

1 **Genomic testing of landlocked Kildin cod (*Gadus morhua kildinensis*) for its**
2 **ancestral state: stationary or migratory ecotype?**

3 Anastasia A. Teterina and Lev A. Zhivotovsky

4 N.I. Vavilov Institute of General Genetics, 3 Gubkin St., 119991, Moscow, Russia

5 **Corresponding author:**

6 Teterina AA

7 Email: teterina.anastasia@gmail.com

8

9 **Keywords:** Atlantic cod, reproductive isolation, meromictic lake, small population, Barents
10 Sea, ecotypes

11

12 **Running title:** Origins and genomic variation of landlocked Kildin cod.

13

14 **ABSTRACT**

15 Kildin cod is a small landlocked population of Atlantic cod reproductively isolated from
16 marine counterparts for around 1500-2000 years. The Kildin cod lives in a shallow
17 meromictic lake in the five-meter intermediate layer of water with sharp gradients of oxygen
18 and salinity. The cod had an effective population size of around one hundred individuals and
19 evolved unique physiological, morphological and behavioral features. The marine Atlantic
20 cod has two ecologically distinct forms: the stationary (coastal) and migratory (deep-water)
21 ecotypes that differ in migratory behavior and habitat preferences (the depth, oxygen content,
22 salinity and temperature). To understand the origin and genetic properties of Kildin cod, we
23 scrutinized genomic regions associated with the cod ecotypes differentiation (LG1, LG2, and
24 LG7) and found out that Kildin cod's regions LG2 and LG7 were fixed with the migratory
25 variants, whereas polymorphic LG1 had a higher frequency of the stationary variant, that
26 could be explained by the possible strong genetic drift. The lake cod investigated had four
27 times lesser genome diversity than marine population. Our finding suggests that Kildin cod
28 originated from the migratory ecotype of the marine cod.

29 **INTRODUCTION**

30 Exploring ways of species adaptation to various environmental conditions is an important
31 issue of evolutionary biology. Many species have ecological forms with different life
32 strategies that occupy different niches (Engelhard *et al.* 2010, Losos 2010, Jones *et al.* 2012,
33 Nadeau *et al.* 2012, Rocab *et al.* 2003).

34 Two different ecotypes, “migratory” and “stationary”, have been described in Atlantic cod
35 *Gadus morhua* L., a well-studied marine teleost widespread in the continental shelves and
36 coastal waters (Rollefsen 1933, Nordeide *et al.* 2011) on both sides of the Atlantic
37 Ocean (Bradbury *et al.* 2010). The migratory ecotype inhabits the depths of hundreds meters,
38 is prone to extensive spawning and feeding migrations between the coastal and deep waters,
39 while the stationary ecotype prefers lower depths and does not leave the coastal area
40 (Nordeide *et al.* 2011). The ecotypes differ in the rate of growth and age of maturation (Berg
41 & Albert 2003, Brander 1995) and the number of vertebrae (Templeman 1981). Furthermore,
42 the migratory ecotype prefers a higher salinity, lower oxygen content and lower water
43 temperature than the stationary one (Bradbury *et al.* 2010).

44 On the genetic level, vigorous differences in genomic “islands of divergence” on linkage
45 groups LG1, LG2, LG7 and LG12 were found between the ecotypes (Bradbury *et al.* 2010,
46 Sodeland *et al.* 2016). These genomic regions are extended chromosomal inversions of the
47 lengths of 18.5, 5, 9.5, and 13 Mb respectively (Kirubakaran *et al.* 2016, Berg *et al.* 2015,
48 2016) that strongly correlated with environmental conditions (Bradbury *et al.* 2010, Berg *et al.*
49 *et al.* 2015) and migratory behavior (Samuk 2016). The LG1 region contains the pantophysin
50 gene PanI, for which genetic divergence between the ecotypes was first observed (Pogson
51 2001), and is associated with adaptation to the deep sea habitat (Kirubakaran *et al.* 2016). The
52 inversion in LG1 was detected only in the eastern part of the cod range and proved to be
53 under strong directional selection within the migratory ecotype (Samuk 2016, Berg *et al.*
54 2016). LG2 and LG7 are correlated with the oxygen level and the salinity, and were essential
55 for the differentiation between the two ecotypes (Berg *et al.* 2016). LG12 is associated with
56 adaptation to temperature conditions (Berg *et al.* 2015), but unlikely related to the ecotypes
57 divergence (Berg *et al.* 2016).

