Open access • Journal Article • DOI:10.1093/GENETICS/IYAB001

Heritability of DNA methylation in threespine stickleback (Gasterosteus aculeatus) — Source link

Juntao Hu, Juntao Hu, Sara J S Wuitchik, Sara J S Wuitchik ...+5 more authors

Institutions: McGill University, Fudan University, University of Calgary, Boston University ...+1 more institutions

Published on: 03 Mar 2021 - Genetics (Oxford Academic)

Topics: DNA methylation, Epigenomics, Reduced representation bisulfite sequencing, Epigenetics and Methylation

Related papers:

- Heritability of DNA methylation in threespine stickleback (Gasterosteus aculeatus)
- Genome-Wide DNA Methylation Patterns in Wild Samples of Two Morphotypes of Threespine Stickleback (Gasterosteus aculeatus)
- Persistent and plastic effects of temperature on DNA methylation across the genome of threespine stickleback (Gasterosteus aculeatus)
- Unique DNA Methylation Profiles Are Associated with cis-Variation in Honey Bees.
- DNA methylation variation in Arabidopsis has a genetic basis and shows evidence of local adaptation

Share this paper:

View more about this paper here: https://typeset.io/papers/heritability-of-dna-methylation-in-threespine-stickleback-5f31ids2e2

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1 Title: Heritability of DNA methylation in threespine stickleback (*Gasterosteus aculeatus*)

2

3 Running title: Heritability of DNA methylation in stickleback

4

5 Authors: JUNTAO HU, ‡,* SARA J. S. WUITCHIK ,†,**,†† TEGAN N. BARRY,† SEAN M.

6 ROGERS,† HEATHER A. JAMNICZKY§ and ROWAN D. H. BARRETT*

7

8 Initials: J.H., S.J.S.W., T.N.B., S.M.R, H.A.J. and R.D.H.B.

9

10 Keywords: Epigenetics, reduced representation bisulfite sequencing, intergenerationally stable

11 methylation, meQTL, *Gasterosteus aculeatus*

12

13 ‡Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering,

14 Coastal Ecosystems Research Station of the Yangtze River Estuary, Institute of Biodiversity

15 Science, School of Life Sciences, Fudan University, Shanghai, China 200438

16 *Redpath Museum and Department of Biology, McGill University, Montreal, Quebec, Canada

17 H3A 0C4

18 **Informatics Group, Harvard University, 38 Oxford St, Cambridge, MA, 02138

19 ††Department of Biology, Boston University, 5 Cummington Mall, Boston, MA, 02215

20 †Department of Biological Sciences, University of Calgary, Calgary, Alberta, Canada T2N 1N4

21 §Department of Cell Biology and Anatomy, McCaig Institute for Bone and Joint Health,

22 University of Calgary, Calgary, Alberta, Canada T2N 1N4

23

24 Correspondence: Juntao Hu, E-mail: juntao_hu@fudan.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25 **Abstract**

26 Epigenetic mechanisms underlying phenotypic change are hypothesized to contribute to
27 population persistence and adaptation in the face of environmental change. To date, few studies
28 have explored the heritability of intergenerationally stable methylation levels in natural
29 populations, and little is known about the relative contribution of *cis*- and *trans*-regulatory
30 changes to methylation variation. Here, we explore the heritability of DNA methylation, and
31 conduct methylation quantitative trait loci (meQTL) analysis to investigate the genetic
32 architecture underlying methylation variation between marine and freshwater ecotypes of
33 threespine stickleback (*Gasterosteus aculeatus*). We quantitatively measured genome-wide DNA
34 methylation in fin tissue using reduced representation bisulfite sequencing of F1 and F2 crosses,
35 and their marine and freshwater source populations. We identified cytosines (CpG sites) that
36 exhibited stable methylation levels across generations. We found that genetic variance explained
37 an average of 24 to 35% of the methylation variance, with a number of CpG sites possibly
38 autonomous from genetic control. Finally, we detected both *cis*- and *trans*-meQTLs, with only
39 *trans*-meQTLs overlapping with previously identified genomic regions of high differentiation
40 between marine and freshwater ecotypes, as well as identified the genetic architecture underlying
41 two key CpG sites that were differentially methylated between ecotypes. These findings
42 demonstrate a potential role for DNA methylation in facilitating adaptation to divergent
43 environments and improve our understanding of the heritable basis of population epigenomic
44 variation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

45 **Introduction**

46 DNA methylation is a chemical modification to DNA that typically occurs at cytosines within
47 CpG dinucleotides in animals (Suzuki & Bird 2008). It has been suggested that DNA
48 methylation can play a number of biological roles, including gene expression regulation
49 (expression, repression, alternative splicing, and spurious transcription prevention), cell-fate
50 decision, and phenotypic evolution and adaptation to divergent environments (Bird 2007;
51 Bossdorf *et al.* 2008; Maunakea *et al.* 2010; Feil & Fraga 2012; Jones 2012; Verhoeven *et al.*
52 2016; Neri *et al.* 2017; Richards *et al.* 2017). Recent genome-wide studies have revealed that
53 DNA methylation variation is widely observed between closely related animal species and
54 populations that have adapted to ecologically divergent environments (Massicotte *et al.* 2011;
55 Liebl *et al.* 2013; Smith *et al.* 2015; Lea *et al.* 2016; Artemov *et al.* 2017; Le Luyer *et al.* 2017;
56 Hu *et al.* 2018; Hu *et al.* 2019; Laporte *et al.* 2019; Heckwolf *et al.* 2020). In addition,
57 methylation variation has been shown to have a substantial heritable component that selection
58 can act on (Lim & Brunet 2013; Heard & Martienssen 2014; Taudt *et al.* 2016). Modification of
59 the methylome may therefore be an important mechanism underlying phenotypic variation,
60 adaptive evolution, and possibly ecological speciation (Jaenisch & Bird 2003; Turck & Coupland
61 2014; Verhoeven *et al.* 2016).

62 While theoretical studies have suggested that the evolutionary relevance of methylation
63 variation is partially related to its heritability, experimental studies investigating heritable DNA
64 methylation and its role in adaptive evolution are in their initial stages (Verhoeven *et al.* 2016;
65 Hu & Barrett 2017; Richards *et al.* 2017). Although it is clear that DNA methylation levels can
66 sometimes be intergenerationally stable (Jablonka & Raz 2009; Daxinger & Whitelaw 2012;
67 Heard & Martienssen 2014), results have mainly come from plant studies, and the small number
68 of animal studies have typically used isogenic lab lines (Morgan *et al.* 1999; Rakyan *et al.* 2003;
69 but see Nätt *et al.* 2012; Weyrich *et al.* 2016; Weyrich *et al.* 2018; Heckwolf *et al.* 2020). The

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

70 homogenous genetic backgrounds of these isogenic lines may mean that they are not
71 representative of the methylation patterns occurring in more genetically heterogeneous
72 populations (Herman *et al.* 2014; Verhoeven & Preite 2014). In addition, most studies in non-
73 model species have so far been limited to describing broad patterns based on anonymous markers
74 of DNA methylation (Schrey *et al.* 2013; Hu & Barrett 2017; Richards *et al.* 2017), which has
75 hindered understanding of the functional relevance and genetic basis of stable methylation in
76 these species.
77 Methylation variation is mainly under genetic control, which can be caused by DNA
78 sequence variation in both *cis*- and *trans*-regulatory elements (Taudt *et al.* 2016; Hu & Barrett
79 2017). Recently, methylation quantitative trait loci (meQTL) analysis has found both *cis*- and
80 *trans*-acting genetic variants underlying methylation variation (Dubin *et al.* 2015; Orozco *et al.*
81 2015; Kawakatsu *et al.* 2016; Meng *et al.* 2016; Taudt *et al.* 2016). *Cis*-regulatory genetic
82 variation typically affects methylation patterns of only one or a few nearby sites and is less
83 pleiotropic, whereas genetic variants in *trans*-regulatory elements can simultaneously change the
84 methylation levels of multiple sites (Taudt *et al.* 2016; Do *et al.* 2017; Schulz *et al.* 2017;
85 Hannon *et al.* 2018; Gupta *et al.* 2019). However, with the exception of a few studies (Fan *et al.*
86 2019) almost all meQTL studies have been conducted in model species, and thus, the prevalence
87 of *cis*- and/or *trans*-meQTLs, and their role in adaptive evolution in natural populations remains
88 unclear.
89 To explore the stability of epigenetic modification between generations, and to study the
90 genetic architecture of methylation variation between natural populations adapted to distinct
91 environments, we used threespine stickleback (*Gasterosteus aculeatus*), an abundant fish species
92 in both marine and freshwater habitats in the Northern Hemisphere. Since the end of the last ice
93 age, marine stickleback colonized freshwater lake and stream habitats that were uplifted and
94 landlocked, resulting in replicate freshwater populations that show repeated evolution of a suite

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

95 of locally adapted traits (Bell & Foster 1994). The repeated adaptive divergence between marine
96 and freshwater populations makes this a powerful system to study the ecology and genetic
97 architecture of adaptation (Jones *et al.* 2012). In the last decade, a variety of genetic and genomic
98 resources have been developed for this species (Baird *et al.* 2008; Hohenlohe *et al.* 2010; Jones
99 *et al.* 2012; Ishikawa *et al.* 2017; Peichel & Marques 2017). In addition, genome-wide
100 methylation variation between marine and freshwater populations (Smith *et al.* 2015) and
101 between males and females (Metzger & Schulte 2018) have been characterised, as well as the
102 demonstration of methylation responses to environmental change (Artemov *et al.* 2017; Metzger
103 & Schulte 2017; Heckwolf *et al.* 2020). However, the intergenerational stability of methylation
104 in stickleback, and the genetic architecture underlying methylation variation between marine and
105 freshwater ecotypes, remain unclear.

106 We address these gaps by performing an epigenomic survey of fin tissue from sticklebacks
107 under a common garden experimental design with controlled crosses. We first examined
108 methylation divergence between marine and freshwater ecotypes. We then explored levels of
109 methylation and its genetic basis across two generations of the marine-freshwater hybrid lines,
110 and performed meQTL analysis with two F2 families to characterise the genetic architecture of
111 methylation variation between ecotypes. We investigate four specific questions: (1) Is variation
112 in DNA methylation stable between generations? (2) What is the genetic heritability of
113 intergenerationally stable CpG sites? (3) What is the genetic architecture of DNA methylation
114 differences between the stickleback ecotypes? (4) What are the relative contributions of *cis*- and
115 *trans*-meQTLs to DNA methylation differences? Answering these questions will help to provide
116 a baseline for understanding the heritability of methylation variation, and the role of methylation
117 variation in facilitating population persistence and potentially local adaptation in natural
118 populations.

119

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

120 **Materials and methods**

121 *Sampling and husbandry*

122 We collected adult threespine stickleback from one marine (Bamfield Inlet, BI,
123 48°49'12.69"N, 125° 8'57.90"W), and two freshwater (Hotel Lake, HL, 49°38'26.94"N, 124°
124 3'0.69"W, and Klein Lake, KL, 49°43'32.47"N, 123°58'7.83"W) locations in Southwestern
125 British Columbia, Canada in May 2015 (Fig. 1). We transported all fish to an aquatic facility at
126 the University of Calgary, and separated them into population-specific 113 L glass aquaria. We
127 maintained a common garden environment at a density of approximately 20 fish per aquarium,
128 salinity of 4-6 ppt, water temperature of 15 ± 2 °C, and a photoperiod of 16 L: 8 D for one year
129 before making crosses. This period of time should minimize any effects of transportation and
130 allow sufficient time for marine populations to acclimate to hypoosmotic conditions (McCairns
131 & Bernatchez 2010; Morris *et al.* 2014; Wang *et al.* 2014; Artemov *et al.* 2017). We kept each
132 aquarium as a closed system with its own filter, air pump, water supply, and temperature
133 regulator. We fed all fish *ad libitum* once per day with thawed bloodworms (Hikari Bio-Pure
134 Frozen Bloodworms).
135

136 *Crossing design*

137 Threespine stickleback are typically found in either marine or freshwater habitats, but distinct
138 marine and freshwater ecotypes can hybridize, which can facilitate the detection of associations
139 between genotype and phenotype (Jones *et al.* 2012). We generated genetically heterogeneous
140 marine-freshwater F1 families from wild-caught parents by collecting eggs from one marine
141 female and extracting testes from one freshwater male per cross (Fig. S1). To generate a cross,
142 we first equally distributed the eggs into a Petri dish containing fresh water. We then euthanized
143 the male using an overdose of eugenol and removed the testes. We crushed the testes in a Petri
144 dish, with the water activating the released sperm and allowing fertilization. Fertilized eggs were

