bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

1 # Multiple selection signatures in farmed Atlantic salmon adapted
2 # to different environments across Hemispheres

3

4 **Running title:** Selection signatures in farmed Atlantic salmon

5

6 López M.E.[a,b], Linderoth T.[c], Norris, A.[d], Lhorente J.P.[e], Neira R.[f], & Yáñez J.M.[a,e,g*]

7

8 [a] Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile, Santiago, Chile.
9
10 [b] Department of Animal Breeding and Genetics, Swedish University of Agricultural
11 Sciences, Uppsala, Sweden.
12
13 [c] Department of Integrative Biology, University of California, Berkeley, CA, USA.
14
15 [d] Marine Harvest, Kindrum, Fanad, C. Donegal, Ireland.
16
17 [e] Benchmark Genetics Chile, Puerto Montt, Chile.
18
19 [f] Facultad de Ciencias Agronómicas, Universidad de Chile, Santiago, Chile.
20
21 [g] Núcleo Milenio INVASAL, Concepción, Chile.
22

23

24

25 **CORRESPONDENCE:**
26 Dr. José Manuel Yáñez
27 jmayanez@uchile.cl

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

28 **1. ABSTRACT**

29 Domestication of Atlantic salmon started approximately forty years ago, using both
30 artificial and natural selection strategies. Such selection methods are likely to have imposed
31 distinctive selection signatures on the salmon genome. Therefore, identifying differences in
32 selection signatures may give insights into the mechanism of selection and candidate genes
33 of biological and productive interest. Here, we used two complementary haplotype-based
34 statistics, the within-population integrated Haplotype Score test (|iHS|) and the cross-
35 population Extended Haplotype Homozygosity test (XP-EHH) to compare selection
36 signatures in four populations of Atlantic salmon with a common genetic origin. Using
37 |iHS| we found 24, 14, 16 and 26 genomic regions under selection in Pop-A, Pop-B, Pop-C,
38 and Pop-D, respectively. While using the XP-EHH test we identified 27, 25 and 15
39 potential selection regions in Pop-A/Pop-B, Pop-A/Pop-C and Pop-A/Pop-D, respectively.
40 These genomic regions harbor important genes such *igf1r* and *sh3rf1* which have been
41 associated with growth related traits in other species. Our results contribute to the detection
42 of candidate genes of interest and help to understand the evolutionary and biological
43 mechanisms for controlling complex traits under selection in Atlantic salmon.
44
45

46

47 **Keywords**

48 Selection signatures, *Salmo salar,* domestication, SNP data, artificial selection.

49

50

51

52

53

54
55
56

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

57 ## 2. BACKGROUND

58 Atlantic salmon (*Salmo salar* L) were first farmed in Norway during the 1960s, and have
59 now become one of the most important aquaculture species (FAO 2016). Despite a
60 generation interval of three to four years, breeding programs have achieved rapid
61 improvement of economically important traits such as growth, sexual maturation and
62 disease resistance (GJEDREM *et al.* 2012). One of the first farmed populations named Mowi
63 strain, was established with fish from west coast rivers in Norway, with major contributions
64 from River Bolstad in the Vosso watercourse, River Årøy and possibly from the
65 Maurangerfjord area (VERSPOOR *et al.* 2007). Salmon from the Vosso and Årøy rivers are
66 characterized by large size and late maturity (VERSPOOR *et al.* 2007). Phenotypic selection
67 for growth, late maturation and fillet quality was the focus in this population until 1999
68 (GLOVER *et al.* 2009). Ova from this population were imported into Fanad Peninsula,
69 Ireland between 1982 and 1986 to establish an Irish-farmed population (NORRIS 1999).
70 Similarly, ova from this Irish farmed population were introduced to Chile in the early
71 1990s. These stocks were subsequently adapted to the biotic and abiotic factors present in
72 southern hemisphere conditions. Artificial selection and adaptation to captive environments
73 has left detectable genomic patterns in farmed Atlantic salmon populations, as evidenced
74 by differences between wild and farmed populations for several traits, such as growth rate
75 (THODESEN *et al.* 1999; GLOVER *et al.* 2009; SOLBERG *et al.* 2012), predator awareness
76 (EINUM AND FLEMING 1997) and gene transcription patterns (ROBERGE *et al.* 2006; BICSKEI
77 *et al.* 2014; CHRISTIE *et al.* 2016).
78
79 Domestication processes are likely to have exerted selection pressures on certain genomic
80 regions that underlie traits of human interest or other traits involved in adaptation to captive
81 environments. Accordingly, positive selection pressures will cause the frequency of alleles
82 underlying favorable traits to increase rapidly in these domesticated populations. Linkage
83 disequilibrium between favorable mutations and neighboring loci will increase and spread,
84 given there is little opportunity for recombination over the brief time since the onset of
85 intense selection (SABETI *et al.* 2002). Analyses of these selection signatures in domestic
86 animals can provide further insights into the genetic basis of adaptation to diverse
87 environments and genotype/phenotype relationships (OLEKSYK *et al.* 2010; ANDERSSON

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

88 2012). Access to genomic data through next-generation sequencing and high-throughput
89 genotyping technologies have made the comparison of genomic patterns of SNP variation
90 between different livestock breeds possible, allowing for the identification of putative
91 genomic regions and genes under selection in various species including cattle (FLORI *et al.*
92 2009), horses (PETERSEN *et al.* 2013; FRISCHKNECHT *et al.* 2016), sheep (KIJAS *et al.* 2012;
93 FARIELLO *et al.* 2014), pigs (AMARAL *et al.* 2011), Atlantic salmon (VASEMÄGI *et al.* 2005;
94 VASEMÄGI *et al.* 2012; MÄKINEN *et al.* 2014; GUTIERREZ *et al.* 2015; LÓPEZ *et al.* 2018)
95 and tilapia (HONG XIA *et al.* 2015).
96
97 There are several approaches for detecting selection signatures in the genome, one of which
98 relies on the length or variability of haplotypes. Directional selection acting on a new
99 beneficial mutation results in the haplotype harboring the mutation to increase in frequency
100 and to be longer than average. In order to exploit this, Sabeti et al (2002), proposed the
101 extended haplotype homozygosity (EHH) statistic to detect of positive selection in a
102 population, which is specifically the probability that two randomly selected haplotypes are
103 identical-by-descent over their entire length around a core SNP (Sabeti et al 2002). This
104 concept forms the basis for other haplotype homozygosity based metrics, such as the
105 relative EHH (REHH) (SABETI *et al.* 2002) and the widely-used integrated Haplotype Score
106 (|iHS|) (VOIGHT *et al.* 2006). |iHS| compares EHH between derived and ancestral alleles
107 within a population and has the most power to detect selection when the selected allele is at
108 intermediate frequencies in the population (SABETI *et al.* 2006; VOIGHT *et al.* 2006). To
109 detect selection signatures between populations, the cross-population Extended Haplotype
110 Homozygosity test (XP-EHH) compares the integrated EHH profiles between two
111 populations at the same SNP. It was designed to detect ongoing or nearly fixed sites
112 harboring selection in one population (SABETI *et al.* 2007).

