Methylation patterns reveal cryptic structure and a pathway for adaptation in a panmictic carnivore

Justin B. Johnson*[1], Dennis L. Murray[1,2], Aaron B.A. Shafer[1,3]

1. Environmental and Life Sciences Graduate Program, Trent University, K9J 7B8 Peterborough, Canada
2. Biology Department, Trent University, K9J 7B8 Peterborough, Canada
3. Forensics Department, Trent University, K9J 7B8 Peterborough, Canada

***corresponding author**

Summary

Determining molecular signatures of population divergence is a fundamental component of evolutionary biology. Identifying divergence is particularly challenging between populations of highly mobile species that undergo substantial gene flow, such as the Canada lynx (*Lynx canadensis*), where populations are considered panmictic when only neutral genetic markers are considered. Here, we used high-throughput bisulfite sequencing to examine the environmental determinants of methylation structure across the distributional range of Canada lynx. Despite a high degree of genetic similarity among mainland populations, epigenetic structure did not mirror these neutral patterns, instead showing environmental associations and high levels of divergence in the peripheral Newfoundland and Alaskan populations. Interestingly, a disproportionate number of genes related to body-size were hypermethylated on the island of Newfoundland, providing a possible mechanism for adaptive evolution and the observed island effect on organism size. Our results indicate that epigenetic modifications, specifically DNA methylation, are powerful markers to investigate adaptive divergence and rapid evolutionary response.

Background

Investigations into divergent selection have often relied on quantifying phenotypic variation and the heritability of such traits, the latter assumed to be genetic polymorphisms transmitted via an organism's DNA[1]. However, detecting adaptive divergence in species that experience high rates of gene flow is

25 challenging due to the homogenization of genomic regions that are neutral or under weak selection[2].
26 Environmental conditions can be a powerful driver of adaptive divergence, where relationships are
27 traditionally ascertained by correlating allele frequencies to environmental variation[3]. An aspect of
28 adaptive molecular evolution that is undetected by standard genetic sequencing involves direct
29 modifications to the structure of DNA. Epigenetic modifications like DNA methylation are influenced by
30 environmental conditions, directly affect gene expression, and may be indicative of early adaptive
31 divergence due to local adaptation[4–6]. DNA methylation could play a role in local adaptation due to its
32 regulatory role in transcription by modifying chromatin structure, repressing transcription factors, or
33 recruiting protein complexes that block transcriptional machinery, especially around CpG islands[7–9]. CpG
34 islands are dense clusters of cytosine-guanine dinucleotides and frequently occur near the transcription
35 start site of genes and have functional role with gene expression[10–12]. Consequently, DNA methylation,
36 especially around CpG islands, could explain the molecular basis of local adaptation due to its regulatory
37 function on gene expression, and might be an overlooked molecular marker of adaptive divergence and
38 rapid evolutionary adaptation.

39 Here, we assessed whether environmental variation, geographic distance, or insularity were
40 determinants of DNA methylation structure in a free-ranging carnivore. Our study species, the Canada
41 lynx, (*Lynx canadensis*) is a mid-sized felid that is highly mobile and whose neutral genetic variation (i.e.,
42 microsatellites) exhibits low levels of genetic differentiation across the mainland, with divergent island
43 populations[13,14]. Despite this low degree of overall genetic structure, two potential mechanisms might
44 drive epigenetic divergence in Canada lynx. First, allele frequencies are correlated to climatic gradients
45 in both population time-series and fine-scale genetic analyses[13,15,16], suggesting climate might also
46 influence patterns of methylation in lynx. Second, Canada lynx show a subtle cline in body size with
47 larger individuals in Alaska[17] to smaller individuals in insular populations including Newfoundland and
48 Cape Breton Island[18]. Body size changes in island populations appear to be consistent with the 'Island

49 Rule'[18], where insular mammals are smaller in size compared to their mainland counterparts[19–22]. If
50 functional genes related to body size are repressed in geographically isolated populations, then
51 epigenetic modifications might be an underlying mechanism driving population divergence[23]. Based on
52 these two mechanisms (climate and island isolation), we tested two predictions: i) climatic conditions
53 are the driving force behind spatial epigenetic structuring[24]; and ii) patterns of DNA methylation over
54 CpG islands and gene bodies are more correlated with biogeographical substructuring than methylation
55 patterns over DNA of unknown function, due to the regulation of transcription arising from adaptation
56 to local conditions[25].

57 **Results**

58 We identified differential methylation at base-pair resolution across the genome in 95 Canada lynx
59 epidermal tissue samples from four populations (n = 23-24 per sampling area) across North America,
60 including one insular population in Newfoundland (Figure 1; Supplemental Table S4). The sampled
61 populations have a wide geographical spread, with an average minimum distance between populations
62 (Québec and Newfoundland) of 1,158 km and a maximum distance (Alaska and Newfoundland) of
63 5,520 km. Habitats around these populations present a dynamic range of environmental conditions,
64 ranging from 32 to 432 mm of winter precipitation and a mean annual temperature range[26] of -6.3 to
65 4.7° C. To determine associations between these environmental conditions and genome-wide patterns
66 of DNA methylation, we created a reduced representation bisulfite sequencing (RRBS) library (full
67 protocol in Supplemental). Paired-end sequencing on a HiSeq2500 generated a total of 210,773,612
68 filtered and demultiplexed reads that were aligned to the domestic cat genome (85.0% average mapping
69 success; Supplemental Table S4; *Felis catus*; NCBI felcat9.0) and variants were called using specially
70 designed software for bisulfite converted reads[27]. We mapped our reads to the human (*Homo sapiens*)
71 and lambda phage genomes to rule out contamination (2.7% and <0.1% success, respectively), and
72 assessed bisulfite conversion efficiency by including non-methylated lambda phage DNA in the

73 sequencing lane. Furthermore, we quantified the temporal effects of methylation, the ramifications of
74 missing data, and model parameterization with sensitivity analyses, with no implications on overall
75 inferences (Supplemental Figures S5-S8).