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

145 left within the Petri dish for 20 minutes before being suspended in a well-aerated mesh-bottom
146 container within 37 L glass aquaria, with an air stone for oxygenation and a sponge filter. In
147 total, we produced one F1 family of BIxHL hybrids (hereafter referred to as HL_F1), and three
148 F1 families of BIxKL hybrids (hereafter referred to as KL_F1). After hatching, the larval fish
149 from the same family were reared in the same 37 L aquaria until reaching approximately 1 cm
150 total length (TL), at which time the families were equally split into aquaria to maintain low
151 densities. The fish and fry were fed twice daily with live *Artemia spp*. nauplii. At approximately
152 2 cm TL, juvenile stickleback fish were transitioned to a diet of chopped thawed bloodworms
153 once per day *ad libitum*. They were then transitioned to an adult diet of full thawed bloodworms
154 gradually. We sampled caudal fin clips (hereafter referred to as 'fin clips') when individuals
155 reached a 3.5 cm TL or more. In addition to the fish we used to generate the F1 crosses, we also
156 sampled extra parental fish from the same marine or freshwater population. Fin clips were stored
157 in 70% ethanol in microcentrifuge tubes at room temperature until extraction of genetic material.
158 To generate F2 families, we randomly selected and crossed one male and one female
159 sibling within an F1 family from each hybrid line (HL or KL) using the same crossing methods.
160 We produced one F2 family of HL hybrids (hereafter referred to as HL_F2) and one F2 family of
161 KL hybrids (hereafter referred to as KL_F2). Fish were raised as described above. We randomly
162 selected fish from HL_F2 and KL_F2 families, and sampled fin clips when individual reached
163 approximately 3.5 cm TL. We stored all fin clips as described above. In addition to the fish we
164 used to make the F2 crosses, we also randomly sampled extra F1 fish from all F1 families. In
165 total, we sampled 94 fish, including 11 parental fish (six marine females; two HL and three KL
166 freshwater males), 19 F1 fish (7 HL_F1 and 12 KL_F1), and 64 F2 fish (28 HL_F2 and 36
167 KL_F2). Detailed information about sex and family is included in Table S1. All sampling,
168 crossing, and housing protocols were approved by the University of Calgary Life and

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

169 Environmental Science Animal Care Committee (AC13-0040 and AC17-0050) following the
170 ethical standards maintained by the Canadian Council for Animal Care.
171
172 *Tissue choice*
173 The choice of tissue used for genome-wide mapping of cytosines can influence the interpretation
174 of methylation patterns (Stricker *et al.* 2017). We conducted our analyses using caudal fin tissue
175 for several reasons. It has been shown that fin position, caudal depth, caudal fin size are different
176 between marine and freshwater stickleback, and that this phenotypic difference is heritable and
177 associated with repeated adaptation to divergent marine and freshwater environments (Walker
178 1997; Jones *et al.* 2012). Because methylation is tissue-specific, choosing a tissue showing
179 phenotypic differences between ecotypes increases the likelihood of finding meQTLs that
180 contribute to this ecotype divergence. Caudal fins can also be dissected quickly and consistently,
181 and the excision of fin tissue does not affect survival.
182
183 *DNA extraction and sex determination*
184 We extracted DNA from caudal fin using phenol:chloroform:isoamyl alcohol (25:24:1), and
185 assessed the quality and quantity using Tecan Infinite® 200 NanoQuant and Quant-iT
186 PicoGreen® dsDNA assay kit (ThermoFisher Scientific). We determined the sex of fish
187 following Peichel *et al.* (2004).
188
189 *Reduced representation bisulfite sequencing*
190 To measure genome-wide DNA methylation levels, we used reduced representation bisulfite
191 sequencing (RRBS) (Meissner *et al.* 2008; Gu *et al.* 2011), following Boyle *et al.* (2012) with
192 some minor modifications. For each individual, we created a library from 120 ng of genomic
193 DNA, and ligated the *MspI*-digested fragments in each library with unique Illumina TruSeq

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

194 adapters. We targeted fragments of 160-340bp (including ~120bp of adapter sequence) using
195 NaCl-PEG diluted SpeedBeads (Rohland & Reich 2012). We split the libraries into four pools
196 (three pools of 24 libraries and one pool of 22 libraries), and treated the pools with sodium
197 bisulfite (EpiTect, Qiagen) following a protocol for formalin-fixed paraffin-embedded samples
198 (Gu *et al.* 2011). After two rounds of bisulfite treatment to ensure complete conversion of
199 unmethylated cytosines, each pool was amplified with Illumina primers, and loaded in four lanes
200 (100-bp single-end reads) of a Hiseq 2500 at the McGill University and Genome Quebec
201 Innovation Centre. In total, we sequenced all 94 fish sampled across three generations (Table
202 S1). Each sample was sequenced to a mean depth (± SD) of 8.094 ± 2.532 million reads. The
203 average mapping efficiency was 61.4 ± 4.7% (± SD). We quantified methylation at non-CpG
204 motifs and found less than 1% non-CpG cytosines were methylated, suggesting a highly efficient
205 bisulfite conversion.
206
207 *Read filtering and mapping*
208 To remove adapter contamination, low-quality bases, and bases artificially introduced during
209 library construction, we trimmed reads using Trim Galore! v0.4.4
210 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), with the 'rrbs'option. We
211 then used the program Bowtie2 v2.2.9 (Langmead & Salzberg 2012), implemented in Bismark
212 v0.17.0 (Krueger & Andrews 2011) to align trimmed reads for each sample to the stickleback
213 genome (ENSEMBL version 98) with default settings, except for tolerating one non-bisulfite
214 mismatch per read. We only included reads that mapped uniquely to the reference genome, and
215 cytosines that had at least 10x coverage in downstream analyses. Only CpG context cytosine
216 methylation was analysed because CpG methylation is the most common functional methylation
217 in vertebrates (Suzuki & Bird 2008).
218

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

219 *General methylation patterns*

220 To identify general methylation patterns, we first performed a principal component analysis
221 (PCA) on methylation levels in all samples using the *prcomp* function in R (R Core Team, 2018,
222 v3.4.3). We ran the analysis by first identifying cytosines that were covered in all samples using
223 the R package methylKit v1.4.1 (Akalin *et al.* 2012). Read coverage was then normalized
224 between samples, using the median read coverage as the scaling factor. A minimum of ten reads
225 in all samples was required at a CpG site for that site to be analysed. We removed CpG sites that
226 were in the 99.9th percentile of coverage from the analysis to account for potential PCR bias. We
227 calculated the methylation levels by extracting the total amount of methylation-supporting reads,
228 and the total coverage of each CpG site, using the *percMethylation* function in the R package
229 methylKit. To improve methylation estimates, we corrected for SNPs, which could have resulted
230 in an incorrect methylation call if C-to-T and G-to-A SNPs were falsely interpreted as
231 unmethylated cytosines, following Heckwolf *et al.* (2020). We first identified SNPs using the
232 methylation value of each CpG site of all 11 parental individuals for input to Bis-SNP v0.82.2
233 (Liu *et al.* 2012) with the default parameters. Because Bis-SNP is sensitive to the directionality
234 of the RRBS protocol (i.e., whether sequenced reads come from the original forward and reverse
235 strands when calling C-to-T and G-to-A SNPs), we used a directional bisulfite-seq protocol that
236 is similar to Krueger *et al.* (2012). We observed a similar number of reads per individual in our
237 study vs. Heckwolf *et al.* (2020), suggesting that sufficient coverage on both strands was
238 obtained to distinguish SNPs from conversions at C-to-T and G-to-A SNPs. We chose parental
239 samples for identifying SNPs because they are the genetic source of the F1s and F2s, and are the
240 most genetically heterogeneous samples. We used GATK's VariantFiltration and SelectVariants
241 to restrict variants to diallelic sites, and filter variants based on the following GATK variant
242 annotation cut-offs: QD < 2.0, MQ < 40.0, MQRankSum < -12.5, and ReadPosRankSum < -8.0.
243 We then used VCFtools v0.1.16 (Danecek *et al.* 2011) to remove SNPs with a minor allele

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

244 frequency (MAF) greater than 0.0049 (Heckwolf *et al.* 2020), and more than 10% missing data
245 across all 11 parental samples. Using MAF thresholds from 0.001 to 0.01 resulted in similar
246 numbers of filtered SNPs. We calculated pairwise weighted F_{st} between BI, HL and KL fish in
247 the parental generation using VCFtools on filtered SNPs. We produced a list of positions (C-to-T
248 and G-to-A SNPs) for correcting methylation estimates, using custom written Perl scripts from
249 Heckwolf *et al.* (2020) and the R package GenomicRanges v.1.30.3 (Lawrence *et al.* 2013). In
250 addition, it has been suggested that sex specific methylation affects less than 0.1% of CpG sites
251 on autosomal chromosomes, but more than 5% of CpGs on the sex chromosome in stickleback
252 (Metzger & Schulte 2018). Therefore, to exclude a potential sex bias, we removed all CpGs
253 located on the sex chromosome (group XIX). In total, we retained 52,940 CpG sites that passed
254 the filtering step. To perform PCA, methylation levels at each CpG site were taken as input
255 variables, whereas each point in multidimensional space represented a stickleback individual.
256 Finally, to compare DNA methylation variation levels between F1 and F2 fish in each hybrid
257 line, we calculated DNA methylation levels in 7,840 1-kb tiling windows (step = 1 kb; size = 1
258 kb) compiled from the same 52,940 CpG sites, and compared the standard deviations of
259 methylation levels for each genomic window within each hybrid line.
260
261 *Analysis of methylation divergence between ecotypes*
262 To examine methylation divergence between ecotypes, we performed a differential methylation
263 analysis between marine and freshwater populations from the parental generation, using the
264 52,940 CpG sites that passed the filtering step above. CpG sites were considered to be
265 differentially methylated cytosines (DMCs) with a false discovery rate correction Q-value < 0.01
266 and a minimum required methylation difference of 15% between ecotypes, using the R package
267 methylKit with sampling site (i.e., BI for the marine population, and KL and HL for the
268 freshwater populations) as an covariate. We visualized differential methylation patterns across

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

269 individuals and obtained clustering of samples and DMCs in heatmaps with the "complete"
270 clustering method on Euclidian distances, using the R package pheatmap version 1.0.8
271 (https://cran.r-project.org/web/packages/pheatmap/index.html). We clustered hyper- and
272 hypomethylated DMCs between ecotypes using relative percent DNA methylation, which is the
273 normalized percent DNA methylation scaled for each DMC's percent DNA methylation (median
274 percent methylation as 0) of marine and freshwater fish in heatmaps. We also clustered
275 individual fish based on overall methylation patterns across DMCs. We then analysed the
276 proportion of cytosines within genomic features (promoter/exon/intron/intergenic; promoters are
277 defined as regions being upstream 1000 bp and downstream 1000 bp around the transcription
278 start sites (TSSs)) for DMCs, using the R package genomation v1.6.0 (Akalin *et al.* 2015).
279 Because *MspI* restriction sites are not randomly distributed in the genome, we built a null
280 distribution of genomic features based on all filtered CpG sites (i.e., 52,940 CpG sites). We gave
281 precedence to promoters > exons > introns > intergenic regions when features overlapped (Smith
282 *et al.* 2015; Hu *et al.* 2018). Finally, we annotated genes associated with DMCs, using the R
283 packages biomaRt v2.34.2 (Durinck *et al.* 2005; Durinck *et al.* 2009) and ChIPpeakAnno v3.12.7
284 (Zhu *et al.* 2010; Zhu 2013) on the stickleback reference genome from Ensembl 98 database, and
285 performed gene ontology (GO) analysis on DMC-associated genes, using the R package topGO
286 v2.28.0 (Alexa *et al.* 2006). Over-represented GO terms were those with multiple-test corrected
287 *P*-values (Benjamini-Hochberg's false discovery rate) below 0.1, based on a Fisher's exact test.
288 We compared DMC-associated genes with the genes associated with the 52,940 CpGs that
289 passed the filtering step.
290
291 *Analysis of intergenerationally stable methylation*
292 We considered a CpG as intergenerationally stable when the CpG was not significantly
293 differentially methylated between F1 and F2 generations of the same family within the same

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

294 hybrid line (HL or KL) and fulfilled this criterion in both hybrid lines. Note that these 'stable'
295 sites are not necessarily 'heritable' in the sense of methylation variation between individuals
296 being due to additive genetic factors. DMCs between fish in F1 vs. F2 generations were
297 identified using the same method as described above for identifying DMCs between ecotypes,
298 with sequencing lane as a covariate. We identified 137 and 82 DMCs within HL and KL hybrid
299 lines, respectively. These sites were removed from the 52,940 CpG sites that passed the filtering
300 step to provide the dataset of intergenerationally stable sites. We clustered fish based on the
301 similarity of their DNA methylation profiles, with the "ward" clustering method on Pearson's
302 correlation distances, using the *clusterSamples* function in the R package methylKit. We also
303 compared the locations of DMCs between ecotypes with the locations of intergenerationally
304 stable CpG sites to assess which of the sites involved in methylation divergence between
305 ecotypes are stable across generations.

