

Linking Allele-Specific Expression And Natural Selection In Wild Populations

Romuald Laso-Jadart^{1,6*}, Kevin Sugier¹, Emmanuelle Petit², Karine Labadie², Pierre Peterlongo³, Christophe Ambroise⁴, Patrick Wincker^{1,6}, Jean-Louis Jamet⁵, Mohammed-Amin Madoui^{1,6*}

¹Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France.

²CEA, Genoscope, Institut de Biologie François Jacob, Université Paris-Saclay, Evry, 91057, France.

³Univ Rennes, CNRS, Inria, IRISA - UMR 6074, F-35000 Rennes.

⁴LaMME, CNRS, Univ Evry, Université Paris-Saclay, Evry, France

⁵Université de Toulon, Aix-Marseille Université, CNRS/INSU/IRD, Mediterranean Institute of Oceanology MIO UMR 110, CS 60584, 83041 Toulon cedex 9, France.

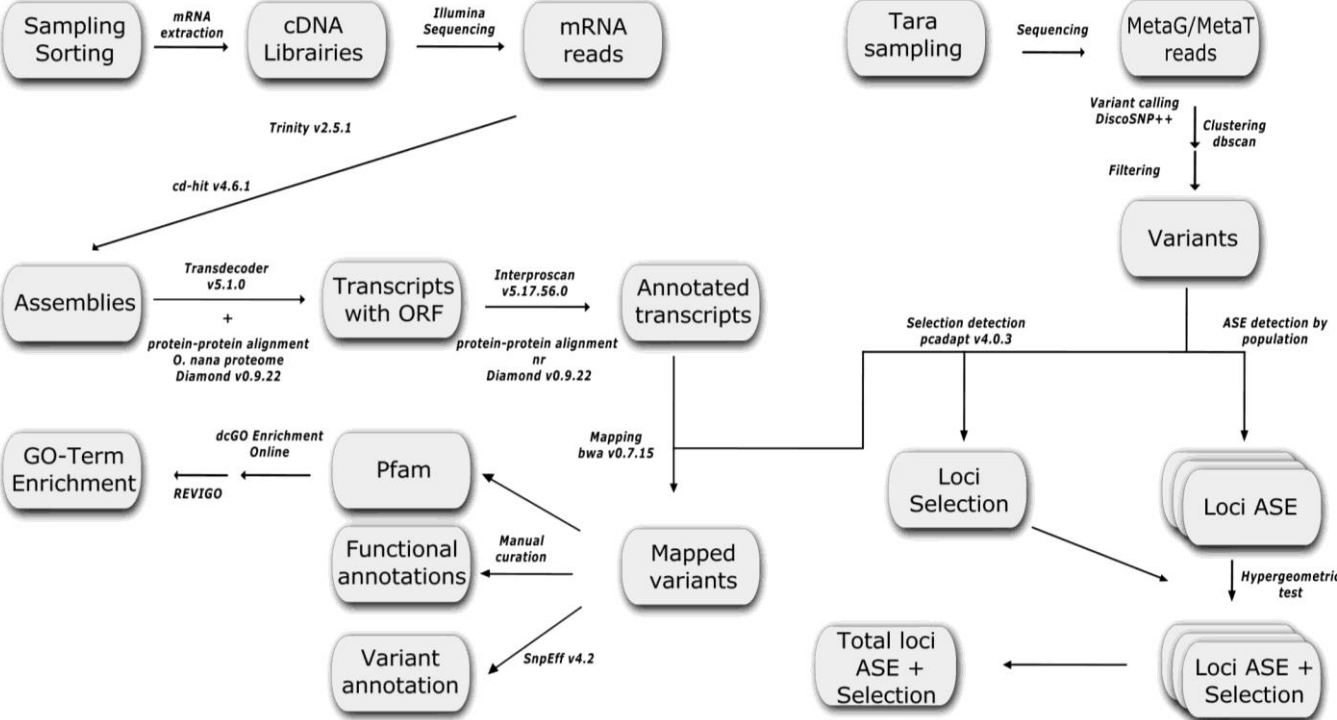
⁶Research Federation for the study of Global Ocean Systems Ecology and Evolution, FR2022/Tara Oceans GO-SEE, 3 rue Michel-Ange, 75016 Paris, France

*Corresponding authors. Emails: rlasojad@genoscope.cns.fr & amadoui@genoscope.cns.fr

