



HAL
open science

Investigation of the molecular signatures of selection on ATP synthase genes in the marine bivalve *Limecola balthica*

Eric Pante, Vanessa Becquet, Amélia Viricel, Pascale Garcia

► **To cite this version:**

Eric Pante, Vanessa Becquet, Amélia Viricel, Pascale Garcia. Investigation of the molecular signatures of selection on ATP synthase genes in the marine bivalve *Limecola balthica*. *Aquatic Living Resources*, 2019, 32, pp.3. 10.1051/alr/2019001 . hal-02383588

HAL Id: hal-02383588

<https://univ-rochelle.hal.science/hal-02383588v1>

Submitted on 8 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 **Investigation of the molecular signatures of selection on ATP synthase genes in the marine**
2 **bivalve *Limecola balthica***

3 Eric Pante^{1*}, Vanessa Becquet¹, Amélia Viricel¹, and Pascale Garcia¹

4 ¹ Littoral, Environnement et Sociétés (LIENSs), UMR 7266 CNRS – Université de La Rochelle, 2 rue
5 Olympe de Gouges, 17000 La Rochelle, France * Corresponding author. E-mail: epante@univ-lr.fr;
6 Phone: +33.05.46.50.76.37; Fax: +33.05.46.50.76.63.

7 **Abstract**

8 We used transcriptomic sequence data to describe patterns of divergence and selection across
9 different populations of a marine bivalve (*Limecola balthica*). Our analyses focused on a nuclear
10 gene (*atp5c1*) that was previously detected in an F_{ST} scan as highly structured among populations
11 separated by the Finistère Peninsula in France. This gene encodes the gamma subunit of the F₀/F₁
12 ATP synthase, a multi-protein complex that is paramount to cellular respiration and energy
13 production. Analysis of non-synonymous to synonymous mutation ratios revealed that 65% of the
14 gene is highly conserved ($dN/dS \leq 0.1$, min = 0), while 6% of the gene is likely under positive
15 selection ($dN/dS \geq 1$, max = 2.03). All replacement mutations are clustered on a 46 residues portion
16 of the protein, within an inter-peptide interaction zone. Comparative genomics suggests that these
17 mutations are evolutionarily stable, and we hypothesize that they are involved in inter-population
18 genetic incompatibilities with other subunits of the ATP synthase complex. The protein stability of
19 the gamma subunit conferred by southern variants was inferred to be higher under warmer
20 temperatures, suggesting that environmental conditions may contribute to the strength of genetic
21 barriers in *L. balthica*.

22 **Keywords:** Molecular evolution; local adaptation; genetic incompatibilities; selection; ATP
23 synthase; *Macoma balthica*

24 **Running title:** Barriers to gene flow in *Limecola balthica*

25

26 Introduction

27 It is now well-accepted that marine organisms can show local adaptation to environmental
28 conditions despite their high dispersal potential, which was long thought to prevent adaptive
29 differentiation (e.g. review of Sanford and Kelly 2011). In fact, recent research has shown that
30 adaptations can emerge and be maintained in spite of gene flow (Tigano & Friesen 2016). For
31 instance, focusing on the high-gene flow sea urchin *Strongylocentrotus purpuratus*, Pespeni et al
32 (2013) found evidence for local adaptation to temperature in an environmental mosaic, at multiple
33 protein-coding genes that were previously detected using a genome scan.

34 *Limecola balthica* is a highly dispersive marine bivalve, broadly distributed in the northwestern
35 Europe (Väinölä and Varvio, 1989). Its pelagic larval phase is estimated to last 2 to 5 weeks (Caddy
36 1967). In the Northern Atlantic Ocean, two lineages of *L. balthica* occur, due to multiple events of
37 trans-arctic dispersal from the northern Pacific Ocean (Väinölä 2003, Nikula et al 2007). A Pacific
38 lineage (*L. balthica balthica*) occurs in the White and Baltic Seas, and an Atlantic lineage (*L. balthica*
39 *rubra*) is found along the Atlantic coasts from Norway to France and around the British Isles
40 (Väinölä 2003; Luttikhuisen et al 2003; Nikula et al 2007; Becquet et al 2012). The *rubra* lineage
41 extends to the Gironde Estuary, France (with sparse populations down to Arcachon Basin), which
42 corresponds to the present-day range limit for the species (Bachelet 1980). A hybrid zone between
43 *rubra* and *balthica* was detected at the entrance of the Baltic Sea (Kattegat Strait; Nikula et al 2008).
44 Within *rubra*, populations north and south of the Finistère Peninsula (France) show significant
45 mitochondrial and nuclear (microsatellite loci) genetic differentiation (Figure 1b; Becquet et al
46 2012).

