Supplementary Material

Population structure of the swordfish, *Xiphias gladius*, across the Indian Ocean using Next Generation Sequencing

Thomas Chevrier (0009-0006-4108-5698)[1,2*], Dominique A. Cowart (0000-0002-2581-0355)[2], Anne-Elise Nieblas (0000-0002-3974-8484)[2], Grégory Charrier (0000-0002-8730-9267)[3], Serge Bernard (0000-0003-1772-0592)[4], Hugues Evano[1], Blandine Brisset (0009-0003-3549-1698)[1], Jérémie Chanut (0009-0002-8808-9689)[2], Sylvain Bonhommeau (0000-0002-0882-5918)[1]

[1] IFREMER Délégation Océan Indien, Rue Jean Bertho, 97822 Le Port, La Réunion, France

[2]Company for Open Ocean Observations and Logging (COOOL), Saint-Leu, La Réunion, France

[3]Univ. Brest, CNRS, IRD, Ifremer, LEMAR, Rue Dumont d'Urville, 29280 Plouzané, France

[4]LIRMM-CNRS, rue Ada, 34000 Montpellier, France

[*]Corresponding author :

t.chevrier.cooolresearch@gmail.com

Tables

Figures

Supplemental protocol :

DArTseq™ represents a combination of DArT complexity reduction methods and next generation sequencing platforms (Sansaloni et al, 2011; Kilian et al, 2012; Courtois et al, 2013; Raman et al. 2014; Cruz et al. 2013). Therefore, DArTseq™ represents a new implementation of sequencing of complexity reduced representations (Altshuler et al, 2000) and more recent applications of this concept on the next generation sequencing platforms (Baird et al, 2008; Elshire et al, 2011). Similarly, to DArT methods based on array hybridisations the technology is optimized for each organism and application by selecting the most appropriate complexity reduction method (both the size of the representation and the fraction of a genome selected for assays). Based on testing several enzyme combinations for complexity reduction the PstI-SphI method was selected for Xiphias. DNA samples were processed in digestion/ligation reactions principally as per Kilian et al (2012) but replacing a single PstI-compatible adaptor with two different adaptors corresponding to two different Restriction Enzyme (RE) overhangs. The PstI-compatible adapter was designed to include Illumina flowcell attachment sequence, sequencing primer sequence and "staggered", varying length barcode region, similar to the sequence reported by Elshire et al, 2011). Reverse adapter contained flowcell attachment region and SphI-compatible overhang sequence.

Only "mixed fragments" (PstI-SphI) were effectively amplified in 30 rounds of PCR using the following reaction conditions:

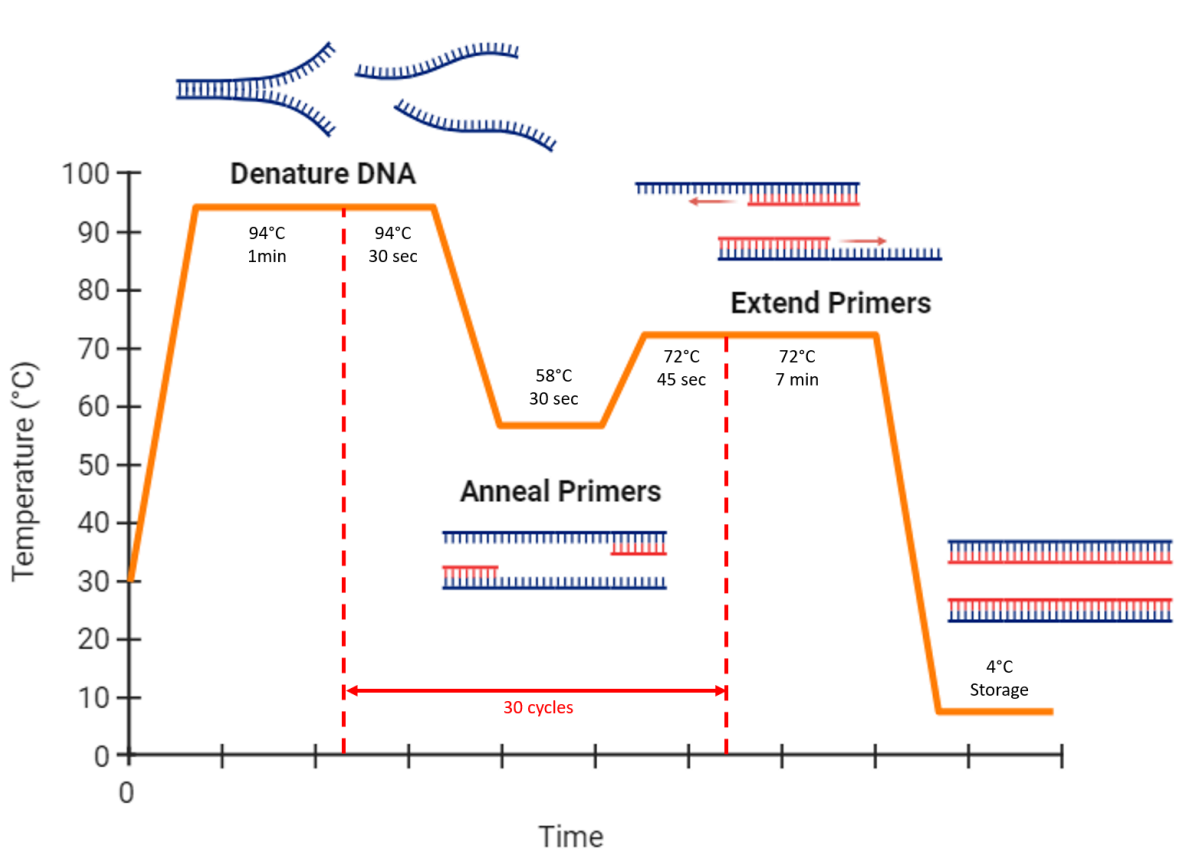


Figure S1: Specification for the temperature cycle for the PCR

After PCR equimolar amounts of amplification products from each sample of the 96-well microtiter plate were bulked and applied to c-Bot (Illumina) bridge PCR followed by

sequencing on Illumina NovaSeq6000. The sequencing (single read) was run for 100 cycles.
Sequences generated from each lane were processed using proprietary DArT analytical pipelines. In the primary pipeline the fastq files were first processed to filter away poor quality sequences, applying more stringent selection criteria to the barcode region compared to the rest of the sequence. In that way the assignments of the sequences to specific samples carried in the "barcode split" step were very reliable.

Filtering was performed on the raw sequences using the parameters described in Table S1.