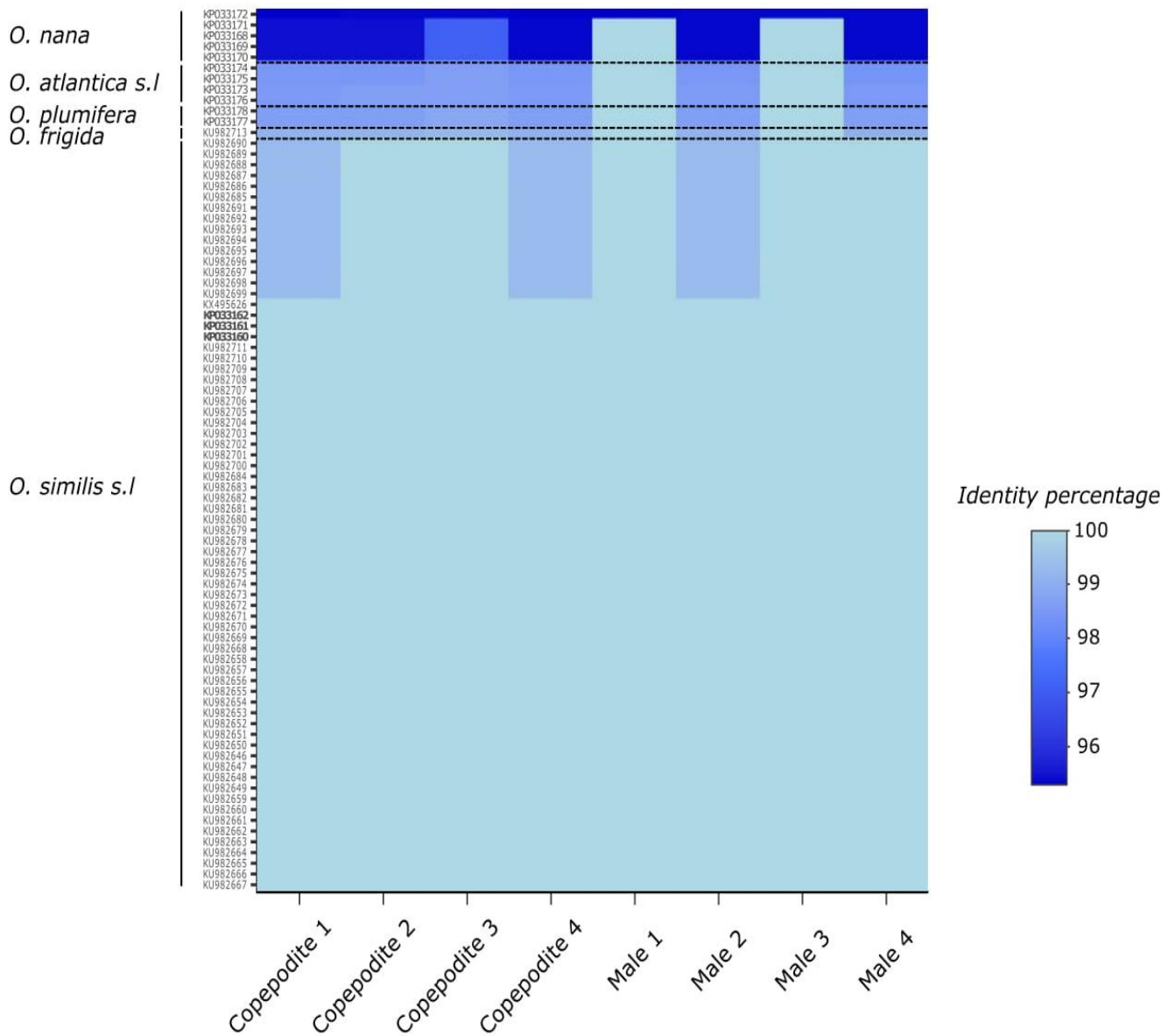
Supplementary figures

Supplementary Figure 1: Method pipeline overview	3
Supplementary Figure 2: Validation of taxonomic assignation.....	4
Supplementary Figure 3: <i>Oithona similis</i> depth of coverage of biallelic loci in seven <i>Tara</i> Oceans samples.	5
Supplementary Figure 4: Metagenomic coverage distribution of the seven <i>Tara</i> Oceans samples.	6
Supplementary Figure 5: Genomic differentiation of Arctic Seas <i>Oithona similis</i> populations.	7
Supplementary Figure 6: Allele-specific expression detection.....	8
Supplementary Figure 7: Functional analysis of <i>Oithona similis</i> transcripts targeted by ASE and selection.....	9
Supplementary Table 1: <i>Oithona similis</i> Mediterranean transcriptomes summary.	10
Supplementary Table 2: <i>Tara</i> Oceans and <i>Oithona similis</i> Mediterranean transcriptomes samples accession numbers.....	11
Supplementary Table 5: Variant annotation by SNPeff.	12
References.....	13

