

Evaluating genetic traceability methods for captive bred marine fish and their applications in fisheries management and wildlife forensics

Jonas Bylemans*, Gregory E. Maes, Eveline Diopere, Alessia Cariani, Helen Senn, Martin I. Taylor, Sarah Helyar, Luca Bargelloni, Alessio Bonaldo, Gary Carvalho, Ilaria Guarniero, Hans Komen, Jann Th. Martinsohn, Einar E. Nielsen, Fausto Tinti, Filip A.M. Volckaert, Rob Ogden

*Corresponding author: Jonas.Bylemans@canberra.edu.au

Aquaculture Environment Interactions 8: 131–145 (2016)

SUPPLEMENTS

The supplementary material includes a complete list of all samples (Supplement 1) and all genetic markers (Supplement 2) used in the analysis. Details about the comparative analyses between the simulated datasets and an overview of the parentage analyses performed to reconstruct parent-offspring relationships within the farmed sole samples are given (Supplements 3 and 4, respectively). Additionally, the results of the Discriminant Analysis of Principal Components (Supplement 5) are provided.

Supplement 1. Sampling details.

Table S1: Summary information on location, position, number of individuals and sampling year for the empirical samples used in the traceability analysis for Atlantic cod, Atlantic and Mediterranean populations of sole. Sample code refers to the abbreviations used for the population samples in the EU FP7 project FISHPOPTRACE. Sample type indicates the method employed to obtain samples; A) scientific cruise or scientific collection in case of aquaculture populations, B) contracted collection by commercial fishermen. (na = information not available)

Species	ICES/FAO region and sampling location	Sample code	Sample type	Latitude	Longitude	Number of individuals	Sampling year
<i>Solea solea</i>	Wild populations						
	27.III.a - Skagerak and Kattegat						
	Belt Sea	STO	A	55.65	10.76	40	2007
	27.IV.b - Central North Sea						
	German Bight	GER	A	54.52	7.89	40	2007
	27.IV.c - Southern North Sea						
	Norfolk	NOR	A	52.92	2.24	28	2008
	Belgian Coast	BEL2009	A	51.22	2.83	24	2009
	Thames Estuary	THA	A	51.47	1.33	40	2007
	27.VII.d - Eastern English Channel						
	Eastern English Channel	ENG	A	50.78	1.48	40	2008
	27.VII.a - Irish Sea						
	Bristol Channel	IS	A	52.21	-5.33	40	2008
	27.VIII.a - Bay of Biscay - North						
	Pertuis Breton	GAS	A	45.92	-1.69	40	2009
	37.1.3 - Sardinia						
	Viareggio, Northern Tyrrhenian Sea	THY	A	43.30	9.54	40	2009
	37.2.1 - North Adriatic						
	Chioggia Lagoon, North Adriatic	ADR1	A	44.73	13.27	40	2009
	37.2.2 - Ionian						

	South Adriatic Albanian Coast	ADR3	B	41.28	19.13	14	2000
	South Adriatic Italian Coast	ADR2	A	42.02	15.40	19	2000
	<i>37.3.1 - Aegean</i>						
	Gulf of Kavala, Northern Greece	GRE	A	40.85	24.49	40	2009
	<i>37.3.2 - Levant</i>						
	Turkish Coast	TUR2009	A	36.75	33.87	27	2009
	Aquaculture populations						
	<i>Solea BV, the Netherlands</i>						
	Broodstock	A _{BS-SOLE}	na	na	na	4	2003-2005
	Offspring	A _{FI-SOLE}	na	na	na	92	2006
	<i>UNIBO DVP HAP, Italy</i>						
	Broodstock	M _{BS-SOLE}	na	na	na	26	2006
	Offspring	M _{FI-SOLE}					
	Batch 1	M _{FI-SOLE} -B1	na	na	na	24	2008
	Batch 2	M _{FI-SOLE} -B2	na	na	na	24	2008
	Batch 3	M _{FI-SOLE} -B3	na	na	na	24	2008
	Batch 4	M _{FI-SOLE} -B4	na	na	na	24	2009
<i>Gadus morhua</i>	Wild populations						
	<i>27.IV.a - Northern North Sea</i>						
	Northern North Sea	MF03	A	58.00	-3.00	39	2003
	<i>27.IV.b - Central North Sea</i>						
	Northeastern North Sea	NO07	B	57.75	5.50	40	2007
	Southern North Sea	SC06	B	54.29	0.02	40	2006
	<i>27.V.a - Icelandic Grounds</i>						
	Iceland south, offshore	IS	A	63.20	-19.30	39	2002
	<i>27.V.b1 - Faroe Plateau</i>						
	Faroe Plateau	FP02	A	62.53	-6.16	40	2002
	<i>27.V.b2 - Faroe Bank</i>						
	Faroe Bank	FB02	A	60.95	-8.49	40	2002
	<i>27.VII.a - Irish Sea</i>						
	Irish Sea	IR06	A	54.62	-5.46	39	2006
	<i>27.VII.d - English Channel</i>						
	English Channel	EK05	A	50.79	0.48	40	2005
	<i>27.VII.f - Bristol Channel</i>						
	Celtic Sea	CS98	B	50.50	-5.16	39	1998
	<i>27.XII.a - Norwegian Sea</i>						
	Lofoten (NEAC)	SK03	A	68.35	12.14	39	2003
	Aquaculture populations						
	<i>Fiskeaaling A/S, Faeroe Islands</i>						
	Broodstock	A _{BS-COD}	B	na	na	92	2009