58 The cod from Lake Mogilnoe (Kildin Island, Barents Sea), well known as Kildin cod, is one
59 of eight cod populations inhabiting the Arctic meromictic lakes. Mogilnoe is a small coastal
60 lake of 9 ha and was separated from the adjacent coastal waters of Barents Sea by a natural

61 dam more than 1000-1500 years ago (Gurevich & Liiva 1975, Figure 1). Lake Mogilnoe
62 consists of three water layers: the upper 4 meters layer is freshwater (1-3 ‰), the middle layer
63 from 4 to 8 meters is saline (5-30‰) and the bottom layer from 8 to 15 meters is poisoned by
64 hydrogen sulfide (Derjugin 1925). The lake is also characterized with step gradients of
65 temperature, oxygen content, pH and other parameters. Freshwater organisms inhabit the
66 upper layer while marine species, and the cod as well, live in the middle saline layer (Derugin
67 1925). Lake Mogilnoe is a nature monument of the Russian Federation, and the Kildin cod is
68 in the national Red list (Shilin 2001). Kildin cod is represented by a small population with the
69 effective size of about a hundred individuals, a several orders of magnitude less than the
70 population of Barents sea cod (Zhivotovsky *et al.* 2016, Andreev *et al.* 2015) and much less
71 than effective sizes of other long-term isolated lake cod populations. The cod from Lake
72 Mogilnoe differs from a marine cod by morphological, behavioral, physiological and genetic
73 characteristics (Stroganov 2015). It inhabits a limited area, 4 meters deep layer of the lake,
74 feeds on primarily crustaceans *Gammarus duebeni* and some other invertebrates (Tseeb
75 1975); the adults largely prey on cod juveniles (Hardie & Hutchings 2015). Kildin cod is an
76 extremely prolific and early maturing, in 3-5 years, and does not live long, up to 8-9 years,
77 in contrast to marine fish: 7-9 years and 16-18, respectively (Mukhina *et al.* 2002).
78 Significant genetic divergence of the lake cod from marine populations was shown by
79 microsatellite and mitochondrial DNA markers (Stroganov *et al.* 2011, Zhivotovsky *et al.*
80 2016, Andreev *et al.* 2015).

81 Among other Atlantic cod populations studied from meromictic lakes are well-studied
82 isolated *Gadus morhua ogac* populations from Lake Ogak, cod from Lake Qasigialiminiq,
83 Lake Tariujarusiq (all in Baffin Island, Canada), and scarcely studied possibly extinct and/or
84 not completely isolated cod populations from two Norwegian lakes, a Green Land lake, and a
85 lake in the Novaya Zemlya (Hardie *et al.* 2008). However, Kildin cod is under complete
86 reproductive isolation for a few hundred generations and has a smallest population size, about
87 100 individuals, as compared with the known long-term isolated cod populations. Particularly,
88 it is four orders of magnitude lower than for the Canadian lake cod populations (Zhivotovsky
89 *et al.* 2016, Andreev *et al.* 2015, Hardie *et al.* 2006).

90 The aim of the present study is to estimate the genome-wide variation of Kildin cod and
91 understand which of the two Atlantic cod ecotypes is its ancestor. We performed whole
92 genome pooled sequencing of the cod samples from Lake Mogilnoe and compared it with
93 the described genomic regions LG1, LG2, LG7 (Berg *et al.* 2016, Kirubakaran *et al.* 2016)

94 and the whole genome data from the open sea and coastal populations of Atlantic cod from
95 previous studies (Karlsen *et al.* 2013).

96

97 **Materials and Methods**

98 *Sample collection and sequencing*

99 Dorsal fin clips from 33 individuals of Atlantic cod from Lake Mogilnoe, Kildin Island,
100 Barents sea (69°19N, 34°20E) and 24 fish from Kildin Strait, Barents sea (69°18N, 34°26E)
101 were collected in 2012 and stored in 96% ethanol. All fish were returned back to their habitat,
102 the order of Rosprirodnadzor of 19.04.2012 N44. Total DNA was extracted with the Diatom
103 DNA prep 200 kit (ISOGEN LAB LTD). DNA concentration was measured by Qubit 2.0
104 (Life Technologies). DNA probes from 33 Kildin cod individuals were mixed in equal
105 proportions for a total amount of 3 mcg. Paired-end sequencing of pooled genomic DNA was
106 performed by Beijing Genomics Institute (BGI, <http://www.genomics.cn/>) on the Illumina
107 HiSeq 200 platform (Illumina Inc., San Diego, CA, USA). In addition, we used data on
108 SOLID sequences of two Atlantic cod population samples: from the coastal (Norwegian
109 coastal cod, NCC) and open sea waters (Northeast Arctic cod, NEAC) (Karlsen *et al.* 2013).