Table S1: Parameters used to filter the raw sequencing data

Filter	Filter Parameters
Barcode region	Min Phred pass score 30, Min pass percentage 75
Whole read	Min Phred pass score 10, Min pass percentage 50

Approximately 2,500,000 sequences per barcode/sample were identified and used in marker calling. Finally, identical sequences were collapsed into "fastqcoll files". The fastqcoll files were "groomed" using DArT PL's proprietary algorithm which corrects low quality base from singleton tag into a correct base using collapsed tags with multiple members as a template. The "groomed" fastqcoll files were used in the secondary pipeline for DArT PL's proprietary SNP and SilicoDArT (presence/absence of restriction fragments in representation) calling algorithms (DArTsoft14). For SNP calling all tags from all libraries included in theDArTsoft14 analysis are clustered using DArT PL's C++ algorithm at the threshold distance of 3, followed by parsing of the clusters into separate SNP loci using a range of technical parameters, especially the balance of read counts for the allelic pairs. Additional selection criteria were added to the algorithm based on analysis of approximately 1,000 controlled cross populations. Testing for Mendelian distribution of alleles in these populations facilitated selection of technical parameters discriminating well true allelic variants from paralogous sequences. In addition, multiple samples were processed from DNA to allelic calls as technical replicates and scoring consistency was used as the main selection criteria for high quality/low error rate markers.

Calling quality was assured by high average read depth per locus, average across all markers was over 10 reads/locus. Approximately 10% of the samples had technical replicates which were used to estimate reproducibility of reported markers (>99%). The Average Reproducibility was calculated as a fraction of allele calls which are consistent among the technical replicates (libraries) generated from the same DNA samples in a fully independent manner. Reproducibility fraction was calculated for each of the two alleles and averaged for the marker.

References

Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, et al. (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407: 513-516.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, et al. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3: e3376.

Courtois B, Audebert A, Dardou A, Roques S, Ghneim- Herrera T, et al. (2013) Genome-Wide Association Mapping of Root Traits in a Japonica Rice Panel. PLoS ONE 8(11): e78037

Cruz VM, Kilian A, Dierig DA (2013) Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop lesquerella and related

species. PLoS One 8(5): e64062.

Elshire RJ, Glaubitz JC , Sun Q, Poland JA, Kawamoto K, et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One, 6(5):e19379.
Kilian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., Caig, V., et al. (2012). Diversity arrays technology: a generic genome profiling technology on open platforms. Methods in Molecular Biology, 888, 67-89. doi:10.1007/978-1-61779-870-2

Raman H, Raman R, Kilian A, Detering F, Carling J, et al. (2014) Genome-Wide Delineation of Natural Variation for Pod Shatter Resistance in Brassica napus. PLoS ONE 9(7). e101673.

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A (2011) Diversity Arrays Technology (DArT) and next generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. BMC Proceedings 5(Suppl 7):P54

Supplemental table 1 : Sample selection from IOSSS project.

Location	IOSSS code	Year	Number of fish
Australia	SW	2009	16
Glorioso Islands	GL	2006	27
Madagascar	IA	2009 - 2010	122
	IB	2009	69
	IC	2009	142
	ID	2009	58
	XA	2009	248
Mayotte	TD	2009	66
	TF	2010	82
Mozambique Channel	TB	2010	27
Reunion Island	IE	2010	57
	IF	2010	36
	XB	2010	90
	XC	2010	108
Seychelles	SA	2010 - 2011	88
	SB	2009	71
	TE	2011	55
Sri Lanka	FA	2009 - 2010	13
	FB	2010	48
South Africa	BB	2009	46
	BC	2009 - 2010	35
	BE	2009 - 2010	76
Thailand	AA	2010 - 2011	40
	AC	2009 - 2010	74

Supplementary table 2 : DArT metadata.

Information	Meaning
SNP	Mutational change and its position in the sequence tag referenced from zero
SNP Position	Position (zero is position 1) in the sequence tag of the defined SNP variant base
Trimmed Sequence	The sequence containing the SNP or SNPs (the sequence tag) trimmed of adaptors
Call Rate	Proportion of samples for which the genotype call is nonmissing (that is not "-")
OneRatioRef	Proportion of samples for which the genotype is 0
OneRatioSNP	Proportion of samples for which the genotype is 2
FreqHomRef	Proportion of samples homozygous for Reference allele
FreqHomSNP	Proportion of samples homozygous for the Alternate (SNP) allele
FreqHets	Proportion of samples which score as heterozygous that is scored as 1
PICRef	Polymorphism information content (PIC) for the reference allele
PICSnp	Polymorphism information content (PIC) for the SNP
AvgPIC	Average of the polymorphism information content (PIC) of the Reference and SNP alleles
AvgCountRef	Sum of the tag read count for all samples, divided by the number of samples with non-zero tag read count, for the Reference allele row
AvgCountSnp	Sum of the tag read count for all samples, divided by the number of samples with non-zero tag read count, for the Alternate (SNP) allele row
RepAvg	Proportion of technical replicate assay pairs for which the marker score is consistent

Supplementary table 3 : Filtering workflow.

Parameters	Threshold	Loci filtered	Individuals filtered
Total samples submitted : 2068			
Total samples submitted : 2068			2068
Technical replicates	/	/	189
Total sample files remaining	/	/	2227
Samples not sequenced	/	/	30
Non-swordfish samples	/	/	8
Final swordfish samples/files for downstream processing : 2030			
Population structure bioinformatic process (2030 individuals & 86 409 SNPs)			
Sequencing depth (*dartR*)	Between 20 and 145	43 110	/
Linkage disequilibrium (dartR)	> 1	22 640	/
Reproducibility (*dartR*)	0.95	65	/
Call Rate by locus (*dartR*)	0.99	7 634	/
Call Rate by individual (*dartR*)	0.95	/	2
Minor Allele Frequency (MAF) (*dartR*)	0.05	10 480	/
DNA quality and contamination (*kinference*)	/	/	327
Monomorphic loci	/	0	/
Under selected loci (outflank)	0.05	74*	/
Delete IOSSS group with not enough individuals	< 10 fish	/	12
HWE equilibrium *The minimum number of populations where the same locus has to be out of H-W proportions to be removed (1/4)*	alpha = 0.05	37 / 33*	/
Final dataset for population structure : 1 694 fish	**ALL =** 2,443 SNPs (**OUTLIERS** + **NEUT**) **OUTLIERS** = 70 SNPs **NEUT** = 2,373 SNPs (neutral SNPs)*		

* Only apply to the dataset with neutral loci

Supplementary table 4 : Confidence interval at 95% for FST value calculated using 10 000 bootstrap samples for the three dataset. The yellow cells indicate the lower limit and blue cells indicate the upper limit.