76 **Pronounced population structure revealed by methylation patterns**

77 To examine regions with putative regulatory function, CpG islands were first bioinformatically identified
78 *de novo* (N = 28,127) using hidden Markov models and contained an average GC content of 59.2% with a
79 posterior probability of observed-to-expected GC content ($CpG_{o/e}$) of 1.13. Our DNA methylation analysis
80 was broken into two subsets based on proximity to genomic features, with identical filtering parameters
81 for each subset to maximize comparative inferences. Our subset over CpG islands and gene bodies
82 contained 329 5kb windows with 4,611 CpG positions, while the unannotated dataset contained 376 kb
83 windows with 5,031 CpG positions (Supplemental Table S6). Qualitatively, population structure was
84 most pronounced in DNA methylation patterns over CpG islands and gene bodies compared to
85 methylation patterns over regions of unknown function. The first two axes of a principal coordinates
86 analysis (PCoA) on a Euclidean distance matrix summarizing CpG island and gene body methylation
87 (PCoA1 = 6.2%, PCoA2 = 5.0% variation explained) distinctly cluster Alaska from the remaining mainland
88 populations (Figure 2). Similar, but less distinct patterns were seen in the ordination of methylation
89 patterns over unannotated regions of the genome (Figure 2). Analogous population-level trends
90 between the datasets were further confirmed with Pearson and Spearman's rank coefficients of the first
91 PCoA axis ($r = 0.91$; $\rho = 0.88$).

92 **Environmental variation associated with epigenetic structure**

93 For both methylation datasets, model selection identified three significant variables in our db-RDA
94 analysis: geographic distance, climate and a binary variable representing insularity for the
95 Newfoundland population (pseudo-$\boldsymbol{F}$ = 16.26 - 16.27; adjusted $R^2 = 0.33$; all $p = 0.001$). Tree cover (p =

96 0.13 - 0.50) and a randomly-generated numeric variable to assess the effect of noise ($p = 0.64 – 0.78$)
97 added no explanatory power to either model (Supplemental Table S11). Collinearity was low between all
98 retained variables (VIF = 2.09 – 3.77; Supplemental Table S12). We examined the explanatory power of
99 each variable independently using partial db-RDAs (Figure 4), which revealed similar trends between
100 both methylation datasets for geographic distance (pseudo-$\boldsymbol{F}$ = 15.72 – 17.47; adjusted R^2 = 0.11 – 0.12;
101 $p = 0.001$) and climate (pseudo-$\boldsymbol{F}$ = 6.75 – 6.90; adjusted R^2 = 0.04; $p = 0.001$). However, methylation
102 patterns over CpG islands and gene bodies were more strongly associated with the insular
103 Newfoundland population (pseudo-$\boldsymbol{F}$ = 13.66; adjusted R^2 = 0.09; $p = 0.001$) than with methylation
104 patterns over unannotated regions (pseudo-$\boldsymbol{F}$ = 10.03; adjusted R^2 = 0.07; $p = 0.001$), suggesting
105 stronger epigenetic divergence in putatively regulatory regions of the genome in insular Canada lynx
106 (Figure 3).

107 **Differential methylation over genes related to morphology**

108 We identified differential methylation directly over CpG islands and gene bodies using beta regressions
109 ($n = 329$), with population as an explanatory variable (i.e. windows with > 10% overall difference in
110 methylation and p-value < 0.001). Of the 16 gene regions identified, 11 were significantly differentiated
111 in the Newfoundland population (Table 1). All differentially methylated regions with associated
112 morphological function were hypermethylated, as well as genes with direct epigenetic function with
113 morphological interactions (i.e. transcription regulation, DNA-binding transcription factors; Table 1). The
114 11 differentially methylated regions in the insular population have a difference in methylation of 14%-
115 40% compared to mainland populations, which suggests that morphological divergence in the insular
116 population might be regulated by repressed gene expression due to DNA methylation. Additionally, we
117 identified three differentially methylated regions in the Alaskan population related to spectrin,
118 carbohydrate, and ATP binding, suggesting differential expression of genes with metabolic function.

119 **Genetic structure driven by insular and climatic divides**

120 We examined the relationship between neutral SNPs, environmental variation, geographic distance, and
121 insularity by first performing a principal coordinates analysis on a Euclidean distance dissimilarity matrix
122 of 489 genomic SNPs identified using a Bayesian wildcard algorithm with conservative calling and
123 filtering parameters[28]. Consistent with previous research on population structure in Canada lynx[13,29], our
124 SNP data analysis indicated substantial genetic separation of the insular Newfoundland population (F_{ST} =
125 0.10 – 0.13). We complemented our F_{ST} analysis with a null model approach that calculated the relative
126 differentiation between populations, and mainland populations were relatively similar (F_{ST} = 0.005 –
127 0.024) with Newfoundland as the most segregated (F_{ST} = 0.096). Consistent with previous research[14],
128 gene diversity was lowest in the Newfoundland population (H_e = 0.224). An analysis of molecular
129 variance (AMOVA) identified substantial genetic variation between populations, which explained 8.58%
130 of the total variation in SNP data (σ = 5.84, p = 0.01). Outlier detection with Bayescan v2.0[30] (false
131 discovery rate > 0.05) identified no loci under selection.

132 We determined relationships between genetic data and environmental variables by performing a
133 db-RDA step-wise model selection on the retained PCoA axes, and identified only climate (pseudo-***F*** =
134 51.13; p-value = 0.001) and insularity (pseudo-***F*** = 26.33; p-value = 0.001) as important for explaining
135 genetic variation (Supplemental Figure S11). Geographic distance (p-value = 0.31), tree cover (p-value =
136 0.95), and the randomly generated numeric variable (p-value = 0.33) were all dropped during the model
137 step-selection. The first axis of the db-RDA was strongly representative of both climate (db-RDA1 = -
138 0.86) and insularity (db-RDA1 = 0.99), while the second axis primarily summarized genetic variation
139 associated with climate (db-RDA2 = 0.51). Partial db-RDAs identified a substantial amount of the net
140 variation was explained by the mainland – Newfoundland divide (pseudo-***F*** = 26.3; adjusted R^2 = 0.15; p-
141 value = 0.001), while climatic variation explained much of the variation seen in mainland populations
142 (pseudo-***F*** = 13.70; adj. R^2 = 0.08; p-value = 0.001). We identified a stronger correlation between SNP

143 data and non-coding region methylation (r = -0.78; ρ = -0.62) than with CpG island and gene body
144 methylation (r = -0.74; ρ = -0.61) from the first axes of PCoAs using Pearson's and Spearman's ranking
145 coefficients, respectively.