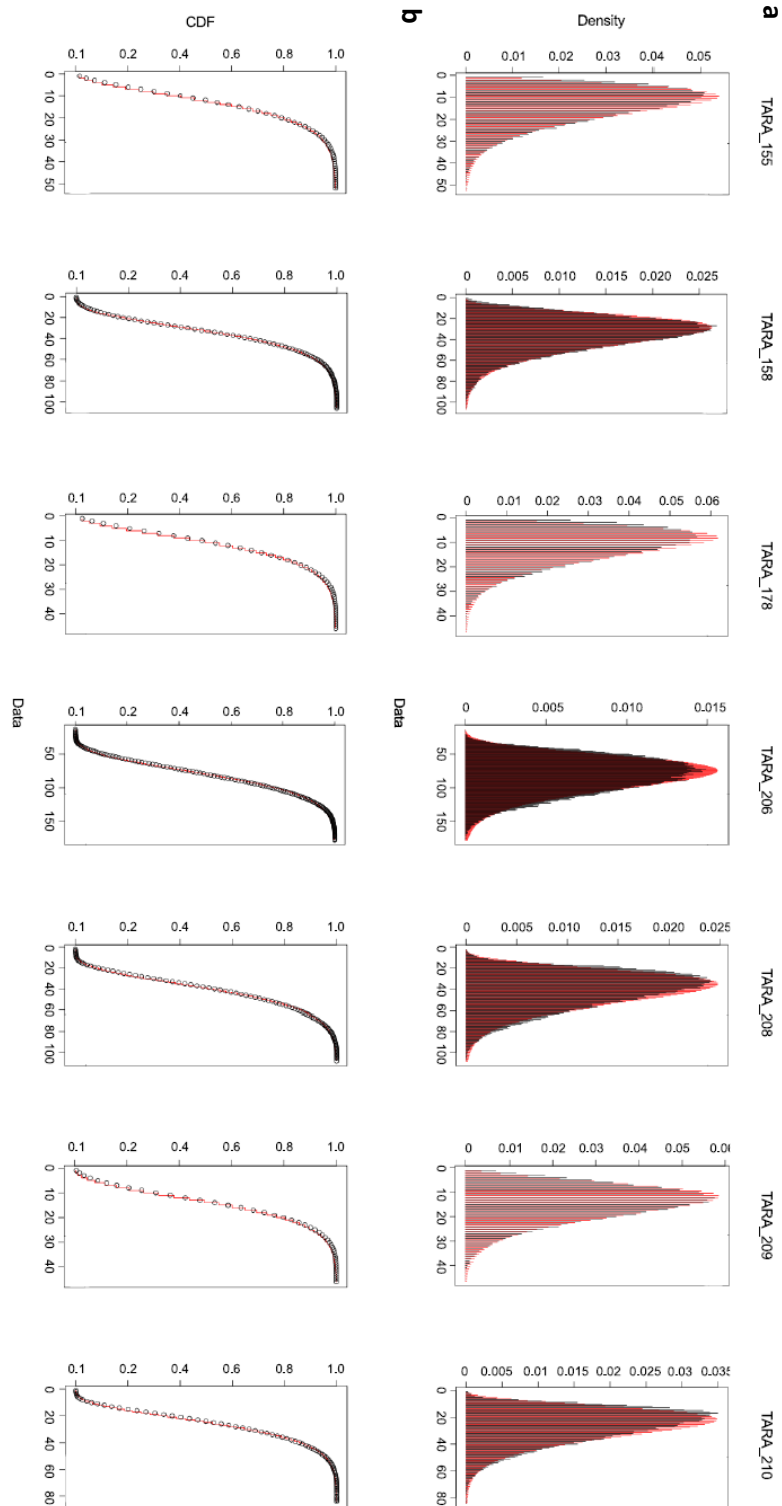
Supplementary Figure 1: Method pipeline overview



