

Population Genomics Unveils the Century-Old Invasion of the Seagrass *Halophila stipulacea* in the Mediterranean Sea

Catalina A. García-Escudero ^{1, 2}, Costas S. Tsigenopoulos ³, Tereza Manousaki ³, Alexandros Tsakogiannis ³, Núria Marbà ⁴, Salvatrice Vizzini ^{5, 6}, Carlos M. Duarte ⁷, Eugenia T. Apostolaki ^{1 *}

¹ Department of Biology , University of Crete, Voutes University Campus, P. O. Box 2208, 70013, Heraklion, Crete, Greece.

² Institute of Oceanography, Hellenic Centre for Marine Research, P. O. Box 2214, 71003, Heraklion, Crete, Greece.

³ Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, P. O. Box 2214, 71003, Heraklion, Crete, Greece.

⁴ Global Change Research Group, IMEDEA (CSIC-UIB), Institut Mediterrani d'Estudis Avançats, Miguel Marques 21, 07190, Esporles, Illes Balears, Spain

⁵ Dipartimento di Scienze della Terra e del Mare, Università degli Studi di Palermo, via Archirafi, 18, 90123, Palermo, Italy

⁶ CoNISMa, Consorzio Nazionale Interuniversitario per le Scienze del Mare, Piazzale Flaminio 9, 00196 Rome, Italy

⁷ Red Sea Research Centre (RSRC) and Computational Bioscience Research Center (CBRC), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

*Correspondence: Eugenia T. Apostolaki. Email: eapost@hcmr.gr

Haplotype-based analysis

The *denovo_map.pl* function from STACKS (Catchen et al. 2011, 2013) was utilized to call SNPs, preserving all SNPs per RAD locus. Haplotype-level statistics, including Φ_{ST} , haplotype count, haplotype diversity and gene diversity were derived from the output generated by the "populations" module in STACKS. These values were then averaged or summed for each population accordingly. For population structure analysis, the recommended approach for de-novo analysis outlined in the paper by Rivera-Colón and Catchen (2022) was followed. First, RADpainter was employed to compute the co-ancestry matrix, summarizing the nearest-neighbour haplotype relationships among individuals across the RAD loci (Malinsky et al. 2018). Subsequently, the co-ancestry matrix was integrated into the fineSTRUCTURE Markov chain Monte Carlo (MCMC) clustering algorithm. This involved 100,000 burn-in iterations (-x), followed by 100,000 sample iterations after the burn-in (-y), with an output thinning interval of 1000 (-z). Finally, fineSTRUCTURE was used to construct a tree, enabling a detailed exploration of the relationships between the inferred populations (Lawson et al. 2012; Malinsky et al. 2018).

In terms of genetic diversity, both haplotype and gene diversity exhibit higher values in the Red Sea compared to the Mediterranean Sea (Table SI1). Among Mediterranean populations, LIM (Cyprus) stands out with the highest values, while other populations within the Mediterranean show relatively consistent levels of diversity. Based on the pairwise Φ_{ST} (Table SI2) the genetic divergence increase following the direction of the spread from the Red Sea to the Western Mediterranean. Moreover, TER and SAN populations both situated in Italy, exhibit negative Φ_{ST} values. These values suggest that genetic differentiation at the haplotype level is lower than what would be expected by random chance, mirroring the patterns observed with F_{ST} obtained from the biallelic SNP analysis. Based on the clear separation in

the dendrogram between exotic and native populations, together with the low coancestry values between them, indicate that exotic populations are genetically distinct from the native populations (Fig. SI1), increasing by following the direction of the spread, as seen previously with the biallelic SNP results.

