between
372 Helgoland and all other populations of *L. digitata* (KS tests, D statistic 0.11-0.23).

373 The average multilocus estimates of F_{IS} within populations ranged from 0.12 to 0.43 for *L.*
374 *digitata* (Table S1, Supporting Information) whereas mean values were mainly negative within
375 populations in *L. rodriguezii*, with a range from -0.47 in Bonifacio to -0.02 at Banc Magaud_1
376 (Table 3A). After applying the correction for repeated genotypes, the average multilocus
377 estimates of F_{IS} increased to a range from $F_{IS} = -0.26$ at Bonifacio to $F_{IS} = -0.01$ at Banc
378 Magaud_2 (Table 3B). After the correction for repeated genotypes, the heterozygosity excess
379 only remained significant at Bonifacio (Score [U] test, $P < 0.001$).

380 The distribution of the index of multilocus association $\overline{r_d}$ (Figure 5) was nearly identical
381 whatever the number of SNPs subsampled (i.e. 1000; 2000 and 3000 random SNPs). The mean
382 $\overline{r_d}$ was statistically significant ($p < 0.001$) within all populations of both species. For *L.*
383 *rodriguezii*, the $\overline{r_d}$ index reached high levels within all populations (mean $\overline{r_d} = 0.08$ to 0.27),
384 except at Banc Magaud_1 (mean $\overline{r_d} = 0.01$) for which there was no repeated sampling of a

385 clonal lineage (Table 3). The values of $\overline{r_d}$ decreased slightly by keeping only one representative
386 of each clonal lineage (Table 3, Figure 5). For *L. digitata*, the mean $\overline{r_d}$ ranged between 0.01 to
387 0.03 within all populations (Table S1, Figure 5), except for Helgoland that displayed a
388 remarkably high level of linkage disequilibrium similar to values observed in *L. rodriguezii*
389 (mean $\overline{r_d} = 0.20$).

390 **Population differentiation in *Laminaria rodriguezii***

391 *Laminaria rodriguezii* exhibited high levels of population structure with a global F_{ST} of 0.28
392 when considering all individuals ($n = 43$). After applying the clone correction (i.e. one sample
393 per clonal lineage), the estimate of genetic differentiation decreased (global $F_{ST} = 0.18$). After
394 correcting for multiple testing, genotypic differentiation among populations of *L. rodriguezii*
395 was significant (HMP significance threshold, $\alpha_L \omega_R = 0.032$; HMP P-value, $P_R = 0$), with and
396 without the clone correction.

397 **Negative F_{IS} and high F_{ST} in *L. rodriguezii* over SNP filtering methods**

398 For *L. rodriguezii*, the mean F_{IS} and overall F_{ST} varied between the reference dataset and the
399 filtered datasets (Table 4). However, the mean F_{IS} was always negative, with high variance
400 among loci, whatever the levels of stringency. In the same way, the overall F_{ST} stayed high and
401 it was never below 0.27. The maximum rate of missing data allowed per SNP (max-missing)
402 had a very limited effect on mean F_{IS} and overall F_{ST} , which ranged from -0.06 to -0.07 and
403 from 0.27 to 0.30, respectively. The highest variance was associated with MAF cutoff as the
404 filter excluded the lowest polymorphic loci, and subsequently increased the level of genetic
405 variation in the dataset. Therefore, we choose to keep the MAF cutoff at 2% so as not to inflate
406 the levels of genetic variation in *L. rodriguezii*. The SNP error rate decreased from 5.22% in the
407 reference dataset to 2.74% in the dataset which was filtered by excluding loci with low
408 coverage (i.e. meanDP > 15X). Genotyping errors were probably non-randomly distributed and
409 occurred mostly in loci with low coverage, so we excluded them by applying meanDP > 15X.

410

411

412 **DISCUSSION**

413 We used RAD-sequencing to analyze the genome-wide diversity of the rare *Laminaria*
414 *rodriguezii*. We found that intra-individual genomic diversity is a powerful tool to identify
415 clonal lineages. We also proposed that the observed F_{IS} distribution and linkage disequilibrium
416 at the genome scale can be compared with theoretical models to search for a discernable
417 signal of clonality, even in this partially clonal reproducing species. Finally, we will discuss here
418 the findings of high heterozygosity within some individuals that raise questions about the
419 mechanisms allowing the persistence of such genomic features.

420 **RAD-seq as a powerful tool for detection of clonal lineages**

421 The identification of clonal lineages in natural populations was addressed by assessing the
422 probability that identical or slightly distinct genotypes belong to the same clonal lineage. That
423 can be achieved through the development of standardized methods allowing for investigation
424 of the influence of somatic mutations and scoring errors (Arnaud-Haond et al., 2007).
425 However, while this methodological framework is particularly robust for a small number of
426 loci, RAD-seq and the amount of data generated by Next Generation Sequencing (NGS) brings
427 new challenges. The main concern is that increasing the number of loci is equivalent to
428 increasing the number of missing data and genotyping errors that can subsequently affect the
429 detection of clonal lineages. This is especially the case for reduced representation sequencing
430 (e.g. RAD-seq) sensitive to missing data and genotyping error, thus necessitating appropriate
431 filtering strategy (e.g. Mastretta-Yanes et al., 2015; Shafer et al., 2017; Andrews et al., 2018;
432 Boscari et al., 2019). How then can we find an adequate distance threshold for collapsing
433 slightly distinct genotypes into clonal lineages? For that purpose, we can define a threshold
434 using the pairwise genetic distances, a threshold below which genotypes are assumed to
435 belong to the same clonal lineage. Some studies have fixed a cut-off based on the gap in
436 pairwise genetic distances (e.g. Shrestha et al., 2014), while others have defined an arbitrary
437 95% genotype similarity threshold (e.g. Locatelli & Drew, 2019). Nevertheless, such
438 approaches can be inaccurate, particularly when different sources of noise (i.e. genotyping
439 error, missing data, somatic mutation) can accentuate differences between clonal genotypes
440 above the detection threshold. This also depends on the genetic diversity of the populations:

441 clonal lineages are more easily identified in populations where the diversity among sexually
442 produced individuals is high, i.e. in partially clonal species. In this study, we proposed defining
443 a threshold on the basis of the pairwise genetic distance across replicates. While this approach
444 seems to be useful in evaluating the genetic distance expected as a result of genotyping errors
445 and somatic mutations, it has rarely been reported, except in *Leptopsammia pruvoti* (Boscari
446 et al., 2019). For this organism genotyped with 1386 SNPs (with the 2b-RAD protocol), slightly
447 distinct genotypes below a distance threshold set at $d = 0.043$ were collapsed into clonal
448 lineages. For *L. rodriguezii*, this cut-off was lower ($d = 0.022$), even if of a similar order as in *L.*
449 *pruvoti*. This difference highlights the importance of not considering a common threshold but
450 rather of adjusting it using genotype mismatch across replicates. Note that using the same
451 method on both species, the threshold was very similar between *L. rodriguezii* and *L. digitata*.

452 **The genomic patterns of clonality**

453 Our results show how the distribution of F_{IS} among loci can be used to study species and
454 populations with various signals of clonality including partial clonality. Signatures of clonality
455 in natural populations are usually inferred from repeated genotypes, along with mean
456 negative values of F_{IS} , and linkage disequilibrium across loci (Balloux et al., 2003; Halkett et al.,
457 2005). Nevertheless, recent studies reported that one of these signatures, genotypic richness,
458 had poor relevance for measuring the relative importance of sexual versus clonal reproduction
459 (Arnaud-Haond et al., 2020), even using relatively large sample sizes (from 100 to 500
460 individuals) (Stoeckel et al., 2019). This is because the estimates of genotypic richness are
461 strongly dependent on sampling strategy and density, basically decreasing with increasing
462 sampling effort (Arnaud-Haond et al., 2007; Gorospe et al., 2015; Becheler et al., 2017). In our
463 study, by looking at the F_{IS} distribution in the genome, we propose a new analysis which seems
464 useful to detect partial clonal reproduction. Our approach confirmed the prediction that inter-
465 locus variance of F_{IS} (rather than an average value) increased as clonal rates increased (Balloux
466 et al., 2003; Halkett et al., 2005; Stoeckel & Masson 2014; Reichel et al., 2016; Stoeckel et al.,
467 2019). We also reported (1) contrasted distribution of F_{IS} between species, with F_{IS} shifted
468 towards negative values in the partially clonal species *L. rodriguezii*, and towards positive
469 values in the fully sexually reproducing species *L. digitata*; (2) striking differences in
470 distributions of F_{IS} among populations of *L. rodriguezii*, with higher inter-locus variance in F_{IS}

471 at Bonifacio (Corsica) than in the three localities of Eastern Provence. Remarkably, the
472 empirical distributions of F_{IS} in Eastern Provence matched well with the simulation results of
473 Stoeckel & Masson (2014) which predicted that most F_{IS} values are concentrated between -
474 0.3 and 0.1 below intermediate levels of clonality.

475 Genetic drift is another possible explanation for negative F_{IS} values. In a finite population with
476 random mating, genetic drift due to sampling effect indeed tends to create a slightly negative
477 mean F_{IS} , but the effect is small, of order $1 / (2N)$ where N is the size of the population (see
478 Figures 3 and 4 in Stoeckel & Masson, 2014). For the fully sexual dioicous species *L. digitata*,
479 this drift effect could explain why the distribution mode of F_{IS} is between -0.1 and 0 whatever
480 the population considered. However, the probability of highly negative values increases when
481 genetic drift is combined with intermediate rates of clonality (Stoeckel & Masson, 2014).
482 Nevertheless, Stoeckel & Masson (2014) showed that values below -0.3 remain rare if the
483 clonality rate is below 0.9. This is indeed what is observed for all *L. rodriguezii* populations,
484 suggesting that their clonality rates are intermediate, except at Bonifacio, the Corsican
485 population for which we observed the highest proportion of F_{IS} below -0.3 and even below -
486 0.8. In addition, the estimates of genotypic richness (R), Pareto (β), and linkage disequilibrium
487 ($\overline{r_d}$) emphasized the degree of distinction between the Eastern Provence and Bonifacio
488 populations, thus strengthening our hypothesis of a higher level of clonality in Corsica than in
489 Eastern Provence. A more in-depth investigation of the combined effects of drift and clonal
490 reproduction on the distribution of F_{IS} would be useful in this context. Parameters other than
491 reproductive mode differ between the two species compared here, such as the levels of
492 genetic drift, the investigated geographical range, and potentially their demographic history.

