

Supplementary Information

Title: Admixture of Holothurian species in the Hellenic Seas (eastern Mediterranean) as revealed by RADseq

Georgios A. Gkafas, Joanne Sarantopoulou, Chrysoula Apostologamvrou, Chryssanthi Antoniadou, Athanasios Exadactylos, Georgios Fleris, Dimitris Vafidis

Material and methods

S1. AIC calculations and model comparison

The Akaike Information Criterion (AIC) was used to evaluate the 5 demographic scenarios. The chosen model is the one that minimizes the Kullback-Leibler distance between the model and the truth. We used the formula:

$$AIC = -2 (\ln (\text{likelihood})) + 2 K$$

, where likelihood is the probability of the data given a model and K is the number of free parameters in the model. AIC scores are shown as ΔAIC scores (so the best model has a ΔAIC of zero).

Akaike weights represent the relative likelihood of a model. To calculate them, for each model we first calculate the relative likelihood of the model, using the formula:

$$\exp(-0.5 * \Delta AIC),$$

The Akaike weight for each model is this value divided by the sum of these values across all models. According to Burnham and Anderson (2002), as a rule of thumb, a $\Delta AIC < 2$ suggests substantial evidence for the model. In this case, indeed model 2, 3 and 4 meet the criteria. Further we ascribe a probability to the models via

$$p = \exp (-\Delta AIC / 2),$$

, which provides a relative probability (Burnham and Anderson, 2004).

Finally, Model averaging is then calculated by

$$\text{mod_avg} = p \times \text{weightedAIC}.$$

The model 4 is slightly better supported than model 2 (Table S4).

However, it is still not clear which of these two is supported. For this reason we performed a F-test comparison of the two models (2 and 4) based on the sum of squares of the fastsimcoal output. Results show a better support for model 4 ($p < 0.001$, $F = 7.84$).

Table S1. Filtering steps that were carried out to generate the final high-quality SNP dataset starting from the raw SNPs output by the GATK pipeline. The number of retained SNPs after each step is reported. GQ: Genotype quality; DP: Genotype depth of coverage; IGR: Individual genotyping rate; maxDP: Depth of coverage (twice the mean depth of coverage of the raw dataset); MAF: Minimum Allele Frequency

Filtering	<i>H. tubulosa</i>	<i>H.poli</i>
Raw SNPs	14,221,157	10,744,004
Biallelic SNPs	14,102,846	10,659,309
GQ > 5	7,089	7,748
DP > 5	7,089	7,748
IGR > 80%	4,268	4,777
maxDP	4,268	4,777
MAF > 0.05	4,253	4,758

Table S2. Genetic indices of the *H.tubulosa* and *H.poli* geographical populations as calculated in Sambar (ref). Numbers in parantheses indicate number of individuals.

	Population	Private alleles	Polymorphic sites	H _{obs} +/- s.d.	H _{exp} +/- s.d.	F _{is} +/- s.d.
<i>H.poli</i>	Toronaïos (10)	492	3,960	0.484 +/- 0.004	0.298 +/- 0.002	- 0.046 +/- 0.018
	Sporades (17)	1,107	3,954	0.497 +/- 0.004	0.345 +/- 0.002	- 0.010 +/- 0.026
	Dodekanisa (17)	378	2,553	0.459 +/- 0.004	0.305 +/- 0.002	- 0.001 +/- 0.022
	Thermaikos (6)	168	1,387	0.418 +/- 0.005	0.238 +/- 0.003	- 0.009 +/- 0.010
	Cyclades (13)	301	3,633	0.433 +/- 0.005	0.274 +/- 0.002	- 0.007 +/- 0.018
	Ionio (4)	260	3,690	0.433 +/- 0.005	0.254 +/- 0.002	- 0.003 +/- 0.009
<i>H. tubulosa</i>	Thermaikos (7)	185	3,974	0.332 +/- 0.005	0.190 +/- 0.002	- 0.008 +/- 0.01
	Dodekanisa (13)	643	4,022	0.385 +/- 0.003	0.272 +/- 0.002	0.042 +/- 0.017
	Cyclades (14)	450	3,701	0.383 +/- 0.003	0.258 +/- 0.002	0.026 +/- 0.015
	Kavala (11)	469	3,117	0.388 +/- 0.003	0.262 +/- 0.002	0.030 +/- 0.015
	Pagasetikos (8)	72	3,412	0.319 +/- 0.005	0.177 +/- 0.002	0.001 +/- 0.001
	Toronaïos (7)	143	4,005	0.327 +/- 0.004	0.187 +/- 0.002	- 0.007 +/- 0.009
	Ionio (4)	151	2,401	0.311 +/- 0.004	0.181 +/- 0.002	0.014 +/- 0.006
	Kalamata (9)	812	2,366	0.392 +/- 0.003	0.258 +/- 0.002	0.012 +/- 0.011
	Sporades (17)	1,048	3,728	0.417 +/- 0.003	0.303 +/- 0.002	0.041 +/- 0.021

Table S3. Structure results of the likelihood value of the different K values and ΔK as implementing in Evanno method.