CI 95 ALL Dataset

	AA	AC	BB	BC	BE	FA	FB	GL	IA	IB	IC	ID	IE	IF	SA	SW	TB	TD	TF	XA	XB	XC
AA		0.004	0.017	0.017	0.017	0.007	0.010	0.012	0.018	0.016	0.016	0.016	0.018	0.017	0.010	0.027	0.018	0.016	0.016	0.016	0.017	0.017
AC	0.000		0.011	0.011	0.012	0.002	0.006	0.007	0.013	0.012	0.010	0.010	0.013	0.013	0.005	0.021	0.013	0.010	0.011	0.011	0.012	0.012
BB	0.008	0.005		0.000	0.001	0.011	0.003	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.011	0.002	0.000	0.001	0.002	0.002	0.002
BC	0.008	0.006	0.000		0.001	0.011	0.003	0.002	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.010	0.001	0.000	0.000	0.002	0.002	0.002
BE	0.008	0.006	0.000	0.000		0.012	0.003	0.002	0.001	0.001	0.001	0.001	0.001	0.002	0.003	0.010	0.001	0.001	0.001	0.001	0.001	0.002
FA	0.000	0.000	0.005	0.004	0.005		0.007	0.008	0.013	0.012	0.011	0.011	0.013	0.012	0.006	0.021	0.012	0.010	0.011	0.011	0.011	0.012
FB	0.003	0.003	0.000	0.000	0.001	0.002		0.001	0.003	0.002	0.002	0.002	0.003	0.003	0.001	0.011	0.004	0.003	0.002	0.002	0.002	0.002
GL	0.005	0.003	0.000	0.000	0.000	0.002	0.000		0.003	0.002	0.001	0.002	0.003	0.004	0.001	0.012	0.003	0.002	0.002	0.002	0.002	0.003
IA	0.009	0.006	0.000	0.000	0.000	0.005	0.001	0.001		0.000	0.001	0.001	0.001	0.001	0.003	0.009	0.000	0.001	0.001	0.001	0.001	0.001
IB	0.007	0.006	0.000	0.000	0.000	0.005	0.000	0.000	0.000		0.000	0.000	0.001	0.000	0.002	0.009	0.000	0.001	0.000	0.000	0.001	0.001
IC	0.007	0.005	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000		0.000	0.001	0.001	0.002	0.010	0.001	0.001	0.001	0.000	0.001	0.001
ID	0.007	0.005	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000		0.001	0.002	0.002	0.010	0.001	0.001	0.001	0.000	0.001	0.001
IE	0.008	0.006	0.000	0.000	0.000	0.006	0.001	0.001	0.000	0.000	0.000	0.000		0.002	0.003	0.010	0.001	0.002	0.002	0.001	0.001	0.001
IF	0.007	0.006	0.000	0.000	0.000	0.005	0.000	0.001	0.000	0.000	0.000	0.000	0.000		0.003	0.010	0.002	0.002	0.002	0.001	0.001	0.001
SA	0.004	0.003	0.001	0.001	0.001	0.003	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.001		0.011	0.003	0.002	0.002	0.002	0.002	0.003
SW	0.014	0.011	0.004	0.003	0.004	0.010	0.004	0.004	0.003	0.003	0.004	0.004	0.004	0.004	0.005		0.010	0.010	0.009	0.009	0.010	0.010
TB	0.008	0.006	0.000	0.000	0.000	0.005	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003		0.001	0.001	0.001	0.001	0.001
TD	0.007	0.005	0.000	0.000	0.000	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.004	0.000		0.001	0.001	0.002	0.002
TF	0.007	0.006	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.000	0.000		0.001	0.001	0.001
XA	0.007	0.005	0.000	0.000	0.000	0.005	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.000	0.000		0.001	0.000
XB	0.008	0.006	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.000	0.000	0.000	0.000		0.001
XC	0.008	0.006	0.000	0.000	0.000	0.005	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.000	0.000	0.000	0.000	0.000	

