Species and population genomic differentiation in *Pocillopora* corals (Cnidaria, Hexacorallia)

Aurelle Didier [1, 2, 3, 15, *], Pratlong Marine [1, 4, 16], Oury Nicolas [5], Haguenauer Anne [2, 6], Gélin Pauline [5], Magalon Hélène [5], Adjeroud Mehdi [5, 6], Romans Pascal [7], Vidal-Dupiol Jeremie [8], Claereboudt Michel [9], Noûs Camille [10], Reynes Lauric [1], Toulza Eve [8, 17], Bonhomme François [11], Mitta Guillaume [12, 17], Pontarotti Pierre [13, 14]

[1] Aix Marseille Univ, Université de Toulon, CNRS, IRD, MIO, Marseille, France
[2] Aix Marseille Univ, Avignon Université, CNRS, IRD, IMBE, Marseille, France
[3] Institut de Systématique, Evolution, Biodiversité (ISYEB), Muséum National d'Histoire Naturelle, CNRS, Sorbonne Université, EPHE, 57 rue Cuvier, 75005, Paris, France
[4] Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France
[5] UMR ENTROPIE (Université de La Réunion, IRD, IFREMER, Université de Nouvelle-Calédonie, CNRS), Université de La Réunion, SaPSL UNIV PARISnt-Denis La Réunion, France
[6] PSL Université Paris, USR CRIOBE—EPHE-UPVD-CNRS, 3278, Perpignan, France
[7] Centre de Ressources Biologiques Marines, Sorbonne Universités, UPMC Univ Paris 06, Observatoire Océanologique, UMS 2348, Banyuls/Mer, France
[8] IHPE, Univ. Montpellier, CNRS, Ifremer, Univ. Perpignan Via Domitia, Montpellier France, Montpellier, France
[9] Department of Marine Science and Fisheries, College of Agricultural and Marine Sciences, Sultan Qaboos University, Al-Khod, 123, Muscat, Sultanate of Oman
[10] Laboratoire Cogitamus, Paris, France
[11] ISEM, Univ Montpellier, CNRS, IRD, EPHE, 34000, Montpellier, France
[12] EIO, Univ Polynesie Francaise, ILM, IRD, Ifremer, 98719, Tahiti, French Polynesia, France
[13] Aix Marseille Université, IRD, APHM, MEPHI, IHU Méditerranée Infection, 19–21 Boulevard Jean Moulin, 13005, Marseille, France
[14] SNC5039 CNRS, 19–21 Boulevard Jean Moulin, 13005, Marseille, France
[15] Campus de Luminy—Oceanomed, 13288, Marseille cedex 09, France
[16] Equipe Evolution Biologique et Modélisation, Marseille, France
[17] IHPE, Univ. Montpellier, CNRS, Ifremer, Univ. Perpignan Via Domitia, Perpignan, France

* Corresponding author : Didier Aurelle, email address : didier.aurelle@univ-amu.fr

Abstract :

Correctly delimiting species and populations is a prerequisite for studies of connectivity, adaptation and conservation. Genomic data are particularly useful to test species differentiation for organisms with few informative morphological characters or low discrimination of cytoplasmic markers, as in Scleractinians. Here we applied Restriction site Associated DNA sequencing (RAD-sequencing) to the study of species differentiation and genetic structure in populations of Pocillopora spp. from Oman and French Polynesia, with the objectives to test species hypotheses, and to study the genetic structure among sampling sites within species. We focused here on coral colonies morphologically similar to P. acuta (damicornis type β). We tested the impact of different filtering strategies on the stability of the results. The main genetic

differentiation was observed between samples from Oman and French Polynesia. These samples corresponded to different previously defined primary species hypotheses (PSH), i.e., PSHs 12 and 13 in Oman, and PSH 5 in French Polynesia. In Oman, we did not observe any clear differentiation between the two putative species PSH 12 and 13, nor between sampling sites. In French Polynesia, where a single species hypothesis was studied, there was no differentiation between sites. Our analyses allowed the identification of clonal lineages in Oman and French Polynesia. The impact of clonality on genetic diversity is discussed in light of individual-based simulations.

Introduction

Anthozoans, i.e. hexacorals and octocorals, are ecologically key species in various marine ecosystems, from tropical shallow reefs to deep ecosystems. They are the subject of numerous studies on the impact of climate change, as heat waves can lead to bleaching or necrosis events in tropical and temperate species (Garrabou et al. 2009; Hughes et al. 2018). Anthozoans are also important models in evolutionary biology, from phylogenetic studies to better understand their long-term evolution (Kayal et al. 2018; Pratlong et al. 2017b), to population genetic studies dealing with dispersal, parentage analysis or sex determinism (Underwood et al. 2007; Ledoux et al. 2010; Mokhtar-Jamaï et al. 2013; Pratlong et al. 2017a; Sheets et al. 2018). Population genetic studies should rely on adequate species delineation, as evidenced in studies assessing the diversity of thermotolerance and making inferences on population connectivity (Pante et al. 2015b; Brener-Raffali et al. 2019). Nevertheless, in Anthozoans species limits can be difficult to infer because of morphological plasticity, slow evolution of mitochondrial DNA, and hybridization (Calderón et al. 2006; Marti-Puig et al. 2014; Aurelle et al. 2017; Gélin et al. 2017b).

Hexacorals of the *Pocillopora* genus (Lamarck, 1816), such as the morpho-species *P. damicornis*, *P. grandis*, and *P. acuta*, are common corals found in shallow waters of the Red Sea, Indian and Pacific Oceans. The taxonomy of the *Pocillopora* genus is complicated by the effects of an important morphological variability, plasticity, and hybridization (Veron 2013; Schmidt-Roach et al. 2014). Considering these difficulties, several studies have used genetic data to refine species delineation in *Pocillopora*. Pinzón et al. (2013) used internal transcribed spacer 2 (ITS2), and the so-called mitochondrial ORF (which corresponds to the *tmp362* gene; Banguera-Hinestroza et al. 2019). The different lineages identified with these markers were not correlated with the morphology of *Pocillopora* corals, and there was no or reduced gene flow among lineages found in sympatry. Schmidt-Roach et al. (2014) made a taxonomic revision of the *P. damicornis* species complex on the basis of morphological characters, including micromorphology, and on mitochondrial ORF and nuclear HSP70B loci. These authors have shown a concordance between mitochondrial lineages and morphology, though the separation of some genetic lineages was in some cases blurred by morphological variability. Gélin et al. (2017b) put the study of species delineation within *Pocillopora* in a framework of testing species hypotheses. They proposed primary species hypotheses (PSH) based on mitochondrial DNA, and then tested secondary species hypotheses (SSH) with microsatellite data (see Figure 1 in Pante et al. 2015b, for the PSH and SSH process). Gélin et al. (2017b) have shown that PSH and SSH defined with molecular markers were not always congruent with species hypotheses based on morphology (i.e. morpho-species).

Genomic data can now be used to test species delineations on the basis of previous species hypotheses based on morphology or on a reduced number of markers. Restriction Sites Associated DNA sequencing (RAD-sequencing), allows the simultaneous discovery and genotyping of Single

113 Nucleotide Polymorphism (SNPs) in non-model organisms (Baird et al. 2008). RAD-sequencing

114 has been used to test species delineations in octocorals (Pante et al. 2015a) and hexacorals (Forsman

115 et al. 2017), including *Pocillopora*. The first RAD-sequencing study dealing with *Pocillopora*

116 corals suggested the possibility of hybridization of *P. damicornis* (Linnaeus, 1758) with *P. grandis*

117 Dana, 1846, and *P. elegans* Dana, 1846 (Combosch and Vollmer 2015). Based on RAD-Sequencing

118 analyses of seven *Pocillopora* species (*P. acuta, P. damicornis, P. grandis, P. ligulata, P. meandrina,*

119 *P. verrucosa, P.* sp. B), Johnston et al. (2017) found a good concordance with the phylogenetic

120 relationships inferred from mitochondrial DNA. Their results suggested a possibility of

121 hybridization between the closest sister species corresponding to mitochondrial haplotypes 4 (*P.*

122 *damicornis*) and 5 (*P. acuta*). RAD-sequencing has also been used to assess levels of intraspecific

123 variation in *P. damicornis,* demonstrating genetic by environment interactions, and therefore

124 potential local adaptation, when comparing populations in the Great Barrier Reef (van Oppen et al.

125 2018).

126 Additional genomic studies of *Pocillopora* spp. species and populations may provide insights into

127 important evolutionary questions regarding the population dynamics and species delineation of

128 corals. First, genomic data would allow testing more precisely the possibility of hybridization (i.e.

129 the existence of individuals with mixed ancestry) or introgression (directional gene flow) among

130 putative species, particularly in situation of sympatry. Second, on the basis of sound species

131 delineations, it would be useful to study levels of genetic differentiation and connectivity among

132 populations within species (Pante et al. 2015b). Third, RAD-sequencing could be used to study the

133 potential impact of clonality on the genomic diversity of these corals. Indeed, as in many other

134 scleractinian species, clonality has been demonstrated in *P. acuta* (Gélin et al., 2017a). It is expected

135 that clonality will have a significant impact on the populations genetic structure (Balloux et al.

136 2003; Adjeroud et al. 2014).

137 Here we used a hierarchical sampling design to study the genomic diversity of *Pocillopora* lineages

138 at different scales. Specifically, we sampled *Pocillopora* spp. in two distant regions located at the

139 margins of the distribution range of the genus: French Polynesia, Central Pacific Ocean, and Oman,

140 Northwestern Indian Ocean, with multiple sampling sites within each region. This sampling scheme

141 corresponded to a previous study on the diversity of thermotolerance in *Pocillopora* associated with

142 different thermal regimes (see Brener-Raffali et al. 2022). Despite sampling morphologically

143 similar colonies in French Polynesia and Oman, our samples included different mitochondrial

144 haplotypes associated with distinct putative species. In order to minimize taxonomic assumptions,

145 we will use throughout the manuscript the nomenclature of mitochondrial lineages from Pinzón et

146 al. (2013). We indicate in Table 1 the correspondence between the nomenclature of mitochondrial

147 lineages used in main text, the ORF haplotype number, the PSH and SSH defined based on Gélin et

148 al. (2017b), and potential nominal species. Specifically, in this study we mainly compared

149 individuals of mitochondrial lineages 5 (corresponding to *P. acuta;* PSH 5) in French Polynesia, and

150 3 and 7 *(corresponding to *P. verrucosa*; PSHs 13 and 12 respectively) in Oman (see Results).

151 Previous studies with microsatellite loci showed a distinction between mitochondrial lineages 5 and

152 3/7; however, mitochondrial lineages 3 and 7 were partly separated as distinct SSHs (Gélin et al.

153 2017b). Conversely, microsatellite data did not separate individuals with mitochondrial haplotypes

154 3 and 7 in the Red Sea and Arabian Gulf (Pinzón et al. 2013). We therefore applied RAD-

155 sequencing to these samples to test genetic differentiation among lineages identified with

156 mitochondrial sequences, and among sampling sites. We further simulated data with different levels

157 of clonal reproduction to help in the interpretation of results obtained with RAD-sequencing.

158

159 **Materials and methods**

160 ***Sampling and DNA extraction***

161 Five sites were sampled in Oman (hereafter identified as O1 to O5 sample sites, export CITES n°

162 37/2014 / import CITES n° FR1406600081-I), and six sites at two islands in French Polynesia (with

163 sites MH, MV and MT, at Moorea, and TF, TV and TT at Tahiti; export CITES n° FR1398700171-E

164 / import CITES n° FR1306600053-I). The list and location of sampling sites are presented in Table

165 2 and in Supplementary Figure S1. Thirty colonies were sampled in each site in Oman (except at

166 O3, which included 13 sampled colonies), and ten colonies per site in French Polynesia. The

167 sampling was focused on coral colonies morphologically similar to *P. acuta (damicornis* type *β*)-

168 like *corallum* morphology. Both in Oman and French Polynesia, we also sampled additional

169 colonies potentially belonging to other species, to be used as outgroups. The corresponding species

170 hypotheses were checked through sequencing of part of the mitochondrial DNA (see below).

171 After sampling, all colony fragments were bleached with menthol according to previously described

172 protocols (Wang et al. 2012; Vidal-Dupiol et al. 2020). Samples were then preserved in 95 %

173 ethanol and stored at -20 °C until DNA extraction. Total genomic DNA was extracted according to

174 the protocol described by Sambrook et al. (1989). After precipitation of DNA by isopropanol and

175 ethanol, DNA was resuspended in 50 µL water. This DNA solution was then purified with DNeasy

176 blood and tissue spin columns (Qiagen, Hilden, Germany) according to the manufacturer's protocol,

177 with elution in 100 µL water, which were put on the column for a second centrifugation. Genomic

178 DNA concentration was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad,

179 CA).

180

181 ***Mitochondrial ORF sequencing and microsatellite genotyping:***

182 Based on Gélin et al. (2017b), we used mitochondrial sequencing and microsatellite genotyping to

183 assign the colonies to Primary Species Hypothesis (PSH) and to Secondary Species Hypothesis

184 (SSH), respectively. The mitochondrial locus ORF was amplified with the FATP6.1 (5′-

185 TTTGGGSATTCGTTTAGCAG-3′) and RORF (5′-SCCAATATGTTAAACASCATGTCA-3′)

186 primers (Flot and Tillier 2007) and submitted for Sanger sequencing in both directions. GenBank

187 accession numbers for all mitochondrial haplotypes used for PSH definition (obtained here and

188 from previous works) are indicated in Gélin et al. (2017b). Nucleotide sequences were analyzed

189 using MEGA version 6 (Tamura et al. 2013). We built a network of ORF sequences with the

190 minimum spanning method using PopART v.1.7 (Bandelt et al. 1999; Leigh and Bryant 2015).

191 Haplotypes were colored and sized according to localities, and numbers of occurrences,

192 respectively. Characterization of haplotypes were based on comparisons with the sequences used in

193 Gélin et al. (2017b). Additionally, a subset of colonies (N = 165) were genotyped with 13

194 microsatellite loci, and assigned to SSHs with Bayesian clustering as described in Gélin et al.

195 (2017b). Furthermore, as PSH05 (*P. acuta* or *P. damicornis* type *β*) is known to propagate asexually

196 (Gélin et al. 2017a, 2018), microsatellite genotyping was used to search for repeated multilocus

197 genotypes (MLGs) as a benchmark for the delineation of clonal lineages with RAD-sequencing (see

198 Gélin et al. 2017a for description of methods).

199

200 ***RAD sequencing and analyses***

201 The preparation and sequencing of RAD-sequencing libraries was performed with the PstI

202 restriction enzyme as described in Pratlong et al. (2021). We started with an initial number of 211

203 individuals, distributed among seven libraries. Libraries were sequenced on an Illumina HiSeq2000

204 using 100 bp single-end reads, at the Biology Institute of Lille (France, IBL, UMR 8199 CNRS)

205 and at the MGX sequencing platform in Montpellier (France). The sequences were first

206 demultiplexed, filtered by quality, and searched for adapters contamination using iPyrad v0.7.28

207 with default parameters (Eaton and Overcast 2020; https://ipyrad.readthedocs.io/). We then checked

208 the quality of the sequences and the absence of adapters with FastQC (Andrews 2010). The

209 assembly of RAD loci was performed with Stacks 2.3 (Catchen et al. 2013). We used a published

210 genome of *P. damicornis* (Cunning et al. 2018) as a reference to map the reads with BWA using

211 default parameters (Li and Durbin 2009). The mean length of RAD loci was 95 bp. At that step, the

212 numbers of reads and percentages of missing data were very uneven among the 211 samples (see

213 Results). As a consequence, we chose to use different assembling strategies leading to different

214 datasets. Following preliminary assembly analyses, we removed individuals with less than 900 000

215 reads, as their inclusion led to datasets with very low numbers of SNPs. The resulting dataset

216 comprised 140 individuals, including 104 from Oman and 36 from French Polynesia. Then we used

217 the module populations of Stacks to assemble three datasets: 1) one considering all 140 individuals

218 grouped by mitochondrial lineage ("All" dataset), 2) one with only samples from Oman and

219 grouping by mitochondrial lineage, and 3) one with only samples from French Polynesia and

220   grouping by sampling sites (apart from outgroups, all individuals from French Polynesia shared the
221   same mitochondrial haplotype; see Results). In the populations module we removed SNPs with
222   allele frequencies lower than 0.01, and sites which were present in less than 50% of individuals in
223   each group (i.e. mitochondrial lineage or sampling site depending on the dataset). At that stage, the
224   individuals of the All and Oman datasets for which we did not get any mitochondrial sequence were
225   put in an additional group for assembly ("unknown"). We then retained the first SNP of each RAD
226   locus with a custom script. Second,we used Tassel 5.0 (Bradbury et al. 2007) to filter the
227   corresponding VCF files according to missing data, with two consecutive filters on missing data in
228   each dataset. In the first case, we first retained loci present in at least 75 % of the individuals. Then
229   we retained individuals which had genotypes for at least 75 % of the loci. This first strategy resulted
230   in less than 25 % of missing data both for loci and individuals (hereafter strategy 75-75). In the
231   second case, we first retained loci present in at least 95 % of the individuals, and then individuals
232   with data for at least 75 % of the loci (hereafter strategy 95-75). These filtering led to the removal
233   of the less informative samples or loci. The 75-75 strategy allowed to retain more loci with less
234   individuals compared to the 95-75 strategy (Table 3). The final sampling sizes per sampling site and
235   mitochondrial lineage for the six datasets (with the two strategies applied to the whole dataset, and
236   separately to Oman and French Polynesia datasets) are described in Supplementary Table S1.

237

238   ***Phylogenetic inference and genetic differences among individuals***

239   We used two complementary approaches to study the genetic differences among individuals with
240   RAD-sequencing data. First we performed a phylogenetic reconstruction. For that purpose, among
241   the 140 individuals retained for the processing of RAD-sequencing data, we chose up to 10
242   representatives of each mitochondrial lineages (those with the highest number of raw reads), and
243   three outgroups: we thus limited potential biases due to uneven representations of lineages in the
244   phylogenetic inference. We used the population module of Stacks to build a Phylip file, including
245   all sequences as well as variable sites, with only loci present in all individuals, and sites with allele
246   frequencies above 0.01. We then used IQ-TREE 2.1.1 for phylogenetic inference, with the
247   following options: ModelFinder Plus, 1 000 boostraps with the ultrafast boostrap approximation,
248   and the hill-climbing nearest neighbor interchange search (Minh et al. 2013; Hoang et al. 2018).
249   The resulting tree was formatted using FigTree 1.4.4 (Rambaut 2012). Second we built networks of
250   mitochondrial haplotypes with the NeighborNet option of SPLITSTREE 4.14.8 (Huson and Bryant
251   2006). The distance matrix used to build networks was the percentage of nucleotidic divergence
252   among individuals computed with the poppr R package (Kamvar et al. 2014). This network
253   reconstruction was used to analyze the previously defined 75-75 and 95-95 filtered datasets, first
254   including all samples, and then separately for the Oman and French Polynesia samples.

255

Population genetic analyses

257 We first estimated genetic diversity on the dataset not corrected for the presence of repeated

258 multilocus genotypes (MLGs; see below). We used the GENEPOP R package (Rousset 2008) to

259 compute gene diversity within individuals (*1-Qintra*; corresponding to observed heterozygosity)

260 and among individuals within sampling sites or lineages (*1-Qinter*; corresponding to expected

261 heterozygosity), and F_{IS} (Weir and Cockerham 1984). We used VCFTOOLS 0.1.15 (Danecek et al.

262 2011) to compute an estimate of inbreeding coefficient, *F*, which compares the observed number of

263 homozygous sites to its expectation under panmixia.

264 We tested the presence of repeated multilocus genotypes (MLGs) and multilocus lineages.

265 Multilocus lineages (MLLs) correspond to genotypes separated by a varying number of mutations

266 and reflecting apparent divergence among MLGs either because of sequencing or genotyping errors,

267 or because of somatic mutations. We used the R package poppr to analyze MLGs and MLLs

268 (Kamvar et al. 2015). The choice of thresholds to delineate MLLs was made according to two

269 criteria: first, we used MLGs obtained with microsatellite loci (data not shown) for a subset of

270 individuals to define an MLL threshold. We also used the distribution of genetic distances among

271 individuals to look for lowly differentiated individuals that could belong to the same MLL. The

272 genetic distances among individuals were measured by the percentage of nucleotide divergence

273 computed with poppr. According to their respective levels of diversity, the retained MLL threshold

274 was different for the different datasets (see Results).

275 One representative of each MLL was kept for clustering and F_{ST} analyses. We analyzed the genetic

276 disequilibrium among loci by computing the modified index of association $\bar{r}_d$ (Agapow and Burt

277 2001) with the poppr R package. To keep reasonable computing time, we first randomly

278 subsampled the All_75_75 and Oman_75_75 to 25 000 SNPs. Then we computed $\bar{r}_d$ on datasets

279 comprising randomly subsampled 200 SNPs (this number allowed enough different resampling with

280 the smallest dataset), and with 10 000 repetitions of this subsampling. With this approach we could

281 analyze the linkage disequilibrium in datasets with a high number of SNPs. To take into account the

282 impact of subpopulation structure (i.e. Wahlund effect) on this analysis of linkage disequilibrium,

283 we performed the analysis at two levels (i.e. the "strata" levels used in poppr): first at the level of

284 the whole corresponding dataset, and second at the level of mitochondrial lineages for the All and

285 Oman datasets, or of sampling sites for the French Polynesia and Oman datasets.

286 Genetic differentiation among populations was measured with the F_{ST} estimator of Weir and

287 Cockerham (1984) computed with VCFTOOLS. The differentiation among individuals was

288 visualized based on a Principal Component Analysis (PCA) performed with the R package adegenet

289 (Jombart 2008). Missing data were replaced by the mean allele frequency as in the adegenet tutorial

290 (https://adegenet.r-forge.r-project.org/files/tutorial-basics.pdf). As a complementary analysis to

291 PCA, in order to identify the main genetic groups in the dataset, we analyzed the partition in K

292 independent units with the snmf function of the R package LEA (Frichot and François 2015). This

293 approach performs a least squares estimates of ancestry proportions (Frichot et al. 2014). We tested

294 K values from 1 to 10, with ten replicates for each K value.

295

296 *Simulations*

297 To help the interpretation of our results on individual inbreeding coefficient F, on $\bar{r}_d$ and on F_{ST}, we

298 performed simulations to analyze the variability of these estimates, with a focus on the impact of

299 partial clonality. We used SLiM 3 to build genetically explicit individual-based simulations (Haller

300 and Messer 2019). We simulated two populations, each with 100 individuals, and connected

301 through reciprocal gene flow at a rate of m = 0.01 per generation. The genetic data were modeled

302 with 2 000 loci of 100 bp each, mutating at a rate of $\mu = 10^{-4}$ mutation per site per generation. This

303 high mutation rate is a way to model enough genetic diversity with a moderate number of

304 individuals and memory usage. After an initialization phase of 5 000 generations with panmixia

305 within each population, we performed 50 000 generations with one of the following reproductive

306 modes: panmixia (within population), clonality at a rate of c = 0.1, 0.5 or 0.9, selfing at a rate of s =

307 0.1, and a combination of 0.1 selfing rate and clonality rates of 0.1 or 0.5. At the end of the

308 simulations, 30 simulated individuals were sampled per population, and 30 replicates were

309 performed for each simulation configuration. The output VCF files were analyzed with

310 VCFTOOLS to compute the estimate of individual inbreeding coefficients F and F_{ST}. For each

311 simulation we computed the mean, minimum and maximum values of F and F_{ST} over individuals

312 and loci, respectively. We computed $\bar{r}_d$ separately on each of the two population of the simulations.