306 Our criterion for identifying CpG sites with intergenerationally stable methylation is such
307 that a type 2 error in our differential methylation test between generations (a false negative for
308 differential methylation between F1 and F2 within a line) will lead to a false positive for stable
309 methylation. To investigate the potential importance of this type of error in our data, we
310 conducted a power analysis using a simulated data set in methylKit, following Wreczycka *et al.*
311 (2017) with some minor modifications. To be conservative, our simulated dataset consisted of
312 eight samples (four F1s and four F2s, matching the minimum number of fish in a line from each
313 generation in the empirical dataset). We modelled the read coverage following a binomial
314 distribution and defined the methylation levels following a beta distribution with parameters
315 alpha = 0.4, beta= 0.5 and theta =10. We ran simulations of differential methylation at 1% of
316 52,940 CpG sites, with effect sizes of 5%, 10%, 15%, 20% and 25% differential methylation,
317 respectively. After correcting for the covariate of sequencing lane (the same sequencing lane was
318 assigned to two samples within F1 or F2 generation, for a total of four sequencing lanes), we

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

319 adjusted *P*-values for multiple testing using the *Q*-value method (Storey & Tibshirani 2003), and
320 considered CpGs to be DMCs with a false discovery rate correction *Q*-value < 0.01. Finally, we
321 calculated the proportion of CpGs that were falsely identified as non-DMCs (false negatives)
322 among all CpG sites under each effect size (5% to 25%) above.

323 We distinguished between three categories of methylated sites that were stable between F1
324 and F2 generations: 1) constitutively hypermethylated sites, which are CpG sites with average
325 DNA methylation levels greater than 0.9 in all samples, 2) constitutively hypomethylated sites,
326 which are CpG sites with average DNA methylation levels less than 0.1 in all samples (Lam *et*
327 *al.* 2012; Lea *et al.* 2016), and 3) methylated sites with average DNA methylation levels between
328 0.1 and 0.9 (hereafter referred to 'variable sites'). Finally, we analysed the proportion of
329 cytosines within genomic features for CpG sites in each category and annotated genes associated
330 with all intergenerationally stable CpG sites by testing for overlap between the locations of CpG
331 sites and genomic regions of genes, following the same method as described above for ecotype
332 DMCs.

333

334 *Heritability of stable methylation*

335 To determine the genetic heritability of the intergenerationally stable CpG sites across F1 and F2
336 generations, we first identified SNPs using the aligned reads of all F1 and F2 individuals for
337 input to Bis-SNP, following the same SNP calling and filtering steps as described above with
338 some minor modifications. We retained three sets of SNPs by filtering the 92,983 SNPs using a
339 constant MAF cut-off (0.005) and three missing data cut-offs (10%, 30%, 50%) across all F1 and
340 F2 individuals. We used BCFtools (https://github.com/samtools/bcftools) to exclude sites that
341 were under linkage disequilibrium (LD, pairwise $r^2 > 0.8$ within a window of 1Mb) or on the sex
342 chromosome. Finally, we used a linear mixed model implemented in PyLMM
343 (http://genetics.cs.ucla.edu/pylmm/index.html) to test whether variation at SNPs is significantly

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

344 associated with methylation levels at stable CpG sites in F1 and F2 generations, after correcting
345 for sequencing lane variation and kinship based on the SNP data. We adjusted multiple-test *P*-
346 values using Benjamini-Hochberg's false discovery rate and considered an association to be
347 significant when the corrected *P*-value < 0.05.

348 We estimated the narrow sense heritability of DNA methylation levels for individual CpG
349 sites of all F1 and F2 individuals using a linear mixed model approach (Yang *et al.* 2010)
350 implemented in the R package lmmlite (https://github.com/kbroman/lmmlite). We treated the
351 methylation levels at individual CpG sites of all F1 and F2 individuals as phenotypes, and
352 assumed each phenotype $\mathbf{y}$ can be modelled as $\mathbf{y} = \mathbf{1_n}\boldsymbol{\mu} + \mathbf{u} + \mathbf{e}$, where the random variable $\mathbf{u}$
353 follows a normal distribution centred at zero with variance σ_g^2K, and $\mathbf{e}$ represents an independent
354 noise component with variance σ_e^2. The matrix K is the same kinship matrix as calculated above.
355 For each trait we estimated σ_g^2 and σ_e^2 using the restricted maximum likelihood (REML)
356 approach, with correction for the covariate of sequencing lane, and calculated the heritability as
357 $h^2 = \sigma_g^2/(\sigma_g^2 + \sigma_e^2)$. Finally, we calculated the average heritability by taking the mean of
358 heritability values of all CpG sites.

359

360 *meQTL analysis*

361 To identify the genetic architecture of methylation divergence between marine and freshwater
362 stickleback, we performed meQTL mapping of the methylome in two F2 families of marine-
363 freshwater hybrids. Due to the distinct methylation patterns that may be caused by genetic
364 variation between the two hybrid lines (Fig. S2), we performed mapping separately for each
365 hybrid line. We first filtered SNPs that were not located on the sex chromosome, had less than
366 10% missing data, and had low LD in HL_F2 or KL_F2 samples, using the same SNP filtering
367 steps as described above. We then compiled a percentage methylation level matrix among
368 HL_F2 or KL_F2 samples containing the 52,940 CpG sites that passed these filtering steps.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

369 Finally, we truncated these sites by the 10% minimum range of methylation variation across
370 samples to reduce non-informative sites that could possibly inflate test statistics and create
371 spurious SNP-CpG pairs. After filtering, we retained 525 SNPs and 27,614 CpGs in HL_F2, and
372 330 SNPs and 27,039 CpGs in KL_F2 for meQTL analysis, with no overlap between the retained
373 SNPs in each line. We tested all genome-wide SNP-CpG pairs using the R package MatrixEQTL
374 v2.2 (Shabalin 2012). This package enables rapid computation of QTLs by only retaining those
375 that are significant at a pre-defined threshold. We fit an additive linear model to test if the
376 number of alleles (coded as 0, 1, 2) predicted percentage DNA methylation levels (value ranging
377 from 0 to 1) at each CpG site, including sequencing lane as a covariate. We used a Bonferroni-
378 corrected multiple-test corrected threshold, set it to genome-wide significance for GWAS and
379 divided by the number of CpG sites tested (i.e., HL_F2: $5 \times 10^{-8}/27,614 = 1.81 \times 10^{-12}$; KL_F2: 5
380 $\times 10^{-8}/27,039 = 1.85 \times 10^{-12}$). We chose this stringent threshold to call meQTLs to minimize the
381 possibility of false positives (Orozco *et al.* 2015). We calculated the distance between a SNP and
382 a CpG site within a significant meQTL and defined a SNP as *cis*-acting if the SNP was located
383 within 1Mb from its associated CpG site or *trans*-acting if the SNP was located more than 1Mb
384 from its associated CpG site (Zhang *et al.* 2014). We then performed GO analysis on genes
385 associated with unique SNPs within significant meQTLs and identified over-represented GO
386 terms, using the same method as described above. The gene pools against which we compared
387 the unique SNPs were the genes associated with the SNPs that passed the filtering step. Because
388 previous studies have suggested that meQTLs and eQTLs are likely to co-occur in close genomic
389 proximity, we compared locations of significant meQTLs in our study to significant eQTLs
390 identified in Ishikawa *et al.* (2017), which also used a marine-freshwater hybrid design. Ishikawa
391 *et al.* (2017) identified eQTLs under a range of salinities, but we only used eQTLs that they
392 identified under 3.1 ppt, which is a similar salinity level to our experimental conditions. Finally,
393 to investigate the role of meQTLs in adaptation to different habitats in stickleback, we compared

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

394 locations of unique SNPs within significant meQTLs to previously documented regions of
395 parallel genomic divergence between marine and freshwater sticklebacks (Hohenlohe *et al.* 2010;
396 Jones *et al.* 2012; Terekhanova *et al.* 2014), and identified genetic architecture of DMCs
397 between ecotypes by comparing the locations of DMCs and unique CpG sites associated with
398 significant meQTLs. In addition to performing the meQTL analyses using MatrixEQTL as
399 described above, we also validated meQTL results within each hybrid line using the R package
400 R/qtl v.1.46-2 (Broman *et al.* 2003) with default settings, except for rescaling the basepair
401 positions of SNPs by multiplying by a constant of 3.11 x 10^{-6} due to the genome-wide
402 recombination rate of 3.11cM/Mb in stickleback (Roesti *et al.* 2013). We calculated genome-
403 wide logarithm of the odds (LOD) thresholds through 1,000 permutations, using the n.perm
404 function in the R package R/qtl and set the 95th percentile LOD score as the significance
405 threshold (Hoglund *et al.* 2020).
406
407 *Data Accessibility*
408 Raw Illumina sequencing reads for the 94 analysed individuals can be downloaded from the
409 NCBI Short Read Archive (BioProject ID: PRJNA587332). The cytosine coverage files (.cov)
410 for the 94 analysed individuals and codes used for analyses in this study are available through
411 Github (https://github.com/barrettlabecoevogeno/Heritability_DNA_methylation_sticklebacks).
412 Additional supplemental material is available at figshare.
413
414 **Results**
415 *General methylation patterns*
416 To identify general methylation patterns, we performed principal component analyses (PCA) on
417 the methylation levels of filtered CpG sites represented in all samples (Fig. 2). PC1 reflected
418 sequencing lane chemistry (Fig. S2; Table S2) and so we included sequencing lane as a covariate

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

419 in all downstream analyses. When analysing all samples, PC2 (variance explained: 5.1%) clearly
420 separated parental and F2 samples, with F1 samples filling the intermediate space between
421 parental and F2 samples. Furthermore, the PCA separated the samples by sire (HL vs. KL) in F1
422 populations along PC3, which accounted for 3.2% of the variance observed in the data set (Fig.
423 2a). In the parental fish, the PCA separated samples mainly by their habitat along PC3 (Fig. 2b),
424 whereas the F1 generation shows clustering based on family (Fig. 2c). The PCA also revealed
425 some clustering between the HL and KL hybrid lines in the F2 generation, although there is no
426 clear separation between lines (Fig. 2d). Within families, we found significantly higher mean
427 DNA methylation variance in the F2 generation than the F1 generation of families in the HL line
428 ($W = 2.94 \times 10^7$; $P = 5.61 \times 10^{-6}$), but not the KL line ($W = 3.08 \times 10^7$; $P = 0.775$).
429 The average pairwise F_{st} calculated between populations was 0.03 (BI vs. HL), 0.04 (BI vs.
430 KL) and 0.01 (HL vs. KL). These values are comparable to what has been reported between
431 other marine and freshwater populations of stickleback using SNPs extracted from RAD-seq
432 (e.g., Hohenlohe *et al.* 2010; Catchen *et al.* 2013; Lescak *et al.* 2015; Garcia-Elfring *et al.*
433 Unpublished) and whole genome sequencing (Shanfelter *et al.* 2019), suggesting that our use of
434 SNPs identified from RRBS should not bias estimates of genetic differentiation relative to other
435 methods.
436
437 *Methylation divergence between ecotypes in the parental generation*
438 We identified 891 DMCs between parental fish sampled from marine vs. freshwater
439 habitats after false discovery rate correction. Based on Euclidean distances, individual fish
440 clustered by their ecotypes, with the freshwater fish further clustered by their sampling site (HL
441 vs. KL; Fig. 3a). When analysing the methylation patterns of the 891 CpGs across generations,
442 we found two major clusters, with the first cluster only including marine fish from the parental
443 generation, and the second cluster including all freshwater fish from the parental generations and

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

444 all F1 and F2 hybrids (Fig. S3). When comparing the mean methylation levels of the 891 ecotype
445 DMCs in F1s versus parents, we found that in both hybrid F1 lines, there was a significantly
446 greater number of CpG sites with methylation levels that were intermediate between the values
447 of the parents than the number of CpG sites with methylation levels outside the values observed
448 in the parents (KL_F1: $G = 294$, df = 1, $P < 2.20 \times 10^{-16}$; HL_F1: $G = 124$, df = 1, $P < 2.20 \times 10^{-16}$
449). In the F2 generation, we observed a greater proportion of sites showing methylation values
450 outside the range observed in their F1 parents relative to the pattern between F1s and their wild
451 parents (KL_F2: $G = 241$, df = 1, $P < 2.20 \times 10^{-16}$; HL_F2: $G = 9.72$, df = 1, $P = 1.82 \times 10^{-3}$). In
452 addition, we found a marginally greater proportion of sites showing bias in methylation levels
453 towards those of the mother in HL_F1 ($G = 4.10$, df = 1, $P = 0.0428$) but not KL_F1 ($G = 0.576$,
454 df = 1, $P = 0.448$) when compared to the parental generation. However, sex is confounded with
455 parental habitat in this comparison (marine fish are always female, and freshwater fish are
456 always male). This confounding effect is not present in the F1 generation, where the same
457 analysis found a marginally greater proportion of sites showing bias toward the F1 mother in
458 KL_F2 ($G = 3.85$, df = 1, $P = 0.0497$) but not HL_F2 ($G = 0.0627$, df = 1, $P = 0.802$).