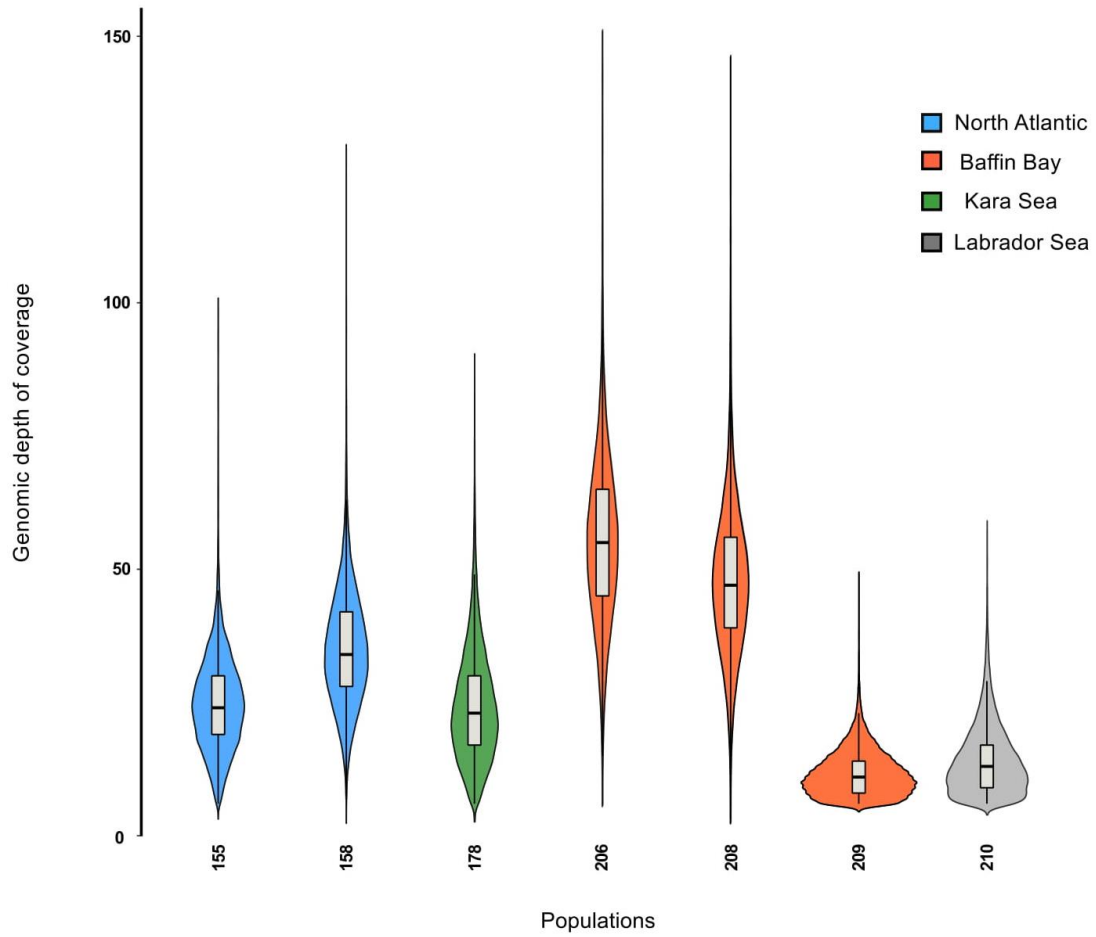
Supplementary Figure 2: Validation of taxonomic assignment. In rows are represented the 82 accession numbers of *Oithona* species 28S sequences. In bold, type localities of *O. similis* as described in Cornils et al., 2017. In columns are represented ribosomal read sets of the eight individuals.



Supplementary Figure 3: *Oithona similis* depth of coverage of biallelic loci in seven *Tara* Oceans samples. In red is represented the theoretical depth of coverage. In black is represented the observed depth of coverage. **a**, Comparison of coverage distribution. **b**, Comparison of the cumulative distribution functions.