313 For computing reasons, the mean and standard deviation of $\bar{r}_d$ were computed with 50 re-samplings

314 of 1 000 SNPs.

315

316 **Results**

317 ***Assignment to species hypotheses according to mitochondrial sequences and microsatellite***

318 ***genotypes:***

319 Out of the 140 individuals retained in the final RAD-sequencing dataset, we did not get any usable

320 mitochondrial sequence for 18 individuals (three from French Polynesia and 15 from Oman). The

321 sequences obtained in this study allowed a clear assignment of individuals to previously defined

322 sequence groups and corresponding primary species hypotheses. The network of mitochondrial

323 sequences is presented in Figure S2. In French Polynesia, two individuals sampled as outgroups on

324 the basis of morphology were highly divergent from individuals characterized by RAD-Seq, and

325 corresponded to mitochondrial lineages type 1a and type 2. The high divergence of these individuals

326 from other samples blurred the analysis of the differentiation among the other lineages with RAD-

327 sequencing data, especially on multivariate analyses (data not shown). Giving this signal and the

328 small sample size for these outgroups, we did not retain them in the following analyses. Apart from

329 these two individuals, French Polynesia included only samples from mitochondrial lineage 5a (PSH

330 5). Oman included samples from mitochondrial lineages 7a (PSH12), 3e and 3g (both in SSH 13a),

331 and only one individual from mitochondrial lineage 5a (Supplementary Table S1). The assignments

332 to species hypotheses were consistent with those inferred from microsatellite loci (data not shown).

333 With microsatellites, we did not detect any repeated multilocus genotype (MLG) in French

334 Polynesia; i.e., each individual corresponded to a unique 13 loci genotype. Among individuals from

335 mitochondrial lineage 7a in Oman, which included over 64 individuals for which we got a complete

336 13 loci genotype, only 53 distinct MLGs were retrieved, with one MLG repeated five times, another

337 one four times in O2, and three MLGs repeated two times (one in O2 and two in O5). Among

338 individuals from mitochondrial lineage 3g, one MLG was repeated two times in O5.

339

340 *RAD sequencing data*

341 The initial number of sequences obtained per individual was very uneven among samples, varying

342 from 5 735 to 30 394 029 reads (Supplementary Table S2; Figure S3A). The mean number of reads

343 per individual was higher for samples from Oman (mean 5 647 233) compared to French Polynesia

344 (mean 2 309 860). The percentage of reads aligned to the *Pocillopora* genome was more regular,

345 with a mean of 85.1 %, but it was still higher in Oman (mean 85.6%) than in French Polynesia

346 (84.1%; Table S2; Figure S3B). The lowest percentages of alignment were obtained for the

347 outgroup samples. This heterogeneity among samples and sampling regions motivated our different

348 assembly strategies. Table 3 presents the characteristics of the six final datasets, with a total of 140

349 samples distributed among the different datasets. The highest numbers of SNPs, with one SNP per

350 RAD locus, were obtained for the 75 % filtering on loci missing data in the All (194 370 SNPs) and

351 Oman datasets (134 307 SNPs). The separate assembly of Oman and Polynesia allowed the

352 recovery of more SNPs than the All assembly with the 95_75 strategy, whereas these two separate

353 assemblies led to less SNPs than the All assembly with the 75_75 strategy. The All_95_75 dataset

354 had the highest number of individuals (132) and the lowest number of SNPs (320; Table 3).

355

356 *Genetic differences among individuals and repeated MLLs*

357 The ML tree of relationships among individuals representative of the different mitochondrial

358 lineages is presented in Figure S4. Apart from the two outgroups from French Polynesia, the main

359 separation corresponded to the split between samples from French Polynesia, with mitochondrial

360 lineage 5a, and samples from Oman, with mitochondrial lineages 3e, 3g and 7a. This divergence

361   corresponded to highly divergent monophyletic groups well supported by bootstraps analyses

362   (100%). There was no clear phylogenetic grouping by mitochondrial lineage among the samples in

363   Oman.

364   The networks based on the percentages of differences among individuals with RAD-sequencing are

365   presented in Figure 1 for the 95_75 datasets which had the highest number of individuals. The

366   corresponding histograms of the distribution of genetic distances are presented in Supplementary

367   Material (Figure S5). The All_95_75 network showed a clear distinction between Oman and French

368   Polynesia individuals, which was associated with differences in mitochondrial lineages, separating

369   type 5a versus all other lineages. The single colony of type 5a sampled in Oman (O4_21) grouped

370   with other colonies from Oman in this network. In the separate analysis of Oman and French

371   Polynesia, there was no clear sub-grouping according to sampling location or mitochondrial

372   lineages (Figure 1). Similar results were obtained for the networks based on the 75_75 datasets,

373   except in French Polynesia where we observed small clusters of individuals from the same site,

374   albeit without clear differentiation from other groups (Figure S6).

375   The histograms of pairwise differences among individuals in Oman, and to a lesser extent in French

376   Polynesia, showed a peak of low value distances, potentially reflecting repeated MLLs (Figure S5).

377   We used the distance observed with RAD-sequencing among individuals sharing identical

378   microsatellite MLGs as a threshold to define MLLs in the Oman dataset. In Oman, the highest

379   distance among individuals with repeated microsatellite MLGs reached 11% which seemed too high

380   to define MLLs. Therefore, in this case we retained the second highest distance among

381   microsatellite-defined MLGs to characterize MLLs. It should be noted that this did not change the

382   main results of this study, as only a few pairwise comparisons were then removed from potential

383   MLLs. We could not use a single threshold for all datasets because the levels of divergence differed

384   considerably between the datasets (95_75 vs 75_75, and French Polynesia vs Oman), and we had no

385   repeated microsatellite MLG to be used as a reference in some datasets. Therefore, in cases where

386   no threshold could be defined on the basis of microsatellites, we used a threshold allowing the

387   removal of the closest individuals, as indicated by preliminary tests and by observation of the

388   distribution of pairwise distances among individuals. The number of individuals for each corrected

389   dataset is given in Table 3, and the corresponding thresholds are indicated below the distribution of

390   individual distances (Figure S5).

391

392   ***Heterozygosity and inbreeding coefficients***

393   The parameters of genetic diversity for the different datasets are presented in Table 4. Separate

394   estimates of genetic diversity per lineage and site are presented as Supplementary Materials (Table

395   S3). The *1-Qintra* and the *1-Qinter* statistics indicated a higher genetic diversity in the 75_75

396   datasets compared to the 95_75 ones. For a given assembly strategy, the highest levels of diversity

397   were observed in Oman, followed by the All and the French Polynesia datasets.

398   The F_{IS} were mainly null or positive, apart from the All_95_75 (F_{IS} = -0.15) and the

399   Polynesia_95_75 (F_{IS} = -0.21) datasets, but with important variations among lineages or sites,

400   especially in French Polynesia (Table S3). The estimates of individual inbreeding coefficient F gave

401   highly variable and extreme values (Table 4). The distributions of the F estimates illustrate this

402   wide dispersion, and the shift to more positive values from 95_75 to 75_75 datasets (Figure S7). In

403   the All_95_75 dataset, the F values in Oman (from -0.278 to 0,546) were higher than in French

404   Polynesia (from -1,306 to -0.303). There was no general signal towards higher or lower F values for

405   individuals involved in potential MLLs (i.e. individuals involved in the closest pairwise

406   relationships in the different datasets), apart for the French Polynesia datasets, where these

407   individuals showed among the highest F values (Figure S7).

408

409   ***Linkage disequilibrium***

410   The results of the analysis of linkage disequilibrium with the $\bar{r}_d$ index are presented as

411   Supplementary Materials (Table S4 and Figure S8). For the analyses at the level of the whole

412   datasets, the highest $\bar{r}_d$ values were obtained in the All_75_75 dataset, followed by the French

413   Polynesia_95_75 and All_95_75 datasets. The values obtained in French Polynesia were higher

414   than those in Oman for the 95_75 and the 75_75 datasets. When the analysis was performed at the

415   level of mitochondrial lineages or sites, the highest values were observed for the MH site in French

416   Polynesia both for the 95_75 and the 75_75 datasets (Table S4, Figure S8).

417

418   **F_{ST}** ***estimates***

419   The mean F_{ST} estimates among loci were generally lower for the datasets corrected for MLLs

420   compared to the non-corrected datasets (Table 5), except for the comparison among mitochondrial

421   lineages with All_95_75, and the comparison among populations with All_75_75. The mean F_{ST}

422   between French Polynesia and Oman (by grouping samples from each region) was 0.105 and 0.352

423   for the All_95_75 and the All_75_75 non-corrected datasets, respectively. The distributions of F_{ST}

424   among loci for the Oman / French Polynesia comparison are presented in Figure S9. For this

425   comparison, the All_75_75 non-corrected dataset showed an important proportion of loci with F_{ST}

426   above 0.2, and a peak at F_{ST} = 1, whereas the distribution was mainly restricted to values below 0.2

427   for the All_95_75 non-corrected dataset. The F_{ST} estimates for all datasets indicated very low levels

428   of differentiation within Oman and within Polynesia (Table 5).

429

430   ***Analysis of genetic structure***

431   Multivariate PCA of the All_95_75 and All_75_75 datasets separated the samples from French

432   Polynesia and Oman, with individuals from French Polynesia (type 5a lineage) being more spread

433   apart than those from Oman (Figure S10). The separate PCAs on the Oman and French Polynesia

434   datasets did not reveal any clear structure according to sampling site nor mitochondrial lineage,

435   whatever the filtering strategy. The plots of cross-entropy for all snmf analyses are presented in

436   Figure S11. There was no clear signal for an informative K value on the basis of these cross-entropy

437   plots, apart for the All_95_75 dataset where a first minimum value was observed at K = 2 and a

438   second at K = 5, and for Polynesia_95_75 where a slight minimum was observed at K = 2. We then

439   also analyzed the results with a K value corresponding to the number of mitochondrial lineages (for

440   All and Oman) or the number of sampling sites (for French Polynesia and Oman). The

441   corresponding barplots of coancestry coefficients for the 95_75 datasets are given in Figure 2. For

442   All_95_75, the K = 2 solution clearly separated French Polynesia (mitochondrial lineage type 5a)

443   and Oman samples (other mitochondrial lineages, with one 5a exception). At K = 4, two additional

444   sub-clusters were observed, one in Oman and one in French Polynesia. For the Oman_95_75

445   dataset, the K = 4 solution (corresponding to the number of mitochondrial lineages) led to a major

446   and three minor clusters. These clusters did not separate individuals neither by mitochondrial

447   lineage nor by sampling site. This clustering was nevertheless partly linked with potential MLLs, as

448   the purple cluster of Oman_95_75 grouped individuals of the MLG02 identified with microsatellites

449   (Figures 1 and 2). The K = 5 solution for Oman (corresponding to the number of sites), did not

450   evidence any informative clustering either, again with the separation of minor clusters mainly

451   composed of individuals closely grouped in the network analysis (results not shown). For

452   Polynesia_95_75 the K = 2 and the K = 5 solutions did not separate individuals neither by sampling

453   site type nor by MLL. The snmf analysis of genetic structure with the 75_75 datasets gave similar

454   results, with a separation of Oman and French Polynesia at K = 2 for All_75_75, and a further

455   distinction of a few individuals in two additional clusters in Oman at K = 4 (results not shown).

456

457   ***Simulations***

458   The results of the analyses of individual-based simulations are detailed in the Supplementary

459   Materials section, with a comparison with empirical data. The main results of these analyses

460   revealed that the maximum F values tended to be higher for the configurations with the highest

461   clonality rates (from 0.5), but were much higher for the simulations including selfing. Regarding the

462   minimum F value, a decrease in the distribution was observed for the highest levels of clonality,

463   with more negative values compared to other configurations. An increase in the index of linkage

464   disequilibrium $\bar{r}_d$ was observed with increasing clonality rate, mainly for the highest clonality rate

465   (0.9). Regarding the average F_{ST}, without any variation neither in census size nor migration rate, the

466 resulting values were mostly similar among simulation configurations. A slight decrease and higher
467 variance was nevertheless observed for the highest clonality rate (0.9) simulated.
468
469

470 **Discussion**
471 *Genomic analysis of species hypotheses*
472 When considering the All datasets, we observed a marked differentiation between French Polynesia
473 and Oman populations, which was superimposed on a differentiation between mitochondrial lineage
474 5a and other lineages. The mitochondrial lineage 5a corresponds to *P. acuta*, and is phylogenetically
475 well separated from lineages 3e-3g and 7a, which correspond to *P. verrucosa* (Gélin et al. 2017b).
476 The distinction of the corresponding species hypotheses based on mitochondrial lineages was
477 previously confirmed with microsatellite loci (Gélin et al. 2017b), and here with RAD-sequencing.
478 Nevertheless, our sampling scheme did not allow to test the possibility of hybridization of lineage
479 5a with other lineages in sympatry. There was only one individual bearing mitochondrial type 5a in
480 Oman for which we were able to get RAD-sequencing data: this individual did not separate from
481 other individuals in Oman with different mitochondrial types. This last observation could indicate a
482 possible introgression of mitochondrial type 5a into the 3e-3g-7a gene pool, but this should be
483 tested by considering additional 5a individuals from Oman. We obtained six additional individuals
484 from this lineage, hence indicating that the single 5a haplotype reported here is not likely a result of
485 contamination; however, the RAD sequencing of these samples was not good enough to retain them
486 and confirm a possible introgression.
487 Conversely, we did not observe a differentiation between individuals from Oman assigned to the
488 species hypotheses corresponding to mitochondrial lineages 7 and 3e-3g (i.e., PSH12 and SSH13a
489 in Gélin et al. 2017b, respectively). All methods of species delineation based on mitochondrial DNA
490 used in Gélin et al. (2017b) indeed separated these two species hypotheses. Nevertheless,
491 microsatellite data did not support this distinction (Gélin et al. 2017b; Pinzón et al. 2013). One can
492 note that in the case of Pinzón et al. (2013) this lack of differentiation was observed in the Red Sea
493 and in the Arabian Gulf. With a sampling in sympatry and based on our RAD-Sequencing, we also
494 reject the distinction of PSH12 and SSH13a. The most parsimonious hypothesis here would be that
495 these different lineages correspond to mitochondrial polymorphism present within a given species,
496 here *P. verrucosa*, even when mitochondrial DNA in anthozoans has been shown to evolve slowly
497 (van Oppen et al. 1999; Calderón et al. 2006; Hellberg 2006). Accordingly, one can note that 12
498 over 16 species delineation methods based on mitochondrial ORF did not conclude to separate
499 PSHs for the lineages 3e and 3g (Gélin et al. 2017b). Another possibility could be genetic swamping
500 (e.g. Bog et al. 2017; Kosiński et al. 2019) following a secondary contact between different lineages
501 in Oman. Reticulate evolution has already been proposed as a major factor shaping the current

502 diversity of scleractinian corals (van Oppen et al. 2001; Vollmer and Palumbi 2002). A more precise
503 analysis of the genomic patterns of differentiation would be useful to study the existence, timing
504 and direction of introgression in *Pocillopora* spp. (e.g. Nelson et al. 2020). Our sampling was also
505 not random, with a focus on morphologically similar colonies. We excluded a few very divergent
506 colonies and therefore we did not explore the whole diversity of *Pocillopora* in the studied areas. A
507 more extended sampling in sympatry, random in regard of morphology, would be essential to get a
508 better understanding of the levels of genomic differentiation and interactions among the different
509 *Pocillopora* species.

510 In French Polynesia, neither of the datasets showed evidence of genetic structuring among sites,
511 which were distributed in the two islands of Moorea and Tahiti. These results are consistent with
512 previous studies on the genetic structure of *Pocillopora* (Magalon et al. 2004; Adjeroud et al. 2014).
513 One can note that in *Pocillopora* corals the patterns of genetic structure are evidently dependent on
514 the species hypothesis and locations considered (Oury et al. 2020b). For example, Gélin et al.
515 (2018) observed a high genetic differenciation for *P. acuta* (PSH05 in Gélin et al. 2017b) at
516 different spatial scales in the Western Indian and Tropical Southwestern Pacific Oceans. Similarly,
517 Combosch and Vollmer (2011) have shown a significant genetic structure in *P. damicornis* at a
518 spatial scale of a few tens of kilometers, but this was not a general result for all population
519 comparisons. Conversely, Oury et al. (2021), Robitzch et al. (2015) and Thomas et al. (2014),
520 observed low genetic structure along high distances in different *Pocillopora* lineages (with some
521 exceptions reported by Oury et al., 2021).

522

523 *Impact of filtering strategies on RAD-sequencing data*
524 Our RAD-sequencing analyses provided a dataset with very uneven levels of missing data. High
525 levels of missing data, if not accounted for, can lead to incorrect conclusions regarding genetic
526 structure (Larson et al. 2021). Missing data in RAD-sequencing can have multiple origins, including
527 mutations in enzyme-cutting sites, technical problems associated with library preparation, uneven
528 amplification or sequencing, or errors in the *in silico* identification of homologous sites (Eaton et al.
529 2017; O'Leary et al. 2018). In our study, the lowest read numbers were obtained for outgroup
530 samples, corresponding to mitochondrial lineages 1a and 2, as opposed to mitochondrial lineages
531 5a, 7, 3e and 3g, which were the most frequent in our datasets. For example, seven of these
532 outgroup samples had only around 20 000 reads or less and thus were not retained in our analyses.
533 We did not observe any relationship between the number of raw reads and the number of reads
534 mapped to a Symbiodiniaceae genome (data not shown); therefore, a potential contamination from
535 dinoflagellate genomes cannot explain these results. Despite standard verifications, problems with
536 DNA quantity and quality may have impacted the number of reads obtained, such as for example
537 the presence of partially degraded DNA or PCR inhibitors (O'Leary et al. 2018).

538   Facing these difficulties, we compared different strategies for filtering missing data, which can have

539   important consequences on the obtained results. Our study showed that some results were stable

540   among the different filtering strategies. For example, there was a marked differentiation between

541   samples from Oman and French Polynesia, associated with the species sampled in each region, and

542   a lack of genetic structure among sites within each region. Conversely, the estimates of genetic

543   diversity and structure differed among datasets. The genetic distances among individuals were much

544   higher with more loci in both the All and Oman datasets.

545   If part of missing data are linked to mutations in the cutting sites, filtering loci according to their

546   rates of missing data is expected to reduce the frequency of loci with high mutation rates (Huang

547   and Knowles 2016). This agrees well with our observation of a lower diversity detected under more

548   stringent filtering. Missing data can also lead to allele dropout, which corresponds to the non-

549   observation of a SNP linked to a mutated restriction site. For the retained variable loci, an

550   overestimation of heterozygosity can be expected with allele dropout, particularly if the dropout

551   concerns more ancestral allelic states and therefore leads to an increase in minimum allele

552   frequencies (Gautier et al. 2013). Consistent with this expectation, our analysis showed an increase

553   in heterozygosity for the 75_75 compared to the 95_75 datasets. Finally, allele dropout can

554   overestimate F_{ST} (Gautier et al. 2013), which seems consistent with our results, with higher F_{ST} for

555   the 75_75 compared to the 97_75 All datasets. Regarding linkage disequilibrium, the increase in $\bar{r}_d$

556   observed for All_75_75 compared to All_95_75 can result from a combination of higher Wahlund

557   effect (with the inclusion of more differentiated loci), and of a higher number of physically linked

558   loci. An important question is whether the variance of polymorphism among loci is high enough to

559   explain the observed differences through the aforementioned effects.

560

561   ***Signals of clonality with RAD-sequencing data***

562   Populations of *Pocillopora* corals, notably in *P. acuta* (mitochondrial lineage 5) and *P. damicornis*

563   (mitochondrial lineage 4), can show different levels of clonal reproduction (Pinzón et al. 2012;

564   Torda et al. 2013; Adjeroud et al. 2014), with sometimes different ramets of the same genet

565   separated by several kilometers (Gélin et al. 2017a, 2018). Clonal reproduction in these species can

566   happen through fragmentation of individuals, polyp bail-out or asexual production of larvae

567   (Highsmith 1982; Gélin et al. 2017a; Oury et al. 2019). Clonal reproduction can lead to

568   heterozygote excess compared to panmixia (Balloux et al. 2003; Reichel et al. 2016), and to shifts

569   in the distribution of F_{IS} among loci towards negative values for the highest rates of clonality

570   (Stoeckel and Masson 2014; Reynes et al. 2021). Results from this study show the effect of

571   clonality in the studied populations. First, samples corresponding to MLGs detected with

572   microsatellites were indeed grouped with reduced distance in networks based on RAD-sequencing.

573   Second, the distribution of pairwise differences showed a peak of low divergence, which can be an

574   indication of repeated MLLs. Third, the distribution of estimates of the inbreeding coefficient F

575   showed some individuals with negative values, corresponding to highly heterozygous individuals,

576   especially in French Polynesia. Our results point to the first observation of clonality (either with

577   microsatellites or RAD-sequencing loci) in populations with mitochondrial lineages 7a and 3g

578   (corresponding to *P. verrucosa*). This observation of clonality could be explained by the fact that

579   these individuals were sampled in shallow water, with a *P. acuta*-like corallum macromorphology

580   (i.e. thin branches, highly breakable). In such environment, waves and swell might favor

581   fragmentation. Testing this hypothesis would require studying this lineage in different environments

582   with contrasted levels of energy. In all cases, estimating the rate of clonality in these populations

583   would require a dedicated sampling.

584   We have also observed individuals with very low F values, indicating high rates of heterozygous

585   loci. This was observed in our simulations only with the highest clonal rates which are not

586   compatible with the low frequency of repeated MLLs observed here. Other effects could explain

587   these observations, such as the presence of brooded larvae, or the intra-colonial genetic diversity

588   linked to chimerism or mosaicism (Oury et al. 2020a). Hybrids between divergent lineages could

589   also create such high heterozygosity, but it could not be detected here if the parental lineages were

590   not analyzed.

591   From a methodological point of view, our simulations provide new avenues in the study of clonality

592   with RAD-sequencing data. Our simulations showed a discernible effect of clonality on the

593   distribution of F and $\bar{r}_d$ for the highest clonality level tested here (0.9). This is in line with previous

594   studies demonstrating an effect of clonality only for extreme rates of clonality (Balloux et al. 2003).