493 The population of Helgoland for the fully sexually reproducing species, *L. digitata*, reached a
494 high level of linkage disequilibrium comparable to the Bonifacio population of *L. rodriguezii*
495 for which the signature of clonality was clearly discernible. Nevertheless, Helgoland had a
496 distribution of F_{IS} markedly different from *L. rodriguezii*, with a majority of positive values.
497 Helgoland is a rocky island surrounded by several hundreds of kilometers of sandy seafloor.
498 This population shows the highest levels of positive F_{IS} , the lowest values of pairwise genetic
499 distances between individuals, and the highest value of linkage genetic disequilibrium
500 observed for *L. digitata*. This is most probably explained by the effect of low effective size and
501 maybe of a bottleneck. This population also showed the lowest level of genetic diversity
502 observed for *L. digitata* when genotyped with the microsatellite locus in the study of Liesner

503 et al. (2020). There is a lack of case studies to conclude more precisely on the role of drift and
504 of demographic history in the distribution of inter locus F_{IS} variance. However, the theoretical
505 results of Stoeckel & Masson (2014) and Stoeckel et al. (2019) agree well with our
506 experimental results. This paves the way toward a better assessment of the importance of
507 clonality through genomic studies, but it remains the case that the estimation of clonality rates
508 is still only possible using temporal sampling (Ali et al., 2016; Becheler et al., 2017).

509 **Extreme levels of individual genomic heterozygosity**

510 We reported an intriguing signal of genome-wide heterozygosity in three of the 43 individuals
511 in *L. rodriguezii*. These outlier individuals were characterized by an exceptionally high
512 proportion of heterozygous genotypes compared to other individuals. Given the relatively low
513 sample sizes, we do not really know how these individuals depart from the population
514 distribution in individual heterozygosity, but we can discuss several hypotheses regarding
515 their origin. High heterozygosity at particular loci can correspond to paralogs assembled in the
516 same RAD locus depending on assembly parameters (Verdu et al., 2016; Mckinney et al., 2017;
517 Nadukkalam Ravindran et al., 2018). However, for the following reasons, we would argue that
518 polymorphisms detected within these outlier individuals do not correspond to artefactual loci:
519 firstly, sequencing reads were mapped to the reference genome of a closely related species,
520 followed by stringent filter criteria, including for maximum heterozygosity. Secondly, paired-
521 end sequencing protocols provide additional information on correct read placement since
522 RAD loci are retained when both reads are properly mapped (according to the insert length
523 between forward and reverse reads). Thirdly, if paralogs were responsible for high levels of
524 heterozygosity, we would not expect to observe this in only three individuals. On the other
525 hand, recent studies have revealed that both cross-contamination among samples and
526 environmental contamination, especially during library preparation and/or sequencing, can
527 lead to artifactual results, with the sequencing of loci of different origins in the same sample
528 (Laurin-Lemay et al., 2012; Ballenghien et al., 2017). Here, the three outlier individuals were
529 also characterised by the highest proportion of private alleles at the individual level, and they
530 were not grouped on the network. Such private alleles can then not be explained by cross-
531 contamination among samples of the same species. Rather they could be explained by the
532 accumulation of mutations along with clonal reproduction, thus increasing the divergence
533 among loci in the same individuals, a potential signal of the Meselson effect (Birky, 1996;

534 Welsh & Meselson, 2000; De Meeûs et al., 2007). These outlier individuals could then belong
535 to a relatively ancient asexual lineage. However, we lack temporal sampling to get more
536 precise estimates of the rates of clonal reproduction, and therefore, the hypothesis of
537 clonality to explain these outlier individuals remains highly speculative. A high frequency of
538 heterozygous loci could also be observed following hybridization among well differentiated
539 genetic lineages. Hybridization has indeed been proposed as the main driver of high
540 heterozygosity in parthenogenetic animals (Jaron et al., 2020). Here, the hybrid hypothesis
541 should involve several hybridizations from three different sources to explain the observed
542 pattern of private alleles, as well as the genetic distance among the three outliers. It would
543 therefore be interesting to analyze the genomic diversity of other *L. rodriguezii* populations in
544 the Mediterranean in order to go further on this topic. Nevertheless, the populations of *L.*
545 *rodriguezii* are scarce, and there are no other congeneric species in this area.

546 **Implications for conservation**

547 Our results emphasise that *L. rodriguezii* is spatially structured into genetically distinct
548 populations. At Banc Magaud_2 we also evidenced a significant spatial genetic structure:
549 regarding the low occurrence of clonality in this population, this structure is better explained
550 by reduced dispersal through haploid spores. Although high levels of population genetic
551 structure are not uncommon in kelps (e.g. Billot et al., 2003; Coleman et al., 2009; Durrant et
552 al., 2018), our results contrast with the lack of genetic differentiation among populations of *L.*
553 *digitata* spatially distant by up to 10 km (Billot et al., 2003; Robuchon et al., 2014). In this
554 respect, the spatial scale of dispersal of *L. rodriguezii* is presumably smaller than that reported
555 in other species of the genus *Laminaria* (Santelices, 1991). The patchy distribution and small
556 population sizes of *L. rodriguezii* can probably explain this pattern of high and local genetic
557 differentiation through strong local genetic drift and limited gene flow among populations.
558 The identification of such a local genetic structure has important implications for management
559 and conservation. The high genetic drift and limited gene flow can explain the lower genetic
560 diversity expected in *L. rodriguezii* compared to *L. digitata*, a species with a wider geographical
561 distribution, exhibiting semi-continuous and dense populations (see Hamrick & Godt, 1990;
562 Frankham, 1996; Levy et al., 2016). Furthermore, the range of *L. rodriguezii* has been
563 drastically reduced during the last half-century, especially in the Adriatic Sea, where the
564 species has suffered losses of 85% from its historical range (Žuljević et al., 2016). However, the

565 observed heterozygosity and the percentage of polymorphic loci were higher in *L. rodriguezii*
566 when compared to *L. digitata*. This can be explained by the conservation of heterozygous
567 genotypes with low recombination rates (Judson & Normark, 1996; Balloux et al., 2003; see
568 Meloni et al., 2013 for an empirical example), as genetic drift under clonal reproduction acts
569 on genotype frequencies rather than allele frequencies (Stoeckel & Masson, 2014; Reichel et
570 al., 2016). Further studies would be necessary to test with genomic data whether the current
571 fragmented range and small populations of *L. rodriguezii* correspond to a demographic
572 reduction from a larger ancestral population. The extensive cartography and monitoring of
573 the biocenoses down to 100 m along French coasts do point to the rarity of this species. The
574 conservation of these last isolated populations should become a high priority for stakeholders.

575 **Conclusion**

576 Our study shows that despite limitations on the detection of clonal lineages, SNPs derived
577 from RAD sequencing are interesting markers for population genomic studies of partially
578 clonal species. The interaction between demographic history and reproductive mode shaped
579 the distribution of F_{IS} and linkage disequilibrium in *L. rodriguezii*. Interestingly, even if our loci
580 were mapped to a reference genome for assembly, we did not use the physical location of loci
581 in our analyses, which indicates that similar approaches could be envisioned with *de novo*
582 assembly. In order to decipher the respective impact of demography and reproduction in our
583 results, one next step could be to use summary statistics linked with these distributions (such
584 as higher moments) for evolutionary inferences, using for example Approximate Bayesian
585 Computation (Csilléry et al., 2010).

586

587

588

589

590

591 **ACKNOWLEDGEMENTS**

592 This project was funded, by the MARFOR Biodiversa/0004/2015 project (<http://marfor.eu/>),
593 and by the Agence Nationale de la Recherche, Grant/ Award Number: ANR-18-CE32-0001
594 (Clonix2D). Lauric Reynes also received support from the GDR Génomique Environnementale
595 (GDR3692). The project leading to this publication has received funding from the European
596 FEDER Fund under project 1166-39417. This work was also funded by a Ph.D. Fellowship from
597 Région PACA and the Calanques National Park. The authors would like to thank Julie Deter, in
598 charge of the scientific coordination during sampling, the team of research divers of
599 Andromède Océanologie for the prowess in sampling *Laminaria rodriguezii* at 76 meters
600 depth, the Biogenouest genomics core facility (Genomer Plateforme génomique at the Station
601 Biologique de Roscoff) for their technical support, Claire Daguin-Thiébaud and Quentin
602 Langevin for useful discussions regarding the optimization processes in purpose-extracted
603 high-quality DNA in brown macroalgae, Marck Cock providing the draft genome of *Laminaria*
604 *digitata*, Thierry Gosselin for very useful discussions on the quality control of RADseq data,
605 and finally Solenn Stoeckel for his comments on the manuscript. We deeply thank Thierry De
606 Meeûs, three anonymous reviewers and the corresponding editor, Tatiana Giraud, for their
607 comments that enabled us to greatly improve the current version of this manuscript. Finally,
608 we would like to thank Michael Paul for English language proofreading.

609

610

611

612

613

614

615 **REFERENCES**

- 616 [dataset] Reynes, L., Thibaut, T., Mauger, S., Blanfuné, A., Holon, F., Cruaud, C., ... Aurelle, D.
 617 (2020). Individual high-quality reads (dd-RADsequencing) "Genomic signatures of
 618 clonality in the deep water kelp *Laminaria rodriguezii*". Dryad.
 619 doi:10.5061/dryad.hmgqnk9dq
- 620 Adjeroud, M., Guérécheau, A., Vidal-Dupiol, J., Flot, J.-F., Arnaud-Haond, S., & Bonhomme,
 621 F. (2014). Genetic diversity, clonality and connectivity in the scleractinian coral
 622 *Pocillopora damicornis*: a multi-scale analysis in an insular, fragmented reef system.
 623 *Marine Biology*, 161(3), 531–541. doi: 10.1007/s00227-013-2355-9
- 624 Agapow, P.-M., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. *Molecular*
 625 *Ecology Notes*, 1(1–2), 101–102. doi: 10.1046/j.1471-8278.2000.00014.x
- 626 Alberto, F., Gouveia, L., Arnaud-Haond, S., Perez-Llorens, J. L., Duarte, C. M., & Serrao, E. A.
 627 (2005). Within-population spatial genetic structure, neighbourhood size and clonal
 628 subrange in the seagrass *Cymodocea nodosa*. *Molecular Ecology*, 14(9), 2669–2681.
 629 doi: 10.1111/j.1365-294X.2005.02640.x
- 630 Ali, S., Soubeyrand, S., Gladieux, P., Giraud, T., Leconte, M., Gautier, A., ... Enjalbert, J. (2016).
 631 Cloncase: Estimation of sex frequency and effective population size by clonemate
 632 resampling in partially clonal organisms. *Molecular Ecology Resources*, 16(4), 845–
 633 861. doi: 10.1111/1755-0998.12511
- 634 Andrew, S. (2010). FastQC: a quality control tool for high throughput sequence data.
 635 (Version 0.11.7). Retrieved from
 636 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- 637 Andrews, K. R., Adams, J. R., Cassirer, E. F., Plowright, R. K., Gardner, C., Dwire, M., ... Waits,
 638 L. P. (2018). A bioinformatic pipeline for identifying informative SNP panels for
 639 parentage assignment from RADseq data. *Molecular Ecology Resources*, 18(6), 1263–
 640 1281. doi: 10.1111/1755-0998.12910
- 641 Araújo, R. M., Assis, J., Aguillar, R., Airoidi, L., Bárbara, I., Bartsch, I., ... Sousa-Pinto, I. (2016).
 642 Status, trends and drivers of kelp forests in Europe: an expert assessment.
 643 *Biodiversity and Conservation*, 25(7), 1319–1348. doi: 10.1007/s10531-016-1141-7
- 644 Ardehed, A., Johansson, D., Schagerström, E., Kautsky, L., Johannesson, K., & Pereyra, R. T.
 645 (2015). Complex spatial clonal structure in the macroalgae *Fucus radicans* with both
 646 sexual and asexual recruitment. *Ecology and Evolution*, 5(19), 4233–4245. doi:
 647 10.1002/ece3.1629
- 648 Arnaud-Haond, S., Migliaccio, M., Diaz-Almela, E., Teixeira, S., Van De Vliet, M. S., Alberto, F.,
 649 ... Serrão, E. A. (2007). Vicariance patterns in the Mediterranean Sea: east–west
 650 cleavage and low dispersal in the endemic seagrass *Posidonia oceanica*. *Journal of*
 651 *Biogeography*, 34(6), 963–976. doi: 10.1111/j.1365-2699.2006.01671.x
- 652 Arnaud-Haond, S., Duarte, C. M., Alberto, F., & Serrao, E. A. (2007). Standardizing methods
 653 to address clonality in population studies. *Molecular ecology*, 16(24), 5115–5139. doi:
 654 10.1111/j.1365-294X.2007.03535.x
- 655 Arnaud-Haond, S., Stoeckel, S., & Bailleul, D. (2020). New insights into the population
 656 genetics of partially clonal organisms: when seagrass data meet theoretical
 657 expectations. *Molecular Ecology*, mec.15532. doi: 10.1111/mec.15532
- 658 Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E.
 659 A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers.