113 Although previous studies have already been carried out to detect selection signatures in
114 Atlantic salmon (MÄKINEN *et al.* 2014; GUTIERREZ *et al.* 2015; LIU *et al.* 2016; LÓPEZ *et al.*
115 2018) exploration of selection signatures in additional populations will illuminate how
116 genetic variation among the different strains, adapted to different culture conditions, across
117 hemispheres has not been assessed yet. Herein we used an Affymetrix 200K SNP array
118 dataset to investigate selection signatures in farmed Atlantic salmon populations from the

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

119 same origin, cultivated in Ireland and Chile. We identified several selection signatures
120 using two haplotype-based approaches (|iHS| and XPEHH) at the whole genome level in
121 four Atlantic salmon populations. These findings are important as they highlight regions of
122 the genome that might benefit economically relevant attributes, such as growth, resistance
123 to local diseases and adaptation to specific environmental conditions.

124

125 3. MATERIALS AND METHODS

126 Samples, genotyping and quality control.

127

128 We used a total of 270 individuals from four farmed Atlantic salmon populations of
129 Norwegian origin (Pop-A, n = 40; Pop-B, n = 71; Pop-C, n = 85; Pop-D, n = 74). Pop-A
130 fish are from the Irish strain (Fanad) originating from the west coast Rivers of Norway, as
131 described in the Introduction section. Artificial selection for improving growth, maturity
132 and fillet quality was applied from the beginning in this population (GLOVER *et al.* 2009).
133 We estimated that this population had been under artificial selection for at least ten
134 generations. Pop-B and Pop-C are two different Chilean populations, established with fish
135 from two different year classes of the same Irish strain (Fanad) in the 1990s. Pop-B and
136 Pop-C have been farmed and adapted to the Los Lagos Region, Chile (42°S 72°O). Pop-D
137 is another Chilean population founded with fish from the same Irish farmed strain but
138 adapted to the XIInd Region, Magallanes, Chile (53°S 70°O). Pop-B, Pop-C and Pop-D
139 populations experienced four generations of selective breeding for growth in Chilean
140 farming conditions at the time of sampling.

141

142 Genotyping of all populations was performed using Affymetrix's Atlantic salmon 200K
143 SNP Chip described in YÁÑEZ *et al.* (2016). We assessed SNP quality control using Axiom
144 Genotyping Console (GTC, Affymetrix) and SNPolisher (an R package developed by
145 Affymetrix) *i*) removing SNPs that did not match with high quality clustering patterns,
146 according to the best practices recommended by Affymetrix, *ii*) removing SNPs with call
147 rate lower than 95% and iii) we discarded individuals with genotyping call rate under 90%.
148 We used only SNPs that mapped to chromosomes in the newest version of the Atlantic

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

149 salmon reference genome, ICSAG_v2 (GenBank: GCA_000233375.4). After quality
150 control filtering, 146,102 SNPs remained for downstream analyses.
151
152 **Genetic diversity and population structure**
153
154 We evaluated genetic diversity in terms of the observed heterozygosity (H_O) and expected
155 heterozygosity (H_E) calculated with PLINK v1.07 (PURCELL *et al.* 2007). To investigate
156 population structure based on individual ancestry proportions, we performed model-based
157 clustering assuming no prior knowledge about strain origins in ADMIXTURE 1.2.2
158 (ALEXANDER *et al.* 2009). We performed 10 separate randomly seeded runs for each
159 number K of ancestral populations ($1<K<20$) and selected the optimum K according to the
160 lowest value of the cross-validation error. The aforementioned analyses were conducted
161 using a total of 20,000 SNPs after retaining only those with linkage disequilibrium (LD)
162 values of at most 0.2 to minimize possible confounding effects of LD on the underlying
163 patterns of genetic structure.
164
165 **Selection signatures, gene annotation and functional analyses**
166
167 To detect potentially regions harboring selection signatures, two complementary haplotype-
168 based detection methods, iHS and *XPEHH,* were used for within and between population
169 analyses, respectively.
170 **Detection of within-population selection signatures using *iHS*.** The iHS score is based
171 on the ratio of extended haplotype homozygosity (EHH) for haplotypes anchored with the
172 ancestral versus derived allele. The ancestral allele state for salmon is unknown and so to
173 avoid losing SNPs by trying to polarize them from publicly available outgroup references,
174 we assumed that the major allele represented the ancestral state as used by Bahbahani *et al*
175 (2015). We phased the haplotypes using Beagle (BROWNING AND BROWNING 2009). Single-
176 site iHS values were calculated across the genome for each population. |iHS| scores were
177 calculated using the REHH package (GAUTIER AND VITALIS 2012) and a score threshold of
178 3.0 was used to infer candidate genomic regions under selection.
179

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Detection of between-population selection signatures using *XP-EHH*. The XP-EHH statistic compares the integrated EHH between two populations at the same SNP, in order to identify selection based on overrepresented haplotypes in one of the populations, detecting entirely or approximately fixed sites (SABETI *et al.* 2007). The direction of selection can be determined from the sign of XP-EHH scores, whereby negative XP-EHH scores suggest selection in the 'reference' population, whereas positive scores suggest selection in the 'observed' population. Pop-A was used as the reference population to the other three populations, hence there were three pairs of comparisons.

Gene functional annotation

Genomic regions harboring SNPs showing evidence of selection were annotated based on the ICSAG_v2 reference genome (LIEN *et al.* 2016) using SnpEff (CINGOLANI *et al.* 2012). Gene transcripts from these candidate regions were aligned (using blastx) (ALTSCHUL *et al.* 1990) to the zebra fish (*Danio rerio*) peptide reference database (downloaded from http://www.ensembl.org/) to determine gene identify. As evidence of homology we used an e-value □ 0 and then retrieved the zebra fish gene identifiers and gene ontology (GO) information from the ensembl biomart database (http://www.ensembl.org/biomart).

4. RESULTS

Genetic diversity and structure.