595   Our simulations may be limited in the exploration of the impact of clonality for several reasons.

596   First, we did not explore the impact of sampling scheme on the estimates of genetic diversity and

597   genetic structure. Second, *Pocillopora* corals show overlapping generations, and a given clone may

598   persist over several generations, a scenario that was not specifically implemented in our simulation

599   framework. Third, selective effects can lead to the expansion of one clone (see Gélin et al. 2017a,

600   and references therein) and thus modify the distribution of clones. These simulations were also not

601   used here to estimate clonal rates, which would require dedicated approaches, such as Approximate

602   Bayesian Computation (Csilléry et al. 2010).

603   Regarding genetic structure, our simulations showed a slight decrease of F_{ST} only for a clonal rate of

604   0.9. This agrees well with theoretical expectations (Balloux et al. 2003). The correction of datasets

605   for repetitions of clonal lineages changed the estimated F_{ST} values but this did not change our main

606   conclusions. Our results showed an important differentiation for the All datasets, and low to no

607   genetic structure in Oman and French Polynesia. In some cases such correction for clonal diversity

608   can lead to very different conclusions, from genetic differentiation without correction to near

609   panmixia in *Pocillopora* in French Polynesia (Adjeroud et al. 2014). Therefore, one should not rely

610   on a single analysis strategy (e.g., using all individuals or single genotypes), but consider different

611   sampling and filtering strategies (De Meeûs et al. 2006).

612

613

Data availability:

Raw sequences are available in Genbank under BioProject ID PRJNA689941 and SRA accession number SRA PRJNA689941.

The mitochondrial ORF sequences, microsatellite genotypes and SLiM scripts are available in Zenodo: https://zenodo.org/record/4748346. The scripts used for SLiM simulations are also available at: https://gitlab.osupytheas.fr/aurelle/slim-simulations.

648   The scripts used for the analyses of RAD-sequencing data are available at:

649   https://gitlab.osupytheas.fr/aurelle/rad_pocillopora.git

650 **References**

Adjeroud M, Guérécheau A, Vidal-Dupiol J, et al (2014) Genetic diversity, clonality and connectivity in the scleractinian coral Pocillopora damicornis: a multi-scale analysis in an insular, fragmented reef system. Mar Biol 161:531–541. https://doi.org/10.1007/s00227-013-2355-9

Agapow P, Burt A (2001) Indices of multilocus linkage disequilibrium. Mol Ecol Notes 1:101–102

Andrews S (2010) . FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Aurelle D, Pivotto ID, Malfant M, et al (2017) Fuzzy species limits in Mediterranean gorgonians (Cnidaria, Octocorallia): inferences on speciation processes. Zool Scr 46:767–778

Baird NA, Etter PD, Atwood TS, et al (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS ONE 3:e3376. https://doi.org/10.1371/journal.pone.0003376

Balloux F, Lehmann L, de Meeûs T (2003) The population genetics of clonal and partially clonal diploids. Genetics 164:1635–1644

Bandelt H-J, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol 16:37–48

Banguera-Hinestroza E, Ferrada E, Sawall Y, Flot J-F (2019) Computational Characterization of the mtORF of Pocilloporid Corals: Insights into Protein Structure and Function in Stylophora Lineages from Contrasting Environments. Genes 10:324. https://doi.org/10.3390/genes10050324

Bog M, Bässler C, Oberprieler C (2017) Lost in the hybridisation vortex: high-elevation *Senecio hercynicus* (Compositae, Senecioneae) is genetically swamped by its congener *S. ovatus* in the Bavarian Forest National Park (SE Germany). Evol Ecol 31:401–420. https://doi.org/10.1007/s10682-017-9890-7

Bradbury PJ, Zhang Z, Kroon DE, et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635

Brener-Raffali K, Vidal-Dupiol J, Adjeroud M, et al (2022) Gene expression plasticity and frontloading promote thermotolerance in *Pocillopora* corals. Peer Community Journal e13. https://doi.org/10.24072/pcjournal.79

Calderón I, Garrabou J, Aurelle D (2006) Evaluation of the utility of COI and ITS markers as tools for population genetic studies of temperate gorgonians. J Exp Mar Biol Ecol 336:184–197

Catchen J, Hohenlohe PA, Bassham S, et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. https://doi.org/10.1111/mec.12354

Combosch DJ, Vollmer SV (2015) Trans-Pacific RAD-Seq population genomics confirms introgressive hybridization in Eastern Pacific *Pocillopora* corals. Mol Phylogenet Evol 88:154–162

Combosch DJ, Vollmer SV (2011) Population Genetics of an Ecosystem-Defining Reef Coral *Pocillopora damicornis* in the Tropical Eastern Pacific. PLoS ONE 6:e21200. https://doi.org/10.1371/journal.pone.0021200

Csilléry K, Blum MG, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. Trends Ecol Evol 25:410–418

Cunning R, Bay R, Gillette P, et al (2018) Comparative analysis of the *Pocillopora damicornis* genome highlights role of immune system in coral evolution. Sci Rep 8:16134

Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158

De Meeûs T, Lehmann L, Balloux F (2006) Molecular epidemiology of clonal diploids: a quick overview and a short DIY (do it yourself) notice. Infect Genet Evol 6:163–170

Eaton DA, Overcast I (2020) ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics 36:2592–2594

Eaton DA, Spriggs EL, Park B, Donoghue MJ (2017) Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst Biol 66:399–412

Flot J-F, Tillier S (2007) The mitochondrial genome of *Pocillopora* (Cnidaria: Scleractinia) contains two variable regions: the putative D-loop and a novel ORF of unknown function. Gene 401:80–87

Forsman Z, Knapp I, Tisthammer K, et al (2017) Coral hybridization or phenotypic variation? Genomic data reveal gene flow between *Porites lobata* and *P. compressa*. Mol Phylogenet Evol 111:132–148

Frichot E, François O (2015) LEA: an R package for landscape and ecological association studies. Methods Ecol Evol 6:925–929

Frichot E, Mathieu F, Trouillon T, et al (2014) Fast and efficient estimation of individual ancestry coefficients. Genetics 196:973–983

Garrabou J, Coma R, Bensoussan N, et al (2009) Mass mortality in Northwestern Mediterranean rocky benthic communities: effects of the 2003 heat wave. Glob Change Biol 15:1090–1103. https://doi.org/10.1111/j.1365-2486.2008.01823.x

Gautier M, Gharbi K, Cezard T, et al (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Mol Ecol 22:3165–3178

Gélin P, Fauvelot C, Mehn V, et al (2017a) Superclone expansion, long-distance clonal dispersal and local genetic structuring in the coral *Pocillopora damicornis* type β in Reunion Island, South Western Indian Ocean. PloS One 12:e0169692

Gélin P, Postaire B, Fauvelot C, Magalon H (2017b) Reevaluating species number, distribution and endemism of the coral genus *Pocillopora* Lamarck, 1816 using species delimitation methods and microsatellites. Mol Phylogenet Evol 109:430-446

Gélin P, Pirog A, Fauvelot C, Magalon H (2018) High genetic differentiation and low connectivity in the coral *Pocillopora damicornis* type β at different spatial scales in the Southwestern Indian Ocean and the Tropical Southwestern Pacific. Mar Biol 165:167

Haller BC, Messer PW (2019) SLiM 3: forward genetic simulations beyond the Wright–Fisher model. Mol Biol Evol 36:632–637

Hellberg ME (2006) No variation and low synonymous substitution rates in coral mtDNA despite high nuclear variation. BMC Evol Biol 6:1–8

Highsmith RC (1982) Reproduction by fragmentation in corals. Mar Ecol Prog Ser Oldendorf 7:207–226

Hoang DT, Chernomor O, Von Haeseler A, et al (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522

Huang H, Knowles LL (2016) Unforeseen consequences of excluding missing data from next-generation sequences: simulation study of RAD sequences. Syst Biol 65:357–365

Hughes TP, Kerry JT, Baird AH, et al (2018) Global warming transforms coral reef assemblages. Nature 556:492

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

Johnston EC, Forsman ZH, Flot J-F, et al (2017) A genomic glance through the fog of plasticity and diversification in *Pocillopora*. Sci Rep 7:5991

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405

Kamvar ZN, Tabima JF, Grünwald NJ (2014) Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281

Kamvar ZN, Brooks JC, Grünwald NJ (2015) Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Front Genet 6:208

Kayal E, Bastian B, Pankey MS, et al (2018). Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. BMC Evol Biol 18(1):1-18

Kosiński P, Sękiewicz K, Walas Ł, et al (2019) Spatial genetic structure of the endemic alpine plant Salix serpillifolia: genetic swamping on nunataks due to secondary colonization? Alp Bot 129:107–121

Larson WA, Isermann DA, Feiner ZS (2021) Incomplete bioinformatic filtering and inadequate age and growth analysis lead to an incorrect inference of harvested-induced changes. Evol Appl 14(2):278-289

Ledoux J, Garrabou J, Bianchimani O, et al (2010) Fine-scale genetic structure and inferences on population biology in the threatened Mediterranean red coral, *Corallium rubrum*. Mol Ecol 19:4204–4216. https://doi.org/10.1111/j.1365-294X.2010.04814.x

Leigh JW, Bryant D (2015) POPART: full-feature software for haplotype network construction. Methods Ecol Evol 6:1110–1116

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760

Magalon H, Adjeroud M, Veuille M (2004) Patterns of genetic variation do not correlate with geographical distance in the reef-building coral *Pocillopora meandrina* in the South Pacific. Mol Ecol 14:1861–1868

Marti-Puig P, Forsman ZH, Haverkort-Yeh RD, et al (2014) Extreme phenotypic polymorphism in the coral genus *Pocillopora*; micro-morphology corresponds to mitochondrial groups, while colony morphology does not. Bull Mar Sci 90:211–231. https://doi.org/10.5343/bms.2012.1080

Minh BQ, Nguyen MAT, von Haeseler A (2013) Ultrafast Approximation for Phylogenetic Bootstrap. Mol Biol Evol 30:1188–1195. https://doi.org/10.1093/molbev/mst024

Mokhtar-Jamaï K, Coma R, Wang J, et al (2013) Role of evolutionary and ecological factors in the reproductive success and the spatial genetic structure of the temperate gorgonian *Paramuricea clavata*. Ecol Evol 3:1765–1779

Nelson TC, Stathos AM, Vanderpool DD, et al (2020) Ancient and recent introgression shape the evolutionary history of pollinator adaptation and speciation in a model monkeyflower radiation (*Mimulus* section *Erythranthe*). bioRxiv

O'Leary SJ, Puritz JB, Willis SC, et al (2018) These aren't the loci you'e looking for: Principles of effective SNP filtering for molecular ecologists. Mol Ecol 27:3193–3206

Oury N, Gélin P, Magalon H (2020a) Together stronger: Intracolonial genetic variability occurrence in *Pocillopora* corals suggests potential benefits. Ecol Evol 10:5208–5218

Oury N, Gélin P, Magalon H (2020b) Cryptic species and genetic connectivity among populations of the coral *Pocillopora damicornis* (Scleractinia) in the tropical southwestern Pacific. Mar Biol 167:1–15

Oury N, Gélin P, Magalon H (2021) High connectivity within restricted distribution range in *Pocillopora* corals. J Biogeogr 48:1679–1692

Oury N, Gélin P, Massé L, Magalon H (2019) First study of asexual planulae in the coral *Pocillopora damicornis* type β SSH05c from the southwestern Indian Ocean. Coral Reefs 38:499–503

Pante E, Abdelkrim J, Viricel A, et al (2015a) Use of RAD sequencing for delimiting species. Heredity 114:450–459

Pante E, Puillandre N, Viricel A, et al (2015b) Species are hypotheses: avoid connectivity assessments based on pillars of sand. Mol Ecol 24:525–544

Pinzón JH, Reyes-Bonilla H, Baums IB, LaJeunesse TC (2012) Contrasting clonal structure among *Pocillopora* (Scleractinia) communities at two environmentally distinct sites in the Gulf of California. Coral Reefs 31:765–777. https://doi.org/10.1007/s00338-012-0887-y

Pinzón JH, Sampayo E, Cox E, et al (2013) Blind to morphology: genetics identifies several widespread ecologically common species and few endemics among Indo-Pacific cauliflower corals (*Pocillopora*, Scleractinia). J Biogeogr 40:1595–1608. https://doi.org/10.1111/jbi.12110

Pratlong M, Haguenauer A, Brener K, et al (2021) Separate the wheat from the chaff: genomic scan for local adaptation in the red coral *Corallium rubrum*. Peer Community J 1:e31 https://doi.org/10.24072/pcjournal.12

Pratlong M, Haguenauer A, Chenesseau S, et al (2017a) Evidence for a genetic sex determination in Cnidaria, the Mediterranean red coral (*Corallium rubrum*). R Soc Open Sci 4:. https://doi.org/10.1098/rsos.160880

Pratlong M, Rancurel C, Pontarotti P, Aurelle D (2017b) Monophyly of Anthozoa (Cnidaria): why do nuclear and mitochondrial phylogenies disagree? Zool Scr 46:363–371

Rambaut A (2012) FigTree v1. 4. Available online at: http://tree.bio.ed.ac.uk/software/figtree/

Reichel K, Masson J-P, Malrieu F, et al (2016) Rare sex or out of reach equilibrium? The dynamics of F IS in partially clonal organisms. BMC Genet 17:76

Reynes L, Thibaut T, Mauger S, et al (2021) Genomic signatures of clonality in the deep water kelp *Laminaria rodriguezii*. Mol Ecol 30:1806–1822. https://doi.org/10.1111/mec.15860

Robitzch V, Banguera-Hinestroza E, Sawall Y, et al (2015) Absence of genetic differentiation in the coral *Pocillopora verrucosa* along environmental gradients of the Saudi Arabian Red Sea. Front Mar Sci 2:5

Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour 8:103–106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold spring harbor laboratory press

Schmidt-Roach, S, Miller KJ, Woolsey E, Gerlach G, Baird AH (2012). Broadcast spawning by *Pocillopora* species on the Great Barrier Reef. PLoS One 7:e50847

Schmidt-Roach, S, Lundgren P, Miller KJ, Gerlach G, Noreen AME, Andreakis N (2013). Assessing hidden species diversity in the coral *Pocillopora damicornis* from Eastern Australia. Coral Reefs 32: 161-172.

Schmidt-Roach S, Miller KJ, Lundgren P, Andreakis N (2014) With eyes wide open: a revision of species within and closely related to the *Pocillopora damicornis* species complex (Scleractinia; Pocilloporidae) using morphology and genetics. Zool J Linn Soc 170:1–33

Sheets EA, Warner PA, Palumbi SR (2018) Accurate population genetic measurements require cryptic species identification in corals. Coral Reefs 37:549–563

Stoeckel S, Masson J-P (2014) The exact distributions of F_{IS} under partial asexuality in small finite populations with mutation. PLoS One 9:e85228

Tamura K, Stecher G, Peterson D, et al (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Thomas L, Kendrick G, Stat M, et al (2014) Population genetic structure of the *Pocillopora damicornis* morphospecies along Ningaloo Reef, Western Australia. Mar Ecol Prog Ser 513:111–119

Torda G, Lundgren P, Willis B, van Oppen MJ (2013) Genetic assignment of recruits reveals short- and long-distance larval dispersal in *Pocillopora damicornis* on the Great Barrier Reef. Mol Ecol 22:5821–5834

Underwood JN, Smith LD, Van Oppen MJH, Gilmour JP (2007) Multiple scales of genetic connectivity in a brooding coral on isolated reefs following catastrophic bleaching. Mol Ecol 16:771–784

van Oppen MJ, Bongaerts P, Frade P, et al (2018) Adaptation to reef habitats through selection on the coral animal and its associated microbiome. Mol Ecol 27:2956–2971

van Oppen MJH, McDonald BJ, Willis B, Miller DJ (2001) The evolutionary history of the coral genus *Acropora* (Scleractinia, Cnidaria) based on a mitochondrial and nuclear marker: reticulation, incomplete lineage sorting or morphological convergence? Mol Biol Evol 18:1315–1329

van Oppen MJH, Willis BL, Miller DJ (1999) Atypically low rate of cytochrome b evolution in the scleractinian coral genus *Acropora*. Proc Biol Sci 266:179–183

Veron J (2013) Overview of the taxonomy of zooxanthellate Scleractinia. Zool J Linn Soc 169:485–508

Vidal-Dupiol J, Chaparro C, Pratlong M, et al (2020) Sequencing, de novo assembly and annotation of the genome of the scleractinian coral, *Pocillopora acuta*. bioRxiv 698688

Vollmer S, Palumbi SR (2002) Hybridization and the Evolution of Reef Coral Diversity. Science 296:2023–2025

Wang J-T, Chen Y-Y, Tew KS, et al (2012) Physiological and biochemical performances of menthol-induced aposymbiotic corals. PLOS ONE 7(9): e46406. https://doi.org/10.1371/journal.pone.0046406

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370

651

Table and figure captions :

Table 1: Correspondence between the nomenclature of mitochondrial lineages used in the manuscript (indicated as ORF haplotypes numbers), and primary and secondary species hypotheses (PSH and SSH respectively) in Gélin et al. (2017), the types of Pinzón et al. (2013) and Schmidt-Roach et al. (2012, 2013, 2014), and the potential nominal species associated to these types. Note that the correspondence between mitochondrial lineages and SSH indicated here for our samples is based on the complementary analysis of microsatellite genotypes: it may therefore be different for other samples. See table 1 of Johnston et al., 2017 as well.

Table 2 Characteristics of sampling sites, including region, site, GPS location, and sampling depth and date. The thermal regime gives a qualitative indication of temperature variability at the corresponding sampling site. See Supplementary Table S1 for the final sampling sizes used depending on datasets and mitochondrial lineages,

Table 3 Numbers of SNPs and individuals for the different datasets. The locus threshold corresponds to the minimum proportion of available data among individuals to retain a locus. The individual threshold corresponds to the minimum proportion of available data among loci to retain an individual. The mean depth indicates the mean depth per individual averaged over all individuals in the dataset. The last column indicates the number of retained individuals after correction for the presence of Multiple Multilocus Lineages (MLLs; see text for details). *All* corresponds to dataset including all samples (French Polynesia and Oman).

Table 4 Estimates of gene diversity within (*1-Qintra*) and among individuals (*1-Qinter*), and of F_{IS} averaged over samples. For the All datasets, the average was estimated over Oman and French Polynesia samples. For the Oman and French Polynesia datasets, average was done over corresponding sampling locations. The last three columns provide indicators of the distribution of inbreeding coefficient (mean, minimum, and maximum) computed over all individuals for each dataset.

Table 5 Mean F_{ST} estimates across loci for the datasets without and with corrections for repeated MLLs. The "pop" comparison estimates F_{ST} between sampling sites; for the All dataset, it

679 corresponds to the French Polynesia / Oman differentiation. For the "ORF" comparison, this

680 estimate compares samples of individuals grouped according to their ORF haplotypes (individuals

681 without ORF sequence were not taken into account).

682 **Fig. 1** Network based on the percentage of differences among individuals for the All, Oman and

683 French Polynesia 95_75 datasets. The colors indicate the corresponding species hypothesis

684 according to the sequence of mitochondrial ORF. The red ellipses indicate groupings of individuals

685 corresponding to the same MLG according to microsatellite data.

686 **Fig. 2** Plots of coancestry coefficients inferred with the LEA R package for the 95_75 datasets. For

687 the All and French Polynesia datasets, the cross-entropy gave a first minimum value at K = 2. Plots

688 of coancestry coefficients considering the number of mitochondrial lineages in the dataset gave a

689 minimum value of K = 4 for All and Oman, and K = 5 for the number of populations in French

690 Polynesia. Each bar corresponds to one individual and the colors in the bars correspond to the

691 different ancestry groups inferred by this method. The colors of the dots under barplots indicate the

692 mitochondrial lineage for a subset of individuals.

693 **Table 1:** Correspondence between the nomenclature of mitochondrial lineages used in the

694 manuscript (indicated as ORF haplotypes numbers), and primary and secondary species hypotheses

695 (PSH and SSH respectively) in Gélin et al. (2017), the types of Pinzón et al. (2013) and Schmidt-

696 Roach et al. (2012, 2013, 2014), and the potential nominal species associated to these types. Note

697 that the correspondence between mitochondrial lineages and SSH indicated here for our samples is

698 based on the complementary analysis of microsatellite genotypes: it may therefore be different for

699 other samples. See table 1 of Johnston et al., 2017 as well.

700

ORF	PSH Gélin et al. (2017b)	SSH Gélin et al. (2017b)	Type Pinzón et al. (2013)	Type Schmidt-Roach et al. (2012, 2013, 2014)	Nominal species
27	9	9c	1a	e/m	*P. grandis / meandrina*
1	1	1	2	-	*P.* sp. *B*
18	5	to be defined	5a	β	*P. acuta*
34	12	12	7a	-	
36	13	13a	3e	-	*P. verrucosa*
43	13	13a	3g	-	*P. verrucosa*

701 **Table 2** Characteristics of sampling sites, including region, site, GPS location, and sampling depth

702 and date. The thermal regime gives a qualitative indication of temperature variability at the

703 corresponding sampling site. See Supplementary Table S1 for the final sampling sizes used

704 depending on datasets and mitochondrial lineages,

705

Region	Site	GPS	Code	Sampling date	Depth (m)	Thermal regime
French Polynesia	Moorea Haapiti	17°32'39.27 S 149°53'37.40 W	MH	03/2014	0.5-2	Low variations
French Polynesia	Moorea Tiahura	17°29'17.41 S 149°53'45.58 W	MT	03/2014	0.5-2	Low variations
French Polynesia	Moorea Vaiare	17°31'24.10 S 149°46'33.85 W	MV	03/2014	0.5-2	Low variations
French Polynesia	Tahiti Faratea	17°43'17.61 S 149°18'11.78 W	TF	03/2014	0.5-2	Low variations
French Polynesia	Tahiti Vairao	17°48'20.90 S 149°17'43.13 W	TV	03/2014	0.5-2	Low variations
French Polynesia	Tahiti Tautira	17°45'12.11 S 149° 9'26.68 W	TT	03/2014	0.5-2	Low variations*
Oman	Bandar Al Khayral 1	23°30'54.25 N 58°45'15.70 E	O1	06/2014	2-8	High variations
Oman	Bandar Al Khayral 21	23°31'26.66 N 58°44'2.18 E	O2	06/2014	2-8	High variations
Oman	Bandar Al Khayral 3	23°31'8.90 N 58°45'29.40 E	O3	06/2014	> 12	High variations but less than O1, O2 and O4
Oman	Muscat	23°37'28.61 N 58°36'1.39 E	O4	06/2014	2-8	High variations
Oman	Daymaniat	23°51'25.12 N 58° 6'3.43 E	O5	06/2014	2-8	High variations but less than O1, O2 and O4

706 * see Brener-Raffalli et al. (2022) for further details on thermal regime in Oman.

707 **Table 3** Numbers of SNPs and individuals for the different datasets. The locus threshold

708 corresponds to the minimum proportion of available data among individuals to retain a locus. The

709 individual threshold corresponds to the minimum proportion of available data among loci to retain

710 an individual. The mean depth indicates the mean depth per individual averaged over all individuals

711 in the dataset. The last column indicates the number of retained individuals after correction for the

712 presence of Multiple Multilocus Lineages (MLLs; see text for details). All corresponds to dataset

713 including all samples (French Polynesia and Oman).