Supplement 2. Identification codes and NCBI accession numbers of the SNP loci used

Table S2: Overview of the 96 SNP markers used in the analysis of the Atlantic cod and sole samples.

Atlantic cod samples					
SNP ID	Accession number	SNP ID	Accession number	SNP ID	Accession number
Gm349_1196	rs119054515	cgpGmo-S973	rs119056013	cgpGmo-S1926	rs119055995
cgpGmo-S248a	rs119056000	cgpGmo-S693	rs119055530	cgpGmo-S626b	rs119055882
HbBeta1_1	NA	cgpGmo-S510	rs119055475	cgpGmo-S1098	rs119056384
cgpGmo-S2122	rs119055203	cgpGmo-S1708	rs119056382	cgpGmo-S2187	rs119056259
cgpGmo-S1406	rs119054982	cgpGmo-S459	rs119056202	cgpGmo-S209	rs119056237
cgpGmo-S1644	rs119055520	cgpGmo-S426	rs119055254	Gm240_0209	rs119054548
Gm1154_0166	rs119054629	Gm375_0144	rs119054503	cgpGmo-S13b	rs119056051
cgpGmo-S316	rs119055374	cgpGmo-S1001	rs119055621	cgpGmo-S2058	rs119055972
Gm0738_0160	rs119054733	cgpGmo-S917	rs119055827	cgpGmo-S740	rs119056364
cgpGmo-S1046	rs119055980	cgpGmo-S1391	rs119055071	cgpGmo-S1740	rs119055375
cgpGmo-S1205	rs119055597	cgpGmo-S224	rs119055825	cgpGmo-S87	rs119055065
cgpGmo-S252	rs119055474	cgpGmo-S703	rs119056289	cgpGmo-S831	rs119055880