We investigated genetic diversity within each population using SNPs filtered for missing data per individual (max 10%), missing data per marker (max 5%) and allele frequency (min 5%) as described in the Materials and Methods section. A total of 146,103 SNPs were retained for analyses after these quality control steps. Observed heterozygosity levels were similar across the four domestic populations. And was slightly higher than expected for populations A, B and C, and even higher in population D (See Table 1).

Admixture analysis was used to determine the composition of ancestral lineages among individuals to offer insight into the observed genetic variation. We found K=12 ancestral lineages to be optimal in describing the ancestry of the individuals across the 4 populations (Figure 1).

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

212 **Candidate regions under selection - |iHS|**

213 We used the haplotype-based |iHS| test to look for selection within populations. For each
214 population we defined candidate selection regions using the thresholds of |iHS| > 3 (Figure
215 2 and Table 2). Candidate regions were retained if two SNPs separated by ≤ 500 Kb passed
216 this threshold and were annotated using the positions of the first and last SNP as
217 boundaries, extending 500 Kb to each side. In Pop-A we identified 120 markers putatively
218 under selection among ten chromosomes, Ssa02 and Ssa10 combined had approximately 60
219 SNPs. The highest score (-log(p-value) = 5.04) was found in Ssa05 in a region of 6,7 Kb,
220 associated with the CR762469.1 gene; other high scores were found in Ssa10 and Ssa01,
221 nearby to *mipol1*, *furinb*, *csnk1g2a* and *rs17*. Other candidate genes undergoing selection
222 for this population are shown in Supplementary Table S1.

223

224 In Pop-B fourteen regions passed the threshold, distributed among eight chromosomes
225 (Ssa1, 6, 10, 12, 13, 14, 16, and 27). The highest score was in Ssa06, harboring the SASH1
226 gene. Ssa01 and Ssa13 encompassed 4 and 3 regions under selection, respectively,
227 spanning from 11 Kb to 228 Kb. A total of 24 genes were located in these regions
228 (Supplementary Table S1).

229

230 In Pop-C |iHS| detected 121 SNPs passing the threshold and we annotated sixteen genomic
231 regions. Ssa22 showed the highest scores and larger regions under selection, harboring
232 genes such kcnkf, *sc61a*, *mapk3*, *f264* and *cdh2*. Ssa16 and Ssa19 also exhibited high |iHS|
233 scores spanning regions 3 Kb to 1788 Kb.

234

235 Finally, Pop-D presented the highest number of SNPs (134 SNPs) above the threshold
236 compared with other populations, distributed across 11 chromosomes. We defined 25
237 genomic regions under selection, most of them located in Ssa26, where the highest |iHS|
238 scores were also found. Genes such as uqcrfs1, neto1, itfg1 and phkb were found in these
239 regions. Ssa24 also presented higher |iHS| values in one of its regions associated with tchp,
240 ube3b and myo1ha among others. Details of genes and regions can be found in
241 Supplementary Table S1.

242

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

243 **Candidate regions under selection – *XPEHH***

244 We also looked for selection signatures using the XPEHH test between the following
245 populations pairs: A/B; A/C and A/D (Figure 3). We detected, 437 (A/B), 764 (A/C) and
246 262 (A/D) XPEHH scores outlier SNPs indicative of selection (Table 3). We considered
247 potential genomic regions under selection as those containing two or more consecutive
248 SNPs less than 500 Kb apart and that had XPEHH score > 3. After merging overlapping
249 regions 27, 25 and 15 candidate regions were identified for A/B, A/C and A/D comparisons
250 respectively. The total length of the candidate regions was 10.13 Mb for A/B, 12.11 Mb for
251 A/C and 4.05 Mb for A/D. Comparison between A/C yielded negative results in
252 chromosome Ssa14 and Ssa16, furthermore comparison A/D yielded negative results in
253 Ssa14, Ssa24 and Ssa26, suggesting selection in the reference population (A). The gene
254 annotation revealed in A/B the *plecb* gene on Ssa02 and *myo1cb*, *slc43a2a* and *ywhae1*
255 genes on Ssa09, associated with the highest values of XPEHH. In A/C the chromosome
256 Ssa10 presented a large number of SNPs and regions putatively under selection. Also this
257 chromosome presented the highest scores; genes such as *fnbp1l*, *bcar3*, slc5a9 and *fryl* were
258 associated with these values. The highest values for A/D were also located on Ssa10 with
259 *lhx4*, *shr3rf1* and *ftr33* genes. The negative values of XPEHH harboring genes such *agla*,
260 *kcmf1*, *cds1* and *tshz3b* suggest selection on population A.

261

262 **Gene ontology for candidate genes under selection.**

263 To further explore the functions of the candidate genes nearby markers showing evidence
264 of selection signatures, we annotated the candidate genes detected by both methods using
265 DAVID browser (https://david-d.ncifcrf.gov). These candidate genes were enriched in 14
266 gene ontology terms. None of these categories were common across all four populations,
267 but Developmental process and Multicellular organismal process were common on Pop B-
268 C and D. Regulation of biological process was shared for Pop A –B and D; Single-
269 organism process was common for Pop A –B and C; Biological regulation was found in
270 Pop-B and Pop-D; and Growth and Locomotion were common for Pop-B and Pop-C.
271 Anatomical structure development, Biosynthetic process, Cell growth and Single-
272 multicellular organism process were present only in Pop-A; while Localization and
273 Signaling were found only in Pop-B (4).

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

274 **5. DISCUSSION**

275 In this study two complementary tests were used to detect genome wide selection
276 signatures within and between four Atlantic salmon populations with Norwegian origin.
277 We used |iHS| test to evaluate selection signatures within populations and XPEHH to
278 evaluate across populations. We used the oldest population as the reference population
279 when using XPEHH to evaluate the effect of domestication and artificial selection in three
280 different locations in Chile.