47 An F_{ST} -based genome scan was performed on transcriptomic data from pooled individuals, across
48 the Kattegat and Finistère transition zones, in a preliminary effort to look for loci associated with
49 local adaptation (Pante et al 2012). Three geographically-disjunct populations were sampled
50 (Figure 1a). The sites of Aytré (Bay of Biscay, France) and Somme Bay (English Channel, France)
51 correspond to two discrete populations of the *rubra* lineage (Becquet et al., 2012). Aytré is located
52 near the southern limit of the species distributional range and is characterized by warmer sea
53 surface temperatures than Somme Bay (Figure 1b). Furthermore, the Bay of Biscay is subjected to
54 warming surface water temperatures (Goikoetxea et al., 2009). A third population, corresponding to
55 the *balthica* lineage, was sampled in Gdańsk Bay (Baltic Sea, Poland). This F_{ST} scan revealed multiple
56 genes involved in the oxidative phosphorylation (OXPHO) system (including genes coding for
57 subunits of the F_0/F_1 ATP synthase and NADH dehydrogenase complexes, and an ADP/ATP

58 carrier). In particular, the nuclear gene *atp5c1* encoding the gamma subunit of the ATP synthase F1
59 rotor, may be under strong selection: 23/41 mapped SNPs were detected as F_{ST} outliers (maximum
60 $F_{ST} = 0.838$). This gene is paramount to the good functioning of the F₀/F₁ complex, as its rotary
61 action promotes ATP synthesis (reviewed in Sielaff and Börsch, 2013).

62 In this contribution, we examine molecular signatures of selection at *atp5c1* by further analysing the
63 transcriptomic data of Pante et al (2012). Our goals were to determine whether SNPs that are highly
64 differentiated among populations cause modifications of the protein structure, and whether these
65 changes may have functional repercussions. Given that the F₀/F₁ complex is generally encoded by
66 12 nuclear and 2 mitochondrial genes (reviewed in Rand et al., 2004), we asked whether high- F_{ST} ,
67 non-synonymous mutations fall within predicted sites of inter-protein interactions, suggesting the
68 implication of genetic incompatibilities in maintaining barriers to gene flow between southern and
69 northern populations of *L. balthica*.

70

71 **Material and Methods**

72 Specimen sampling, and preparation of genetic data

73 cDNA libraries were prepared from three pools of 10 individuals (one pool per site) and sequenced
74 on a Roche 454 GS-FLX. Details on sequence quality control and assembly, read mapping,
75 polymorphism detection, and gene annotation can be found in Pante et al. (2012). Here we focus on
76 contig G_c113, detected in this latter study, and identified as coding for *atp5c1* (characterized in
77 Matsuda et al., 1993).

78 Sequence evolution and predicted changes in protein function

79 [Differences in allele frequencies between population pairs were tested with Fisher's Exact tests, using](#)
80 [Popoolation2 \(Kofler et al 2011\)](#). The molecular signature of selection was investigated by looking at
81 the ratio between non-synonymous (dN) and synonymous (dS) substitutions (e.g. Kimura, 1977),
82 calculated using the R package seqinr (Charif and Lobry, 2007). A sliding window (150 nt wide,
83 sliding every 3 nt) was used to detect variation in this ratio along the gene. dN/dS ratios were
84 calculated based on population sequence consensus, as in Barreto et al. (2011). The programs
85 SIFT (Kumar et al., 2009; Ng and Henikoff, 2006) and SNAP (Bromberg and Rost, 2007) were used
86 to test whether variations in amino-acid (AA) compositions may have an impact on protein function.