110

111 *SNP discovery and annotation*

112 All datasets were trimmed by read Phred quality >20 and mapped to the cod reference
113 genome (*gadMor1*, Star *et al.* 2011) using *bowtie2* (Langmead & Salzberg 2012). Sorting and
114 filtering (MQ>30) of mapped reads was carried out with *samtools* (Li *et al.* 2009), PCR
115 duplicates removed by *MarkDuplicates* (*picard-tools*, <http://broadinstitute.github.io/picard/>).
116 For genomic variants, additional indels filtering was conducted using scripts *identify-*
117 *genomic-indel-regions.pl* and *filter-pileup-by-gtf.pl* (PoPoolation.v1, Kofler *et al.* 2011a).
118 Single nucleotide variant calling was performed by *FreeBayes* (Garrison & Marth) with *-F*
119 *0.01* and *--pooled-continuous* settings for genome-wide search. Filtration of single nucleotide
120 polymorphic sites was conducted by *vcffilter* program from *vcftools* package (Danecek *et al.*
121 2011) with parameters “QUAL > 20 & QUAL/AO > 10 & SAF > 0 & SAR > 0 & RPR > 1 &
122 RPL > 1”. Annotation of SNPs and possible effect prediction were made with SnpEff
123 (Cingolani *et al.* 2012) with *gadmor1.81* database.

124 *Genome-wide population analysis*

125 Within-population diversity was estimated using Tajima's π and D using windows and a step
126 size of 1000 nucleotides in PoPoolation.v1. For searching across genomic regions with the
127 coverage between 10 and 40, we selected those with high variability in the Kildin cod
128 population, namely with $\pi > 0.004$. Between-population diversity was estimated using Fst-
129 statistic и Fisher exact test in *PoPoolation.v2* (Kofler *et al.* 2011b). Visualization and follow-
130 up work with the data were performed in R (R Development Core Team, 2008) For function
131 analysis, we used genes from target regions with SNPs annotated as “HIGH” or
132 “MODERATE” (according to SNPEff criteria), Gene Ontology annotation and enrichment
133 analysis were perform in R with biomaRt (Durinck *et al.* 2009) and topGo packages
134 (<http://bioconductor.org/packages/topGO/>).

135

136 *Comparison with ecotypes*

137 Assignment of cod loci to the migratory or stationary ecotype's allelic variants was conducted
138 with the SNP-outliers defined in Berg's study (Berg *et al.* 2015). Allele states and frequencies
139 of the SNPs in the open sea and coastal samples were kindly provided by Paul R. Berg.
140 Pooled reads of Kildin cod were mapped to the outlier markers in the flanking regions
141 obtained from dbSNP (<https://www.ncbi.nlm.nih.gov/SNP/>). Reads mapping and alignment
142 filtering were performed as described above. For detection of ecotype variants in a set of SNP
143 marker by *FreeBayes -F 0.01, --pooled-continuous* and *--report-monomorphic* options were
144 used. Only loci with the coverage in SNP position of more than 20 were used for further
145 analyses. The assignment of cod alleles to one of the two ecotypes was performed by a
146 custom *perl* script using the data on allelic states by classification to “stationary”,
147 “migratory”, and “polymorphic” types. The order of loci in the linkage groups was obtained
148 from Berg's study materials (Berg *et al.* 2015).

149 For individual genotyping we chose loci within the genomic inversions on LG1, LG2, LG7
150 and designed the primers for amplification and sequencing (Teterina, Zhivotovsky, 2017).
151 LG1 was typed with the classic *PanI* locus. For LG2 and LG7, we used the regions that flank
152 SNPs ss1712298457 (inside a homolog of *PACSN3*) and ss1712304271 (inside the *AQP11*
153 gene). Purification and Sanger sequencing of samples were carried out in Syntol Ltd (Russia).
154 Sequence alignment and identification of alleles were conducted in *Geneious 4.5* (Kearse *et*

155 *al.* 2012). Testing the Hardy-Weinberg equilibrium (HWE) was performed with Fisher exact
156 test in *GDA* (Lewis & Zaykin 2001), q-values were calculated in R-package *QVALUE*
157 (<http://github.com/jdstorey/qvalue>). Pairwise distances between populations were calculated
158 as the mean of absolute deviations of the “stationary” allele frequencies at the three loci, the
159 *PCoA* was performed in *GenALEx6.5* (Peakall, Smouse, 2006).

160

161 **Results**

162 *Within-population diversity*

163 Sequencing of the Kildin cod sample produced 61 131 381 paired reads. Trimmed reads have
164 been mapped to the cod reference genome and gave an average coverage of about x18.8. Via
165 the variant calling procedure and filtering we obtained 711082 SNPs; that corresponded to the
166 number of polymorphic positions discovered for the open sea (581692 SNPs) and coastal
167 (700999 SNPs) population samples (Karlsen *et al.* 2013). 314532 sites (44%) were biallelic in
168 the Kildin sample.