459 Identified DMCs between ecotypes showed no significant bias towards hyper- versus
460 hypomethylation (430 hypermethylated and 461 hypomethylated DMCs; $G = 1.21 \times 10^{-3}$, df = 1,
461 $P = 0.972$). However, these DMCs showed significant enrichment within introns when compared
462 to the null distribution of all filtered sites (introns: $G = 8.87$, df = 1, $P = 2.90 \times 10^{-3}$; Fig. 3b). In
463 addition, we found no overlap between the locations of the 891 DMCs and the locations of sex-
464 biased DMCs identified in Metzger & Schulte (2018), suggesting that removing CpGs located on
465 the sex chromosome effectively minimized any potential sex-biased differential methylation. In
466 total, DMCs mapped to 228 genes, with some genes having been shown to be associated with
467 differential expression or methylation between ecotypes in recent studies (e.g., differentially
468 expressed genes in gill: *atp1a2a*, *g6pd*, Artemov *et al.* 2017; differentially methylated genes in

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

469 fillet: *g6pd*, *chchd3a*, Smith *et al.* 2015). GO analysis showed no significant GO term
470 enrichment.
471
472 *Intergenerationally stable methylation*
473 We found that 99.6% of CpG sites (52,729 out of 52,940) were not differentially
474 methylated across generations in both lines, suggesting the vast majority of sites show stable
475 levels of methylation across generations. Our power analysis suggests that a small proportion of
476 sites (less than 1%) are likely to have been falsely identified as non-DMCs (Fig. S4) across all
477 effect size groups, suggesting the influence of type 2 error on our criterion for calling stable
478 methylation would only effect a small number of sites. Based on Pearson's correlation distance
479 calculated from the 52,729 CpG sites, most individuals clustered by generation (F1 vs. F2) and
480 by hybrid line (HL vs. KL) (Fig. 4a).
481 We found no significant enrichment of the stable sites in any of the genomic contexts when
482 compared to the null distribution of all filtered sites (promoters: $G = 4.53 \times 10^{-4}$, df = 1, $P =$
483 0.983; exons: $G = 4.94 \times 10^{-5}$, df = 1, $P = 0.994$; introns: $G = 3.49 \times 10^{-5}$, df = 1, $P = 0.995$;
484 intergenic regions: $G = 1.01 \times 10^{-4}$, $P = 0.992$; Fig. 5a). Among the stable CpG sites, we found
485 6,462, 28,005 and 18,262 CpG sites that were constitutively hypermethylated, constitutively
486 hypomethylated, and variable, respectively. When analysing the genomic context of CpGs from
487 these three categories, we found a significantly biased genomic distribution, with constitutively
488 hypermethylated sites enriched within exons ($G = 29.3$, df = 1, $P = 6.33 \times 10^{-8}$; Fig. 5b),
489 constitutively hypomethylated sites enriched within promoters ($G = 17.8$, df = 1, $P = 2.42 \times 10^{-5}$;
490 Fig. 5c), and variable sites enriched within introns ($G = 9.36$, df = 1, $P = 2.22 \times 10^{-3}$; Fig. 5d)
491 when compared to the null distribution of all filtered CpGs. We found that 94.8% (845 out of
492 891) of the DMCs between ecotypes were also identified as stable sites, a percentage not
493 significantly different than the percentage of stable sites among all filtered sites ($G = 0.171$, df =

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

494 1, $P = 0.679$), suggesting that most sites involved in methylation divergence between ecotypes
495 can be stable across generations, and could therefore be plausibly associated with adaptation to
496 different habitats.

497 To calculate the genetic heritability of stable sites, we first identified 92,983 SNPs, and
498 then filtered this dataset down to 350 SNPs that 1) had less than 10% missing data across all F1
499 and F2 individuals, 2) were not located on the sex chromosome, and 3) had low to no LD with
500 each other. Six of these SNPs showed highly significant associations with the methylation values
501 of F1 and F2 individuals ($Q < 0.05$), and 16,514 out of the 52,729 intergenerationally stable CpG
502 sites had $h^2 > 0$ (Fig. 4b). We also retained 3,007 and 4,203 SNPs after filtering the SNPs by
503 30% and 50% missing data, respectively, with 22 and 28 of these SNPs showing highly
504 significant associations with the methylation values of F1 and F2 individuals, and 16,498 and
505 21,055 intergenerationally stable CpG sites having $h^2 > 0$ (Fig. S5). Finally, the kinship matrix
506 estimated the narrow sense heritability for CpG methylation levels at (on average) 24%, 32% and
507 35% using the 350, 3,007 and 4,203 post-filtering SNPs, respectively.

508

509 *Identification of meQTLs associated with methylation divergence between marine and freshwater*
510 *ecotypes*

511 When analyzing meQTLs within each hybrid line, we identified 968 and 531 significant
512 SNP-CpG pairs in HL_F2 and KL_F2 fish, respectively, corresponding to 335 unique SNPs and
513 75 unique CpG sites in HL_F2, and 201 unique SNPs and 72 unique CpG sites in KL_F2. We
514 found that 85.0% (HL_F2: 823 of 968) and 94.4% (KL_F2: 501 out of 531) of the SNP-CpG
515 pairs were also identified as significant SNP-CpG pairs when using R/qtl, and the *P*-value
516 distributions showed no evidence of test statistic inflation (Fig. S6). A two-dimensional plot of
517 meQTLs indicates that each SNP could regulate multiple CpG sites located across the genome
518 (*trans*-meQTLs indicated as grey dots show large scatter around the diagonal line of *cis*-

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

519 meQTLs, indicated as black dots; Fig. S7). We found meQTLs displayed significantly more
520 *trans*- than *cis*-meQTLs (HL_F2: 965 *trans*- vs. 3 *cis*-meQTLs, $G = 1.30 \times 10^3$, df = 1, $P < 2.20$
521 $\times 10^{-16}$; KL_F2: 522 *trans*- vs. 4 *cis*-meQTLs, $G = 689$, df = 1, $P < 2.20 \times 10^{-16}$), with no
522 significant GO term enrichment for genes annotated with unique SNPs within significant
523 meQTLs in either line. There was relatively low LD between SNPs and CpGs with significant *cis*
524 associations (mean $r^2 = 0.287$). We also found no overlap between CpGs associated with
525 meQTLs and the DMCs identified between ecotypes in HL_F2, likely due to the small number of
526 CpG sites associated with significant meQTLs. Alternatively, this could be due to the small
527 amount (ranging from 24% to 35%) of methylation variation explained by genetic variation,
528 suggesting that a significant proportion of CpGs are likely to be autonomous from genetic
529 variation, and thus not detectable in a meQTL analysis. We found two CpGs associated with nine
530 *trans*-meQTLs that overlapped with DMCs in KL_F2, with all nine SNP-CpG pairs verified by
531 R/qtl. Although the two CpGs did not localise within any genes, they were in close genomic
532 proximity (~15 kb) to zinc finger E-box binding homeobox 1b (*zeb1b*, Ensembl Gene ID
533 ENSGACG00000001002) and centrosomal protein 76 (*cep76*, Ensembl Gene ID
534 ENSGACG00000003686). The nine *trans*-meQTLs were annotated with four genes (Table S3).
535 To assess the co-occurrence of meQTLs and eQTLs, we compared locations of meQTLs
536 identified in our study to eQTL hotspots in Ishikawa *et al.* (2017). We found that an overall of
537 9.14% (HL_F2: 24 out of 335; KL_F2: 25 out of 201) of the unique SNPs overlapped with eQTL
538 locations across HL_F2 and KL_F2 samples. This proportion does not suggest an excess of
539 meQTL-eQTL overlap relative to null expectations that are built from all input SNPs for meQTL
540 analyses ($G = 0.245$, df = 1, $P = 0.621$). Finally, to investigate whether meQTLs might be
541 associated with divergent selection in marine vs. freshwater habitats, we examined whether the
542 unique SNPs within meQTLs overlapped with genomic regions of high differentiation between
543 the two stickleback ecotypes. We found a total of six (four in HL_F2 and two in KL_F2) unique

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

544 SNPs within high differentiation regions, corresponding to 14 (11 and 3 in HL_F2 and KL_F2,
545 respectively) *trans*-acting SNP-CpG pairs. The effect sizes (beta) of meQTLs ranged from 0.105
546 to 0.406 (median = 0.161). The six SNPs were annotated with four genes (Table S4), which
547 encode proteins likely to be relevant to marine-freshwater divergence in stickleback (e.g.,
548 sensing changes in osmoregulatory environment; see below).
549

550 **Discussion**

551 The role of DNA methylation in fundamental ecological and evolutionary processes has received
552 increased attention in recent years (Metzger & Schulte 2016; Verhoeven *et al.* 2016; Hu &
553 Barrett 2017). However, the extent to which variation in DNA methylation is stably transmitted
554 across generations, and the prevalence of *cis*- and/or *trans*-acting genetic variants in contributing
555 to methylome evolution remain poorly understood, particularly in natural animal populations.
556 We used a quantitative, single-base-resolution technique (RRBS) to measure DNA methylation
557 from fin tissue across two generations in threespine stickleback sampled from two distinct
558 environment types. A large majority (99.6%) of CpG sites were identified as being
559 intergenerationally stable, as indicated by consistent methylation levels across F1 and F2
560 generations in two hybrid lines. As a consequence, of the subset of CpG sites that also showed
561 significant divergence between marine and freshwater ecotypes in the grandparental generation,
562 a large majority (94.8%) could be classified as being stable across generations. Epigenetic
563 variation was associated with genetic variation to some extent, with a narrow sense heritability
564 ranging from 24% to 35%. These values are consistent with recent epigenome-wide association
565 studies that have found that genetic variation can explain an average of 7-34% of methylation
566 variation in animals (McRae *et al.* 2014; Orozco *et al.* 2015; Taudt *et al.* 2016; Carja *et al.*
567 2017). We found distinct patterns of genomic context between three categories of stable CpG
568 sites: constitutively hypomethylated and hypermethylated sites were predominantly located

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

569 within promoters and exons, respectively, whereas variable sites were enriched within introns.
570 We also identified meQTLs in marine-freshwater F2 hybrid lines, with some meQTLs
571 overlapping with genomic regions of high differentiation between marine and freshwater
572 ecotypes in stickleback. Finally, we identified the genetic architecture underlying two DMCs
573 between ecotypes that were also shown to have intergenerational stability in their methylation
574 levels. Overall, our study provides the first investigation of the genetic basis of stable epigenetic
575 variation in stickleback and identifies methylation differences that could be associated with local
576 adaptation in marine and freshwater ecotypes.
577
578 *Methylation divergence between ecotypes*
579 We found a similar number of differential methylation sites (891 DMCs) between marine and
580 freshwater ecotypes of threespine stickleback to a recent study (737 DMCs in Smith *et al.* 2015).
581 While we did not find any significantly enriched GO terms, some of these DMCs were annotated
582 with genes that are likely to contribute to adaptation to marine and freshwater environments. For
583 example, we found DMCs overlapped with genes related to osmoregulation (ion channel
584 activity: *trpc1*, *RYR3*, *gria3b*, *kcnq3*), metabolic process (lipid and fatty acid metabolism:
585 *elovl6l*, *scap*; glucose and carbohydrate metabolism: *g6pd*), immune response (hemopoiesis:
586 *kalrna*; myeloid cell and neutrophil differentiation: *satb1b*; erythrocyte maturation: *klf3*), and
587 catalytic activity (*alpl*, *phlpp1*, *sdr39u1*). Because the osmotic environments, parasite
588 communities, and migratory life cycles of marine and freshwater ecotypes differ (Smith *et al.*
589 2015; Huang *et al.* 2016; Artemov *et al.* 2017; Ishikawa *et al.* 2017), the differential methylation
590 of these genes suggests that the methylome could be associated with ecologically important
591 phenotypic differentiation between ecotypes.
592

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

593 *The genetic basis and functions of intergenerationally stable epigenetic variation*

594 To understand how methylation divergence between ecotypes might be involved in the process
595 of local adaptation, we next explored the stability of epigenetic variation across generations.
596 While the approach of using experimental crosses to explore stable epigenetic variation and its
597 underlying genetic basis has been widely applied in plant studies (e.g., Johannes *et al.* 2009;
598 Roux *et al.* 2011; Li *et al.* 2014), very few studies have used this type of experimental design in
599 non-model animal populations (but see Nätt *et al.* 2012; Weyrich *et al.* 2016; Weyrich *et al.*
600 2018). Examination of the genetic basis of methylation sites is valuable for exploring the
601 mechanisms that facilitate animal responses to novel environments, and for predicting the
602 likelihood that populations will be able to evolve in response to environmental change (O'Dea *et*
603 *al.* 2016). We identified 52,729 CpG sites that were not differentially methylated between F1 and
604 F2 generations in both HL and KL hybrid lines (99.6% of all CpG sites). Similar to Heckwolf *et*
605 *al.* (2020), this suggests that the majority of our analysed CpG sites have stable levels of
606 methylation across generations. These CpG sites are not necessarily heritable; it is possible that
607 methylation levels are induced to similar levels across generations due to exposure to a similar
608 environment. This pattern may also be due to non-global DNA methylation reprogramming
609 during embryogenesis in fish, which can provide greater opportunity for transmitting DNA
610 methylation from parents to the offspring (Schmitz *et al.* 2011; Skvortsova *et al.* 2018).