	K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	Ln''(K)	Delta K
H.poli	1	3	-158579.8	1.513	NA	NA	NA
	2	3	-157757.767	8.316	822.033	721.967	86.812
	3	3	-157657.7	181.721	100.067	1293	7.115
	4	3	-158850.633	1138.781	-1192.933	1480.2	1.230
	5	3	-158563.367	1034.095	287.267	72034.233	69.659
	6	3	-230310.333	39250.554	-71746.967	NA	NA
H.tubulosa	1	3	-195194.3	4.557	NA	NA	NA
	2	3	-194225.133	19.769	969.167	1892.833	95.749
	3	3	-195148.8	9.482	-923.667	908.1	95.770
	4	3	-195164.367	362.441	-15.567	905.733	2.499
	5	3	-196085.667	454.646	-921.3	292.467	0.6433
	6	3	-196714.5	1244.874	-628.833	NA	NA

Table S4. The Akaike information criterium (AIC) of each of the five tested demographic scenarios. Maximum Likelihood values are also reported for each model. Also, the weighted AIC values are reported for the close 3 scenarios regarding their original AIC values.

Scenarios	MaxEstLhood	MaxObsLhood	AIC	ΔAIC	Relative likelihood	wAIC	Model probabilities	Model averaging
Ancient gene-flow	7.98×10^{-6}	-4772.87	164.11	76.06	3.046×10^{-17}	1.149×10^{-17}		
Ancient/Recent gene-flow	-4696.94	-460	88.17	0.12	0.9417	0.3554	0.9471	0.3415
Constant gene-flow	-4697.51	-4608.76	88.74	0.69	0.7082	0.2673	0.708	0.2123
Recent gene-flow	-4696.82	-4608.76	88.05	0	1	0.3774	1	0.3774
No gene-flow	-22379.39	-4608.76	17770.63	17682.58	0	0		

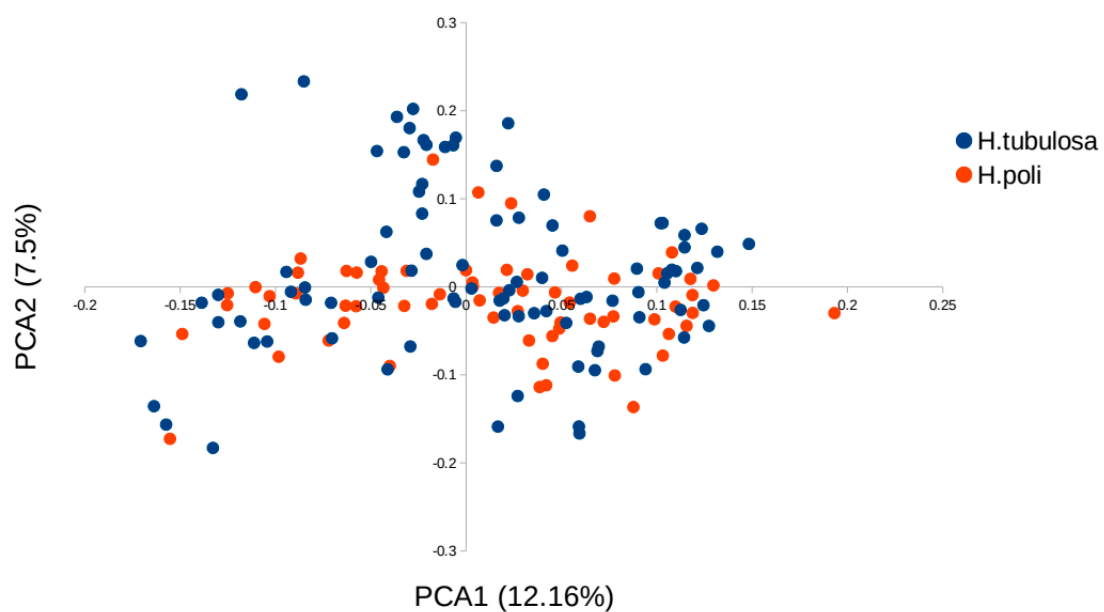


Figure S1. PCA analysis based on both species individuals. Blue dots: *H.tubulosa*; Red dots: *H.poli*