281 **Structure and diversity**

282 To examine genetic population structure and relationships among the major groups of
283 salmon, we conducted an ADMIXTURE analyses based on high-quality SNP data. This
284 analysis revealed twelve clusters, which was expected considering the admixed origin of
285 these populations (VERSPOOR *et al.* 2007). The four populations used in this study come
286 from the Mowi strain, which was created, using samples from several rivers along the west
287 coast of Norway (NORRIS *et al.* 1999). The population with the lowest level of admixture
288 was Pop-A, which was also the population with the lowest genetic diversity, a condition
289 that could reflect a higher intensity of artificial selection in this population. Intense artificial
290 selection causes loss of genetic variation as a consequence of the mating of related
291 individuals (GJEDREM 2005). Pop-B and Pop-C showed very similar patterns of
292 heterozygosity and admixture level, which was expected due to the similar breeding
293 practices and environmental conditions to which they have been subjected. Pop-D,
294 however, showed the highest level of heterozygosity and a more complex pattern of
295 admixture, likely produced by a lower pressure of artificial selection on this population.
296 Recent genetic introgression cannot be discarded for Pop-D given the potential of crosses
297 with a different strain for management issues. The results presented here also reinforce the
298 notion that a few generations (at least four in this particular case) are sufficient to generate
299 large changes in terms of genetic structure in farmed Atlantic salmon populations, with the
300 same genetic origin, which have been subjected to different management and
301 environmental conditions. Estimates of inbreeding coefficient (F_{IS}) showed the lowest
302 value in Pop-D, which is consistent with the heterozygosity level in this population. Pop-A
303 presented the second lowest value, despite the fact that this population has been subjected

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

304 to the most intense selection pressure probably due to a better inbreeding management with
305 the use of DNA fingerprinting technology to know relatedness among individuals in order
306 to avoid inbreeding.
307

308 **Selection signatures**

309 As expected the highest |iHS| scores were found in Pop-A because this population has been
310 subjected to more intense artificial selection pressures for a longer time. The number of
311 SNPs under selection detected by this method was similar in Pop-A, Pop-C and Pop-D, but
312 lower in Pop-B. We suggest that this difference is due to the fact that the |iHS| test has little
313 power to detect signals near fixation (SABETI *et al.* 2007; SIMIANER *et al.* 2010). XPEHH,
314 which is more powerful at detecting selection signatures at or near fixation (Sabeti 2007),
315 detected a similar number of regions putatively under selection in Pop-B and Pop-C, but
316 lower in Pop-D. Conversely, |iHS| detected more SNP in this population, suggesting loci
317 under selection in Pop-D have experienced weaker pressure of artificial selection and a
318 greater impact of natural selection, which has prevented allele fixation. Overlaps among
319 regions detected by |iHS| method, were found only when using pairs of populations, that is,
320 a common region was found between Pop-A/B, Pop-A/C, Pop-A/D, Pop-B/C, Pop-B/D and
321 Pop-C/D. No overlap was found among four populations or when using any combination of
322 three. XPEHH detected a higher number of shared regions among populations, specifically
323 in Ssa02, which was common to all 3 tested populations. In addition, shared regions were
324 found in population pairs B/C and Pop-D/C. A greater number of shared regions detected
325 by XPEHH could be explained by a greater power to detect regions that have experienced
326 older selection events (SABETI *et al.* 2007; KLIMENTIDIS *et al.* 2011) than those detectable
327 by |iHS|. Therefore, these regions may be explaining selection signatures that originated
328 before these populations were brought to Chile.

329 **Domestication traits in salmon**

330 Selection signatures found in this study may be involved in some desirable economic traits
331 in salmon production as well as traits that are typically under the effect of domestication.
332 All populations used in this study have been subjected to artificial selection to improve
333 growth rate. According to the functional annotations of the candidate genes, several

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

334 biological processes were found to be involved with growth and development, such as the
335 Development process and Regulation of Biological/metabolic processes. Additionally,
336 some of the genes identified have been associated with growth traits in other species; such
337 *sh3rf1* in chicken and cattle (HANOTTE *et al.* 2003; RUBIN *et al.* 2010) or *igf1r*, which was
338 previously found to be a size locus of large effect in dogs (SUTTER *et al.* 2007; HOOPES *et*
339 *al.* 2012). We suggest that these genes may be under selection for improving growth related
340 traits in salmon. On the other hand, we also identified genes such *scaper*, *clstn3* and *pex5*
341 related to mental disorders in humans (GLATT *et al.* 2005; PETTEM *et al.* 2013). Other genes
342 related to behavioral traits have be found in other Atlantic salmon strains, as well (Lopez et
343 al, 2018), suggesting that artificial selection acts on behavioral traits in salmon as in other
344 domestic animals (Clutton-Brock 1999).

345

346 **6. CONCLUSIONS**

347 In the present study, several candidate genomic regions with selection signatures were
348 identified using two haplotype based methods, |iHS| and XPEHH in four populations of
349 Atlantic salmon. These genomic regions harbored important genes that enriched G terms
350 including growth, developmental processes, and have been associated with growth and
351 behavior in other species. These finding improve our understanding of genomic variants
352 undergoing selection in domestic populations of Atlantic salmon.

353

354 **Ethics approval and consent to participate**

355 The sampling protocol was previously approved by The Comité de Bioética Animal,
356 Facultad de Ciencias Veterinarias y Pecuarias, Universidad de Chile (Certificate N° 29–
357 2014).

358

359 **Consent for publication**

360 Not applicable

361

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

362 **Availability of data and material**

363 Genotype data for each population is available from the online digital repository *figshare*
364 https://figshare.com/s/83efa70722ed5ada023a

365

366 **Competing Interest**

367 The authors have no conflicts of interest to declare

368

369 **Funding**

370 This work has been conceived on the frame of the grant CORFO (11IEI-12843 and
371 12PIE17669), Government of Chile.

372 **Author´s contributions**

373 MEL and JMY conceived the research idea. MEL drafted the manuscript and carried out
374 the analyses. TL supervised the data analyses and contributed to discussion and writing.
375 TL, AN, JPL, RN, and JMY reviewed the manuscript. All authors read and approved the
376 final manuscript.

377

378

379

380 **Acknowledgements**

381 MEL acknowledges the National Commission of Scientific and Technologic Research
382 (CONICYT) for the funding through the National PhD funding program. JMY is supported
383 by Núcleo Milenio INVASAL funded by Chile's government program, Iniciativa Científica
384 Milenio from Ministerio de Economía, Fomento y Turismo