714

Dataset	Populations	Locus threshold	Individual threshold	Mean depth	SNPs	Individuals before MLL correction	Individuals after MLL correction
All_95_75	All	0.95	0.75	389	320	132	100
All_75_75	All	0.75	0.75	66.9	194370	98	78
Oman_95_75	Oman	0.95	0.75	155	1711	99	82
Oman_75_75	Oman	0.75	0.75	72.7	134307	77	62
Polynesia_95_75	Polynesia	0.95	0.75	433.3	558	31	29
Polynesia_75_75	Polynesia	0.75	0.75	204.2	3285	25	18

715 **Table 4** Estimates of gene diversity within (*1-Qintra*) and among individuals (*1-Qinter*), and of F_{IS}

716 averaged over samples. For the All datasets, the average was estimated over Oman and French

717 Polynesia samples. For the Oman and French Polynesia datasets, average was done over

718 corresponding sampling locations. The last three columns provide indicators of the distribution of

719 inbreeding coefficient (mean, minimum, and maximum) computed over all individuals for each

720 dataset.

721

Dataset	*1-Qintra*	*1-Qinter*	F_{IS}	Mean F	Min F	Max F
All_95_75	0.07	0.06	-0.15	-0.024	-1.306	0.546
All_75_75	0.12	0.15	0.21	0.396	0.043	0.732
Oman_95_75	0.16	0.16	0.00	-0.009	-0.311	0.602
Oman_75_75	0.20	0.24	0.14	0.138	-0.062	0.539
Polynesia_95_75	0.04	0.04	-0.21	0.095	-2.733	0.539
Polynesia_75_75	0.05	0.05	0.02	0.292	-0.838	0.489

722

723 **Table 5** Mean F_{ST} estimates across loci for the datasets without and with corrections for repeated

724 MLLs. The "pop" comparison estimates F_{ST} between sampling sites; for the All dataset, it

725 corresponds to the French Polynesia / Oman differentiation. For the "ORF" comparison, this

726 estimate compares samples of individuals grouped according to their ORF haplotypes (individuals

727 without ORF sequence were not taken into account).

728

Dataset	F_{ST} all individuals	F_{ST} corrected for MLLs
All_95_75 pop	0.105	0.088
All_95_75 ORF	0.178	0.186
All_75_75 pop	0.352	0.355
All_75_75 ORF	0.361	0.312
Oman_95_75 pop	0.011	0.001
Oman_95_75 ORF	0.003	-0.003
Oman_75_75 pop	0.016	0.002
Oman_75_75 ORF	0.005	-0.002
Polynesia_95_75 pop	0.000	-0.007
Polynesia_75_75 pop	0.015	-0.024

729

Figure 1

All_95_75

Oman

Polynesia

0.01

Oman_95_75

MLG02

MLG06
MLG10
MLG07

MLG09
MLG01

0.01

mt ORF :
Unknown
Type 3e
Type 3g
Beta/type 5a
Type 7a

Polynesia_95_75

0.01
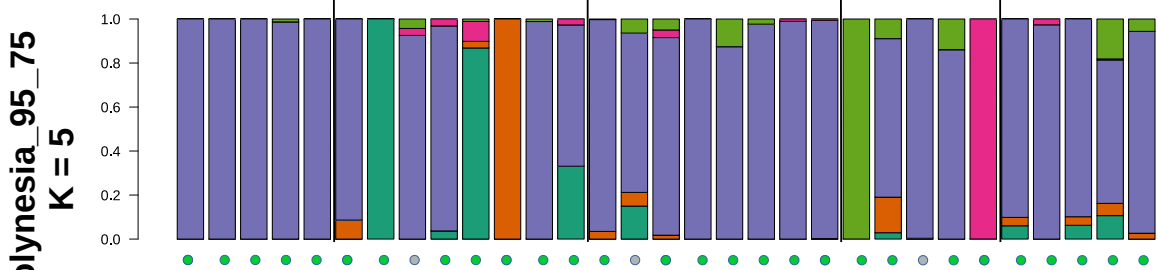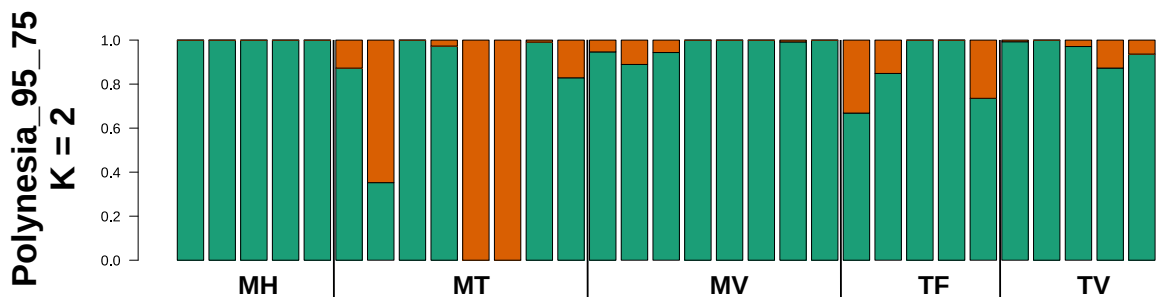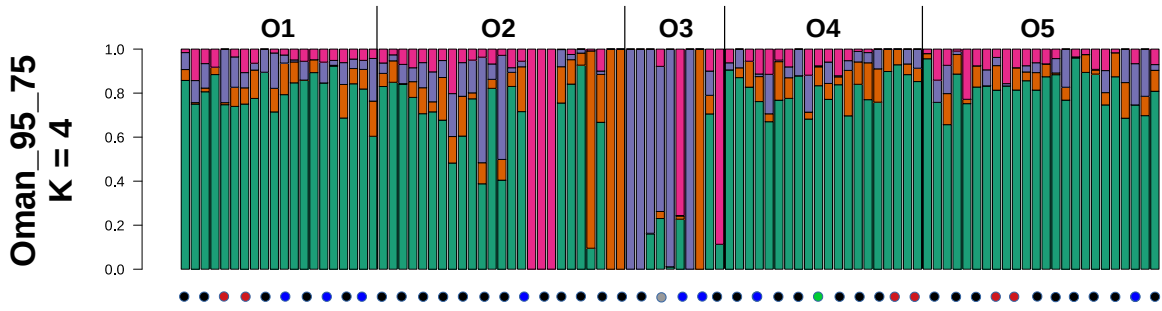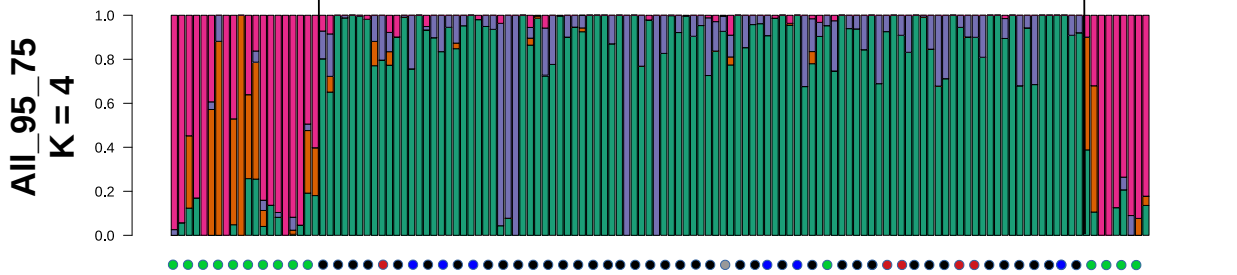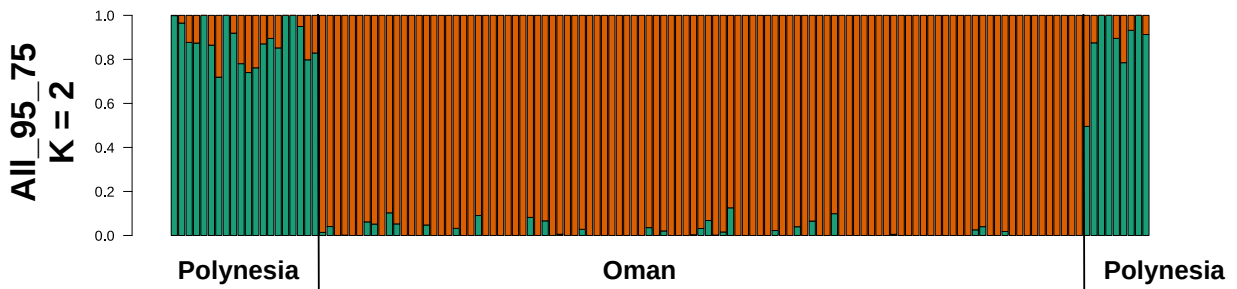
mt lineage: • Unknown • Type 7a • Type 3g • Type 5a • Type 3e

Table S1: samples sizes per species hypothesis (on the basis of mitochondrial ORF) and sampling site for the different datasets. See main text for details on sampling sites and Gélin et al. (2017) for species hypotheses. For French Polynesia, all mitochondrial sequences corresponded to type 5a. Note that some individuals were shared among datasets.

Dataset	Region	Site	type 5a	type 7a	type 3g	type 3e	unknown
All_95_75	Oman	O1		16	3	3	
All_95_75	Oman	O2		24		2	
All_95_75	Oman	O3		6		2	2
All_95_75	Oman	O4	1	13	2	4	1
All_95_75	Oman	O5		19	2	3	
All_95_75	French Polynesia	MH	5				
All_95_75	French Polynesia	MT	6				1
All_95_75	French Polynesia	MV	7				1
All_95_75	French Polynesia	TF	3				1
All_95_75	French Polynesia	TV	5				
All_95_75		**Total**	**27**	**78**	**7**	**14**	**6**
All_75_75	Oman	O1		5	3	3	
All_75_75	Oman	O2		19		2	
All_75_75	Oman	O3		5			2
All_75_75	Oman	O4	1	11	2	3	
All_75_75	Oman	O5		19	2	3	
All_75_75	Polynesia	MH	4				
All_75_75	Polynesia	MV	6				1
All_75_75	Polynesia	TF	1				1
All_75_75	Polynesia	TV	5				
All_75_75		**Total**	**17**	**59**	**7**	**11**	**4**
Oman_95_75	Oman	O1		14	3	3	
Oman_95_75	Oman	O2		23		2	
Oman_95_75	Oman	O3		7		2	1
Oman_95_75	Oman	O4	1	13	2	4	
Oman_95_75	Oman	O5		19	2	3	
Oman_95_75		**Total**	**1**	**76**	**7**	**14**	**1**
Oman_75_75	Oman	O1		5	3	3	
Oman_75_75	Oman	O2		18		2	
Oman_75_75	Oman	O3		5			1
Oman_75_75	Oman	O4	1	10	2	3	
Oman_75_75	Oman	O5		19	2	3	
Oman_75_75		**Total**	**1**	**57**	**7**	**11**	**1**
Polynesia_95_75	French	MH	5				

	Polynesia				
Polynesia_95_75	French Polynesia	MT	7		1
Polynesia_95_75	French Polynesia	MV	7		1
Polynesia_95_75	French Polynesia	TF	4		1
Polynesia_95_75	French Polynesia	TV	5		
Polynesia_95_75		**Total**	**28**		**3**
Polynesia_75_75	French Polynesia	MH	5		
Polynesia_75_75	French Polynesia	MT	6		1
Polynesia_75_75	French Polynesia	MV	7		1
Polynesia_75_75	French Polynesia	TV	5		
Polynesia_75_75		**Total**	**23**		**2**

Table S2 : statistics on read numbers and alignment after initial filtering on sequence quality : mean, minimum, maximum and standard deviation (s.d.) of the initial numbers of reads, and of the percentage of reads mapped on genome par individual.

Samples	Number of individuals	Mean reads	Min reads	Max reads	s.d.	Mean mapped	Min mapped	Max mapped	s.d.
All	211	4540048	5735	30394029	5169991	85.1	70.6	86.7	1.9
Oman	141	5647233	5735	30394029	5669214	85.6	78.3	86.7	1.1
French Polynesia	70	2309860	20495	11800706	2930211	84.1	70.6	86.2	2.7

Table S3 : estimates of gene diversity within (*1-Qintra*) and among individuals (*1-Qinter*), and of F_{IS} over all loci. For each dataset the estimates are given first for the whole dataset and per population (here sampling site) and per ORF haplotype if at least two individuals with the corresponding ORF haplotype have been analysed. The results are given for datasets not corrected for repeated MLLs. For French Polynesia, all individuals for which we got a mitochondrial ORF sequence corresponded to beta/type5a.

Dataset	Population / ORF	*1-Qintra*	*1-Qinter*	F_{IS}
All 95 75	All	0.07	0.06	-0.15
	Population			
	French Polynesia	0.09	0.09	-0.04
	Oman	0.05	0.04	-0.27
	ORF			
	type5a	0.09	0.09	-0.01
	7a	0.05	0.04	-0.27
	3g	0.04	0.03	-0.42
	3e	0.05	0.04	-0.28
All 75 75	All	0.12	0.15	0.21
	Population			
	French Polynesia	0.13	0.17	0.27
	Oman	0.11	0.13	0.15
	ORF			
	type5a	0.13	0.19	0.33
	7a	0.11	0.13	0.14
	3g	0.12	0.13	0.06
	3e	0.11	0.13	0.13
Oman 95 75	All	0.16	0.16	0.00
	Population			
	O1	0.15	0.16	0.07
	O2	0.16	0.15	-0.04
	O3	0.15	0.16	0.11
	O4	0.16	0.16	0.01
	O5	0.19	0.16	-0.16
	ORF			
	7a	0.16	0.16	0
	3g	0.18	0.15	-0.18
	3e	0.16	0.16	-0.05
Oman 75 75	All	0.20	0.24	0.14
	Population			
	O1	0.22	0.24	0.08
	O2	0.19	0.23	0.15
	O3	0.18	0.24	0.24
	O4	0.20	0.24	0.18
	O5	0.23	0.24	0.05
	ORF			
	7a	0.21	0.24	0.13
	3g	0.22	0.24	0.07
	3e	0.21	0.24	0.13
Polynesia 95 75	All	0.04	0.04	-0.21
	Population			
	MH	0.04	0.02	-0.52
	MT	0.05	0.06	0.14
	MV	0.04	0.03	-0.34
	TF	0.05	0.06	0.09
	TV	0.03	0.02	-0.44
Polynesia 75 75	All	0.05	0.05	0.02
	Population			
	MH	0.04	0.04	-0.15
	MT	0.05	0.08	0.39
	MV	0.05	0.05	0.02
	TV	0.05	0.04	-0.17

Table S4 : estimate of the index association $\bar{r}_d$ to study linkage disequilibrium in the different datasets. The first column gives the mean and the second the standard deviation computed over 10 000 replicates of 320 SNPs. The "global" results correspond to analyses done at the level of the whole corresponding dataset. The "strata" results correspond to analyses done at the level of mitochondrial lineages for the All and Oman datasets, and at the level of sampling site for the French Polynesia datasets (i.e. the "strata" levels used in the poppr R package). Note that the analysis was not done for 5a in Oman where only one individual was analysis.

Dataset	mean	s.d.
Global		
All_95_75	0.033	0.005
All_75_75	0.099	0.019
Oman_95_75	0.002	0.001
Oman_75_75	0.004	0.001
Polynesia_95_75	0.035	0.007
Polynesia_75_75	0.017	0.007
Strata All_95_75		
3e	0.003	0.006
3g	0.025	0.025
5a	0.019	0.004
7a	0.005	0.002
Strata All_75_75		
3e	0	0.003
3g	0.049	0.009
5a	0.083	0.017
7a	0.007	0.002
Strata Oman_95_75		
3e	0.003	0.003
3g	0.002	0.006
7a	0.003	0.001
Strata Oman_75_75		
3e	0	0.002
3g	0.05	0.006
7a	0.007	0.001
Strata Polyn_95_75		
MH	0.278	0.154
MT	0.008	0.009
MV	0.014	0.024
TF	0.086	0.043
TV	0.116	0.101
Strata Polyn_75_75		
MH	0.235	0.072
MT	0.01	0.011
MV	0.008	0.014
TV	0.108	0.065

Figure S1 : map of sampling points. See main text for details on sampling sites
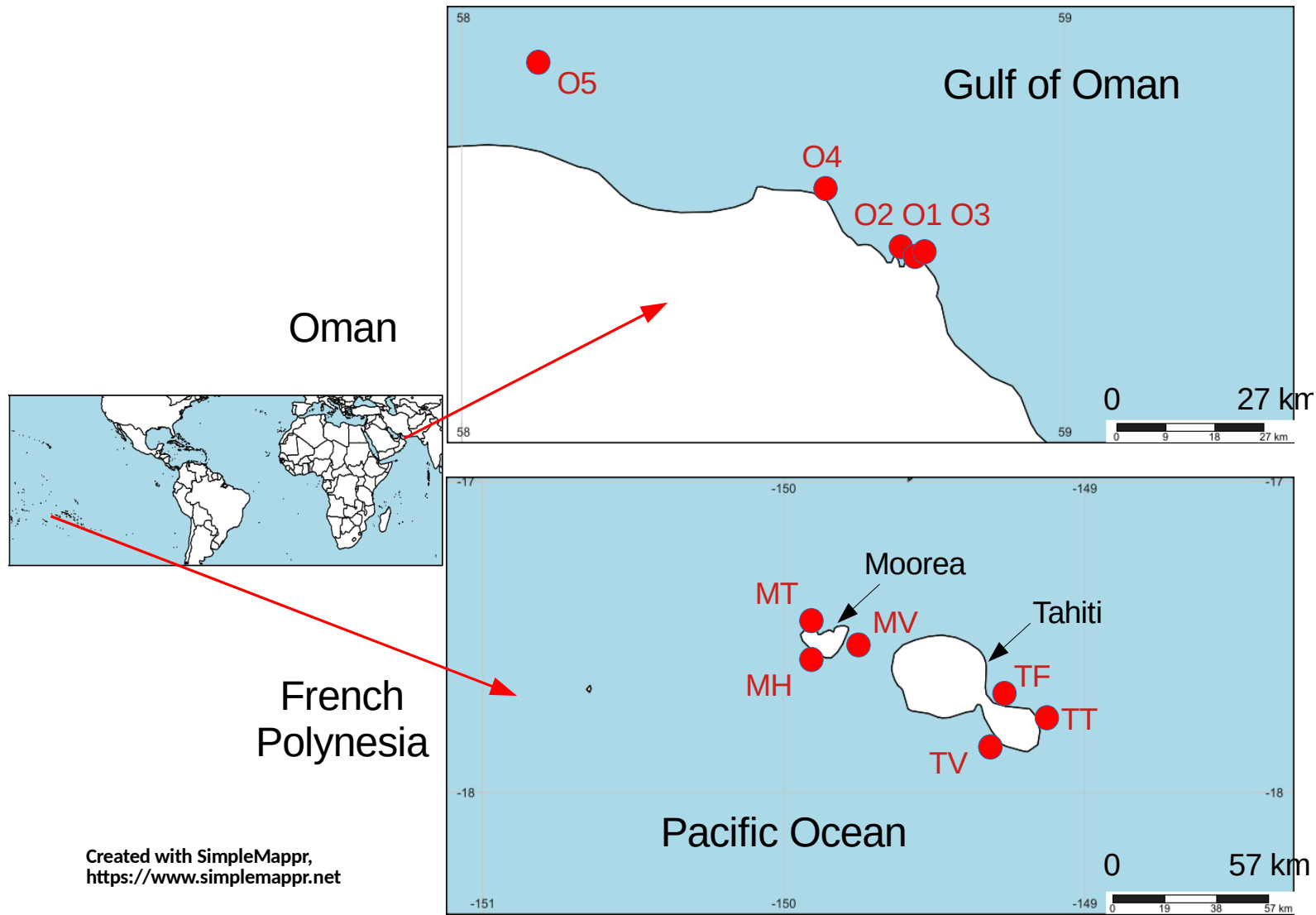
Figure S2 : network of mitochondrial ORF sequences. The colors indicate the origin of sequences (Oman, French Polynesia or other studies). Sequences from previous studies can be found in Gélin et al. (2017b).
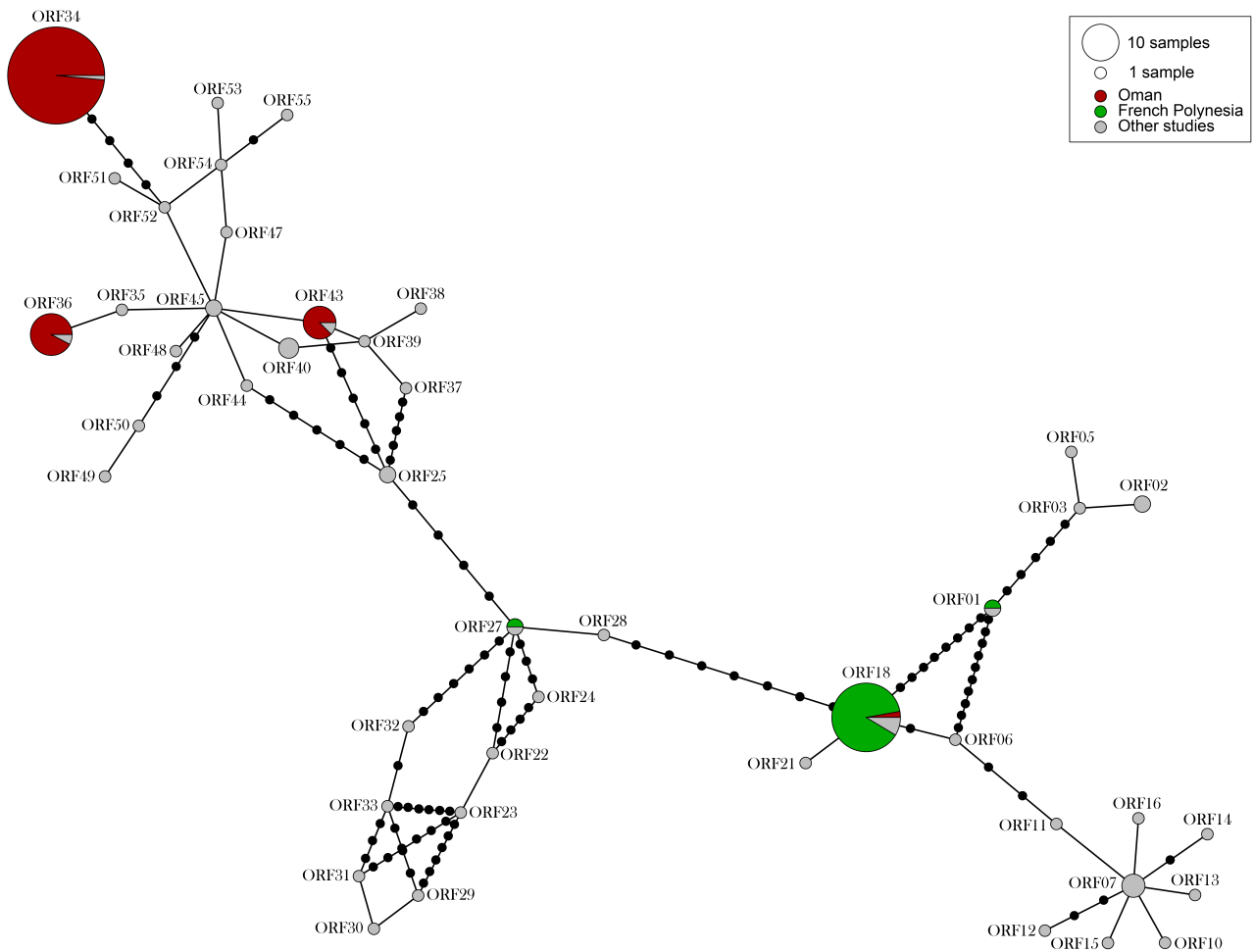
Figure S3: assembly statistics for RAD-Seq data. A) distribution of the number of raw reads, with color according to groups: red: Oman, grey: French Polynesia, blue: outgroups; B) distribution of the percentage of reads mapped to the *Pocillopora* genome for the different samples

