SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

Temporally replicated DNA methylation patterns in great tit using reduced representation bisulfite sequencing

Hannu Mäkinen[1,3,4], Heidi M. Viitaniemi[1], Marcel E. Visser[2], Irene Verhagen[2], Kees van Oers[2] & Arild Husby[1,3,4]

Seasonal timing of reproduction is an important fitness trait in many plants and animals but the underlying molecular mechanism for this trait is poorly known. DNA methylation is known to affect timing of reproduction in various organisms and is therefore a potential mechanism also in birds. Here we describe genome wide data aiming to detect temporal changes in methylation in relation to timing of breeding using artificial selection lines of great tits (*Parus major*) exposed to contrasting temperature treatments. Methylation levels of DNA extracted from erythrocytes were examined using reduced representation bisulfite sequencing (RRBS). In total, we obtained sequencing data from 63 libraries over four different time points from 16 birds with on average 20 million quality filtered reads per library. These data describe individual level temporal variation in DNA methylation throughout the breeding season under experimental temperature regimes and provides a resource for future studies investigating the role of temporal changes in DNA methylation in timing of reproduction.

Background and Summary

In seasonally varying environments, timing of reproduction is under strong selection, as individuals need to adjust the time of reproduction to favorable environmental conditions. Individuals that are reproducing too early or too late in relation to the peak in food abundance may have reduced fitness[1,2]. Understanding how organisms translate environmental cues into phenotypes, such as timing of breeding is therefore important for predicting how individuals and populations respond to changing environmental conditions. A well-known environmental cue that plants and animals use to time their reproduction is photoperiod[3,4]. However, additional cues are also involved and increases in yearly temperatures has led to changes in seasonal timing of breeding in many plants and animals[5]. For example, in some passerine birds, the timing of reproduction has advanced about 0.25 days per year during the last three decades[6,7]. The shift towards earlier breeding in many species in the Northern hemisphere is likely an adaptive response to the increase in temperature and the resulting shift in the timing of emergence of their prey[1,8].

While we have a quite good understanding of the environmental factors and selective agents operating on seasonal timing of reproduction in birds, our knowledge about its genetic basis is poor. Seasonal timing of reproduction is frequently found to be heritable[9–12] but the underlying genes involved and how environmental cues are sensed and translated into a physiological response remains largely unknown.

Recently, several studies have indicated DNA methylation as a potential mechanism that may modulate gene expression via environmental stimuli[3,13]. Due to technical advances in next generation sequencing the characterization of DNA methylation has become a popular tool, also in non-model organisms[14]. The identification of methylated sites is based on bisulfite conversion of un-methylated C's to T's but methylated C's are not affected.

[1]Organismal and Evolutionary Biology Research Programme, University of Helsinki, Helsinki, Finland. [2]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands. [3]Evolutionary Biology, Department of Ecology and Genetics, Uppsala University, Uppsala, Sweden. [4]Centre for Biodiversity Dynamics, Department of Biology, NTNU, Trondheim, Norway. These authors contributed equally: Hannu Mäkinen and Heidi M. Viitaniemi. Correspondence and requests for materials should be addressed to H.M. (email: hannu. makinen@ebc.uu.se) or A.H. (email: arild.husby@ebc.uu.se)