169 Within-population diversity of Kildin cod ($\pi=0.0012\pm 0.0025$) was estimated across 489458
170 window regions (totally, 60% of the cod genome length) – about four times less than π -values
171 in marine samples reported by Karlsen *et al.* (2013) and was equally low throughout the
172 Kildin cod genome (Figure 2). For genomic regions with high variation ($\pi > 0.0041$, the 95th
173 percentile of the total variation) no over-represented functional genes were not found. The
174 average D score in the Kildin population ($D=0.0207\pm 0.5552$) indicated that population
175 genetic processes were selectively neutral. The high standard deviation probably reflects the
176 susceptibility of the D-statistic to population demographic histories (Schmidt and Pool 2002).

177

178 *Between-population differences*

179 The mean *Fst*-values and their standard deviations between the Kildin and fjord populations,
180 the Kildin and Barents Sea populations, and between the two marine populations of Atlantic
181 cod were, respectively, 0.183 ± 0.177 , 0.189 ± 0.178 and 0.030 ± 0.028 (Figure 3). *Fst*-values
182 between the Kildin cod and each of the two marine populations significantly correlated ($z =$
183 49.9921 , $p\text{-value} < 2.2e-16$, $R^2 = 1$). Significant divergence between the Kildin and marine
184 populations were found in 18697 genomic regions (adjusted $p\text{-value} < 0.005$), which

185 corresponded to 3012 scaffolds. Functional annotation and enrichment analysis of genes from
186 highly divergent genomic regions between the lake and marine populations did not identify
187 any excessively represented functional classes of genes.

188

189 *Comparison with cod ecotypes*

190 Population statistics for the three loci in Kildin cod (MOG) and the Kildin Strait (KS) samples
191 are summarized in Supplementary file A. The three inversions on LG1, LG2 and LG7 were
192 studied in the Kildin cod population. After filtering the data by quality and coverage the
193 allelic states were defined for 39 SNPs in LG1, 15 SNPs in LG2, and 37 SNPs in LG7. The
194 LG1 region of Kildin cod carried both stationary and migratory variants. In LG2 and LG7, the
195 lake cod was monomorphic with the migratory variants (Figure 2). The results were validated
196 by individual genotyping of 33 Kildin cod individuals: the LG1 region was polymorphic
197 carrying the stationary allele with frequency of 0.8 and no deviation from HWE, the regions
198 LG2 and LG7 were monomorphic with the migratory variants. The Kildin Strait sample was
199 polymorphic at the inversions: the frequencies of the stationary variants at LG1, LG2, and
200 LG7 were 0.55, 0.15, and 0.06, respectively. There was a significant deviation from the
201 Hardy-Weinberg expectations at the *PanI* locus (LG1).

202

203 **Discussion**

204 Kildin cod is a small population inhabiting a meromictic lake with an effective size of about a
205 hundred. It originated from an ancestral population of Atlantic cod around eighteen hundred
206 years ago, has greatly reduced genetic diversity due to genetic drift and reproductive isolation
207 (Zhivotovsky *et al.* 2016). In the present study, we showed by the wide-genome analysis that
208 the Kildin cod population greatly differed from marine cod populations across the genome.

209 Based on the ecological features of the Kildin cod, one might have thought that the lake
210 population had originated from the stationary ecotype of Atlantic cod because the Kildin cod
211 inhabits the extremely shallow lake close to the mainland coast and is able to occur in the
212 fresh water layer searching for food (Tseeb 1975, our observations); the latter habit was also
213 confirmed by the pattern of Sr/Ca ratio in the lake cod otoliths, characteristic for the fish
214 periodically changing from salt to fresh waters (Stroganov *et al.* 2017). However, the

215 revealed genomic data suggest that the Kildin cod originated from a marine cod population of
216 the migratory ecotype. Our finding was supported by the fixation in the Kildin cod genome of
217 the migratory variants of LG2- and LG7-inversions, associated with the adaptations to the salt
218 and oxygen contents common for deep-water migratory fish and by the absence of the
219 alternative variants linked to the stationary behavior of shallow-water fish. It is worth to note
220 that another well-known landlocked cod population in Lake Ogak was also monomorphic
221 with the migratory variants of LG2 and LG7 (Bradbury *et al.* 2014).

222 As to the LG1-inversion, it associates with an ancient adaptation of eastern Atlantic cod to the
223 open sea's deep waters where this genomic region is subjected to strong selection, while it
224 becomes selectively neutral relative to the alternative ancestral variant in the shallow offshore
225 waters (Berg *et al.* 2016). Therefore this genomic region was expected to be selectively
226 neutral and subjected to genetic drift in the Kildin cod population. The high frequency of
227 *PanI-A* allele could be likely explained by strong genetic drift known to be a cause of
228 dramatic changes in allele frequencies and the following fixation of the allele in this
229 particular population (Zhivotovsky *et al.* 2016). Highly probable that selectively neutral
230 processes dominate in the Kildin cod population, where both rates of reduction in genetic
231 variation and divergence between the lake population and marine counterparts are roughly
232 uniform across all chromosomes (linkage groups).