A)

B)
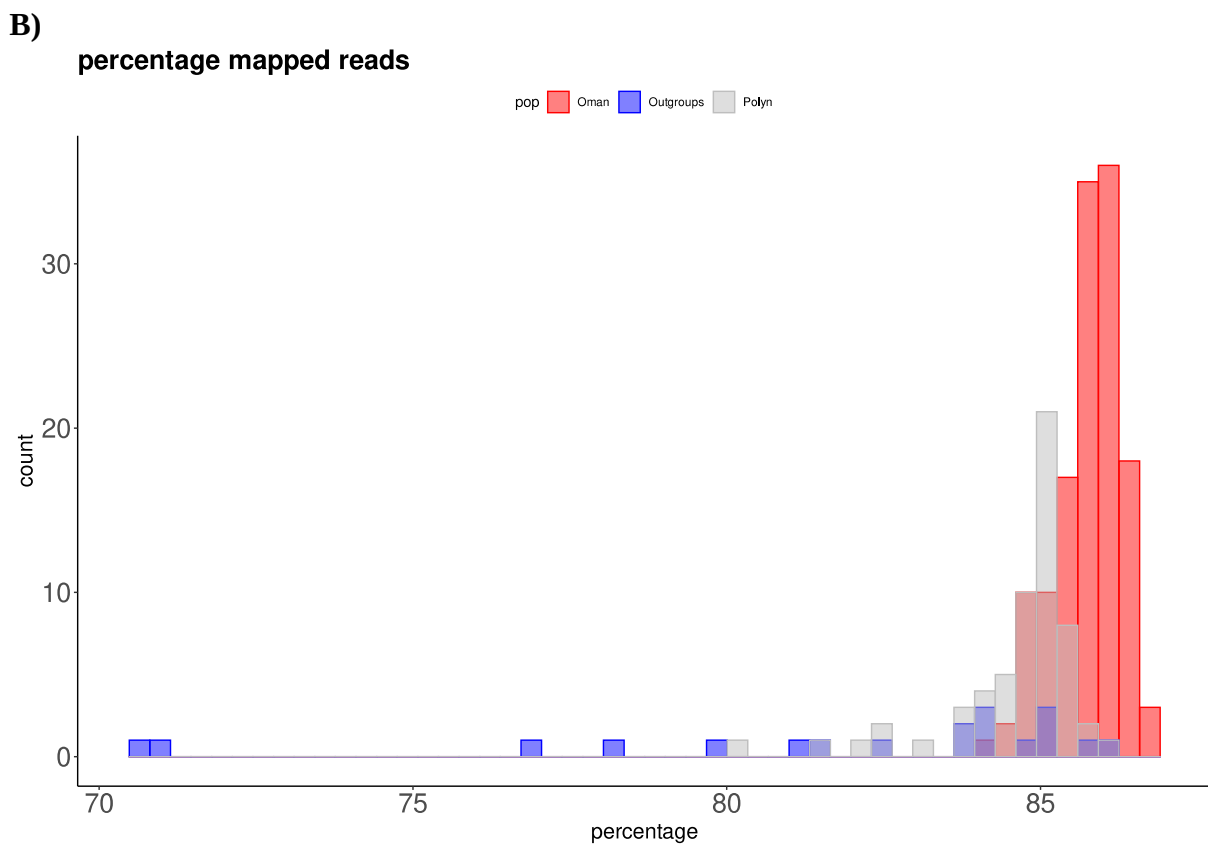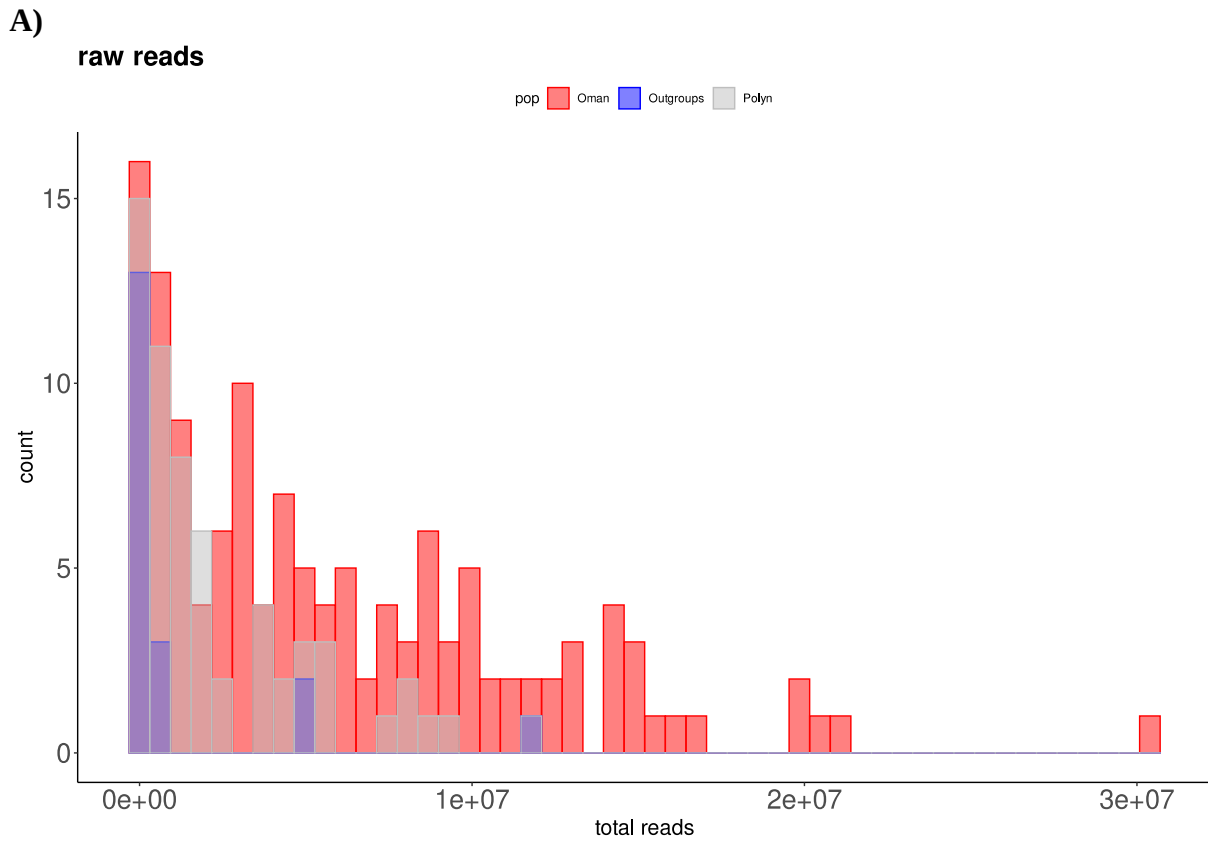
Figure S4: maximum-likelihood phylogenetic inference from RAD-Sequencing on the basis of 80 596 RAD loci, with 1 000 ultrafast bootstraps. The tree has been rooted at mid-point. The model retained by IQ-TREE was the TN+F+R5. The percentages of bootstraps are indicated at the left of the corresponding nodes. The colors indicate the mitochondrial lineage for each sample.
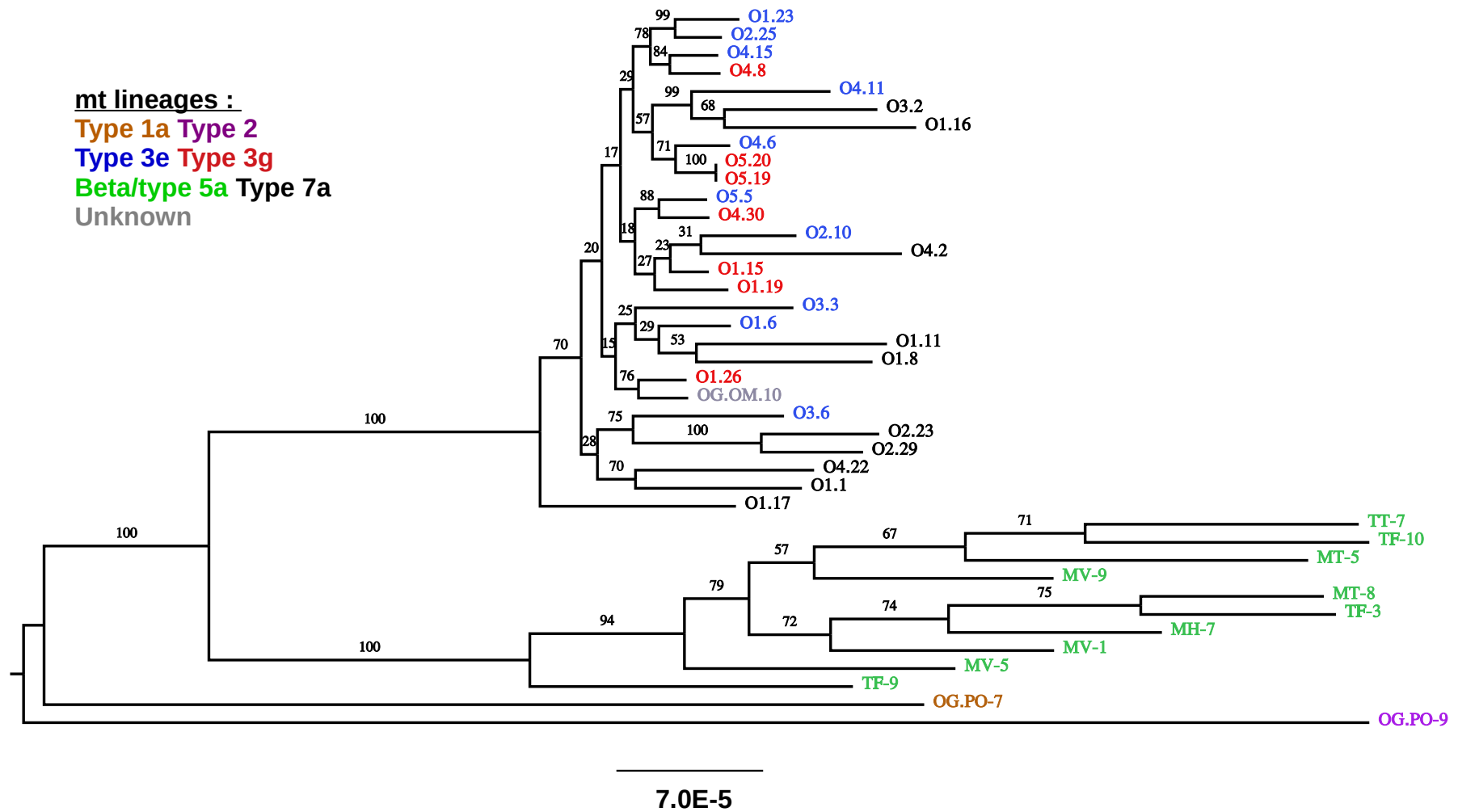
Figure S5 : histograms of genetic distance among individuals for the different datasets. The distances correspond to the proportion of differences among loci computed with poppr. We indicate below each histogram the retained threshold to remove repeated MLLs.

A) All_95_75

% differences All 95 75
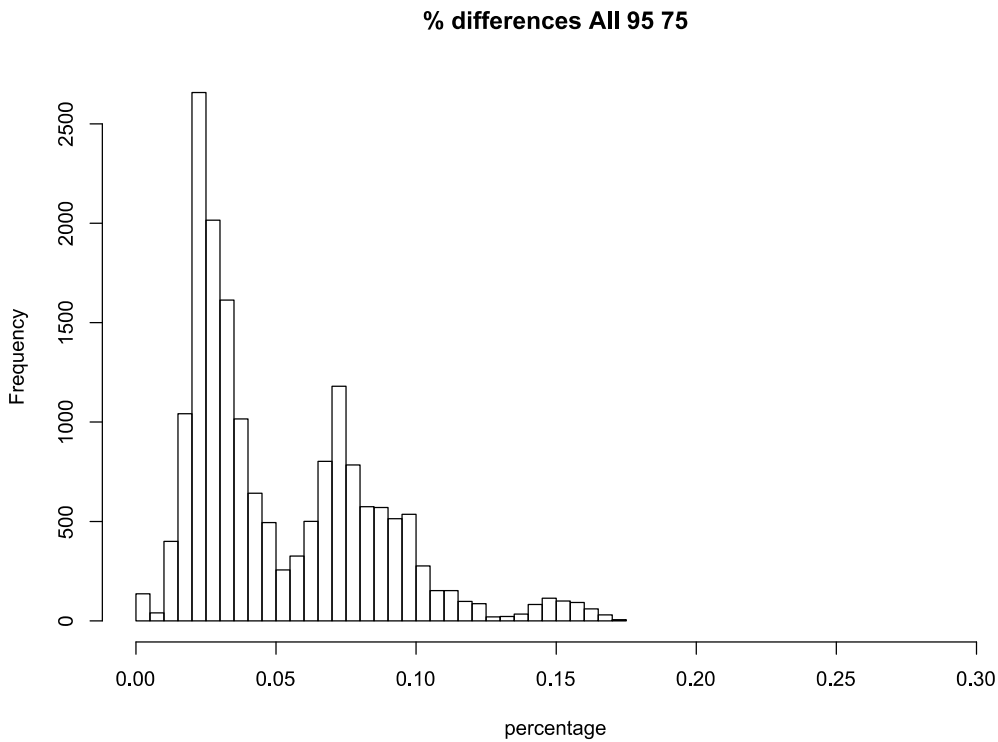
MLL threshold 0.0157

B) All_75_75

% differences All 75 75
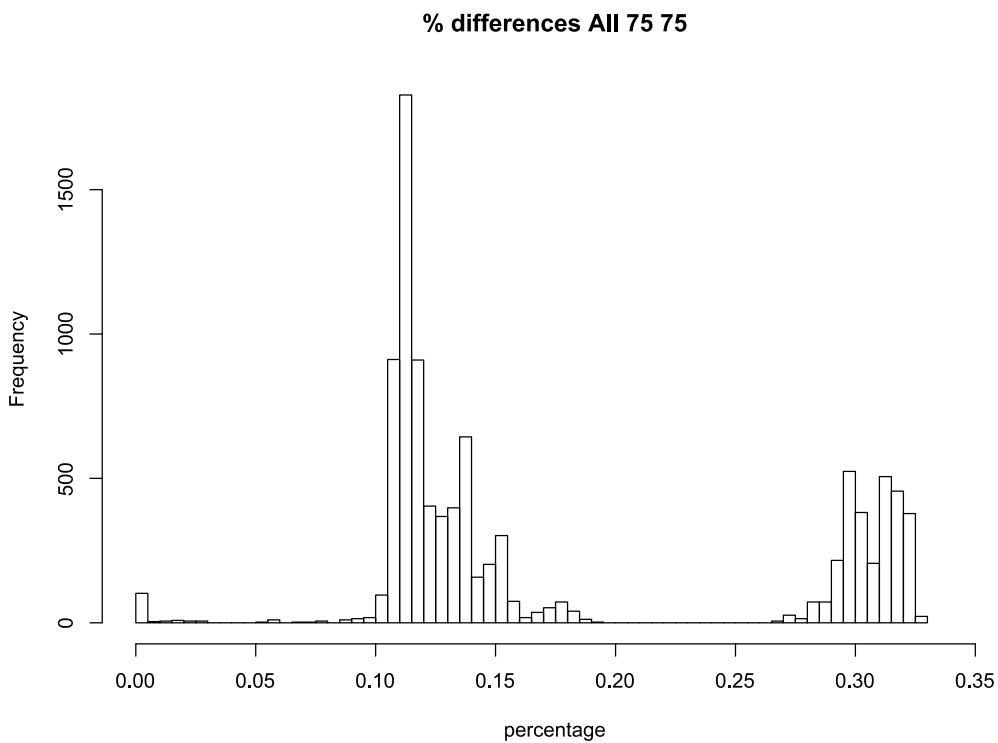
MLL threshold 0.06

C) Oman_95_75

% differences Oman 95 75
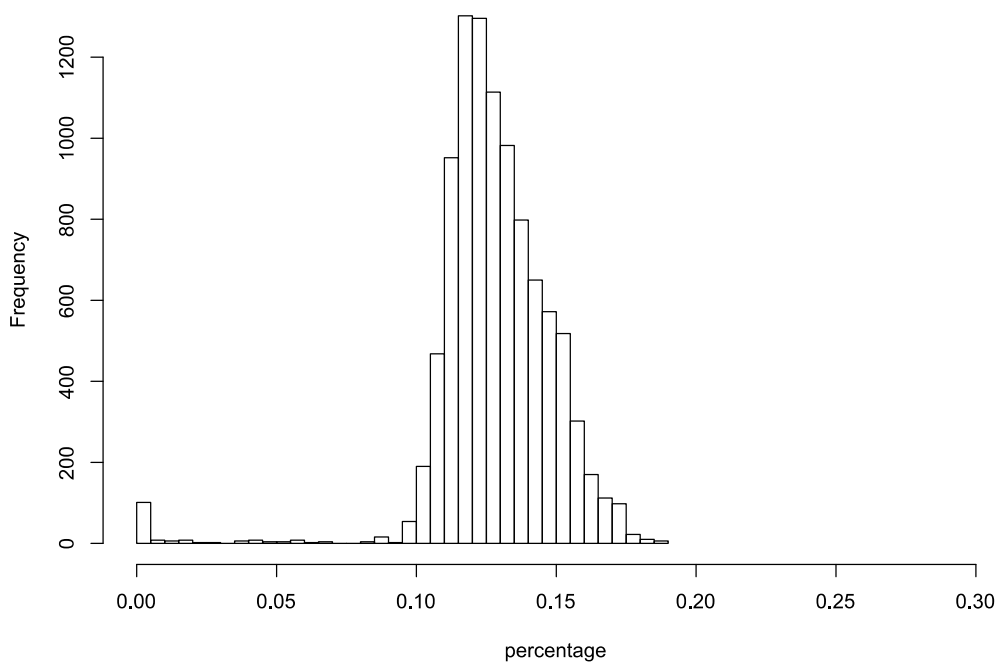
MLL threshold 0.0685

D) Oman_75_75

% differences Oman 75 75
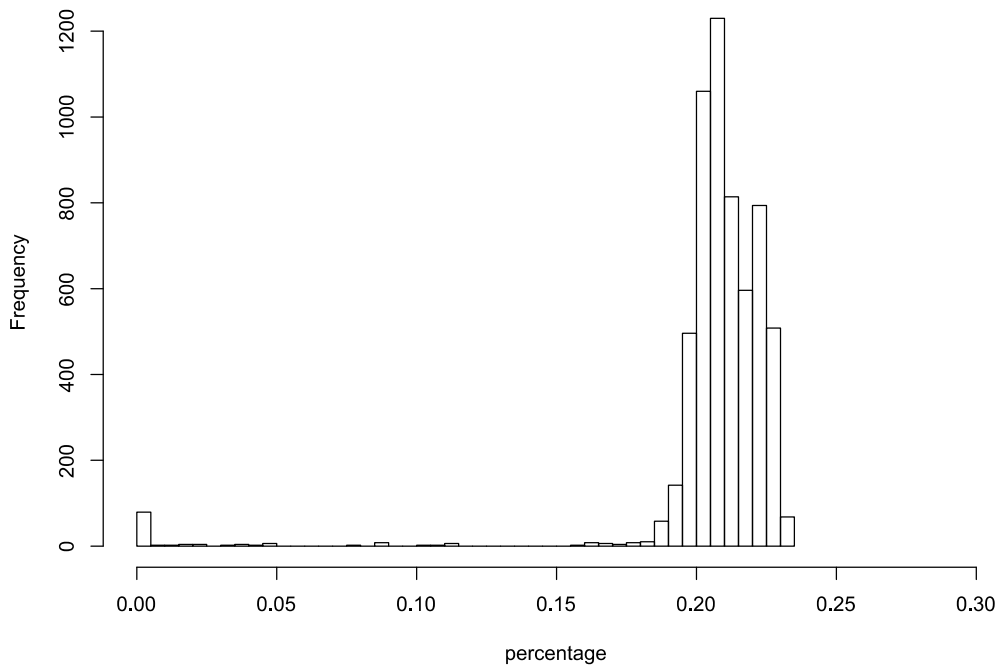
MLL threshold 0.09

E) Polynesia_95_75

% differences Polynesia 95 75
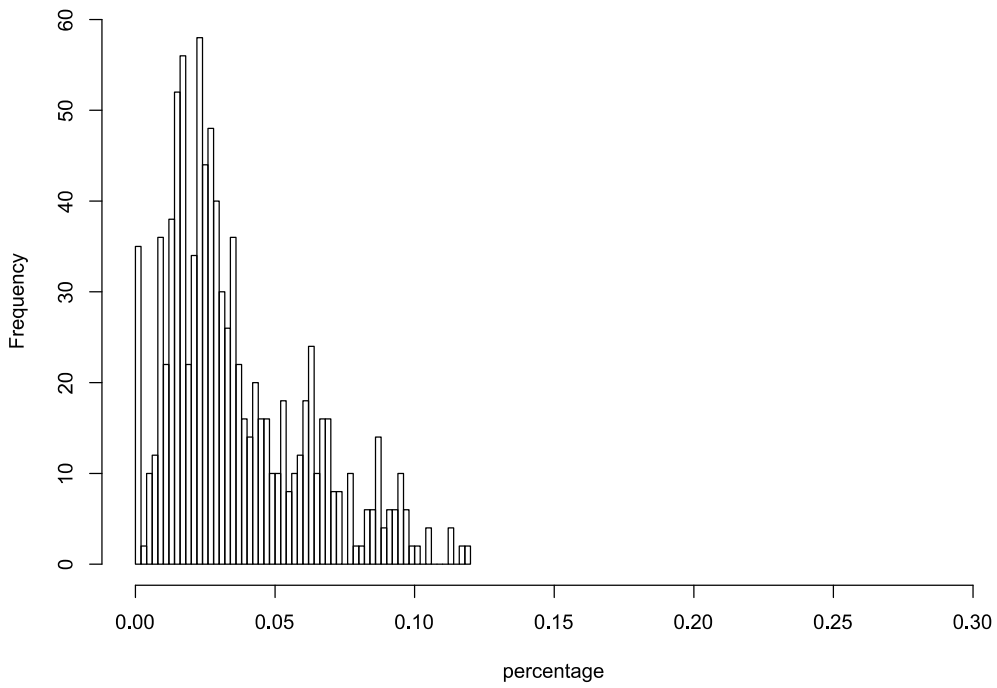
MLL threshold 0.002