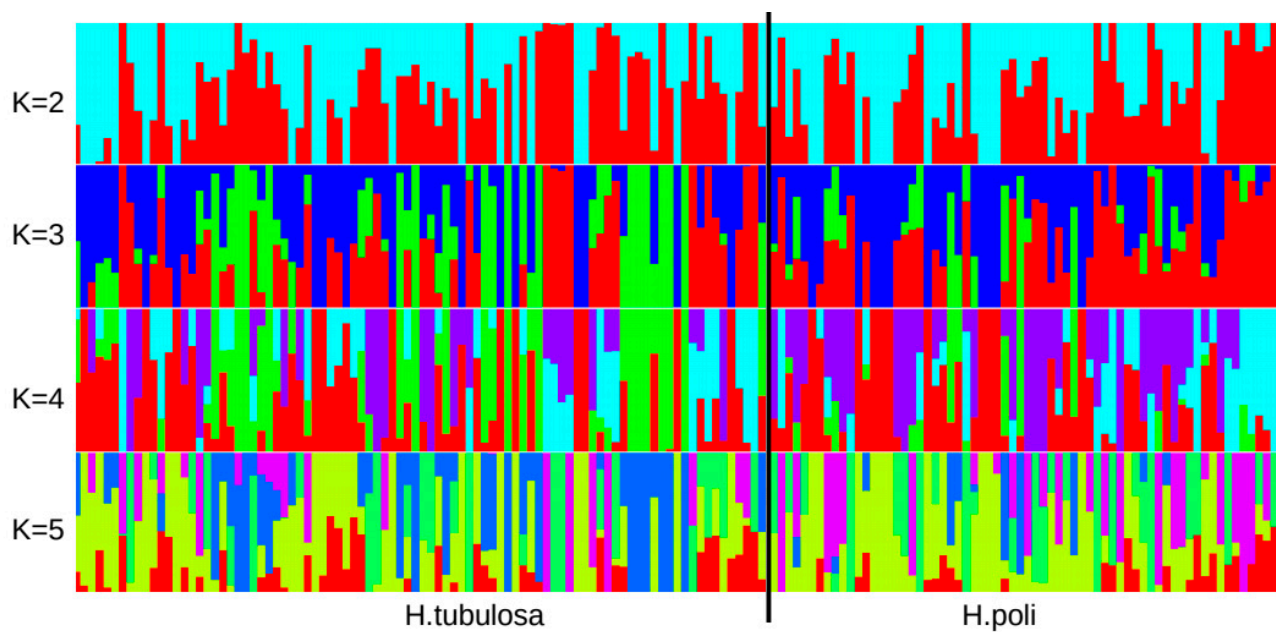


Figure S2. Admixture histogram of the detected clusters and assignment of the individual-level ancestry proportions from each cluster.

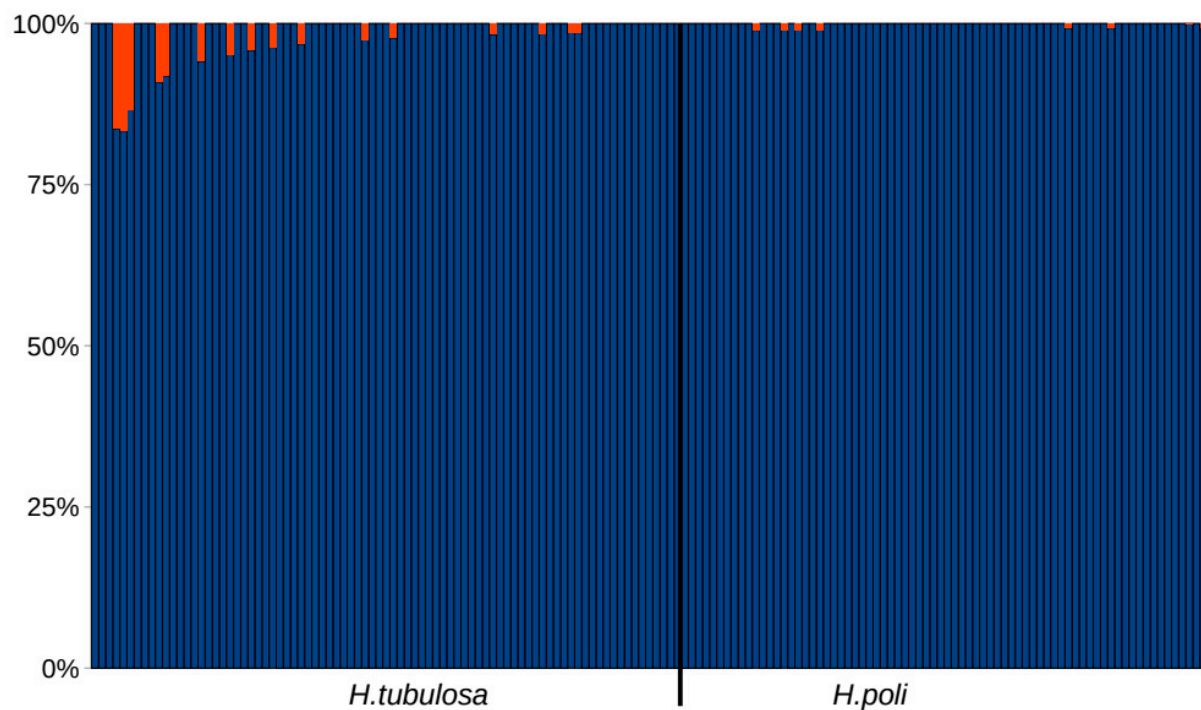


Figure S3. NEWHYBRIDS histogram of the posterior probabilities of the *H. tubulosa* and *H. poli* individuals. Among the 157 individuals analysis assigned them to single category (blue color).

References

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2), 261–304.
<https://doi.org/10.1177/0049124104268644>