385

386

387

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

388 **REFERENCES**
389
390

391 Alexander, D. H., J. Novembre and K. Lange, 2009 Fast model-based estimation of
392 ancestry in unrelated individuals. Genome Research.
393 Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local
394 alignment search tool. J Mol Biol 215**:** 403-410.
395 Amaral, A. J., L. Ferretti, H.-J. Megens, R. P. M. A. Crooijmans, H. Nie *et al.*, 2011
396 Genome-Wide Footprints of Pig Domestication and Selection Revealed through
397 Massive Parallel Sequencing of Pooled DNA. PLoS ONE 6**:** e14782.
398 Andersson, L., 2012 How selective sweeps in domestic animals provide new insight into
399 biological mechanisms. Journal of Internal Medicine 271**:** 1-14.
400 Bahbahani, H., H. Clifford, D. Wragg, M. N. Mbole-Kariuki, C. Van Tassell *et al.*, 2015
401 Signatures of positive selection in East African Shorthorn Zebu: A genome-wide
402 single nucleotide polymorphism analysis. Scientific reports 5.
403 Bicskei, B., J. E. Bron, K. A. Glover and J. B. Taggart, 2014 A comparison of gene
404 transcription profiles of domesticated and wild Atlantic salmon (Salmo salar L.) at
405 early life stages, reared under controlled conditions. BMC Genomics 15**:** 884.
406 Browning, B., and S. Browning, 2009 A unified approach to genotype imputation and
407 haplotype-phase inference for large data sets of trios and unrelated individuals. Am
408 J Hum Genet 84**:** 210 - 223.
409 Christie, M. R., M. L. Marine, S. E. Fox, R. A. French and M. S. Blouin, 2016 A single
410 generation of domestication heritably alters the expression of hundreds of genes.
411 Nature Communications 7**:** 10676.
412 Cingolani, P., A. Platts, L. Wang, M. Coon, T. Nguyen *et al.*, 2012 A program for
413 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff:
414 SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6**:**
415 80-92.
416 Clutton-Brock, J., 1999 *A natural history of domesticated mammals*. Cambridge University
417 Press.
418 Einum, S., and I. Fleming, 1997 Genetic divergence and interactions in the wild among
419 native, farmed and hybrid Atlantic salmon. J Fish Biol 50.
420 FAO, 2016 The State of World Fisheries and Aquaculture 2016.
421 Contributing to food security and nutrition for all.**:** 200 pp.
422 Fariello, M.-I., B. Servin, G. Tosser-Klopp, R. Rupp, C. Moreno *et al.*, 2014 Selection
423 Signatures in Worldwide Sheep Populations. PLoS ONE 9**:** e103813.
424 Flori, L., S. Fritz, F. Jaffre zic, M. Boussaha, I. Gut *et al.*, 2009 The genome response to
425 artificial selection: a case study in dairy cattle. PLoS ONE 4**:** e6595.
426 Frischknecht, M., C. Flury, T. Leeb, S. Rieder and M. Neuditschko, 2016 Selection
427 signatures in Shetland ponies. Animal genetics.
428 Gautier, M., and R. Vitalis, 2012 rehh: an R package to detect footprints of selection in
429 genome-wide SNP data from haplotype structure. Bioinformatics 28**:** 1176-1177.
430 Gjedrem, T., 2005 *Selection and Breeding Programs in Aquaculture*. Springer, Dordrecht,
431 The Netherlands.
432 Gjedrem, T., N. Robinson and M. Rye, 2012 The importance of selective breeding in
433 aquaculture to meet future demands for animal protein: A review. Aquaculture 350–
434 353**:** 117-129.

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

435 Glatt, S. J., I. P. Everall, W. S. Kremen, J. Corbeil, R. Sásik *et al.*, 2005 Comparative gene
436 expression analysis of blood and brain provides concurrent validation of
437 SELENBP1 up-regulation in schizophrenia. Proceedings of the National Academy
438 of Sciences of the United States of America 102**:** 15533-15538.
439 Glover, K., H. Ottera, R. Olsen, E. Slinde, G. Taranger *et al.*, 2009 A comparison of
440 farmed, wild and hybrid Atlantic salmon (Salmo salar L.) reared under farming
441 conditions. Aquaculture 286.
442 Gutierrez, A. P., J. M. Yáñez and W. S. Davidson, 2015 Evidence of recent signatures of
443 selection during domestication in an Atlantic salmon population. Marine Genomics.
444 Hanotte, O., Y. Ronin, M. Agaba, P. Nilsson, A. Gelhaus *et al.*, 2003 Mapping of
445 quantitative trait loci controlling trypanotolerance in a cross of tolerant West
446 African N'Dama and susceptible East African Boran cattle. Proceedings of the
447 National Academy of Sciences of the United States of America 100**:** 7443-7448.
448 Hong Xia, J., Z. Bai, Z. Meng, Y. Zhang, L. Wang *et al.*, 2015 Signatures of selection in
449 tilapia revealed by whole genome resequencing. Scientific Reports 5**:** 14168.
450 Hoopes, B. C., M. Rimbault, D. Liebers, E. A. Ostrander and N. B. Sutter, 2012 The
451 insulin-like growth factor 1 receptor (IGF1R) contributes to reduced size in dogs.
452 Mammalian genome : official journal of the International Mammalian Genome
453 Society 23**:** 780-790.
454 Kijas, J. W., J. A. Lenstra, B. Hayes, S. Boitard, L. R. Porto Neto *et al.*, 2012 Genome-
455 Wide Analysis of the World's Sheep Breeds Reveals High Levels of Historic
456 Mixture and Strong Recent Selection. PLoS Biol 10**:** e1001258.
457 Klimentidis, Y. C., M. Abrams, J. Wang, J. R. Fernandez and D. B. Allison, 2011 Natural
458 selection at genomic regions associated with obesity and type-2 diabetes: East
459 Asians and sub-Saharan Africans exhibit high levels of differentiation at type-2
460 diabetes regions. Human genetics 129**:** 407-418.
461 Lien, S., B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent *et al.*, 2016 The Atlantic
462 salmon genome provides insights into rediploidization. Nature 533**:** 200-205.
463 Liu, L., K. P. Ang, J. A. K. Elliott, M. P. Kent, S. Lien *et al.*, 2016 A genome scan for
464 selection signatures comparing farmed Atlantic salmon with two wild populations:
465 Testing colocalization among outlier markers, candidate genes, and quantitative trait
466 loci for production traits. Evolutionary applications 10**:** 276-296.
467 López, M. E., L. Benestan, J. S. Moore, C. Perrier, J. Gilbey *et al.*, 2018 Comparing
468 genomic signatures of domestication in two Atlantic salmon (Salmo salar L)
469 populations with different geographical origins. Evolutionary Applications 0.
470 Mäkinen, H., A. Vasemägi, P. McGinnity, T. F. Cross and C. R. Primmer, 2014 Population
471 genomic analyses of early-phase Atlantic Salmon (Salmo salar)
472 domestication/captive breeding. Evolutionary Applications 8**:** 93-107.
473 Norris, A., 1999 Microsatellite genetic variation between and within farmed and wild
474 Atlantic salmon (Salmo salar) populations. Aquaculture 180.
475 Norris, A. T., D. G. Bradley and E. P. Cunningham, 1999 Microsatellite genetic variation
476 between and within farmed and wild Atlantic salmon (Salmo salar) populations.
477 Aquaculture 180.
478 Oleksyk, T. K., M. W. Smith and S. J. O'Brien, 2010 Genome-wide scans for footprints of
479 natural selection. Philosophical Transactions of the Royal Society B: Biological
480 Sciences 365**:** 185-205.