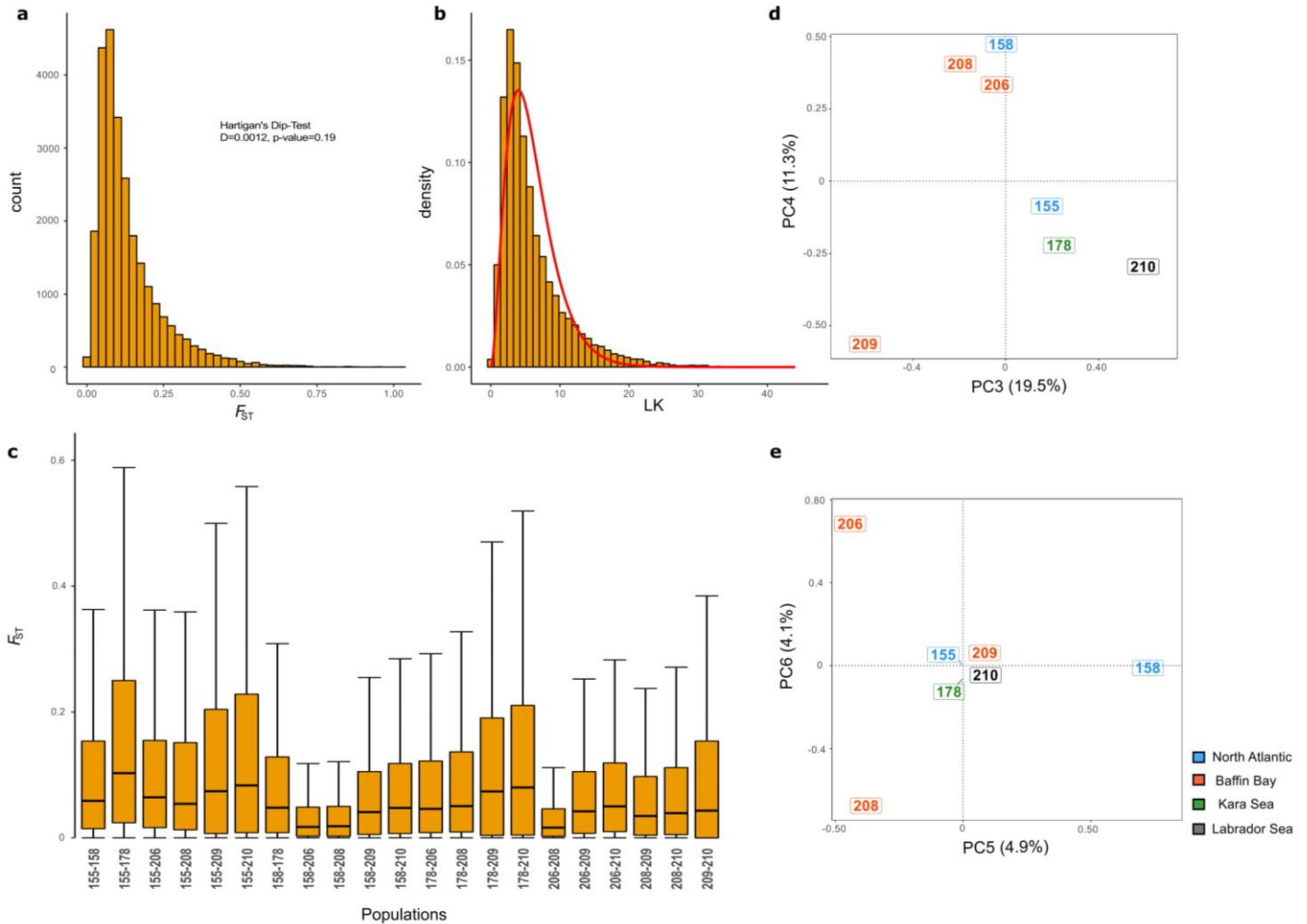


Supplementary Figure 4: Metagenomic coverage distribution of the seven *Tara* Oceans samples. Genomic depth of coverage distributions of the set of 25,768 variants obtained after filtering, by populations.