611 When assessing the contribution of the intergenerationally stable CpG sites to evolutionary
612 processes, we found significant enrichment of constitutively hypomethylated CpG sites within
613 promoters, and significant enrichment of constitutively hypermethylated CpG sites within gene
614 bodies, suggesting that stable DNA methylation may directly regulate gene expression and
615 facilitate alternative splicing, and thus contribute to genomic evolution by providing access to
616 alternative promoter sites and increasing the number of transcriptional opportunities and
617 phenotypes (Roberts & Gavery 2012). Consequently, mutations impacting intergenerationally

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

618 stable methylation could accelerate the exploration of phenotypic space, and therefore allow
619 populations to adapt to the changing environments more efficiently (Klironomos *et al.* 2013).
620 Furthermore, different categories of stable sites showed distinct features of genomic context,
621 with hypermethylated sites enriched within exons and constitutively hypomethylated sites
622 enriched within promoters. The distinct distribution patterns of hyper- and hypomethylated CpG
623 sites are consistent with whole genome assessments of methylation in other fish species and in
624 model animals and plants (Feng *et al.* 2010; Zemach *et al.* 2010; Long *et al.* 2013; Shao *et al.*
625 2014), and suggest a conserved role for constitutive hyper- and hypomethylation in a wide range
626 of species.

627 We found a large proportion (~95%) of DMCs between marine and freshwater ecotypes in
628 the grandparental generation also showed intergenerational stability in methylation levels,
629 suggesting that the genes associated with these DMCs could play a role in facilitating adaptation
630 to different environments. Theoretical work has suggested that environmentally responsive
631 epigenetic changes that can be transmitted to the next generation can be beneficial when the
632 effects of epigenetic variation increase both parental and offspring fitness with low cost (Herman
633 *et al.* 2014). As the functions of several DMC-associated genes identified here are relevant to
634 responses to changes in aquatic environments such as salinity, parasites, and diet, our findings
635 provide evidence for a possible adaptative mechanism in threespine stickleback whereby
636 advantageous epigenetic changes that have been triggered by environmental stimulus are
637 transmitted across generations.

638 There is substantial interest in biomedical and agricultural fields to understand the
639 contribution of genetic variation to population epigenomic variation, with a number of recent
640 genetic studies having quantified the heritable basis of population epigenomic variation in model
641 animals (Taudt *et al.* 2016). When applying a stringent missing data cut-off (10%), we found a
642 reasonably high average heritability of 24% for methylation levels across F1 and F2 generations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

643 When applying more relaxed missing data cut-offs (30% and 50%), we find heritability estimates
644 of 32% and 35%. In addition, we found that 31% to 40% of stable CpG sites had a measurable
645 genetic component (narrow-sense heritability $h^2 > 0$), a percentage similar to previous findings in
646 model species (Taudt *et al.* 2016). Together, our results suggest a plurality of mechanisms are
647 likely contributing to stable levels of methylation variation across generations, including genetic
648 control, epimutation, and exposure to past or current environmental factors.
649
650 *Contribution of meQTL to methylation divergence between ecotypes*
651 We characterised meQTLs in the F2 generation of two marine-freshwater hybrid lines, and
652 detected two CpGs associated with significant meQTLs that overlapped with DMCs between
653 marine and freshwater ecotypes, both in the KL_F2 line. The two CpGs were close to *zeb1b* and
654 *cep76*, which are key genes involved in developmental processes. Z*eb1b* has been hypothesized
655 to be a regulator of interleukin 2, which is associated with differences in parasite load of
656 stickleback inhabiting marine and freshwater environments, and consequently affects their
657 immune responses (Scharsack *et al.* 2016; Verta & Jones 2019). C*ep76* is an important paralog
658 of *CC2D2A*, which is a gene associated with development of the primary cilia, and is relevant to
659 morphological differences between Pacific lamprey populations with distinct migratory
660 behaviours (Hess *et al.* 2014). Because the morphology between marine and freshwater
661 stickleback ecotypes differs significantly (Jones *et al.* 2012), the overlap between meQTL-
662 associated CpGs and ecotype-DMCs suggests that these loci may be under divergent selection in
663 marine versus freshwater habitats.
664 Although methylation variation between marine and freshwater ecotypes can be caused by
665 both *cis*- and *trans*-regulatory changes, we found only *trans*-meQTLs within genomic regions of
666 high differentiation between ecotypes. This is interesting because we expected to detect a bias
667 towards *cis*-meQTLs due to the close genomic proximity of the SNPs and CpGs from the same

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

668 RRBS fragments. Moreover, a number of recent studies have shown greater contribution of *cis*-
669 regulatory than *trans*-regulatory genetic variants in gene expression divergence in the gill, brain
670 and liver tissues of stickleback (Ishikawa *et al.* 2017; Pritchard *et al.* 2017; Verta & Jones 2019).
671 However, predominantly *trans*-regulatory changes in gene expression have also been found in
672 the tooth plate of stickleback and *Drosophila* (McManus *et al.* 2010; Osada *et al.* 2017; Hart *et*
673 *al.* 2018). These contrasting results have been attributed to a number of factors such as inter- vs.
674 intra-specific comparison and tissue heterogeneity, where *tran*s-regulatory effects dominate in
675 intraspecific comparisons and in more heterogeneous tissue (Hart *et al.* 2018). Our findings fit
676 with these explanations in that we conducted an intraspecific comparison using heterogenous
677 caudal fin tissue that consists of epidermis, osteoblasts, dermal fibroblasts, and vascular
678 endothelium (Tu & Johnson 2011).

679 Although selection may initially favour master regulator genes that regulate distant genes
680 through *trans*-acting mechanisms during rapid adaptation, it has been suggested that different
681 evolutionary scenarios and selective contexts may alternatively favour *trans*- and *cis*-acting
682 mechanisms during intraspecific adaptive divergence (Cooper *et al.* 2003; Lemos *et al.* 2008;
683 Stern & Orgogozo 2009; Hart *et al.* 2018). In the case of marine stickleback invading freshwater
684 environments, local adaptation must often occur in the presence of ongoing gene flow (Nelson &
685 Cresko 2018). This scenario may initially favour *trans*-acting mechanisms that are less
686 susceptible to being eroded through recombination. However, as the population reaches later
687 stages of adaptation to the local environment, the advantage of responses mediated by *trans*-
688 regulatory genes may shift to favour *cis*-regulatory mechanisms, where co-evolved mutations are
689 more closely linked to each other and the genes they regulate (Verta & Jones 2019).

690 Our functional analysis identified multiple genes associated with meQTLs that are
691 located in genomic regions that have been shown to have significant differentiation between
692 marine and freshwater populations, and could therefore be relevant for local adaptation

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

693 (Hohenlohe *et al.* 2010; Jones *et al.* 2012; Terekhanova *et al.* 2014) (Table S4). For example, we
694 found genes annotated with osmosis and electrolyte transport (*KCNB2*), and skeletal and
695 fibroblast growth (*Slco5a1a*), which have also been found in previous stickleback studies
696 investigating differential gene expression or methylation in gill or fillet tissue between the
697 marine and freshwater ecotypes (Smith *et al.* 2015; Artemov *et al.* 2017). These results suggest
698 that this collection of genes might be important for facilitating adaptation to these divergent
699 environments in stickleback, although the direction of causality between DNA methylation
700 variation and gene expression remain elusive. Interestingly, it has been shown that genomic
701 regions that are not significantly differentiated between ecotypes can still play an important role
702 in adaptation to novel aquatic environments in stickleback (DeFaveri *et al.* 2011; Leinonen *et al.*
703 2012; Ellis *et al.* 2015; Erickson *et al.* 2016; Ferchaud & Hansen 2016). Our findings suggest
704 that the genetic architecture underlying methylation divergence and physiological adaptation to
705 different aquatic environments in stickleback is complicated and could include SNPs from
706 genomic regions that experience either neutral or selective processes.
707
708 *Limitations*
709 Our study has a number of caveats that should be noted. First, an intrinsic problem of *in vivo*
710 studies using next-generation sequencing techniques such as RRBS is the heterogeneity of
711 analysed tissues. Fin tissues consist of many different cell types including epidermis, osteoblasts,
712 dermal fibroblasts, and vascular endothelium (Tu & Johnson 2011). Therefore, various
713 proportions of different cell types could introduce biases in measures of methylation levels
714 (Kratochwil & Meyer 2015). In addition, we only used fin tissues from a single developmental
715 stage of sticklebacks, whereas methylation and gene expression patterns are known to be
716 development-related and tissue-specific (Wang *et al.* 2009; Feil & Fraga 2012), and thus, overlap
717 between the locations of meQTLs identified in this study and eQTLs in Ishikawa *et al.* (2017)

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

718 should be interpreted with caution. Further studies extending our work to a broader range of
719 tissues and developmental stages will be helpful for a more comprehensive characterisation of
720 methylation variation, and its role in gene regulation and development.

721 Second, the reduced representation genome sequencing method used here can only cover a
722 small proportion of all possible methylation patterns in these populations. Thus, we are
723 inevitably missing a large number of stable CpG sites and SNPs located outside of the regions of
724 the genome represented here. In addition, because the accuracy of SNP calls from bisulfite
725 sequencing data can be affected by the conversion rate of unmethylated cytosines (Barturen *et al.*
726 2014), the SNPs identified in our study could be different than those that would be obtained
727 using a sequencing method that produces independent SNP data (e.g., restriction-site associated
728 DNA sequencing, RAD-seq; Baird *et al.* 2008). Furthermore, while we mainly focused
729 methylation patterns in gene bodies, other regulatory elements such as enhancers and
730 transposons, although less well annotated in stickleback, are also important drivers of regulatory
731 and phenotypic evolution (Wittkopp & Kalay 2011) and thus warrant further research.

732 Third, the number of individuals and families used in our study is limited. Thus, including
733 additional samples from more families would provide additional information on family-level
734 variation, as well as more loci associated with methylation variation that would increase the
735 power of our heritability and meQTL analyses. In addition, we generated marine-freshwater F1
736 families by crossing marine females and freshwater males, and recent studies in zebrafish have
737 suggested that epigenetic patterns at early developmental stages can often reprogram to reflect
738 the paternal state (Jiang *et al.* 2013; Potok *et al.* 2013). Whether such reprogramming is common
739 to teleosts remains unclear (Skvortsova *et al.* 2018), but additional reciprocal crosses using
740 marine males and freshwater females, as well as pure crosses within marine or freshwater
741 populations, would allow a more comprehensive understanding of parental effects on epigenetic
742 inheritance (Laporte *et al.* 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

743 Fourth, while we found *cis* associations between SNPs and CpGs, this does not necessarily
744 indicate that the CpGs are under genetic control. Although average LD between SNPs and CpG
745 sites with significant *cis* associations was relatively low, it is still possible that patterns are driven
746 in part by linkage disequilibrium between epigenetic variation at a locus and its proximal SNP
747 (Taudt *et al.* 2016; Heckwolf *et al.* 2020). It is also possible that the CpGs are autonomous from
748 genetic control and contribute to heritable variation that is shaped by natural selection, and thus
749 will be indistinguishable from genetic variation in a standard heritability analysis (Johannes *et al.*
750 2008; Helanterä & Uller 2010; Tal *et al.* 2010). Thus, the results of our stable methylation and
751 meQTL analyses should be interpreted with consideration of these alternatives.
752 Finally, although we have corrected for the possibility of falsely interpreting C-to-T and G-
753 to-A SNPs as epigenetic variation by excluding them from methylation estimates, it is possible
754 that some SNPs were miscalled. Thus, our results provide a necessarily coarse map of the genetic
755 architecture underlying stable methylation and methylation divergence between marine and
756 freshwater stickleback populations. A wider investigation of regulatory elements in combination
757 with genome-wide sequencing of chromatin modifications (e.g., chromatin immunoprecipitation
758 followed by sequencing (ChIP-seq); Park 2009; Furey 2012) and whole-genome resequencing
759 (e.g., Le Luyer *et al.* 2017) would provide a more comprehensive and precise understanding of
760 the relationship between genetics and DNA methylation, and the role that epigenetic responses
761 may play in facilitating evolutionary change.
762
763 **Conclusions**
764 Here, we provide the first insights into the genetic architecture of DNA methylation in threespine
765 stickleback. Our genome-wide methylation data reveals that the vast majority of CpG sites have
766 stable methylation levels across generations, including the sites that show significant divergence
767 in methylation levels between marine and freshwater ecotypes. Some of these sites show

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

768 evidence of genetic control, while others are likely to be autonomous from genetic variation. We
769 also explored the genomic distribution of methylation in marine-freshwater hybrid populations
770 and found meQTLs that overlap with previously identified genomic regions of high
771 differentiation between marine and freshwater populations. In addition, our data demonstrates
772 different contributions of *cis*- and *trans*-meQTLs to methylome divergence in stickleback. Our
773 study adds to the few studies using non-model, outbred vertebrates to test for the genetic basis of
774 intergenerationally stable methylation and methylation divergence between ecotypes. Our results
775 suggest that methylation could play an important role in facilitating phenotypic plasticity over
776 the short-term, as well as population persistence and adaptation over longer evolutionary time
777 scales.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

778 **Acknowledgements**

779 The authors would like to thank Antoine Paccard and Caroline LeBlond for their assistance in
780 sample preparation, and Paul Gugger for his guidance in performing association analysis. This
781 work was supported by China Scholarship Council Fellowship 201406350023 and Start-up
782 research funding from Fudan University to JH, Queen Elizabeth II Scholarships to SJSW and
783 TNB, Alberta Innovates to TNB, a National Sciences and Engineering Research Council
784 (NSERC) RT735278 to SMR, a NSERC Discovery Grant 418249-2012 to HAJ, and a NSERC
785 Discovery Grant 429955-2013 and Canada Research Chair to RDHB. We would like to
786 acknowledge the Bamfield Marine Sciences Centre for resources in the collection of fishes and
787 during the writing of this manuscript.