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

481 Petersen, J. L., J. R. Mickelson, A. K. Rendahl, S. J. Valberg, L. S. Andersson *et al.*, 2013
482 Genome-wide analysis reveals selection for important traits in domestic horse
483 breeds. PLoS Genet 9.
484 Pettem, K. L., D. Yokomaku, L. Luo, M. W. Linhoff, T. Prasad *et al.*, 2013 The specific α-
485 neurexin interactor calsyntenin-3 promotes excitatory and inhibitory synapse
486 development. Neuron 80**:** 113-128.
487 Roberge, C., S. Einum, H. Guderley and L. Bernatchez, 2006 Rapid parallel evolutionary
488 changes of gene transcription profiles in farmed Atlantic salmon. Molecular
489 Ecology 15**:** 9-20.
490 Rubin, C.-J., M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood *et al.*, 2010 Whole-
491 genome resequencing reveals loci under selection during chicken domestication.
492 Nature 464**:** 587-591.
493 Sabeti, P., D. Reich, J. Higgins, H. Levine, D. Richter *et al.*, 2002 Detecting recent positive
494 selection in the human genome from haplotype structure. Nature 419**:** 832 - 837.
495 Sabeti, P., S. Schaffner, B. Fry, J. Lohmueller, P. Varilly *et al.*, 2006 Positive natural
496 selection in the human lineage. Science 3129**:** 1614 - 1620.
497 Sabeti, P., P. Varilly, B. Fry, J. Lohmueller, E. Hostetter *et al.*, 2007 Genome-wide
498 detection and characterization of positive selection in human populations. Nature
499 449**:** 913 - 918.
500 Simianer, H., S. Qanbari and D. Gianola, 2010 Detection of selection signatures within and
501 between cattle populations. Proceedings of 9th World Congress on Genetics
502 Applied to Livestock Production.
503 Solberg, M. F., B. O. Kvamme, F. Nilsen and K. Glover, 2012 Effects of environmental
504 stress on mRNA expression levels of seven genes related to oxidative stress and
505 growth in Atlantic salmon Salmo salar L. of farmed, hybrid and wild origin. BMC
506 Res Notes 5.
507 Sutter, N. B., C. D. Bustamante, K. Chase, M. M. Gray, K. Zhao *et al.*, 2007 A single IGF1
508 allele is a major determinant of small size in dogs. Science (New York, N.Y.) 316**:**
509 112-115.
510 Thodesen, J., B. Grisdale-Helland, S. J. Helland and B. Gjerde, 1999 Feed intake, growth
511 and feed utilization of offspring from wild and selected Atlantic salmon (Salmo
512 salar). Aquaculture 180**:** 237-246.
513 Vasemägi, A., J. Nilsson, P. McGinnity, T. Cross, P. O'Reilly *et al.*, 2012 Screen for
514 Footprints of Selection during Domestication/Captive Breeding of Atlantic Salmon.
515 Comparative and Functional Genomics 2012**:** 1-14.
516 Vasemägi, A., J. Nilsson and C. R. Primmer, 2005 Expressed sequence tag-linked
517 microsatellites as a source of gene-associated polymorphisms for detecting
518 signatures of divergent selection in Atlantic salmon (Salmo salar L.). Mol Biol Evol
519 22**:** 1067-1076.
520 Verspoor, E., L. Stradmeyer and J. L. Nielsen, 2007 *The Atlantic salmon: genetics,*
521 *conservation and management*. John Wiley & Sons.
522 Voight, B., S. Kudaravalli, X. Wen and J. Pritchard, 2006 A map of recent positive
523 selection in the human genome. PLoS Biology 4**:** e72.
524 Yáñez, J. M., S. Naswa, M. E. López, L. Bassini, K. Correa *et al.*, 2016 Genomewide
525 single nucleotide polymorphism discovery in Atlantic salmon (Salmo salar):
526 validation in wild and farmed American and European populations. Molecular
527 Ecology Resources**:** n/a-n/a.

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 1. Mean genetic diversity (Observed heterozygosity and expected heterozygosity) of four Atlantic salmon populations

Population	H_o	H_e
Pop-A	0.38±0.16	0.37±0.15
Pop-B	0.40±0.15	0.39±0.14
Pop-C	0.40±0.14	0.39±0.13
Pop-D	0.46±0.22	0.37±0.16

573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

611

Table 2. Number of SNPs identified by |iHS| among populations and chromosomes.

	Pop-A	Pop-B	Pop-C	Pop-D
Ssa01	14	11	2	15
Ssa02	31			
Ssa03				9
Ssa04				1
Ssa05	2	1	5	
Ssa06	1	6		
Ssa07	3			
Ssa08				
Ssa09	6		2	
Ssa10	43	3	6	14
Ssa11			2	
Ssa12	3	7		1
Ssa13	4	16	3	2
Ssa14	4	3	4	4
Ssa15	6	2	1	
Ssa16	1	5	32	
Ssa17				1
Ssa18				
Ssa19		2	15	
Ssa20				18
Ssa21				
Ssa22	2		49	
Ssa23				
Ssa24				15
Ssa25				
Ssa26				54
Ssa27		2		
Ssa28				
Ssa29				
Total	120	58	121	134

612
613
614
615
616
617
618
619
620
621

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Table 3. SNPs and regions under selection identified by *XPEHH* among population pairs and chromosomes.

	Pop-A/Pop-B		Pop-A/Pop-C		Pop-A/Pop-D	
	SNPs	Regions	SNPs	Regions	SNPs	Regions
Ssa01						
Ssa02	68	3	82	5	17	2
Ssa03						
Ssa04						
Ssa05	11	1			26	3
Ssa06						
Ssa07						
Ssa08						
Ssa09	289	17	71	3	6	1
Ssa10	16	2	571	13	154	4
Ssa11					10	2
Ssa12						
Ssa13						
Ssa14			1		8	1
Ssa15			7	1		
Ssa16			25	1		
Ssa17						
Ssa18						
Ssa19						
Ssa20						
Ssa21						
Ssa22						
Ssa23						
Ssa24	53	4			8	1
Ssa25						
Ssa26					32	1
Ssa27						
Ssa28						
Ssa29			7	2	1	
Total	437	27	764	25	262	15

622
623
624
625
626
627
628
629
630
631

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure legends

Figure 1. Individual assignment probabilities generated with ADMIXTURE (1•K•12). Each color represents a cluster, and the ratio of vertical lines is proportional to assignment probability of and individual to each cluster.