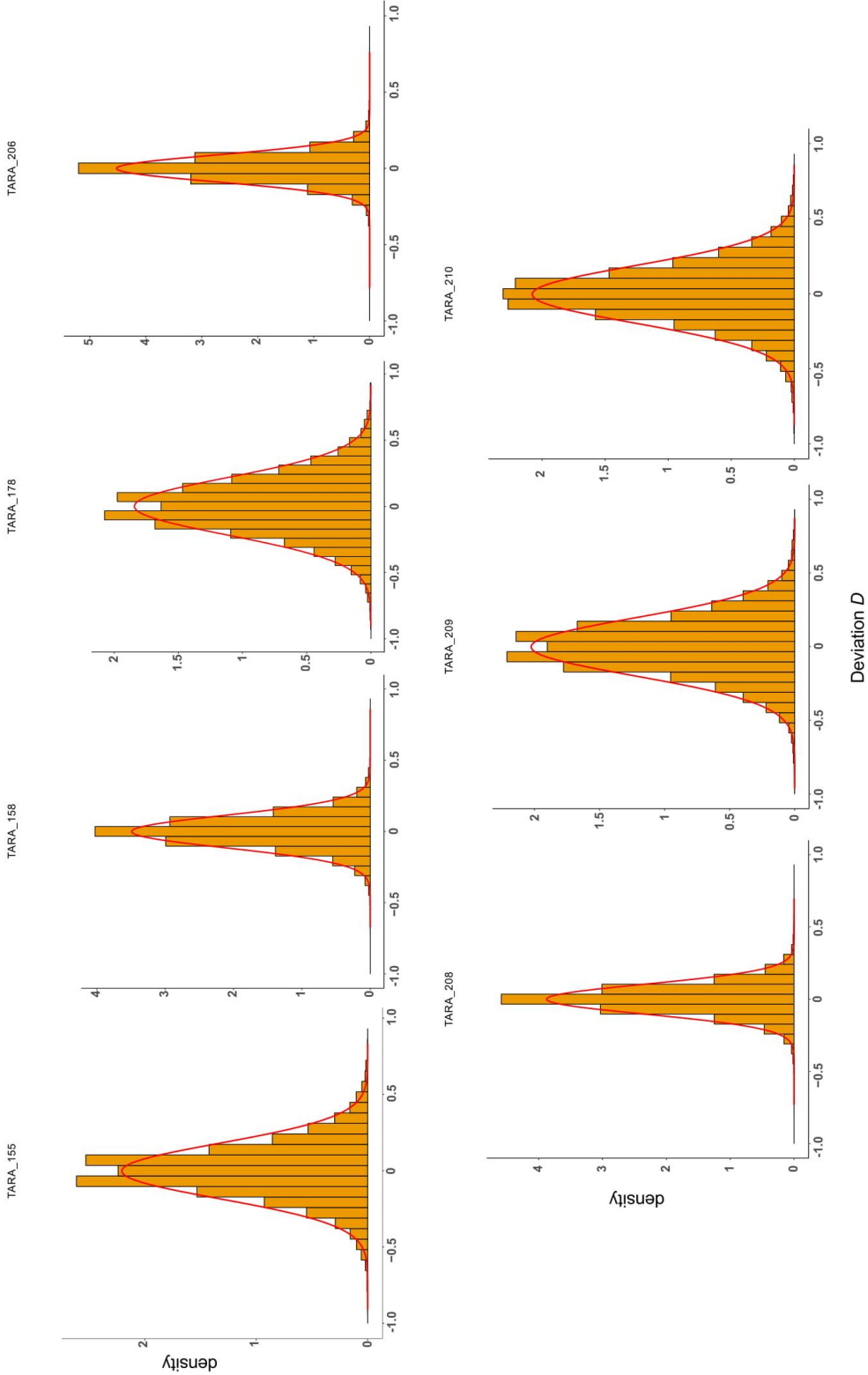


Supplementary Figure 5: Genomic differentiation of Arctic Seas *Oithona similis* populations.