87 SIFT predicts whether AA substitutions can be tolerated by aligning and comparing homologous
88 sequences retrieved by PSI-Blast; protein structure is not taken into account. SNAP uses sequence
89 information, but can incorporate functional and structural annotations if such data are available
90 (Bromberg and Rost, 2007). In SIFT, median conservation of sequences was set at 3.00 and
91 sequences $\geq 90\%$ identical to the query were removed. The hypothesis that local environmental
92 conditions, such as sediment temperature and pH, affect the stability of different alleles of *atp5c1*
93 was tested using I-Mutant 2.0 (Capriotti et al., 2005). Protein stability ($\Delta\Delta G$) was estimated at 5, 15,
94 and 30°C, and at pH 6, 7, and 8 (conditions based on observed data; Ekeboom, 1999 and Lavergne
95 and Barnett, personal communication). In order to further evaluate the impact of AA mutations on
96 protein function, the structure of the gamma subunit was predicted using the I-TASSER server (Roy
97 et al., 2010; Zhang, 2008) and visualized with the UGENE toolkit (Okonechnikov et al., 2012).
98 Finally, we looked for non-synonymous mutations in the other nuclear-encoded F_0/F_1 ATP synthase
99 subunits that could be detected in the transcriptome.

100 **Results**

101 Replacement mutations on *atp5c1*

102 Amino acid sequences of *atp5c1* from populations sampled some 2800 km apart are, overall, highly
103 conserved, with a global dN/dS ratio of 0.08. The sliding window analysis (150 nt wide, sliding
104 every codon; Figure 2a) revealed that 65% of the gene is under negative selection ($dN/dS \leq 0.1$, min
105 = 0), while 6% of the gene is under positive selection ($dN/dS \geq 1$, max = 2.03). Five AA changes
106 occur between residues 126 and 171, meaning that non-synonymous changes are clustered within a
107 46-residue fragment spanning 15% of the sequence. This region was predicted by I-TASSER (best
108 model C-score = 0.04; Figure 2b) to overlap with two known peptide binding sites (Protein Data
109 Bank IDs 2W6I and 2HLD; C-scores of 0.09 and 0.07, respectively). CDD annotation (Conserved
110 Domains Database, Marchler-Bauer et al., 2013) following a BLASTX (Altschul et al., 1990) search
111 for G_c113 suggest that these mutations occur at the interface with the core domain (interaction
112 with the alpha-beta hexamere) and with the epsilon subunit. These non-synonymous changes were
113 predicted to be evolutionarily tolerated based on SIFT and SNAP (expected SNAP accuracy 85-94%).
114 The first four AA changes (L126A, N131D, D136N, H156D; nomenclature: Aytré considered as
115 mutant) discriminate Aytré from Gdańsk Bay and Somme Bay, while the fifth one (T171S)
116 distinguishes Gdańsk from the two French populations. N131D and H156D have a stabilizing effect
117 on the protein, while all others have a destabilizing effect (I-Mutant tests between 15 and 30°C at
118 pH7; Figure 3). In all cases, a relative increase in $\Delta\Delta G$ was observed with increasing temperature

119 and pH (i.e., in warmer conditions, a destabilizing mutation was less destabilizing). In one case
120 (N131D), $\Delta\Delta G$ changed sign (from negative to positive) with increased temperature (in all pH
121 conditions), indicating that the expected effect of the mutation shifted from destabilizing to
122 stabilizing. Considering the 4 AA mutations described above, the Aytré population contained a mix
123 of two alleles, the major allele (85.5% of the reads) being unique to Aytré and one corresponding to
124 the type found at Somme and Gdańsk (14.5% of the reads). The median depth of coverage for contig
125 G_c113 was 55, 45 and 18 for Aytré, Gdańsk and Somme, respectively (Table 1).

126 Potential for incompatibilities with other ATP synthase subunits

127 Contig G c1077 was identified as coding for the alpha subunit and bear one outlier F_{ST} when
128 comparing the samples from Aytré and Somme ($F_{ST} = 0.672$, Pante et al., 2012). This contig could be
129 reliably placed in ORF, which revealed that all nine mutations were synonymous. All mutations
130 segregated Aytré from Somme and Gdańsk. Several contigs were identified as coding for the beta
131 subunit; one (A c2900) was included in our initial F_{ST} scan but did not stand out as bearing outlier
132 SNPs (in addition, this contig could not be reliably placed in ORF). Unfortunately, the epsilon
133 subunit could not be detected in our transcriptome dataset.

