Received: 3 September 2021 | Revised: 29 March 2022 | Accepted: 13 April 2022

DOI: 10.1111/1755-0998.13624

MOLECULAR ECOLOGY RESOURCES WILEY

FROM THE COVER

An ecologist's guide for studying DNA methylation variation in wild vertebrates

Veronika N. Laine[1] | Bernice Sepers[2,3] | Melanie Lindner[2,4] | Fleur Gawehns[2] | Suvi Ruuskanen[5,6] | Kees van Oers[2,3]

[1]Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland

[2]Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, The Netherlands

[3]Behavioural Ecology Group, Wageningen University & Research (WUR), Wageningen, The Netherlands

[4]Chronobiology Unit, Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, The Netherlands

[5]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

[6]Department of Biology, University of Turku, Finland

Correspondence
Veronika N. Laine, Kees van Oers, Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland.
Emails: veronika.laine@helsinki.fi; K.vanOers@nioo.knaw.nl

Funding information
NWO open competition grant, Grant/Award Number: ALWOP.314; Turku Collegium for Science and Medicine

Handling Editor: Shawn Narum

Abstract

The field of molecular biology is advancing fast with new powerful technologies, sequencing methods and analysis software being developed constantly. Commonly used tools originally developed for research on humans and model species are now regularly used in ecological and evolutionary research. There is also a growing interest in the causes and consequences of epigenetic variation in natural populations. Studying ecological epigenetics is currently challenging, especially for vertebrate systems, because of the required technical expertise, complications with analyses and interpretation, and limitations in acquiring sufficiently high sample sizes. Importantly, neglecting the limitations of the experimental setup, technology and analyses may affect the reliability and reproducibility, and the extent to which unbiased conclusions can be drawn from these studies. Here, we provide a practical guide for researchers aiming to study DNA methylation variation in wild vertebrates. We review the technical aspects of epigenetic research, concentrating on DNA methylation using bisulfite sequencing, discuss the limitations and possible pitfalls, and how to overcome them through rigid and reproducible data analysis. This review provides a solid foundation for the proper design of epigenetic studies, a clear roadmap on the best practices for correct data analysis and a realistic view on the limitations for studying ecological epigenetics in vertebrates. This review will help researchers studying the ecological and evolutionary implications of epigenetic variation in wild populations.

KEYWORDS
bisulfite sequencing, DNA methylation, ecology, epigenetics, evolution

1 | INTRODUCTION

The field of molecular ecology is constantly advancing with the development of new powerful technologies, methods and analysis software. Tools for sequencing-based data analyses are increasingly used in ecological and evolutionary research as costs decrease and analyses become more time-efficient (Aguirre et al., 2019; van Gurp et al., 2016; Solares et al., 2018). In particular, the interest for a role of epigenetics in ecology and evolution is rapidly growing (Bossdorf et al., 2008; Hawes et al., 2018; Kilvitis et al., 2014; Ledón-Rettig et al., 2013; Schrey et al., 2013; Sepers et al., 2019; Verhoeven et al., 2016; Vogt, 2021). In ecological epigenetics studies, DNA methylation is the most widely studied epigenetic mechanism. DNA methylation involves the addition of a methyl group to a DNA nucleotide, usually a cytosine (C), and it affects the binding of proteins required for transcription initiation (Yin et al., 2017). This

process usually represses gene expression (Bird, 2002; Goldberg et al., 2007; Moore et al., 2013), especially if methylated sites are located close to the transcription start site in the promoter region (Bird, 2002; Laine et al., 2016; Li et al., 2011; Moore et al., 2013), but the relationship between DNA methylation and gene expression is rather complex and hard to generalize as DNA methylation can also increase transcription (Korochkin, 2006). Because DNA methylation can affect gene expression, this epigenetic mechanism is generally accepted to mediate the expression of phenotypic traits (Law & Jacobsen, 2010).

1.1 | Mechanisms causing variation in DNA methylation and phenotypic consequences

DNA methylation can be induced by genetic variation (Richards, 2006), by spontaneous epimutations (Becker et al., 2011) and by environmental induction (Pértille et al., 2017; Weaver et al., 2004; Zimmer et al., 2017; for more studies see below), although environmentally induced epigenetic variation might depend on genetic variation as well. Given the effects on transcription and subsequently gene expression, DNA methylation is, in theory, able to fine-tune phenotypic expression of ecologically relevant traits under environmental influence (Law & Jacobsen, 2010). This implies that epigenetic mechanisms provide an organism with the opportunity to develop a phenotype that is adaptive, in response to the environment (Jablonka & Lamb, 2007). These epigenetic changes may persist and affect the phenotype throughout an individual's lifetime (Roth et al., 2009; St-Cyr & McGowan, 2015). Thus, DNA methylation can be a mechanism underlying phenotypic plasticity, which can be defined as "the ability of a genotype to produce distinct phenotypes when exposed to different environments throughout its ontogeny" (Pigliucci, 2001, 2005).

Ecologists strive to explain the diversity of ecologically important phenotypic traits and to understand how these traits are shaped by the environment. Hence, DNA methylation has become of great interest and, as a result, the number of studies on the causes and consequences of DNA methylation in natural populations is rising. Studies that aim to provide insights into the origin of variation in DNA methylation focus on a wide range of environmental influences, including pH (Massicotte & Angers, 2012), habitat quality (Hu et al., 2019), parasites (Hu et al., 2018; McNew et al., 2021; Wenzel & Piertney, 2014), and anthropogenic causes such as urbanization (Caizergues et al., 2021; Garcia et al., 2019; McNew et al., 2017; Riyahi et al., 2015; Watson et al., 2021) and contaminants (Laine et al., 2021; Mäkinen et al., 2021; McNew et al., 2021; Nilsen et al., 2016; Pierron et al., 2014; Romano et al., 2017). In addition, studies have now included DNA methylation changes as a possible mechanism causing phenotypic changes due to environmental experiences during early development, such as brood size (Sepers et al., 2021; Sheldon et al., 2018), diet or resource availability (Laubach et al., 2019; Lea et al., 2016; Weyrich et al., 2018), predation risk (Noguera & Velando, 2019) and parental effects (Bentz et al., 2016; Rubenstein et al., 2016; Weyrich et al., 2016). DNA methylation might thereby explain developmental phenotypic plasticity (Watson et al., 2019), referred to as irreversible phenotypic changes that are the result of environmental induction during development (Forsman, 2015).

Other studies aim to provide insights into the link between epigenetic variation and ecologically relevant phenotypic variation. So far, studies have focused on exploratory behaviour (van Oers et al., 2020; Verhulst et al., 2016), novelty-seeking behaviour (Riyahi et al., 2015), salinity tolerance (Heckwolf et al., 2020), stress resilience (Taff et al., 2019), plumage characteristics (Soulsbury et al., 2018; Taff et al., 2019) and patterns of scutes (Caracappa et al., 2016). Another popular topic in ecological studies is how DNA methylation levels change over time due to, for example, ageing (Ito et al., 2018; Paoli-Iseppi et al., 2019; Parrott et al., 2014; Polanowski et al., 2014; Soulsbury et al., 2018; Thompson et al., 2017; Wilkinson et al., 2021), where methylation might accumulate over time or in the context of regulation of temporal plastic changes, such as regulation of timing of migration and reproduction (Baerwald et al., 2016; Lindner et al., 2021; Mäkinen et al., 2019; Saino et al., 2017; Viitaniemi et al., 2019) and hibernation (Alvarado et al., 2015).

1.2 | Role of epigenetic variation in adaptation to changing environments and evolution

Studies have shown that environmentally induced epigenetics patterns can be transmitted stably to future generations with effects on offspring phenotypes (Anway et al., 2005; Franklin et al., 2010). Thus, epigenetic modifications harbour a source of nongenetic phenotypic variation that potentially can be a fast and heritable response to the environment. If epigenetically mediated phenotypic variation is truly heritable, natural selection might act on it and it might hold evolutionary potential. Evolutionary ecologists now study epigenetic mechanisms to provide insights into the potential role of epigenetic variation in adaptation to changing environments. To assess the evolutionary implications, DNA methylation has been studied in relation to heritability (Hu et al., 2021; van Oers et al., 2020), whether it might be under selection (Laine et al., 2016; Skinner et al., 2014) and whether it might be involved in range expansion and adaptation (Caizergues et al., 2021; Gore et al., 2018; Heckwolf et al., 2020; Liebl et al., 2013; Meröndun et al., 2019; Riyahi et al., 2017; Schrey et al., 2012; Sheldon et al., 2018).

DNA methylation is also studied in historical and ancient samples to characterize environmental and regulatory changes that possibly underlie adaptation and speciation (Gokhman et al., 2017; Orlando et al., 2015; Orlando & Willerslev, 2014; Rubi et al., 2019). However, it is important to note that the heritability of DNA methylation and its role in evolution is still an ongoing point of discussion (see Burggren, 2016; Guerrero-Bosagna et al., 2018; Heard & Martienssen, 2014; Laland et al., 2014; Liberman et al., 2019; Lind & Spagopoulou, 2018; Perez & Lehner, 2019; Richards & Pigliucci, 2020; Sarkies, 2020; Vogt, 2021).

1.3 | Methods to study DNA methylation

There are several methods for detecting DNA methylation (Feng & Lou, 2019; Tang et al., 2015) which are based on distinguishing unmethylated cytosines from methylated cytosines (5mC) in the DNA sequence and the three common principles are: (i) digestion of DNA with methylation-sensitive restriction enzymes, (ii) enrichment of methylated genomic DNA fragments using antimethylcytosine antibody or methyl-binding domain (MBD) proteins, and lastly, (iii) sequencing of bisulfite-converted DNA (BSseq). In this review we will focus on the last, but see Box 1 for alternative and upcoming methods.

Currently, ecological epigenetics is a field without well-established best practices compared to other types of molecular studies (Alvarez et al., 2015). However, studying DNA methylation comes with methodological and technical challenges, especially in vertebrate study systems in an evolutionary or ecological context. For example, sampling is often challenging, especially if samples are needed from different developmental stages. Therefore, blood is commonly used in studies in wild vertebrates, and thus may function as a proxy for patterns in other tissues. Yet if blood cells are not the main target, for example if questions are specifically related to liver function or neurological effects, it is important to establish correlations between DNA methylation patterns (of genes of interest) across tissues in the study species (see, Derks et al., 2016; Lindner et al., 2021; McKay et al., 2011). Furthermore, vertebrates are highly mobile, have relatively long generation times and often only a few offspring per generation, which can make studying epigenetics difficult. Many vertebrates also rely on parental care,

BOX 1 Drawbacks of bisulfite sequencing and alternative methods

The sequencing of bisulfite-converted DNA is a very versatile method for methylation calling at single nucleotide resolution that can be used across taxa, which is especially relevant in the context of wild epigenetics. However, the method and especially the treatment of DNA with bisulfite has some drawbacks. The treatment creates a harsh environment for the DNA, which can lead to the degradation of genomic DNA (Grunau et al., 2001) and affect the quality of the sequencing reads. Furthermore, bisulfite-induced DNA degradation is biased towards genomic regions that are enriched for unmethylated cytosines, which can result in an overestimation of global methylation levels. How the bisulfite treatment affects the quality of sequencing reads and biases the estimation of methylation levels differ between bisulfite treatment protocols (Olova et al., 2018). Recently, methods have been developed to overcome the degradation of DNA in BSseq (Wang et al., 2017). Conversion of nonmethylated cytosines can be achieved without the use of bisulfite, either with (i) enzymes in enzymatic methyl-seq (EM-seq), which uses two enzymatic steps to differentiate between cytosine and its modified forms, 5mC and 5hmC, and shows little DNA degradation (Vaisvila et al., 2021), or (ii) using ten-eleven translocation (TET)-assisted pyridine borane sequencing (TAPS) that has been developed for detecting 5mC (Liu, Siejka-Zielińska, et al., 2019). TAPS uses milder conditions for converting 5mC to thymine compared to bisulfite conversion. However, TAPS has multiple steps of enzymatic and chemical reactions and needs more input DNA than for example EM-seq. Moreover, EM-seq was shown to have higher CpG coverage, a better CpG site overlap between samples and higher consistency in methylation levels across input series compared to WGBS (Vaisvila et al., 2021) and PBAT (Han et al., 2021), making EM-seq the preferred nonbisulfite whole genome method to date.

To overcome many of the challenges that BSseq possesses, the third-generation sequencing technologies also show promising results, such as nanopore sequencing by Oxford Nanopore Technologies (ONT) and single-molecule real-time sequencing (SMRT), such as from Pacific BioSciences, PacBio. These technologies offer in addition to single molecule sequencing in real time also the opportunity to detect DNA methylation from the same data sets without any additional DNA treatment (Flusberg et al., 2010; Gouil & Keniry, 2019; Liu, Fang, et al., 2019; Tse et al., 2021). Although both ONT and SMRT sequencing have their challenges, such as sequencing errors, and possible low detection rate of 5mC in SMRT, these issues can be overcome (Liu et al., 2019; Stoiber et al., 2017; Tse et al., 2021) and thus third-generation technologies are promising for future studies also in wild epigenetics (see Gouil & Keniry, 2019 for method comparison).

Lastly, in some cases a more targeted approach is needed where methylation levels are counted from specific areas of the genome. In both bisulfite pyrosequencing and methylation-specific qPCR, single loci or regions can be analysed, for example in the case of candidate gene studies when the sequence of the region of interest is known (De Chiara et al., 2020). In the microarray-based Infinium Methylation Assay, several thousands probes are designed to target CpGs. This method has been mostly used in humans (Bibikova et al., 2011) but recently a custom-made array was developed for other mammals (Arneson et al., 2021) and was used successfully in bats (Wilkinson et al., 2021). There is also a method called MEBS, which enriches methylated sequences by combining the MBD2 (methyl-binding domain2) protein with bisulfite treatment, and next-generation sequencing (Weyrich et al., 2014). Lastly, MALDI-TOF—mass spectrometry of bisulphite-converted DNA (commercialized as Sequenom's EpiTYPER assay using the MassArray system)—can be also used for more targeted approaches (Thompson et al., 2009).

making it even more challenging to separate true environmentally induced transgenerational epigenetic effects from intra-, inter- and multigenerational effects (Heard & Martienssen, 2014; Skinner, 2008; Skvortsova et al., 2018). It is essential to acknowledge these challenges, since they can have huge effects on the reliability of the results and conclusions. Overall, ecological epigenetics studies differ greatly in their experimental setup, study species, tissue, developmental stage, conditions, sample size, sequencing methods, and the choices made during the bioinformatics and statistical analyses (Husby, 2020; Lea et al., 2017), despite the similarities in research questions. A lack of consistency across studies complicates the interpretation of the obtained results and hinders comparisons across those studies. Published information to guide researchers new to the field of ecological epigenetics is scattered across reviews, research papers and manuals. Papers discussing the whole procedure of DNA methylation from start to end are missing for studies on wild vertebrates. Furthermore, several tips and tricks stem from experience and rely on "personal communication" that have not yet been published.

Therefore, in this technical review, we provide a practical guide to researchers new to the field of ecological epigenetics. This technical review provides an inclusive overview of considerations when designing and executing an evolutionary ecological study involving nonmodel vertebrates and BSseq, in a step-by-step manner. First, we outline the different considerations of designing a study, followed by the bioinformatics steps involved in preparing the data for downstream analyses. Next, we discuss statistical considerations for the analysis of BSseq data, and the validation, interpretation and presentation of the results. Overall, this review offers recommendations focusing on BSseq, its potential biases and future directions.