788

789 **Author Contributions**

790 JH and RDHB conceived the study. SJSW, TB, HAJ, and SMR sampled, crossed and reared fish,
791 and collected tissue. JH generated and analysed sequencing data. JH wrote the manuscript with
792 input from SMR, and RDHB. The authors have no conflicts of interest.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

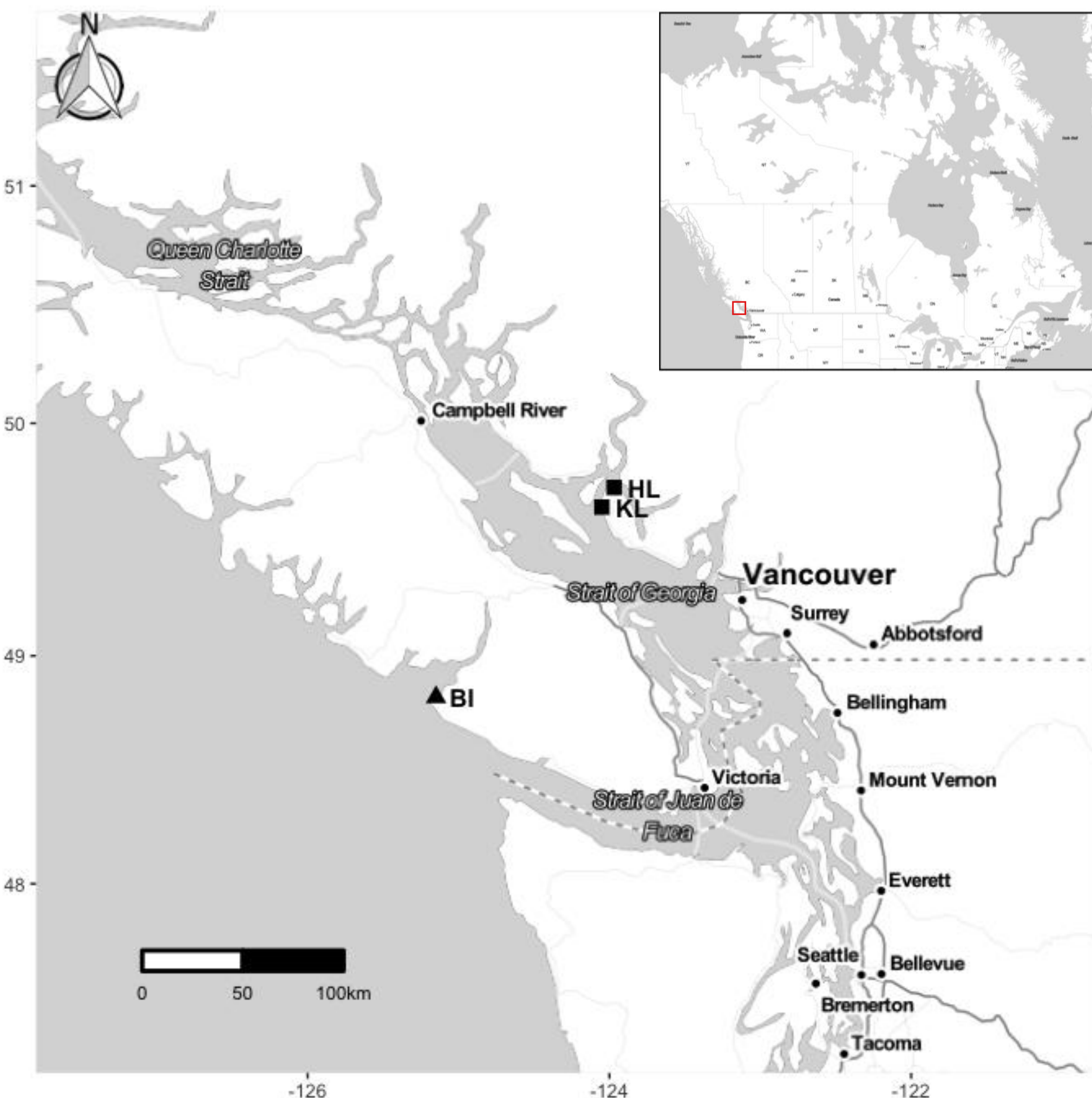

Fig. 1 Geographical location of threespine stickleback populations used in this experiment. Triangle indicates the marine sampling site, squares indicate freshwater sampling sites. BI, Bamfield Inlet (marine); HL, Hotel Lake (freshwater); KL, Klein Lake (freshwater). The red square in the inset shows the location of sampling sites in relation to the broader geographic region (the west coast of Canada).

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

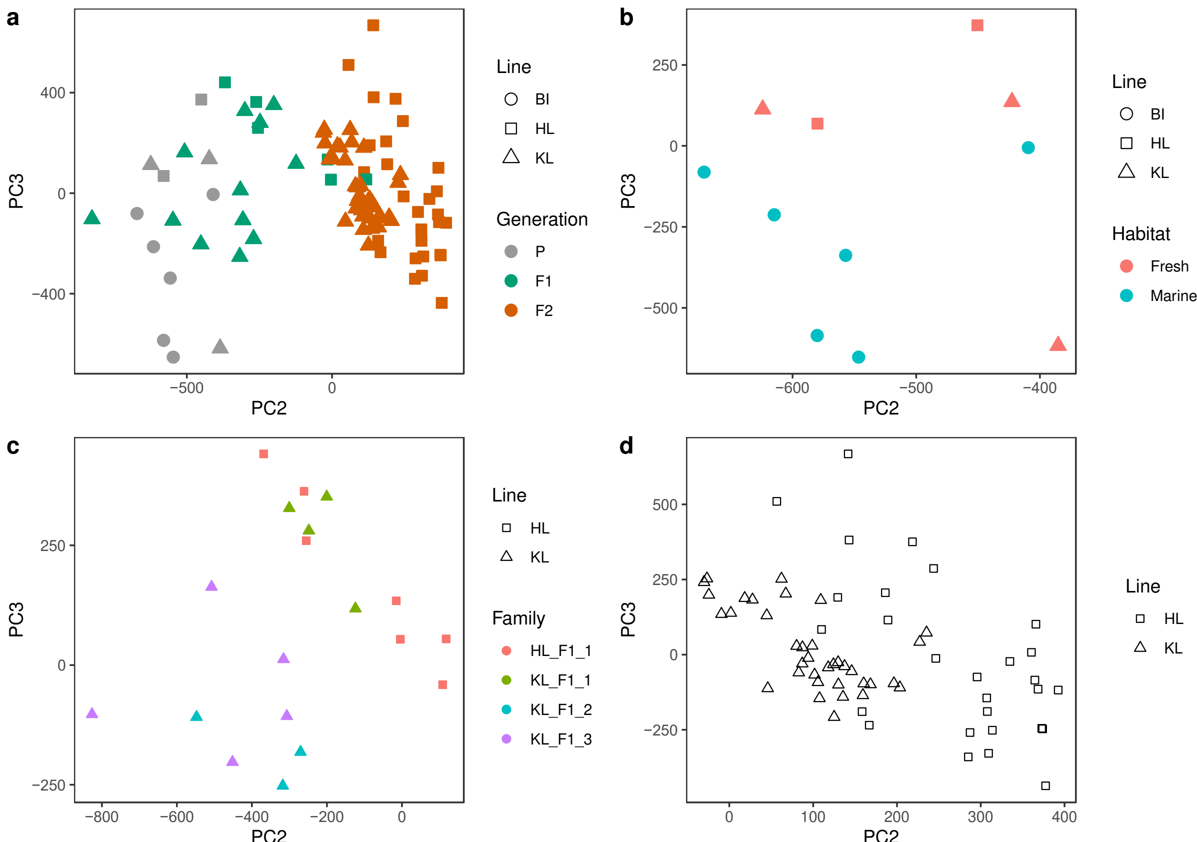

Fig. 2 Principal component analysis (PCA) of DNA methylation profiles based on all CpG sites after filtering (See Methods) in a) all individuals from parental, F1 and F2 generation, b) parental generation, c) F1 generation, and d) F2 generation. Line: Sampling site of parental fish in generation P, parental sire of fish in the F1 generation, and grandparental sire of fish from the F2 generation.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

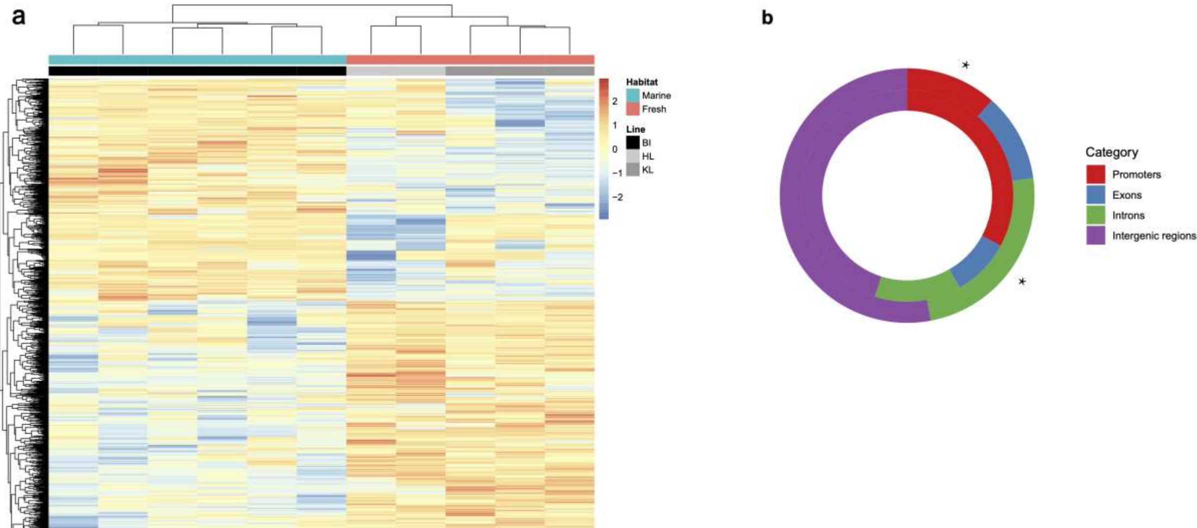

Fig. 3 (a) Heatmap of methylation levels of the 891 DMCs between marine and freshwater ecotypes from the parental generation. Each column represents a colour-coded individual: blue for marine fish, red for freshwater fish; black for marine fish from BI, light grey for freshwater fish from HL, and dark grey for freshwater fish from KL. Each row represents one of the DMCs, which are clustered based on the similarities of the methylation patterns between individuals. Darker red indicates greater methylation in an individual for that DMC. Darker blue indicates lower methylation in an individual for that DMC. Individual dendrogram positions are based on overall methylation patterns across the 891 DMCs. (b) The proportion of genomic features (promoters, exons, introns or intergenic regions) in the 891 DMCs. Outer rings describe the locations of DMCs in each category; inner rings describe the features of null distribution of all filtered CpGs. Asterisks denote significant differences between the features of DMCs in each category vs. the features of null distribution of CpGs across the genome using a G test at $P <$ 0.01.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