233

234 **Conclusions**

235 Genome-wide comparison of Kildin cod with the stationary and migratory ecotypes of the
236 Atlantic cod and validation of the results via individual genotyping showed that Kildin cod
237 were monomorphic with the migratory type of inversions LG2 and LG7 but polymorphic at
238 the LG1-inversion. Our data suggest that Kildin cod originated from the migratory ecotype of
239 the Barents Sea cod.

240 The Kildin cod population was reproductively isolated in a small lake for hundreds of
241 generations and showed signatures of strong genetic drift across the genome that was likely
242 caused by the population's small effective size. Despite these wide-genome random genetic
243 processes, Kildin cod evolved unique physiological, behavioral and morphological
244 characteristics different from those in the marine counterparts that require further genomic
245 studies.

246 **Acknowledgements**

247 The authors are grateful to Kotkin KS, Stroganov AN., Mukhina NV, Tolstenkov OO. The
248 research was supported in part by The Russian Foundation for Basic Research #12-04-00101
249 and #15-29-02421 (grants to LZh).

250

251 **References**

- 252 1. Andreev V, Fokin M, Mogue N, Strelkov P (2015) Long-term persistence and
253 evolutionary divergence of a marine fish population with a very small effective population
254 size (Kildin cod *Gadus morhua kildinensis*). *Marine Biology* **162**, 979-992.
- 255 2. Berg E, Albert OT (2003) Cod in fjords and coastal waters of North Norway: distribution
256 and variation in length and maturity at age. *ICES Journal of Marine Science: Journal du*
257 *Conseil* **60**, 787-797.
- 258 3. Berg PR, Jentoft S, Star B et al. (2015) Adaptation to low salinity promotes genomic
259 divergence in Atlantic cod (*Gadus morhua* L.). *Genome biology and evolution* **7**, 1644-
260 1663.
- 261 4. Berg PR, Star B, Pampoulie C et al. (2016) Three chromosomal rearrangements promote
262 genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific*
263 *reports* **6**, 23246.
- 264 5. Bradbury IR, Bowman S, Borza T (2014) Long distance linkage disequilibrium and
265 limited hybridization suggest cryptic speciation in Atlantic cod. *PloS one* **9**, 106380.
- 266 6. Bradbury IR, Hubert S, Higgins B et al. (2010) Parallel adaptive evolution of Atlantic cod
267 on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal*
268 *Society of London B: Biological Sciences* **277**, 3725-3734.
- 269 7. Brander KM (1995) The effect of temperature on growth of Atlantic cod (*Gadus morhua*
270 L.). *ICES Journal of Marine Science: Journal du Conseil* **52**, 1-10.
- 271 8. Cingolani P, Platts A, Wang LL et al. (2012) A program for annotating and predicting the
272 effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila*
273 *melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80-92.
- 274 9. Derjugin KM (1920) Eine neue Kabeljau-oder Dorschform aus dem See Mogilnoje (Insel
275 Kildin). *Travaux de la Societe des Naturalistes de Petrograde* **LI**, 1-3
- 276 10. Derjugin KM (1925) Relikt lake Mogilnoe (Island Kildin in the Barents Sea). Glavnauka,
277 Leningrad (in Russian).