2 | DESIGNING YOUR BISULFITE SEQUENCING STUDY

2.1 | Before starting your bisulfite project

It is always advisable to start a project with a project and data management plan. Ideally, an experimental setup and design file is created which includes, in addition to the design, the sample names with additional metadata such as treatments, locations and conditions. We also recommend thinking about which technologies and laboratory protocols will be used. This information is often needed for collaborations and during the process of writing the manuscript (at what stage you tend to have forgotten about such details), but also when submitting data files to the repositories. Stage this file in such a way that you can follow the project life cycle from samples all the way to the final results and that you can always get back to this file and update it when needed. It is also crucial to consider the computing resources needed for a specific data set. For genome-wide experiments normal desktop computers are often not sufficient as file sizes can be as big as several hundreds of megabytes or even gigabytes, depending on the genome size of the species and sequencing depth used. Analysing these files will consume a large amount of memory (RAM) and thus high-performance clusters (HPCs) are often needed (Wreczycka et al., 2017).

To make your epigenetic project open and FAIR (Findable, Accessible, Interoperable and Reusable; Wilkinson et al., 2016) and also comparable to other studies, it is important to make both the pipeline including the code used and the (meta)data openly available (Gallagher et al., 2020). While the raw data and general pipeline information are relatively easy to access in the papers and data repositories, the actual code is often not provided even though some journals encourage code-sharing. In past ecological studies code availability has been alarmingly low, which can be an important limiting factor for computational reproducibility in ecology (Culina et al., 2020). In biology in general, code is an important player in analysis pipelines and should be made available to promote openness and FAIRness of the project. Common places for code sharing are Github and Gitlab, which also organize version control of your scripts with git. Zenodo, an open-access repository, can be used for both permanent data and code storage and it also provides citable DOI code. There are several good guidelines for code sharing available (Barnes, 2010; Osborne et al., 2014; Wilson et al., 2017).

2.2 | Sampling

As in designing any scientific study, also in a DNA methylation project, there are some key decisions to ensure that the collected data can actually answer the research questions in both exploratory and experimental study settings. First, taxa differ considerably in the type and extent of DNA methylation variation (reviewed, e.g., in Feng et al., 2010; Zemach et al., 2010), and thus the researcher should be aware of the details of their target species. For example, in many insects DNA methylation levels are low and sporadic, and in plants and fungi DNA methylation regularly occurs outside CpG context (CG dinucleotides) (Feng et al., 2010; de Mendoza et al., 2020; Provataris et al., 2018). In vertebrates, cytosines in CpG context are a major target of DNA methylation (Bernstein et al., 2007; Bird, 2002). However, it seems that non-CpG methylation plays an important role in specific tissues, cell types and developmental states in vertebrates (Ross et al., 2021; Zabet et al., 2017; Ziller et al., 2011). Second, whether a reference genome is available will influence the choice of methods and downstream analysis that can be performed. For a large part of the wild study species a high-quality reference genome is still lacking, yet methodologies have been developed also for species without a reference genome (van Gurp et al., 2016; Klughammer et al., 2015). Third, the timing of sampling needs to be tailored to the study question, for example targeting developing or adult individuals (DNA methylation patterns change over age across organisms; see, e.g. Parrott et al., 2014; Thompson et al., 2017), and to account for seasonal differences in DNA methylation (Viitaniemi et al., 2019). Fourth, given that DNA methylation patterns can differ among tissues (Lindner et al., 2021), the target tissue should also

be chosen appropriately to the study question. Finally, it has been clearly shown in both domestic and captive models, and recently in wild organisms, that the genetic background explains a relatively large proportion of DNA methylation differences (van Oers et al., 2020; Viitaniemi et al., 2019). Therefore, understanding and accounting for any genetic structure and relatedness in the data is crucial (see in-depth discussion in Lea et al., 2017).

2.2.1 | To pool or not to pool?

During study design it is also important to decide whether samples are sequenced individually or pooled before sequencing. Pooling of samples is a cost-effective way of accurately assessing the average methylation level of single CpGs over all samples, without acquiring the DNA methylation level of each individual sample (Docherty et al., 2010). Pooling can also be used when individual sample quantities are small. Using DNA methylation levels calculated as the total number of converted and unconverted reads off all individuals within a pool, however, prohibits assessing variation between individuals. This, therefore, creates uncertainty around the average DNA methylation levels, which potentially leads to false confidence in the average DNA methylation levels. When, for example, certain individuals contribute more in reads to the overall pooled sample than others, the DNA methylation levels of the pool might be biased towards these individuals. Furthermore, answering most contemporary questions in ecology and evolution requires sample sizes exceeding those currently used (Lea et al., 2017). We therefore strongly advise against pooling, as it reduces the power to detect moderate effect size by collapsing the number of biological replicates. This advice is supported by Ziller et al. (2015), who emphasize that biological replicates should be sequenced and analysed separately rather than being pooled to increase power. However, pooling might still be a useful tool for initial exploration of differences between experimental groups or populations, but only if pools are chosen in such a way that relatedness is taken into account (see, van Oers et al., 2020; Sepers et al., 2021) and the variation among pools is minimized. This can be achieved by balancing pools carefully based on factors such as sex, age and starting material.

2.3 | Choosing the methods

2.3.1 | The bisulfite sequencing methods

When treating DNA with bisulfite, cytosines are converted to uracil but 5-methylcytosines are unaffected (Fraga & Esteller, 2002). Thus, DNA that has been treated with bisulfite retains only methylated cytosines. After bisulfite conversion, it is possible to combine PCR with high-throughput sequencing or microarray-based methods to determine the methylation levels of individual CpGs in the samples (also called MethylC-Seq, Figure 1). The two most common approaches for sequencing of bisulfite-converted DNA target the whole genome (whole-genome BSseq, WGBS) or a reduced and biased representation of the genome by using restriction enzymes (reduced representation bisulfite sequencing, RRBS).

WGBS is still widely considered the gold standard for DNA methylation profiling as it captures about 90% of the CpG sites within the

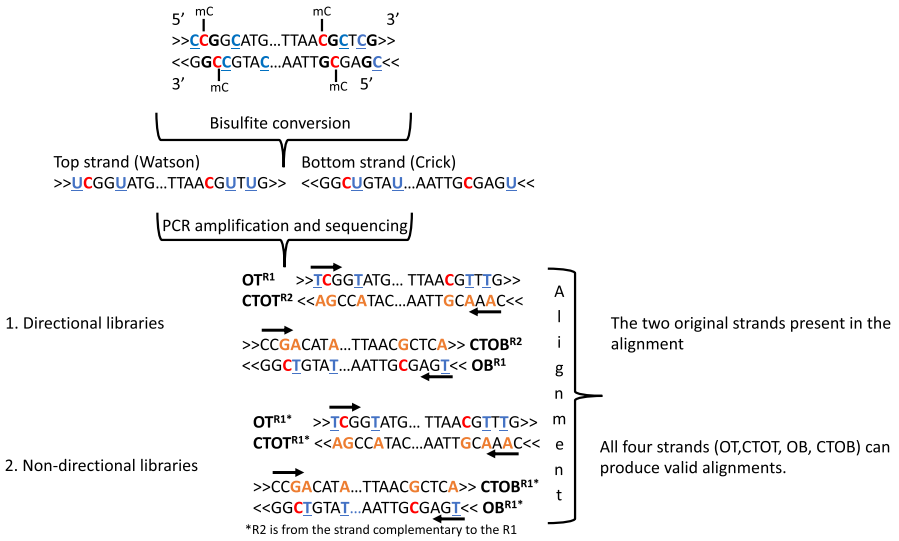

FIGURE 1 Bisulfite conversion of DNA and the PCR amplification results in two PCR products. Unmethylated cytosines (in blue) are converted to uracils and then to thymines, and methylated cytosines (in red) are unaffected. In directional libraries, the first read in paired-end sequencing originates from the original strands. The second read of paired-end sequencing is derived from the complementary strand. In nondirectional methods, the first read can be from any of the four strands and the second read from the strand complementary to the first read. mC, 5-methylcytosine; OT (Watson), original top strand; CTOT, strand complementary to the original top strand; OB (Crick), original bottom strand; and CTOB, strand complementary to the original bottom strand; R1, read 1; R2, read 2; arrows point to the sequencing direction; complementary bases to methylated and unmethylated cytosines are in orange

genome at single base pair resolution (Lister et al., 2009). In addition to traditional BSseq as mentioned above (MethylC-Seq), other methods for WGBS are also available. These include techniques such as "post-bisulfite adaptor tagging" (PBAT; Miura et al., 2012) and tagmentation-based whole-genome bisulfite sequencing (T-WGBS; Wang et al., 2013). PBAT is an amplification-free method, which has shown low degradation bias, insignificant CG-context coverage bias and better matched methylation levels measured by liquid chromatography-tandem mass spectrometry (LC-MS) (Olova et al., 2018). T-WGBS uses Tn5 transposome and bisulfite conversion to study 5mC and is especially suited when starting material is limited (Wang et al., 2013; Weichenhan et al., 2019).

Most current studies on ecological or evolutionary questions require sample sizes that by far exceed the sample sizes used (Lea et al., 2017), highlighting the need for approaches that target a reduced and biased representation of the genome. RRBS (Gu et al., 2011; Meissner et al., 2008), like other reduced representation approaches, makes use of restriction enzymes that nonrandomly cut the DNA at or close to the recognition sequence of the chosen restriction enzyme(s). There is a huge variety in restriction enzymes with different properties concerning, for example, the enzyme cofactor requirements and the nature of their target sequence (Williams, 2003). Depending on the recognition sequence and the frequency of this motif within the genomic sequence, restriction enzymes also differ in how frequently they cut genomic DNA. For example, the restriction enzyme *Msp*I cuts DNA in CG-rich areas often in coding regions and, in this way, RRBS requires a reduced number of reads to obtain a modest coverage of a reproducible fraction of genome-wide CpG sites. These sites are often enriched for promotor regions in vertebrate systems, where CpG methylation is known to affect gene expression (see, Laine et al., 2016). This makes RRBS more cost effective than WGBS and also avoids conducting analyses on hundreds of thousands of CpG sites that are expected to have no functional significance (Sun et al., 2015). Certain studies such as when DNA methylation is studied in other contexts than genes only, however, might require an unbiased representation of DNA methylation at the individual CpG site level across the whole genome, for which WGBS is needed. Furthermore, when looking at differentially methylated regions over sites, more caution is needed in RRBS (see section 4.2.3).

2.3.2 | Sequencing depth

For large-scale projects, there is a trade-off between the number of biological replicates and sequencing depth per replicate when funding is limited. Since both factors are important for statistical power, we need a large sample size to detect small effect sizes and (depending on the sequencing method) a high sequencing depth per replicate to cover CpG sites at sufficient coverage. In general, the sequencing depth of a replicate refers to the total number of sequencing reads per replicate. Often, the average sequencing depth is used, defined as the product of read length (L) and the number of sequencing reads (N) divided by the genome length (G), that is $L \times N/G$ (Sims et al., 2014). Although sequencing depth and coverage are often used interchangeably, the sequencing depth differs from the CpG site coverage, which is defined as the number of unique sequencing reads that cover a certain CpG site. Ziller et al. (2015) provide data-driven guidance on the trade-off between the number of biological replicates and sequencing depth per replicate for high-quality WGBS data sets when a region-based analysis is performed for a between-group comparison; they recommend an average sequencing depth of at least 5–15× per sample depending on the magnitude of DNA methylation differences between groups and the strategy used for identifying regions (smoothing or single CpG site). At 15× depth the fraction of true positives discovered is expected to range up to 70%, while the fraction of false negatives stays below 30%, although using replicates of each sample (i.e., acquiring a depth of 30×) is generally preferred. Ziller et al. (2015) also emphasize that, in terms of power, additional resources would be better spent on increasing the number of replicates rather than sequencing depth. Recently, a tool has been developed, POWEREDBISEQ, that helps to predict the power of specific studies taking in to account read-depth and coverage filtering (Seiler Vellame et al., 2021). This is specifically relevant when trying to assess small effects, when read depth might be limiting, although there is still much debate across research fields whether such small effects are meaningful or not (Breton et al., 2017).

In RRBS, enzymes with different properties and recognition sites might be combined to optimize the number and coverage of CpG sites (or the representation of any other targeted subset of the genome) and the average sequencing depth. The choice of enzymes can be tested *in silico* to optimize the RRBS approach for specific experiments (Fu et al., 2016; Martin-Herranz et al., 2017). The number and coverage of CpG sites and the average sequencing depth not only depend on the choice of restriction enzymes, but also on the number of fragments and the selection of fragment lengths in combination with the chosen restriction enzymes (Fu et al., 2016). Size selection is an important part of the RRBS protocol as it prohibits the sequencing of long and uninformative sequences (e.g., sequences that do not contain CpG sites) and can be performed via cutting the desired fragment length range from a gel (Meissner et al., 2008) or via the use of beads (Boyle et al., 2012).

2.3.3 | Paired-end or single-end?

In both WGBS and RRBS, libraries can be sequenced on single- or paired-end mode (Figure 1). In general, paired-end sequencing offers the potential to better map repeat regions (Grehl et al., 2018) and was reported to reduce error rates and enhance sensitivity (Tsuji & Weng, 2016). Particularly in RRBS approaches where fragment sizes are small, it is important to consider the read length relative to the fragment size to avoid overlap of read-pairs (e.g., when overlap is not removed, a position is covered by the same read-pair twice and counted twice, which results in bias). Yet, in cases where the interest is in assessing strand-specific DNA methylation levels, such as when

studying genomic imprinting (Patiño-Parrado et al., 2017), a paired-end approach is indispensable.

2.3.4 | Choice of library type

For a comprehensive and systematic evaluation of library preparation methods and sequencing platforms for high-throughput BSseq we refer the reader to Olova et al. (2018) and Zhou et al. (2019). One important aspect of bisulfite libraries is directionality (Figure 1). Directional libraries, in which only the original top and bottom strands are sequenced in single-end mode (complementary strands are also covered in paired-end mode), constitute the most common library type. In nondirectional libraries also the strands complementary to original top and bottom strands are sequenced in addition to the original top and bottom strands in single-end mode (Tsuji & Weng, 2016). This means that for nondirectional libraries, alignments to all four strands are required, which can increase processing time and the computational resources needed compared to alignments performed with directional data.

In theory, directional data can be used to prevent the misidentification of C/T substitutions as unmethylated cytosines. Directional library protocols are strand-specific, which means that guanines on the strand opposing a cytosine are not affected by the bisulfite conversion (Krueger et al., 2012). Consequently, reads of a cytosine position that map to the cytosine-strand can be used to quantify the methylation level of that cytosine position, but yield no information on a potential C/T substitution, while reads that map to the guanine-strand do not yield information of the methylation status of that cytosine position but can be used to identify the C/T substitution (Liu et al., 2012).