1
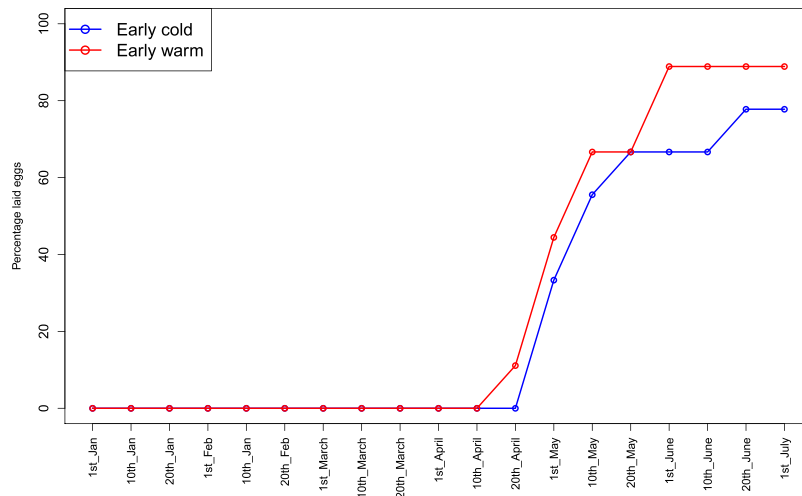
Fig. 1 An illustrated figure in which the percentage of females laying within the warm (red) and cold (blue) treatment are calculated per blood sampling moment. Females from a selection line for early reproduction were exposed to two different temperature regimes warm (red) or cold (blue) housed in climate-controlled aviaries from January until end of July. Blood samples were collected biweekly from each individual during this time. Egg laying was monitored daily from mid-March onwards.

In most methylation analyses sequencing reads are mapped to a reference genome and those sites showing no C to T change are considered as methylated sites. Furthermore, techniques such as reduced representation bisulfite sequencing (RRBS) that use a restriction enzyme (such as MSPI in most vertebrates) enriching for CpG rich regions, allow for cost efficient methylation profiling[15].

The great tit has become an ecological model species for understanding the impact of climate change on many different aspects, including morphological changes[16], population sizes[17] and reproductive related traits[18,19]. Earlier studies have found that great tits adjust their timing of breeding based on local environmental conditions[9,10], indicating that this trait is phenotypically plastic[9–11]. Experimental studies have demonstrated that temperature is causally related to the initiation of timing of breeding in great tits[20–22]. Thus, temperature may result in differences in methylation profiles for individuals exposed to different temperatures causing gene expression differences among these individuals, leading to differences in timing of reproduction. While there are a number of ecological studies examining DNA methylation it is often difficult to rule out potential confounding factors behind observed methylation changes in natural populations. To avoid this problem, we used great tit blood samples from birds originating from a genomic selection experiment for early timing of breeding that were kept in climate-controlled aviaries[23]. This allowed us to control for variation in factors such as food availability and age of the birds that could otherwise confound our results. We characterized DNA methylation patterns using reduced representation bisulfite sequencing (RRBS) of females subjected to either a cold or a warm temperature treatment. We sampled each individual in a temporally replicated manner during the breeding season to better understand potential seasonal variation in DNA methylation as well as in relation to experimental temperature effects.

Methods

Samples for sequencing.   We used samples from great tit females belonging to a large-scale artificial selection experiment (for more exact details see[23]). In short, DNA samples of erythrocyte origin, from offspring (first generation 'F1') from phenotypically early and late wild breeding pairs (parental generation 'P') were collected and genotyped using a 650k SNP chip[23]. Genomic breeding values (GEBVs) were calculated, which were used for bi-directional genomic selection for early or late reproduction. GEBV is the value of an individual in the breeding scheme based on the estimated genomic marker (i.e. SNPs) effects throughout the genome[23]. The individuals carrying the most extreme GEBVs within the two lines, produced the F2 generation. Of the F2, 36 breeding pairs (late selection line n = 18, early selection line n = 18) were housed in climate-controlled aviaries from January until July and subjected to contrasting temperature environments, mimicking a cold (2013) or warm (2014) spring in The Netherlands. Birds were allowed to breed and their egg laying date (the date at which they laid their first egg) was recorded.