F) Polynesia_75_75

% differences Polynesia 75 75
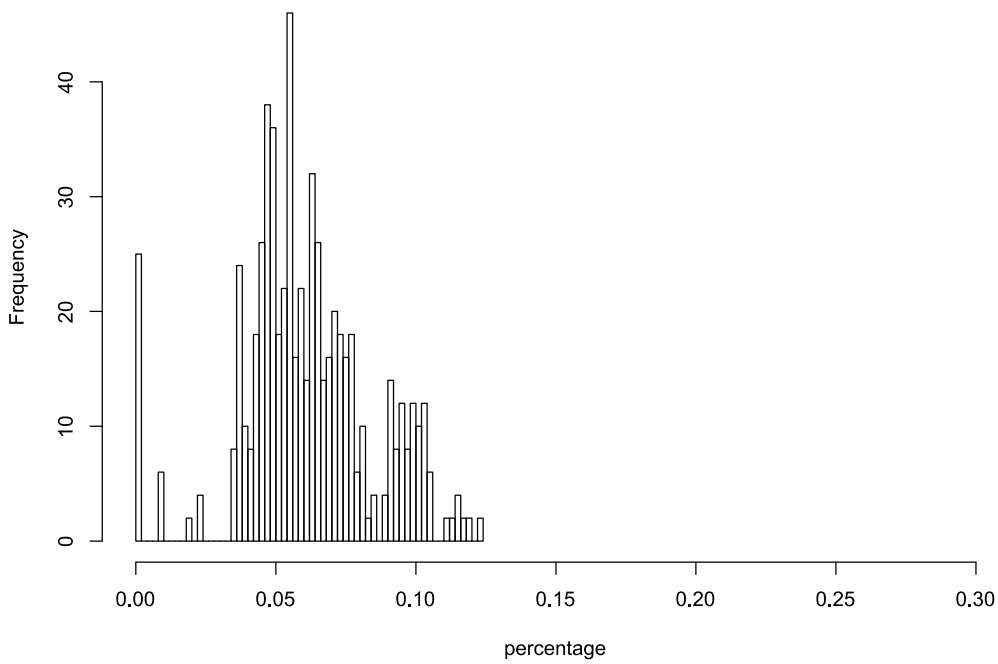
MLL threshold 0.04

Figure S6: networks based on the percentage of difference among individuals for the All, Oman and Polynesia 75_75 datasets. The colors indicate the corresponding mitochondrial lineage.

mt lineages :
Unknown
Type 3e
Type 3g
Beta/type 5a
Type 7a
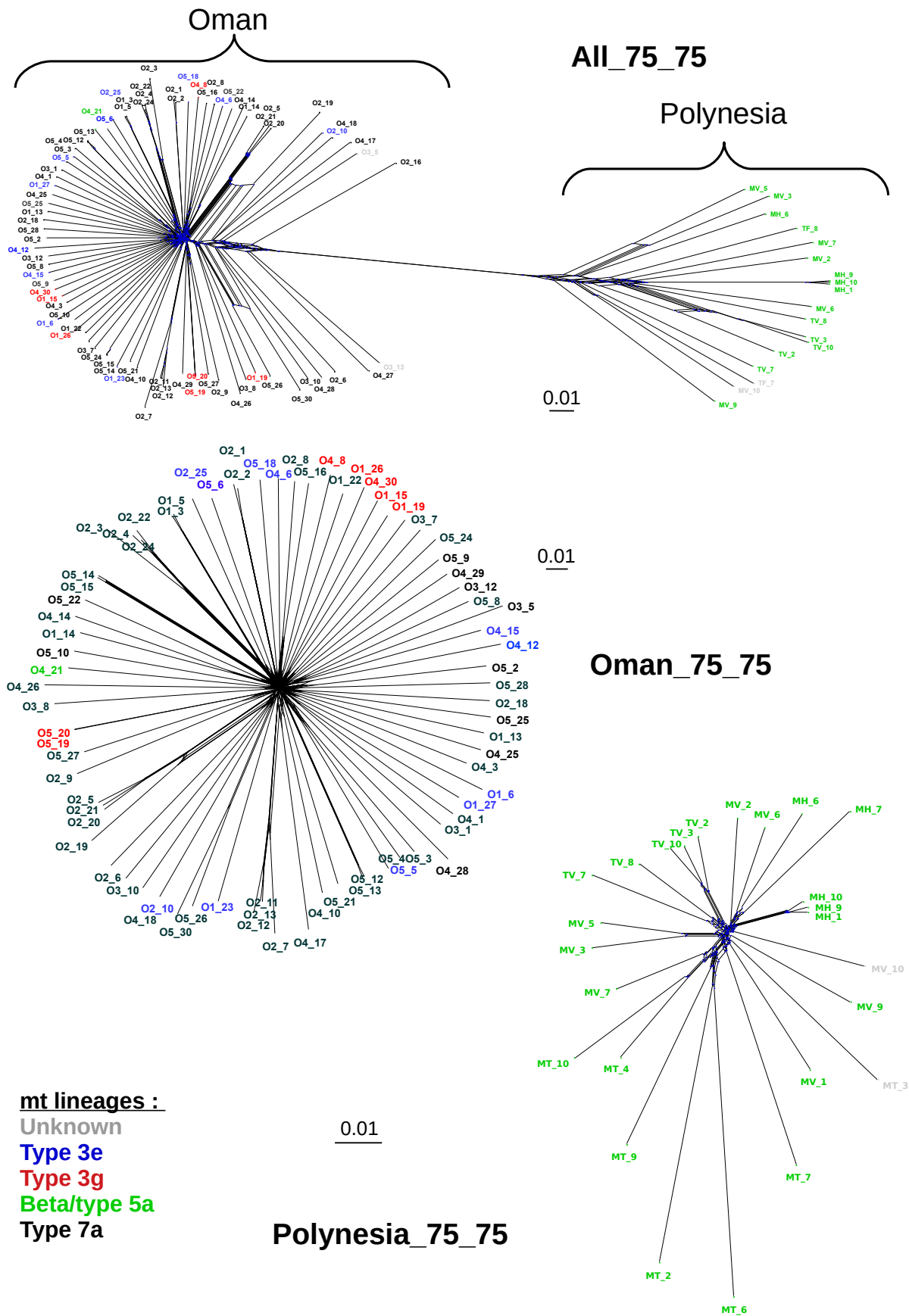
Figure S7 : histograms of individual inbreeding coefficients F estimated with VCFtools for the different datasets. The blue histograms correspond to all individuals, and the yellow histograms correspond to individuals potentially involved in MLLs with the previously defined distance thresholds. For visibility the x axis is different among figures. A red line at x = 0 is given for comparison.

All_95_75
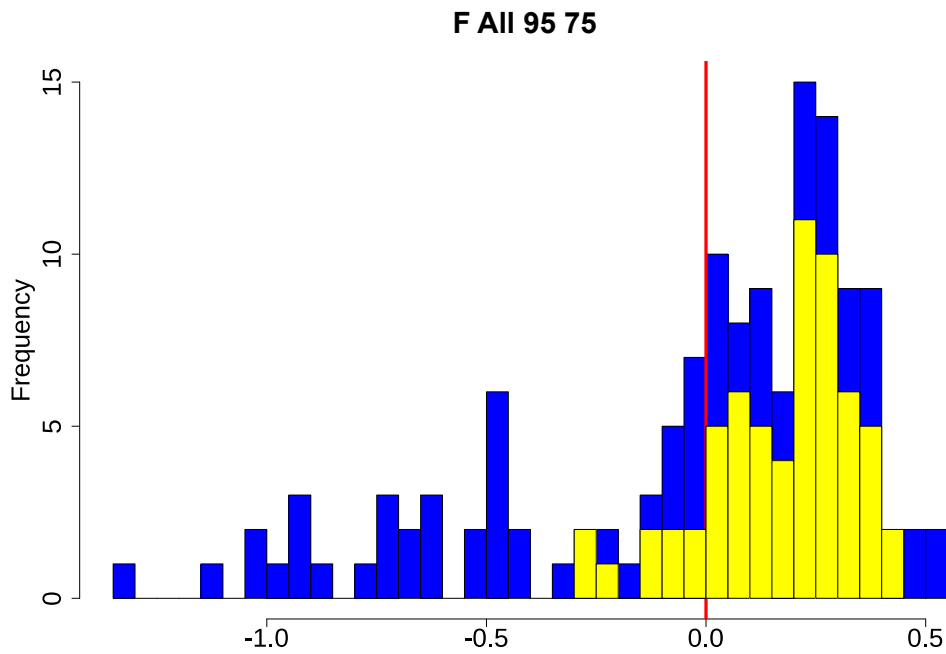
All_75_75
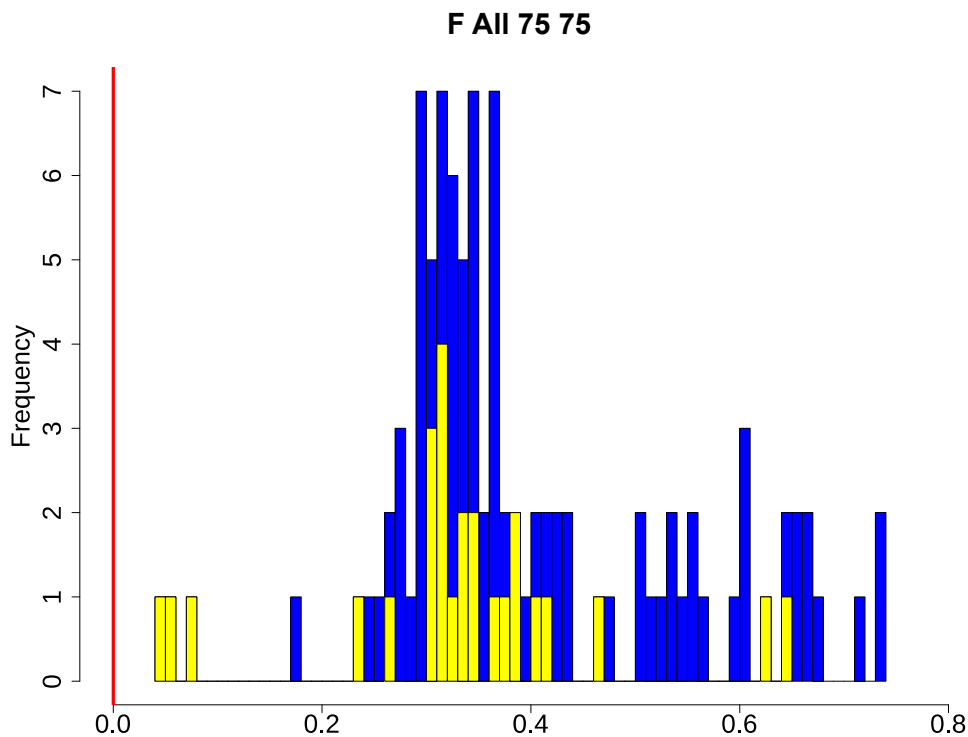
Oman_95_75

F Oman 95 75
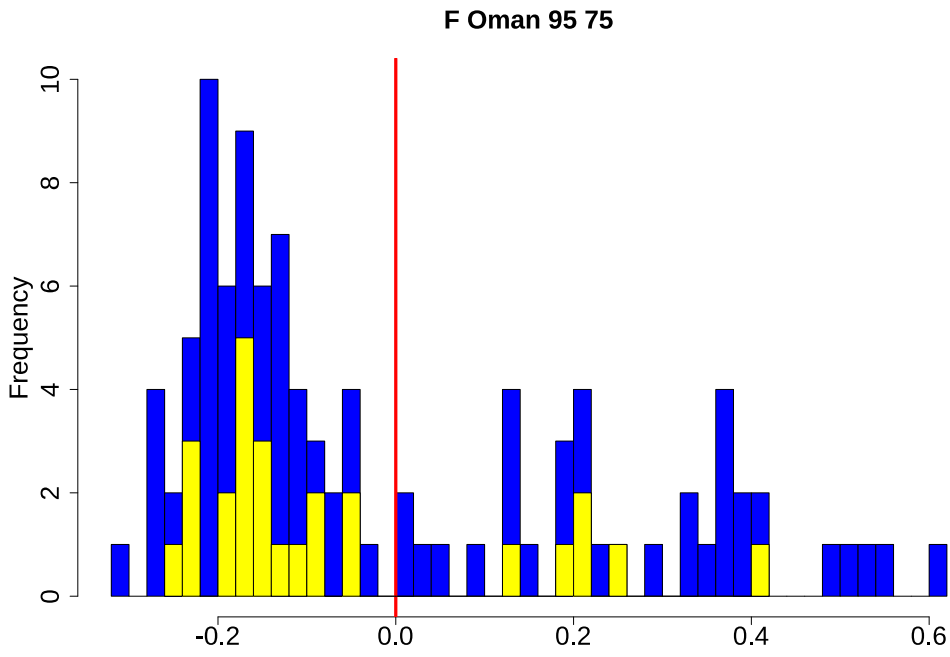
Oman_75_75

F Oman 75 75
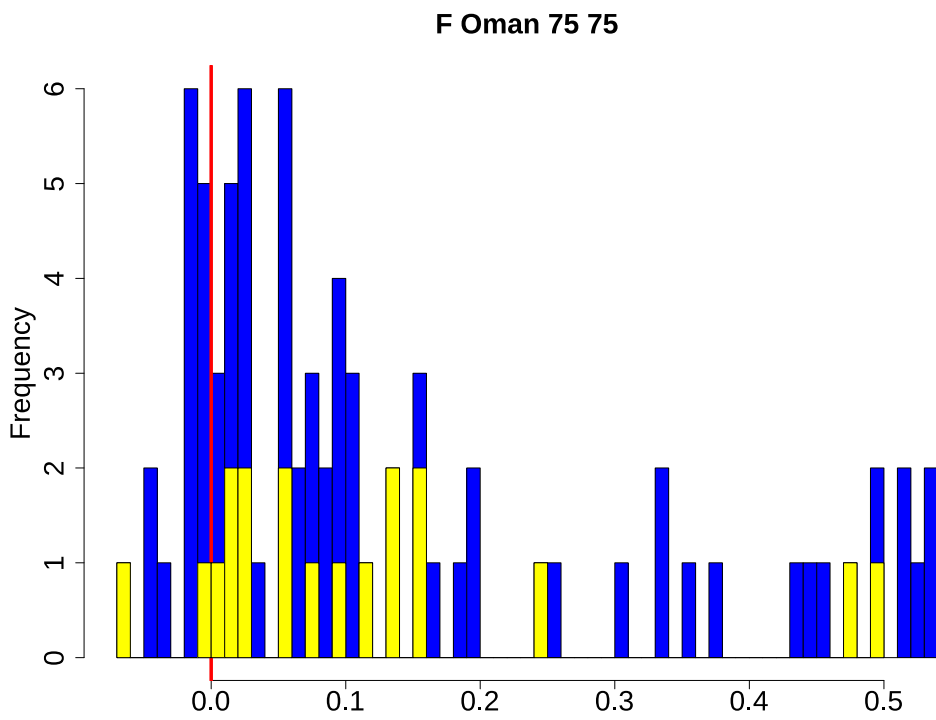
Polynesia_95_75

F Polyn 95 75
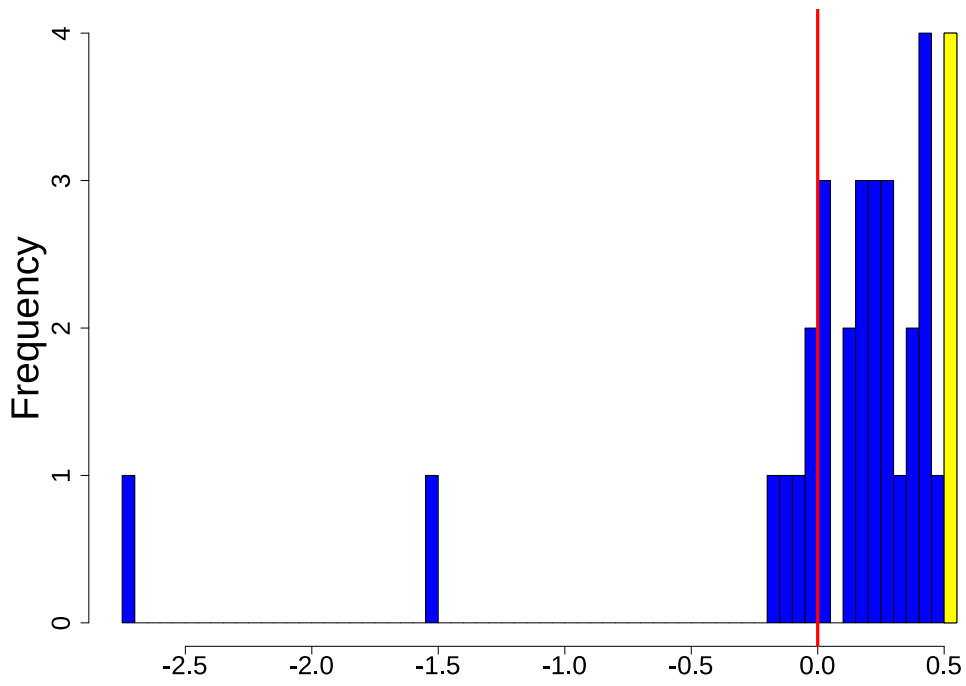
Polynesia_75_75

F Polyn 75 75
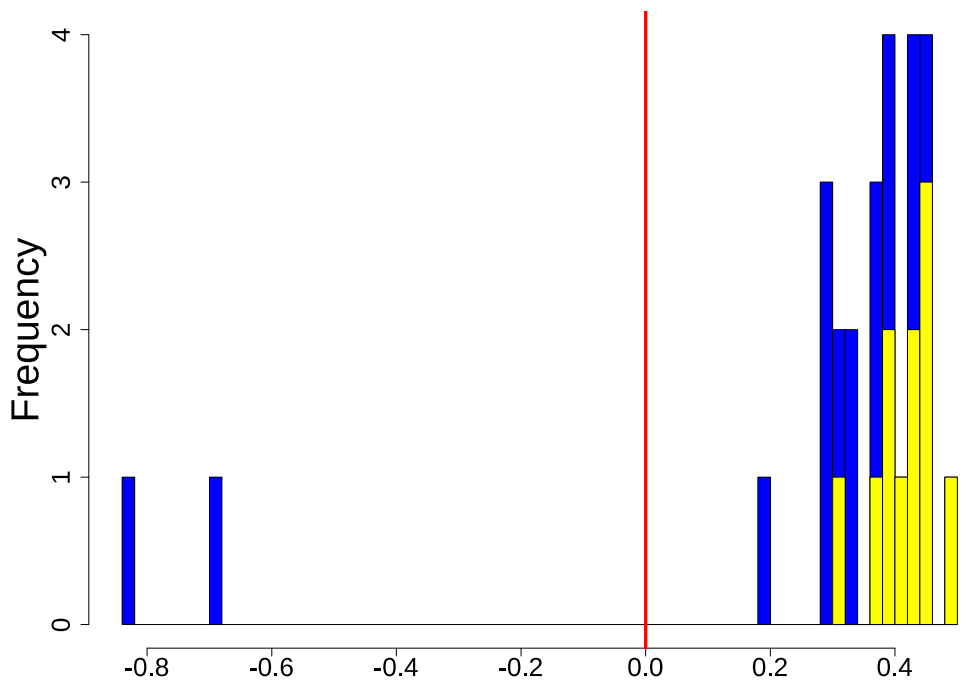
Figure S8 : distribution of the estimates of the index association $\bar{r}_d$ to study linkage disequilibrium in the different datasets : A) on the whole datasets, B) per mitochondrial lineages or per site. See the main text and the legend of Table S4 for details.
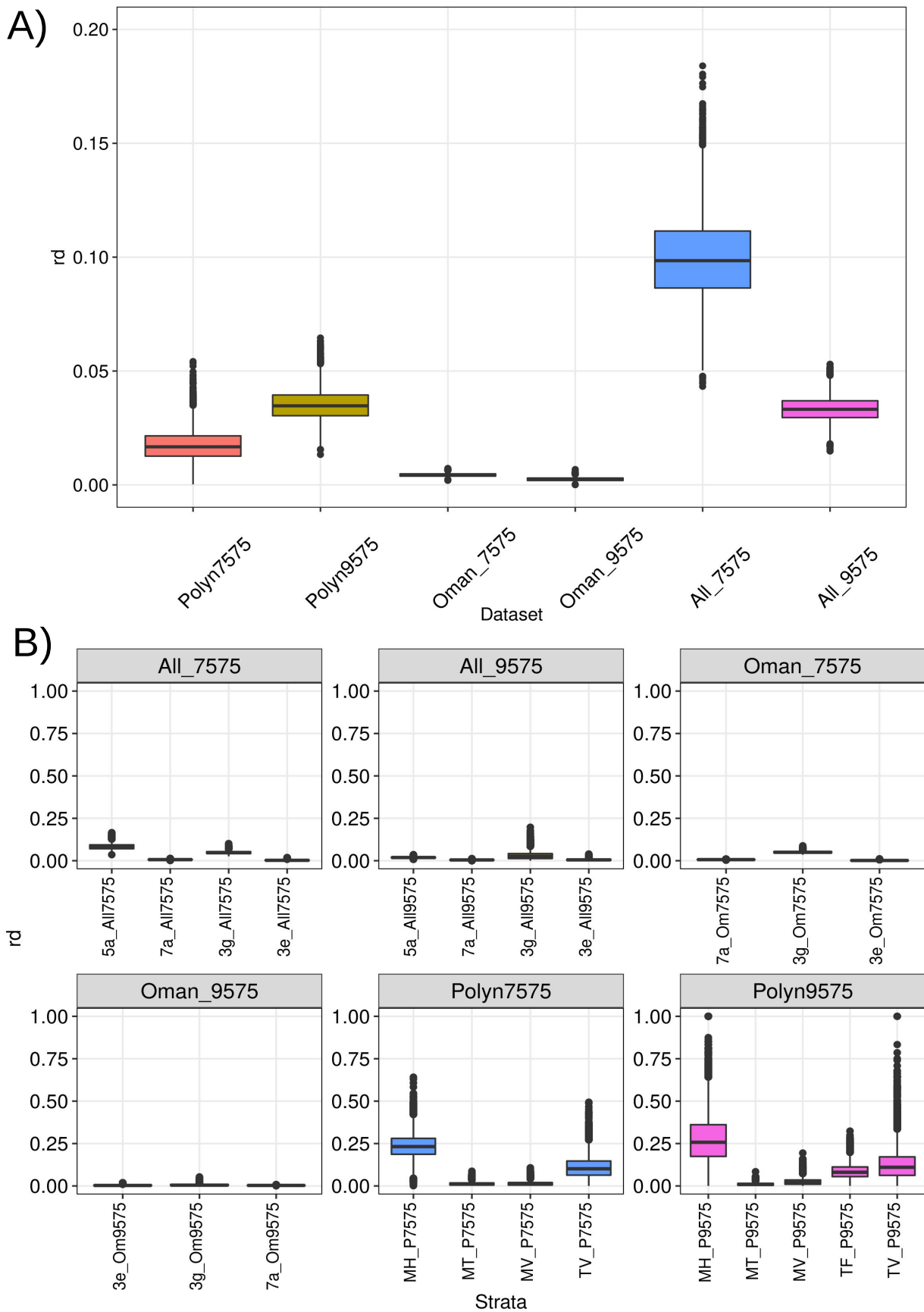
Figure S9: distribution of F_{ST} over all loci for the comparison between Oman and French Polynesia for A) the All_95_75 dataset (mean F_{ST} estimate: 0.105) and B) the All_75_75 dataset (mean F_{ST} estimate: 0.352).