3 | BISULFITE SEQUENCING BIOINFORMATICS

3.1 | Quality control (QC)

QC is an essential step in the analysis to ensure that the data are of sufficient quality and to determine sample-specific biases. There are three technical biases that should be checked for when evaluating the QC reports. Some quality issues are the result of technical biases that are general in next-generation sequencing data. The first one is sequencing into the 3′-end adapter which occurs when the fragment size is smaller than the read length and the sequencing read continues into the adapter sequence of the opposite fragment end (Krueger et al., 2012). As a result, short read sequences will consist of adapter sequence at the 3′ end, which will negatively affect the mapping efficiency of the data. The second general issue is a decrease in sequencing quality towards the 3′ end due to accumulation of sequencing errors. As these errors can affect mapping efficiency and DNA methylation calling accuracy, it is important to pay attention to the quality scores of the reads and nucleotides during QC (see below for trimming). The third issue is restricted to RRBS data. In RRBS, the fragments are end-repaired after digestion to allow adapter ligation. This introduces either an unmethylated or a methylated cytosine on the 3′ end or both ends of the DNA fragments (Bock, 2012). Usually, unmethylated cytosines are introduced. If the fragment size is smaller than the read length and the sequencing process includes the inserted cytosine position, this will introduce a biased DNA methylation estimation at the read ends. Therefore, it is important to check for the presence of such a bias and, if necessary, to trim the introduced cytosine position off by removing at least two bases. QC will reveal a nucleotide content bias, but a biased DNA methylation estimation will only be visible after methylation calling (see section 3.3).

FASTQC (Andrews, 2010) can be used to check for quality of the reads and adapter content and FASTQ screen (Wingett & Andrews, 2018) in bisulfite mode to detect possible contaminations. The report files can be summarized using MULTIQC (Ewels et al., 2016) (Figure 2). Please be aware that bisulfite treatment affects the results of GC and per-base sequence graphs in FASTQC due to the treatment converting most of the cytosines to thymines. After alignment (see section 3.2), there are tools available for more in-depth inspection of the quality of bisulfite data, such as BSEQC (Lin et al., 2013).

3.1.1 | Estimating bisulfite conversion efficiency

The DNA methylation analysis quality of BSseq-based studies relies on the efficiency of the bisulfite treatment. Two types of errors can arise when aiming at converting unmethylated Cs. First, over-conversion occurs when methylated Cs are inappropriately converted to uracils. This occurs predominantly when the conversion of unmethylated cytosines is almost complete (Genereux et al., 2008), but it is also affected by a range of factors, such as DNA quality and quantity and bisulfite conversion settings (Olova et al., 2018). Over-conversion rates are difficult to measure (Liu et al., 2012), but are typically far below 1% (Genereux et al., 2008). Second, if conversion of unmethylated Cs fails, DNA methylation will be overestimated. Hence, we highly recommend calculation of the bisulfite conversion rate.

In general, two different approaches can be distinguished to estimate the number of unconverted cytosines based on the analysis of (i) native methylation patterns or spike-in controls or (ii) non-native spike-in controls. For vertebrates it is often assumed that Cs outside a CpG context remain always unmethylated (Reik et al., 2003). So, the conversion rate of a sequence can be calculated as the ratio between the number of correctly converted Cs outside a CpG context divided by the sum of converted and unconverted Cs outside a CpG context as done by, for example the BIQ ANALYZER (Bock et al., 2005). However, this assumption is not true as mentioned above in section 2.2. Alternatively, a native spike-in approach was developed for vertebrates that uses telomeric sequences, in which "TTAGGG" is repeated ~3000 times. Its complementary DNA strand contains "CCCTAA" repeats, which have three non-CpG sites (one CpT and

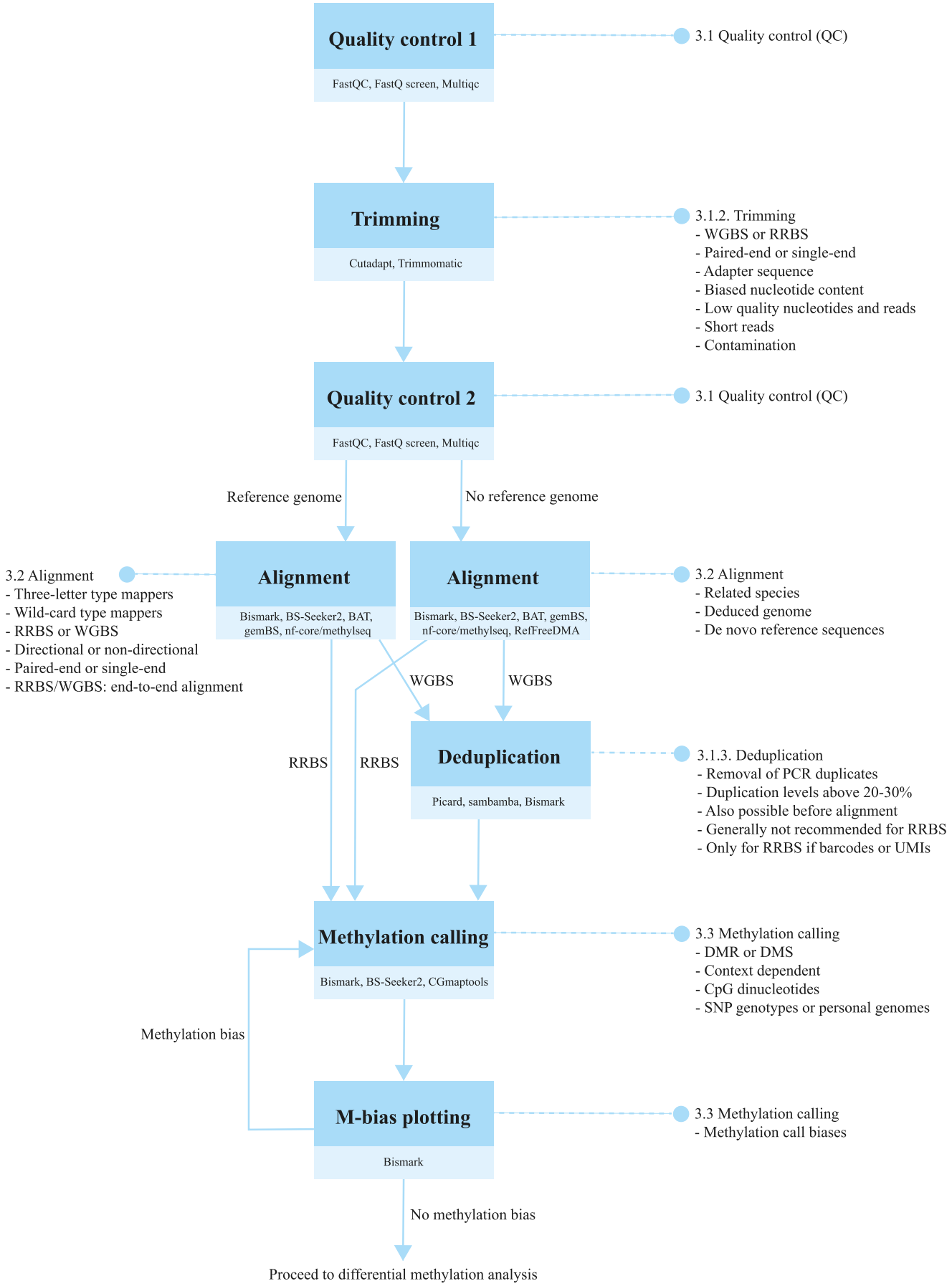
Quality control 1
FastQC, FastQ screen, Multiqc
3.1 Quality control (QC)
Trimming
Cutadapt, Trimmomatic
3.1.2. Trimming
- WGBS or RRBS
- Paired-end or single-end
- Adapter sequence
- Biased nucleotide content
- Low quality nucleotides and reads
- Short reads
- Contamination
Quality control 2
FastQC, FastQ screen, Multiqc
3.1 Quality control (QC)
Reference genome
No reference genome
3.2 Alignment
- Three-letter type mappers
- Wild-card type mappers
- RRBS or WGBS
- Directional or non-directional
- Paired-end or single-end
- RRBS/WGBS: end-to-end alignment
Alignment
Bismark, BS-Seeker2, BAT, gemBS, nf-core/methylseq
Alignment
Bismark, BS-Seeker2, BAT, gemBS, nf-core/methylseq, RefFreeDMA
3.2 Alignment
- Related species
- Deduced genome
- De novo reference sequences
WGBS
WGBS
RRBS
RRBS
Deduplication
Picard, sambamba, Bismark
3.1.3. Deduplication
- Removal of PCR duplicates
- Duplication levels above 20-30%
- Also possible before alignment
- Generally not recommended for RRBS
- Only for RRBS if barcodes or UMIs
Methylation calling
Bismark, BS-Seeker2, CGmaptools
3.3 Methylation calling
- DMR or DMS
- Context dependent
- CpG dinucleotides
- SNP genotypes or personal genomes
Methylation bias
M-bias plotting
Bismark
3.3 Methylation calling
- Methylation call biases
No methylation bias
Proceed to differential methylation analysis

FIGURE 2 The BSseq DNA methylation analysis workflow. Blue boxes indicate main steps, light blue boxes indicate examples of programs for each main step, dotted lines refer to the parts where the main step is discussed and related points that require attention. In general, the first part of the analysis of BSseq sequencing data consists of the following steps: quality control (QC), read trimming, second QC, read mapping, methylation calling, M-bias plotting

two CpC sites) for each repeat. These non-CpG sites are used in BCREVAL to calculate the bisulfite conversion rate in WGBS studies (Zhou et al., 2020). Aligners BS-SEEKER and BS-SEEKER2 also use native methylation patterns to analyse bisulfite conversion, but their approach is based on the existence of two different distribution groups of reads with unconverted cytosine sites: reads with sporadically distributed sites and reads with densely distributed sites, from which the dense group is more common (Chen et al., 2010; Guo et al., 2013). Densely unconverted non-CpG sites can be identified and removed or used to calculate the bisulfite conversion rate.

In the second group of bisulfite conversion rate estimations, non-native DNA with known methylation state is added and sequenced to calculate either incomplete- or over-conversion. The disadvantage of this approach is that such controls might show different conversion properties compared to the DNA that is analysed (Krueger et al., 2012). Instead of non-native DNA, single unmethylated cytosines can also be used. These nucleotides can naturally be part of restriction enzyme overhangs or added to the sequencing adapters at known positions as done in EPIGBS (van Gurp et al., 2016) and EPIGBS2 (Gawehns et al., 2022). In contrast, when methylated cytosines are spiked-in, over-conversion rates can be calculated.

3.1.2 | Trimming

To improve the mapping efficiency, reads should be trimmed for quality (e.g., PHRED score <20) and adapter sequence and short reads (e.g., <20 bp) should be removed. Common trimming programs are TRIMMOMATIC (Bolger et al., 2014) and CUTADAPT (Martin, 2011), which can be used as a standalone program, but also within TRIM GALORE ("Babraham Bioinformatics - Trim Galore!"); a wrapper tool around CUTADAPT and FASTQC (Figure 2). Trimming programs or wrapper tools can offer options to remove the cytosine that was introduced during end-repair as well, such as the --rrbs option in TRIM GALORE. During the trimming steps, it is important to know the data well, because a non-directional mode and paired-end-specific options are often offered.

3.1.3 | Deduplication

High duplication levels of reads might indicate the presence of PCR duplicates, which can arise by excessive PCR amplification. Such PCR duplicates result in a high number of methylation calls and more power to detect statistically significant changes in DNA methylation. However, as the power is artificially inflated for some but not for all reads, a large number of false positives might be picked up and PCR duplicates should be avoided. In WGBS data, duplication levels above 20–30% indicate a high presence of PCR duplicates and deduplication should be applied. In WGBS, deduplication can be performed before or after alignment and it can be done during trimming (Figure 2), with programs such as PRINSEQ (Schmieder & Edwards, 2011) and FASTUNIQ (Xu et al., 2012). PCR duplicates can also be removed after alignment in WGBS using PICARD (Picard Toolkit, 2019), SAMBAMBA (Tarasov et al., 2015) or the *deduplicate_bismark* command in BISMARK (Krueger & Andrews, 2011).

In comparison, RRBS data duplication levels are usually higher (70–90%) due to nonrandom fragmentation of the genome. In RRBS data it is hard to differentiate between PCR and biological duplicates (i.e., overlapping reads) because, as a consequence of the restriction enzyme cut site, the 5′ end of reads always have the same coordinates. Therefore, it is not recommended to use deduplication with RRBS data. PCR duplicates in RRBS data can be removed using barcodes or unique molecular identifiers (UMIs), with for example the --barcode option in BISMARK or UMIBAM (Krueger, 2020). The most commonly used method to account for PCR bias in RRBS is to remove CpG sites that have an extremely high coverage via so-called percentile filtering, where typically those sites that fall within the 0.01 highest percentile of coverage are filtered out with, for example, the *filterByCoverage* command in METHYLKIT (Akalin et al., 2012) (see section 4.). It is good to note that in restriction-site associated DNA sequencing (RADseq) paired-end sequencing can be used for PCR-duplicate removal (Rochette et al., 2019) and this could be explored more for other BSseq methods such as RRBS as well.

3.2 | Alignment

Once the reads are of satisfactory quality, they can be aligned to a reference genome (Figure 2). The bisulfite treatment has converted unmethylated cytosines to uracils (becoming thymines during amplification) and has left methylated cytosines unchanged. Therefore, to account for the C/T conversion, reads are either aligned in a three-letter space or a wild-card algorithm (Lee et al., 2015). In the case of three-letter type mappers, all Cs in the reads and reference genome are converted to Ts and therefore the alignment focuses on the three letters A, G and T. A second index is created by converting Gs to As, which allows the complementary reads to be aligned. In the case of wild-card type mappers (also called four-letter aligners), all Cs of the reference genome are converted to the wild-card letter Y, to which the Cs and Ts in the reads align or the alignment scoring matrix is modified to allow for mismatches between Cs in the reference genome and Ts in the reads (Bock, 2012). There are many alignment programs available of which most have been extensively reviewed (Nunn et al., 2021; Shafi et al., 2017; Sun et al., 2018). The most commonly used alignment tool, especially for paired-end data, is BISMARK (Krueger & Andrews, 2011), which uses BOWTIE2 (Langmead

& Salzberg, 2012) to align the reads in a three-letter space. Other commonly used programs are BS-SEEKER2 (Guo et al., 2013), GEM3 (Marco-Sola et al., 2012) and BSMAP (Xi & Li, 2009). The choice of alignment tool might depend on the study species and the quality of its reference genome (Nunn et al., 2021), and thus we recommend testing several tools before deciding on the final pipeline especially if BSseq studies have not been conducted before on the studied species.

Regardless of the aligner, it is important to differentiate between RRBS and WGBS and to know the directionality of the library. In addition, it is important to know whether the sequencing data are single- or paired-end to ensure that possible overlap between R1 and R2 will be removed (e.g., by using the --no_overlap flag in BISMARK methylation calling). Furthermore, data can be aligned in local or end-to-end (or global) mode. In the case of end-to-end alignment, all read characters will be included during the alignment. Hence, it is important to adapter- and/or quality-trim the reads before alignment in the case of end-to-end alignment. In the case of local alignment, only part of a read has to align and the ends of the read are soft-clipped if they do not align to the reference genome well. This increases the total number of aligned reads, but lowers the accuracy compared to end-to-end alignment as reads may be incorrectly trimmed and mistakenly aligned to mainly repetitive regions (Krueger, 2016). Therefore, only end-to-end alignment is recommended for WGBS and RRBS data.