cgpGmo-S936	rs119056133	Gm1156_0573	rs119054627	cgpGmo-S1024	rs119055966
cgpGmo-S251	rs119056385	Gm394_0364	rs119054493	cgpGmo-S1085a	rs119056019
Gm1339_0238	rs119054574	cgpGmo-S474	rs119055080	cgpGmo-S78	rs119055665
cgpGmo-S1112	NA	cgpGmo-S1076a	rs119056360	cgpGmo-S18	rs119056212
cgpGmo-S689	rs119055763	Hsp90	rs267733128	Gh_2_1	NA
cgpGmo-S742b	rs119055042	cgpGmo-S1094	rs119056042	cgpGmo-S1497	rs119055607
cgpGmo-S430a	rs119055124	cgpGmo-S1338	rs119056054	cgpGmo-S471	rs119055806
cgpGmo-S261b	rs119055116	Gm374_0856	rs119054504	cgpGmo-S2182	rs119055658
cgpGmo-S1751	rs119055210	cgpGmo-S594	rs119055234	cgpGmo-S1219b	rs119056340
cgpGmo-S535b	rs119055277	cgpGmo-S944	rs119055207	cgpGmo-S1051	rs119054998
cgpGmo-S814b	rs119055258	cgpGmo-S968	rs119055670	cgpGmo-S965	rs119056203
cgpGmo-S466	rs119055814	cgpGmo-S1978	rs119056150	cgpGmo-S241	rs119055894
cgpGmo-S875b	rs119055511	cgpGmo-S1418	rs119055063	cgpGmo-S1664	rs119055669
cgpGmo-S879	rs119055741	cgpGmo-S312	rs119055737	cgpGmo-S544	rs119055570
cgpGmo-S624	rs119055557	cgpGmo-S1104	rs119055218	cgpGmo-S127	rs119055331
cgpGmo-S408	rs119055633	cgpGmo-S760	rs119056399	cgpGmo-S2093	rs119056129
cgpGmo-S1743	rs119055928	Gm1002_0428	rs119054666	cgpGmo-S1423a	rs119055003
cgpGmo-S350	rs119056090	cgpGmo-S967b	rs119055979	cgpGmo-S1085b	rs119055638
cgpGmo-S905	rs119055666	cgpGmo-S1362	rs119056422	cgpGmo-S603	rs119055501
Rhod_1_1	NA	cgpGmo-S515	rs119055332	cgpGmo-S2229	rs119055872
Atlantic sole samples					
SNP ID	Accession number	SNP ID	Accession number	SNP ID	Accession number
SNP1012	ss1026565503	SNP1355	ss1026565675	SNP520	ss1026565844
SNP1018	ss1026565506	SNP1388	ss1026565687	SNP570	ss1026565857
SNP1030	ss1026565516	SNP1400	ss1026565690	SNP600	ss1026565867
SNP1033	ss1026565518	SNP1413	ss1026565697	SNP642	ss1026565879
SNP1038	ss1026565521	SNP147	ss1026565713	SNP652	ss1026565883
SNP1068	ss1026565537	SNP1472	ss1026565715	SNP725	ss1026565899
SNP1070	ss1026565539	SNP1478	ss503772168	SNP726	ss1026565900
SNP1091	ss1026565547	SNP1489	ss1026565719	SNP73	ss1026565901
SNP1106	ss1026565552	SNP1496	ss1026565723	SNP776	ss1026565918
SNP1114	ss503772271	SNP1512	ss503772216	SNP779	ss1026565919
SNP1125	ss1026565561	SNP1519	ss1026565731	SNP788	ss1026565925
SNP1127	ss1026565562	SNP1531	ss1026565736	SNP809	ss1026565936
SNP1129	ss503772195	SNP1536	ss1026565737	SNP821	ss1026565945
SNP1137	ss1026565567	SNP1546	ss1026565739	SNP831	ss1026565948
SNP1159	ss1026565577	SNP184	ss1026565753	SNP844	ss1026565952
SNP1160	ss1026565578	SNP199	ss1026565758	SNP845	ss1026565953
SNP1169	ss1026565584	SNP220	ss1026565764	SNP850	ss1026565955
SNP1184	ss1026565590	SNP228	ss503772147	SNP855	ss1026565959
SNP1190	ss1026565593	SNP235	ss503772240	SNP864	ss1026565966
SNP1191	ss1026565594	SNP276	ss1026565777	SNP877	ss1026565969
SNP1200	ss1026565600	SNP284	ss503772263	SNP88	ss1026565972
SNP1213	ss503772234	SNP35	ss1026565787	SNP898	ss1026565979
SNP1262	ss1026565626	SNP376	ss1026565794	SNP899	ss1026565980
SNP1269	ss1026565629	SNP383	ss1026565796	SNP915	ss503772245
SNP1293	ss1026565639	SNP386	ss1026565797	SNP920	ss1026565986
SNP1294	ss1026565640	SNP398	ss1026565805	SNP923	ss503772160
SNP1320	ss1026565658	SNP399	ss1026565806	SNP932	ss503772200
SNP1331	ss1026565661	SNP418	ss1026565809	SNP935	ss1026565993
SNP1337	ss1026565665	SNP455	ss1026565823	SNP948	ss1026565998
SNP134	ss1026565666	SNP464	ss1026565827	SNP963	ss1026566005
SNP1343	ss1026565668	SNP488	ss1026565835	SNP977	ss1026566014
SNP1346	ss1026565669	SNP499	ss503772166	SNP992	ss1026566020
Mediterranean sole samples					
SNP ID	Accession number	SNP ID	Accession number	SNP ID	Accession number
SNP1003	ss503772179	SNP1319	ss1026565656	SNP609	ss1026565870
SNP1010	ss1026565501	SNP1359	ss503772231	SNP633	ss1026565875
SNP1022	ss1026565509	SNP1376	ss1026565680	SNP638	ss1026565877
SNP1024	ss1026565511	SNP1383	ss1026565686	SNP640	ss1026565878
SNP1029	ss1026565515	SNP1388	ss1026565687	SNP645	ss1026565880