CI 95 Outliers Dataset

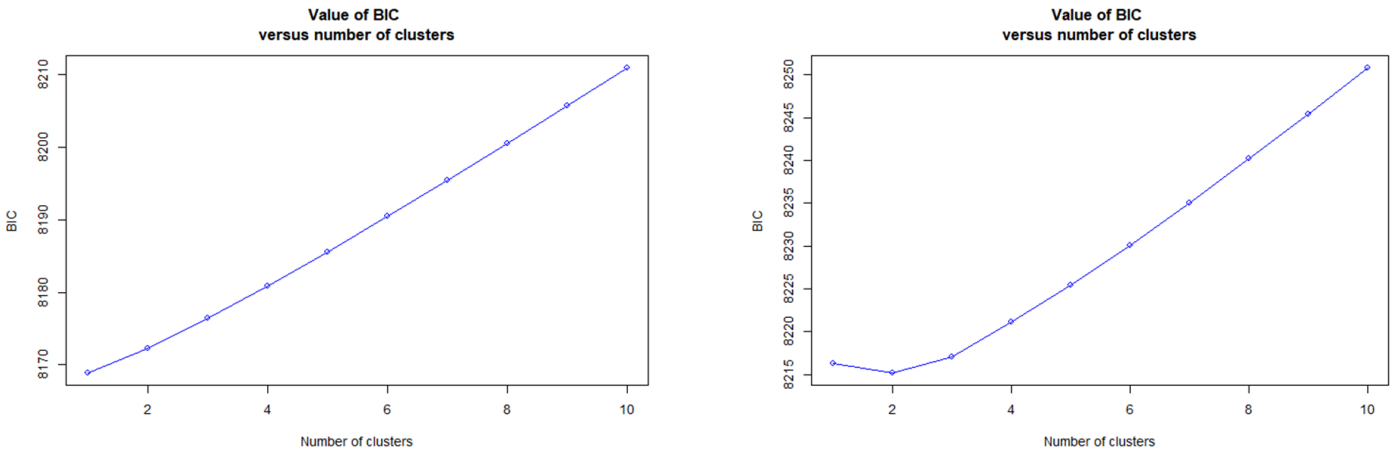
	AA	AC	BB	BC	BE	FA	FB	GL	IA	IB	IC	ID	IE	IF	SA	SW	TB	TD	TF	XA	XB	XC
AA		0.060	0.341	0.311	0.329	0.113	0.225	0.230	0.352	0.329	0.301	0.294	0.346	0.349	0.205	0.455	0.358	0.295	0.307	0.315	0.328	0.339
AC	0.010		0.263	0.252	0.262	0.012	0.141	0.140	0.282	0.266	0.231	0.227	0.279	0.284	0.111	0.391	0.297	0.223	0.239	0.242	0.258	0.265
BB	0.198	0.150		0.003	0.013	0.259	0.083	0.054	0.043	0.033	0.028	0.029	0.045	0.055	0.059	0.248	0.028	0.003	0.013	0.045	0.050	0.061
BC	0.173	0.134	0.000		0.007	0.237	0.088	0.048	0.031	0.027	0.027	0.024	0.044	0.049	0.059	0.211	0.015	0.000	0.001	0.043	0.046	0.062
BE	0.187	0.142	0.002	0.000		0.259	0.091	0.050	0.017	0.016	0.020	0.015	0.025	0.035	0.072	0.214	0.007	0.009	0.003	0.032	0.038	0.048
FA	0.018	0.000	0.145	0.126	0.143		0.138	0.141	0.283	0.257	0.226	0.218	0.275	0.281	0.113	0.398	0.287	0.214	0.235	0.243	0.252	0.264
FB	0.115	0.069	0.036	0.027	0.031	0.069		0.028	0.088	0.067	0.042	0.051	0.072	0.075	0.016	0.250	0.116	0.068	0.069	0.045	0.050	0.063
GL	0.123	0.075	0.019	0.015	0.017	0.073	0.006		0.068	0.052	0.030	0.033	0.063	0.064	0.012	0.251	0.070	0.032	0.037	0.043	0.051	0.055
IA	0.203	0.161	0.017	0.011	0.004	0.166	0.041	0.030		0.012	0.015	0.011	0.013	0.011	0.092	0.204	0.008	0.039	0.024	0.016	0.019	0.023
IB	0.180	0.148	0.011	0.008	0.002	0.145	0.031	0.023	0.001		0.002	0.003	0.002	0.002	0.075	0.208	0.015	0.029	0.017	0.007	0.009	0.012
IC	0.163	0.127	0.009	0.006	0.003	0.127	0.021	0.011	0.006	0.000		0.001	0.006	0.007	0.048	0.214	0.027	0.023	0.016	0.003	0.007	0.010
ID	0.162	0.126	0.012	0.006	0.002	0.125	0.022	0.010	0.002	0.000	-0.002		0.007	0.009	0.053	0.208	0.020	0.021	0.015	0.005	0.009	0.012
IE	0.192	0.159	0.016	0.015	0.005	0.158	0.036	0.028	0.000	0.000	0.002	0.000		0.005	0.087	0.217	0.027	0.042	0.027	0.006	0.006	0.006
IF	0.188	0.154	0.019	0.014	0.008	0.157	0.035	0.026	0.001	0.000	-0.001	0.000	-0.004		0.089	0.216	0.027	0.050	0.032	0.005	0.009	0.004
SA	0.106	0.064	0.030	0.024	0.032	0.061	0.002	0.003	0.048	0.036	0.022	0.025	0.043	0.042		0.242	0.096	0.042	0.053	0.062	0.069	0.084
SW	0.264	0.209	0.100	0.078	0.075	0.236	0.087	0.097	0.069	0.075	0.076	0.077	0.078	0.073	0.098		0.224	0.227	0.208	0.209	0.207	0.220
TB	0.206	0.164	0.007	0.000	0.000	0.165	0.045	0.024	0.000	0.000	0.005	0.003	0.002	0.001	0.044	0.073		0.028	0.012	0.035	0.037	0.047
TD	0.163	0.123	0.000	0.000	0.001	0.118	0.024	0.010	0.016	0.009	0.006	0.007	0.018	0.016	0.019	0.087	0.006		0.006	0.040	0.046	0.059
TF	0.167	0.131	0.000	0.000	0.000	0.125	0.025	0.011	0.009	0.004	0.003	0.004	0.008	0.009	0.024	0.071	0.000	0.000		0.027	0.031	0.042
XA	0.169	0.129	0.017	0.012	0.007	0.132	0.021	0.017	0.005	0.000	0.001	0.000	0.000	0.000	0.029	0.075	0.005	0.013	0.007		0.003	0.002
XB	0.170	0.134	0.019	0.013	0.012	0.131	0.021	0.021	0.007	0.001	0.002	0.002	0.001	0.001	0.029	0.068	0.007	0.017	0.008	0.000		0.006
XC	0.181	0.144	0.022	0.019	0.012	0.146	0.028	0.022	0.007	0.002	0.002	0.002	0.000	0.000	0.038	0.082	0.008	0.021	0.012	0.000	0.001	