Table SI1. Summary of haplotype-based genetic diversity estimates per population

Pop ID	Haplotype count	Haplotype diversity	Gene diversity
YAN	1,308	0.422	0.209
HAD	1,312	0.510	0.257
LIM	1,283	0.357	0.180
MAR	1,306	0.288	0.154
LIO	1,300	0.276	0.141
SOU	1,307	0.326	0.154
TER	1,312	0.278	0.140
SAN	1,309	0.273	0.144

Table SI2. Pairwise genetic distances (Φ_{ST}) between sampling sites. Site codes: YAN and HAD (Saudi Arabia), LIM (Cyprus), MAR, LIO, and SOU (Greece) and TER and SAN (Italy).

	YAN	HAD	LIM	MAR	LIO	SOU	TER	SAN
YAN								
HAD	0.192							
LIM	0.240	0.135						
MAR	0.305	0.207	0.134					
LIO	0.299	0.206	0.129	0.224				
SOU	0.303	0.214	0.156	0.208	0.187			
TER	0.345	0.251	0.230	0.255	0.292	0.268		
SAN	0.322	0.226	0.202	0.237	0.278	0.254	-0.143	

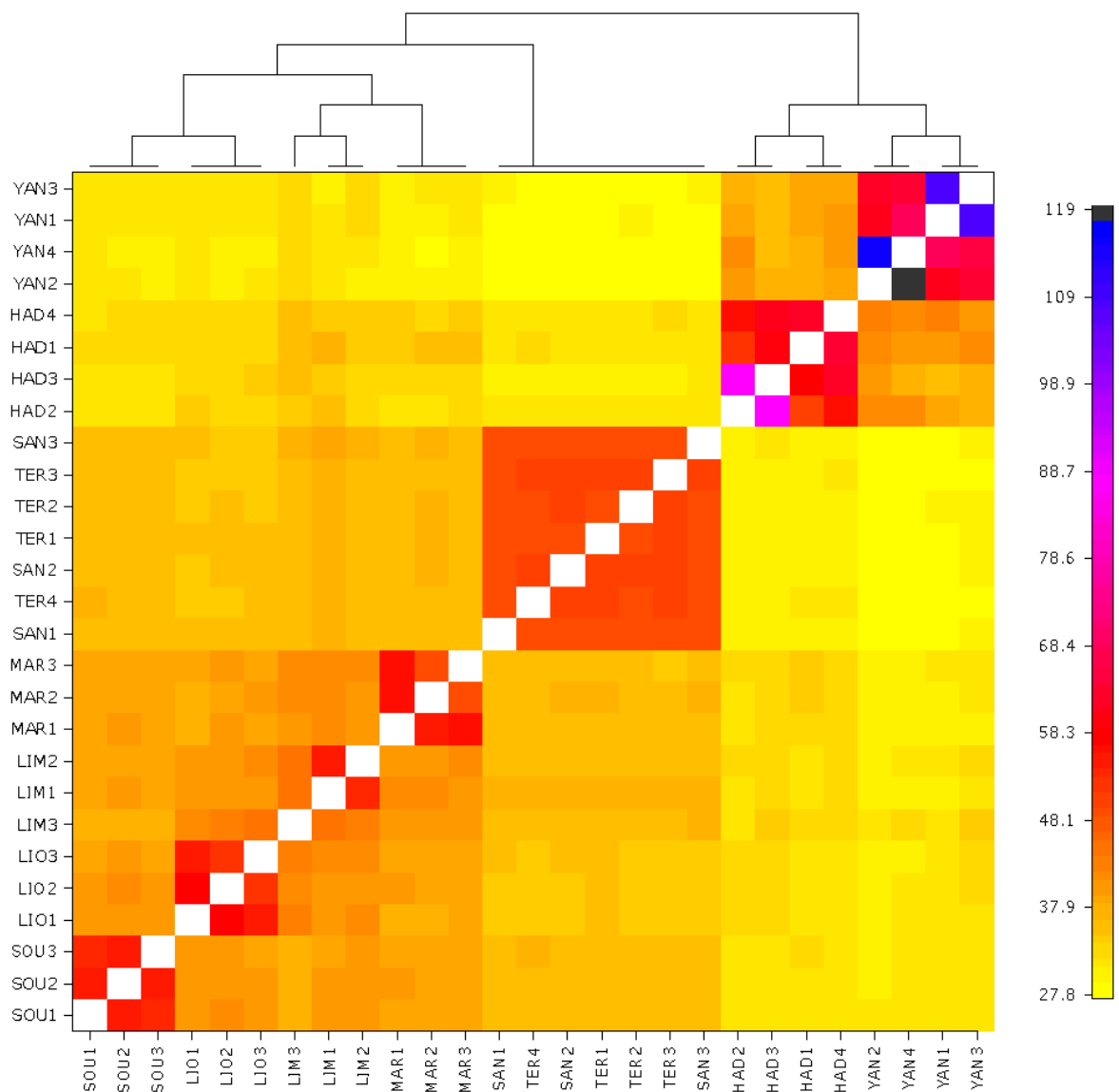


Fig. S1 Haplotype-based clustered fineRADstructure coancestry matrix between all 27 MLGs. The scale indicates the estimated coancestry (RAD loci). Site codes: YAN and HAD (Saudi Arabia), LIM (Cyprus), MAR, LIO, and SOU (Greece) and TER and SAN (Italy)

Table SI3. Upper and lower confidence interval based on 1000 permutations for F_{IS} per site. Site codes: YAN and HAD (Saudi Arabia), LIM (Cyprus), MAR, LIO, and SOU (Greece) and TER and SAN (Italy).

	F_{IS}	lower limit	upper limit
YAN	-0.305	-0.384	-0.252
HAD	-0.079	-0.104	-0.002
LIM	-0.285	-1.471	-0.902
MAR	-0.420	-0.486	-0.341
LIO	-0.488	-0.724	-0.541
SOU	-0.966	-1.075	-0.980
TER	-0.944	-1.022	-0.935