Blood samples were collected from the birds every other week during the experiment (from 28-01 until 07-07-2016), with half the birds sampled in odd weeks and the other half sampled in even weeks (Fig. 1). For every sampling moment, all four line × treatment combinations were represented. After weighing, a blood sample (max. 150 μl) was taken from the jugular vein with a syringe (Easy Touch Insulin, 0.3 ml with 31 G). All birds were sampled within 10 minutes of capture. Plasma was separated from red blood cells with a Hamilton syringe after centrifuging at 14000 rpm for 10 minutes and the red blood cells were stored in Queens buffer (0.01 M Tris, 0.01 M NaCl, 0.01 M EDTA, 1% n-lauroylsarcosine, pH 8.0)[24] at room temperature until being processed. The experiment was performed under the approval by the Animal Experimentation Committee of the Royal Academy of Sciences (DEC-KNAW), Amsterdam, The Netherlands, protocol NIOO 14.10.
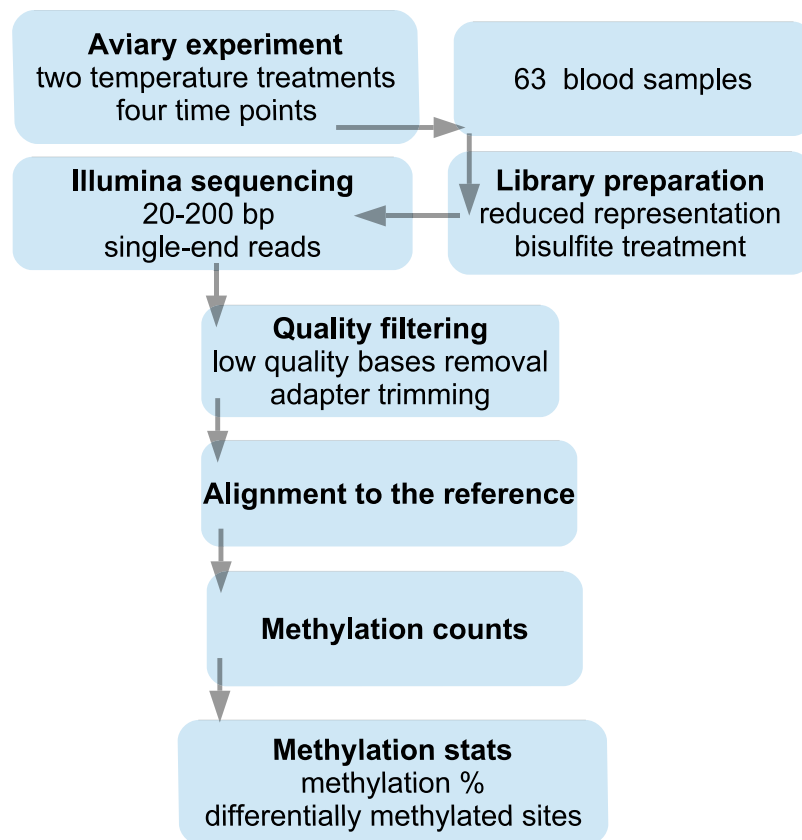
Fig. 2 A schematic overview of the work flow to generate the data. First great tit offspring were collected from the wild were used to generate F1 and F2 generations for the aviary experiment. Blood samples from F2 females were collected from four time points and two thermal treatments. Standard Illumina protocols were used for library preparation and sequencing. Several quality control steps were performed for raw sequencing reads for subsequent filtering steps. Quality filtered reads were mapped against the reference sequence and methylation counts were recorded for statistical analyses.

For this study, we used the red blood cell samples from the early selection line females (n = 16) from both temperature treatments. In addition, four sampling time points were chosen for analysis based on the known effects of photoperiod on reproduction[19] and the realized lay dates of the individuals; (1) the day when day light length >12hrs (time point 1), (2) the day when 25% of the females from the warm environment had initiated laying (time point 2), (3) the day when 25% and 50% of the females from the cold and warm environment respectively had initiated laying (time point 3) and (4) the day when 50% of the females from the cold treatment had initiated laying. As these time points do not coincide with the days of blood sampling we chose blood samples closest to (+/−7 days) the four time points. One female (warm treatment, time point 4) was incubating at the time so we could not take a blood sample at this stage. The total number of samples is therefore 63. Because of the blood sampling scheme and that females were incubating eggs (and hence were not sampled) there is a one- to two-week difference between the exact sampling days within a time point (Fig. 1).