CI 95 NEUT Dataset

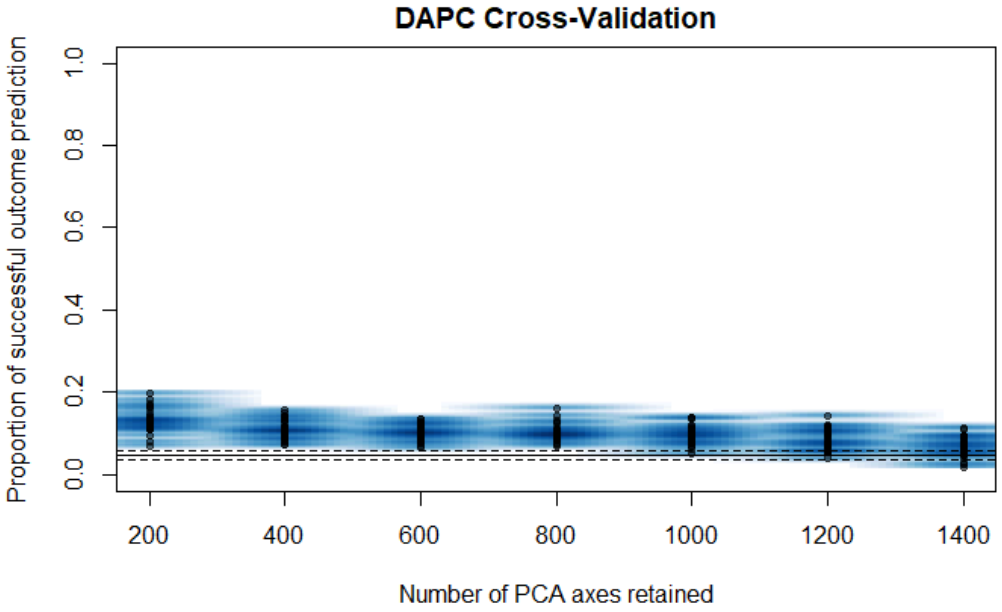
	AA	AC	BB	BC	BE	FA	FB	GL	IA	IB	IC	ID	IE	IF	SA	SW	TB	TD	TF	XA	XB	XC
AA		0.003	0.005	0.006	0.005	0.005	0.004	0.005	0.005	0.004	0.005	0.005	0.005	0.004	0.004	0.011	0.005	0.006	0.005	0.005	0.005	0.005
AC	0.000		0.002	0.003	0.002	0.002	0.002	0.003	0.002	0.002	0.002	0.002	0.003	0.003	0.002	0.007	0.002	0.002	0.002	0.002	0.002	0.002
BB	0.000	0.001		0.000	0.000	0.003	0.000	0.001	0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.004	0.001	0.000	0.001	0.001	0.001	0.000
BC	0.001	0.001	0.000		0.001	0.003	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.003	0.001	0.001	0.000	0.001	0.001	0.000
BE	0.000	0.001	0.000	0.000		0.004	0.001	0.001	0.001	0.000	0.001	0.001	0.001	0.001	0.001	0.003	0.001	0.001	0.001	0.000	0.000	0.001
FA	0.000	0.000	0.000	0.000	0.000		0.003	0.004	0.003	0.003	0.003	0.004	0.004	0.002	0.004	0.007	0.002	0.003	0.003	0.003	0.003	0.003
FB	0.000	0.001	0.000	0.000	0.000	0.000		0.001	0.000	0.000	0.001	0.001	0.001	0.001	0.000	0.004	0.001	0.001	0.001	0.000	0.001	0.001
GL	0.000	0.001	0.000	0.000	0.000	0.000	0.000		0.001	0.001	0.001	0.001	0.002	0.002	0.001	0.004	0.002	0.001	0.001	0.001	0.001	0.001
IA	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.001	0.001	0.000	0.003	0.000	0.000	0.000	0.000	0.000	0.000
IB	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.001	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.001	0.000
IC	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.001	0.001	0.000	0.004	0.001	0.001	0.000	0.000	0.001	0.000
ID	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.001	0.002	0.001	0.004	0.000	0.001	0.000	0.000	0.001	0.001
IE	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.002	0.001	0.003	0.000	0.001	0.001	0.001	0.001	0.001
IF	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001		0.001	0.004	0.001	0.001	0.001	0.001	0.001	0.001
SA	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.004	0.001	0.001	0.000	0.000	0.000	0.000
SW	0.003	0.004	0.000	0.000	0.001	0.002	0.001	0.000	0.000	0.000	0.001	0.001	0.000	0.001	0.001		0.004	0.003	0.003	0.003	0.004	0.003
TB	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.001	0.001	0.000	0.001	0.000
TD	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.001	0.000	0.001	0.001
TF	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000
XA	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000
XB	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000		0.001
XC	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	

Supplementary table 5 : Outliers SNPs and their biological functions

Chromosome	Position	Gene	Biological functions
NC_053402.1_chromosome_3	4727491	rbfox1	The Fox-1 family of RNA-binding proteins is evolutionarily conserved, and regulates tissue-specific alternative splicing in metazoa. Fox-1 recognizes a (U)GCAUG stretch in regulated exons or in flanking introns. The protein binds to the C-terminus of ataxin-2 and may contribute to the restricted pathology of spinocerebellar ataxia type 2 (SCA2). Ataxin-2 is the product of the SCA2 gene which causes familial neurodegenerative diseases.
	5070105	**Intron**	
	6240648	Intron	
NC_053404.1_chromosome_5	3549766	Intron	
	3661571	**Intron**	
	4166928	kcnh2b	Potassium voltage-gated channel
	4380445	**LOC120789773**	
	4575679	Intron	
	4749120	**LOC1200790173**	
	5323226	Intron	
	5557111	**stk17a**	**(Serine/Threonine Kinase 17a): Involved in stress response and cell survival, which could be significant for adapting to environmental stressors in different marine habitats.**
	5751240	tgfbr1b	(Transforming Growth Factor Beta Receptor 1b): Plays a role in developmental processes and immune response, potentially influencing growth or reproductive traits
	8360449	**Intron**	
	8550525	txndc5	Encodes a member of the disulfide isomerase (PDI) family of endoplasmic reticulum (ER) proteins that catalyze protein folding and thiol-disulfide interchange reactions
	8941913	**ofcc1**	
	9269038	LOC120789622	
	9628111	**Intron**	
	9849221	Intron	
	9993033	**zfhx4**	**zinc finger homeobox 4 : Predicted to enable DNA-binding transcription factor activity, RNA polymerase II-specific and RNA polymerase II cis-regulatory region sequence-specific DNA binding activity. Predicted to be involved in regulation of transcription by RNA polymerase II.**
	10340557	tpk1	The protein encoded by this gene functions as a homodimer and catalyzes the conversion of thiamine to thiamine pyrophosphate, a cofactor for some enzymes of the glycolytic and energy production pathways. Defects in this gene are a cause of thiamine metabolism dysfunction syndrome-5.
	10365758	**Intron**	
	10391664	Intron	
	10595728	**cntnap2a**	**Predicted to act upstream of or within protein localization to juxtaparanode region of axon. Predicted to be located in membrane and paranodal junction. Predicted to**
	10694578	**cntnap2a**	**be integral component of membrane. Is expressed in anterior lateral line ganglion and posterior lateral line ganglion.**
	11033047	shrprbck1r	Predicted to enable ubiquitin binding activity and ubiquitin-protein transferase activity. Acts upstream of or within embryonic cranial skeleton morphogenesis and otic vesicle morphogenesis. Predicted to be part of LUBAC complex. Is expressed in inner ear; midbrain hindbrain boundary; otic vesicle; and pharyngeal arch.
	11177750	**si:dkey12|12.1**	**Predicted to be located in membrane. Predicted to be integral component of membrane.**
	11342298	LOC120789951	
	11506115	**sema5a**	**This gene belongs to the semaphorin gene family that encodes membrane proteins containing a semaphorin domain and several thrombospondin type-1 repeats. Members of this family are involved in axonal guidance during neural development. This gene has been implicated as an autism susceptibility gene.**
	11555123	Intron	
	11710541	**Intron**	
	11734943	Intron	
	11904497	**rttn**	**This gene encodes a large protein whose specific function is unknown. Absence of the orthologous protein in mouse results in embryonic lethality with deficient axial**
	11923541	**rttn**	**rotation, abnormal differentiation of the neural tube, and randomized looping of the heart tube during development. In human, mutations in this gene are associated**
	11925938	**rttn**	**with polymicrogyria with seizures. In human fibroblasts this protein localizes at the ciliary basal bodies. Given the intracellular localization of this protein and the phenotypic effects of mutations, this gene is suspected of playing a role in the maintenance of normal ciliary structure which in turn effects the developmental process of left-right organ specification, axial rotation, and perhaps notochord development.**
	12020360	Intron	
	12072479	**jph1a**	**Predicted to be located in endoplasmic reticulum and membrane. Predicted to be integral component of membrane. Predicted to be part of junctional membrane complex. Predicted to be active in endoplasmic reticulum membrane; plasma membrane; and sarcoplasmic reticulum. Human ortholog(s) of this gene implicated in Charcot-Marie-Tooth disease axonal type 2K.**
	12278371	kcnb2b	(Potassium Channel, Two Pore Domain Subfamily K Member 2): This gene regulates neuronal excitability and membrane potential, which might be relevant for sensory perception and navigating varied oceanic conditions.
	12381819	**LOC120790135**	
	12483066	Intron	
	12626026	**LOC120789639**	
	12626144	LOC120789639	
	12668985	**sulf1**	**(Sulfatase 1): Modifies extracellular matrix components and may affect tissue development or regeneration, with implications for adaptation to different ecological niches.**
	12727757	Intron	
	12933292	**arfgef1**	**ADP-ribosylation factors (ARFs) play an important role in intracellular vesicular trafficking. The protein encoded by this gene is involved in the activation of ARFs by accelerating replacement of bound GDP with GTP. It contains a Sec7 domain, which may be responsible for guanine-nucleotide exchange activity and also brefeldin A inhibition.**
	12996699	ppp1r42	(Protein Phosphatase 1 Regulatory Subunit 42): Plays a role in cellular signaling, potentially affecting growth and metabolism.
	13304700	**spata13**	**Enables guanyl-nucleotide exchange factor activity and identical protein binding activity. Involved in cell migration; plasma membrane bounded cell projection assembly; and regulation of cell migration. Located in several cellular components, including filopodium; lamellipodium; and ruffle membrane.**
	13570958	Intron	
	14018455	**vill**	**The protein encoded by this gene belongs to the villin/gelsolin family. It contains 6 gelsolin-like repeats and a headpiece domain. It may play a role in actin-bundling.**
	14111485	oxsr1b	(Oxidative-Stress Responsive Kinase 1b): Involved in responding to oxidative stress, which could be critical for fish living in environments with varying levels of oxygen or
	14122397	oxsr1b	other stressors.
	14449505	**Intron**	
	15089375	dlgap1b	Predicted to enable molecular adaptor activity. Predicted to be involved in regulation of postsynaptic neurotransmitter receptor activity. Predicted to act upstream of or within chemical synaptic transmission. Predicted to be located in synapse. Predicted to be active in glutamatergic synapse and postsynaptic specialization.
	15169078	**Intron**	
NC_053405.1_chromosome_6	14935393	LOC120790323	
	15064461	**ttll77**	**Enables alpha-tubulin binding activity; beta-tubulin binding activity; and tubulin-glutamic acid ligase activity. Involved in protein polyglutamylation.**
	15151513	LOC120790868	
NC_053408.1_chromosome_9	**2639398**	**LOC120794859**	
NC_053410.1_chromosome_11	13150849	cuedc2	Predicted to enable ubiquitin binding activity. Acts upstream of or within negative regulation of cytokine production involved in inflammatory response and negative regulation of macrophage cytokine production. Located in cytosol; nuclear membrane; and nucleoplasm.
NC_053415.1_chromosome_16	**27616376**	**LOC120801590**	
NC_053417.1_chromosome_18	14029826	zgc:195245	
NC_053418.1_chromosome_19	**19814568**	**adgra2**	**Involved in sensory perception and immune response, which could relate to behavioral traits or responses to environmental stimuli.**