3.3 | DNA methylation calling

After alignment the DNA methylation levels are called from the obtained alignment files. Often methylation calling is part of the alignment tool such as in BISMARK and BS-SEEKER2 (Figure 2). There are also separate calling programs such as CGMAPTOOLS (Guo et al., 2018), which is made especially for BS-SEEKER2 alignments, and which is not recommended for, for example, BISMARK alignments. Although efforts have been made in unifying file formats such as having SAM/BAM for alignment format, different tools might make subtle changes inside these formats, such as adding more information in the header of SAM files, making the output file not compatible for downstream programs. Trying to integrate results from several tools can be time consuming because of extra efforts in unifying data format potentially creating errors. Thus, we recommend staying in one, well-documented analysis pipeline to do both the alignment and methylation calling, rather than combining different tools from different pipelines.

In general, methylation calls count how many reads are covering every single C, both methylated and unmethylated, after which these Cs are written to an output, based on context (CpG, CHG and CHH). Sometimes, both Cs of each strand at a specific CpG site (CpG dinucleotide) are covered and reported, as BSseq is strand-specific. These can be treated individually in downstream analyses or combined to one site, depending on whether DNA methylation is assumed to be strand-specific. For example, in METHYLKIT it is possible to merge reads on both strands of a CpG dinucleotide with the command *destrand=TRUE*. This will provide higher coverage for each site. However, METHYLKIT advises this for CpG methylation only. We recommend combining the read counts from both strands (Schultz et al., 2012) as DNA methylation between strands should be highly correlated in CpG sites in vertebrates (Sharif & Koseki, 2018).

Both sequencing and base-calling errors, which are common and well characterized in high-throughput sequencing data sets in general (Taub et al., 2010) and in BSseq (Hansen et al., 2012), can affect downstream results. To overcome these issues, a method called methylation bias or M-bias plotting has been developed (Hansen et al., 2012). This allows inspecting the DNA methylation proportion across each possible position in the read, to help in deciding if the data need further read trimming. Methylation levels should not depend on read position, and hence these plots should show a flat horizontal line. However, sometimes increased or decreased DNA methylation is observed at the start or the end of the reads. These parts of the reads should then be trimmed. This M-bias method has been implemented in BISMARK and the trimming step can be added to the methylation calling command.

Lastly, single nucleotide variants, insertions and deletions can cause bias in methylation calling (Wulfridge et al., 2019). If samples have "CG-to-TG" single nucleotide polymorphisms (SNPs) compared to the reference genome, this would be called as an unmethylated CpG dinucleotide in "TG" samples, which would lead to underestimation of DNA methylation levels. Given that almost 80% of SNPs at CpG sites are "CG-to-TG" SNPs (Tomso & Bell, 2003), this constitutes an important error source for methylation calling that can greatly bias the data. This bias can be corrected by providing a list of known SNP genotypes when calling methylation (Krueger et al., 2012; Liu et al., 2012) that will be excluded from further analysis. Alternatively, personal genomes for each sample can be generated, although this is often not feasible in nonmodel systems (Wulfridge et al., 2019). Simultaneous whole genome and BSseq can also provide good individual SNP information, which can be used in the subsequent analyses (see more about SNP calling in BSseq in Lindner et al., 2022).

4 | STATISTICAL CONSIDERATIONS

4.1 | Quality requirements

Once the methylation status of CpGs has been called, the derived methylation levels can be used for subsequent downstream analyses. Several assumptions are made that need to be taken into account before drawing conclusions from the results. First, we assume we are able to call DNA methylation without error variance, and therefore measure the actual DNA methylation level of the site. Although the calling of methylation seems very straightforward with absolute precision, when assessing methylation levels of CpG sites error variance may arise from various causes. For example, during bisulfite conversion, not all nonmethylated cytosines

will be converted and bisulfite will not only target nonmethylated cytosines. In general, the conversion success is high and will ideally range between 98% and 100%, depending on the used library kit and tissue (Holmes et al., 2014). Conversion efficiency, how reliably unmethylated cytosines are converted to uracils, may also be correlated to the conversion specificity, whether bisulfite conversion only targets cytosines and not also other nucleotides. If these are indeed correlated, under optimal bisulfite conversion circumstances, the number of thymines that are falsely converted increases (Holmes et al., 2014), but that typically does not exceed 1%. Another reason for error variance in methylation call rates is the occurrence of sequence errors. Sequence errors not only affect the mappability of the bisulfite converted reads, because with high error rates mapping percentages will go down irrespective of the already lower mapping success due to bisulfite conversion (see Box 1 for more information about bisulfite conversion drawbacks). They also affect the accuracy at which methylation is called (Otto et al., 2012). The last source of variation has to do with the location and characteristics of the CpG site itself. In cases when DNA methylation rates are moderate and when they are situated on CG-island shores and in exons, more errors seem to occur than when DNA methylation rates are high and when the CpGs are situated in other genomic contexts (Sun et al., 2018). This induces biased error rates across the genome.

Second, also the coverage of a CpG site affects the accuracy at which methylation is called (see also section 2.3.2). At 10× coverage, a typical coverage used in ecological studies, the methylation level of a site for a specific individual can only be assessed at a 10% accuracy if methylation calls are 100% accurate. However, it is often unknown what percentage of change is needed to affect gene expression on an organismal level and in some this might be lower than 10%. For example, in great tits (*Parus major*), genetic variation in DNA methylation between two lines selected for opposite levels of exploratory behaviour showed differences in CpG methylation in the promoter region only, as opposed to gene body methylation. These differences, assessed by pyrosequencing, were between 3% and 10% (Verhulst et al., 2016), and are probably not picked up with genome-wide methods with lower accuracy and low sample sizes such as in a study on the same lines by van Oers et al. (2020). At the population level, this issue may be partly solved by increasing the number of samples to assess differences between two or more groups in order to be able to sequence at lower depths. This is also the case when methylation levels are required not at the CpG level, but when average DNA methylation is calculated over larger fractions of the genome, such as at the chromosome level (Crary-Dooley et al., 2017).

A third assumption when analysing DNA methylation levels is that the methylation calls for each CpG site are independent measures. However, there is a clear structure in DNA methylation levels across the genome. In vertebrates, DNA methylation levels in intergenic regions are higher compared to levels within and around genes (see, Laine et al., 2016). Furthermore, the promoter region, but especially the region directly flanking the transcription starting site (TSS), has much lower DNA methylation levels compared to the rest of the genome. Hence, it is unclear whether we can take methylation calls of individual CpG sites as independent measures. This has consequences, because we might ascribe more value to a series of significant sites within one gene than to a single isolated site. On the other hand, taking a more regional approach, such as when condensing several CpG sites into regions to assess if these regions are differentially methylated (DMRs), assumes that a possible functional change, such as gene expression changes, are due to regional changes, rather than single sites. These two types of analyses (see below) should therefore be seen as complementary and dependent on the a priori expectations at which level functional changes are expected.

4.2 | Differential DNA methylation analyses

Differential DNA methylation can be analysed per CpG site (differentially methylated site/cytosine, DMS/DMC) or by combining the sites into regions (DMRs) and there are several analysis packages available to study DNA methylation differences between groups, such as METHYLKIT and DSS (Feng et al., 2014). There are several reviews that provide information on the pros and cons of the different packages and tools that can be used to study differential DNA methylation and other alternatives such as Markov-chain approaches (Carmona et al., 2017; Chatterjee et al., 2012; Zhang et al., 2018) or approaches that aim to minimize coverage issues by assessing DNA methylation at a regional level. For example, smoothing based-approaches such as BSSMOOTH (Hansen et al., 2012) for WGBS and BISEQ (Hebestreit et al., 2013) for RRBS produce region-wide methylation values taking into account both the coverage and location of individual CpG sites. We do, however, recommend that when data sets include structure, such as related or repeated samples, building mixed models in R (R Core Team, 2020) provides the user with the highest flexibility and control.

4.2.1 | Filtering

After methylation calling and before any statistical analyses it is often needed to do some filtering based on coverage levels and DNA methylation percentage. If the samples are suffering, for example, from high levels of PCR bias, it is important to discard bases with very high read coverage as these would create bias in the statistical analyses (often called percentile filtering). Additionally discarding bases with low read coverage (<10×) will help with the power of the statistical tests. For example, METHYLKIT provides descriptive statistics and easy filtering options for the methylation data sets. Typically, 0.1% of the sites with the highest coverage are omitted from the data set. Moreover, many sites show no variation in DNA methylation and are either completely methylated or unmethylated in all cells of all individuals. These sites should also be omitted from further analyses.

4.2.2 | Structure of the data

More and more evidence is accumulating that most of the observed variation in DNA methylation has a genetic origin and relatives will therefore be more similar to each other compared to unrelated individuals (see section 2.2). Before starting to decide on an analysis method, any structure in the individuals used can be explored using PCA plots, provided by METHYLKIT as well. Subsequently one can decide to omit individuals with unusual similarity when analyses assume dependence of the samples. Alternatively, one can include pedigree information using MACAU (Lea et al., 2015) or by creating generalized linear mixed models in R (R Core Team, 2020). In the latter, random effects, such as relatedness, can be included to correct for structure or experimental design. On some occasions it might be necessary to omit outliers, when single individuals show a very different methylation structure compared to the rest of the individuals, which might be due to lower read counts for those individuals, or a lower quality in general.

4.2.3 | Differentially methylated sites vs. regions

A way of correcting for the dependence of methylation levels between neighbouring sites is to use DMRs instead of single CpG sites as the unit for analysis. Differences between individual sites may be small, but if they are constant across larger regions, the statistical power to detect them may be greater when combining them in regions. However, defining these regions can be challenging, because the methylome of vertebrates has a specific structure and DNA methylation is not evenly distributed along the methylome (Bird, 2002). Therefore, also the sequencing method used needs to be taken into account. Especially when using RRBS methods, defining a region needs to be done carefully because not the whole genome is represented and the use of restriction enzymes will generate biases towards certain features of the genome, such as CpG islands (Meissner et al., 2008).

There are two methods for determining DMRs: the use of predefined regions or the definition of regions of differential methylation (DM) (Robinson et al., 2014; Schultz et al., 2012; Shafi et al., 2017; Srivastava et al., 2019). In the first method, methylation variation of CpG sites that are situated in predefined regions or genome features, such as CpG islands, TSS regions, untranslated regions (UTRs) and introns, can be condensed into DMR-specific DNA methylation levels and used for statistical testing. For example, in METHYLKIT it is possible to do regional analyses when providing the predefined regions or by conducting a tiling window analysis, where METHYLKIT summarizes DNA methylation information inside user-specified windows of set length. In the second method, when defining DM regions from the data set, several ways have been described to do this (Robinson et al., 2014; Shafi et al., 2017; Sun et al., 2015). For example, in the package DSS, regions are defined using a combination of *p*-value thresholds, the number of CpG sites and distance between CpG sites and the length of regions (Feng et al., 2014). Machine learning can be also used in defining DMRs which was implemented successfully in the Histogram Of MEthylation (HOME) method (Srivastava et al., 2019). Some algorithms such as BSmooth/bsseq conduct a smoothing step before defining DMRs, which smooth out outlier CpGs and utilize CpGs with a low coverage (Hansen et al., 2012; Shafi et al., 2017).

The danger with all these methods is that comparing the different DMRs defined by different methods could often lead to very different DNA methylation values even when using the same WGBS data. To overcome this an R package called DMRcaller has been developed which implements several methods to detect DMRs between two samples (Catoni et al., 2018). It has been shown that different algorithms are required to compute DMRs and the most appropriate algorithm in each case depends on the sequence context and levels of DNA methylation. Creating DMRs should therefore be done only with specific information about the functionality of the DMRs that are created.

4.2.4 | Significance thresholds

In general, under an appropriate null model, one expects the *p*-value distribution to be approximately uniform, whereas significant effects would show up as an excess of low *p*-values (see, Barton et al., 2013). Deviations from these predictions, such as U-shaped *p*-value distributions, point to issues with model assumptions or with the raw data and should be investigated (see, Mäkinen et al., 2021). Since most ecological studies will not include functional validation steps, analysis strategies will only detect candidates and not causal mechanisms. An important distinction should therefore be made between studies that have an exploratory nature or when specific hypotheses are experimentally tested, especially in how rigid they take the correction for multiple testing (see, Mansell et al., 2019). In studies with an exploratory nature, significance thresholds can be relaxed by using false discovery rates (FDRs; Verhoeven et al., 2005), in order to include more true positive results and to avoid any false negatives. The FDR identifies those sites that most likely meet the chosen criteria, but also assumes a uniform distribution of *p*-values in the differential methylation analysis, a criterium that often is not met. FDR methods are therefore not very suitable for comparing studies (Mansell et al., 2019), but an FDR approach will lead to lists of potential candidate loci that should be verified in subsequent studies. When studies aim at testing hypotheses through experiments, a more stringent testing strategy should be used, such as Bonferroni corrections for the number of CpG sites used in the analysis, to avoid false positives. Bonferroni corrections are very conservative and will probably cause some true differentially methylated sites to be undetected. This effect is often even exaggerated, as not all CpGs tested are truly unrelated, since multiple sites can be affected by single factors (see, Chen & Riggs, 2005).

4.3 | Validation of the results

4.3.1 | Overdispersion

There are different ways in which DNA methylation can be modelled. Since the output of BSseq provides us with the number of methylated and unmethylated Cs at a CpG site, one can investigate whether a site is methylated or not. For example, in plant studies it is common practice first to do a binomial test, while taking into account the nonconversion rate (see, Takuno & Gaut, 2012). If significant methylation is present at a site, one could convert this to a percentage by taking the fraction of methylated Cs. We do, however, strongly advise taking into account the accuracy at which this methylation percentage is calculated. To do so, read depth should be taken into account when conducting statistical analyses. This is done by using the number of methylated Cs over the read depth as the dependent variable in a binomial-type analysis. In general, vertebrate studies use either a logistic regression-type approach or a beta-binomial approach, since these are embedded in most popular packages for analysing differential methylation. A factor that plays an important role here is over- and underdispersion (Wreczycka et al., 2017). Due to the binary nature of bisulfite data in combination with varying coverages for each CpG site, the observed variances often deviate from those calculated by the models. Some packages, such as METHYLKIT, that use a logistic regression approach have a function to correct for overdispersion. Other packages that use a beta-binomial approach are less sensitive to overdispersion, such as DSS. Nevertheless, it is advised to calculate an overdispersion parameter (λ) for each analysis and if needed correct for overdispersion if this deviates greatly from 1. This can be done by using the residuals of the individual statistical models that are used for calculating differential methylation by summing the residuals and dividing that by twice the number of fixed and random factors in the model. Sites that fall outside 95% of the posterior density interval of λ should be removed to correct for this overdispersion (van Oers et al., 2020).

4.3.2 | Graphical validation of the results

An essential step to check whether deduplication has sufficiently removed any bias is plotting the coverage of the sites against their significance value to evaluate whether a coverage-dependent significance is present in coverage plots and QQ plots (Figure 3). The power to assess differential methylation at the level of individual CpG sites relies on the coverage of each of these sites. Results should therefore not be biased towards sites with higher coverage (Shafi et al., 2017), since smaller differences will be assessed as being significant when the number of reads covering a certain site is higher. Therefore, we also urge authors to provide these steps in their papers.