146 **Discussion**

147 **A new chapter in Canada lynx population structure**

148 We have shown that despite seemingly panmictic populations across the mainland at the genetic level,
149 epigenetic structure is substantial and might be an underlying mechanism driving local adaptation. Row
150 et al. showed a minimal barrier effect of the Rocky Mountains to Canada lynx[13], but it remains possible
151 that distinct environmental variation might be driving unique patterns of functionally important DNA
152 methylation at the range margins, including Alaska and Newfoundland. Despite the lack of genetic
153 structure between mainland populations[29], we see distinct patterns of DNA methylation that
154 differentiate Alaska from mid- and east-continental populations. Furthermore, we identified variation
155 that is associated by trends in winter precipitation and temperature that is not accounted for by
156 geographic distance.

157 The differentiated DNA methylation patterns in the Newfoundland population suggest a molecular
158 pathway for morphological differences between mainland and island populations, as predicted by the
159 Island Rule. Evidence for this adaptive divergence was observed in disparities between methylation in
160 CpG islands/gene bodies and DNA of unknown function and specific genes. We identified specific genes
161 (ZEB1 and HDAC9) that may underlie differences in body size between mainland and Newfoundland
162 lynx. The hypermethylated HDAC9 gene is of particular interest because downregulation in HDAC9 is
163 associated with diminished body mass[31]. Commonly, insular mammalian carnivores are smaller in size[22],
164 which is consistent with existing data on the insular population of Canada lynx[17,18].

165 **An overlooked marker for examining rapid evolution and adaptive divergence**

166 Our results indicate that DNA methylation can detect population divergence despite substantial gene
167 flow, and can be a powerful marker to examine cryptic population structure in species that seem
168 genetically panmictic. Identifying the molecular basis of adaptive divergence between populations has
169 relied on comparing allele frequencies or isolating outlier loci under selection[32,33]. These results suggest
170 that epigenetic modifications increase resolution when defining population structure and are a useful
171 marker for examining subtle differentiation. With time, divergent methylation patterns should lead
172 linked SNPs to act as if under directional selection, and thus could be factored into the working model of
173 speciation[2]. Divergent methylation patterns also have implications for conservation. For example,
174 seemingly genetically homogenous individuals could be epigenetically adapted to local conditions[34], and
175 thus, greater care would be needed to select individuals for translocations[35]. These results indicate that
176 epigenetic modifications play a role in defining population structure, with a putative relationship to
177 adaptation.

178 For decades, the adaptive role of epigenetic modifications in evolution has been controversial[36–38].
179 Our data provide evidence that epigenetic modifications are a mechanism underlying rapid evolutionary
180 change. The evolutionary history of Canada lynx on Newfoundland is unclear, yet evidence suggests
181 post-glacial colonization (10,000 years)[13]. The disproportionate levels of differential methylation in the
182 smaller Newfoundland lynx, particularly over genes with morphological significance, suggest an
183 epigenetic pathway for rapid evolutionary responses to geographic isolation[36]. Insular populations of
184 mammals generally have notable phenotypic differentiation[22,39], and DNA methylation serves as a
185 potential mechanism driving insular phenotypic divergence due to its plasticity to environmental
186 conditions[40]. Collectively, our results suggest that epigenetic modifications like DNA methylation are an
187 overlooked molecular signature of population divergence and local adaptation, and offer a putative
188 mechanism underlying rapid evolutionary response.

189 **Methods**

190 **Sample Acquisition and Reduced-Representation Bisulfite Sequencing (RRBS) Library Preparation**

191 Georeferenced Canada lynx tissue samples were collected from fur auction houses throughout eastern
192 North America from dried pelts (North American Fur Auctions, Fur Harvester's Auctions, Inc.).
193 Individuals from four geographic locations were chosen for this study, spanning the longitudinal and
194 latitudinal range of the species and consistent with previous research on the Canada lynx system
195 (Rueness et al 2003). Tissue was consistently taken from the same morphological location from each
196 adult-sized pelt. Genomic DNA was isolated using a MagneSil® (Promega Corporation) Blood Genomic
197 Max Yield System a JANUS® workstation (PerkinElmer, Inc.) and quantified and standardized to 20 ng/µl
198 with a Quant-iT PicoGreen® ds-DNA assay using manufacturer's instructions (Thermo Fisher Scientific).
199 We adapted an existing reduced representation bisulfite sequencing library preparation workflow
200 designed for multiplexed high-throughput sequencing[41]. Genomic DNA (400 ng) for 95 Canada lynx
201 samples was digested with NsiI and AseI restriction enzymes overnight and subsequently ligated with
202 methylated adapters (Supplemental Table S1). An individual sample of completely non-methylated
203 lambda phage genomic DNA (200 ng; Sigma-Aldrich – D3654) with a unique barcode was included to
204 assess bisulfite conversion efficiency (Supplemental Table S7). Barcoded samples were then combined
205 into eight pools to ensure consistent reaction environments for the entire library using a QIAquick PCR
206 Purification Kit (Qiagen, Valencia, CA, USA) following manufacturer's instructions and 10 µL of 3M NaAc
207 was added to neutralize pH. We then performed a SPRI size selection on each pool (0.8x volume ratio)
208 with Agencourt® AMPure® XP beads (Beckman Coulter, Inc.). Nicks between the 3' fragment overhang
209 and the 5' non-phosphorylated adapter nucleotide were repaired with DNA polymerase I and bisulfite
210 conversion was performed on each pool using an EZ DNA Methylation-Lightning Kit™ (Zymo Research)
211 with a 20-minute desulphonation time.

212 Pools were amplified in three separate PCRs to mitigate stochastic differences in amplification and
213 were subsequently concentrated using a QIAquick PCR Purification Kit (Qiagen, Valencia, CA, USA). The

214 eight pools were quantified with Qubit 3.0® (Thermo Fisher Scientific) and appropriate amounts were
215 added to a final super-pool for equal weighting. A final magnetic bead clean-up (0.8x volume ratio) was
216 performed to remove any adapter dimer. We checked final library concentration and fragment
217 distribution with a Qubit® 3.0 (ThermoFisher Scientific) and Hi-Sense Bioanalyzer 2100 chip (Agilent),
218 respectively. Paired-end 125-bp sequencing was performed on a single lane on an Illumina HiSEq 2500
219 platform at the Centre for Applied Genomics at the Hospital for Sick Children (Toronto, Ontario,
220 Canada). Raw sequence data FastQ files are available on the Sequence Read Archive (ID Pending).