660 *PLoS ONE*, 3(10), e3376. doi: 10.1371/journal.pone.0003376

661 Ballenghien, M., Faivre, N., & Galtier, N. (2017). Patterns of cross-contamination in a
662 multispecies population genomic project: detection, quantification, impact, and
663 solutions. *BMC Biology*, 15(1), 25. doi: 10.1186/s12915-017-0366-6

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

666 Balloux, F., Lehmann, L., & De Meeûs, T. (2003). The Population Genetics of Clonal and
667 Partially Clonal Diploids. *Genetics Society of America*, 164(4), 1635-1644.

668 consequences of a major earthquake on asexual and sexual populations. *Evolutionary*
669 *Applications*. doi: 10.1111/eva.12967

670 Becheler, R., Masson, JP., Arnaud-Haond, S., Halkett, F., Mariette, S., Guillemin, ML., ...
671 Stoeckel, S. (2017). ClonEstiMate, a Bayesian method for quantifying rates of
672 clonality of populations genotyped at two-time steps. *Molecular Ecology Resources*,
673 17(6), e251–e267. doi: 10.1111/1755-0998.12698

674 Becheler, R., Guillemin, M.-L., Stoeckel, S., Mauger, S., Saunier, A., Brante, A., ... Valero, M.
675 (2020). After a catastrophe, a little bit of sex is better than nothing: Genetic
676 Beichman, A. C., Koepfli, K.-P., Li, G., Murphy, W., Dobrynin, P., Kliver, S., ... Wayne, R. K.
677 (2019). Aquatic Adaptation and Depleted Diversity: A Deep Dive into the Genomes of
678 the Sea Otter and Giant Otter. *Molecular Biology and Evolution*, 36(12), 2631–2655.
679 doi: 10.1093/molbev/msz101

680 Billot, C., Engel-Gautier, C., Rousvoal, S., Kloareg, B., & Valero, M. (2003). Current patterns,
681 habitat discontinuities and population genetic structure: the case of the kelp
682 *Laminaria digitata* in the English Channel. *Marine Ecology Progress Series*, 253, 111–
683 121. doi: 10.3354/meps253111

684 Bringloe, T. T., Starko, S., Wade, R. M., Vieira, C., Kawai, H., Clerck, O. D., ... Verbruggen, H.
685 (2020). Phylogeny and Evolution of the Brown Algae. *Critical Reviews in Plant*
686 *Sciences*, 39(4), 281–321. doi: 10.1080/07352689.2020.1787679

687 Birky, C. W. (1996). Heterozygosity, heteromorphy, and phylogenetic trees in asexual
688 eukaryotes. *Genetics*, 144(1), 427-437.

689 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina
690 sequence data. *Bioinformatics*, 30(15), 2114–2120. doi:
691 10.1093/bioinformatics/btu170

692 Boscari, E., Abbiati, M., Badalamenti, F., Bavestrello, G., Benedetti-Cecchi, L., Cannas, R., ...
693 Congiu, L. (2019). A population genomics insight by 2b-RAD reveals populations'
694 uniqueness along the Italian coastline in *Leptopsammia pruvoti* (Scleractinia,
695 Dendrophylliidae). *Diversity and Distributions*, 25(7), 1101–1117. doi:
696 10.1111/ddi.12918

697 Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks:
698 Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes*,
699 *Genomes, Genetics*, 1(3), 171–182. doi: 10.1534/g3.111.000240

700 Coleman, M. A., Gillanders, B. M., & Connell, S. D. (2009). Dispersal and gene flow in the
701 habitat-forming kelp, *Ecklonia radiata*: relative degrees of isolation across an east–
702 west coastline. *Marine and Freshwater Research*, 60(8), 802–809. doi:
703 10.1071/MF08268

704 Coleman, M., & Wernberg, T. (2018). Genetic and morphological diversity in sympatric kelps
705 with contrasting reproductive strategies. *Aquatic Biology*, 27, 65–73. doi:
706 10.3354/ab00698

707 Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., & François, O. (2010). Approximate Bayesian
708 Computation (ABC) in practice. *Trends in Ecology & Evolution*, *25*(7), 410–418. doi:
709 10.1016/j.tree.2010.04.001

710 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R.
711 (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. doi:
712 10.1093/bioinformatics/btr330

713 Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L.
714 (2011). Genome-wide genetic marker discovery and genotyping using next-
715 generation sequencing. *Nature Reviews Genetics*, *12*(7), 499–510. doi:
716 10.1038/nrg3012

717 De Meeûs, T., Lehmann, L., & Balloux, F. (2006). Molecular epidemiology of clonal diploids: A
718 quick overview and a short DIY (do it yourself) notice. *Infection, Genetics and
719 Evolution*, *6*(2), 163–170. doi: 10.1016/j.meegid.2005.02.004

720 De Meeûs, T., Prugnolle, F., & Agnew, P. (2007). Asexual reproduction: Genetics and
721 evolutionary aspects. *Cellular and Molecular Life Sciences*, *64*(11), 1355–1372. doi:
722 10.1007/s00018-007-6515-2

723 De Meeûs, T. (2012). Initiation à la génétique des populations naturelles: Applications aux
724 parasites et à leurs vecteurs. IRD Editions.

725 Demes, K. W., & Graham, M. H. (2011). Abiotic Regulation of Investment in Sexual Versus
726 Vegetative Reproduction in the Clonal Kelp *Laminaria Sinclairii* (Laminariales,
727 Phaeophyceae)1. *Journal of Phycology*, *47*(3), 463–470. doi: 10.1111/j.1529-
728 8817.2011.00981.x

729 Dorken, M. E., & Eckert, C. G. (2001). Severely reduced sexual reproduction in northern
730 populations of a clonal plant, *Decodon verticillatus* (Lythraceae). *Journal of Ecology*,
731 *89*(3), 339–350. doi: 10.1046/j.1365-2745.2001.00558.x

732 Duminil, J., Fineschi, S., Hampe, A., Jordano, P., Salvini, D., Vendramin, G. G., & Petit, R. J.
733 (2007). Can Population Genetic Structure Be Predicted from Life-History Traits? *The
734 American Naturalist*, *169*(5), 662–672. doi: 10.1086/513490

735 Durrant, H. M. S., Barrett, N. S., Edgar, G. J., Coleman, M. A., & Burrige, C. P. (2018).
736 Seascape habitat patchiness and hydrodynamics explain genetic structuring of kelp
737 populations. *Marine Ecology Progress Series*, *587*, 81–92. doi: 10.3354/meps12447

738 Ercegović, A. (1960). Quelques traits caractéristiques de la végétation des algues de
739 l'Adriatique. *Acta Botanica Croatica*, *18*–19.

740 Frankham, R. (1996). Relationship of Genetic Variation to Population Size in Wildlife.
741 *Conservation Biology*, *10*(6), 1500–1508. doi: 10.1046/j.1523-1739.1996.10061500.x

742 Goecke, F., Klemetsdal, G. and Ergon, Å. (2020). Cultivar Development of Kelps for
743 Commercial Cultivation—Past Lessons and Future Prospects. *Frontiers in Marine
744 Science* *7*, 110.

745 González de León, S., Herrera, I., & Guevara, R. (2016). Mating system, population growth,
746 and management scenario for *Kalanchoe pinnata* in an invaded seasonally dry
747 tropical forest. *Ecology and Evolution*, *6*(13), 4541–4550. doi: 10.1002/ece3.2219

748 Gorospe, K. D., Donahue, M. J., & Karl, S. A. (2015). The importance of sampling design:
749 spatial patterns and clonality in estimating the genetic diversity of coral reefs. *Marine
750 Biology*, *162*(5), 917–928. doi: 10.1007/s00227-015-2634-8

751 Goudet, J., Raymond, M., De Meeûs, T., & Rousset, F. (1996). Testing differentiation in
752 diploid populations. *Genetics*, *144*(4), 1933–1940. doi: 10.1093/genetics/144.4.1933

753 Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics.
754 *Molecular Ecology Notes*, 5(1), 184–186. doi: 10.1111/j.1471-8286.2004.00828.x

755 Guillemin, M.-L., Faugeron, S., Destombe, C., Viard, F., Correa, J. A., & Valero, M. (2008).
756 Genetic variation in wild and cultivated populations of the haploid– diploid red alga
757 *Gracilaria chilensis*: how farming practices favor asexual reproduction and
758 heterozygosity. *Evolution*, 62(6), 1500–1519. doi: 10.1111/j.1558-5646.2008.00373.x

759 Halkett, F., Plantegenest, M., Prunier-Leterme, N., Mieuxet, L., Delmotte, F., & Simon, J. C.
760 (2005). Admixed sexual and facultatively asexual aphid lineages at mating sites:
761 admixed sexual and asexual lineages in aphids. *Molecular Ecology*, 14(1), 325–336.
762 doi: 10.1111/j.1365-294X.2004.02358.x

763 Halkett, F., Simon, J.-C., & Balloux, F. (2005). Tackling the population genetics of clonal and
764 partially clonal organisms. *Trends in Ecology & Evolution*, 20(4), 194–201. doi:
765 10.1016/j.tree.2005.01.001

766 Hamrick, J. L., & Godt, M. J. W. (1990). Allozyme diversity in plant species. *Plant Population*
767 *Genetics, Breeding, and Genetic Resources.*, 43–63.