SNP1031	ss1026565517	SNP1404	ss1026565691	SNP652	ss1026565883
SNP1033	ss1026565518	SNP1415	ss503772192	SNP7	ss503772184
SNP1046	ss1026565526	SNP1432	ss1026565702	SNP726	ss1026565900
SNP1052	ss503772171	SNP1436	ss1026565703	SNP747	ss1026565905
SNP106	ss1026565532	SNP1439	ss1026565705	SNP749	ss1026565907
SNP1060	ss1026565533	SNP1491	ss1026565720	SNP750	ss1026565908
SNP1070	ss1026565539	SNP1492	ss1026565721	SNP767	ss1026565912
SNP1074	ss1026565540	SNP1496	ss1026565723	SNP776	ss1026565918
SNP1091	ss1026565547	SNP1512	ss503772216	SNP780	ss503772187
SNP1114	ss503772271	SNP1519	ss1026565731	SNP788	ss1026565925
SNP1129	ss503772195	SNP158	ss1026565742	SNP800	ss503772266
SNP1137	ss1026565567	SNP201	ss1026565760	SNP806	ss1026565934
SNP1160	ss1026565578	SNP228	ss503772147	SNP844	ss1026565952
SNP1182	ss1026565588	SNP232	ss1026565767	SNP848	ss1026565954
SNP1184	ss1026565590	SNP235	ss503772240	SNP850	ss1026565955
SNP1190	ss1026565593	SNP246	ss1026565768	SNP879	ss1026565971
SNP1203	ss503772190	SNP275	ss1026565776	SNP88	ss1026565972
SNP1214	ss503772211	SNP350	ss503772258	SNP891	ss1026565978
SNP1236	ss503772209	SNP357	ss1026565789	SNP90	ss1026565981
SNP1240	ss1026565615	SNP394	ss1026565802	SNP914	ss503772228
SNP1250	ss503772203	SNP418	ss1026565809	SNP920	ss1026565986
SNP1251	ss1026565620	SNP422	ss1026565811	SNP925	ss1026565988
SNP1260	ss1026565624	SNP464	ss1026565827	SNP928	ss1026565989
SNP1261	ss1026565625	SNP466	ss1026565829	SNP935	ss1026565993
SNP1284	ss1026565635	SNP486	ss1026565834	SNP953	ss1026566001
SNP1302	ss1026565644	SNP503	ss1026565839	SNP962	ss503772213
SNP1310	ss1026565650	SNP520	ss1026565844	SNP992	ss1026566020

Supplement 3. Comparative analysis between the simulated datasets.

The effect of the simulation software on the overall genetic diversity of the simulated datasets was assessed to evaluate the comparability between the datasets. Since no strong deviation between H_{obs} and H_{exp} were observed in the simulated data, H_{obs} could be used as a proxy for the overall genetic diversity within the datasets. Values of H_{obs} were calculated with Genetix v4.05 for the simulated data generated with HYBRIDLAB and NOOKIE. The data of the two most extreme simulation series ($N_e = 5$ (4 for the PBT) and $N_e = 50$) were used and the H_{obs} was calculated for each series and for all simulated generations (P_{1-SIM} , F_{1-SIM} , F_{2-SIM} , F_{3-SIM} and F_{4-SIM}).

Results (Figure S1) show that both programs performed similarly and yielded similar estimates of H_{obs} for all simulated data sets. Furthermore, as expected, a decline in H_{obs} with an increasing number of captive bred generations (F_{n-SIM}) was observed and this decline was more pronounced at low N_e .