- 278 11. Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the
279 integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature*
280 *protocols* **4**, 1184-1191.
- 281 12. Engelhard GH, Ellis JR, Payne MR *et al.* (2011) Ecotypes as a concept for exploring
282 responses to climate change in fish assemblages. *ICES Journal of Marine Science* **68**,
283 *1189-1198*.
- 284 13. Garrison E, Marth G (2012) Haplotype-based variant detection from short-read
285 sequencing. *arXiv preprint arXiv:1207.3907*.
- 286 14. Gurevich VI, Liiva AA (1975) Age of the lake Mogilnoe. *Relict Lake Mogilnoe. Nauka,*
287 *Leningrad*,102-104.
- 288 15. Hardie DC, Gillett RM, Hutchings JA (2006) The effects of isolation and colonization
289 history on the genetic structure of marine-relict populations of Atlantic cod (*Gadus*
290 *morhua*) in the Canadian Arctic. *Canadian Journal of Fisheries and Aquatic Sciences* **63**,
291 *1830-1839*.
- 292 16. Hardie DC, Hutchings JA (2015) Cannibalistic growth polyphenism in Atlantic cod.
293 *Evolutionary Ecology Research* **16**, 569-580.
- 294 17. Hardie DC, Renaud CB, Ponomarenko VP *et al.* (2008) The isolation of Atlantic cod,
295 *Gadus morhua* (*Gadiformes*), populations in northern meromictic lakes—a recurrent
296 arctic phenomenon. *Journal of Ichthyology* **48**, 230-240.
- 297 18. Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution
298 in threespine sticklebacks. *Nature* **484**, 55-61.
- 299 19. Karlsen BO, Klingan K, Emblem A *et al.* (2013) Genomic divergence between the
300 migratory and stationary ecotypes of Atlantic cod. *Molecular ecology* **22**, 5098-5111.
- 301 20. Kearsse M, Moir R, Wilson A *et al.* (2012) Geneious Basic: an integrated and extendable
302 desktop software platform for the organization and analysis of sequence data.
303 *Bioinformatics* **28**, 1647-1649.
- 304 21. Kirubakaran TG, Grove H, Kent MP *et al.* (2016) Two adjacent inversions maintain
305 genomic differentiation between migratory and stationary ecotypes of Atlantic cod.
306 *Molecular ecology* **25**, 2130–2143.
- 307 22. Kofler R, Orozco-ter Wengel P, De Maio N *et al.* (2011a) PoPoolation: a toolbox for
308 population genetic analysis of next generation sequencing data from pooled individuals
309 *PloS one* **6**, 15925.

- 310 23. Kofler R, Pandey RV, Schlotterer C (2011b) PoPoolation2: identifying differentiation
311 between populations using sequencing of pooled DNA samples (Pool-Seq).
312 *Bioinformatics* **27**, 3435-3436.
- 313 24. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature*
314 *methods* **9**, 357-359.
- 315 25. Lewis PO, Zaykin D (2001) Genetic Data Analysis: Computer program for the analysis of
316 allelic data. Version 1.0 (d16c). Free program distributed by the authors over the internet
317 from <http://lewis.eeb.uconn.edu/lewishome/software.html>
- 318 26. Li H, Handsaker B, Wysoker A et al. (2009) The sequence alignment/map format and
319 SAMtools. *Bioinformatics* **25**, 2078-2079.
- 320 27. Losos JB. (2010). Adaptive radiation, ecological opportunity, and evolutionary
321 determinism. *The American Naturalist* **175**, 623-639.
- 322 28. Mukhina NV, Lepesevich NA, Filina EA (2002) Biological State of Kildin Cod, in
323 Reliktovoe ozero Mogil'noe (studies of 1997–2000 yy.). Relic Lake Mogil'noe (Studies
324 of 1997–2000), Murmansk: Izd. PINRO, 2002, 88–109, (in Russian).
- 325 29. Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in
326 hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing.
327 *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*
328 **367**, 343-353.
- 329 30. Nordeide JT, Johansen SD, Jørgensen TE *et al.* (2011) Population connectivity among
330 migratory and stationary cod *Gadus morhua* in the Northeast Atlantic- A review of 80
331 years of study. *Marine Ecology Progress Series* **435**, 269-283.
- 332 31. Peakall ROD, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population
333 genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.
- 334 32. Pogson GH (2001) Nucleotide polymorphism and natural selection at the pantophysin
335 (*Pan I*) locus in the Atlantic cod, *Gadus morhua* (L.). *Genetics* **157**, 317–330.
- 336 33. Rocap G, Larimer FW, Lamerdin J *et al.* (2003) Genome divergence in two
337 *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**, 1042-1047.
- 338 34. Rollefsen G (1933) The otoliths of the cod. *Fiskeridirektoratets Skrifter. Serie*
339 *Havundersøkelser* **4**, 1-4.
- 340 35. Samuk K (2016) Inversions and the origin of behavioral differences in cod. *Molecular*
341 *ecology* **25**, 2111-2113.