768 Huson, D. H. (1998). SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*.
769 14(1), 68–73. doi: 10.1093/bioinformatics/14.1.68

770 Huvé, H. (1955). Présence de *Laminaria rodriguezii* Bornet sur les cotes françaises de la
771 Méditerranée. *Recueil Des Travaux de La Station Marine d'Endoume*, 15.

772 Jaron, K. S., Bast, J., Nowell, R. W., Ranallo-Benavidez, T. R., Robinson-Rechavi, M., &
773 Schwander, T. (2020). Genomic features of parthenogenetic animals. BioRxiv doi:
774 10.1101/497495

775 Judson, O. P., & Normark, B. B. (1996). Ancient asexual scandals. *Trends in Ecology &*
776 *Evolution*, 11(2), 41–46. doi: 10.1016/0169-5347(96)81040-8

777 Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-
778 wide population genetic data with emphasis on clonality. *Frontiers in Genetics*, 6. doi:
779 10.3389/fgene.2015.00208

780 Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic
781 analysis of populations with clonal, partially clonal, and/or sexual reproduction.
782 *PeerJ*, 2, e281. doi: 10.7717/peerj.281

783 Kettenring, K. M., & Mock, K. E. (2012). Genetic diversity, reproductive mode, and dispersal
784 differ between the cryptic invader, *Phragmites australis*, and its native conspecific.
785 *Biological Invasions*, 14(12), 2489–2504. doi: 10.1007/s10530-012-0246-5

786 Koffi, M., Meeûs, T. D., Bucheton, B., Solano, P., Camara, M., Kaba, D., ... Jamonneau, V.
787 (2009). Population genetics of *Trypanosoma brucei gambiense*, the agent of sleeping
788 sickness in Western Africa. *Proceedings of the National Academy of Sciences*, 106(1),
789 209–214. doi: 10.1073/pnas.0811080106

790 Laurin-Lemay, S., Brinkmann, H., & Philippe, H. (2012). Origin of land plants revisited in the
791 light of sequence contamination and missing data. *Current Biology*, 22(15), R593–
792 R594. doi: 10.1016/j.cub.2012.06.013

793 Le Cam, S., Daguin-Thiébaud, C., Bouchemousse, S., Engelen, A. H., Mieszkowska, N., & Viard,
794 F. (2019). A genome-wide investigation of the worldwide invader *Sargassum*
795 *muticum* shows high success albeit (almost) no genetic diversity. *Evolutionary*
796 *Applications*, 13(3), 500-514. doi: 10.1111/eva.12837

797 Levy, E., Byrne, M., Coates, D. J., Macdonald, B. M., McArthur, S., & van Leeuwen, S. (2016).
798 Contrasting Influences of Geographic Range and Distribution of Populations on

799 Patterns of Genetic Diversity in Two Sympatric Pilbara Acacias. *PLoS ONE*, 11(10). doi:
800 10.1371/journal.pone.0163995

801 Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler
802 transform. *Bioinformatics*, 25(14), 1754–1760. doi: 10.1093/bioinformatics/btp324

803 Liesner, D., Fouqueau, L., Valero, M., Roleda, M. Y., Pearson, G. A., Bischof, K., ... Bartsch, I.
804 (2020). Heat stress responses and population genetics of the kelp *Laminaria digitata*
805 (Phaeophyceae) across latitudes reveal differentiation among North Atlantic
806 populations. *Ecology and Evolution*, 10(17), 9144–9177. doi: 10.1002/ece3.6569

807 Locatelli, N., & Drew, J. (2019). Population structure and clonal prevalence of scleractinian
808 corals (*Montipora capitata* and *Porites compressa*) in Kaneohe Bay, Oahu. *BioRxiv*
809 doi: 10.1101/2019.12.11.860585

810 Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: an R package to perform genome scans
811 for selection based on principal component analysis. *Molecular Ecology Resources*,
812 17(1), 67–77. doi: 10.1111/1755-0998.12592

813 Marshall, D. R., & Weir, B. S. (1979). Maintenance of genetic variation in apomictic plant
814 populations. *Heredity*, 42(2), 159–172. doi: 10.1038/hdy.1979.20

815 Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C.
816 (2015). Restriction site-associated DNA sequencing, genotyping error estimation and
817 de novo assembly optimization for population genetic inference. *Molecular Ecology*
818 *Resources*, 15(1), 28–41. doi: 10.1111/1755-0998.12291

819 McKinney, G. J., Waples, R. K., Seeb, L. W., & Seeb, J. E. (2017). Paralogs are revealed by
820 proportion of heterozygotes and deviations in read ratios in genotyping-by-
821 sequencing data from natural populations. *Molecular Ecology Resources*, 17(4), 656–
822 669. doi: 10.1111/1755-0998.12613

823 Meloni, M., Reid, A., Caujapé-Castells, J., Marrero, Á., Fernández-Palacios, J. M., Mesa-Coelo,
824 R. A., & Conti, E. (2013). Effects of clonality on the genetic variability of rare, insular
825 species: the case of *Ruta microcarpa* from the Canary Islands. *Ecology and Evolution*,
826 3(6), 1569–1579. doi: 10.1002/ece3.571

827 Nadukkalam Ravindran, P., Bentzen, P., Bradbury, I. R., & Beiko, R. G. (2018). PMERGE:
828 Computational filtering of paralogous sequences from RAD-seq data. *Ecology and*
829 *Evolution*, 8(14), 7002–7013. doi: 10.1002/ece3.4219

830 Nebavi, F., Ayala, F. J., Renaud, F., Bertout, S., Eholie, S., Moussa, K., ... De Meeûs, T. (2006).
831 Clonal population structure and genetic diversity of *Candida albicans* in AIDS patients
832 from Abidjan (Cote d’Ivoire). *Proceedings of the National Academy of Sciences*,
833 103(10), 3663–3668. doi: 10.1073/pnas.0511328103

834 Nougé, O., Rode, N. O., Jabbour-zahab, R., Ségard, A., Chevin, L.-M., Haag, C. R., &
835 Lenormand, T. (2015). Automixis in *Artemia*: Solving a century-old controversy.
836 *Journal of Evolutionary Biology*, 28(12), 2337–2348. doi: 10.1111/jeb.12757

837 O’Leary, S. J., Puritz, J. B., Willis, S. C., Hollenbeck, C. M., & Portnoy, D. S. (2018). These aren’t
838 the loci you’re looking for: Principles of effective SNP filtering for molecular ecologists.
839 *Molecular Ecology*, 27(16), 3193–3206. doi: <https://doi.org/10.1111/mec.14792>

840 Omilian, A. R., Cristescu, M. E. A., Dudycha, J. L., & Lynch, M. (2006). Asexual recombination
841 in asexual lineages of *Daphnia*. *Proceedings of the National Academy of Sciences*,
842 103(49), 18638–18643. doi: 10.1073/pnas.0606435103

843 Pardo, C., Guillemin, M.-L., Peña, V., Bárbara, I., Valero, M., & Barreiro, R. (2019). Local

844 Coastal Configuration Rather Than Latitudinal Gradient Shape Clonal Diversity and
845 Genetic Structure of *Phymatolithon calcareum* Maerl Beds in North European
846 Atlantic. *Frontiers in Marine Science*, 6, 149. doi: 10.3389/fmars.2019.00149

847 Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest
848 RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in
849 Model and Non-Model Species. *PLoS ONE*, 7(5). doi: 10.1371/journal.pone.0037135

850 Raymond, M., & Rousset, F. (1995). An Exact Test for Population Differentiation. *Evolution*,
851 49(6), 1280–1283. doi: <https://doi.org/10.1111/j.1558-5646.1995.tb04456.x>

852 Reichel, K., Masson, J.-P., Malrieu, F., Arnaud-Haond, S., & Stoeckel, S. (2016). Rare sex or
853 out of reach equilibrium? The dynamics of FIS in partially clonal organisms. *BMC*
854 *Genetics*, 17(1), 76. doi: 10.1186/s12863-016-0388-z

855 Robuchon, M., Le Gall, L., Mauger, S. and Valero, M. 2014. Contrasting genetic diversity
856 patterns in two sister kelp species co-distributed along the coast of Brittany, France.
857 *Molecular Ecology* 23:2669-2685. doi:10.1111/mec.12774

858 Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for
859 paired-end sequencing improve RADseq-based population genomics. *Molecular*
860 *Ecology*, 28(21), 4737–4754. doi: 10.1111/mec.15253

861 Rousset, F., & Raymond, M. (1995). Testing heterozygote excess and deficiency. *Genetics*,
862 140(4), 1413–1419.

863 Rousset, François. (2008). genepop'007: a complete re-implementation of the genepop
864 software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103–106. doi:
865 10.1111/j.1471-8286.2007.01931.x

866 Santelices, B. (1990). Patterns of reproduction, dispersal and recruitment in seaweeds.
867 *Oceanography and Marine Biology*, 28, 177–276.

868 Schaefer, I., Domes, K., Heethoff, M., Schneider, K., Schon, I., Norton, R. A., ... Maraun, M.
869 (2006). No evidence for the “Meselson effect” in parthenogenetic oribatid mites
870 (*Oribatida*, *Acari*). *Journal of Evolutionary Biology*, 19(1), 184–193. doi:
871 10.1111/j.1420-9101.2005.00975.x

872 Schwander, T., Henry, L., & Crespi, B. J. (2011). Molecular Evidence for Ancient Asexuality in
873 *Timema* Stick Insects. *Current Biology*, 21(13), 1129–1134. doi:
874 10.1016/j.cub.2011.05.026

875 Schwander, T. (2016). Evolution: The end of an ancient asexual scandal. *Current Biology*,
876 26(6), 233–235. doi: 10.1016/j.cub.2016.01.034

877 Sekhon, J. S. (2008). Multivariate and propensity score matching software with automated
878 balance optimization: The Matching package for R (SSRN Scholarly Paper No. ID
879 1009044). Rochester, NY: Social Science Research Network.