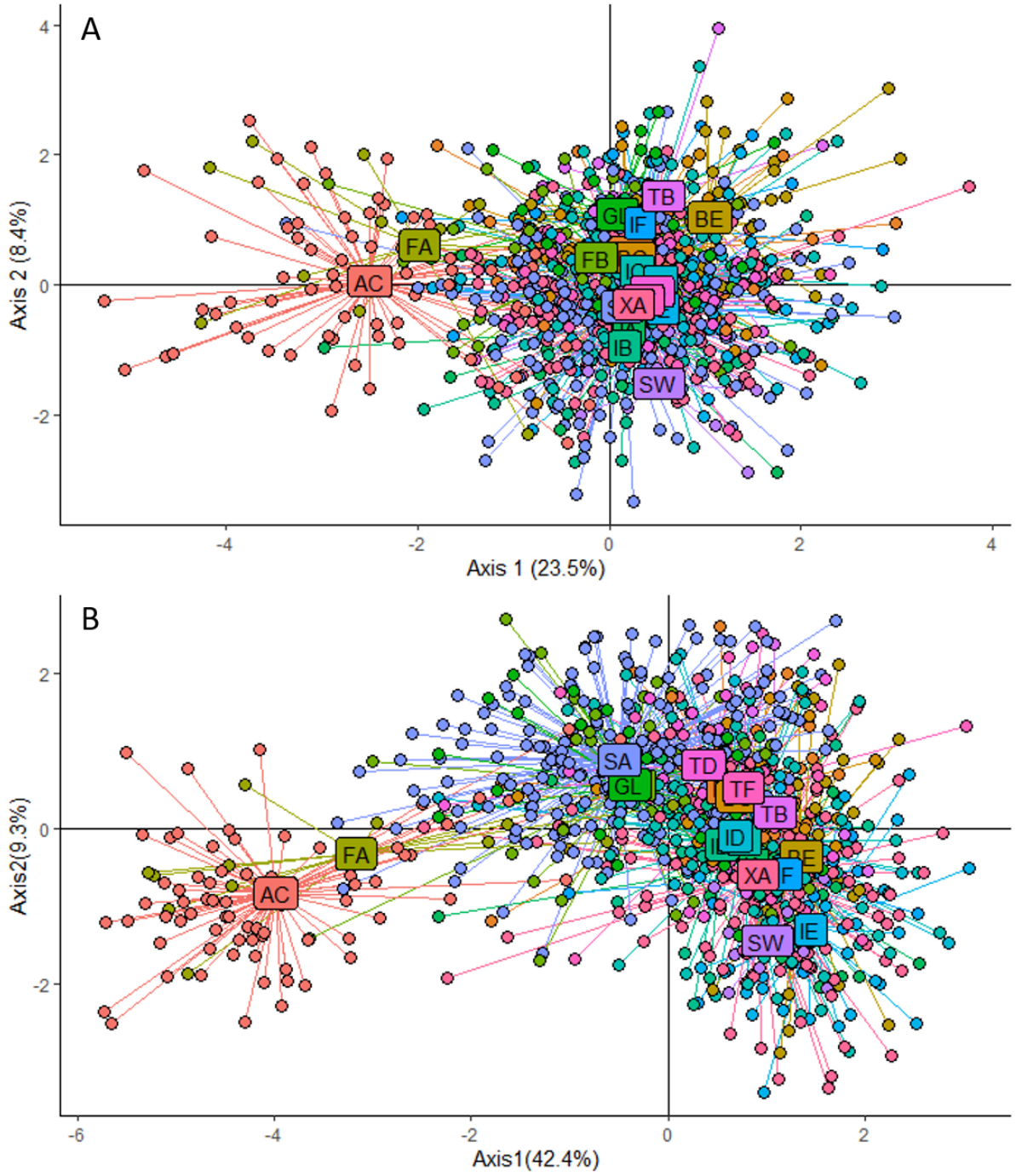
Supplemental Figure 1: Goodness of fit (Bayesian Information Criteria) for the different numbers of clusters assessed according to the *K*-means clustering method using the *adegenet* package for (A) the dataset with only neutral loci and (B) the dataset with loci under potential selection.