These considerations urge for a rigorous QC, not only of the raw data (see above), but also regarding the outcomes of our statistical models. A way to assess the quality of the analyses lies in presenting both Manhattan and Volcano plots. Manhattan plots show the location of the differentially methylated sites and may indicate the violation of any of the above-mentioned assumptions. A Volcano plot, in which the difference in DNA methylation is plotted against the significance of the site, provides information on the order of the magnitude of the difference in DNA methylation between the samples or groups in relation to its significance. To indicate the extent to which false conclusions could be drawn from a DNA methylation study, Figure 4 shows the Manhattan and Volcano plots of the same WGBS data set from which we have already presented the coverage and QQ plot. Not only are the −log(*p*) values heavily inflated, but also significance is given to sites that only vary very marginally (Figure 4).

5 | CONCLUSIONS

With this practical guide we point towards these best practices and give guidelines for conducting sound ecological end evolutionary epigenetic studies from the study idea until publication. Good experimental planning and good bioinformatics and biostatistics practices are essential for minimizing the many sources of error

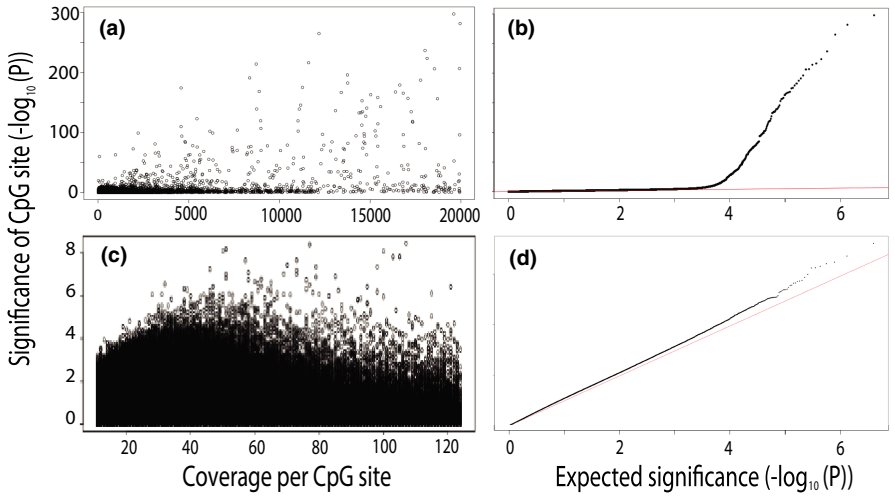

FIGURE 3 Coverage visualizations and QQ plots of a WGBS data set, before (a, b) and after (c, d) coverage filtering and overdispersion correction (data from van Oers et al., 2020). Before coverage filtering a clear association exists between the coverage of a CpG site and its significance level in a differential methylation analysis. After filtering this association disppears. (b, d) QQ plots showing inflated significance values (b) that disappear after filtering (d)

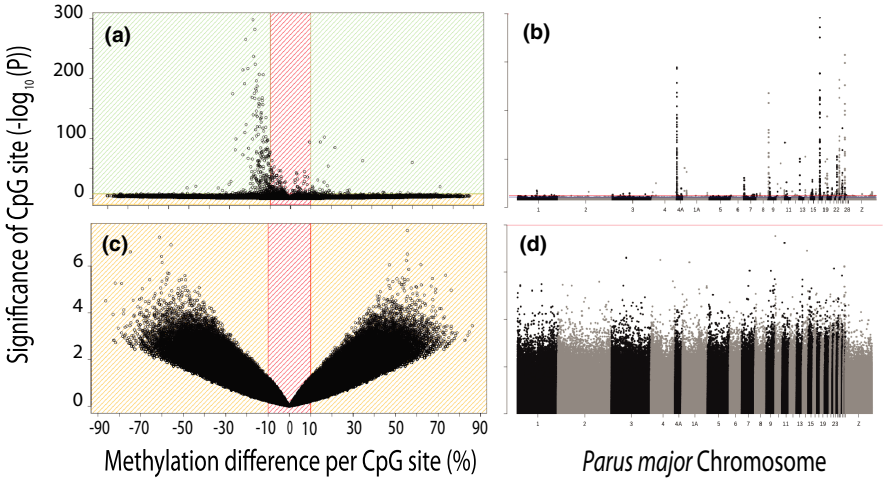

FIGURE 4 Differential methylation visualizations (a, c) and Manhattan plots (b, d) of a WGBS data set, before (a, b) and after (c, d) coverage filtering and overdispersion correction (data from van Oers et al., 2020). Before coverage filtering significance values are inflated (a, b). From the Manhattan plots it is apparent that these CpG sites with inflated significance values are not randomly distributed. These results could therefore be falsely interpreted as DMS or DMR at specific sites on the genome. After filtering no association is significant after Bonferroni correction (c, d)

associated with the use of BSseq in ecological studies. Apart from this, we also urge for unified and validated methods to compare different studies and draw general conclusions. The shortcomings of using bisulfite methods in ecological epigenetics should, however, also be considered. Moreover, although we cannot prove causality in correlational DNA methylation studies, such exploratory studies remain important for ecological epigenetics. An important future challenge will be to increase sample sizes and to replicate studies to draw conclusions on the true impact of DNA methylation in ecology and evolution. Moreover, to draw conclusions on causality, validation studies on nonmodel organisms are required by using common garden approaches or controlled studies.

ACKNOWLEDGEMENTS

We thank Arild Husby, Marcel Visser, Koen Verhoeven, Christa Mateman and the members of the NIOO-KNAW "ecological epigenetics" theme group for useful discussions. An Editor and three anonymous reviewers helped improving the manuscript with very constructive comments. This work was supported by an NWO open competition grant (ALWOP.314) to K.V.O. and B.S. S.R. was supported by Turku Collegium for Science and Medicine.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable—no new data were generated, or the article describes entirely theoretical research.

ORCID

Veronika N. Laine https://orcid.org/0000-0002-4516-7002
Bernice Sepers https://orcid.org/0000-0002-5723-5455
Melanie Lindner https://orcid.org/0000-0003-2931-265X
Fleur Gawehns https://orcid.org/0000-0002-9236-966X
Suvi Ruuskanen https://orcid.org/0000-0001-5582-9455
Kees van Oers https://orcid.org/0000-0001-6984-906X

REFERENCES

Aguirre, N., Filippi, C., Zaina, G., Rivas, J., Acuña, C., Villalba, P., García, M., González, S., Rivarola, M., Martínez, M., Puebla, A., Morgante, M., Hopp, H., Paniego, N., & Marcucci Poltri, S. (2019). Optimizing ddRADseq in non-model species: A case study in *Eucalyptus dunnii* maiden. *Agronomy*, *9*(9), 484. https://doi.org/10.3390/agronomy9090484

Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F. E., Figueroa, M. E., Melnick, A., & Mason, C. E. (2012). methylKit: A comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome Biology*, *13*(10), R87. https://doi.org/10.1186/gb-2012-13-10-r87

Alvarado, S., Mak, T., Liu, S., Storey, K. B., & Szyf, M. (2015). Dynamic changes in global and gene-specific DNA methylation during hibernation in adult thirteen-lined ground squirrels, *Ictidomys tridecemlineatus*. *Journal of Experimental Biology*, *218*(11), 1787–1795. https://doi.org/10.1242/jeb.116046

Alvarez, M., Schrey, A. W., & Richards, C. L. (2015). Ten years of transcriptomics in wild populations: What have we learned about their ecology and evolution? *Molecular Ecology*, *24*(4), 710–725. https://doi.org/10.1111/mec.13055

Andrews, S. (2010). *FastQC: A quality control tool for high throughput sequence data*. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Anway, M. D., Cupp, A. S., Uzumcu, M., & Skinner, M. K. (2005). Epigenetic transgenerational actions of endocrine disruptors and male fertility. *Science*, *308*(5727), 1466–1469. https://doi.org/10.1126/science.1108190

Arneson, A., Haghani, A., Thompson, M. J., Pellegrini, M., Kwon, S. B., Vu, H., Yao, M., Li, C. Z., Lu, A. T., Barnes, B., Hansen, K. D., Zhou, W., Breeze, C. E., Ernst, J., & Horvath, S. (2021). A mammalian methylation array for profiling methylation levels at conserved sequences. BioRxiv, 2021.01.07.425637. https://doi.org/10.1101/2021.01.07.425637

Babraham Bioinformatics—Trim Galore! (n.d.). Retrieved June 17, 2021, from https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

Baerwald, M. R., Meek, M. H., Stephens, M. R., Nagarajan, R. P., Goodbla, A. M., Tomalty, K. M. H., Thorgaard, G. H., May, B., & Nichols, K. M. (2016). Migration-related phenotypic divergence is associated with epigenetic modifications in rainbow trout.

Molecular Ecology, *25*(8), 1785–1800. https://doi.org/10.1111/mec.13231

Barnes, N. (2010). Publish your computer code: It is good enough. *Nature*, *467*(7317), 753. https://doi.org/10.1038/467753a

Barton, S. J., Crozier, S. R., Lillycrop, K. A., Godfrey, K. M., & Inskip, H. M. (2013). Correction of unexpected distributions of P values from analysis of whole genome arrays by rectifying violation of statistical assumptions. *BMC Genomics*, *14*(1), 161. https://doi.org/10.1186/1471-2164-14-161

Becker, C., Hagmann, J., Müller, J., Koenig, D., Stegle, O., Borgwardt, K., & Weigel, D. (2011). Spontaneous epigenetic variation in the *Arabidopsis thaliana* methylome. *Nature*, *480*(7376), 245–249. https://doi.org/10.1038/nature10555

Bentz, A. B., Sirman, A. E., Wada, H., Navara, K. J., & Hood, W. R. (2016). Relationship between maternal environment and DNA methylation patterns of estrogen receptor alpha in wild Eastern Bluebird (*Sialia sialis*) nestlings: A pilot study. *Ecology and Evolution*, *6*(14), 4741–4752. https://doi.org/10.1002/ece3.2162

Bernstein, B. E., Meissner, A., & Lander, E. S. (2007). The Mammalian Epigenome. *Cell*, *128*(4), 669–681. https://doi.org/10.1016/j.cell.2007.01.033

Bibikova, M., Barnes, B., Tsan, C., Ho, V., Klotzle, B., Le, J. M., Delano, D., Zhang, L. U., Schroth, G. P., Gunderson, K. L., Fan, J.-B., & Shen, R. (2011). High density DNA methylation array with single CpG site resolution. *Genomics*, *98*(4), 288–295. https://doi.org/10.1016/j.ygeno.2011.07.007

Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes & Development*, *16*(1), 6–21. https://doi.org/10.1101/gad.947102

Bock, C. (2012). Analysing and interpreting DNA methylation data. *Nature Reviews Genetics*, *13*(10), 705–719. https://doi.org/10.1038/nrg3273

Bock, C., Reither, S., Mikeska, T., Paulsen, M., Walter, J., & Lengauer, T. (2005). BiQ Analyzer: Visualization and quality control for DNA methylation data from bisulfite sequencing. *Bioinformatics*, *21*(21), 4067–4068. https://doi.org/10.1093/bioinformatics/bti652

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bossdorf, O., Richards, C. L., & Pigliucci, M. (2008). Epigenetics for ecologists. *Ecology Letters*, *11*(2), 106–115. https://doi.org/10.1111/j.1461-0248.2007.01130.x

Boyle, P., Clement, K., Gu, H., Smith, Z. D., Ziller, M., Fostel, J. L., Holmes, L., Meldrim, J., Kelley, F., Gnirke, A., & Meissner, A. (2012). Gel-free multiplexed reduced representation bisulfite sequencing for large-scale DNA methylation profiling. *Genome Biology*, *13*(10), R92. https://doi.org/10.1186/gb-2012-13-10-R92

Breton, C. V., Marsit, C. J., Faustman, E., Nadeau, K., Goodrich, J. M., Dolinoy, D. C., Herbstman, J., Holland, N., LaSalle, J. M., Schmidt, R., Yousefi, P., Perera, F., Joubert, B. R., Wiemels, J., Taylor, M., Yang, I. V., Chen, R., Hew, K. M., Freeland, D. M. H., … Murphy, S. K. (2017). Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: The Children's Environmental Health and Disease Prevention Research Center's Epigenetics Working Group. *Environmental Health Perspectives*, *125*(4), 511–526. https://doi.org/10.1289/EHP595

Burggren, W. (2016). Epigenetic inheritance and its role in evolutionary biology: Re-evaluation and new perspectives. *Biology*, *5*(2), 24. https://doi.org/10.3390/biology5020024

Caizergues, A. E., Luyer, J. L., Grégoire, A., Szulkin, M., Senar, J.-C., Charmantier, A., & Perrier, C. (2021). Testing for parallel genomic and epigenomic footprints of adaptation to urban life in a passerine bird. BioRxiv. 2021.02.10.430452. https://doi.org/10.1101/2021.02.10.430452

Caracappa, S., Pisciotta, A., Persichetti, M. F., Caracappa, G., Alduina, R., & Arculeo, M. (2016). Nonmodal scutes patterns in the Loggerhead Sea Turtle (*Caretta caretta*): A possible epigenetic effect? *Canadian Journal of Zoology*, *94*(5), 379-383. https://doi.org/10.1139/cjz-2015-0248

Carmona, J. J., Accomando, W. P., Binder, A. M., Hutchinson, J. N., Pantano, L., Izzi, B., Just, A. C., Lin, X., Schwartz, J., Vokonas, P. S., Amr, S. S., Baccarelli, A. A., & Michels, K. B. (2017). Empirical comparison of reduced representation bisulfite sequencing and Infinium BeadChip reproducibility and coverage of DNA methylation in humans. *Npj Genomic Medicine*, *2*(1), 13. https://doi.org/10.1038/s41525-017-0012-9

Catoni, M., Tsang, J. M., Greco, A. P., & Zabet, N. R. (2018). DMRcaller: A versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and nonCpG contexts. *Nucleic Acids Research*, *46*(19), e114. https://doi.org/10.1093/nar/gky602

Chatterjee, A., Stockwell, P. A., Rodger, E. J., & Morison, I. M. (2012). Comparison of alignment software for genome-wide bisulphite sequence data. *Nucleic Acids Research*, *40*(10), e79. https://doi.org/10.1093/nar/gks150

Chen, P.-Y., Cokus, S. J., & Pellegrini, M. (2010). BS Seeker: Precise mapping for bisulfite sequencing. *BMC Bioinformatics*, *11*(1), 203. https://doi.org/10.1186/1471-2105-11-203

Chen, Z., & Riggs, A. D. (2005). Maintenance and regulation of DNA methylation patterns in mammals. *Biochemistry and Cell Biology*, *83*(4), 438–448. https://doi.org/10.1139/o05-138

Crary-Dooley, F. K., Tam, M. E., Dunaway, K. W., Hertz-Picciotto, I., Schmidt, R. J., & LaSalle, J. M. (2017). A comparison of existing global DNA methylation assays to low-coverage whole-genome bisulfite sequencing for epidemiological studies. *Epigenetics*, *12*(3), 206–214. https://doi.org/10.1080/15592294.2016.1276680

Culina, A., van den Berg, I., Evans, S., & Sánchez-Tójar, A. (2020). Low availability of code in ecology: A call for urgent action. *PLOS Biology*, *18*(7), e3000763. https://doi.org/10.1371/journal.pbio.3000763