134 **Discussion**

135 Adaptive divergence at OXPHO loci

136 Mitochondrial function can be significantly influenced by temperature (e.g. Dahlhoff and Somero,
137 1993). In Dutch populations of *L. balthica*, respiration rate was shown to increase with temperature
138 intra-seasonally (Hummel et al., 2000). In addition, Dutch specimens transplanted in the Spanish
139 estuary of Bidasoa, 200 km south of the known species range limit (i.e. the Gironde Estuary, some
140 80 km from our sampling site in Aytré, France) showed respiration rates significantly higher than in
141 their native populations (Hummel et al., 2000). Our analyses on the effect of temperature on the
142 stability of subunit gamma suggest a small, positive effect of each replacement mutation in the
143 southern population. While only five replacement mutations were detected, previous studies have
144 shown that even few differences in amino acid composition can have a significant impact on
145 mitochondrial performance. In the marine copepod *Tigriopus californicus*, three amino acid changes
146 at the nuclear cytochrome c, separating populations characterized by different thermal regimes,
147 correlated with significant differences in cytochrome c oxidase activity (Rawson and Burton, 2002).
148 In the seed beetle *Callosobruchus maculatus*, metabolic rates were detected between cytotypes that

149 differed by one amino acid at COI and one amino acid at Cyt-B (Arnqvist et al., 2010, and see below).
150 In killer whales (*Orcinus orca*), single amino acid changes at Cyt-B are associated with Antarctic
151 ecotypes, and are therefore possibly implicated in improved mitochondrial performance in polar
152 waters (Foote et al., 2011). Cyt-B was also recently implicated in thermal adaptation in the
153 European anchovy (Silva et al., 2014). Finally, the synergistic impact of temperature and few
154 replacement mutations on protein stability and function was recorded at the nuclear locus encoding
155 cytoplasmic malate dehydrogenase (cMDH) in limpets (Dong and Somero, 2009). As the surface
156 water of the Bay of Biscay is known to be warming at rate of 0.26°C/decade (1977-2007 time
157 period, Goikoetxea et al., 2009) and the geographical range (Jansen et al., 2007) of southern *L.*
158 *balthica* is thought to be receding, it becomes increasingly important from a conservation
159 standpoint to understand how genetic diversity at OXPHO genes relates to adaptive potential.
160 Indeed, as gene flow is limited between southern and northern *rubra* populations at *atp5c1*, a
161 narrowing geographical range could result in the loss of adaptive alleles if the southern population
162 does not persist.

163 Implications for intrinsic genetic incompatibilities among ATP synthase subunits

164 Our data suggest that five replacement mutations may interact with other subunits, likely epsilon
165 (encoded by *atp5e* in humans, Tu et al., 2000). This subunit sits on the F₀ stator, which is composed
166 of nuclear- and mitochondrial-encoded genes and is embedded in the mitochondrial membrane.
167 While *atp5e* was not detected in our dataset, characterizing genetic variation at *atp5e* and other ATP
168 synthase subunits (including mitochondrial and nuclear ones) across the Finistère transition zone
169 seems central to understanding how intrinsic genetic incompatibilities are involved in maintaining
170 barriers to gene flow in *L. balthica*. The sequencing of mitochondrial genomes from individuals
171 sampled on either sides of the Baltic and Finistère hybrid zones (Saunier et al., 2014) will help us
172 shed some light on possible incompatibilities between the nuclear and mitochondrial genes coding
173 for ATP synthase subunits. In addition, the recent discovery of sex-linked heteroplasmy in *L.*
174 *balthica* suggests that mitochondria in this species are characterised by Doubly Uniparental
175 Inheritance (DUI; Pante et al., 2017, and unpublished male mitogenome draft). In DUI species, the
176 somatic tissues of males are characterized by a ‘female’ mitotype (as in all tissues of females), while
177 the male germ line is characterized by a ‘male’ mitotype that is passed on from fathers to sons
178 (reviewed in Zouros, 2013). Female and male mitotypes sampled from a single individual can be
179 highly divergent, reaching up to 52% in freshwater unionoid mussels (Doucet-Beaupré et al 2010).
180 In *L. balthica*, genetic incompatibilities could therefore occur in multiple ways, as OXPHO epistatic

181 interactions among nuclear genes, nuclear and female mitochondrial genes, and among nuclear and
182 male mitochondrial genes in the sperm of interpopulational hybrids.