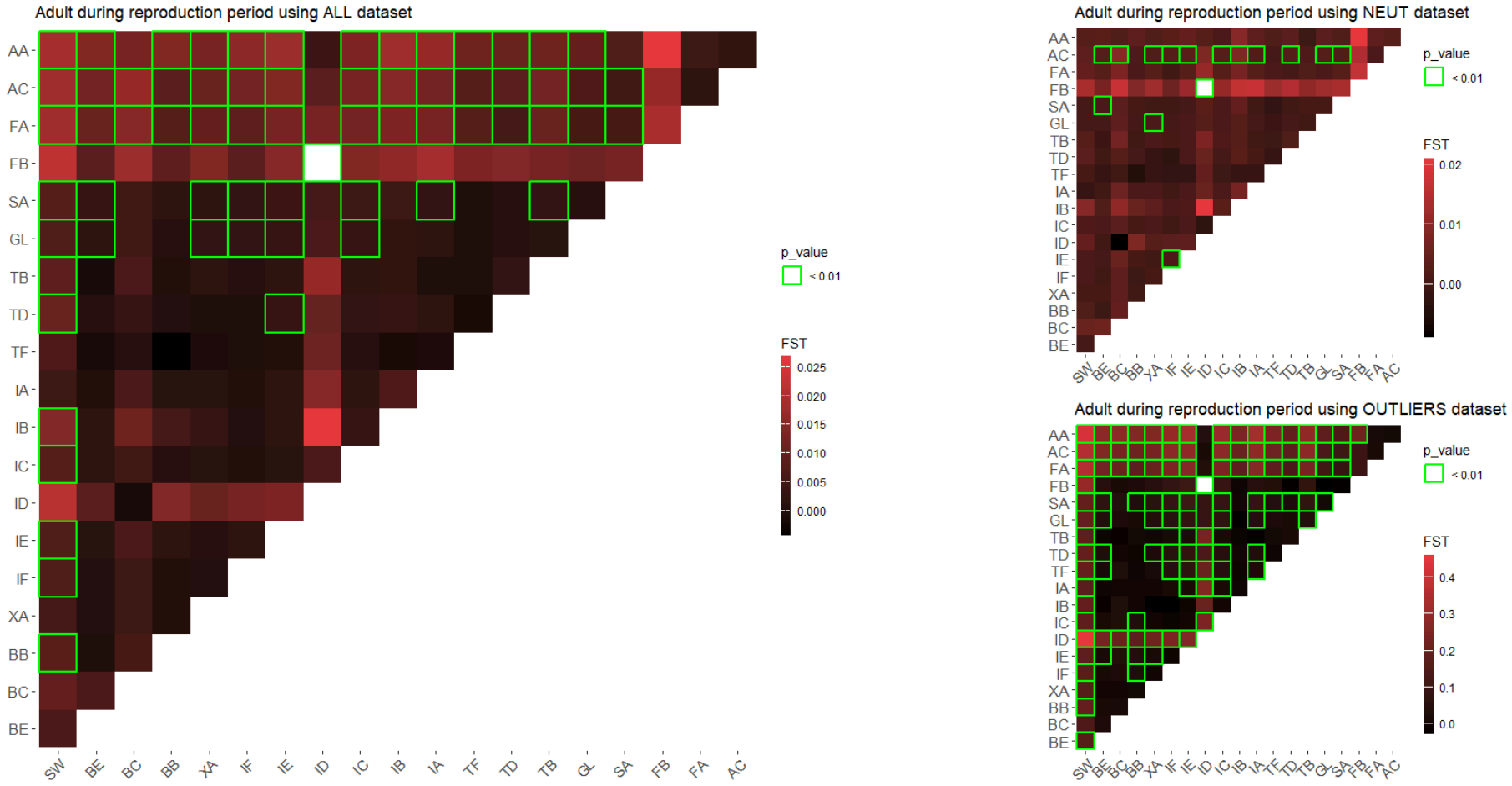
Supplemental Figure 2: DAPC cross-validation plot with the number of PCs retained in each DAPC.

To carry out a DAPC the number of retained PCs must be determined. The number of PCs can have a substantial impact on the results of the analysis. Cross-validation (carried out with the function *xvalDapc*) provides an objective optimization procedure for identifying the 'goldilocks point' in the trade-off between retaining too few and too many PCs in the model.