A)

Fst Oman - Polynesia All 95 75

B)

Fst Oman - Polynesia All 75 75
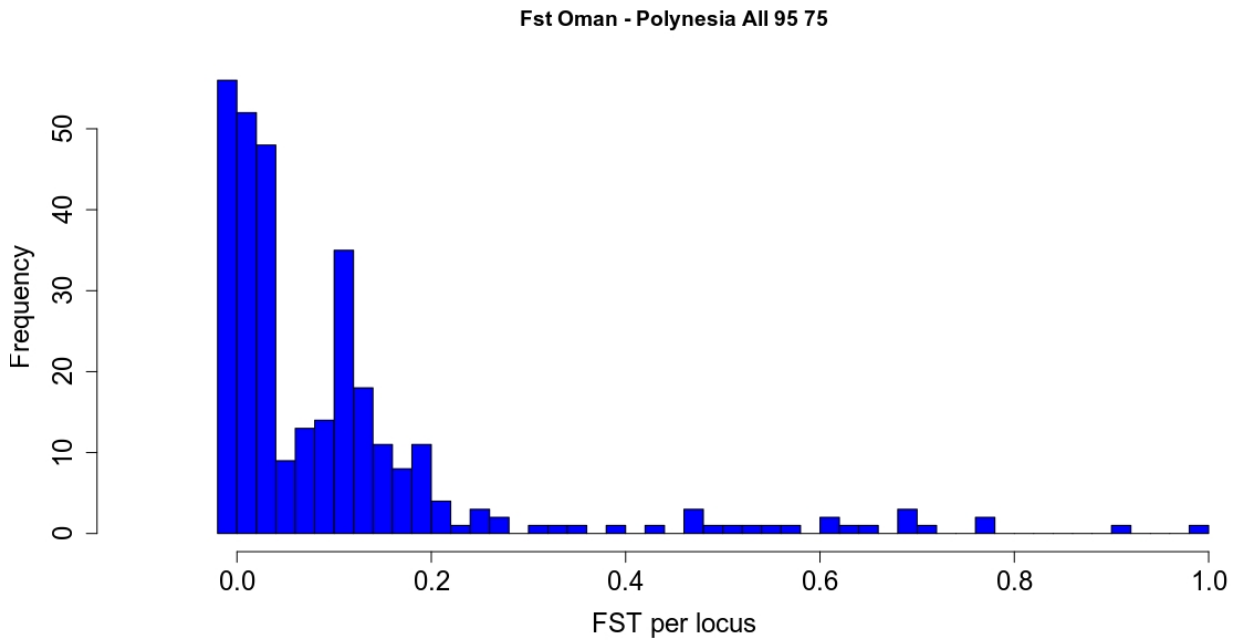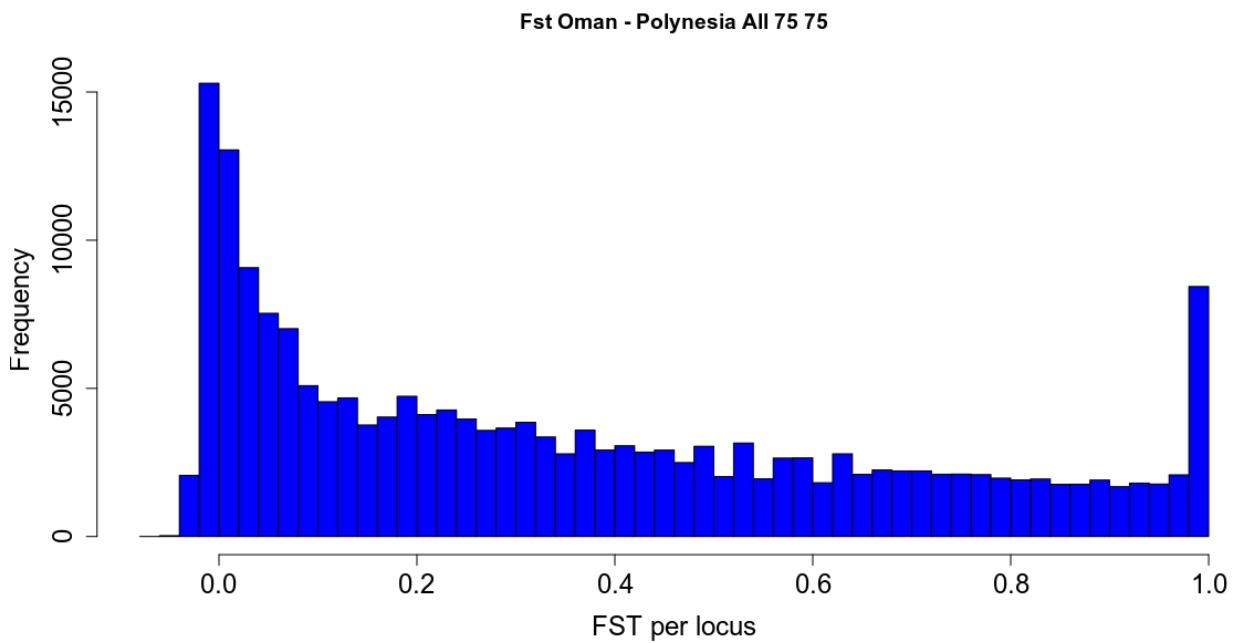
Figure S10: plot of individuals on the PCA axes 1 and 2 for A) the All_95_75, and B) the All_75_75 datasets. For All_95_75, the percentage of inertia was 9.2 for axis 1 and 5.6 for axis 2. For All_75_75, the percentage of inertia was 25.1 for axis 1 and 3.9 for axis 2. The individual dots are colored according to their ORF mitochondrial lineage, and a focus on the central part is given under each plot.

A) All_95_75

B) All_75_75

axis 2

axis 1
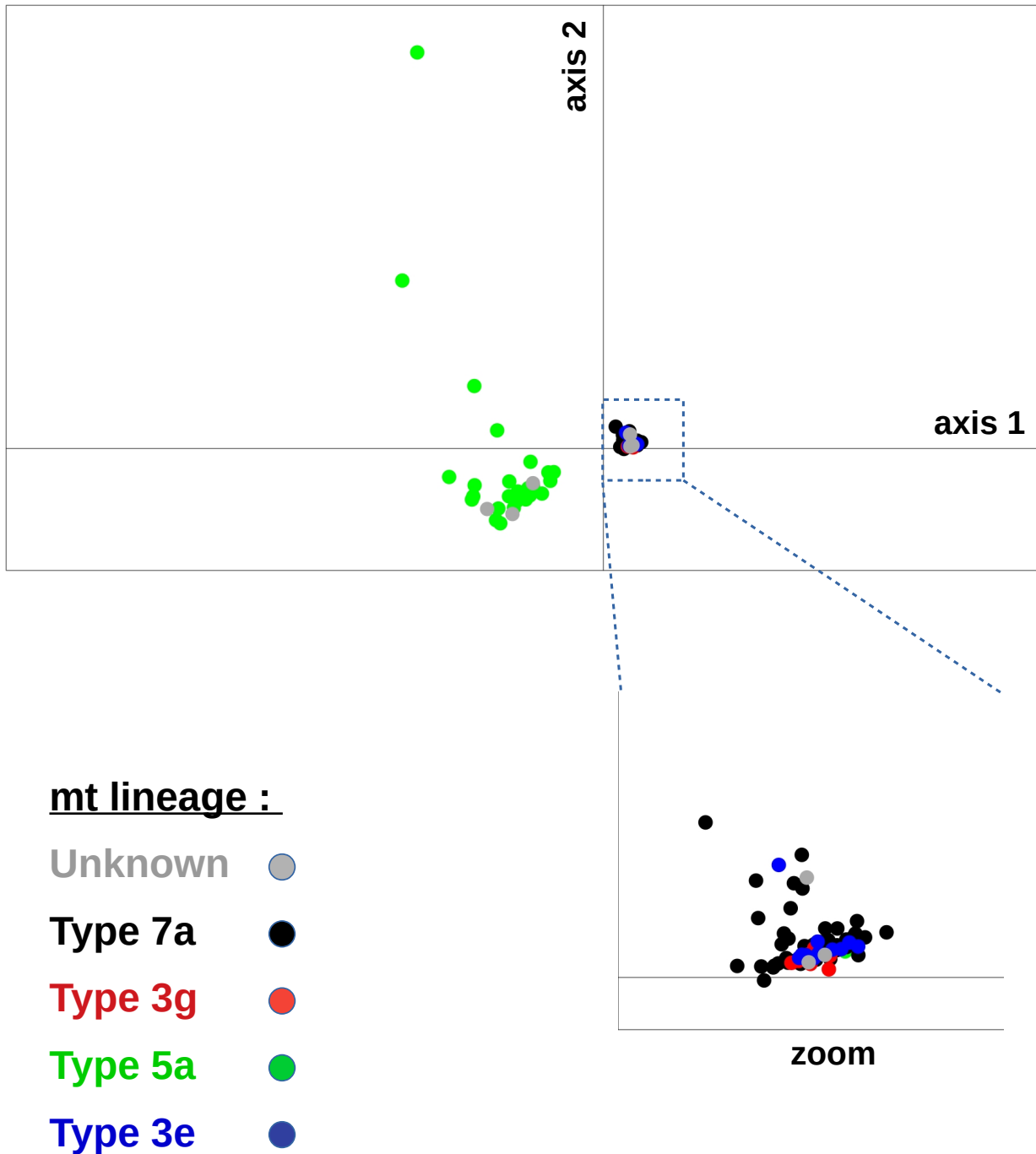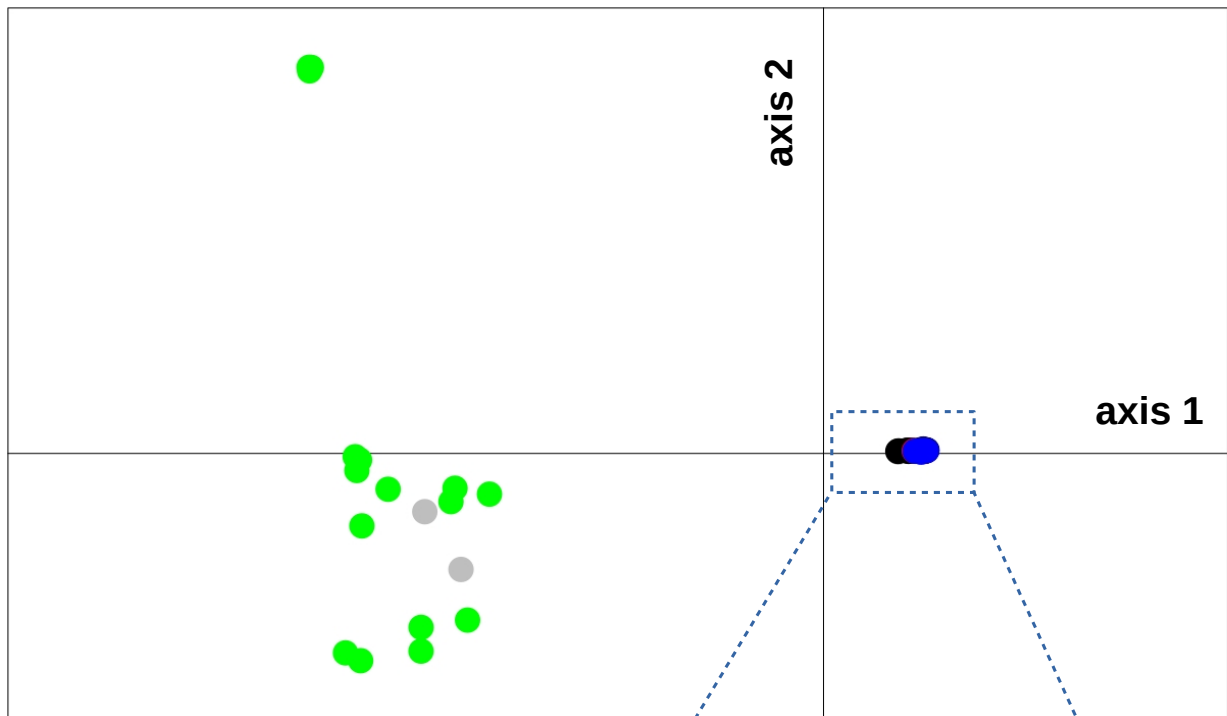
mt lineage :

Unknown ⬤

Type 7a ⬤

Type 3g ⬤

Type 5a ⬤

Type 3e ⬤
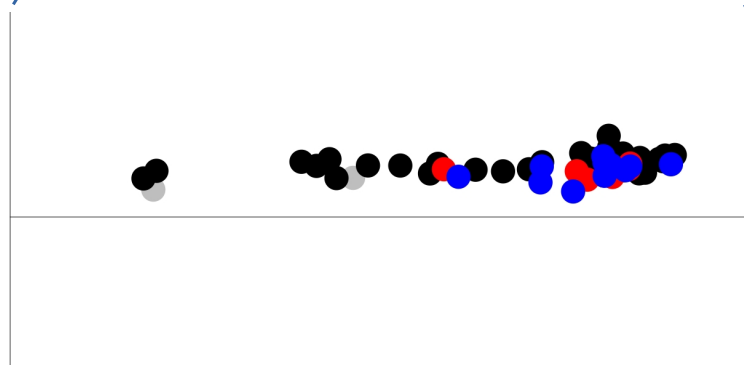
zoom

Figure S11: cross-entropy plots according to the number of clusters K for the snmf analyses on the different datasets: A) All_95_75, B) All_75_75, C) Oman_95_75, D) Oman_75_75, E) Polynesia_95_75, F) Polynesia_75_75.
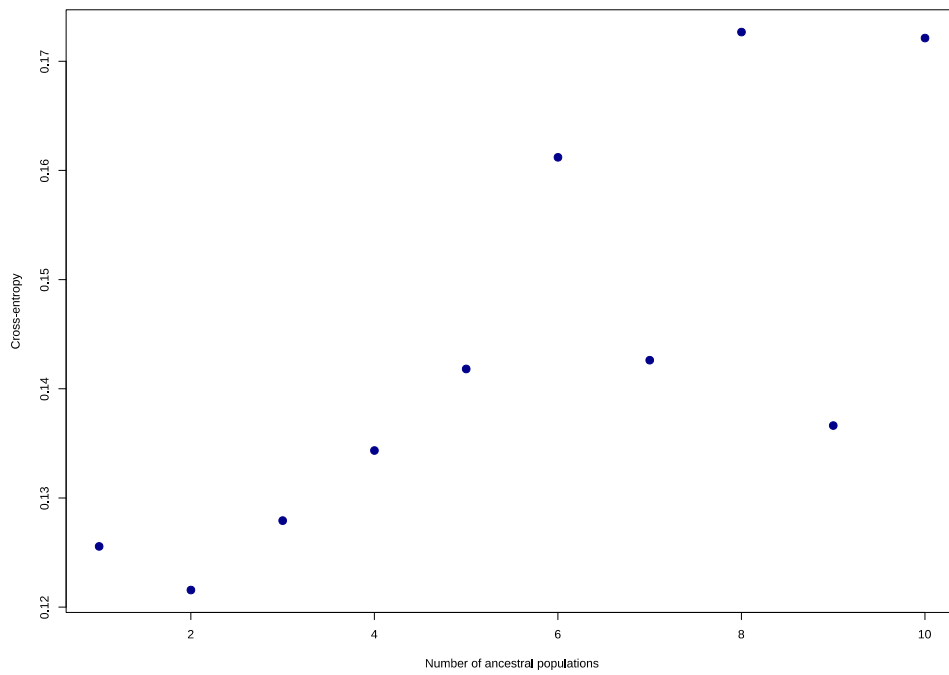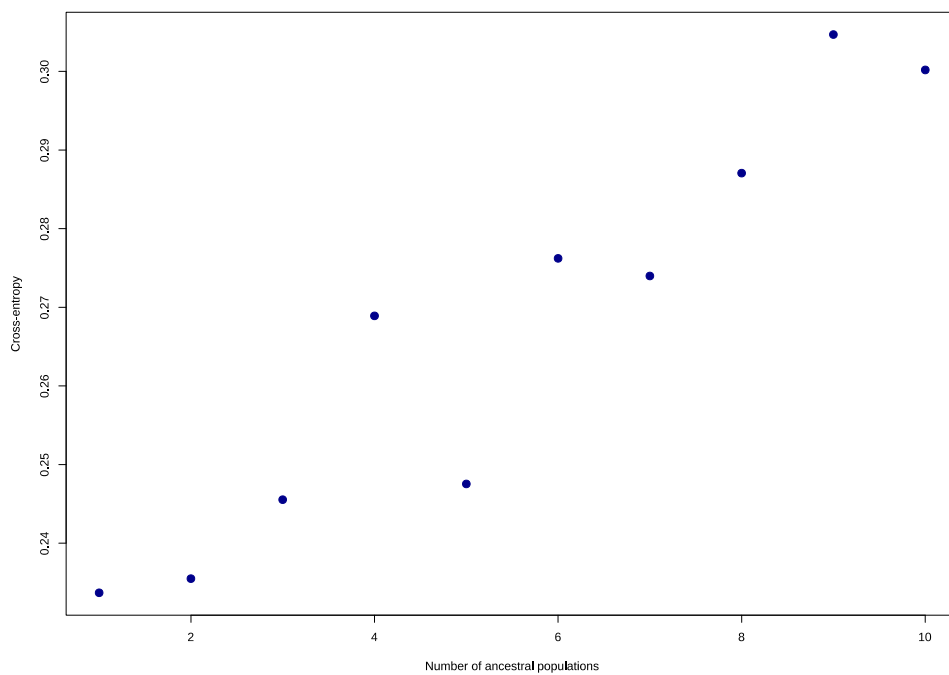
A)
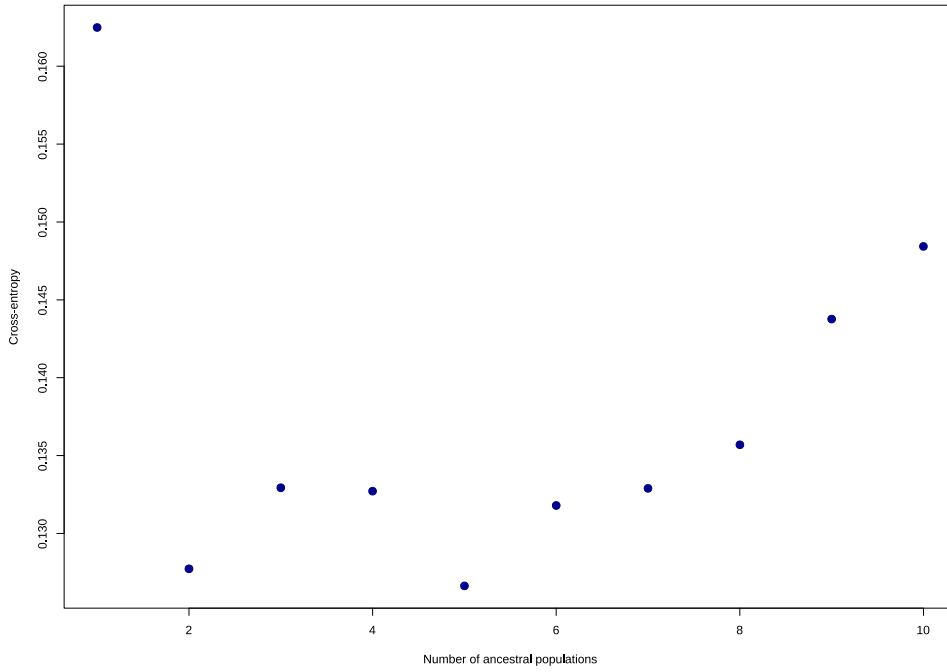
B)
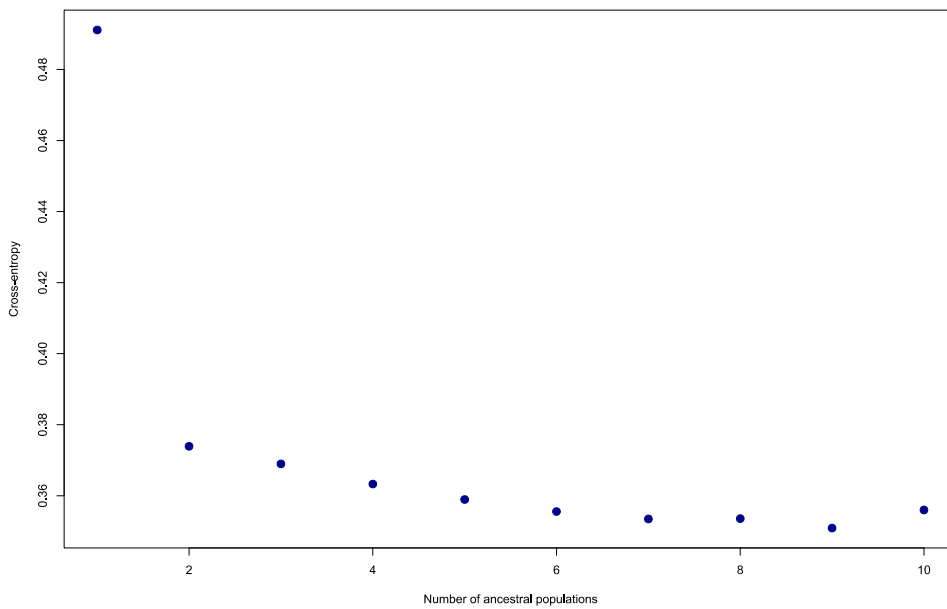
C)
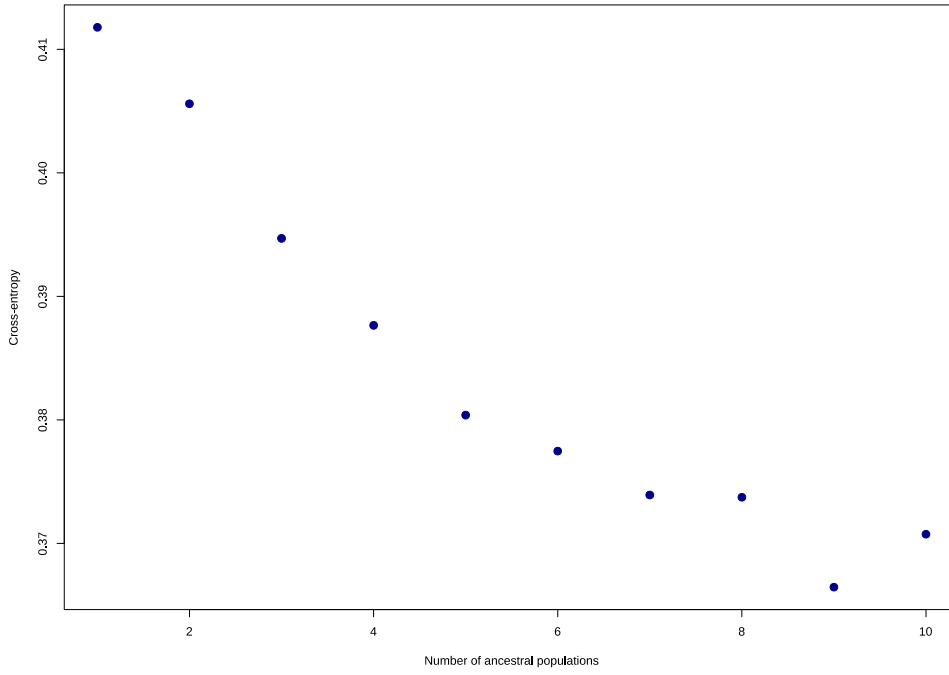
D)
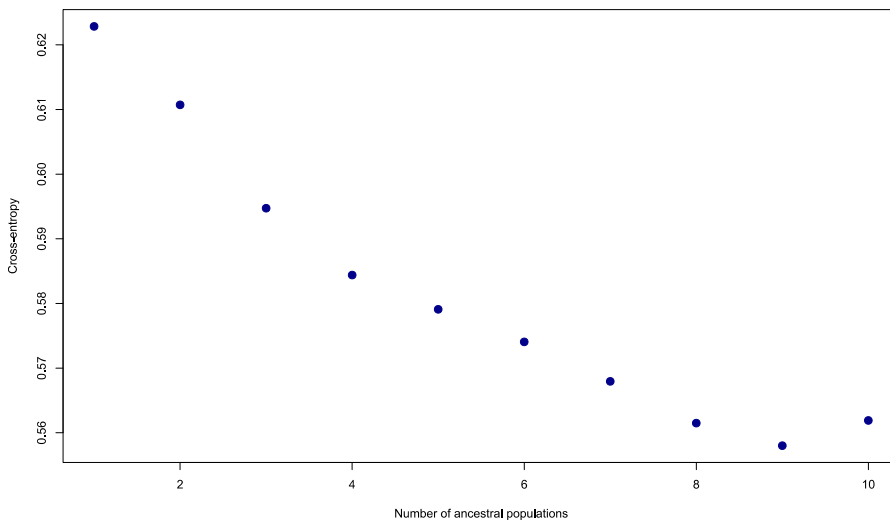
E)


F)

Results of individual-based simulations :

In this part we detail the results obtained for the analyses of individual-based simulations. We first present the figures, before developing the presentation of the results, by comparison with empirical data.