183 Interplay between intrinsic genetic incompatibilities and temperature

184 One fascinating avenue for research is the characterization of interactions between intrinsic genetic
185 incompatibilities and extrinsic environmental forces. Arnqvist et al. (2010) were the first to
186 demonstrate environmental effects on mitonuclear epistatic interactions by crossing mitochondrial
187 and nuclear genomes of *C. maculatus*. In this very elegant experiment, the authors showed that the
188 negative effects of genetic incompatibilities on metabolic rate were only detectable when hybrids
189 were exposed to different temperatures (Arnqvist et al., 2010). In *L. balthica*, the replacement
190 mutations mapped on *atp5c1* (i) are located in an inter-peptide interaction zone (most likely in the
191 area where the gamma subunit interacts with epsilon), and (ii) seem to influence the stability of the
192 gamma subunit depending on temperature and pH. This preliminary study is based on very limited
193 data, as we used few, pooled individuals from three populations to investigate the molecular
194 signatures of selection on *atp5c1*. Future research should therefore focus on determining, with
195 larger population sampling, if subunits of the F₀/F₁ ATP synthase complex are indeed involved in
196 mitonuclear genetic incompatibilities enforcing genetic barriers, and whether these putative
197 incompatibilities are affected by the environment.

198 **Acknowledgements**

199 We thank Patrick Triplet, Antoine Meirland (Association GEMEL, Picardie) and Rafal Lasota for
200 specimen collections, and the Molecular Core Facility at the University of La Rochelle. We also thank
201 two anonymous reviewers for improving the manuscript. This work was funded by the French
202 Agence Nationale de la Recherche (Hi-Flo project ANR-08-BLAN-0334; HySea project ANR-12-BSV7-
203 0011); salary for EP was covered by a "Contrat de Projet Etat-Région." Funding from the Université
204 de La Rochelle (ACI JCJC) allowed the authors to present at PhysioMar17.

205 **References**

- 206 Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990). Basic local alignment search tool.
207 *Journal of Molecular Biology*, 215:403–410.
- 208 Arnqvist, G., Dowling, D., Eady, P., Gay, L., Tregenza, T., Tuda, M., and Hosken, D. (2010). Genetic
209 architecture of metabolic rate: environment specific epistasis between mitochondrial and

210 nuclear genes in an insect. *Evolution*, 64(12):3354–3363.

211 Bachelet, G. (1980). Growth and recruitment of the tellinid bivalve *Macoma balthica* at the southern
212 limit of its geographical distribution, the Gironde Estuary (SW France). *Marine Biology* 59,
213 105–117.

214 Barreto, F. S., Moy, G. W., and Burton, R. S. (2011). Interpopulation patterns of divergence and
215 selection across the transcriptome of the copepod *Tigriopus californicus*. *Molecular Ecology*,
216 20(3):560–572.

217 Becquet, V., Simon-Bouhet, B., Pante, E., Hummel, H., and Garcia, P. (2012). Glacial refugium versus
218 range limit: conservation genetics of *Macoma balthica*, a key species in the Bay of Biscay
219 (France). *Journal of Experimental Marine Biology and Ecology*, 432–433:73–82.

220 Bromberg, Y. and Rost, B. (2007). SNAP: predict effect of non-synonymous polymorphisms on
221 function. *Nucleic Acids Research*, 35(11):3823–3835.

222 Caddy, J.F. (1967). Development of mantle organs, feeding and locomotion in postlarval *Macoma*
223 *balthica* (L.) (Lamellibranchiata). *Canadian Journal of Zoology*. 47, 609–617.

224 Capriotti, E., Fariselli, P., and Casadio, R. (2005). I-Mutant2.0: predicting stability changes upon
225 mutation from the protein sequence or structure. *Nucleic Acids Research*, 33(Web Server
226 issue):W306–10.