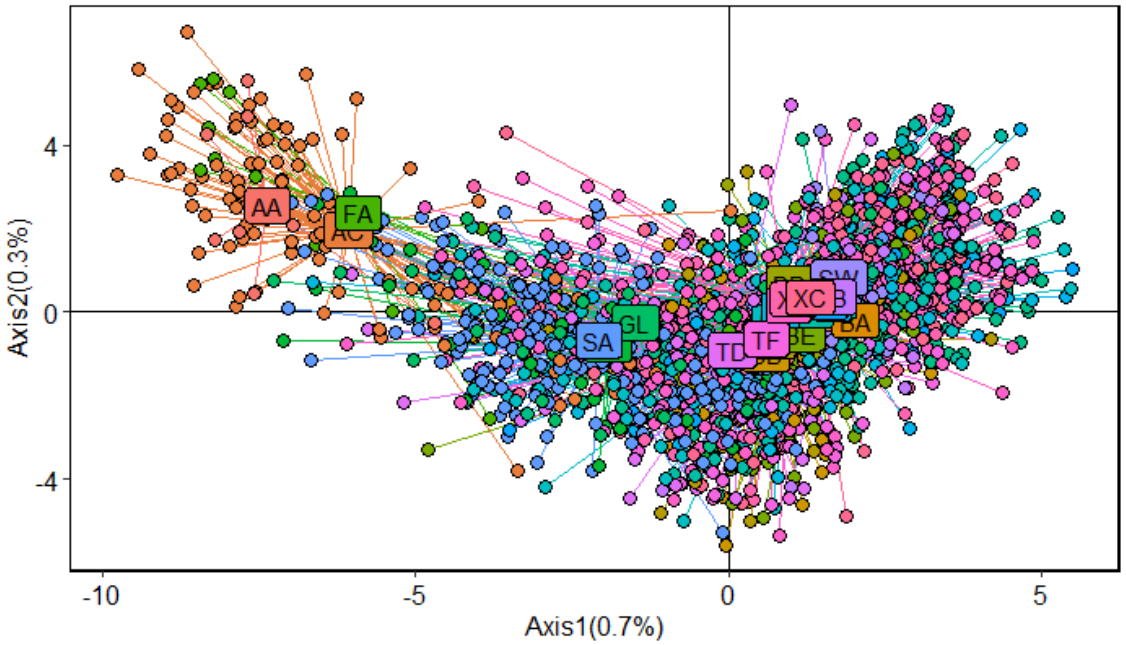
In cross-validation, the data are divided into two sets: a training set (typically comprising 90% of the data) and a validation set (which contains the remainder (by default,10%) of the data). With *xvalDapc*, the validation set is selected by stratified random sampling. This ensures that at least one member of each group or population in the original data is represented in both training and validation sets. DAPC is carried out on the training set with variable numbers of PCs retained, and the degree to which the analysis is able to accurately predict the group membership of excluded individuals (those in the validation set) is used to identify the optimal number of PCs to retain. At each level of PC retention, the sampling and DAPC procedures are repeated n.reptimes. When xval.plot is TRUE, a scatter plot of the DAPC cross-validation is generated. The number of PCs retained in each DAPC varies along the x-axis, and the proportion of successful outcome prediction varies along the y-axis. Individual replicates appear as points, and the density of those points in different regions of the plot is displayed in blue. Based on the model validation literature, it is recommended to use the number of PCs associated with the lowest root mean squared error (RMSE) as the 'optimum' n.pca in the DAPC analysis. In our case, the optimum n.pca was obtained at 200 PCA axes.



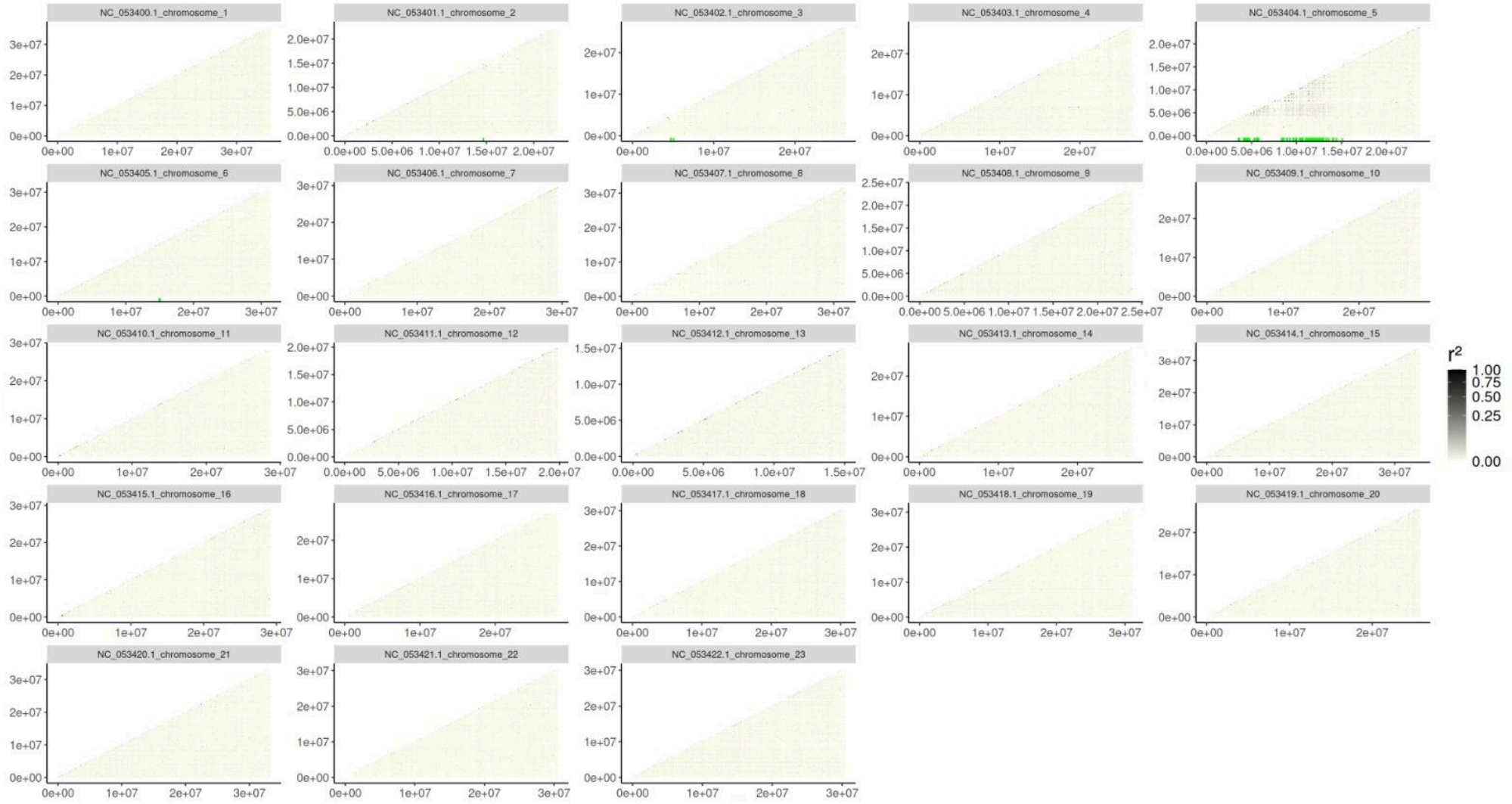
Supplemental Figure 3: DAPC for samples from swordfish (*X. gladius*) adults during reproductive period (November - April) for (A) Neutral SNPs and (B) Neutral + SNPs with a selection signature.



Supplemental Figure 4 : Heatmap representations of the pairwise fixation index (FST) between the different sampling areas for adult swordfish (*X. gladius*) sampled during their reproduction period (November - April) for (Left) ALL dataset (Top right) NEUT dataset and (Bottom right) OUTLIERS dataset.



Supplemental Figure 5 : Principal component analysis (PCA) according to the first two axes, with samples grouped by their sampling areas for dataset with neutral and outliers loci



Supplemental Figure 6 : Linkage Disequilibrium (LD) across each chromosome calculating r² between each pair of SNPs. The LD plot was not made for the 24th chromosome due to a lack of data.