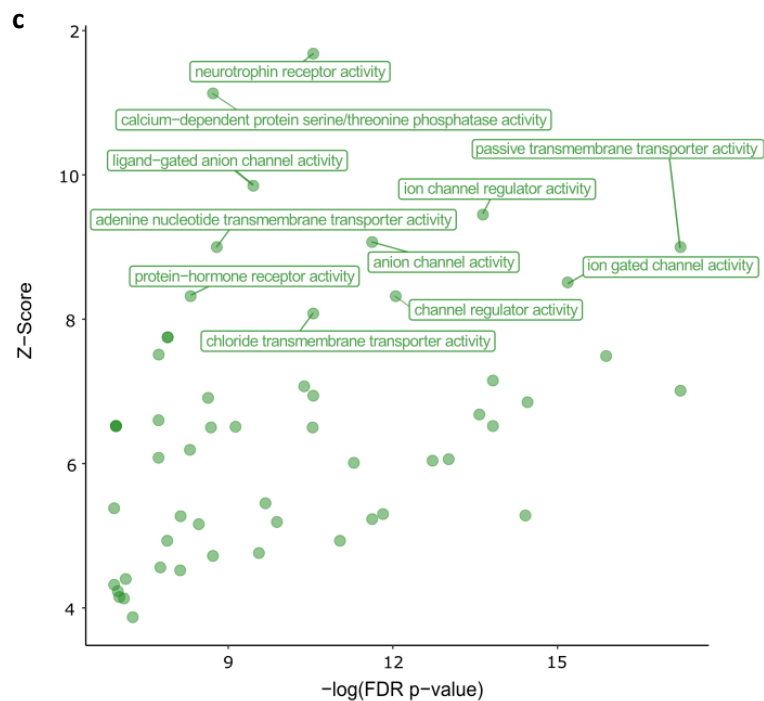
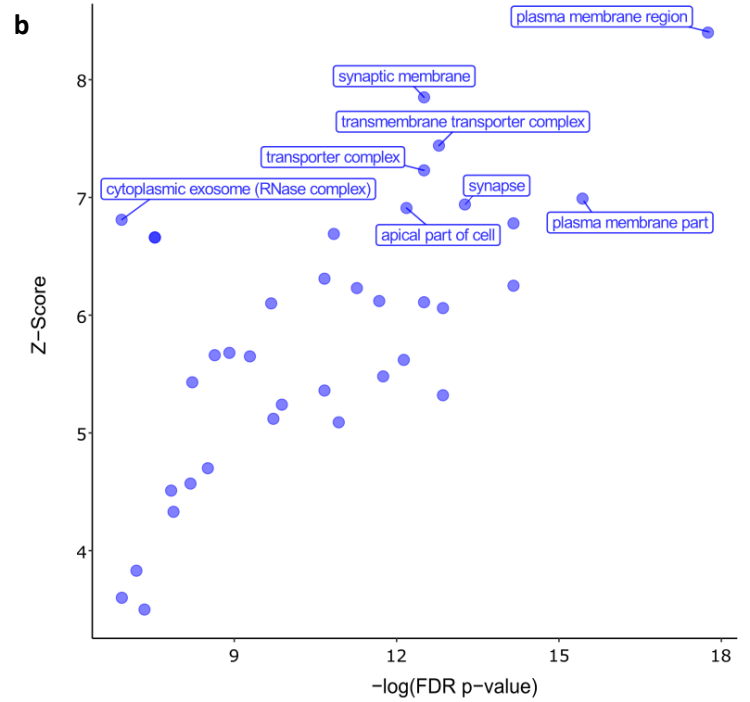
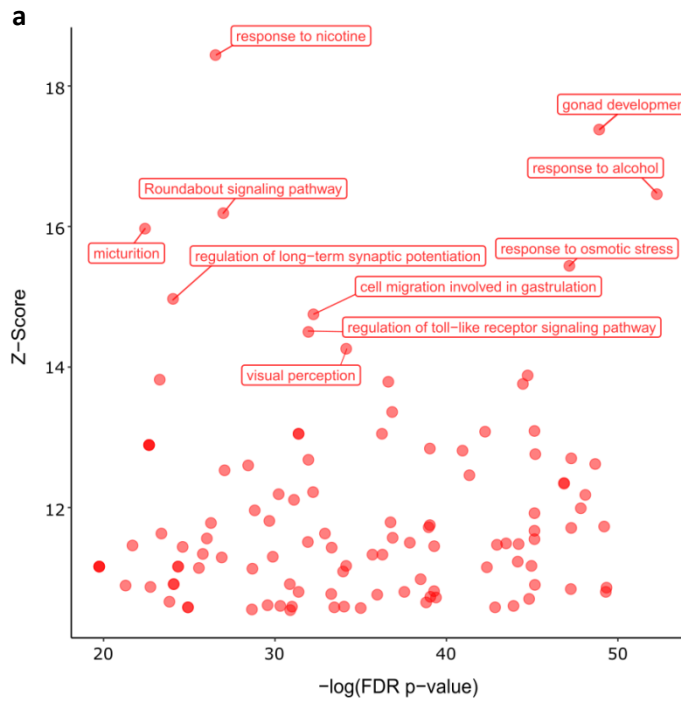
The three computations are based on the B-allele frequency matrix of the seven populations. **a**, Global F_{ST} distribution. **b**, LK distribution ². The red line represents the χ^2 theoretical distribution (df = 6). **c**, Pairwise- F_{ST} distributions. **d** and **e** Principal Component Analysis from *pcadapt* ³ respectively displaying PC3 against PC4 and PC5 against PC6.



Supplementary Figure 6: Allele-specific expression detection. Red lines represent normal distribution estimated from the data.



Supplementary Figure 7: Functional analysis of *Oithona similis* transcripts targeted by ASE and selection. The axes correspond to statistical metrics computed by *dcGO Enrichment*. Are highlighted the most significantly enriched terms in **a**, Biological Process GO-terms, **b**, Cellular Component GO-terms, **c**, Molecular Functions GO-terms.



Supplementary Table 1: *Oithona similis* Mediterranean transcriptomes summary.

	Copepodite				Male			
	1	2	3	4	1	2	3	4
Number of reads	18,052,409	21,048,656	19,129,821	17,624,872	17,823,526	19,237,768	18,292,442	13,933,841
Number of transcripts	81,802	82,811	131,747	83,244	117,951	79,938	115,170	79,038
Number of transcripts with ORF prediction	20,389	20,963	42,404	20,746	38,580	19,488	34,179	20,482

Supplementary Table 2: *Tara Oceans* and *Oithona similis* Mediterranean transcriptomes samples accession numbers