227 Charif, D. and Lobry, J. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical
228 computing devoted to biological sequences retrieval and analysis. In Bastolla, U., Porto, M.,
229 Roman, H., and Vendruscolo, M., editors, *Structural approaches to sequence evolution:*
230 *Molecules, networks, populations, Biological and Medical Physics, Biomedical Engineering*,
231 pages 207–232. Springer Verlag, New York.

232 Dahlhoff, E. and Somero, G. (1993). Effects of temperature on mitochondria from abalone (genus
233 *Haliotis*): adaptive plasticity and its limits. *Journal of Experimental Biology*, 185:151–168.

234 Dong, Y. and Somero, G. N. (2009). Temperature adaptation of cytosolic malate dehydrogenases of
235 limpets (genus *Lottia*): differences in stability and function due to minor changes in sequence
236 correlate with biogeographic and vertical distributions. *Journal of Experimental Biology*,
237 212:169–177.

238 Doucet-Beaupré H., Breton S., Chapman E.G., Blier P.U., Bogan A.E., Stewart D.T. and Hoeh W.R.
239 (2010). Mitochondrial phylogenomics of the Bivalvia (Mollusca): searching for the origin and
240 mitogenomic correlates of doubly uniparental inheritance of mtDNA. BMC Evolutionary
241 Biology 201010:50 doi:10.1186/1471-2148-10-50.

242 Ekebom, J. (1999). Heterotrophic nanoflagellates and bacteria in sediment of a brackish water sill
243 basin in the Baltic Sea. Hydrobiologia, 393:151161.

244 Foote, A. D., Morin, P. A., Durban, J. W., Pitman, R. L., Wade, P., Willerslev, E., Gilbert, M. T. P., and da
245 Fonseca, R. R. (2011). Positive selection on the killer whale mitogenome. Biology Letters,
246 7(1):116–118.

247 Goikoetxea, N., Borja, A., Fontán, A., González, M., and Valencia, V. (2009). Trends and anomalies in
248 sea surface temperature, observed over the last 60 years, within the southeastern Bay of
249 Biscay. Continental Shelf Research, 29:1060–1069.

250 Hummel, H., Bogaards, R., Bachelet, G., Caron, F., Sol, J., and Amiard-Triquet, C. (2000). The
251 respiratory performance and survival of the bivalve *Macoma balthica* (L.) at the southern limit
252 of its distribution area: a translocation experiment. Journal of Experimental Marine Biology
253 and Ecology, 251:85–102.

254 Jansen, J. M., Pronker, A. E., Bonga, S. W., and Hummel, H. (2007). *Macoma balthica* in Spain, a few
255 decades back in climate history. Journal of Experimental Marine Biology and Ecology,
256 344:161–169.

257 Kimura, M. (1977). Preponderance of synonymous changes as evidence for the neutral theory of
258 molecular evolution. Nature, 267:275–276.

259 Kofler, R., Vinay Pandey, R. and Schlötterer, C. (2011) PoPoolation2: Identifying differentiation
260 between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics,
261 27(24):3435–3436.

262 Kumar, P., Henikoff, S., and Ng, P. C. (2009). Predicting the effects of coding non-synonymous
263 variants on protein function using the SIFT algorithm. Nature Protocols, 4(7):1073–1081.

264 Luttikhuisen P.C., Drent J., Baker A.J. (2003) Disjunct distribution of highly diverged mitochondrial
265 lineage clade and population subdivision in a marine bivalve with pelagic larval dispersal.