- 342 36. Schmidt D, Pool J (2002) The effect of population history on the distribution of the
343 Tajima's D statistic. *Preprint, Available online at <http://people.mbi.ohio->*
344 *state.edu/dschmidt/Publications.html*.
- 345 37. Shilin NI (2001) Kildin Cod *Gadus morhua kildinensis* Derugin, 1920, in Krasnaya kniga
346 Rossiiskoi Federatsii. Zhivotnye (Red Data Book of the Russian Federation. Animals),
347 Mos cow: AST Astrel', 308 (in Russian).
- 348 38. Sodeland M, Jorde PE, Lien S et al. 2016. "Islands of Divergence" in the Atlantic Cod
349 Genome Represent Polymorphic Chromosomal Rearrangements. *Genome biology and*
350 *evolution*, **8**, 1012-1022.
- 351 39. Star B, Nederbragt AJ, Jentoft S et al. (2011) The genome sequence of Atlantic cod
352 reveals a unique immune system. *Nature* **477**, 207-210.
- 353 40. Stroganov AN, Afanasiev KI, Burmenskaya VA et al. (2017) About mechanism of
354 adaptation of Kildin cod *Gadus morhua kildinensis* to the specific conditions of Lake
355 Mogilnoe. *Russian Journal of Marine Biology*, in press.
- 356 41. Stroganov AN, Afanasiev KI, Rubtsova GA et al. (2011) Data on variation of
357 microsatellite loci in Kildin cod *Gadus morhua kildinensis* (*Gadidae*). *Journal of*
358 *Ichthyology* **51**, 500-507.
- 359 42. Stroganov AN, Kriksunov EA, Zuykova NV et al. (2015) The biological features of the
360 Kildin cod, *Gadus morhua kildinensis* Derjugin, 1920 (*Gadidae*). *Russian Journal of*
361 *Marine Biology* **41**, 424-431.
- 362 43. Templeman W (1981) Vertebral numbers in Atlantic cod, *Gadus morhua*, of the
363 Newfoundland and adjacent areas, 1947–1971, and their use for delineating cod stocks.
364 *Journal of the Northwest Atlantic Fishery Science* **2**, 21-45.
- 365 44. Teterina AA, Zhivotovsky LA (2017) The development of markers for genetic
366 identification of ecotypes of Atlantic cod *Gadus morhua* *Russian Journal of Genetics*, in
367 press
- 368 45. Tseeb RYA (1975) Feeding. *Relict Lake Mogilnoe. Nauka, Leningrad*, 250-258, (in
369 Russian).
- 370 46. Zhivotovsky LA, Teterina AA, Mukhina NV et al. (2016) Effects of genetic drift in a
371 small population of Atlantic cod (*Gadus morhua kildinensis* Derjugin) landlocked in a
372 meromictic lake: genetic variation and conservation measures. *Conservation Genetics* **17**,
373 229-238.

374

375 **Data accessibility**

376 DNA sequences of pooled Kildin cod sample available from NCBI SRA: SRX1837663.
377 Supplementary file A, filtered VCF file mapped to selected SNPs with ecotype variants, VCF
378 file mapped to gadMor1 genome, genotypes of MOG and KS samples by *PanI*, *PAC3IN3* and
379 *AQP11* loci in GDA format, results from PoPoolation1 and PoPoolation2, and custom perl
380 script available from DRYAD <http://dx.doi.org/10.5061/dryad.108r0> . Pooled sequencing data
381 from Karlsen et al. 2013 available from SRA (SRX110586 and SRX110509). All SNPs could
382 be uploaded from dbSNP.