Figures:

Distribution of *F* estimates :

The figures below show the distribution of the average (A), maximum (B) and minimum (C) of the *F* parameter over individuals for different simulation configurations, with 30 simulations each. These distributions are compared to observed average, maximum and minimum values displayed as horizontal lines for the different datasets (see legends). Notation of the simulation configurations : panm : panmixia ; clon01, clon05, and clon09 : clonality present at rates 0.1, 0.5 and 0.9 respectively ; self_clon01 and self_clon05 : selfing present at rate 0.1, and clonality present at rates 0.1 and 0.5 respectively.

A) average of *F* parameter over simulation replicates ; horizontal lines give observed values for the different datasets : green : all, red : Oman, blue : French Polynesia, continuous line : 75_75, dotted line : 95_75.

F inbreeding average
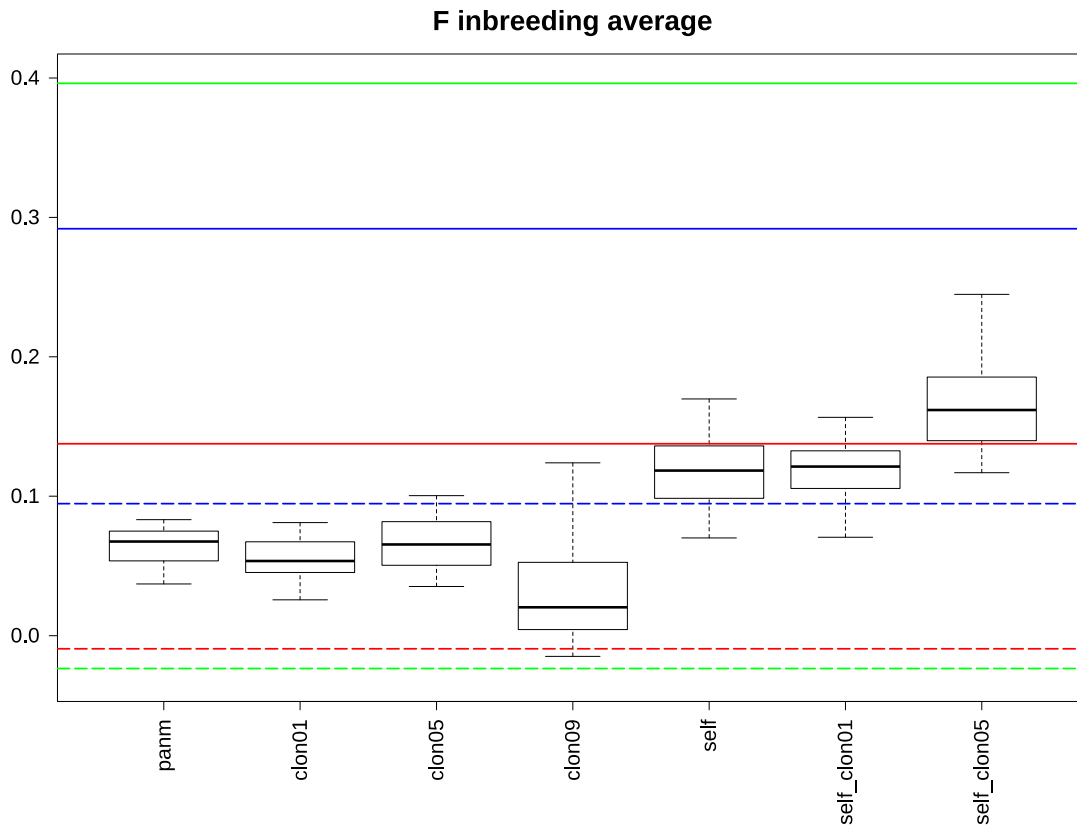
B) maximum of *F* parameter over simulation replicates ; horizontal lines give observed values for the different datasets : green : all, red : Oman, blue : French Polynesia, continuous line : 75_75, dotted line : 95_75.

F inbreeding max
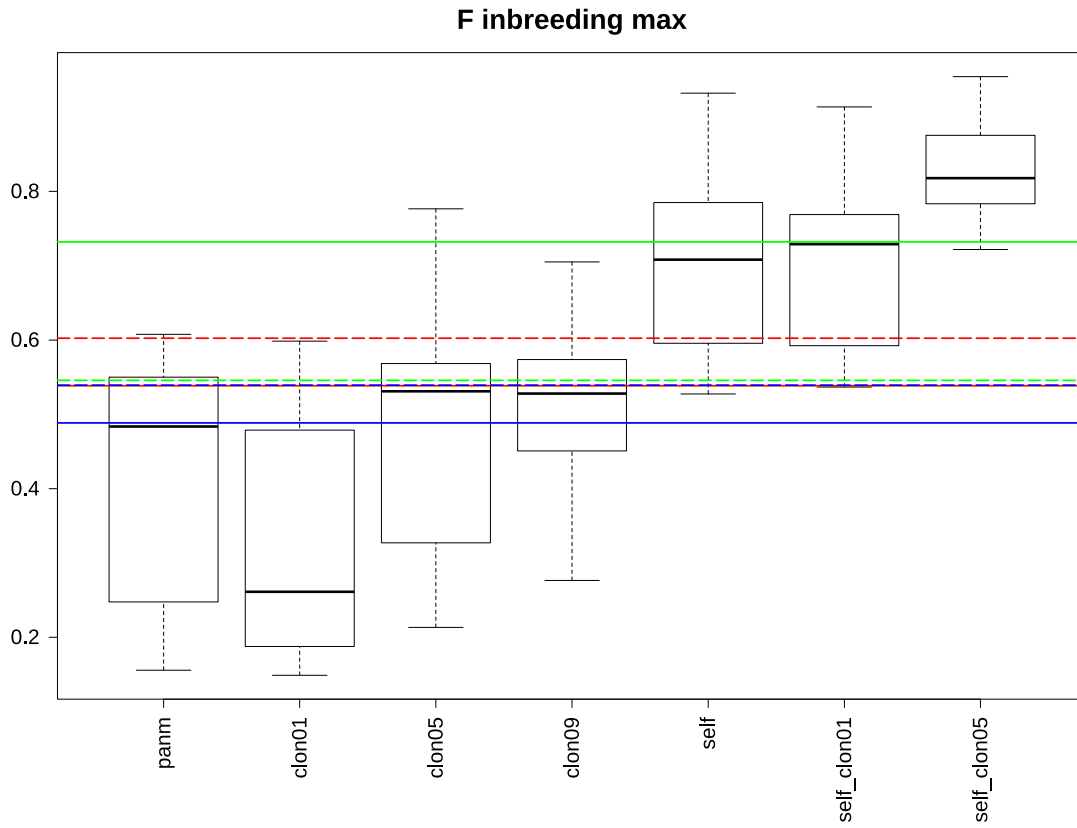
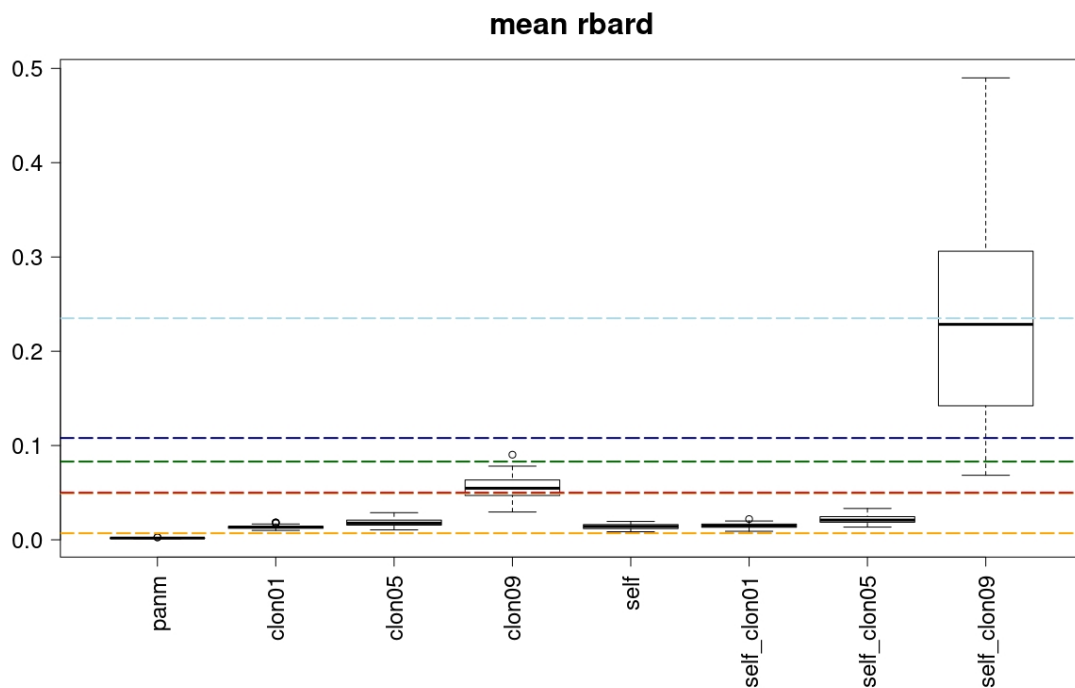**C)** minimum of *F* parameter over simulation replicates ; horizontal lines give observed values for the different datasets : green : all ; red : Oman ; blue : French Polynesia ; continuous line : 75_75 ; dotted line : 95_75.


F inbreeding min
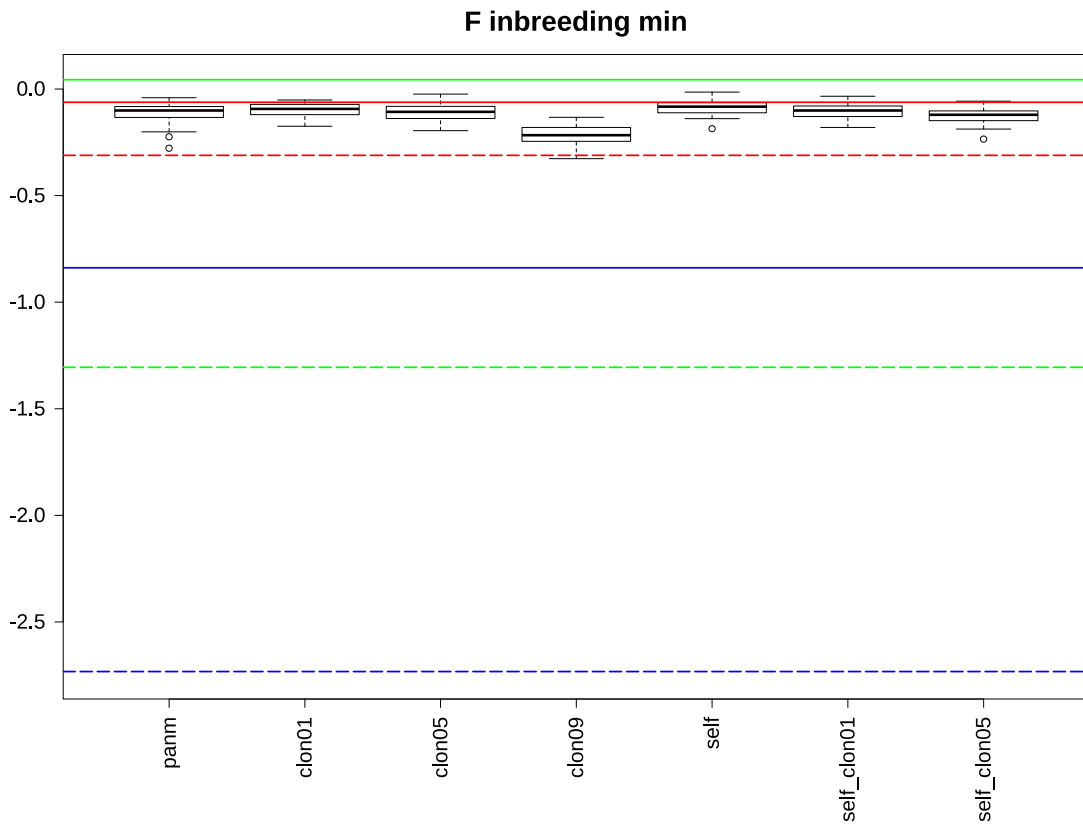
Analysis of linkage disequilibrium :

The figures below show the distribution of the mean (A) and standard deviation (B) of $\bar{r}_d$ over simulations. Note that $\bar{r}_d$ was here computed separately for the two simulated populations in each simulation,with mean and standard deviations computed across subsampling replicates in each population. These distribution are compared to samples with the two highest observed mean values obtained in the 75_75 datasets at the level of populations or mitochondrial lineages : All_75_75 : 3g (light green) and 5a (dark green) ; Oman_75_75 : 3g (red) and 7a (orange) ; Polynesia_75_75 : MH (light blue) and TV (dark blue) ; close values

A) mean of $\bar{r}_d$; the lines for All_75_75 3g (light green) and Oman_75_75 3g overlap (red).

mean rbard

B) standard deviation of $\bar{r}_d$; for clarity we did not plot the lines for Polynesia_75_75 MH (s.d. = 0.072) and TV (s.d. = 0.065).
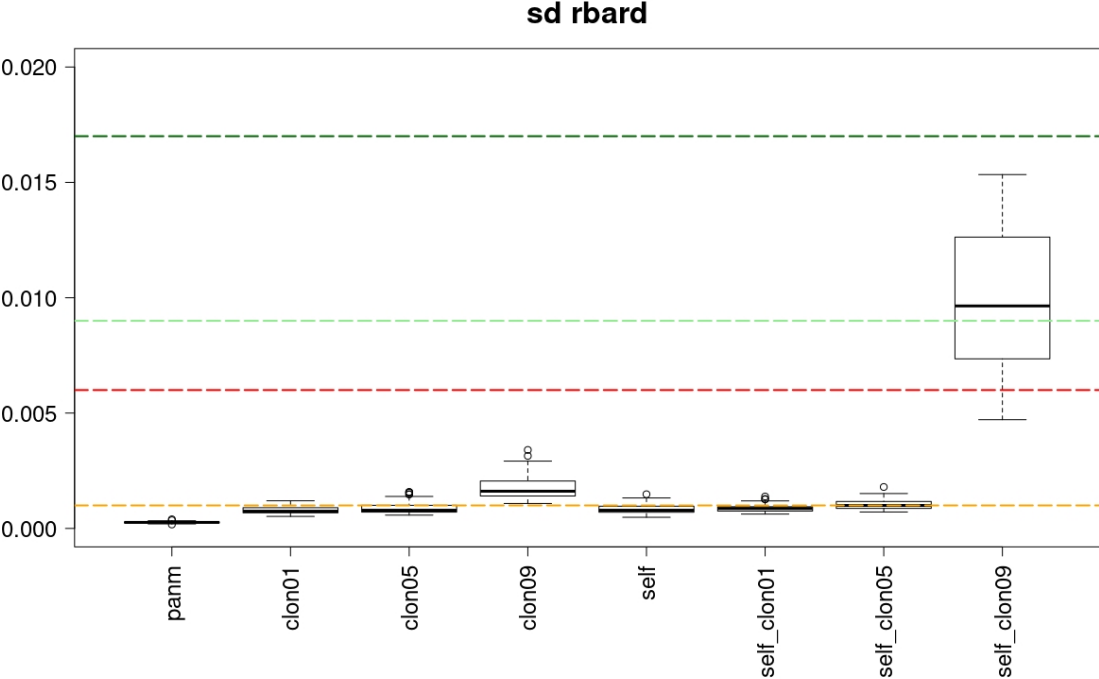
Distribution of F_{ST} estimate :

The figures below show the distribution of the F_{ST} (average over loci) between the two populations over 30 simulations, for the different simulation configurations. These distribution are compared to observed average, maximum and minimum values displayed as horizontal lines for the different datasets : green : all, red : Oman, blue : French Polynesia, continuous line : 75_75, dotted line : 95_75.

FST average
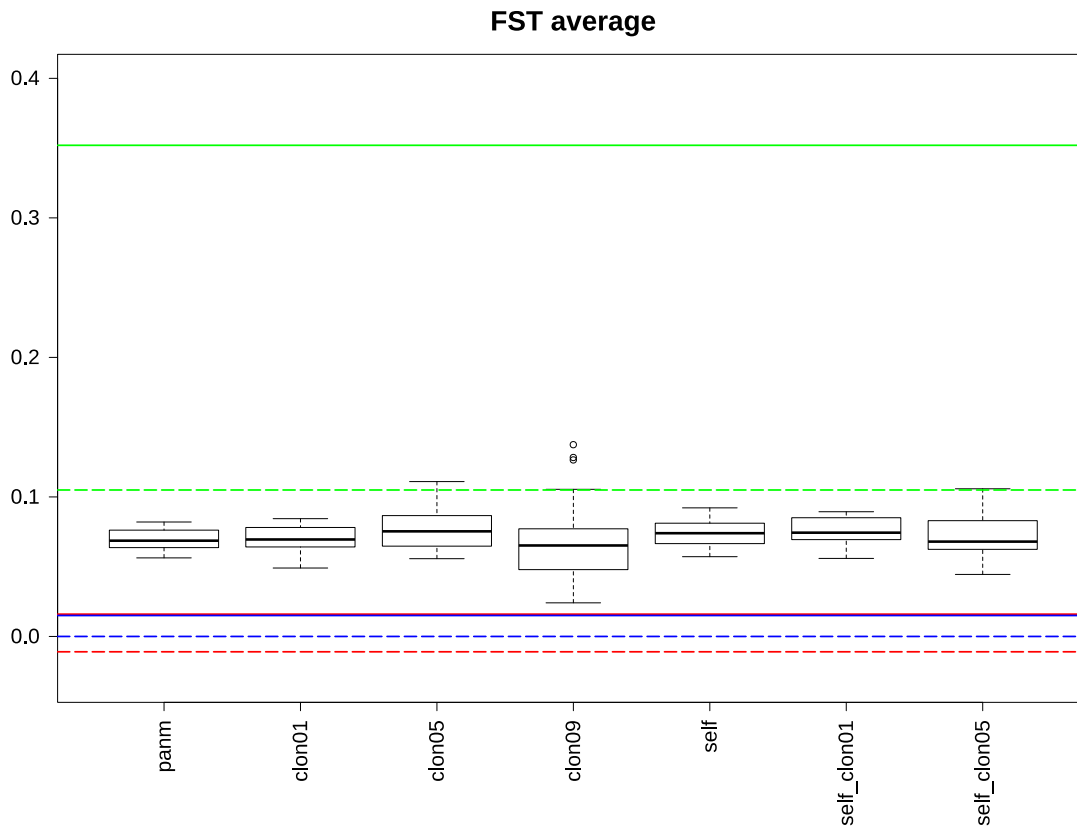
Discussion of the results of individual-based simulations:

The average and F values were higher for the simulations including selfing. The maximum F values were highly variable and tended to be higher for the configurations with the highest clonality rates (from 0.5), but were much higher for the simulations including selfing. Regarding the minimum F value, a decrease in the distribution was observed for the highest levels of clonality compared to other configurations. When comparing these results with observed data, one should note that for the All dataset, the F estimates were based on the pool of Oman and French Polynesia samples, whereas for the simulations we analysed the two simulated populations separately. The average F obtained in the All_75_75 and Polynesia_75_75 was higher than all values obtained with simulations. Regarding the maximum F values, the highest values observed in All_75_75 and Oman_95_75 appeared only compatible with simulations integrating selfing. For the minimum F values, no configuration allowed to recover such highly negative F as those observed in French Polynesia (both datasets), nor in All_95_75.

We analysed the linkage disequilibrium within each of the two simulated populations with the $\bar{r}_d$ index. An increase in $\bar{r}_d$ was observed with increasing clonality rate, mainly for the highest clonality rate tested here (0.9). A much higher increase both in mean and standard deviation was observed for a combination of selfing and a 0.9 clonality rate. We compared these values to the highest observed values obtained with the 75_75 datasets. Apart from one value in Oman, these mean observed values were similar or higher to those obtained with simulations at clonal rates of 0.9. The standard deviation for the observed estimates were usually very high, and in most cases these values were only approached by simulations with selfing and a 0.9 clonal rate.

Regarding the average F_{ST}, without any variation neither in census size nor migration rate, the resulting values were mostly similar among simulation configurations. A slight decrease and higher variance was nevertheless observed for the highest clonality rate (0.9). The observed F_{ST} for the Oman and French Polynesia datasets were lower than those obtained in all simulations, while the F_{ST} for the All datasets were higher than those obtained in almost all simulations, except a for a few simulations performed with the highest clonality rates.