- 266 Molecular Ecology 12:2215-2229.
- 267 Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M. K., Geer, L. Y., Geer, R. C., Gonzales, N. R.,
268 Gwadz, M., Hurwitz, D. I., Lanczycki, C. J., Lu, F., Lu, S., Marchler, G. H., Song, J. S., Thanki, N.,
269 Yamashita, R. A., Zhang, D., , and Bryant, S. H. (2013). CDD: conserved domains and protein
270 three-dimensional structure. *Nucleic Acids Research*, 41(D1):D348–D352.
- 271 Matsuda, C., Endo, H., Ohta, S., and Kagawa, Y. (1993). Gene structure of human mitochondrial ATP
272 synthase gamma-subunit. Tissue specificity produced by alternative RNA splicing. *Journal of*
273 *Biological Chemistry*, 268(33):24950–24958.
- 274 Ng, P. C. and Henikoff, S. (2006). Predicting the effects of amino acid substitutions on protein
275 function. *Annual Review of Genomics and Human Genetics*, 7:61–80.
- 276 Nikula, R., P. Strelkov, and R. Väinölä. 2007. Diversity and trans-arctic invasion history of
277 mitochondrial lineages in the North Atlantic *Macoma balthica* complex (Bivalvia: Tellinidae).
278 *Evolution* 61:928–941.
- 279 Nikula, R., P. Strelkov, and R. Väinölä. 2008. A broad transition zone between an inner Baltic hybrid
280 swarm and a pure North Sea subspecies of *Macoma balthica* (Mollusca, Bivalvia). *Molecular*
281 *Ecology* 17:1505–22.
- 282 Okonechnikov, K., Golosova, O., Fursov, M., and UGENE team (2012). Unipro UGENE: a unified
283 bioinformatics toolkit. *Bioinformatics*, 28(8):1166–1167.
- 284 Pante, E., Poitrimol, C., Saunier, A., Becquet, V., and Garcia, P. (2017). Putative sex-linked
285 heteroplasmy in the tellinid bivalve *Limecola balthica* (Linnaeus, 1758). *Journal of Molluscan*
286 *Studies*, 83(2):226–228.
- 287 Pante, E., Rohfritsch, A., Becquet, V., Belkhir, K., Bierne, N., and Garcia, P. (2012). SNP detection from
288 De Novo transcriptome sequencing in the bivalve *Macoma balthica*: marker development for
289 evolutionary studies. *PLoS ONE*, 7(12):e52302.
- 290 Pespeni M.H., Palumbi S.R. (2013). Signals of selection in outlier loci in a widely dispersing species
291 across an environmental mosaic. *Molecular Ecology* 22:3580-3597.
- 292 Rand, D. M., Haney, R. A., and Fry, A. J. (2004). Cytonuclear coevolution: the genomics of cooperation.

293 Trends in Ecology & Evolution, 19(12):645–53.

294 Rawson, P. D. and Burton, R. S. (2002). Functional coadaptation between cytochrome c and
295 cytochrome c oxidase within allopatric populations of a marine copepod. Proceedings of the
296 National Academy of Sciences of the United States of America, 99(20):12955–12958.

297 Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G. (2007). Daily high-
298 resolution-blended analyses for sea surface temperature. Journal of Climate, 20:5473–5496.

299 Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein
300 structure and function prediction. Nature Protocols, 5(4):725–738.

301 Sanford E., Kelly M.W. (2011). Local Adaptation in Marine Invertebrates Annual Review of Marine
302 Science 3:509-535.

303 Saunier, A., Garcia, P., Becquet, V., Marsaud, N., Escudié, F., and Pante, E. (2014). Mitochondrial
304 genomes of the Baltic clam *Macoma balthica* (Bivalvia: Tellinidae): setting the stage for
305 studying mito-nuclear incompatibilities. BMC Evolutionary Biology, 14:259.

306 Sielaff, H. and Börsch, M. (2013). Twisting and subunit rotation in single FOF1-ATP synthase.
307 Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences,
308 368:1471–2970.

309 Silva, G., Lima, F. P., Martel, P., and Castilho, R. (2014). Thermal adaptation and clinal mitochondrial
310 DNA variation of European anchovy. Proceedings of the Royal Society of London Series B-
311 Biological Sciences, 281(1792).

312 Tigano A., Friesen V.L. (2016). Genomics of local adaptation with gene flow. Molecular Ecology
313 25:2144-64.

314 Tu, Q., Yu, L., Zhang, P., Zhang, M., Zhang, H., Jiang, J., Chen, C., and Zhao, S. (2000). Cloning,
315 characterization and mapping of the human ATP5E gene, identification of pseudogene
316 ATP5EP1, and definition of the ATP5E motif. Biochemical Journal, 347:17–21.

317 Väinölä, R. 2003. Repeated trans-Arctic invasions in littoral bivalves: molecular zoogeography of the
318 *Macoma balthica* complex. Marine Biology 143:935–946.