Data analysis. The overall work flow of this study is described in Fig. 2. Total genomic DNA was extracted using FavorPrepT M 96-well Genomic DNA Kit (Favorgen). RNA was removed with an RNAse treatment. DNA Quality and quantity was assessed using a Nanodrop 2000 (Agilent Biotechnologies) and by 1% agarose gel electrophoresis. Approximately 1 ug of total genomic DNA was used for library preparation. A reduced representation library preparation protocol was used according to manufacturer's protocol (Illumina). DNA samples were first digested with restriction enzyme MspI to generate CCGG overhangs. Fragmented DNA was then bisulfite treated, which converts un-methylated cytosine nucleotides to thymine nucleotides. Fragmented and bi-sulfite treated DNA was then end-repaired with DNA polymerase I and A-overhangs were added to the 3' ends of each fragment for adapter ligation. Standard Illumina adapters containing individual barcodes were used for identification of sequencing reads in the downstream analyses after the sequencing. The libraries were size-selected for fragment sizes 20–200 basepairs (bp), and concentrations were determined by quantitative PCR. Sixteen libraries were pooled into the same sequencing lane of a flow cell by randomizing individuals, sampling days and treatments to avoid lane effects[25]. Altogether eight lanes were used for sequencing such that each pooled set was sequenced on two lanes with 100 bp from single end reads. All the pools were run on the same flow cell on a HiSeq. 2500 sequencer using a HiSeq SBS sequencing kit version 4 (Illumina). An internal positive control (PhiX)
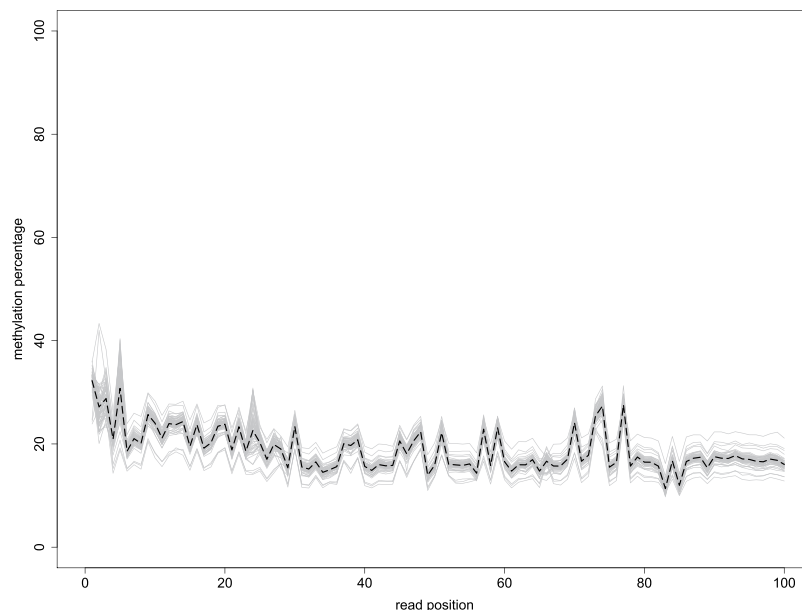
Fig. 3 Methylation bias plot in CpG context. The grey line shows the mean methylation percentage across all sequence reads in one sequence library in each position of the sequence read. The black dashed line shows the mean methylation percentage across all libraries.

library	no hit	aves	mammalia	other
BD 27012_1	4759	10	6	255
BD 27012_2	4957	8	6	29
BD 27012_3	4955	8	6	31
BD 27012_4	4952	12	1	35

Table 1. Results of the Blast searches of unmapped reads from four libraries.

was used to obtain reliable sequence generation in the sequencing processing. Library preparation and sequencing were performed at the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign, USA.

The quality of the sequencing reads was investigated as implemented in the FastQC 0.11.2 quality control tool[25,26]. The quality control analysis indicated presence of low quality bases in the 3' end of the reads[25]. The low quality bases and adapter contamination were trimmed using Trim Galore! 0.4.2[27] with default parameters[25]. In order to obtain methylation counts sequencing reads were aligned against the Great tit reference genome v1.1[28] using Bismark 0.16.3 aligner[29]. Sites in sequence reads containing Cs in comparison to reference sequence were taken as methylated sites whereas Ts were taken as un-methylated sites. The estimation of methylation percentage was based on the relative proportion of methylated and un-methylated sites. The methylation bias (M-bias)[30], i.e. if the methylation level at different position of the read varies, was examined by plotting the average methylation percentage in each position along the read in CpG context (Fig. 3). As can be seen from Fig. 3 there is relatively little technical variation in methylation across the reads, although, depending on the application of these data, the first 5 bp should possibly be removed, as methylation levels are a bit higher in this region.