880 Shafer, A. B. A., Peart, C. R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C. W., & Wolf, J. B. W.
881 (2017). Bioinformatic processing of RAD-seq data dramatically impacts downstream
882 population genetic inference. *Methods in Ecology and Evolution*, 8(8), 907–917. doi:
883 <https://doi.org/10.1111/2041-210X.12700>

884 Shrestha, S., Hu, J., Fryxell, R. T., Mudge, J., & Lamour, K. (2014). SNP markers identify widely
885 distributed clonal lineages of *Phytophthora colocasiae* in Vietnam, Hawaii and Hainan
886 Island, China. *Mycologia*, 106(4), 676–685. doi: 10.3852/13-165

887 Stoeckel, S., & Masson, J.-P. (2014). The Exact Distributions of FIS under Partial Asexuality in
888 Small Finite Populations with Mutation. *PLoS ONE*, 9(1). doi:

889 10.1371/journal.pone.0085228
890 Stoeckel, S., Porro, B., & Arnaud-Haond, S. (2019). The discernible and hidden effects of
891 clonality on the genotypic and genetic states of populations: improving our
892 estimation of clonal rates. *arXiv*, 1902.09365, ver. 4 peer-reviewed and
893 recommended by Peer Community in Evolutionary Biology.
894 <https://arxiv.org/abs/1902.09365v4> submitted to Molecular Ecology Resources
895 Verdu, C. F., Guichoux, E., Quevauvillers, S., De Thier, O., Laizet, Y., Delcamp, A., ... Mariette,
896 S. (2016). Dealing with paralogy in RADseq data: in silico detection and single
897 nucleotide polymorphism validation in *Robinia pseudoacacia* L. *Ecology and*
898 *Evolution*, 6(20), 7323–7333. doi: 10.1002/ece3.2466
899 Weir, W., Capewell, P., Foth, B., Clucas, C., Pountain, A., Steketee, P., ... MacLeod, A. (2016).
900 Population genomics reveals the origin and asexual evolution of human infective
901 trypanosomes. *ELife*, 5, e11473. doi: 10.7554/eLife.11473
902 Weir, B. S., & Cockerham, C. C. (1984). Estimating f-statistics for the analysis of population
903 structure. *Evolution*, 38(6), 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x
904 Wilson, D. J. (2019). The harmonic mean p -value for combining dependent tests.
905 *Proceedings of the National Academy of Sciences*, 116(4), 1195–1200. doi:
906 10.1073/pnas.1814092116
907 Žuljević, A., Peters, A. F., Nikolić, V., Antolić, B., Despalatović, M., Cvitković, I., ... Küpper, F. C.
908 (2016). The Mediterranean deep-water kelp *Laminaria rodriguezii* is an endangered
909 species in the Adriatic Sea. *Marine Biology*, 163(4), 69. doi: 10.1007/s00227-016-
910 2821-2

911

912

913

914

915

916

917

918

919 **DATA ACCESSIBILITY**

920 Individual high-quality reads (dd-RADseq) that support the findings of this study are openly
921 available in a Dryad Digital Repository (doi:10.5061/dryad.hmgqnk9dq). The home-made
922 scripts used to generate the distribution of F_{IS} values into classes, the histograms of pairwise
923 genetic distances, and the graphs representing the percentage of polymorphic loci are
924 available in the public repository: <https://github.com/LauricReynes>

925 **AUTHOR CONTRIBUTION**

926 Thierry Thibaut and Didier Aurelle devised the project and were responsible for the main
927 conceptual ideas. Florian Holon performed the sampling of *Laminaria rodriguezii*. Aurélie
928 Blanfuné contributed to sample preparation. Stéphane Mauger and Lauric Reynes performed
929 the RAD-sequencing experiment and analyzed sequencing reads. Corinne Cruaud and Arnaud
930 Couloux provided the draft genome of *Laminaria digitata*. Lauric Reynes performed research,
931 analyzed data, and wrote the manuscript in close collaboration with Didier Aurelle and
932 Myriam Valero. All authors provided critical feedback and helped to shape the research,
933 analysis, and manuscript.

934

935

936

937

938

939

940

941

942 **TABLES AND FIGURES**

943 **Table captions**

944 **Table 1.** Quality filters used to build the final RAD-seq datasets of *Laminaria rodriguezii* and
945 *Laminaria digitata*. For each filtering step, the number of SNP passing the filter (SNPs count),
946 as well as the mean missing data rate per individual (missingness) and the mean sequencing
947 depth per individual (coverage) are reported. Loci were initially built in both species from the
948 draft genome of *L. digitata* using the Stacks v.2.4 bioinformatics pipeline. The threshold of
949 missing genotypes per individual < 30% has deleted four and nine individuals in *L. rodriguezii*
950 and *L. digitata* datasets respectively. See Methods for details.

951 **Table 2.** Comparison of population genetic summary statistics (mean and range of values) in
952 *Laminaria rodriguezii* and *Laminaria digitata*. %P, average percentage of polymorphic loci
953 within populations rarified at $n = 10$ by random resampling (see methods for details), H_o ,
954 observed heterozygosity, H_e , expected heterozygosity, F_{IS} , fixation index.

955 **Table 3.** Genetic variation within populations of *Laminaria rodriguezii* (A) before and (B) after
956 applying the clonal correction. Clonal correction consists of keeping only one representative
957 of clonal lineages. N, number of individuals successfully genotyped, H_o , observed
958 heterozygosity, H_e , expected heterozygosity, F_{IS} , fixation index, $\overline{r_d}$, average multilocus linkage
959 disequilibrium. F_{IS} and $\overline{r_d}$ values highlighted in bold indicate significant heterozygote excess
960 within populations ($P\text{-val} < 0.001$) and significant linkage disequilibrium ($P\text{-val} < 0.001$),
961 respectively.

962 **Table 4.** The effect of SNP filtering cutoffs on the mean (\pm SD) F_{IS} over loci, overall (\pm SD) F_{ST}
963 over loci and the SNP error rate in *Laminaria rodriguezii*. As thresholds, we considered the
964 maximum rate of missing data allowed per SNP (max-missing), the Minor Allele Frequency
965 (MAF), the mean read depth per locus across all individuals (meanDP). For each cutoff value
966 we reported the number of SNP passing the filter (SNPs count), the mean missing data rate
967 per individual (missingness), and the mean read depth per individual (coverage). We
968 considered an additional set of SNPs (matching genotypes) for which replicates have exactly
969 the same multi-locus genotypes (SNP error rate = 0).

970

971

972

973

974

975

976

977

978 **Figure captions**

979 **Figure 1.** Sampling map of *Laminaria rodriguezii* in Eastern Provence and Southern Corsica,
980 reporting the depth in meters of sampling sites and the number of individuals sampled per
981 site (N).

982 **Figure 2.** Frequency of pairwise genetic distances between individuals of (A) *Laminaria*
983 *rodriguezii* and (B) *Laminaria digitata*. Intra-population (Intra-pop), inter-population
984 (Inter-pop), within Helgoland (Intra-Hel) and replicates (Replicates) comparisons are colored
985 according to the legend. The frequency of pairwise genetic distances across replicates was
986 multiplied by 40 to create an optimum bin size. The vertical dotted line shows the distance
987 threshold ($d = 0.022$) as reported as the maximum value of replicate pairs.

988 **Figure 3.** Phylogenetic networks of *Laminaria rodriguezii* represented with the neighbor-net
989 method. Each dot represents an individual, Individuals belonging to the same clone (MLLs,
990 Multi-Locus Lineages) are colored according to the legend. Individuals characterized by the
991 highest proportion of heterozygous genotypes and private alleles are indicated as outliers.

992 **Figure 4.** Distributions of the fixation index F_{IS} among loci within populations of (A) *Laminaria*
993 *rodriguezii* and (B) *Laminaria digitata* computed across 4 077 and 7364 putatively neutral
994 SNPs, respectively. The F_{IS} values were distributed into 10 discrete classes and the number of
995 occurrences for each class is indicated on top of histogram bars. The variance of F_{IS} among loci
996 (Var) is indicated in brackets next to population labels.

997 **Figure 5.** Distributions of the standardized index of multilocus association \bar{r}_d within
998 populations of *Laminaria rodriguezii* before (highlighted in red) and after (highlighted in
999 green) clonal correction (cc). Comparison of the index value was done with five populations
1000 of *Laminaria digitata* (highlighted in blue). The \bar{r}_d values were estimated using 10 000
1001 iterations by random subsampling of 1000; 2000 and 3000 SNPs. The abbreviations used for
1002 populations are those reported in Table 3 for *L. rodriguezii* and Table S1 for *L. digitata*.

1003

1004

1005

1006

1007

1008

1009

1010

1011 **Table 1**

| | %P | | H _o | | | H _e | | | F _{IS} | | | |
|------------------------------|-------|---|----------------|------|---|----------------|------|---|-----------------|-------|---|-------|
| <i>L. rodriguezii</i> | | | | | | | | | | | | |
| Average | 54.42 | | 0.16 | | | 0.14 | | | -0.14 | | | |
| Range | 40.96 | - | 72.14 | 0.11 | - | 0.19 | 0.11 | - | 0.18 | -0.47 | - | -0.02 |
| <i>L. digitata</i> | | | | | | | | | | | | |
| Average | 39.39 | | 0.11 | | | 0.15 | | | 0.26 | | | |
| Range | 21.44 | - | 54.89 | 0.09 | - | 0.14 | 0.12 | - | 0.19 | 0.12 | - | 0.43 |

1012

1013

1014

1015

1016 **Table 2**

1017

| Filtering thresholds | <i>Laminaria rodriguezii</i> | | | <i>Laminaria digitata</i> | | |
|--|------------------------------|-------------|----------|---------------------------|-------------|----------|
| | SNP counts | missingness | coverage | SNP counts | missingness | coverage |
| Locus genotyped > 75% of the samples > 75% of the populations SNP kept observed heterozygosity < 80% | 70 026 | 13.97 % | 11.29 | 609 699 | 12.26 % | 14.11 |
| Linkage disequilibrium single SNP per locus | 27 769 | 13.18 % | 11.88 | 58 830 | 13.45 % | 14.36 |
| Minor Allele Frequency (MAF) MAF > 2% | 16 865 | 13.51 % | 12.18 | 21 090 | 15.01 % | 14.43 |
| Missingness per individual < 30% per SNP < 20% | 13 235 | 4.46 % | 13.91 | 15 150 | 4.73 % | 17.24 |
| Coverage mean read depth > 15 X mean read depth < 100 X | 4 366 | 2.27 % | 22.17 | 7 538 | 3.40 % | 21.63 |
| Outliers detection putatively neutral | 4 077 | 2.29 % | 22.04 | 7364 | 3.39 % | 21.64 |

1018

1019

1020

1021

1022

1023

1024

1025

1026

1027

1028

1029

1030

1031

1032

1033

1034

Table 3

| (A) | | | | | | |
|-------------------|--------------|----------|----------------------|----------------------|-----------------------|-------------------------------|
| Population | Label | N | H_o | H_e | F_{IS} | \bar{r}_d |
| Banc Magaud_1 | BM_1 | 10 | 0.18 | 0.17 | -0.02 | 0.01 |
| Banc Magaud_2 | BM_2 | 11 | 0.19 | 0.18 | -0.04 | 0.08 |
| Cap Camarat | CAM | 11 | 0.11 | 0.11 | -0.05 | 0.10 |
| Bonifacio | BON | 11 | 0.16 | 0.11 | -0.47 | 0.27 |
| (B) | | | | | | |
| Population | Label | N | H_o | H_e | F_{IS} | \bar{r}_d |
| Banc Magaud_1 | BM_1 | 10 | 0.18 | 0.17 | -0.02 | 0.01 |
| Banc Magaud_2 | BM_2 | 9 | 0.19 | 0.19 | -0.01 | 0.05 |
| Cap Camarat | CAM | 10 | 0.11 | 0.11 | -0.02 | 0.10 |
| Bonifacio | BON | 3 | 0.20 | 0.16 | -0.26 | 0.20 |