319 Väinölä, R. and Varvio, S.-L. (1989). Biosystematics of *Macoma balthica* in northwestern Europe. In

- 320 Ryland, J. and Tyler, P., editors, *Reproduction, genetics and distributions of marine organisms*,
321 23rd Eur Mar Biol Symp, pages 309–316, Fredensborg. Olsen and Olsen.
- 322 Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, 9:40.
- 323 Zouros, E. (2013). Biparental inheritance through uniparental transmission: the doubly
324 uniparental inheritance (DUI) of mitochondrial DNA. *Evolutionary Biology*, 40:1–31.
- 325 Zouros E (2013) Biparental inheritance through uniparental transmission: The Doubly Uniparental
326 Inheritance (DUI) of Mitochondrial DNA. *Evolutionary Biology*, 40: 1–31.

327 **Table 1.** Read counts at each position of codons containing non-synonymous mutations, for each
 328 population. The significance of differences in allele frequencies are presented as $-\log_{10}$ of p-values
 329 from Fisher's Exact tests (significance for at least one pair of populations: *: $-\log_{10}(p)>1$; **: $-\log_{10}(p)>10$).
 330
 331

Locus	Pos	Aytré (A)		Somme (S)		Gdansk (G)		Fisher's Exact Test			
		Major	Minor	Major	Minor	Major	Minor	A-S	A-G	S-G	
L126A	376	G (47)	T (14)	T (25)		T (50)		11.3	18.4	0.0	**
	377	C (51)	T (15)	T (26)		T (51)		12.0	19.2	0.0	**
	378	A (49)	G (15)	G (26)		G (52)		11.7	18.8	0.0	**
N131D	391	G (50)	A (13)	A (24)		A (48)		11.7	19.0	0.0	**
	392	A (58)		A (24)		A (47)		0.0	0.0	0.0	
	393	C (55)		C (24)		C (44)		0.0	0.0	0.0	
D136N	406	A (45)	G (13)	G (27)		G (51)		12.0	18.4	0.0	**
	407	A (58)		A (27)		A (53)		0.0	0.0	0.0	
	408	C (58)		C (25)		C (53)		0.0	0.0	0.0	
H156D	466	G (36)	C (14)	C (26)		C (48)		9.8	14.9	0.0	**
	467	A (50)	G (1)	A (25)	G (1)	A (49)		0.0	0.0	0.5	
	468	C (43)	T (5)	C (23)		C (44)		0.8	1.2	0.0	*
T171S	511	A (53)		A (21)		A (42)		0.0	0.0	0.0	
	512	G (51)	C (2)	G (21)		C (35)	G (7)	0.0	15.9	10.4	**
	513	C (56)		C (21)		C (40)		0.0	0.0	0.0	

332

333 **Figure captions**

334 **Figure 1.** (a) Sampling map displaying the location of two hybrid zones. (b) Heat map of the Bay of
335 Biscay and English Channel with sampling sites of this study (black dots) and of Becquet et al.
336 (2012)(white dots). SST are monthly averages for July, for 1982, 1992, 2002 and 2012; high
337 Resolution SST data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA (Reynolds et al.,
338 2007). (c) Mitochondrial haplotype frequencies at *cox1* (data from Becquet et al., 2012, site
339 locations on panel b) and allele frequencies at the nuclear *atp5c1* (this study). For panels (b) and (c),
340 the sites that are within the putative Finistère hybrid zone (Becquet et al. 2012 and unpublished
341 data) are identified with bolder outer circles.

342 **Figure 2.** (a) Sliding window analysis of the dN/dS ratio along *atp5c1* (window 50 AA wide, sliding
343 every codon). The position of replacement mutations is provided on the right (grey dots), as well as
344 the position of conserved interaction sites detected by CDD (Marchler-Bauer et al., 2013) (black,
345 blue and red: interactions with core domain, delta and epsilon subunits, respectively). (b) Three-
346 dimensional model of the ATP synthase gamma subunit of *Limecola balthica* (I-TASSER prediction
347 based on reference contig from Gdańsk). The yellow segment represents the 46 residue-long
348 fragment bearing replacement mutations (grey dots).

349 **Figure 3.** Effect of temperature and pH on the stability (represented as $\Delta\Delta G$) of the ATP synthase
350 gamma subunit. Error bars represent one standard deviation of three measures of $\Delta\Delta G$ per
351 temperature, using different pH values (see methods). The name of replacement mutations is
352 provided on the right.