De Chiara, L., Leiro-Fernandez, V., Rodríguez-Girondo, M., Valverde, D., Botana-Rial, M. I., & Fernández-Villar, A. (2020). Comparison of bisulfite pyrosequencing and methylation-specific qPCR for methylation assessment. *International Journal of Molecular Sciences*, *21*(23), 9242. https://doi.org/10.3390/ijms21239242

de Mendoza, A., Lister, R., & Bogdanovic, O. (2020). Evolution of DNA Methylome Diversity in Eukaryotes. *Journal of Molecular Biology*, *S0022-2836*(19), 30659-X. https://doi.org/10.1016/j.jmb.2019.11.003

Derks, M. F. L., Schachtschneider, K. M., Madsen, O., Schijlen, E., Verhoeven, K. J. F., & van Oers, K. (2016). Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genomics*, *17*(1), 332. https://doi.org/10.1186/s12864-016-2653-y

Docherty, S. J., Davis, O. S. P., Haworth, C. M. A., Plomin, R., & Mill, J. (2010). DNA methylation profiling using bisulfite-based epityping of pooled genomic DNA. *Methods*, *52*(3), 255–258. https://doi.org/10.1016/j.ymeth.2010.06.017

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Feng, H., Conneely, K. N., & Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Research*, *42*(8), e69. https://doi.org/10.1093/nar/gku154

Feng, L., & Lou, J. (2019). DNA methylation analysis. In Q. Zhang (Ed.), *Nanotoxicity* (pp. 181–227). Springer New York. https://doi.org/10.1007/978-1-4939-8916-4_12

Feng, S., Cokus, S. J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M. G., Hetzel, J., Jain, J., Strauss, S. H., Halpern, M. E., Ukomadu, C., Sadler, K. C., Pradhan, S., Pellegrini, M., & Jacobsen, S. E. (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*(19), 8689–8694. https://doi.org/10.1073/pnas.1002720107

Flusberg, B. A., Webster, D. R., Lee, J. H., Travers, K. J., Olivares, E. C., Clark, T. A., Korlach, J., & Turner, S. W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods*, *7*(6), 461–465. https://doi.org/10.1038/nmeth.1459

Forsman, A. (2015). Rethinking phenotypic plasticity and its consequences for individuals, populations and species. *Heredity*, *115*(4), 276–284. https://doi.org/10.1038/hdy.2014.92

Fraga, M. F., & Esteller, M. (2002). DNA methylation: A profile of methods and applications. *BioTechniques*, *33*(3), 632, 634, 636–649. https://doi.org/10.2144/02333rv01

Franklin, T. B., Russig, H., Weiss, I. C., Gräff, J., Linder, N., Michalon, A., Vizi, S., & Mansuy, I. M. (2010). Epigenetic transmission of the impact of early stress across generations. *Biological Psychiatry*, *68*(5), 408–415. https://doi.org/10.1016/j.biopsych.2010.05.036

Fu, Y.-B., Peterson, G. W., & Dong, Y. (2016). Increasing genome sampling and improving SNP genotyping for genotyping-by-sequencing with new combinations of restriction enzymes. *G3: Genes|genomes|genetics*, *6*(4), 845–856. https://doi.org/10.1534/g3.115.025775

Gallagher, R. V., Falster, D. S., Maitner, B. S., Salguero-Gómez, R., Vandvik, V., Pearse, W. D., Schneider, F. D., Kattge, J., Poelen, J. H., Madin, J. S., Ankenbrand, M. J., Penone, C., Feng, X., Adams, V. M., Alroy, J., Andrew, S. C., Balk, M. A., Bland, L. M., Boyle, B. L., … Enquist, B. J. (2020). Open Science principles for accelerating trait-based science across the Tree of Life. *Nature Ecology & Evolution*, *4*(3), 294–303. https://doi.org/10.1038/s41559-020-1109-6

Garcia, M. J., Rodríguez-Brenes, S., Kobisk, A., Adler, L., Ryan, M. J., Taylor, R. C., & Hunter, K. L. (2019). Epigenomic changes in the túngara frog (*Physalaemus pustulosus*): Possible effects of introduced fungal pathogen and urbanization. *Evolutionary Ecology*, *33*(5), 671–686. https://doi.org/10.1007/s10682-019-10001-8

Gawehns, F., Postuma, M., Antro, M. V., Nunn, A., Sepers, B., van Gurp, T. P., Wagemaker, N. C. A. M., Mateman, A. C., Milanovic-Ivanovic, S., Groβe, I., van Oers, K., Vergeer, P., & Verhoeven, K. J. F. (2022). epiGBS2: Improvements and evaluation of highly multiplexed, epiGBS-based reduced representation bisulfite sequencing. *Molecular Ecology Resources*, 1–8. https://doi.org/10.1111/1755-0998.13597

Genereux, D. P., Johnson, W. C., Burden, A. F., Stoger, R., & Laird, C. D. (2008). Errors in the bisulfite conversion of DNA: Modulating inappropriate- and failed-conversion frequencies. *Nucleic Acids Research*, *36*(22), e150. https://doi.org/10.1093/nar/gkn691

Gokhman, D., Malul, A., & Carmel, L. (2017). Inferring past environments from ancient epigenomes. *Molecular Biology and Evolution*, *34*(10), 2429–2438. https://doi.org/10.1093/molbev/msx211

Goldberg, A. D., Allis, C. D., & Bernstein, E. (2007). Epigenetics: A landscape takes shape. *Cell*, *128*(4), 635–638. https://doi.org/10.1016/j.cell.2007.02.006

Gore, A. V., Tomins, K. A., Iben, J., Ma, L. I., Castranova, D., Davis, A. E., Parkhurst, A., Jeffery, W. R., & Weinstein, B. M. (2018). An epigenetic mechanism for cavefish eye degeneration. *Nature Ecology & Evolution*, *2*(7), 1155–1160. https://doi.org/10.1038/s41559-018-0569-4

Gouil, Q., & Keniry, A. (2019). Latest techniques to study DNA methylation. *Essays in Biochemistry*, *63*(6), 639–648. https://doi.org/10.1042/EBC20190027

Grehl, C., Kuhlmann, M., Becker, C., Glaser, B., & Grosse, I. (2018). How to design a whole-genome bisulfite sequencing experiment. *Epigenomes*, *2*(4), 21. https://doi.org/10.3390/epigenomes2040021

Grunau, C., Clark, S. J., & Rosenthal, A. (2001). Bisulfite genomic sequencing: Systematic investigation of critical experimental parameters. *Nucleic Acids Research*, *29*(13), e65. https://doi.org/10.1093/nar/29.13.e65

Gu, H., Smith, Z. D., Bock, C., Boyle, P., Gnirke, A., & Meissner, A. (2011). Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols*, *6*(4), 468–481. https://doi.org/10.1038/nprot.2010.190

Guerrero-Bosagna, C., Morisson, M., Liaubet, L., Rodenburg, T. B., de Haas, E. N., Košťál, Ľ., & Pitel, F. (2018). Transgenerational epigenetic inheritance in birds. Environmental. *Epigenetics*, *4*(2), https://doi.org/10.1093/eep/dvy008

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M. Q., Chen, P.-Y., & Pellegrini, M. (2013). BS-Seeker2: A versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*, *14*(1), 774. https://doi.org/10.1186/1471-2164-14-774

Guo, W., Zhu, P., Pellegrini, M., Zhang, M. Q., Wang, X., & Ni, Z. (2018). CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. *Bioinformatics*, *34*(3), 381–387. https://doi.org/10.1093/bioinformatics/btx595

Han, Y., Zheleznyakova, G. Y., Marincevic-Zuniga, Y., Kakhki, M. P., Raine, A., Needhamsen, M., & Jagodic, M. (2021). Comparison of EM-seq and PBAT methylome library methods for low-input DNA. *Epigenetics*, *1–10*, https://doi.org/10.1080/15592294.2021.1997406

Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: From whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology*, *13*(10), R83. https://doi.org/10.1186/gb-2012-13-10-r83

Hawes, N. A., Fidler, A. E., Tremblay, L. A., Pochon, X., Dunphy, B. J., & Smith, K. F. (2018). Understanding the role of DNA methylation in successful biological invasions: A review. *Biological Invasions*, *20*(9), 2285–2300. https://doi.org/10.1007/s10530-018-1703-6

Heard, E., & Martienssen, R. A. (2014). Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell*, *157*(1), 95–109. https://doi.org/10.1016/j.cell.2014.02.045

Hebestreit, K., Dugas, M., & Klein, H.-U. (2013). Detection of significantly differentially methylated regions in targeted bisulfite sequencing data. *Bioinformatics*, *29*(13), 1647–1653. https://doi.org/10.1093/bioinformatics/btt263

Heckwolf, M. J., Meyer, B. S., Häsler, R., Höppner, M. P., Eizaguirre, C., & Reusch, T. B. H. (2020). Two different epigenetic information channels in wild three-spined sticklebacks are involved in salinity adaptation. *Science Advances*, *6*(12), eaaz1138. https://doi.org/10.1126/sciadv.aaz1138

Holmes, E. E., Jung, M., Meller, S., Leisse, A., Sailer, V., Zech, J., Mengdehl, M., Garbe, L.-A., Uhl, B., Kristiansen, G., & Dietrich, D. (2014). Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS One*, *9*(4), 15. https://doi.org/10.1371/journal.pone.0093933

Hu, J., Askary, A. M., Thurman, T. J., Spiller, D. A., Palmer, T. M., Pringle, R. M., & Barrett, R. D. H. (2019). The epigenetic signature of colonizing new environments in anolis lizards. *Molecular Biology and Evolution*, *36*(10), 2165–2170. https://doi.org/10.1093/molbev/msz133

Hu, J., Pérez-Jvostov, F., Blondel, L., & Barrett, R. D. H. (2018). Genome-wide DNA methylation signatures of infection status in Trinidadian guppies (*Poecilia reticulata*). *Molecular Ecology*, *27*(15), 3087–3102. https://doi.org/10.1111/mec.14771

Hu, J., Wuitchik, S. J. S., Barry, T. N., Jamniczky, H. A., Rogers, S. M., & Barrett, R. D. H. (2021). Heritability of DNA methylation in threespine stickleback (*Gasterosteus aculeatus*). *Genetics*, *217*(1), 1–15. https://doi.org/10.1093/genetics/iyab001

Husby, A. (2020). On the use of blood samples for measuring DNA methylation in ecological epigenetic studies. *Integrative and Comparative Biology*, *60*(6), 1558–1566. https://doi.org/10.1093/icb/icaa123

Ito, H., Udono, T., Hirata, S., & Inoue-Murayama, M. (2018). Estimation of chimpanzee age based on DNA methylation. *Scientific Reports*, *8*(1), 9998. https://doi.org/10.1038/s41598-018-28318-9

Jablonka, E., & Lamb, M. J. (2007). Précis of evolution in four dimensions. *Behavioral and Brain Sciences*, *30*(4), 353–365. https://doi.org/10.1017/S0140525X07002221

Kilvitis, H. J., Alvarez, M., Foust, C. M., Schrey, A. W., Robertson, M., & Richards, C. L. (2014). Ecological epigenetics. In C. R. Landry, & N. Aubin-Horth (Eds.), *Ecological genomics* (pp. 191–210). Springer Netherlands. https://doi.org/10.1007/978-94-007-7347-9_10

Klughammer, J., Datlinger, P., Printz, D., Sheffield, N. C., Farlik, M., Hadler, J., Fritsch, G., & Bock, C. (2015). Differential DNA methylation analysis without a reference genome. *Cell Reports*, *13*(11), 2621–2633. https://doi.org/10.1016/j.celrep.2015.11.024

Korochkin, L. I. (2006). What is epigenetics. *Russian Journal of Genetics*, *42*(9), 958–965. https://doi.org/10.1134/S102279540609002X

Krueger, F. (2016). *QC Fail Sequencing » Soft-clipping of reads may add potentially unwanted alignments to repetitive regions*. Retrieved June 17, 2021, from https://sequencing.qcfail.com/articles/soft-clipping-of-reads-may-add-potentially-unwanted-alignments-to-repetitive-regions/

Krueger, F. (2020). *Umi-Grinder [Perl]*. Retrieved from https://github.com/FelixKrueger/Umi-Grinder

Krueger, F., & Andrews, S. R. (2011). Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*, *27*(11), 1571–1572. https://doi.org/10.1093/bioinformatics/btr167

Krueger, F., Kreck, B., Franke, A., & Andrews, S. R. (2012). DNA methylome analysis using short bisulfite sequencing data. *Nature Methods*, *9*(2), 145–151. https://doi.org/10.1038/nmeth.1828

Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., de Jager, V., Megens, H.-J., Warren, W. C., Minx, P., Crooijmans, R. P. M. A., Corcoran, P., Sheldon, B. C., Slate, J., Zeng, K., van Oers, K., Visser, M. E., & Groenen, M. A. M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, *7*(1), 10474. https://doi.org/10.1038/ncomms10474

Laine, V. N., Verschuuren, M., van Oers, K., Espín, S., Sánchez-Virosta, P., Eeva, T., & Ruuskanen, S. (2021). Does arsenic contamination affect DNA methylation patterns in a wild bird population? An experimental approach. *Environmental Science & Technology*, *55*(13), 8947–8954. https://doi.org/10.1021/acs.est.0c08621

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G. A., Hoekstra, H. E., Futuyma, D. J., Lenski, R. E., Mackay, T. F. C., Schluter, D., & Strassmann, J. E. (2014). Does evolutionary theory need a rethink? *Nature*, *514*(7521), 161–164. https://doi.org/10.1038/514161a

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Laubach, Z. M., Faulk, C. D., Dolinoy, D. C., Montrose, L., Jones, T. R., Ray, D., Pioon, M. O., & Holekamp, K. E. (2019). Early life social and ecological determinants of global DNA methylation in wild spotted hyenas. *Molecular Ecology*, *28*(16), 3799–3812. https://doi.org/10.1111/mec.15174

Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, *11*(3), 204–220. https://doi.org/10.1038/nrg2719

Lea, A. J., Altmann, J., Alberts, S. C., & Tung, J. (2016). Resource base influences genome-wide DNA methylation levels in wild baboons (*Papio cynocephalus*). *Molecular Ecology*, *25*(8), 1681–1696. https://doi.org/10.1111/mec.13436

Lea, A. J., Tung, J., & Zhou, X. (2015). A Flexible, efficient binomial mixed model for identifying differential DNA methylation in bisulfite sequencing data. *PLOS Genetics*, *11*(11), e1005650. https://doi.org/10.1371/journal.pgen.1005650

Lea, A. J., Vilgalys, T. P., Durst, P. A. P., & Tung, J. (2017). Maximizing ecological and evolutionary insight in bisulfite sequencing data sets. *Nature Ecology & Evolution*, *1*(8), 1074–1083. https://doi.org/10.1038/s41559-017-0229-0

Ledón-Rettig, C. C., Richards, C. L., & Martin, L. B. (2013). Epigenetics for behavioral ecologists. *Behavioral Ecology*, *24*(2), 311–324. https://doi.org/10.1093/beheco/ars145

Lee, J.-H., Park, S.-J., & Kenta, N. (2015). An integrative approach for efficient analysis of whole genome bisulfite sequencing data. *BMC Genomics*, *16*(Suppl 12), S14. https://doi.org/10.1186/1471-2164-16-S12-S14

Li, Q., Li, N., Hu, X., Li, J., Du, Z., Chen, L. I., Yin, G., Duan, J., Zhang, H., Zhao, Y., Wang, J., & Li, N. (2011). Genome-wide mapping of DNA methylation in chicken. *PLoS One*, *6*(5), e19428. https://doi.org/10.1371/journal.pone.0019428