Sampling	Type	Latitude	Longitude	Number of reads	Submission accession	Study accession	Sample accession	Experiment accession	Run accession
<i>Tara Oceans</i>	TARA_155 MetaG	54.5305	-16.9377	182,925,026	-	-	-	-	-
<i>Tara Oceans</i>	TARA_155 MetaT			179,585,384	-	-	-	-	-
<i>Tara Oceans</i>	TARA_158 MetaG	67.1675	0.2177	138,512,807	-	-	-	-	-
<i>Tara Oceans</i>	TARA_158 MetaT			168,629,428	-	-	-	-	-
<i>Tara Oceans</i>	TARA_178 MetaG	77.1604	73.2057	207,442,505	-	-	-	-	-
<i>Tara Oceans</i>	TARA_178 MetaT			170,190,939	-	-	-	-	-
<i>Tara Oceans</i>	TARA_206 MetaG	70.9618	-53.603	223,300,333	-	-	-	-	-
<i>Tara Oceans</i>	TARA_206 MetaT			189,447,541	-	-	-	-	-
<i>Tara Oceans</i>	TARA_208 MetaG	69.1136	-51.5086	223,058,371	-	-	-	-	-
<i>Tara Oceans</i>	TARA_208 MetaT			184,221,525	-	-	-	-	-
<i>Tara Oceans</i>	TARA_209 MetaG	64.7127	-53.0106	181,915,803	-	-	-	-	-
<i>Tara Oceans</i>	TARA_209 MetaT			178,106,374	-	-	-	-	-
<i>Tara Oceans</i>	TARA_210 MetaG	61.5442	-55.9865	182,303,005	-	-	-	-	-
<i>Tara Oceans</i>	TARA_210 MetaT			179,419,896	-	-	-	-	-
Toulon	Transcriptome Copepodite 1	43.1001	5.9482	17,823,526	ERA1801976	ERP114539	ERS3342094	ERX3282627	ERR3255855
Toulon	Transcriptome Copepodite 2	43.1001	5.9482	19,237,768	ERA1801976	ERP114539	ERS3342094	ERX3282628	ERR3255856
Toulon	Transcriptome Copepodite 3	43.1001	5.9482	18,292,442	ERA1801976	ERP114539	ERS3342094	ERX3282629	ERR3255857
Toulon	Transcriptome Copepodite 4	43.1001	5.9482	13,933,841	ERA1801976	ERP114539	ERS3342094	ERX3282630	ERR3255858
Toulon	Transcriptome Male 1	43.1001	5.9482	18,052,409	ERA1801976	ERP114539	ERS3342094	ERX3282631	ERR3255859
Toulon	Transcriptome Male 2	43.1001	5.9482	21,048,656	ERA1801976	ERP114539	ERS3342094	ERX3282632	ERR3255860
Toulon	Transcriptome Male 3	43.1001	5.9482	19,129,821	ERA1801976	ERP114539	ERS3342094	ERX3282633	ERR3255861
Toulon	Transcriptome Male 4	43.1001	5.9482	17,624,872	ERA1801976	ERP114539	ERS3342094	ERX3282634	ERR3255862

Supplementary Table 5: Variant annotation by SNPeff.

SNPeff prediction	Variants under ASE	Variants under selection	Variants under ASE & selection	Total
Missense	54* (12.4%)	66* (12.6%)	15 (9.9%)	2529 (9.8%)
Synonymous	170 (39.1%)	213 (40.8%)	59 (38.8%)	11999 (46.6%)
Start lost	0	0	0	4 (0.02%)
Stop gained	1 (0.2%)	0	0	4 (0.02%)
Stop retained	0	0	0	2 (0.01%)
3'UTR	75 (17.2%)	111* (21.3%)	31 (20.4%)	4480 (17.4%)
5'UTR	80 (18.4%)	79 (15.1%)	29 (19.1%)	4110 (16%)
5'UTR premature start	12 (2.8%)	16 (3.1%)	3 (2%)	778 (3%)
No Transdecoder Hit	23 (5.3%)	20 (3.8%)	8 (5.3%)	1142 (4.4%)
Not mapped	20 (4.6%)	17 (3.3%)	7 (4.6%)	720 (2.8%)
TOTAL	435	522	152	25768

References

1. Cornils, A., Wend-Heckmann, B. & Held, C. Global phylogeography of *Oithona similis* s.l. (Crustacea, Copepoda, Oithonidae) – A cosmopolitan plankton species or a complex of cryptic lineages? *Mol. Phylogenet. Evol.* **107**, 473–485 (2017).
2. R.C., L. & J., K. Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics* **74**, 175–195 (1973).
3. Luu, K., Bazin, E. & Blum, M. G. B. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* **17**, 67–77 (2017).