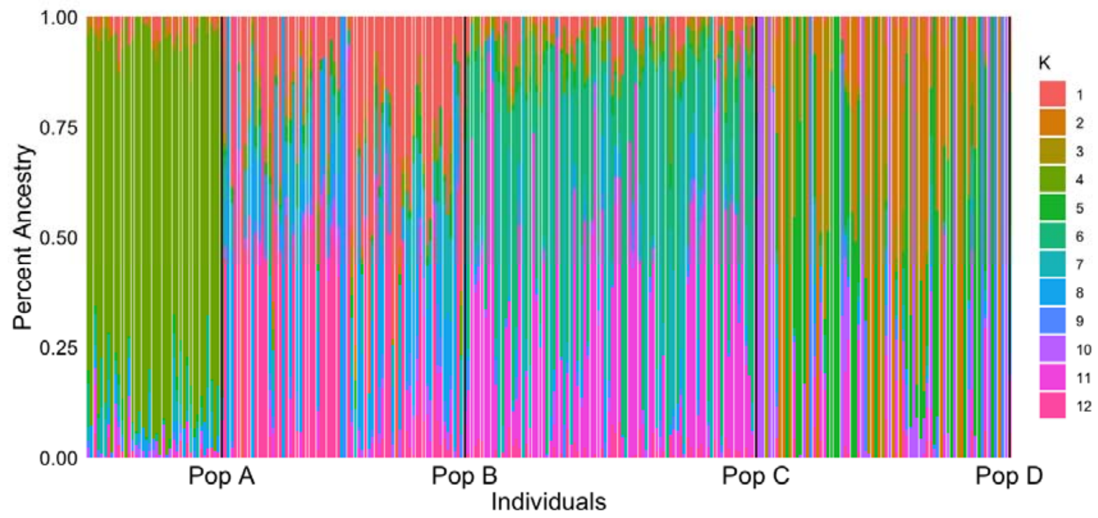
Figure 2. Genome-wide distribution of -log10(p-value) of standardized Integrated Haplotype Score |iHS| among Atlantic salmon populations.

Figure 3. Genome-wide distribution of -log10(p-value) of standardized cross-population extended haplotype homozygosity (XP-EHH) scores in pairwise Atlantic salmon populations.

Figure 4. GO enrichment analysis of genes with evidence of selection in Atlantic salmon. GO functional classification was performed using the DAVID browser.