Liberman, N., Wang, S. Y., & Greer, E. L. (2019). Transgenerational epigenetic inheritance: From phenomena to molecular mechanisms. *Current Opinion in Neurobiology*, *59*, 189–206. https://doi.org/10.1016/j.conb.2019.09.012

Liebl, A. L., Schrey, A. W., Richards, C. L., & Martin, L. B. (2013). Patterns of DNA methylation throughout a range expansion of an introduced songbird. *Integrative and Comparative Biology*, *53*(2), 351–358. https://doi.org/10.1093/icb/ict007

Lin, X., Sun, D., Rodriguez, B., Zhao, Q., Sun, H., Zhang, Y., & Li, W. (2013). BSeQC: Quality control of bisulfite sequencing experiments. *Bioinformatics*, *29*(24), 3227–3229. https://doi.org/10.1093/bioinformatics/btt548

Lind, M. I., & Spagopoulou, F. (2018). Evolutionary consequences of epigenetic inheritance. *Heredity*, *121*(3), 205–209. https://doi.org/10.1038/s41437-018-0113-y

Lindner, M., Gawehns, F., te Molder, S., Visser, M. E., van Oers, K., & Laine, V. N. (2022). Performance of methods to detect genetic variants from bisulfite sequencing data in a non-model species. *Molecular Ecology Resources*, *22*, 834–846. https://doi.org/10.1111/1755-0998.13493

Lindner, M., Verhagen, I., Viitaniemi, H. M., Laine, V. N., Visser, M. E., Husby, A., & van Oers, K. (2021). Temporal changes in DNA methylation and RNA expression in a small song bird: Within- and between-tissue comparisons. *BMC Genomics*, *22*(1), 36. https://doi.org/10.1186/s12864-020-07329-9

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q.-M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., & Ecker, J. R. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*, *462*(7271), 315–322. https://doi.org/10.1038/nature08514

Liu, Q., Fang, L., Yu, G., Wang, D., Xiao, C.-L., & Wang, K. (2019). Detection of DNA base modifications by deep recurrent neural network on Oxford Nanopore sequencing data. *Nature Communications*, *10*, https://doi.org/10.1038/s41467-019-10168-2

Liu, Q., Georgieva, D. C., Egli, D., & Wang, K. (2019). NanoMod: A computational tool to detect DNA modifications using Nanopore long-read sequencing data. *BMC Genomics*, *20*(Suppl 1). https://doi.org/10.1186/s12864-018-5372-8

Liu, Y., Siegmund, K. D., Laird, P. W., & Berman, B. P. (2012). Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biology*, *13*(7), R61. https://doi.org/10.1186/gb-2012-13-7-r61

Liu, Y., Siejka-Zielińska, P., Velikova, G., Bi, Y., Yuan, F., Tomkova, M., Bai, C., Chen, L., Schuster-Böckler, B., & Song, C.-X. (2019). Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nature Biotechnology*, *37*(4), 424–429. https://doi.org/10.1038/s41587-019-0041-2

Mäkinen, H., van Oers, K., Eeva, T., & Ruuskanen, S. (2021). The effect of experimental lead pollution on DNA methylation in a wild bird

population. *Epigenetics*, 1–17. https://doi.org/10.1080/15592294.2021.1943863

Mäkinen, H., Viitaniemi, H. M., Visser, M. E., Verhagen, I., van Oers, K., & Husby, A. (2019). Temporally replicated DNA methylation patterns in great tit using reduced representation bisulfite sequencing. *Scientific Data*, *6*(1), 136. https://doi.org/10.1038/s41597-019-0136-0

Mansell, G., Gorrie-Stone, T. J., Bao, Y., Kumari, M., Schalkwyk, L. S., Mill, J., & Hannon, E. (2019). Guidance for DNA methylation studies: Statistical insights from the Illumina EPIC array. *BMC Genomics*, *20*(1), 366. https://doi.org/10.1186/s12864-019-5761-7

Marco-Sola, S., Sammeth, M., Guigó, R., & Ribeca, P. (2012). The GEM mapper: Fast, accurate and versatile alignment by filtration. *Nature Methods*, *9*(12), 1185–1188. https://doi.org/10.1038/nmeth.2221

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12. https://doi.org/10.14806/ej.17.1.200

Martin-Herranz, D. E., Ribeiro, A. J. M., Krueger, F., Thornton, J. M., Reik, W., & Stubbs, T. M. (2017). cuRRBS: Simple and robust evaluation of enzyme combinations for reduced representation approaches. *Nucleic Acids Research*, *45*(20), 11559–11569. https://doi.org/10.1093/nar/gkx814

Massicotte, R., & Angers, B. (2012). General-purpose genotype or how epigenetics extend the flexibility of a genotype. *Genetics Research International*, *2012*, 1–7. https://doi.org/10.1155/2012/317175

McKay, J. A., Xie, L., Harris, S., Wong, Y. K., Ford, D., & Mathers, J. C. (2011). Blood as a surrogate marker for tissue-specific DNA methylation and changes due to folate depletion in post-partum female mice. *Molecular Nutrition & Food Research*, *55*(7), 1026–1035. https://doi.org/10.1002/mnfr.201100008

McNew, S. M., Beck, D., Sadler-Riggleman, I., Knutie, S. A., Koop, J. A. H., Clayton, D. H., & Skinner, M. K. (2017). Epigenetic variation between urban and rural populations of Darwin's finches. *BMC Evolutionary Biology*, *17*(1), 183. https://doi.org/10.1186/s12862-017-1025-9

McNew, S. M., Boquete, M. T., Espinoza-Ulloa, S., Andres, J. A., Wagemaker, N. C. A. M., Knutie, S. A., Richards, C. L., & Clayton, D. H. (2021). Epigenetic effects of parasites and pesticides on captive and wild nestling birds. *Ecology and Evolution*, https://doi.org/10.1002/ece3.7606

Meissner, A., Mikkelsen, T. S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., & Lander, E. S. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature*, *454*(7205), 766–770. https://doi.org/10.1038/nature07107

Meröndun, J., Murray, D. L., & Shafer, A. B. A. (2019). Genome-scale sampling suggests cryptic epigenetic structuring and insular divergence in Canada lynx. *Molecular Ecology*, *28*(13), 3186–3196. https://doi.org/10.1111/mec.15131

Miura, F., Enomoto, Y., Dairiki, R., & Ito, T. (2012). Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Research*, *40*(17), e136. https://doi.org/10.1093/nar/gks454

Moore, L. D., Le, T., & Fan, G. (2013). DNA Methylation and its basic function. *Neuropsychopharmacology*, *38*(1), 23–38. https://doi.org/10.1038/npp.2012.112

Nilsen, F. M., Parrott, B. B., Bowden, J. A., Kassim, B. L., Somerville, S. E., Bryan, T. A., Bryan, C. E., Lange, T. R., Delaney, J. P., Brunell, A. M., Long, S. E., & Guillette, L. J. (2016). Global DNA methylation loss associated with mercury contamination and aging in the American alligator (*Alligator mississippiensis*). *Science of the Total Environment*, *545–546*, 389–397. https://doi.org/10.1016/j.scitotenv.2015.12.059

Noguera, J. C., & Velando, A. (2019). Bird embryos perceive vibratory cues of predation risk from clutch mates. *Nature Ecology & Evolution*, *3*(8), 1225–1232. https://doi.org/10.1038/s41559-019-0929-8

Nunn, A., Otto, C., Stadler, P. F., & Langenberger, D. (2021). Comprehensive benchmarking of software for mapping whole genome bisulfite data: From read alignment to DNA methylation analysis. *Briefings in Bioinformatics*, *22*, bbab021. https://doi.org/10.1093/bib/bbab021

Olova, N., Krueger, F., Andrews, S., Oxley, D., Berrens, R. V., Branco, M. R., & Reik, W. (2018). Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biology*, *19*(1), 33. https://doi.org/10.1186/s13059-018-1408-2

Orlando, L., Gilbert, M. T. P., & Willerslev, E. (2015). Reconstructing ancient genomes and epigenomes. *Nature Reviews Genetics*, *16*(7), 395–408. https://doi.org/10.1038/nrg3935

Orlando, L., & Willerslev, E. (2014). An epigenetic window into the past? *Science*, *345*(6196), 511–512. https://doi.org/10.1126/science.1256515

Osborne, J. M., Bernabeu, M. O., Bruna, M., Calderhead, B., Cooper, J., Dalchau, N., Dunn, S.-J., Fletcher, A. G., Freeman, R., Groen, D., Knapp, B., McInerny, G. J., Mirams, G. R., Pitt-Francis, J., Sengupta, B., Wright, D. W., Yates, C. A., Gavaghan, D. J., Emmott, S., & Deane, C. (2014). Ten simple rules for effective computational research. *PLOS Computational Biology*, *10*(3), e1003506. https://doi.org/10.1371/journal.pcbi.1003506

Otto, C., Stadler, P. F., & Hoffmann, S. (2012). Fast and sensitive mapping of bisulfite-treated sequencing data. *Bioinformatics*, *28*(13), 1698–1704. https://doi.org/10.1093/bioinformatics/bts254

Paoli-Iseppi, R. D., Deagle, B. E., Polanowski, A. M., McMahon, C. R., Dickinson, J. L., Hindell, M. A., & Jarman, S. N. (2019). Age estimation in a long-lived seabird (*Ardenna tenuirostris*) using DNA methylation-based biomarkers. *Molecular Ecology Resources*, *19*(2), 411–425. https://doi.org/10.1111/1755-0998.12981

Parrott, B. B., Bowden, J. A., Kohno, S., Cloy-McCoy, J. A., Hale, M. D., Bangma, J. T., Rainwater, T. R., Wilkinson, P. M., Kucklick, J. R., & Guillette, L. J. (2014). Influence of tissue, age, and environmental quality on DNA methylation in Alligator mississippiensis. *Reproduction*, *147*(4), 503–513. https://doi.org/10.1530/REP-13-0498

Patiño-Parrado, I., Gómez-Jiménez, Á., López-Sánchez, N., & Frade, J. M. (2017). Strand-specific CpG hemimethylation, a novel epigenetic modification functional for genomic imprinting. *Nucleic Acids Research*, *45*(15), 8822–8834. https://doi.org/10.1093/nar/gkx518

Perez, M. F., & Lehner, B. (2019). Intergenerational and transgenerational epigenetic inheritance in animals. *Nature Cell Biology*, *21*(2), 143–151. https://doi.org/10.1038/s41556-018-0242-9

Pértille, F., Brantsæter, M., Nordgreen, J., Coutinho, L. L., Janczak, A. M., Jensen, P., & Guerrero-Bosagna, C. (2017). DNA methylation profiles in red blood cells of adult hens correlate with their rearing conditions. *Journal of Experimental Biology*, *220*(19), 3579–3587. https://doi.org/10.1242/jeb.157891

Picard toolkit [Java] (2019). *Broad Institute*. Retrieved from https://github.com/broadinstitute/picard

Pierron, F., Bureau du Colombier, S., Moffett, A., Caron, A., Peluhet, L., Daffe, G., Lambert, P., Elie, P., Labadie, P., Budzinski, H., Dufour, S., Couture, P., & Baudrimont, M. (2014). Abnormal ovarian DNA methylation programming during gonad maturation in wild contaminated fish. *Environmental Science & Technology*, *48*(19), 11688–11695. https://doi.org/10.1021/es503712c

Pigliucci, M. (2001). *Phenotypic plasticity: Beyond nature and nurture*. JHU Press.

Pigliucci, M. (2005). Evolution of phenotypic plasticity: Where are we going now? *Trends in Ecology & Evolution*, *20*(9), 481–486. https://doi.org/10.1016/j.tree.2005.06.001

Polanowski, A. M., Robbins, J., Chandler, D., & Jarman, S. N. (2014). Epigenetic estimation of age in humpback whales.

Molecular Ecology Resources, *14*(5), 976–987. https://doi.org/10.1111/1755-0998.12247

Provataris, P., Meusemann, K., Niehuis, O., Grath, S., & Misof, B. (2018). Signatures of DNA methylation across insects suggest reduced DNA methylation levels in Holometabola. *Genome Biology and Evolution*, *10*(4), 1185–1197. https://doi.org/10.1093/gbe/evy066

R Core Team (2020). *R Foundation for Statistical Computing*. R foundation for Statistical Computing. Retrieved from http://www.R-project.org

Reik, W., Santos, F., & Dean, W. (2003). Mammalian epigenomics: Reprogramming the genome for development and therapy. *Theriogenology*, *59*(1), 21–32. https://doi.org/10.1016/S0093-691X(02)01269-4

Richards, C. L., & Pigliucci, M. (2020). Epigenetic inheritance. A decade into the extended evolutionary synthesis. *Paradigmi*, *38*(3):463–494. https://doi.org/10.30460/99624

Richards, E. J. (2006). Inherited epigenetic variation—Revisiting soft inheritance. *Nature Reviews Genetics*, *7*(5), 395–401. https://doi.org/10.1038/nrg1834

Riyahi, S., Sánchez-Delgado, M., Calafell, F., Monk, D., & Senar, J. C. (2015). Combined epigenetic and intraspecific variation of the DRD4 and SERT genes influence novelty seeking behavior in great tit *Parus major*. *Epigenetics*, *10*(6), 516–525. https://doi.org/10.1080/15592294.2015.1046027

Riyahi, S., Vilatersana, R., Schrey, A. W., Ghorbani Node, H., Aliabadian, M., & Senar, J. C. (2017). Natural epigenetic variation within and among six subspecies of the house sparrow, *Passer domesticus*. *Journal of Experimental Biology*, *220*(21), 4016–4023. https://doi.org/10.1242/jeb.169268

Robinson, M. D., Kahraman, A., Law, C. W., Lindsay, H., Nowicka, M., Weber, L. M., & Zhou, X. (2014). Statistical methods for detecting differentially methylated loci and regions. *Frontiers in Genetics*, *5*, 324. https://doi.org/10.3389/fgene.2014.00324

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. https://doi.org/10.1111/mec.15253

Romano, A., De Giorgio, B., Parolini, M., Favero, C., Possenti, C. D., Iodice, S., Caprioli, M., Rubolini, D., Ambrosini, R., Gianfranceschi, L., Saino, N., & Bollati, V. (2017). Methylation of the circadian Clock gene in the offspring of a free-living passerine bird increases with maternal and individual exposure to PM10. *Environmental Pollution*, *220*, 29–37. https://doi.org/10.1016/j.envpol.2016.08.060

Ross, S. E., Hesselson, D., & Bogdanovic, O. (2021). Developmental accumulation of gene body and transposon Non-CpG methylation in the zebrafish brain. *Frontiers in Cell and Developmental Biology*, *9*, 643603. https://doi.org/10.3389/fcell.2021.643603

Roth, T. L., Lubin, F. D., Funk, A. J., & Sweatt, J. D. (2009). Lasting epigenetic influence of early-life adversity on the BDNF gene. *Biological Psychiatry*, *65*(9), 760–769. https://doi.org/10.1016/j.biopsych.2008.11.028

Rubenstein, D. R., Skolnik, H., Berrio, A., Champagne, F. A., Phelps, S., & Solomon, J. (2016). Sex-specific fitness effects of unpredictable early life conditions are associated with DNA methylation in the avian glucocorticoid receptor. *Molecular Ecology*, *25*(8), 1714–1728. https://doi.org/10.1111/mec.13483