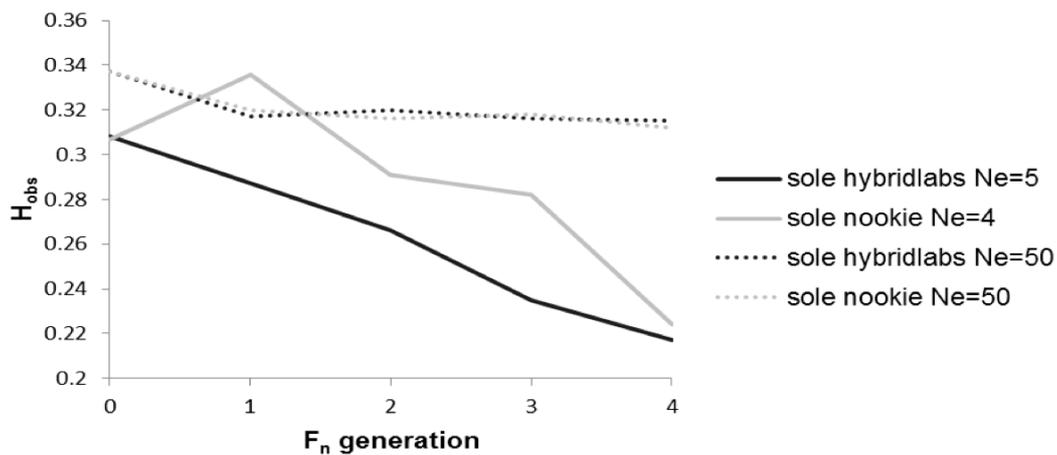


Figure S1: Plot of the H_{obs} in the simulated datasets of sole for the two most extreme effective population sizes ($N_e = 5$ (4) and $N_e = 50$) and simulated with the software NOOKIE v. 1.0 and HYBRIDLAB v.1.0. $F_0 = P_{1-SIM}$ and F_n ($n \neq 0$) = F_{n-SIM} (see text).

Supplement 4. Determining parent-offspring relations in the empirical aquaculture samples.

Parent-offspring relations in the captive bred Atlantic sole

In order to validate the earlier obtained parent-offspring relations, parentage analysis were performed using the SNP genotype data of the Atlantic farmed sole and the software program CERVUS v3.0. Parentage analyses were performed using default parameters. Results of the analysis show that a minimum of 21 highly polymorphic SNPs was sufficient to obtain the same full-sib family structure as in Blonk *et al.* (2009) under strict confidence levels (95%). Furthermore, increasing the number of SNP used in the analysis did not result in another outcome (Table S3). Hence, parent-offspring relations as defined by Blonk *et al.* (2009) do reflect the real mating pattern within the captive bred Atlantic sole samples and thus can be used to evaluate the efficiency of the traceability methods employed in the further analysis.

Table S3: Percentage correctly assigned offspring using Cervus for parentage analysis using the SNP data from the Atlantic farmed sole samples.

Number of SNPs	% correctly assigned offspring under a 95% confidence level
50	100
30	100
21	100
20	98
15	89

Parent-offspring relations in the captive bred Mediterranean sole

Reconstructing the parent-offspring relations within the Mediterranean farmed sole samples was complicated by the absence of SNP genotypes for five broodstock individuals that contributed to the F₁. However, all 24 candidate parents and a subset of F₁'s were genotyped at seven selected microsatellite markers (Table S4). By comparing parentage analysis based on both SNP and microsatellite datasets the most successful parental individuals can be determined and the missing SNP genotypes of highly reproductive parents can be reconstructed.

Table S4: Overview of the seven microsatellite markers for which genotyping data was available for the Mediterranean farmed sole samples.

Marker ID	Reference
F8-ICA9	Iyengar <i>et al.</i> (2000)
F8F8-IGAA7	Iyengar <i>et al.</i> (2000)
F8-ITG11	Iyengar <i>et al.</i> (2000)
F8-IIGT15	Iyengar <i>et al.</i> (2000)
F13-II8/4/7	Iyengar <i>et al.</i> (2000)
Sos(AC)6	Garoia <i>et al.</i> (2006)
Sos(AC)45	Garoia <i>et al.</i> (2006)

Analysis based on the microsatellite data

Complete parent-offspring information was obtained with an initial parentage analysis using the microsatellite data and the software package CERVUS. From these results two parental individuals could be identified that were not SNP-genotyped but did have a relatively high contribution to the F₁ generation. Firstly, a female individual (mother7) did have a high reproductive success within M_{F1-sole}-B2, M_{F1-sole}-B3 and M_{F1-sole}-B4. Additionally, a male individual (father3) could be identified as a successful spawner in M_{F1-sole}-B2.

Analysis based on the SNP data

The software package COLONY (Jones and Wang, 2010) was used to perform an initial parentage assignment analysis using all available loci (181 good quality SNPs) of the Mediterranean farmed sole samples. COLONY relies on sibship reconstruction to determine parent-offspring relations and potential parental genotypes can be

incorporated into the analysis. COLONY will thus determine parent-offspring relations taking into account the genotypes of potential parents but also taking into account that some parental genotypes might not be incorporated. Hence, it is possible to determine whether or not some missing parental genotypes did contribute to the F₁ generation. In addition, COLONY can also be used to reconstruct the genotypes of these missing parental genotypes. The initial analysis based on the SNP data indicated that one missing female genotype (mother #1) was highly successful within M_{F1-sole}-B2, M_{F1-sole}-B3 and M_{F1-sole}-B4 and one missing male genotype (father *3) had a relatively high reproductive success in M_{F1-sole}-B2. Since these results are highly comparable with the results obtained from the parentage analysis based on the microsatellite data we concluded that mother7 = mother #1 and father3 = father *3. The SNP genotypes of these individuals were subsequently reconstructed with COLONY and added to the M_{BS-sole} genotypes.