221 **Bioinformatics – quality checks and SNP calling**

222 We assessed sequencing success as well as removed adapter and low-quality reads via FastQC[42]
223 (Andrews, 2010) and Cutadapt[43] implemented in Trim_Galore! V0.4.4[44] (Krueger, 2014). Individuals were
224 demultiplexed using python scripts[41]. Paired-ends reads for Canada lynx samples were initially aligned to
225 several genomes (*Felis catus, Homo sapiens,* Lambda Phage) using Bismark[27] to assess mapping
226 efficiency and contamination (Supplemental Table S4). Paired-end reads for downstream analyses were
227 aligned to the domestic cat genome with Bismark, using relaxed mapping parameters (score_min L,0,-
228 0.6)[27].

229 SNPs were called from indexed BAM files with CGmapTools[28] using a coupled Bayesian wildcard
230 algorithm with a conservative 0.01 error rate and a static 0.001 p-value for calling variant sites, which
231 generated variant call files (VCFs). VCF files were indexed, merged, and filtered using VCFtools[45] for bi-
232 allelic loci with a sequencing depth of at least five, and were shared between at least 50% of the
233 individuals (max-missing 0.5) and a minor allele frequency (maf) of > 0.001. Sites were further filtered by
234 removing any variants out of Hardy-Weinberg equilibrium within populations (p-value < 0.05). A pair-
235 wise Euclidean dissimilarity (distance) matrix was computed on the SNP data using the function *daisy*
236 within the package *cluster*[46] using R v3.4.2[47]. This dissimilarity matrix was then summarized in a

237 principal coordinates analysis (PCoA) using the *dudi.pco* function in *adegenet*[48]. Missing SNP data was
238 imputed by mean allele at a population level. Pairwise F_{ST} was calculated using StAMPP (Supplemental
239 Figure S2) and relative F_{ST} was measured using a null model Bayesian approach implemented in
240 BayeScan v2.0[30] (Supplemental Figure S2). AMOVA between populations was computed using the
241 *poppr.amova* function within *poppr*[49], and heterozygosity was assessed with *adegenet.*

242 **Bioinformatics – DNA methylation**

243 We identified methylated and non-methylated positions by first filtering BAM files for incomplete
244 bisulfite conversion based on reads containing more than three methylated positions in a CHH or CHG
245 context. In the remaining reads, methylated positions in a CpG context with a sequencing depth of at
246 least five were extracted with Bismark[27], truncating the last two bases of the forward mate-paired reads
247 (R1) and the first two bases of the reverse mate-paired (R2) reads. Methylation polymorphisms in areas
248 of overlap between read pairs were extracted only once. After confirming that our non-directional
249 library contained roughly equal reads for all possible amplified DNA strands, we proceeded to analysis.

250 We first generated a custom CpG island annotation track using hidden Markov models based on
251 $CpG_{o/e}$ implemented in *makeCGI*[50]. Only islands with a calculated posterior probability greater than
252 99.5% were retained for analysis, based on $CpG_{o/e}$. Mapped and extracted methylated sites were then
253 imported into Seqmonk[51] using the generic text importer, and raw data was qualitatively visualized
254 against the annotated domestic cat genome (felCat9.0). We analyzed DNA methylation over CpG islands
255 and gene bodies by creating 5,000-bp running windows directly over and 25,000-bp upstream of gene
256 bodies, combined with windows directly over CpG islands. Each window was assigned a methylated
257 percentage score based on the overall ratio of methylated to non-methylated bases within the feature.
258 We filtered this window-set for regions that had at least one CpG and equal representation from each
259 population. This process was repeated for our second subset of analyses that investigated methylation

260 patterns of unannotated regions of the genome. For this analysis, we created 5,000-bp running windows
261 across the entire genome and removed any windows that overlapped with the initial windows over CpG
262 islands and gene bodies by more than 1%. Data was filtered, exported, and summarized in the same way
263 as the first analysis.

264 **Sensitivity analyses**

265 To determine the implications of missing data and PCoA axis retention thresholds, we performed two
266 sensitivity analyses. The first analysis examined the effects of missing data by repeating all analyses
267 using a subset of the top 10 individuals per population (N = 40), and again with the top 6 individuals per
268 population (N = 24), across all three datasets (Supplemental Figure S5-S6, Table S8). Overall trends in
269 explanatory effects (adjusted R^2) and qualitative inferences (PCoA clustering) were investigated and no
270 change in inferences were determined. We examined the implications of arbitrary PCoA axis retention
271 by repeating all analyses, but instead using different cumulative variation explained thresholds as
272 response variables. We performed a number of db-RDAs using all axes explaining 30%, 50%, 75%, and
273 95% cumulative variation as response variables, and results were qualitatively similar regardless of axis
274 retention (Supplemental Figure S8, Table S10).

275 **Quantifying environmental associations**

276 To determine if patterns of DNA methylation and genomic variability could be explained by macro-scale
277 climatic conditions, geographic distance, or insular divergence, we performed a distanced-based
278 redundancy analysis on the summarized SNP and DNA methylation data. Meaningful axes explaining >
279 30% of the cumulative variation in the data were used as response variables in a distance-based
280 redundancy analysis (db-RDA) conducted in *vegan*[52], using variables that putatively describe the
281 environmental determinants of population structure in Canada lynx. Our covariates included a binary
282 variable of insularity, which identified the Newfoundland population against mainland populations and

283 was used to describe the largely impassable barrier of the Strait of Belle Isle between Newfoundland
284 and mainland Labrador[53]. A variable of geographic distance was included which was simply the first axis
285 of a principal coordinates analysis (PCoA) on a Euclidean distance matrix of latitude and longitude (PCo1
286 = 99.7% of the variation). In addition to the geographic variables, we included a biotic variable of
287 percent tree cover[54], a randomly generated numerical variable to assess the effect of noise, and a
288 climate variable. For the climate variable, we performed a PCA on climate data to prevent multi-
289 collinearity, which reduced annual temperature ranges, winter precipitation, and minimum coldest
290 temperature to a single PCA axis[26] (PC1 = 85.6% of the variation). Linearity was confirmed between
291 response and explanatory variables, and multi-collinearity between explanatory variables was assessed
292 using the VIF and any variables > 4 were removed (Supplemental Table S12). Step-wise model selection
293 using the function *ordistep* within *vegan*[52] was performed to isolate the best overall model using a QR
294 decomposition technique based on p-values (Supplemental Table S11). To isolate the individual
295 explanatory power of each variable, we performed partial distanced-based redundancy analyses (p-
296 dbRDAs) on the variables that were identified as significant in the full db-RDA.