SAN	-0.932	-0.989	-0.924

Table SI4. Upper (above diagonal) and lower (below diagonal) confidence interval based on 1000 permutations for pairwise genetic distances (F_{ST}) between sampling sites. Site codes: YAN and HAD (Saudi Arabia), LIM (Cyprus), MAR, LIO, and SOU (Greece) and TER and SAN (Italy).

	YAN	HAD	LIM	MAR	LIO	SOU	TER	SAN
YAN		0.335	0.432	0.479	0.487	0.500	0.551	0.528
HAD	0.276		0.288	0.353	0.370	0.393	0.424	0.390
LIM	0.354	0.212		0.302	0.320	0.431	0.475	0.462
MAR	0.408	0.286	0.214		0.387	0.396	0.441	0.422
LIO	0.417	0.303	0.235	0.313		0.393	0.489	0.477
SOU	0.436	0.328	0.343	0.329	0.317		0.464	0.462
TER	0.487	0.359	0.394	0.370	0.415	0.397		0.008
SAN	0.463	0.325	0.383	0.350	0.404	0.395	-0.002	

REFERENCES

- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology* 22:3124–3140. doi: 10.1111/mec.12354
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, genomes, genetics* 1:171–182. doi: 10.1534/g3.111.000240
- Lawson DJ, Hellenthal G, Myers S, Falush D (2012) Inference of Population Structure using Dense Haplotype Data. *PLOS Genetics* 8:e1002453. doi: 10.1371/journal.pgen.1002453
- Malinsky M, Trucchi E, Lawson DJ, Falush D (2018) RADpainter and fineRADstructure: Population Inference from RADseq Data. *Molecular Biology and Evolution* 35:1284–1290. doi: 10.1093/molbev/msy023
- Rivera-Colón AG, Catchen J (2022) Population Genomics Analysis with RAD, Reprised: Stacks 2. *Methods Mol Biol* 2498:99–149. doi: 10.1007/978-1-0716-2313-8_7