Estimation of bisulphite conversion rate was based on proportion of C/T in the mitochondria to the reference genome. As mitochondria is mostly un-methylated it can be used to estimate conversion efficiency by comparing the amount of C/T in the reference sequence to what was acquired by RRBS[31]. Similarly, non-CpG methylation is negligible in great tit red blood cells[28,32] and these sites can also be used to estimate bisulphite conversion efficiency. Using mitochondria, the conversion efficiency is 99.8% and based on non-CpG methylation it is 98.6%[25].

We obtained on average 20.16 million raw sequencing reads per library, of which 20.05 million remained after quality filtering and with average coverage of 13.6x per CpG site per sample[25]. As reads were of good quality prior to trimming, read lengths after trimming were hardly affected[25]. Of these trimmed reads, 10.50 million quality-filtered reads were mapped uniquely to the great tit reference genome[25], resulting in average mapping efficiency of $52.0 \pm 2.0\%$. There are several reasons that could explain the rather large proportion of unmapped reads. First, unmapped reads can be contamination from another organism during e.g. library preparation or field sampling. To examine this, we took a subset of first 5,000 unmapped reads from four sequencing libraries and we used Blast search against the non-redundant nucleotide database available at GenBank. The majority of the unmapped reads did not have a Blast hit indicating that contamination of our libraries was unlikely (Table 1). Second, read mapping could be compromised due to incomplete reference genome. The published great tit genome contains a large number (~ 1,500) unordered scaffolds, indicating that some parts of the genome are

	No coverage filtering	1x	3x	5x	10x
CpG sites	11 057 685	2 653 390	2 217 299	1 730 250	522 645
Methylation % selected	21.54	18.8	17.45	16.45	13.82
Methylation % excluded			25.65	23.2	20.02

Table 2. Numbers of sites identified across all samples with different total coverage cut-offs (no cut off, 1x, 3x, 5x and 10x) and methylation levels in the included and excluded sites. Coverage cut off was required for all 16 individuals across the 4 time points.
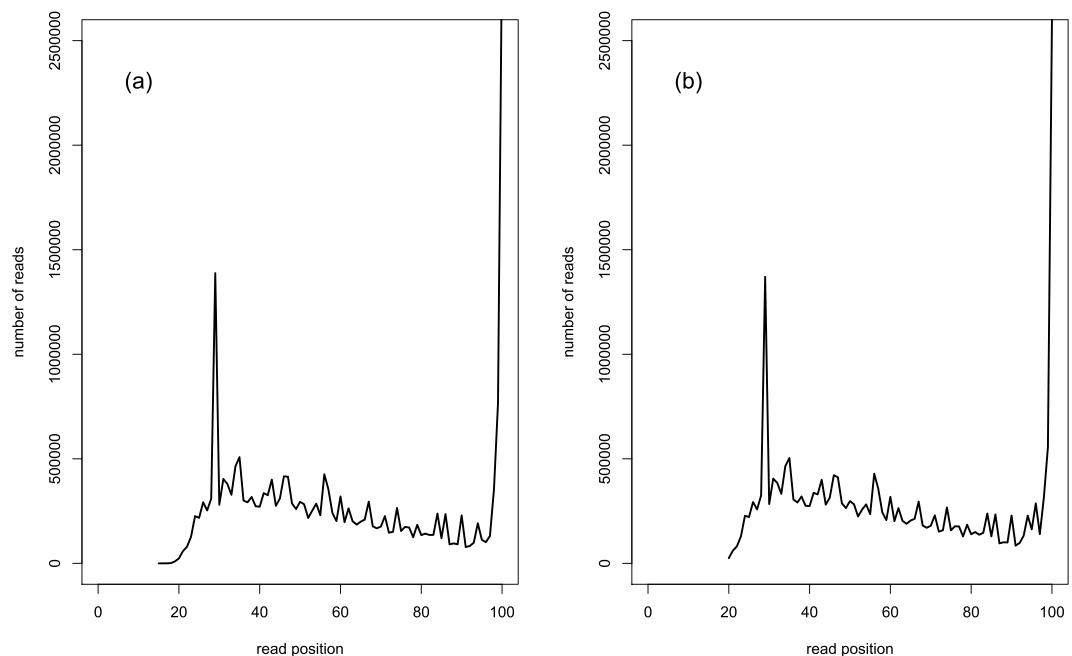
Fig. 4 Representative plots of read length distributions. (**a**) Shows the read length distribution in a raw sequence library. (**b**) Shows the read length distribution in a filtered sequence library. On the X axis is the read length and on the Y axis is number of occurrences of reads of specified lengths.