297 **Differentially methylated regions and gene ontology**

298 We identified differentially methylated regions with functional biological correlates by performing beta-
299 regressions on windows over CpG islands and gene bodies for all 95 individuals with percent
300 methylation as the response variable and population as the explanatory variable. Beta regressions are
301 appropriate for proportion or percentage data[55], and we set an alpha threshold at conservative levels
302 seen in similar studies[34] (p-value < 0.001). Overall differences in percent methylation were calculated,
303 and genes were identified as either hypermethylated or hypomethylated based on their relative degree
304 of methylation compared to other populations. Direct overlap between our differentially methylated
305 regions and the felcat9.0 gene annotations were extracted using Seqmonk[51]. We then identified
306 functional associations by searching UniProt (https://www.uniprot.org) by gene name against the

database of genes in the domestic cat genome using the search term "organism: *'Felis catus* (Cat) (*Felis silvestris catus*) (9685)'".

Acknowledgements

The authors thank Niels C.A.M. Wagemaker for his assistance with library preparation and his generous donation of methylated adapters. We would also like to thank Felix Krueger and Sibelle Torres Vilaça for their assistance with bioinformatics. We would like to thank Joe M. Northrup (OMNR), Elizabeth M. Kirepka, Lynne E. Beaty, Angela R. Eads, and Melanie R. Boudreau for their invaluable statistical advice and friendly manuscript reviews. We thank the Natural Resources DNA Profiling and Forensic Centre (NRDPFC) for their assistance with extractions and laboratory equipment, and the Centre for Applied Genomics with sequencing troubleshooting. This work was funded by the CFI-JELF (358-47; A.B.A.S.), ComputeCanada (GME-665-01; A.B.A.S.), NSERC Discovery Grants (D.L.M., A.B.A.S.) and CREATE-enviro (J.B.J.).

Author Contributions

All authors contributed equally to this work.

Competing Interests

The authors declare no competing interests with this work.

Materials and Correspondence

Inquiries can be directed to the corresponding author, J.B.J.

References

1. Watson, J. & Crick, F. Molecular structure of nucleic acids. *Nature* **171,** 737–8 (1953).

2. Feder, J. L., Egan, S. P. & Nosil, P. The genomics of speciation-with-gene-flow. *Trends Genet.* **28,**

328 342–350 (2012).

329 3. Shafer, A. B. A. & Wolf, J. B. W. Widespread evidence for incipient ecological speciation: A meta-
330 analysis of isolation-by-ecology. *Ecol. Lett.* **16,** 940–950 (2013).

331 4. Jones, P. A. & Takai, D. The role of DNA methylation in mammalian epigenetics. *Science (80-.).*
332 **293,** 1068–1070 (2001).

333 5. Dubin, M. J. *et al.* DNA methylation in Arabidopsis has a genetic basis and shows evidence of local
334 adaptation. *Elife* **4,** 1–23 (2015).

335 6. Artemov, A. V. *et al.* Genome-Wide DNA Methylation Profiling Reveals Epigenetic Adaptation of
336 Stickleback to Marine and Freshwater Conditions. *Mol. Biol. Evol.* **34,** 2203–2213 (2017).

337 7. Fujita, N. *et al.* Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1
338 heterochromatic complex for DNA methylation-based transcriptional repression. *J. Biol. Chem.*
339 **278,** 24132–24138 (2003).

340 8. Lorincz, M. C., Dickerson, D. R., Schmitt, M. & Groudine, M. Intragenic DNA methylation alters
341 chromatin structure and elongation efficiency in mammalian cells. *Nat. Struct. Mol. Biol.* **11,**
342 1068–1075 (2004).

343 9. Maunakea, A. K. *et al.* Conserved role of intragenic DNA methylation in regulating alternative
344 promoters. *Nature* **466,** 253–257 (2010).

345 10. Wutz, A. & P. Barlow, D. Imprinting of the mouse Igf2r gene depends on an intronic CpG island.
346 *Mol. Cell. Endocrinol.* **140,** 9–14 (1998).

347 11. Han, L., Su, B., Li, W.-H. & Zhao, Z. CpG island density and its correlations with genomic features
348 in mammalian genomes. *Genome Biol.* **9,** R79 (2008).

12. Jones, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13,** 484–492 (2012).

13. Row, J. R. *et al.* Dispersal promotes high gene flow among Canada lynx populations across mainland North America. *Conserv. Genet.* **13,** 1259–1268 (2012).

14. Prentice, M. B. *et al.* Selection and drift influence genetic differentiation of insular Canada lynx (Lynx canadensis) on Newfoundland and Cape Breton Island. *Ecol. Evol.* **7,** 3281–3294 (2017).

15. Stenseth, N. C. *et al.* Common Dynamic Structure of Canada Lynx Populations Within Three Climatic Regions. (1999).

16. Row, J. R. *et al.* The subtle role of climate change on population genetic structure in Canada lynx. *Glob. Chang. Biol.* **20,** 2076–2086 (2014).

17. Van Zyll De Jong, C. G. Differentiation of the Canada lynx, Felis (Lynx) canadensis subsolana, in Newfoundland. *Can. J. Zool.* **53,** 699–705 (1975).

18. Khidas, K., Duhaime, J. & Huynh, H. M. Morphological Divergence of Continental and Island Populations of Canada Lynx. *Northeast. Nat.* **20,** 587–608 (2013).

19. Foster, J. B. Evolution of mammals on Islands. *Nature* **202,** 234–235 (1964).

20. Van Valen, L. Pattern and the balance of nature. *Evol. Theory* **1,** 31–49 (1973).

21. Wayne, R. K. *et al.* A morphologic and genetic study of the Island fox, Urocyon littoralis. *Evolution (N. Y).* **45,** 1849–1869 (1991).