1035

1036

Table 4

| Quality filter | cutoff | SNPs count | missingness (%) | coverage | mean (\pm SD) F_{IS} | overall (\pm SD) F_{ST} | SNP error rate (%) |
|--|--------|------------|-----------------|----------|---------------------------|------------------------------|--------------------|
| Reference dataset | | 21 357 | 4.11 | 13.00 | -0.07(0.20) | 0.27(0.18) | 5.22 |
| max-missing (%) | < 15 | 21 114 | 3.97 | 13.08 | -0.07(0.20) | 0.27(0.18) | 5.17 |
| | < 10 | 19 468 | 3.33 | 13.52 | -0.07(0.20) | 0.27(0.18) | 4.94 |
| | < 5 | 14 480 | 1.95 | 14.88 | -0.07(0.19) | 0.28(0.18) | 4.30 |
| | 0 | 5 479 | 0.00 | 18.12 | -0.06(0.18) | 0.30(0.19) | 2.90 |
| MAF (%) | > 2 | 12 735 | 4.28 | 13.31 | -0.08(0.26) | 0.30(0.22) | 7.07 |
| | > 5 | 5 094 | 4.54 | 16.18 | -0.12(0.34) | 0.37(0.26) | 4.92 |
| | > 10 | 3 428 | 4.56 | 17.29 | -0.14(0.37) | 0.42(0.26) | 3.86 |
| meanDP (X) | > 5 | 19 245 | 3.59 | 13.71 | -0.08(0.20) | 0.28(0.18) | 4.88 |
| | > 10 | 11 773 | 2.52 | 17.41 | -0.11(0.22) | 0.31(0.21) | 3.92 |
| | > 15 | 6 163 | 1.90 | 21.85 | -0.13(0.23) | 0.33(0.22) | 2.74 |
| max-missing cutoff for fixed MAF (2%) and meanDP (>15) | < 20 | 3 911 | 2.14 | 21.83 | -0.15(0.29) | 0.34(0.25) | 3.71 |
| | < 10 | 3 850 | 1.99 | 21.81 | -0.14(0.28) | 0.35(0.25) | 3.70 |
| | < 5 | 3 464 | 1.41 | 21.90 | -0.12(0.28) | 0.35(0.25) | 3.53 |
| | 0 | 1 813 | 0.00 | 22.71 | -0.10(0.27) | 0.37(0.26) | 2.68 |
| matching genotypes | | 3 638 | 2.14 | 22.07 | -0.15(0.29) | 0.35(0.25) | 0.00 |

FRANCE

ITALY

SPAIN

Mediterranean Sea

5 km

Eastern
Provence

Cap Camarat
(N=12); -65 to -73m

Banc Magaud_1
(N=12); -76m

Banc Magaud_2
(N=11); -71m

N

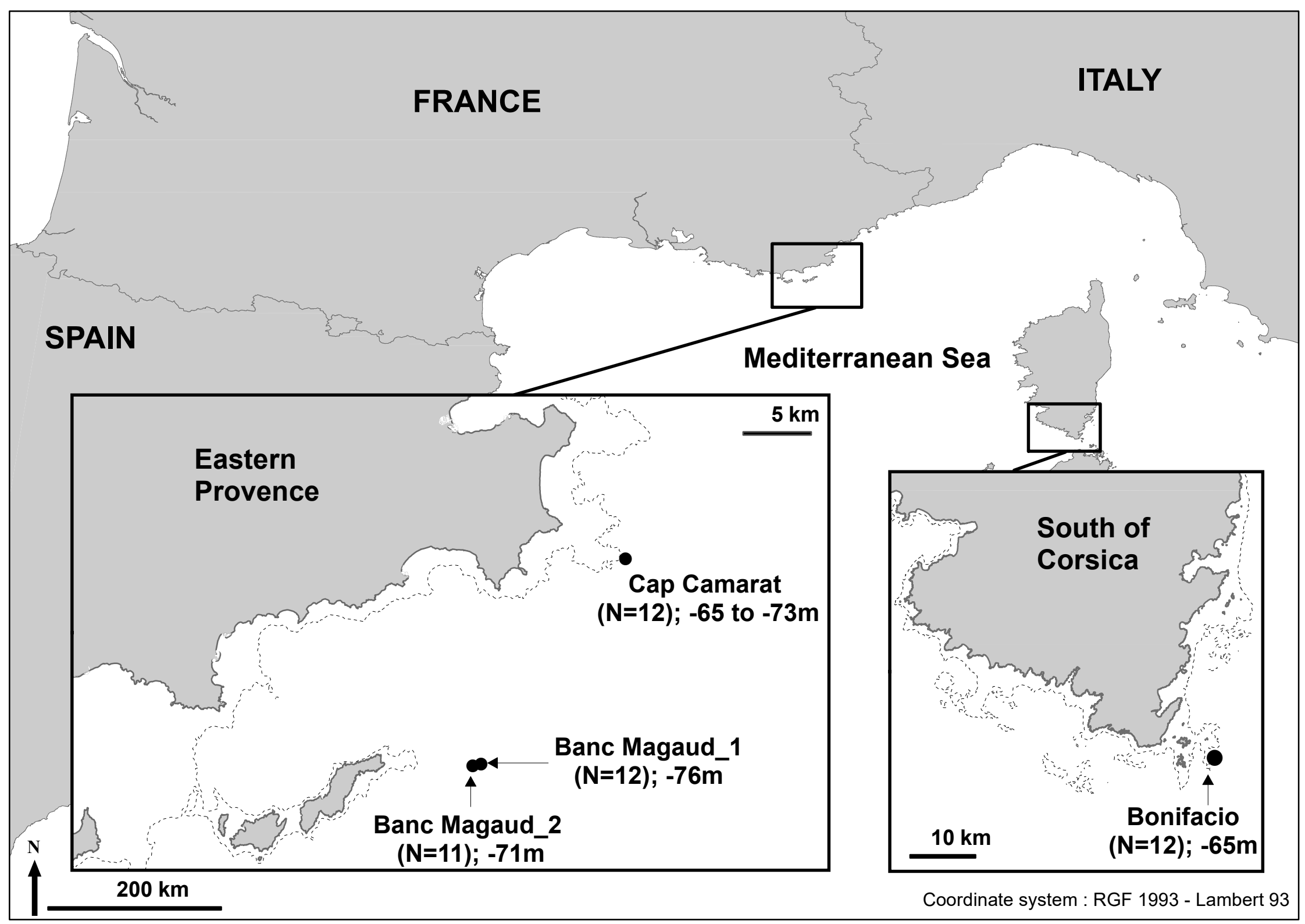
200 km

South of
Corsica

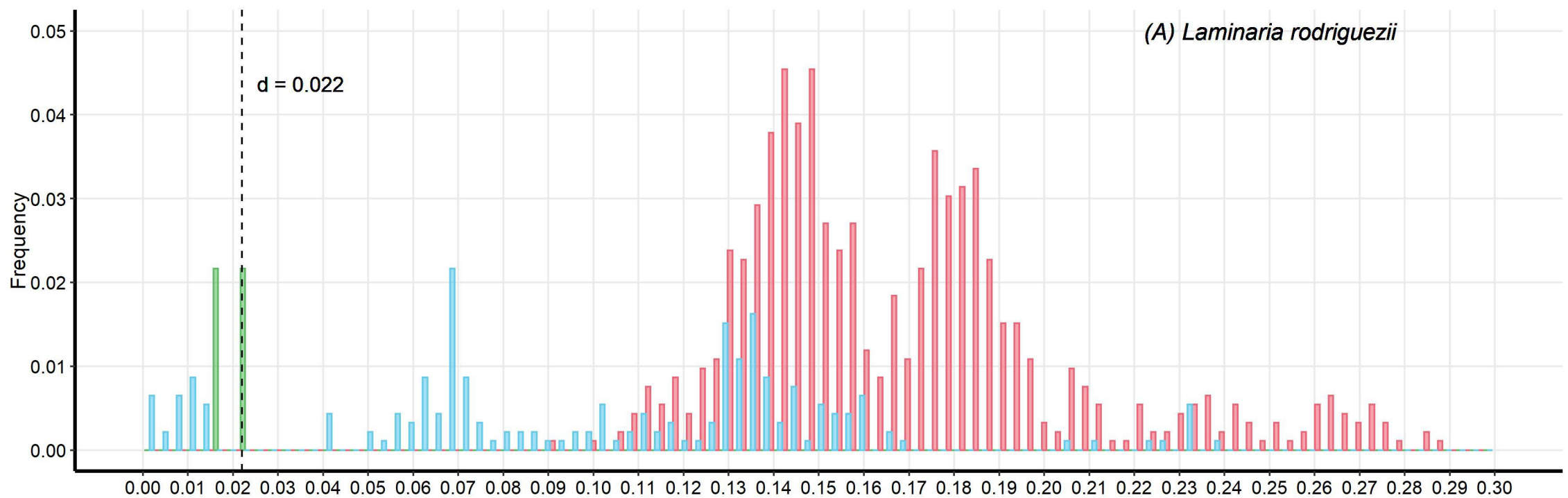
Bonifacio
(N=12); -65m

10 km

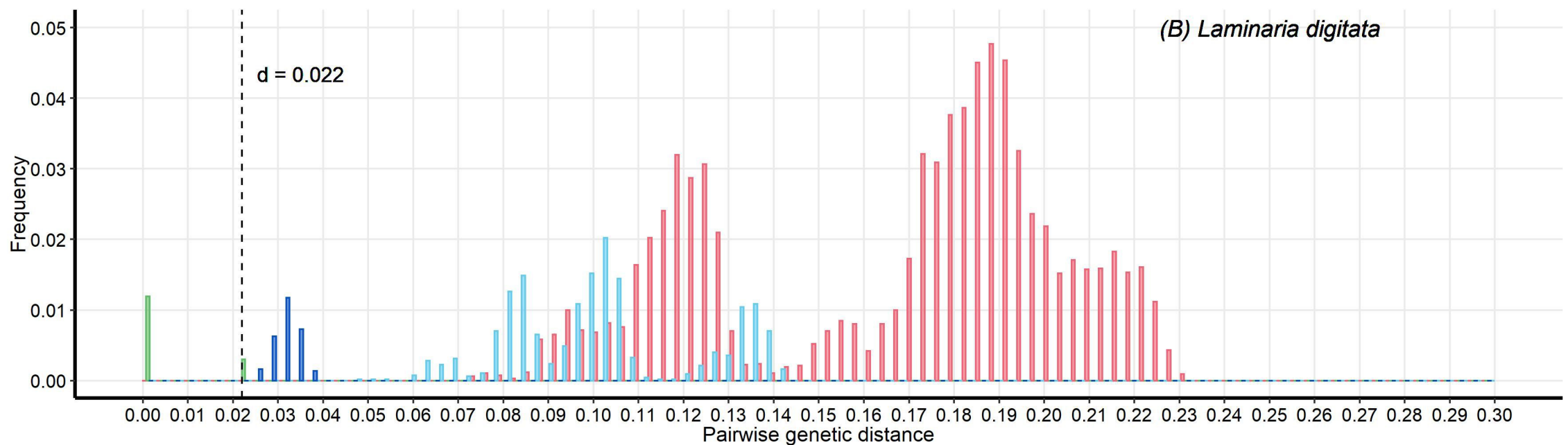
Coordinate system : RGF 1993 - Lambert 93