383

384 **Competing financial interests**

385 The authors declare no competing financial interests.

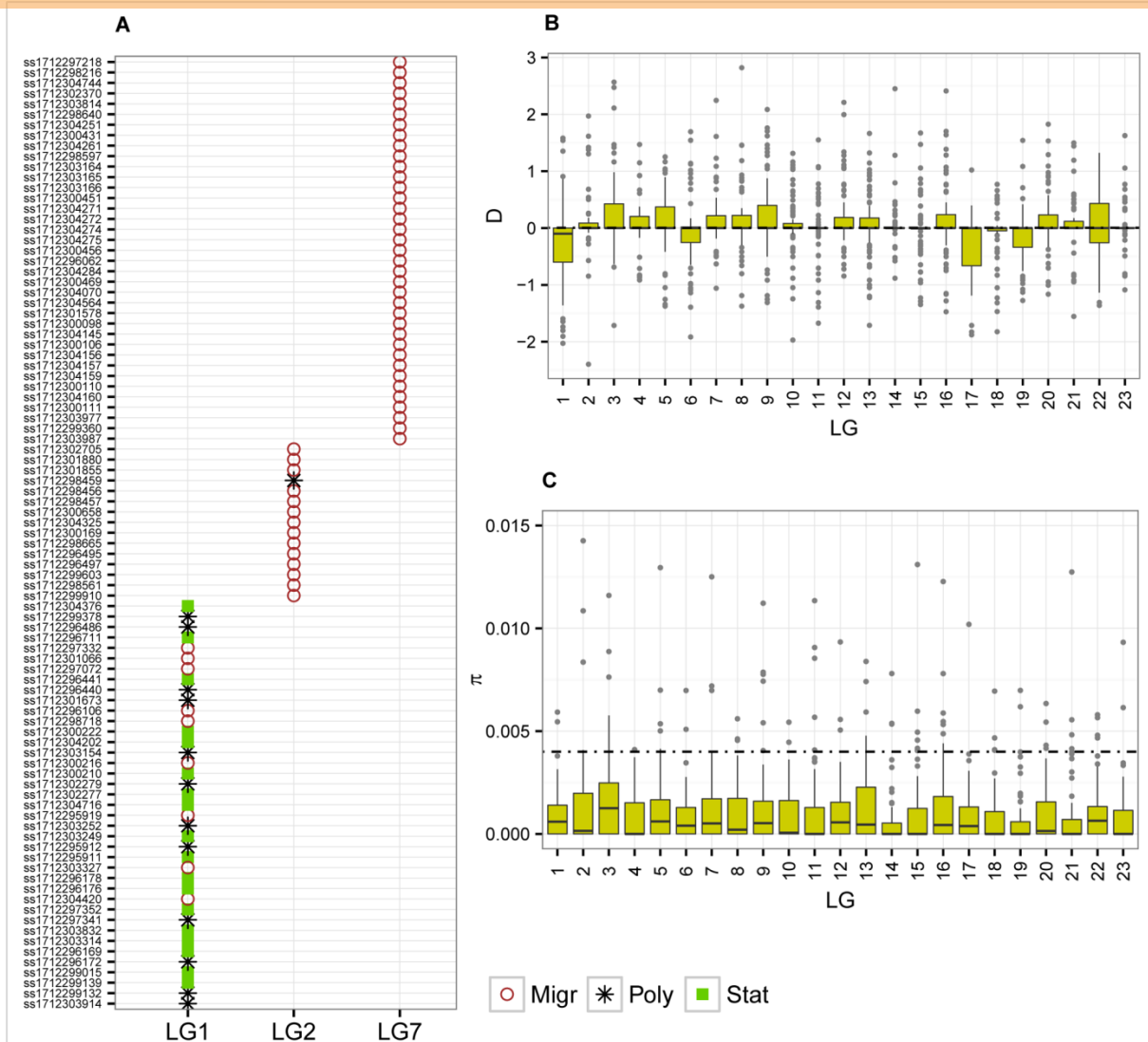
386



387

388 **Figure 1.** Location of Lake Mogilnoe. Lake Mogilnoe separated from Kildin Strait (Barents
389 Sea) by the natural dam by (photo by Stroganov AN, July 20, 2012).

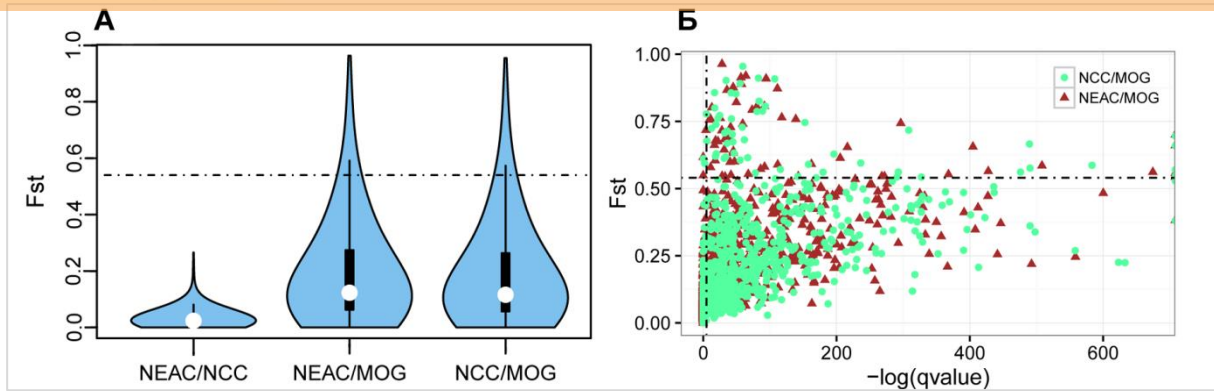
390



391

392 **Figure 2.** A- Determination of alleles at ecotype outlier loci in Kildin cod population from
 393 pool-seq data. Circle marks indicates migratory variant, asterisks – polymorphic sites, solid
 394 squares–stationary type. B – Within-population diversity (Tajima π) of Kildin cod by genomic
 395 linkage groups. Horizontal line marks 95th percentile. C - Within-population diversity (Tajima
 396 D) of lake cod by the linkage groups.

397



398

399 **Figure 3.** A- Violin plot of distribution of Fst values between NEAC vs NCC, NEAC vs Lake
400 Mogilnoe (MOG) and NCC vs Mogilnoe. A dotted line marks 95th percentile of last two
401 groups of comparison. B – Fst values and qvalues of all divergent regions. Horizontal line
402 denotes 95th percentile, a vertical line indicates 5 value, to the right of it there were regions
403 where significant differences (SNPs) of lake cod from the marine cod detected.