22. Lomolino, M. V. Body size evolution in insular vertebrates: Generality of the island rule. *J. Biogeogr.* **32,** 1683–1699 (2005).

23. Flatscher, R., Frajman, B., Schönswetter, P. & Paun, O. Environmental Heterogeneity and

370 Phenotypic Divergence: Can Heritable Epigenetic Variation Aid Speciation? *Genet. Res. Int.* **2012,**
371 1–9 (2012).

372 24. Stenseth, N. C. *et al.* Snow conditions may create an invisible barrier for lynx. *Proc. Natl. Acad.*
373 *Sci. U. S. A.* **101,** 10632–4 (2004).

374 25. Keller, T. E., Lasky, J. R. & Yi, S. V. The multivariate association between genomewide DNA
375 methylation and climate across the range of Arabidopsis thaliana. *Mol. Ecol.* **25,** 1823–1837
376 (2016).

377 26. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global
378 land areas. *Int. J. Climatol.* **37,** 4302–4315 (2017).

379 27. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq
380 applications. *Bioinformatics* **27,** 1571–1572 (2011).

381 28. Guo, W. *et al.* CGmapTools improves the precision of heterozygous SNV calls and supports allele-
382 specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics* **34,**
383 381–387 (2017).

384 29. Rueness, E. K. *et al.* Ecological and genetic spatial structuring in the Canadian lynx. *Nature* **425,**
385 69–72 (2003).

386 30. Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both
387 dominant and codominant markers: A Bayesian perspective. *Genetics* **180,** 977–993 (2008).

388 31. Chatterjee, T. K. *et al.* HDAC9 knockout mice are protected from adipose tissue dysfunction and
389 systemic metabolic disease during high-fat feeding. *Diabetes* **63,** 176–187 (2014).

390 32. Beaumont, M. A. & Balding, D. J. Identifying adaptive genetic divergence among populations

391 from genome scans. *Mol. Ecol.* **13,** 969–980 (2004).

392 33. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484,**
393 55–61 (2012).

394 34. Le Luyer, J. *et al.* Parallel epigenetic modifications induced by hatchery rearing in a Pacific
395 salmon. *Proc. Natl. Acad. Sci.* **114,** 12964–12969 (2017).

396 35. Murray, D. L., Steury, T. D. & Roth, J. D. Assessment of Canada Lynx Research and Conservation
397 Needs in the Southern Range: Another Kick at the Cat. *J. Wildl. Manage.* **72,** 1463–1472 (2008).

398 36. Ho, M. W. & Saunders, P. T. Beyond neo-Darwinism-an epigenetic approach to evolution. *J.*
399 *Theor. Biol.* **78,** 573–591 (1979).

400 37. Jablonka, E. & Raz, G. Transgenerational epigenetic inheritance: prevalence, mechanisms, and
401 implications for the study of heredity and evolution. *Q. Rev. Biol.* **84,** 131–76 (2009).

402 38. Robertson, M. & Richards, C. Opportunities and challenges of next-generation sequencing
403 applications in ecological epigenetics. *Mol. Ecol.* **24,** 3799–3801 (2015).

404 39. Meiri, S., Dayan, T. & Simberloff, D. Area, isolation and body size evolution in insular carnivores.
405 *Ecol. Lett.* **8,** 1211–1217 (2005).

406 40. Lea, A. J., Altmann, J., Alberts, S. C. & Tung, J. Resource base influences genome-wide DNA
407 methylation levels in wild baboons (Papio cynocephalus). *Mol. Ecol.* **25,** 1681–1696 (2016).

408 41. van Gurp, T. P. *et al.* epiGBS: reference-free reduced representation bisulfite sequencing. *Nat.*
409 *Methods* **13,** 322–4 (2016).

410 42. Shafer, A. B. A., Nielsen, S. E., Northrup, J. M. & Stenhouse, G. B. Linking genotype, ecotype, and
411 phenotype in an intensively managed large carnivore. *Evol. Appl.* **7,** 301–312 (2014).

412 43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
413 *EMBnet.journal* **17,** 10 (2011).

414 44. Shafer, A. B. A., Northrup, J. M., Wikelski, M., Wittemyer, G. & Wolf, J. B. W. Forecasting
415 Ecological Genomics: High-Tech Animal Instrumentation Meets High-Throughput Sequencing.
416 *PLoS Biol.* **14,** 1–11 (2016).

417 45. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27,** 2156–2158 (2011).

418 46. Shafer, A. B. A., Cullingham, C. I., Côté, S. D. & Coltman, D. W. Of glaciers and refugia: A decade of
419 study sheds new light on the phylogeography of northwestern North America. *Mol. Ecol.* **19,**
420 4589–4621 (2010).

421 47. Shafer, A. B. A., White, K. S., Côté, S. D. & Coltman, D. W. Deciphering translocations from relicts
422 in Baranof Island mountain goats: Is an endemic genetic lineage at risk? *Conserv. Genet.* **12,**
423 1261–1268 (2011).

424 48. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers.
425 *Bioinformatics* **24,** 1403–1405 (2008).

426 49. Kamvar, Z. N., Tabima, J. F. & Grünwald, N. J. *Poppr* : an R package for genetic analysis of
427 populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* **2,** e281 (2014).

428 50. Wu, H., Caffo, B., Jaffee, H. A., Irizarry, R. A. & Feinberg, A. P. Redefining CpG islands using hidden
429 Markov models. *Biostatistics* **11,** 499–514 (2010).

430 51. Britt, M., Haworth, S. E., Johnson, J. B., Martchenko, D. & Shafer, A. B. A. The importance of non-
431 academic coauthors in bridging the conservation genetics gap. *Biol. Conserv.* **218,** 118–123
432 (2018).

52. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14,** 927–930 (2003).

53. Koen, E. L., Bowman, J. & Wilson, P. J. Isolation of peripheral populations of Canada lynx (*Lynx canadensis*). *Can. J. Zool.* **93,** 521–530 (2015).

54. DeFries, R. S., Hansen, M. C., Townshend, J. R. G., Janetos, A. C. & Loveland, T. R. A new global 1-km dataset of percentage tree cover derived from remote sensing. *Glob. Chang. Biol.* **6,** 247–254 (2000).