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

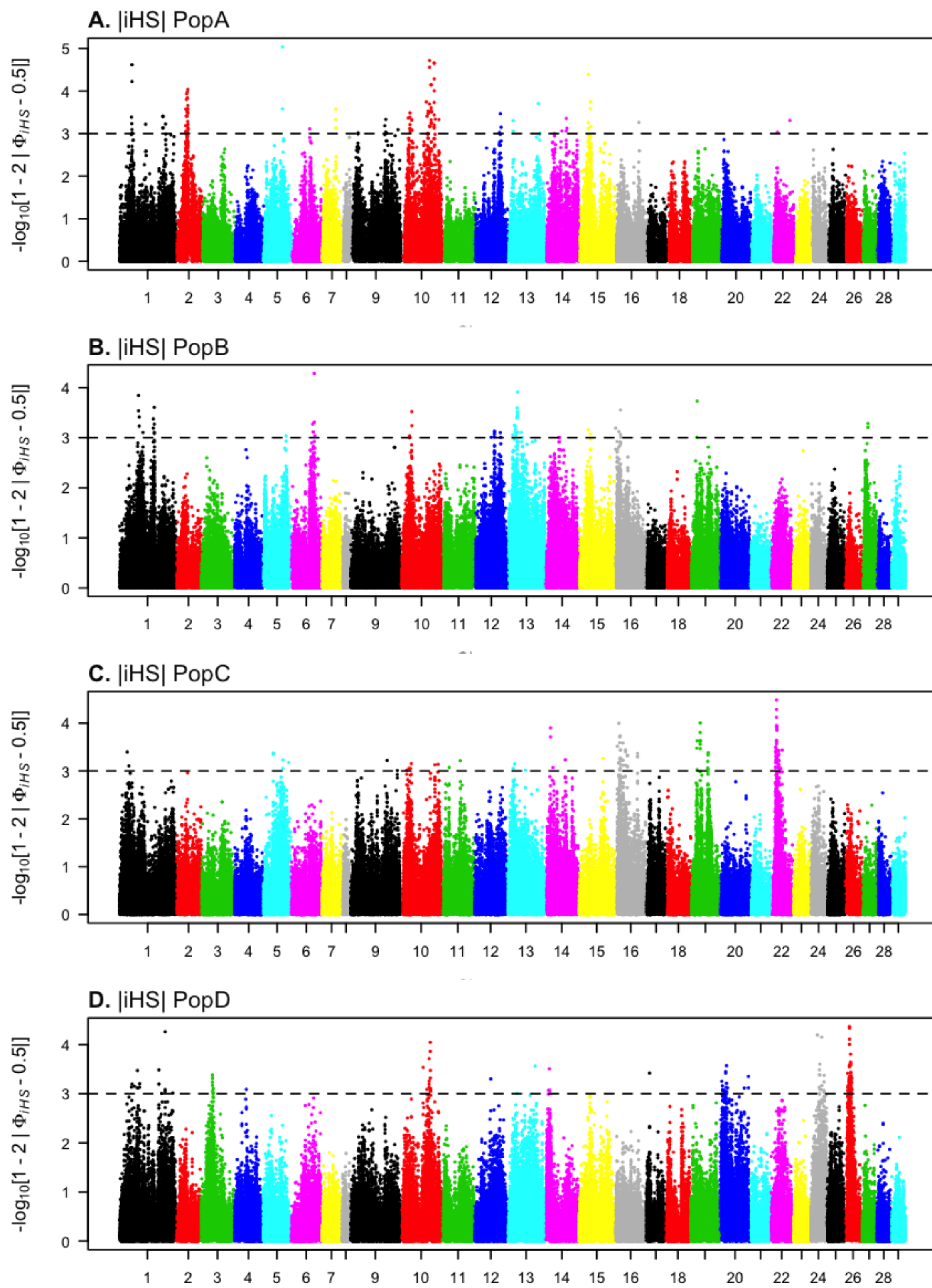
679 **Figure 1**



680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

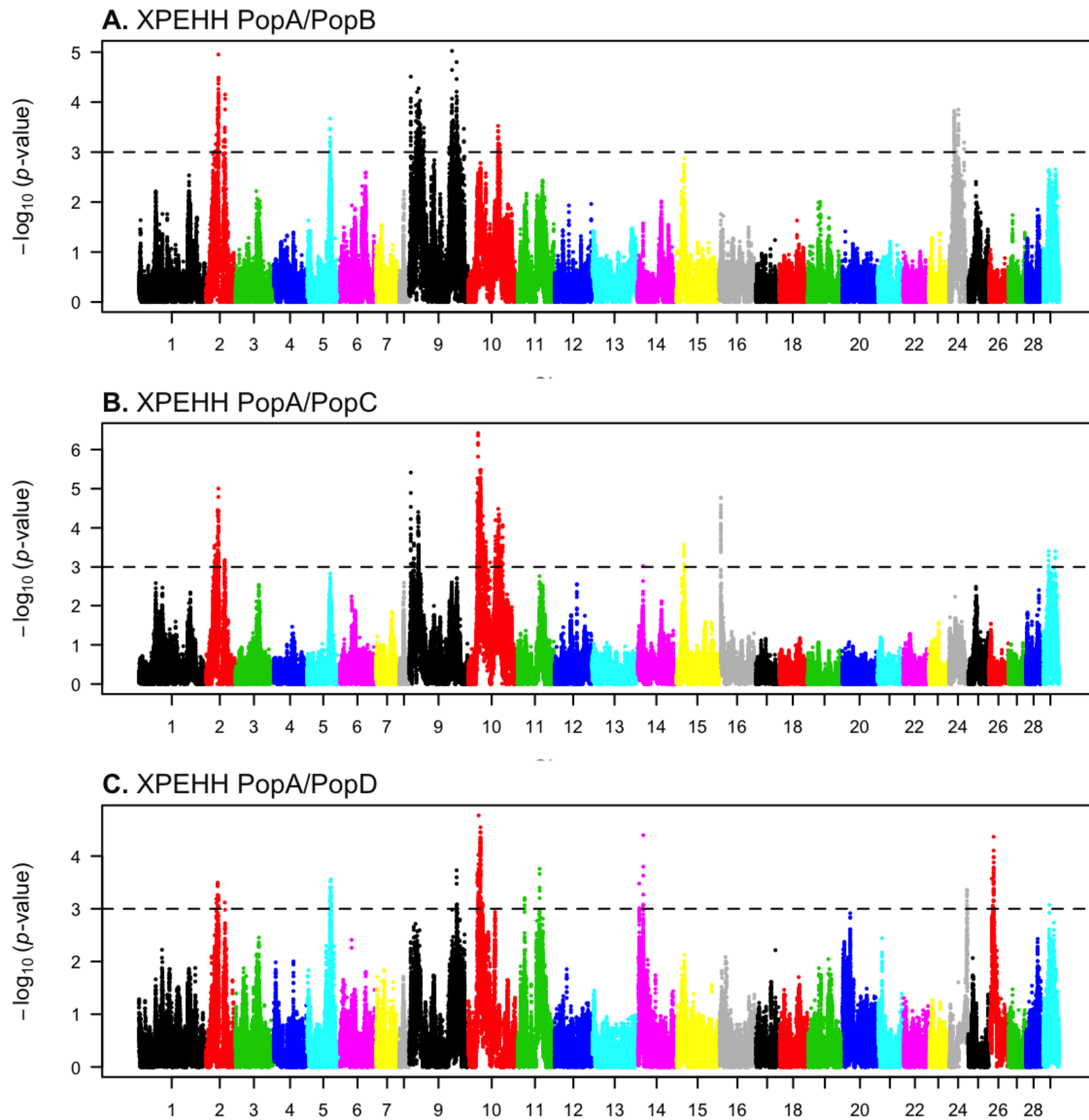
709 **Figure 2**
710
711



712

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

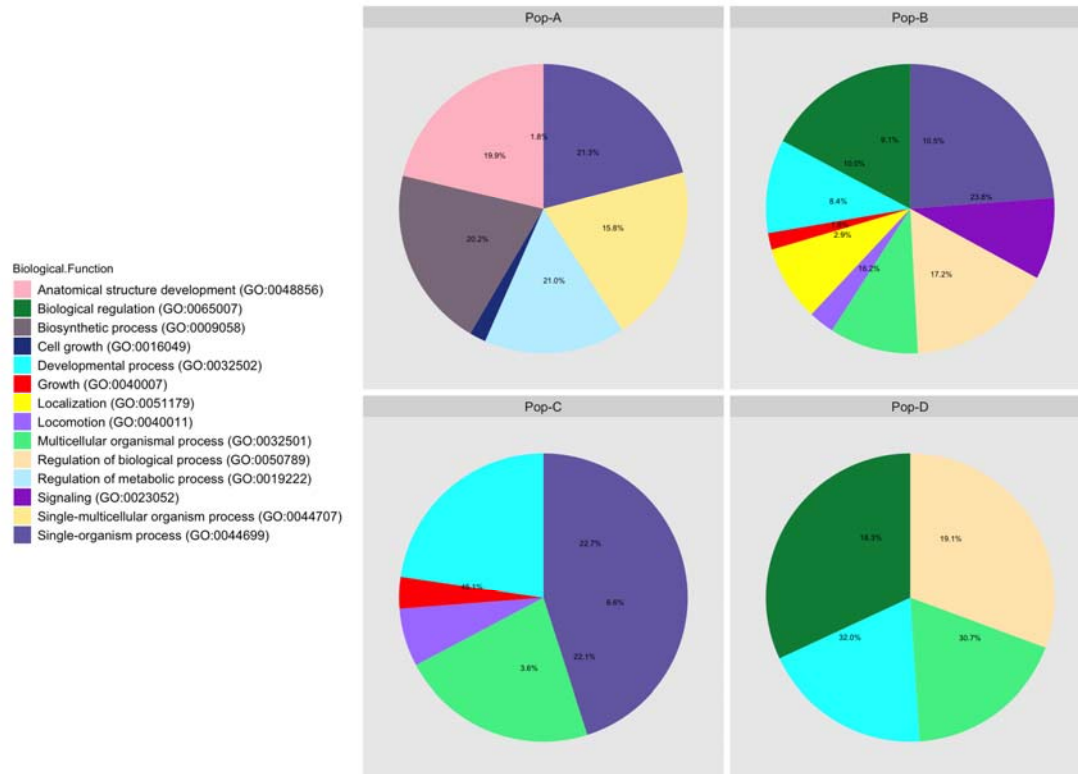
713 **Figure 3**
714
715



716
717
718
719
720
721
722
723
724
725
726
727
728

bioRxiv preprint doi: https://doi.org/10.1101/525550; this version posted January 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

729 **Figure 4**
730
731
732
733



734