Using the complete SNP dataset of the Mediterranean farmed sole (M_{BS-sole} + 2 reconstructed parental genotypes and M_{F1-sole}), new parentage assignment analysis were performed. Before parentage assignment was performed, all SNPs deviating from HW-equilibrium were excluded and a total of 62 SNPs remained. Both COLONY and CERVUS were employed to determine parent-offspring relations based on the 62 SNPs. All relations determined by both software packages under strict (95%) confidence were considered to be the effective parent-offspring trios; these were used in PBT analysis. Relatedness could be reconstructed for 38 individuals of which 34 F₁'s and 4 broodstock individuals (Table S5).

Table S5: Overview of the number of offspring per batch for which both parental genotypes could be identified with sufficient confidence.

Batch ID	Father 3				Father 34			
	B1	B2	B3	B4	B1	B2	B3	B4
Mother 7	0	7	0	0	0	0	13	6
Mother 10	0	8	0	0	0	0	0	0

References

Blonk RJW, Komen J, Kamstra A, Crooijmans RPMA, van Arendonk JAM (2009). Levels of inbreeding in group mating captive broodstock populations of Common sole, (*Solea solea*), inferred from parental relatedness and contribution. *Aquaculture* 289: 26–31.

Supplement 5. Discriminant Analysis of Principal Components

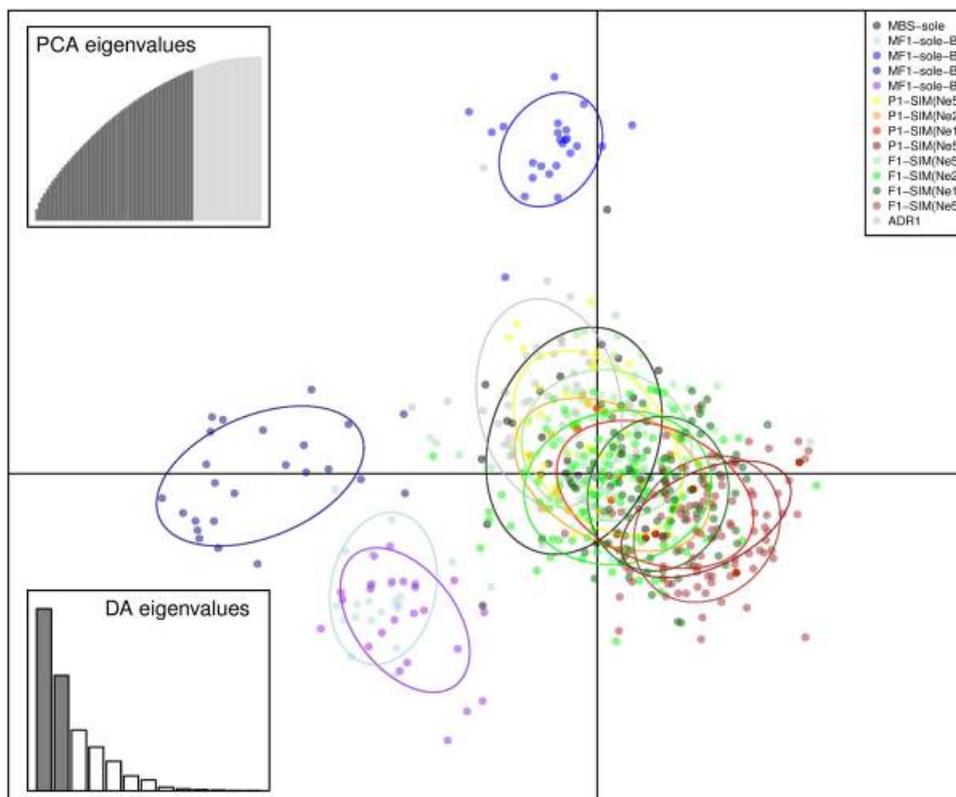


Figure S2: Discriminant Analysis of Principal Components (DAPC) plot for all empirical and simulated Mediterranean aquaculture samples and the wild population ADR1.