55. Ferrari, S. L. P. & Cribari-Neto, F. Beta regression for modelling rates and proportions. *J. Appl. Stat.* **31,** 799–815 (2004).

452

Table 1. List of differentially methylated regions that have direct overlap with annotated genes. Differentially methylated regions were identified using beta regressions with a p-value < 0.001 and when methylation levels differed by more than 10% between the identified population and the others. Hyper- and hypo-methylation is relative to the other 3 populations. GO annotated functions were retrieved from UniProt (www.uniprot.org).

453

Population	Gene	UniProt ID	P-Value	% Difference	Hyper/Hypo-Methylated	GO Annotated Function
Alaska	ADGRL2	M3XFT6_FELCA	7.58E-07	20.17	Hyper	Carbohydrate binding, G-protein coupled receptor activity
	ANK2	M3WXH8_FELCA	2.41E-05	16.10	Hyper	Spectrin binding, maintenance of cytoskeletal structure
	ATP8A1	M3WM63_FELCA	0.000701	14.99	Hyper	ATP binding, magnesium ion binding
Manitoba	NRG3		0.000216	25.43	Hypo	No Uniprot Annotation
	RWDD1	M3XFU2_FELCA	1.86E-05	17.61	Hyper	Cytoplasmic translation
Newfoundland	CDH18	M3WJ78_FELCA	2.83E-05	21.23	Hypo	Calcium ion binding
	CENPU	M3W3F1_FELCA	0.000895	13.71	Hyper	Embryonic development
	CRISPLD1	M3W2H3_FELCA	0.000337	19.82	Hyper	Face morphogenesis
	DCC	M3WHY3_FELCA	0.000399	37.36	Hyper	Spinal cord ventral commissure morphogenesis
	FAM35A	M3WZ20_FELCA	0.000641	19.62	Hyper	No Uniprot Annotation
	HDAC9	A0A2I2U9G1_FELCA	4.40E-05	39.60	Hyper	Histone deacetylase; transcription regulation; downregulation associated with diminished body mass, adaptive thermogenesis1
	LOC101091724		0.000698	26.54	Hypo	lncRNA
	LOC109493917		0.000332	38.44	Hypo	lncRNA
	PBX3	M3WKS0_FELCA	3.21E-05	26.60	Hypo	DNA binding transcription factor activity
	TMOD2	M3X1Q9_FELCA	0.000847	29.43	Hyper	Myofibril assembly, muscle contraction, epithelial cell morphology3
	ZEB1	A0A2I2UAU6_FELCA	5.73E-05	37.60	Hyper	DNA-binding, transcription factor, regulation of adipose tissue mass2