not included in the assembly in their correct genomic location although they are included in the reference used in the alignment. Thus, it is possible that reads do not perfectly align to the scaffolds, especially at the ends of them. Also, while mapping efficiency seems low it is comparable to other methylation studies in great tits. For example, Derks *et al.*[32] used whole genome methylation sequencing and found that 52% (brain) and 64% (blood) of the reads were mapped against the same great tit reference genome as used in this study. We did however observe a large proportion of very short (about 30 bp) reads among the unmapped reads which are known to be challenging to map against any reference for the current alignment software[33] and thus might explain some of the reason for the relatively low mapping success.

Altogether 11,057,686 methylated sites across 63 samples were identified for differential methylation analyses[25]. These sites covered 71.9% of all known CpG sites in the great tit genome. The mean methylation level was $21.54 \pm 1.45\%$[25], which is similar to that observed in a single male individual using whole-genome bisulfite sequencing[28] and in another RRBS dataset on great tits[32]. Non-CpG methylation in the samples was low ($0.46 \pm 0.15\%$, mean and sd)[25]. The identified CpG sites covered 80% of the genes in the current great tit annotation (version 1.1) and encompassed different genomic locations, estimated using the R packages GenomicFeatures[34] and rtracklayer[35]. Identified sites were annotated to introns (39.9%), exons (34.3%), promoters (10.3%) and intergenic regions (15.4%). From earlier work on the great tit, we know that gene expression is associated differentially with these regions[32,36]. Depending on total coverage cut-off, numbers of sites shared across all samples drop quite quickly (Table 2.). Mean methylation level in included and excluded sites also drops when requiring higher total coverage per site and the site to present with required coverage in all 16 samples and all 4 sampling time points. This is mainly because of two things: (1) when filtering for coverage across all samples we are excluding single sites which have high methylation level but which are present in some individuals only and (2) estimation of methylation level is based on low total coverage and can thus be erroneous. Thus, we encourage the use of total coverage when further filtering the data set for downstream analysis to allow more accurate calling of methylation level for any type of downstream analysis.

The temporally replicated DNA methylation dataset reported here is one of very few genome-wide characterizations of DNA methylation in ecological model species available and will serve as an important resource

for future studies examining the stability and repeatability of methylation and the link between methylation and timing of reproduction. It will also be an important resource for future comparative studies of DNA methylation patterns in birds.

Data Records

The data described here consists of sequence reads deposited in NCBI Sequence Read Archive[37]. All the libraries and the individual fastq files contained in them are deposited under accessions SRX3209916-SRX3209919[37–40]. The Figshare records[25] comprise of four files. First file is a summary table including information on sequencing design, mapping statistics and methylation levels in different contexts. Second and third file show sequencing quality reports before and after quality trimming, respectively. The fourth is a table reporting raw methylation counts and methylation level for each CpG site.