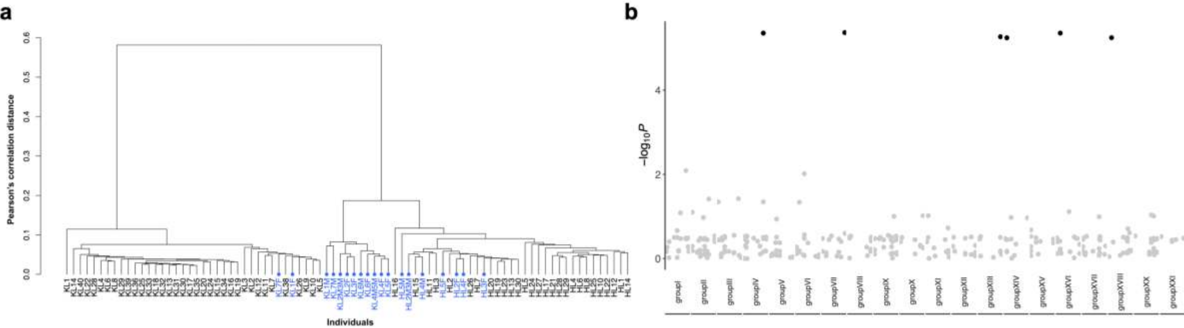

Fig. 4 (a) Dendrogram of methylation levels for all fish in F1 and F2 generations. The y-axis is the Pearson's correlation distance after hierarchical clustering of the percent methylation levels of the 52,729 intergenerationally stable CpG sites. F1 fish are shown in blue, and F2 fish are shown in black. (b) Manhattan plot showing the -log*P* of correlations between each single nucleotide polymorphism (SNPs) (columns) and the 52,729 intergenerationally stable CpG sites when filtering SNPs using a 10% missing data cut-off. Black points are statistically significant SNPs ($Q < 0.05$) after adjusting for multiple testing using the Benjamini-Hochberg's false discovery rate method. SNPs (n = 43) from unassembled scaffolds are without significant hits, and thus are not shown here.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

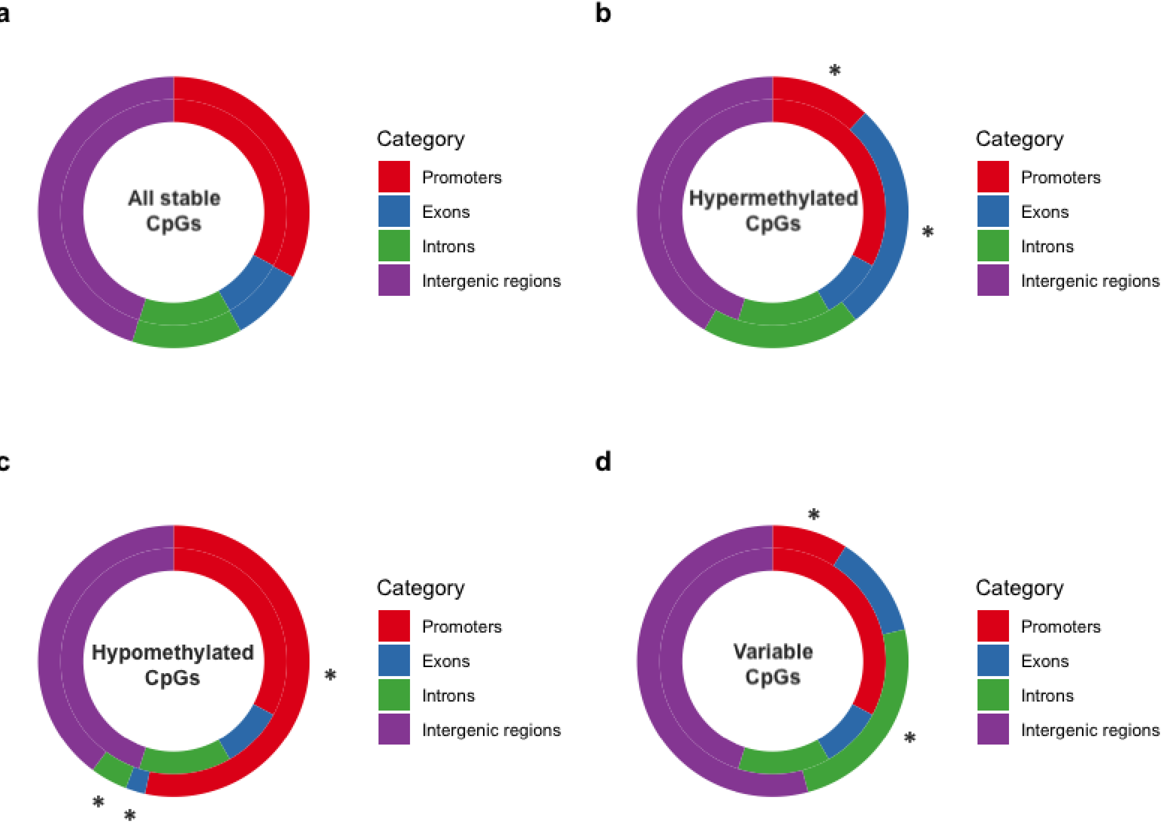

Fig. 5 The proportion of genomic features (promoters, exons, introns or intergenic regions) in intergenerationally stable CpGs compared with null distribution of all filtered CpGs in (a) all intergenerationally stable sites, (b) constitutively hypermethylated sites, (c) constitutively hypomethylated sites, and (d) variable sites. Outer rings describe the locations of CpGs in each category; inner rings describe the features of null distribution of all filtered CpGs. Asterisks denote significant differences between the features of CpGs in each category vs. the features of null distribution of CpGs across the genome using a G test at $P < 0.01$.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

Akalin A, Franke V, Vlahoviček K, Mason CE, Schübeler D (2015) genomation: a toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics*, **31**, 1127-1129.

Akalin A, Kormaksson M, Li S *et al.* (2012) methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, **13**, R87.

Alexa A, Rahnenführer J, Lengauer T (2006) Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics*, **22**, 1600-1607.

Artemov AV, Mugue NS, Rastorguev SM *et al.* (2017) Genome-wide DNA methylation profiling reveals epigenetic adaptation of stickleback to marine and freshwater conditions. *Molecular Biology and Evolution*, **34**, 2203-2213.

Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barturen G, Rueda A, Oliver JL, Hackenberg M (2014) MethylExtract: High-Quality methylation maps and SNV calling from whole genome bisulfite sequencing data. *F1000Research*, **2**, 217.

Bell MA, Foster SA (1994) *The evolutionary biology of the threespine stickleback.* Oxford University Press, New York.

Bird A (2007) Perceptions of epigenetics. *Nature*, **447**, 396-398.

Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. *Ecology Letters*, **11**, 106-115.

Boyle P, Clement K, Gu H *et al.* (2012) Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biology*, **13**, R92.

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. *Bioinformatics*, **19**, 889-890.

Carja O, MacIsaac JL, Mah SM *et al.* (2017) Worldwide patterns of human epigenetic variation. *Nature Ecology & Evolution*, **1**, 1577-1583.

Catchen J, Bassham S, Wilson T *et al.* (2013) The population structure and recent colonization history of Oregon threespine stickleback determined using restriction-site associated DNA-sequencing. *Molecular Ecology*, **22**, 2864-2883.

Cooper TF, Rozen DE, Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in Escherichiacoli. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 1072-1077.

Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156-2158.

Daxinger L, Whitelaw E (2012) Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nature Reviews Genetics*, **13**, 153-162.

DeFaveri J, Shikano T, Shimada Y, Goto A, Merila J (2011) Global analysis of genes involved in freshwater adaptation in threespine sticklebacks (*Gasterosteus aculeatus*). *Evolution*, **65**, 1800-1807.

Do C, Shearer A, Suzuki M *et al.* (2017) Genetic–epigenetic interactions in *cis*: a major focus in the post-GWAS era. *Genome Biology*, **18**, 120.

Dubin MJ, Zhang P, Meng D *et al.* (2015) DNA methylation in *Arabidopsis* has a genetic basis and shows evidence of local adaptation. *Elife*, **4**, e05255.

Durinck S, Moreau Y, Kasprzyk A *et al.* (2005) BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**, 3439-3440.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Durinck S, Spellman PT, Birney E, Huber W (2009) Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, **4**, 1184-1191.

Ellis NA, Glazer AM, Donde NN *et al.* (2015) Distinct developmental genetic mechanisms underlie convergently evolved tooth gain in sticklebacks. *Development*, **142**, 2442-2451.

Erickson PA, Glazer AM, Killingbeck EE *et al.* (2016) Partially repeatable genetic basis of benthic adaptation in threespine sticklebacks. *Evolution*, **70**, 887-902.

Fan Y, Vilgalys TP, Sun S *et al.* (2019) IMAGE: high-powered detection of genetic effects on DNA methylation using integrated methylation QTL mapping and allele-specific analysis. *Genome Biology*, **20**, 220.

Feil R, Fraga MF (2012) Epigenetics and the environment: emerging patterns and implications. *Nature Reviews Genetics*, **13**, 97-109.

Feng S, Cokus SJ, Zhang X *et al.* (2010) Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 8689-8694.

Ferchaud A-L, Hansen MM (2016) The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: three-spine sticklebacks in divergent environments. *Molecular Ecology*, **25**, 238-259.

Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nature Reviews Genetics*, **13**, 840-852.

Gu H, Smith ZD, Bock C *et al.* (2011) Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols*, **6**, 468-481.

Gupta R, van Dongen J, Fu Y *et al.* (2019) Epigenome-wide association study of serum cotinine in current smokers reveals novel genetically driven loci. *Clinical Epigenetics*, **11**, 1.

Hannon E, Gorrie-Stone TJ, Smart MC *et al.* (2018) Leveraging DNA-methylation quantitative-trait loci to characterize the relationship between methylomic variation, gene expression, and complex traits. *The American Journal of Human Genetics*, **103**, 654-665.

Hart JC, Ellis NA, Eisen MB, Miller CT (2018) Convergent evolution of gene expression in two high-toothed stickleback populations. *PLoS Genetics*, **14**, e1007443-e1007443.

Heard E, Martienssen RA (2014) Transgenerational epigenetic inheritance: myths and mechanisms. *Cell*, **157**, 95-109.

Heckwolf MJ, Meyer BS, Häsler R *et al.* (2020) Two different epigenetic information channels in wild three-spined sticklebacks are involved in salinity adaptation. *Science advances*, **6**, eaaz1138.

Helanterä H, Uller T (2010) The Price equation and extended inheritance. *Philosophy and Theory in Biology*, **2**, 1-17.

Herman JJ, Spencer HG, Donohue K, Sultan SE (2014) How stable 'should'epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution*, **68**, 632-643.

Hess JE, Caudill CC, Keefer ML *et al.* (2014) Genes predict long distance migration and large body size in a migratory fish, Pacific lamprey. *Evol Appl*, **7**, 1192-1208.

Hoglund A, Henriksen R, Fogelholm J *et al.* (2020) The methylation landscape and its role in domestication and gene regulation in the chicken. *Nature Ecology & Evolution*.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Hu J, Askary AM, Thurman TJ *et al.* (2019) The epigenetic signature of colonizing new environments in *Anolis* lizards. *Molecular Biology and Evolution*, **36**, 2165-2170.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Hu J, Barrett RDH (2017) Epigenetics in natural animal populations. *Journal of Evolutionary Biology*, **30**, 1612-1632.

Hu J, Pérez-Jvostov F, Blondel L, Barrett RDH (2018) Genome-wide DNA methylation signatures of infection status in Trinidadian guppies (*Poecilia reticulata*). *Molecular Ecology*, **27**, 3087-3102.

Huang Y, Chain FJ, Panchal M *et al.* (2016) Transcriptome profiling of immune tissues reveals habitat-specific gene expression between lake and river sticklebacks. *Molecular Ecology*, **25**, 943-958.

Ishikawa A, Kusakabe M, Yoshida K *et al.* (2017) Different contributions of local- and distant-regulatory changes to transcriptome divergence between stickleback ecotypes. *Evolution*, **71**, 565-581.

Jablonka E, Raz G (2009) Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution. *Quarterly Review of Biology*, **84**, 131-176.

Jaenisch R, Bird A (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, **33**, 245-254.

Jiang L, Zhang J, Wang J-J *et al.* (2013) Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell*, **153**, 773-784.

Johannes F, Colot V, Jansen RC (2008) Epigenome dynamics: a quantitative genetics perspective. *Nature Reviews Genetics*, **9**, 883-890.

Johannes F, Porcher E, Teixeira FK *et al.* (2009) Assessing the impact of transgenerational epigenetic variation on complex traits. *PLoS Genetics*, **5**, e1000530.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55-61.

Jones PA (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics*, **13**, 484-492.