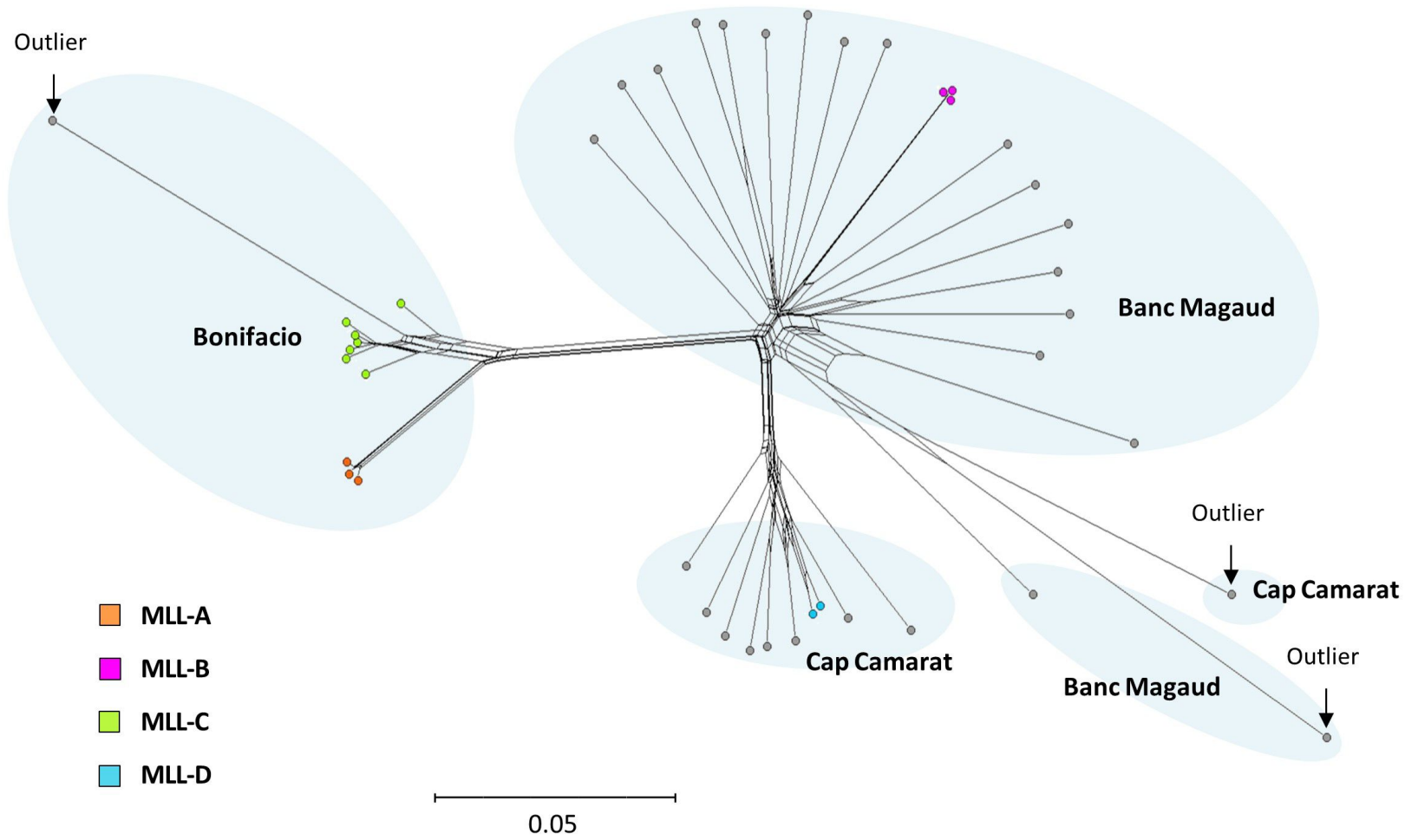
(A) *Laminaria rodriguezii*

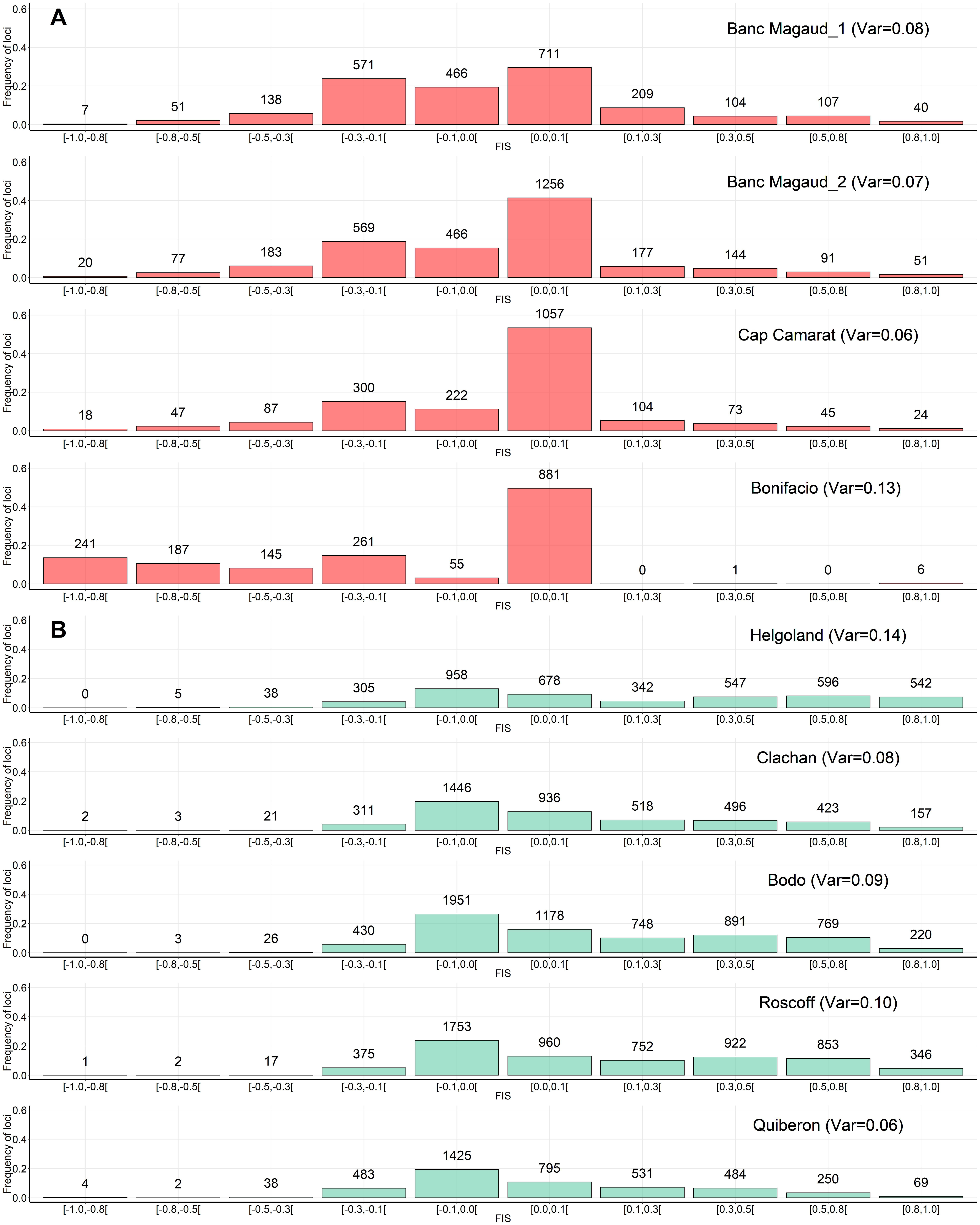


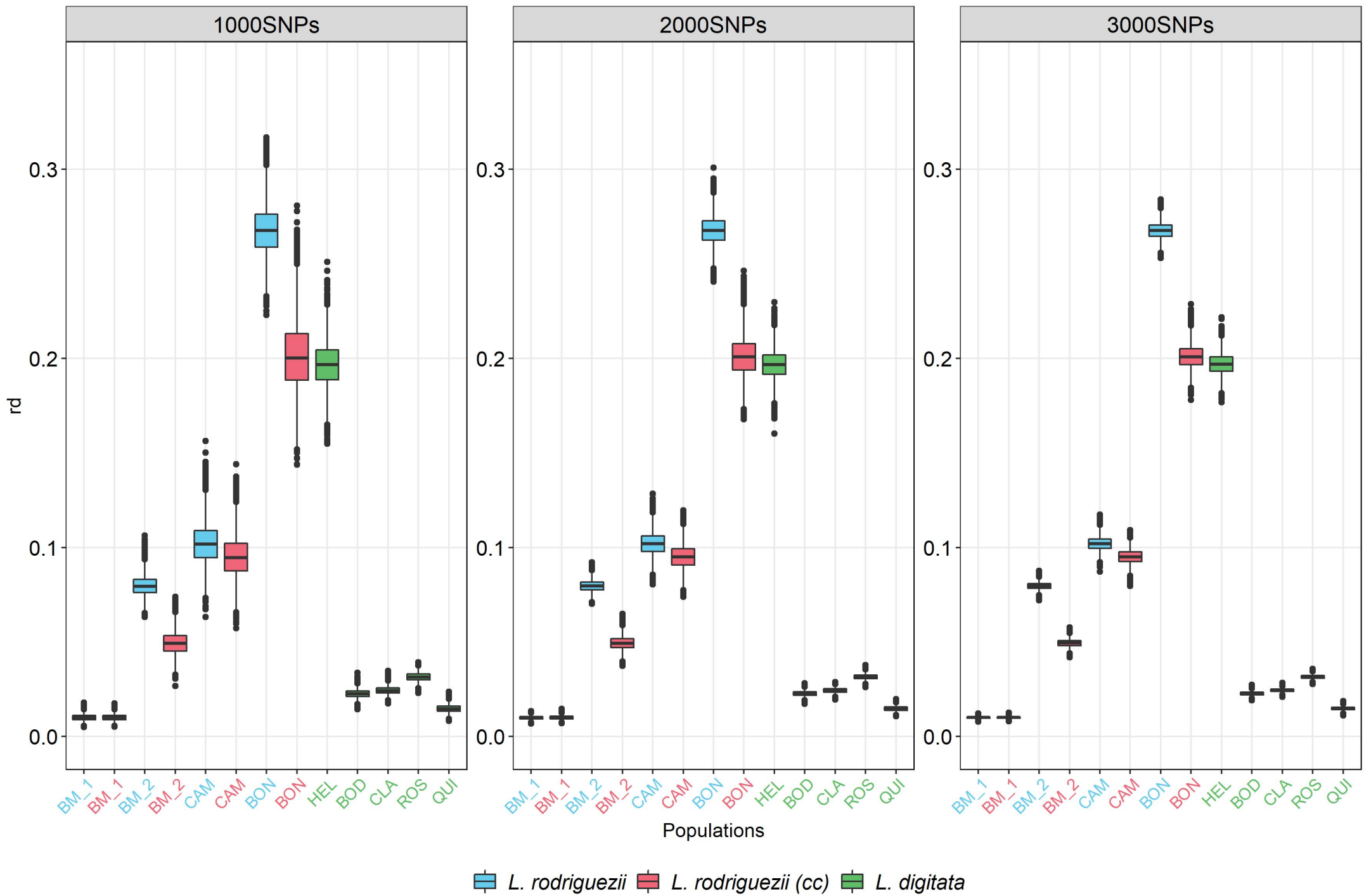
(B) *Laminaria digitata*



Intra_Hel Intra_pop Inter_pop Replicates







Supplemental Information for:

Genomic signatures of clonality in the deep water kelp *Laminaria rodriguezii*

Lauric Reynes, Thierry Thibaut, Stéphane Mauger, Aurélie Blanfuné, Florian Holon, Corinne Cruaud, Arnaud Couloux, Myriam Valero, Didier Aurelle

Table of Contents:

Supplementary materials

| | |
|---|---------|
| Table S1. Genetic variation within populations of <i>Laminaria digitata</i> . N, number of individuals successfully genotyped, Ho, observed heterozygosity, He, expected heterozygosity, FIS, average fixation index, $\overline{r_d}$, average multilocus linkage disequilibrium. FIS and $\overline{r_d}$ values highlighted in bold indicate significant heterozygote deficit within populations (P-val < 0.001) and significant linkage disequilibrium (P-val < 0.001), respectively. | Page S3 |
| Table S2. Genotypic richness (R) and Pareto (β) indexes estimated for populations of <i>Laminaria rodriguezii</i> . No β value is reported at Banc Magaud_1 due to the absence of clonal lineages. | Page S4 |
| Figure S1. Spatial analyses at Banc Magaud_2. (A) the pairwise genetic distances were those previously used to assess clonal identity and (B) the spatial distance matrix (m) | Page S5 |

| | |
|--|-------------------|
| <p>was generated from the positions of individuals within the transect. Mantel tests between spatial and genetic distances were significant ($R = 0.40$, $P < 0.001$) after 10 000 permutations. LRBM2P12, LRBM2P13, and LRBM2P14 belong to the same clone, while LRBM2P2 is one of the outliers highlighted in the study. (C) The linear regression between genetic and spatial distances by including (red) or excluding (blue) pairwise comparisons with the outlier.</p> | |
| <p>Figure S2. Genomic patterns of the three outliers as comparison to other individuals of <i>Laminaria rodriguezii</i>. (A) Percentage of private alleles per individual and (B) percentage of heterozygosity per individual.</p> | <p>Page S6</p> |
| <p>Figure S3. Average percentage of polymorphic loci (%P) within populations of (A) <i>Laminaria rodriguezii</i> and (B) <i>Laminaria digitata</i> obtained using a random sampling approach implemented in a custom bash script (available at https://github.com/rebecca-cj/revegetation). The percentage of polymorphic loci was calculated for each sample size from 100 random resamples. Standard error from the mean from replicates is represented as error bars.</p> | <p>Page S7</p> |
| <p>Materials S1</p> | <p>Pages S8-9</p> |

Table S1

| Population | Label | Lat. | Long. | N | H_o | H_e | F_{IS} | $\overline{r_d}$ |
|-------------------|--------------|-------------|--------------|----------|----------------------|----------------------|-----------------------|------------------------------------|
| Bodo | BOD | 67.2765 | 14.5747 | 23 | 0.14 | 0.18 | 0.23 | 0.02 |
| Clachan | CLA | 56.3171 | -5.5832 | 25 | 0.09 | 0.12 | 0.22 | 0.02 |
| Helgoland | HEL | 54.1779 | 7.8926 | 20 | 0.09 | 0.15 | 0.43 | 0.20 |
| Quiberon | QUI | 48.7275 | 4.0079 | 23 | 0.11 | 0.13 | 0.12 | 0.01 |
| Roscoff | ROS | 47.4700 | -3.0914 | 25 | 0.13 | 0.19 | 0.28 | 0.03 |

Table S2

| Population | β | R |
|-------------------|---------------------------|----------|
| Banc Magaud_1 | NA | 1.00 |
| Banc Magaud_2 | 2.00 | 0.80 |
| Cap Camarat | 3.32 | 0.90 |
| Bonifacio | 0.56 | 0.20 |

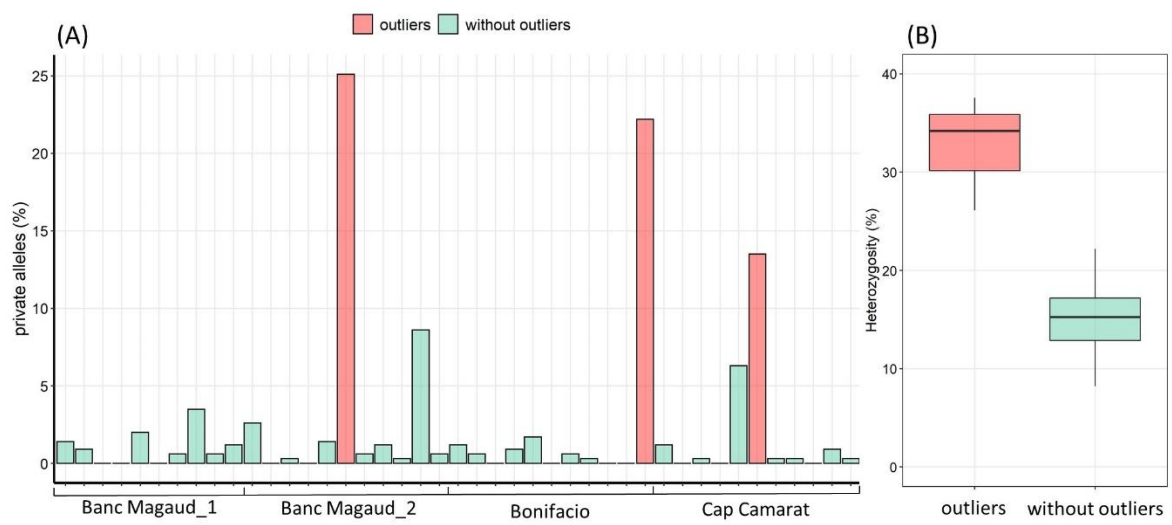
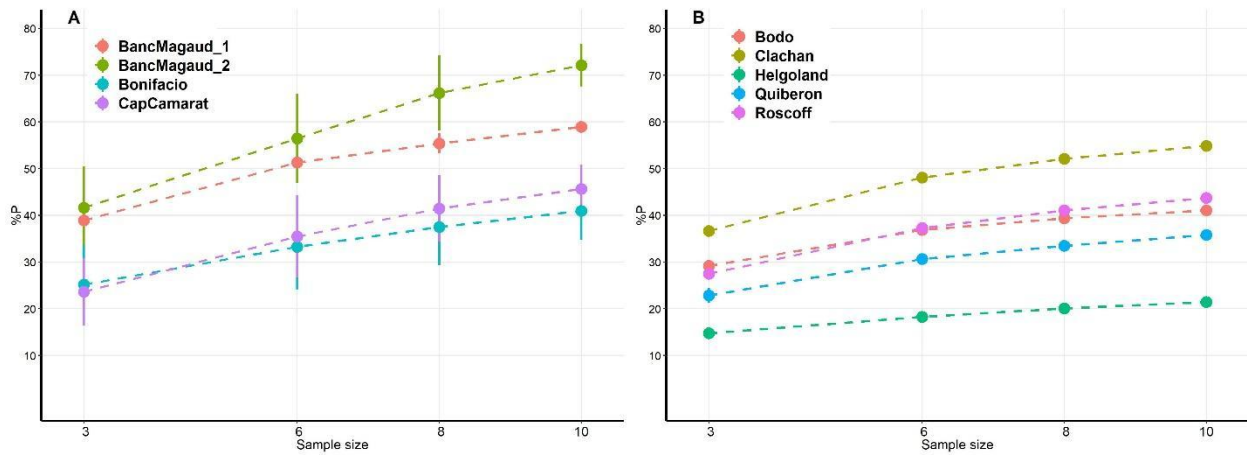


Figure S3



Materials S1

To address the effect of SNP filtering we started from the output of the catalog of 867 513 loci of *L. rodriguezii*, in which 283 864 SNPs were initially reported using the Stacks v2.4 bioinformatics pipeline (Rochette et al., 2019). Then, we selected a subset of informative SNPs by requiring a locus to be present at a frequency of greater than 80% at each of the four study sites. Additionally, we excluded SNPs with more than 80% heterozygosity and selected the first SNP of each RAD locus. As reported in the present study, we eliminated four individuals (with from 35% to 63% of missing genotypes) for which RAD-sequencing has been unsuccessful. At the end of this step, 21 357 informative SNPs were genotyped among the 43 individuals of the present study, thus constituting the reference dataset for the following analyses. We applied many filtering cutoffs (see Table 4) to the reference dataset, using VCFtools v0.1.17 (Danecek et al., 2011), thus producing 15 VCF files. Using the same software, we computed the number of SNP passing the filter (SNPs count), the mean missing data rate per individual (missingness), and the mean read depth per individual (coverage) for each dataset. Finally, we computed both F_{IS} and F_{ST} values (Weir & Cockerham, 1984) using the HierFstat R package (Goudet, 2005) and the mean SNP error rate between replicates with the R function developed by Mastretta-Yanes et al., (2015). We also considered a dataset including SNPs for which genotypes are strictly identical between replicates (matching genotypes) to consider the effect of potentially problematic SNPs on population genetic summary statistics. The results of these analyses are presented in the main manuscript (see Table 4).

REFERENCES

- 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
- Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184–186. doi: 10.1111/j.1471-8286.2004.00828.x
- Mastretta-Yanes, A., Arrigo, N., Alvarez, N., Jorgensen, T. H., Piñero, D., & Emerson, B. C. (2015). Restriction site-associated DNA sequencing, genotyping error estimation and de novo assembly optimization for population genetic inference. *Molecular Ecology Resources*, 15(1), 28–41. doi: 10.1111/1755-0998.12291
- Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, 28(21), 4737–4754. doi: 10.1111/mec.15253
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. doi: 10.1111/j.1558-5646.1984.tb05657.x