Technical Validation

Quality filtering steps taken to ensure the sequences[25] are of good quality are described in Methods section. The effect of quality filtering is presented in[25] with counts of the raw reads as well as reads after trimming for low quality bases at the 3' end[25] and length distributions of raw reads and after trimming (Fig. 4). We acknowledge the potential presence of PCR duplicates in the dataset resulting in low sequence complexity. As most of the reads in the data will have identical start-stop coordinates as a result of RRBS library preparation, deduplication based on just coordinates is not recommended[41]. Finally, masking the genome for G/C polymorphisms might lead to more accurate calling of methylated loci[41]. Such effect has been observed in humans when inter-population divergence was taken into account[42]. However, the samples in this study originate from the same natural population and we expect that the effect of background polymorphisms on the identification methylation calling is not greatly compromised.

References

1. Visser, M. E., Holleman, L. J. M. & Gienapp, P. Shifts in caterpillar biomass phenology due to climate change and its impact on the breeding biology of an insectivorous bird. *Oecologia* **147**, 164–172 (2006).
2. Both, C., Bouwhuis, S., Lessells, C. M. & Visser, M. E. Climate change and population declines in a long-distance migratory bird. *Nature* **441**, 81–83 (2006).
3. Stevenson, T. J. & Prendergast, B. J. Reversible DNA methylation regulates seasonal photoperiodic time measurement. *Proc. Natl. Acad. Sci.* **110**, 16651–16656 (2013).
4. Searle, I. & Coupland, G. Induction of flowering by seasonal changes in photoperiod. *EMBO J.* **23**, 1217–1222 (2004).
5. Peñuelas, J. *et al.* Evidence of current impact of climate change on life: A walk from genes to the biosphere. *Glob. Chang. Biol.* **19**, 2303–2338 (2013).
6. Källander, H. *et al.* Variation in laying date in relation to spring temperature in three species of tits (Paridae) and pied flycatchers Ficedula hypoleuca in southernmost Sweden. *J. Avian Biol.* **48**, 83–90 (2017).
7. Visser, M. E., te Marvelde, L. & Lof, M. E. Adaptive phenological mismatches of birds and their food in a warming world. *J. Ornithol.* **153**, 75–84 (2012).
8. Burgess, M. D. *et al.* Tritrophic phenological match-mismatch in space and time. *Nat. Ecol. Evol.* **2**, 970–975 (2018).
9. Husby, A. *et al.* Contrasting patterns of phenotypic plasticity in reproductive traits in two great tit (Parus major) populations. *Evolution (N. Y).* **64**, 2221–2237 (2010).
10. Husby, A., Visser, M. E. & Kruuk, L. E. B. Speeding Up Microevolution: The Effects of Increasing Temperature on Selection and Genetic Variance in a Wild Bird Population. *PLoS Biol.* **9**, e1000585 (2011).
11. Nussey, D. H. Selection on Heritable Phenotypic Plasticity in a Wild Bird Population. *Science (80-.).* **310**, 304–306 (2005).
12. Gienapp, P., Laine, V. N., Mateman, A. C., van Oers, K. & Visser, M. E. Environment-Dependent Genotype-Phenotype Associations in Avian Breeding Time. *Front. Genet.* **8**, 1–9 (2017).
13. Meijón, M., Feito, I., Valledor, L., Rodríguez, R. & Cañal, M. J. Promotion of flowering in azaleas by manipulating photoperiod and temperature induces epigenetic alterations during floral transition. *Physiol. Plant.* **143**, 82–92 (2011).
14. Laird, P. W. Principles and challenges of genome-wide DNA methylation analysis. *Nat. Rev. Genet.* **11**, 191–203 (2010).
15. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).
16. Husby, A., Hille, S. M. & Visser, M. E. Testing Mechanisms of Bergmann's Rule: Phenotypic Decline but No Genetic Change in Body Size in Three Passerine Bird Populations. *Am. Nat.* **178**, 202–213 (2011).
17. Reed, T. E., Grotan, V., Jenouvrier, S., Saether, B.-E. & Visser, M. E. Population Growth in a Wild Bird Is Buffered Against Phenological Mismatch. *Science (80-.).* **340**, 488–491 (2013).
18. Husby, A., Kruuk, L. E. B. & Visser, M. E. Decline in the frequency and benefits of multiple brooding in great tits as a consequence of a changing environment. *Proc. R. Soc. B Biol. Sci.* **276**, 1845–1854 (2009).
19. Charmantier, A. *et al.* Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. *Science (80-.).* **320**, 800–803 (2008).
20. Visser, M. E., Holleman, L. J. M. & Caro, S. P. Temperature has a causal effect on avian timing of reproduction. *Proc. R. Soc. B Biol. Sci.* **276**, 2323–2331 (2009).
21. Schaper, S. V. *et al.* Increasing Temperature, Not Mean Temperature, Is a Cue for Avian Timing of Reproduction. *Am. Nat.* **179**, E55–E69 (2012).
22. Visser, M. E. *et al.* Genetic variation in cue sensitivity involved in avian timing of reproduction. *Funct. Ecol.* **25**, 868–877 (2011).
23. Verhagen, Irene *et al.* Genetic and phenotypic responses to genomic selection for timing of breeding in a wild songbird. *Funct. Ecol.* **00**, 1–14. https://doi.org/10.1111/1365-2435.13360 (2019).
24. Seutin, G., White, B. N. & Boag, P. T. Preservation of avian blood and tissue samples for DNA analyses. *Can. J. Zool.* **69**, 82–90 (1991).
25. Mäkinen, H. *et al.* Supplementary data for Temporally replicated dna methylation patterns in great tit using reduced representation bisulfite sequencing. *Figshare*, https://doi.org/10.6084/m9.figshare.c.4511198 (2019).
26. Andrews, S. FastQC 0.11.2 quality control tool, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (2016).
27. Krueger, F. Trim Galore! 0.4.2, https://github.com/FelixKrueger/TrimGalore (2016).
28. Laine, V. N. *et al.* Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.* **7**, 10474 (2016).
29. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).

30. Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol.* **13**, R83 (2012).
31. Mechta, M., Ingerslev, L. R., Fabre, O., Picard, M. & Barrès, R. Evidence suggesting absence of mitochondrial DNA methylation. *Front. Genet.* **8**, 1–9 (2017).
32. Derks, M. F. L. *et al.* Gene and transposable element methylation in great tit (Parus major) brain and blood. *BMC Genomics* **17**, 1–13 (2016).
33. Tran, H., Porter, J., Sun, M.-A., Xie, H. & Zhang, L. Objective and Comprehensive Evaluation of Bisulfite Short Read Mapping Tools. *Adv. Bioinformatics* **2014**, 1–11 (2014).
34. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* **9**, 1–10 (2013).
35. Lawrence, M., Gentleman, R. & Carey, V. rtracklayer: An R package for interfacing with genome browsers. *Bioinformatics* **25**, 1841–1842 (2009).
36. Verhulst, E. C. *et al.* Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD4 DNA methylation. *Mol. Ecol.* **25**, 1801–1811 (2016).
37. *NCBI Sequence Read Archive*, https://identifiers.org/ncbi/insdc.sra:SRX3209916 (2018).
38. *NCBI Sequence Read Archive*, https://identifiers.org/ncbi/insdc.sra:SRX3209917 (2018).
39. *NCBI Sequence Read Archive*, https://identifiers.org/ncbi/insdc.sra:SRX3209918 (2018).
40. *NCBI Sequence Read Archive*, https://identifiers.org/ncbi/insdc.sra:SRX3209919 (2018).
41. Wreczycka, K. *et al.* Strategies for analyzing bisulfite sequencing data. *J. Biotechnol.* **261**, 105–115 (2017).
42. Daca-Roszak, P. *et al.* Impact of SNPs on methylation readouts by Illumina Infinium HumanMethylation450 BeadChip Array: Implications for comparative population studies. *BMC Genomics* **16**, 1–13 (2015).

Acknowledgements

Author Contributions

H.M.V. and H.M. conducted the bioinformatic analyses and H.M. wrote the first draft of the manuscript with comments and input from all authors. A.H. and M.E.V. conceived and secured funding for the project and together with K.V.O. supervised all aspects of the work. IV performed the experiment and collected the blood samples.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.