Kawakatsu T, Huang SS, Jupe F *et al.* (2016) Epigenomic diversity in a global collection of *Arabidopsis thaliana* accessions. *Cell*, **166**, 492-505.

Klironomos FD, Berg J, Collins S (2013) How epigenetic mutations can affect genetic evolution: model and mechanism. *BioEssays*, **35**, 571-578.

Kratochwil CF, Meyer A (2015) Mapping active promoters by ChIP-seq profiling of H3K4me3 in cichlid fish - a first step to uncover *cis*-regulatory elements in ecological model teleosts. *Molecular Ecology Resources*, **15**, 761-771.

Krueger F, Andrews S (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, **27**, 1571-1572.

Krueger F, Kreck B, Franke A, Andrews S (2012) DNA methylome analysis using short bisulfite sequencing data. *Nat Methods*, **9**, 145 - 151.

Lam LL, Emberly E, Fraser HB *et al.* (2012) Factors underlying variable DNA methylation in a human community cohort. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 17253-17260.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**, 357-359.

Laporte M, Le Luyer J, Rougeux C *et al.* (2019) DNA methylation reprogramming, TE derepression, and postzygotic isolation of nascent animal species. *Science advances*, **5**, eaaw1644.

Lawrence M, Huber W, Pages H *et al.* (2013) Software for computing and annotating genomic ranges. *PLoS Computational Biology*, **9**, e1003118.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Le Luyer J, Laporte M, Beacham TD *et al.* (2017) Parallel epigenetic modifications induced by hatchery rearing in a Pacific salmon. *Proceedings of the National Academy of Sciences of the United States of America*, **114**, 12964-12969.

Lea AJ, Altmann J, Alberts SC, Tung J (2016) Resource base influences genome-wide DNA methylation levels in wild baboons (*Papio cynocephalus*). *Molecular Ecology*, **25**, 1681-1696.

Leinonen T, McCairns RJ, Herczeg G, Merila J (2012) Multiple evolutionary pathways to decreased lateral plate coverage in freshwater threespine sticklebacks. *Evolution*, **66**, 3866-3875.

Lemos B, Araripe LO, Fontanillas P, Hartl DL (2008) Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 14471-14476.

Lescak EA, Bassham SL, Catchen J *et al.* (2015) Evolution of stickleback in 50 years on earthquake-uplifted islands. *Proc Natl Acad Sci U S A*, **112**, E7204-7212.

Li Q, Eichten SR, Hermanson PJ, Springer NM (2014) Inheritance patterns and stability of DNA methylation variation in maize near-isogenic lines. *Genetics*, **196**, 667-676.

Liebl AL, Schrey AW, Richards CL, Martin LB (2013) Patterns of DNA methylation throughout a range expansion of an introduced songbird. *Integrative and Comparative Biology*, **53**, 351-358.

Lim JP, Brunet A (2013) Bridging the transgenerational gap with epigenetic memory. *Trends in Genetics*, **29**, 176-186.

Liu Y, Siegmund KD, Laird PW, Berman BP (2012) Bis-SNP: combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biology*, **13**, R61.

Long HK, Sims D, Heger A *et al.* (2013) Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. *Elife*, **2**, e00348.

Massicotte R, Whitelaw E, Angers B (2011) DNA methylation: A source of random variation in natural populations. *Epigenetics*, **6**, 421-427.

Maunakea AK, Nagarajan RP, Bilenky M *et al.* (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature*, **466**, 253-257.

McCairns RJ, Bernatchez L (2010) Adaptive divergence between freshwater and marine sticklebacks: insights into the role of phenotypic plasticity from an integrated analysis of candidate gene expression. *Evolution*, **64**, 1029-1047.

McManus CJ, Coolon JD, Duff MO *et al.* (2010) Regulatory divergence in Drosophila revealed by mRNA-seq. *Genome Research*, **20**, 816-825.

McRae AF, Powell JE, Henders AK *et al.* (2014) Contribution of genetic variation to transgenerational inheritance of DNA methylation. *Genome Biology*, **15**, R73.

Meissner A, Mikkelsen T, Gu H *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, **454**, 766 - 770.

Meng D, Dubin M, Zhang P *et al.* (2016) Limited contribution of DNA methylation variation to expression regulation in *Arabidopsis thaliana*. *PLoS Genetics*, **12**, e1006141.

Metzger DC, Schulte PM (2016) Epigenomics in marine fishes. *Marine Genomics*, **30**, 43-54.

Metzger DCH, Schulte PM (2017) Persistent and plastic effects of temperature on DNA methylation across the genome of threespine stickleback (*Gasterosteus aculeatus*). *Proceedings of the Royal Society B: Biological Sciences*, **284**, 20171667.

Metzger DCH, Schulte PM (2018) The DNA methylation landscape of stickleback reveals patterns of sex chromosome evolution and effects of environmental salinity. *Genome Biology and Evolution*, **10**, 775-785.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Morgan HD, Sutherland HG, Martin DI, Whitelaw E (1999) Epigenetic inheritance at the agouti locus in the mouse. *Nature Genetics*, **23**, 314-318.

Morris MR, Richard R, Leder EH *et al.* (2014) Gene expression plasticity evolves in response to colonization of freshwater lakes in threespine stickleback. *Mol Ecol*, **23**, 3226-3240.

Nätt D, Rubin C-J, Wright D *et al.* (2012) Heritable genome-wide variation of gene expression and promoter methylation between wild and domesticated chickens. *BMC Genomics*, **13**, 59.

Nelson TC, Cresko WA (2018) Ancient genomic variation underlies repeated ecological adaptation in young stickleback populations. *Evolution Letters*, **2**, 9-21.

Neri F, Rapelli S, Krepelova A *et al.* (2017) Intragenic DNA methylation prevents spurious transcription initiation. *Nature*, **543**, 72.

O'Dea RE, Noble DWA, Johnson SL, Hesselson D, Nakagawa S (2016) The role of non-genetic inheritance in evolutionary rescue: epigenetic buffering, heritable bet hedging and epigenetic traps. *Environmental Epigenetics*, **2**, dvv014.

Orozco LD, Morselli M, Rubbi L *et al.* (2015) Epigenome-wide association of liver methylation patterns and complex metabolic traits in mice. *Cell Metabolism*, **21**, 905-917.

Osada N, Miyagi R, Takahashi A (2017) Cis-and trans-regulatory effects on gene expression in a natural population of Drosophila melanogaster. *Genetics*, **206**, 2139-2148.

Park PJ (2009) ChIP-Seq: advantages and challenges of a maturing technology. *Nature Reviews Genetics*, **10**, 669-680.

Peichel CL, Marques DA (2017) The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, **372**, 20150486.

Peichel CL, Ross JA, Matson CK *et al.* (2004) The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Current Biology*, **14**, 1416-1424.

Potok ME, Nix DA, Parnell TJ, Cairns BR (2013) Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell*, **153**, 759-772.

Pritchard VL, Viitaniemi HM, McCairns RJ *et al.* (2017) Regulatory architecture of gene expression variation in the threespine stickleback *Gasterosteus aculeatus*. *G3*, **7**, 165-178.

Rakyan VK, Chong S, Champ ME *et al.* (2003) Transgenerational inheritance of epigenetic states at the murine Axin(Fu) allele occurs after maternal and paternal transmission. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 2538.

Richards CL, Alonso C, Becker C *et al.* (2017) Ecological plant epigenetics: Evidence from model and non-model species, and the way forward. *Ecology Letters*, 1576-1590.

Roberts SB, Gavery MR (2012) Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Frontiers in Physiology*, **2**, 116.

Roesti M, Moser D, Berner D (2013) Recombination in the threespine stickleback genome—patterns and consequences. *Molecular Ecology*, **22**, 3014-3027.

Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research*, **22**, 939-946.

Roux F, Colomé-Tatché M, Edelist C *et al.* (2011) Genome-wide epigenetic perturbation jump-starts patterns of heritable variation found in nature. *Genetics*, **188**, 1015-1017.

Scharsack JP, Franke F, Erin NI *et al.* (2016) Effects of environmental variation on host-parasite interaction in three-spined sticklebacks (*Gasterosteus aculeatus*). *Zoology*, **119**, 375-383.

Schmitz RJ, Schultz MD, Lewsey MG *et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, **334**, 369-373.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Schrey AW, Alvarez M, Foust CM *et al.* (2013) Ecological epigenetics: Beyond MS-AFLP. *Integrative and Comparative Biology*, **53**, 340-350.
Schulz H, Ruppert A-K, Herms S *et al.* (2017) Genome-wide mapping of genetic determinants influencing DNA methylation and gene expression in human hippocampus. *Nature Communications*, **8**, 1511.
Shabalin AA (2012) Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics*, **28**, 1353-1358.
Shanfelter AF, Archambeault SL, White MA (2019) Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biology and Evolution*, **11**, 1573-1585.
Shao C, Li Q, Chen S *et al.* (2014) Epigenetic modification and inheritance in sexual reversal of fish. *Genome Research*, **24**, 604-615.
Skvortsova K, Iovino N, Bogdanović O (2018) Functions and mechanisms of epigenetic inheritance in animals. *Nature Reviews Molecular Cell Biology*, **19**, 774-790.
Smith G, Smith C, Kenny JG, Chaudhuri RR, Ritchie MG (2015) Genome-wide DNA methylation patterns in wild samples of two morphotypes of threespine stickleback (*Gasterosteus aculeatus*). *Molecular Biology and Evolution*, **32**, 888-895.
Stern DL, Orgogozo V (2009) Is genetic evolution predictable? *Science*, **323**, 746-751.
Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9440-9445.
Stricker SH, Koferle A, Beck S (2017) From profiles to function in epigenomics. *Nature Reviews Genetics*, **18**, 51-66.
Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, **9**, 465-476.
Tal O, Kisdi E, Jablonka E (2010) Epigenetic contribution to covariance between relatives. *Genetics*, **184**, 1037-1050.
Taudt A, Colome-Tatche M, Johannes F (2016) Genetic sources of population epigenomic variation. *Nature Reviews Genetics*, **17**, 319-332.
Terekhanova NV, Logacheva MD, Penin AA *et al.* (2014) Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genetics*, **10**, e1004696.
Tu S, Johnson Stephen L (2011) Fate restriction in the growing and regenerating zebrafish fin. *Developmental Cell*, **20**, 725-732.
Turck F, Coupland G (2014) Natural variation in epigenetic gene regulation and its effects on plant developmental traits. *Evolution*, **68**, 620-631.
Verhoeven KJF, Preite V (2014) Epigenetic variation in asexually reproducing organisms. *Evolution*, **68**, 644-655.
Verhoeven KJF, vonHoldt BM, Sork VL (2016) Epigenetics in ecology and evolution: what we know and what we need to know. *Molecular Ecology*, **25**, 1631-1638.
Verta J-P, Jones FC (2019) Predominance of *cis*-regulatory changes in parallel expression divergence of sticklebacks. *Elife*, **8**, e43785.
Walker JA (1997) Ecological morphology of lacustrine threespine stickleback *Gasterosteus aculeatus* L. (Gasterosteidae) body shape. *Biological Journal of the Linnean Society*, **61**, 3-50.
Wang G, Yang E, Smith KJ *et al.* (2014) Gene expression responses of threespine stickleback to salinity: implications for salt-sensitive hypertension. *Frontiers in Genetics*, **5**, 312.
Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, **10**, 57-63.

bioRxiv preprint doi: https://doi.org/10.1101/2020.11.26.400531; this version posted November 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Weyrich A, Jeschek M, Schrapers KT *et al.* (2018) Diet changes alter paternally inherited epigenetic pattern in male Wild guinea pigs. *Environmental Epigenetics*, **4**, dvy011-dvy011.

Weyrich A, Lenz D, Jeschek M *et al.* (2016) Paternal intergenerational epigenetic response to heat exposure in male Wild guinea pigs. *Molecular Ecology*, **25**, 1729-1740.

Wittkopp PJ, Kalay G (2011) *Cis*-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nature Reviews Genetics*, **13**, 59.

Wreczycka K, Gosdschan A, Yusuf D *et al.* (2017) Strategies for analyzing bisulfite sequencing data. *Journal of Biotechnology*, **261**, 105-115.

Yang J, Benyamin B, McEvoy BP *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, **42**, 565.

Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, **328**, 916-919.

Zhang H, Wang F, Kranzler HR *et al.* (2014) Identification of methylation quantitative trait loci (mQTLs) influencing promoter DNA methylation of alcohol dependence risk genes. *Human Genetics*, **133**, 1093-1104.

Zhu LJ (2013) Integrative analysis of ChIP-Chip and ChIP-Seq dataset. In: *Tiling Arrays: Methods and Protocols* (eds. Lee T-L, Shui Luk AC), pp. 105-124. Humana Press, Totowa, NJ.

Zhu LJ, Gazin C, Lawson ND *et al.* (2010) ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, **11**, 237.