454

455

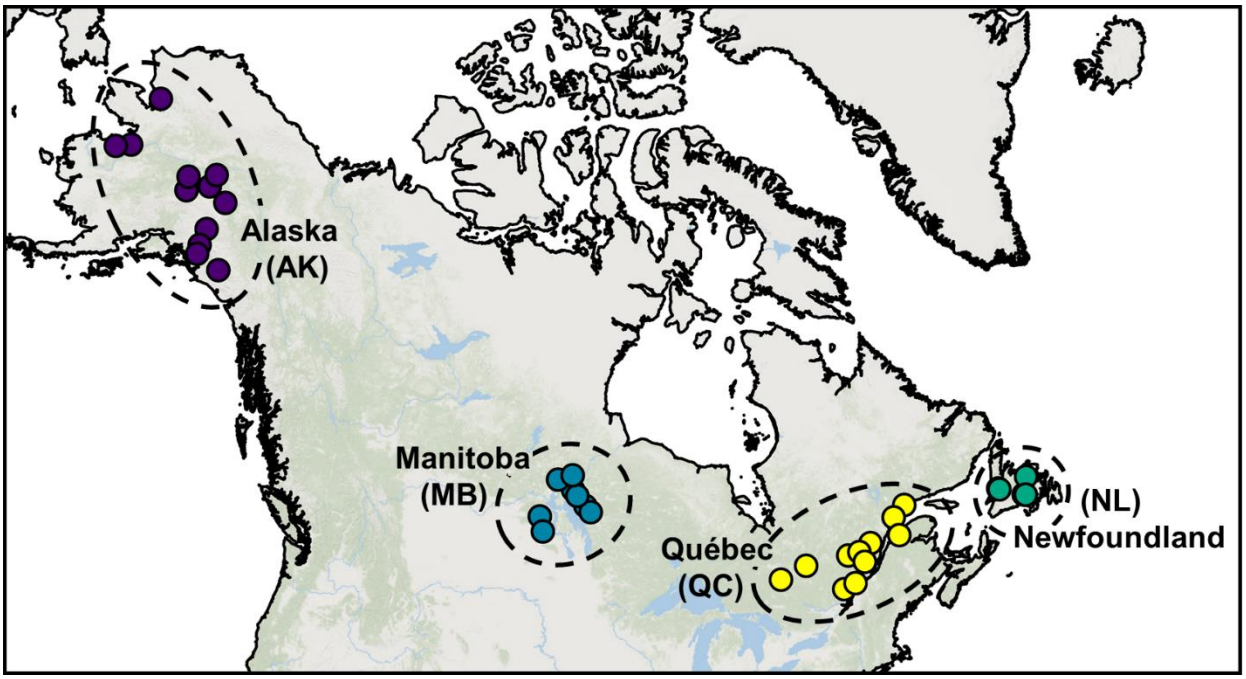
456

457

458

459

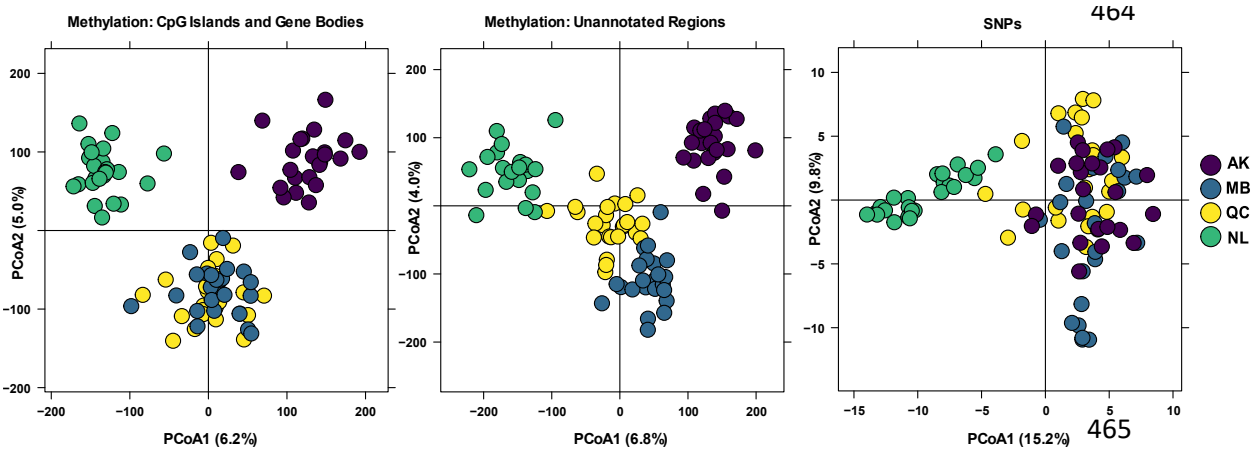
460



461

Figure 1. Distribution of 95 Canada lynx (*Lynx canadensis*) samples across North America, used for high-throughput bisulfite sequencing. All four populations are delineated by colour and include 24 individuals, except Alaska (n = 23).

463



464

465

Figure 2. Principal coordinates analysis (PCoA) plots of variation across three molecular marker datasets, with individuals as single circles and populations delineated by colour. All molecular data was summarized with a pair-wise Euclidean dissimilarity matrix. Methylation was summarized with 5,000-bp running windows over CpG islands and gene bodies (n = 329) and over unannotated regions (n = 376). SNP variants were called from bisulfite converted reads and reflect unstructured mainland populations (n = 496).

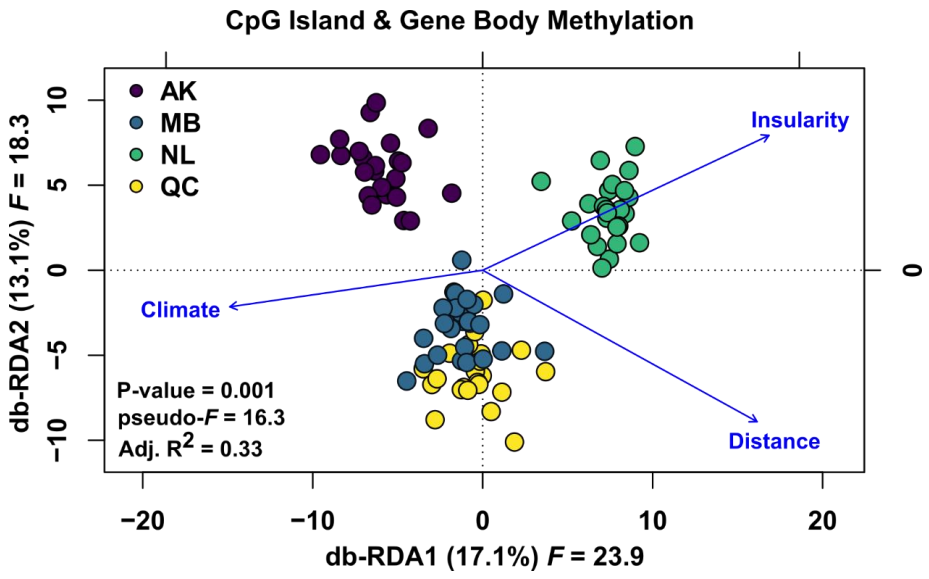


Figure 3. Distance-based redundancy analysis (db-RDA) on DNA methylation data over CpG islands and gene bodies, with population delineated by colour. The axes of a principal coordinates analysis (PCoA) were used a response variable to determine biogeographical relationships. The explanatory variables included a distance variable (the first axis of a PCoA on latitude and longitude); insularity (a binary variable distinguishing the Newfoundland island population); and climate (the first axis of a PCA summarizing winter precipitation, annual temperature ranges, and coldest minimum temperature).

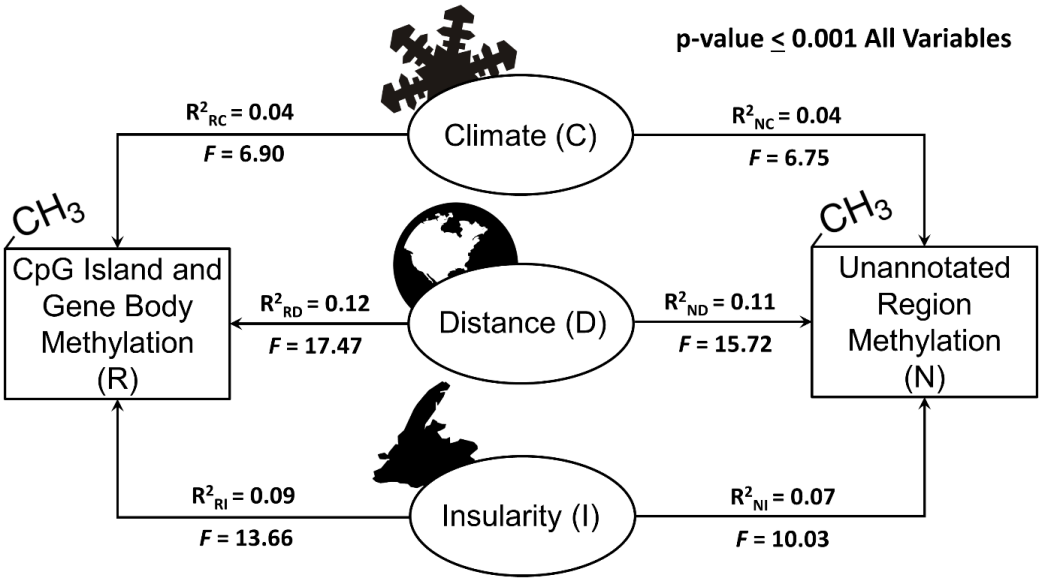


Figure 4. Visual depiction of partial distance-based redundancy analyses (p-db-RDAs). The effect sizes indicate the independent explanatory effects of each variable on explaining methylation patterns, subtracted from the effect of any other variable. The effect size is adjusted R^2, (adj. R^2) and the test-statistic is a pseudo-***F*** generated using QR decomposition within *vegan*.