Rubi, T. L., Knowles, L. L., & Dantzer, B. (2019). Museum epigenomics: Characterizing cytosine methylation in historic museum specimens. *Molecular Ecology Resources*, *20*(5), 1161–1170. https://doi.org/10.1111/1755-0998.13115

Saino, N., Ambrosini, R., Albetti, B., Caprioli, M., De Giorgio, B., Gatti, E., Liechti, F., Parolini, M., Romano, A., Romano, M., Scandolara, C., Gianfranceschi, L., Bollati, V., & Rubolini, D. (2017). Migration phenology and breeding success are predicted by methylation of a photoperiodic gene in the barn swallow. *Scientific Reports*, *7*(1), 45412. https://doi.org/10.1038/srep45412

Sarkies, P. (2020). Molecular mechanisms of epigenetic inheritance: Possible evolutionary implications. *Seminars in Cell & Developmental Biology*, *97*, 106–115. https://doi.org/10.1016/j.semcdb.2019.06.005

Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, *27*(6), 863–864. https://doi.org/10.1093/bioinformatics/btr026

Schrey, A. W., Alvarez, M., Foust, C. M., Kilvitis, H. J., Lee, J. D., Liebl, A. L., Martin, L. B., Richards, C. L., & Robertson, M. (2013). Ecological epigenetics: Beyond MS-AFLP. *Integrative and Comparative Biology*, *53*(2), 340–350. https://doi.org/10.1093/icb/ict012

Schrey, A. W., Coon, C. A. C., Grispo, M. T., Awad, M., Imboma, T., McCoy, E. D., Mushinsky, H. R., Richards, C. L., & Martin, L. B. (2012). Epigenetic variation may compensate for decreased genetic variation with introductions: A case study using house sparrows (*Passer domesticus*) on two continents. *Genetics Research International*, *2012*, 1–7. https://doi.org/10.1155/2012/979751

Schultz, M. D., Schmitz, R. J., & Ecker, J. R. (2012). 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends in Genetics*, *28*(12), 583–585. https://doi.org/10.1016/j.tig.2012.10.012

Seiler Vellame, D., Castanho, I., Dahir, A., Mill, J., & Hannon, E. (2021). Characterizing the properties of bisulfite sequencing data: Maximizing power and sensitivity to identify between-group differences in DNA methylation. *BMC Genomics*, *22*(1), 446. https://doi.org/10.1186/s12864-021-07721-z

Sepers, B., Erven, J. A. M., Gawehns, F., Laine, V. N., & van Oers, K. (2021). Epigenetics and early life stress: Experimental brood size affects DNA methylation in great tits (*Parus major*). *Frontiers in Ecology and Evolution*, *9*, 609061. https://doi.org/10.3389/fevo.2021.609061

Sepers, B., van den Heuvel, K., Lindner, M., Viitaniemi, H., Husby, A., & van Oers, K. (2019). Avian ecological epigenetics: Pitfalls and promises. *Journal of Ornithology*, *160*(4), 1183–1203. https://doi.org/10.1007/s10336-019-01684-5

Shafi, A., Mitrea, C., Nguyen, T., & Draghici, S. (2017). A survey of the approaches for identifying differential methylation using bisulfite sequencing data. *Briefings in Bioinformatics*, *19*(5), 737–753. https://doi.org/10.1093/bib/bbx013

Sharif, J., & Koseki, H. (2018). Hemimethylation: DNA's lasting odd couple. *Science*, *359*(6380), 1102–1103. https://doi.org/10.1126/science.aat0789

Sheldon, E. L., Schrey, A., Andrew, S. C., Ragsdale, A., & Griffith, S. C. (2018). Epigenetic and genetic variation among three separate introductions of the house sparrow (*Passer domesticus*) into Australia. *Royal Society Open Science*, *5*(4), 172185. https://doi.org/10.1098/rsos.172185

Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: Key considerations in genomic analyses. *Nature Reviews Genetics*, *15*(2), 121–132. https://doi.org/10.1038/nrg3642

Skinner, M. K. (2008). What is an epigenetic transgenerational phenotype?: F3 or F2. *Reproductive Toxicology*, *25*(1), 2–6. https://doi.org/10.1016/j.reprotox.2007.09.001

Skinner, M. K., Gurerrero-Bosagna, C., Haque, M. M., Nilsson, E. E., Koop, J. A. H., Knutie, S. A., & Clayton, D. H. (2014). Epigenetics and the evolution of Darwin's finches. *Genome Biology and Evolution*, *6*(8), 1972–1989. https://doi.org/10.1093/gbe/evu158

Skvortsova, K., Iovino, N., & Bogdanović, O. (2018). Functions and mechanisms of epigenetic inheritance in animals. *Nature Reviews Molecular Cell Biology*, *19*(12), 774–790. https://doi.org/10.1038/s41580-018-0074-2

Solares, E. A., Chakraborty, M., Miller, D. E., Kalsow, S., Hall, K., Perera, A. G., Emerson, J. J., & Hawley, R. S. (2018). Rapid low-cost assembly of the *Drosophila melanogaster* reference genome using low-coverage, long-read sequencing. *G3: Genes|genomes|genetics*, *8*, 3143–3154. https://doi.org/10.1534/g3.118.200162

Soulsbury, C. D., Lipponen, A., Wood, K., Mein, C. A., Hoffman, J. I., & Lebigre, C. (2018). Age- and quality-dependent DNA methylation correlate with melanin-based coloration in a wild bird. *Ecology and Evolution*, *8*(13), 6547–6557. https://doi.org/10.1002/ece3.4132

Srivastava, A., Karpievitch, Y. V., Eichten, S. R., Borevitz, J. O., & Lister, R. (2019). HOME: A histogram based machine learning approach for effective identification of differentially methylated regions. *BMC Bioinformatics*, *20*(1), 253. https://doi.org/10.1186/s12859-019-2845-y

St-Cyr, S., & McGowan, P. O. (2015). Programming of stress-related behavior and epigenetic neural gene regulation in mice offspring through maternal exposure to predator odor. *Frontiers in Behavioral Neuroscience*, *9*, 145. https://doi.org/10.3389/fnbeh.2015.00145

Stoiber, M., Quick, J., Egan, R., Lee, J. E., Celniker, S., Neely, R. K., & Brown, J. (2017). De novo Identification of DNA modifications enabled by genome-guided nanopore signal processing. BioRxiv, 094672. https://doi.org/10.1101/094672

Sun, X., Han, Y. I., Zhou, L., Chen, E., Lu, B., Liu, Y., Pan, X., Cowley, A. W., Liang, M., Wu, Q., Lu, Y., & Liu, P. (2018). A comprehensive evaluation of alignment software for reduced representation bisulfite sequencing data. *Bioinformatics*, *34*(16), 2715–2723. https://doi.org/10.1093/bioinformatics/bty174

Sun, Z., Cunningham, J., Slager, S., & Kocher, J.-P. (2015). Base resolution methylome profiling: Considerations in platform selection, data preprocessing and analysis. *Epigenomics*, *7*(5), 813–828. https://doi.org/10.2217/epi.15.21

Taff, C. C., Campagna, L., & Vitousek, M. N. (2019). Genome-wide variation in DNA methylation is associated with stress resilience and plumage brightness in a wild bird. *Molecular Ecology*, *28*(16), 3722–3737. https://doi.org/10.1111/mec.15186

Takuno, S., & Gaut, B. S. (2012). Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. *Molecular Biology and Evolution*, *29*(1), 219–227. https://doi.org/10.1093/molbev/msr188

Tang, J., Fang, F., Miller, D. F., Pilrose, J. M., Matei, D., Huang, T.-H.-M., & Nephew, K. P. (2015). Global DNA methylation profiling technologies and the ovarian cancer methylome. In M. Verma (Ed.), *Cancer epigenetics: Risk assessment, diagnosis, treatment, and prognosis* (pp. 653–675). Springer. https://doi.org/10.1007/978-1-4939-1804-1_34

Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J., & Prins, P. (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics*, *31*(12), 2032–2034. https://doi.org/10.1093/bioinformatics/btv098

Taub, M. A., Corrada Bravo, H., & Irizarry, R. A. (2010). Overcoming bias and systematic errors in next generation sequencing data. *Genome Medicine*, *2*(12), 87. https://doi.org/10.1186/gm208

Thompson, M. J., vonHoldt, B., Horvath, S., & Pellegrini, M. (2017). An epigenetic aging clock for dogs and wolves. *Aging*, *9*(3), 1055–1068. https://doi.org/10.18632/aging.101211

Thompson, R. F., Suzuki, M., Lau, K. W., & Greally, J. M. (2009). A pipeline for the quantitative analysis of CG dinucleotide methylation using mass spectrometry. *Bioinformatics*, *25*(17), 2164–2170. https://doi.org/10.1093/bioinformatics/btp382

Tomso, D. J., & Bell, D. A. (2003). Sequence context at human single nucleotide polymorphisms: Overrepresentation of CpG dinucleotide at polymorphic sites and suppression of variation in CpG islands. *Journal of Molecular Biology*, *327*(2), 303–308. https://doi.org/10.1016/s0022-2836(03)00120-7

Tse, O. Y. O., Jiang, P., Cheng, S. H., Peng, W., Shang, H., Wong, J., Chan, S. L., Poon, L. C. Y., Leung, T. Y., Chan, K. C. A., Chiu, R. W. K., & Lo, Y. M. D. (2021). Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(5), e2019768118. https://doi.org/10.1073/pnas.2019768118

Tsuji, J., & Weng, Z. (2016). Evaluation of preprocessing, mapping and postprocessing algorithms for analyzing whole genome bisulfite sequencing data. *Briefings in Bioinformatics*, *16*(6), 938–952.

Vaisvila, R., Ponnaluri, V. K. C., Sun, Z., Langhorst, B. W., Saleh, L., Guan, S., Dai, N., Campbell, M. A., Sexton, B. S., Marks, K., Samaranayake, M., Samuelson, J. C., Church, H. E., Tamanaha, E., Corrêa, I. R., Pradhan, S., Dimalanta, E. T., Evans, T. C., Williams, L., & Davis, T. B. (2021). Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Research*, *31*(7), 1280–1289. https://doi.org/10.1101/gr.266551.120

van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., & Verhoeven, K. J. F. (2016). epiGBS: Reference-free reduced representation bisulfite sequencing. *Nature Methods*, *13*(4), 322–324. https://doi.org/10.1038/nmeth.3763

van Oers, K., Sepers, B., Sies, W., Gawehns, F., Verhoeven, K. J. F., & Laine, V. N. (2020). Epigenetics of animal personality: DNA methylation cannot explain the heritability of exploratory behavior in a songbird. *Integrative and Comparative Biology*, *60*(6), 1517–1530. https://doi.org/10.1093/icb/icaa138

Verhoeven, K. J. F., Simonsen, K. L., & McIntyre, L. M. (2005). Implementing false discovery rate control: Increasing your power. *Oikos*, *108*(3), 643–647. https://doi.org/10.1111/j.0030-1299.2005.13727.x

Verhoeven, K. J. F., vonHoldt, B. M., & Sork, V. L. (2016). Epigenetics in ecology and evolution: What we know and what we need to know. *Molecular Ecology*, *25*(8), 1631–1638. https://doi.org/10.1111/mec.13617

Verhulst, E. C., Mateman, A. C., Zwier, M. V., Caro, S. P., Verhoeven, K. J. F., & van Oers, K. (2016). Evidence from pyrosequencing indicates that natural variation in animal personality is associated with DRD4 DNA methylation. *Molecular Ecology*, *25*(8), 1801–1811. https://doi.org/10.1111/mec.13519

Viitaniemi, H. M., Verhagen, I., Visser, M. E., Honkela, A., van Oers, K., & Husby, A. (2019). Seasonal variation in genome-wide DNA methylation patterns and the onset of seasonal timing of reproduction in great tits. *Genome Biology and Evolution*, *11*(3), 970–983. https://doi.org/10.1093/gbe/evz044

Vogt, G. (2021). Epigenetic variation in animal populations: Sources, extent, phenotypic implications, and ecological and evolutionary relevance. *Journal of Biosciences*, *46*(1), 24. https://doi.org/10.1007/s12038-021-00138-6

Wang, H., Ke, H., Zheng, Y., Lai, J., Luo, Q., & Chen, Q. (2017). A modified bisulfite conversion method for the detection of DNA methylation. *Epigenomics*, *9*(7), 955–969. https://doi.org/10.2217/epi-2016-0174

Wang, Q. I., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bähr, M., Wolf, S., Shendure, J., Eils, R., Plass, C., & Weichenhan, D. (2013). Tagmentation-based whole-genome bisulfite sequencing. *Nature Protocols*, *8*(10), 2022–2032. https://doi.org/10.1038/nprot.2013.118

Watson, H., Powell, D., Salmón, P., Jacobs, A., & Isaksson, C. (2021). Urbanization is associated with modifications in DNA methylation in a small passerine bird. *Evolutionary Applications*, *14*(1), 85–98. https://doi.org/10.1111/eva.13160

Watson, H., Salmón, P., & Isaksson, C. (2019). Dynamic changes in DNA methylation during embryonic and postnatal development of an altricial wild bird. *Ecology and Evolution*, *9*(17), 9580–9585. https://doi.org/10.1002/ece3.5480

Weaver, I. C. G., Cervoni, N., Champagne, F. A., D'Alessio, A. C., Sharma, S., Seckl, J. R., Dymov, S., Szyf, M., & Meaney, M. J. (2004). Epigenetic programming by maternal behavior. *Nature Neuroscience*, *7*(8), 847–854. https://doi.org/10.1038/nn1276

Weichenhan, D., Imbusch, C. D., Wang, Q., Brors, B., & Plass, C. (2019). Generation of Whole Genome Bisulfite Sequencing Libraries from Very Low DNA Input. In R. Küppers (Ed.), *Lymphoma: Methods and protocols* (pp. 229–248). Springer. https://doi.org/10.1007/978-1-4939-9151-8_10

Wenzel, M. A., & Piertney, S. B. (2014). Fine-scale population epigenetic structure in relation to gastrointestinal parasite load in red grouse (*Lagopus lagopus scotica*). *Molecular Ecology*, *23*(17), 4256–4273. https://doi.org/10.1111/mec.12833

Weyrich, A., Benz, S., Karl, S., Jeschek, M., Jewgenow, K., & Fickel, J. (2016). Paternal heat exposure causes DNA methylation and gene expression changes of Stat3 in Wild guinea pig sons. *Ecology and Evolution*, *6*(9), 2657–2666. https://doi.org/10.1002/ece3.1993

Weyrich, A., Jeschek, M., Schrapers, K. T., Lenz, D., Chung, T. H., Rübensam, K., Yasar, S., Schneemann, M., Ortmann, S., Jewgenow, K., & Fickel, J. (2018). Diet changes alter paternally inherited epigenetic pattern in male Wild guinea pigs. *Environmental Epigenetics*, *4*(2), 1–12. https://doi.org/10.1093/eep/dvy011

Weyrich, A., Schüllermann, T., Heeger, F., Jeschek, M., Mazzoni, C. J., Chen, W., Schumann, K., & Fickel, J. (2014). Whole genome sequencing and methylome analysis of the wild guinea pig. *BMC Genomics*, *15*(1), 1036. https://doi.org/10.1186/1471-2164-15-1036

Wilkinson, G. S., Adams, D. M., Haghani, A., Lu, A. T., Zoller, J., Breeze, C. E., Arnold, B. D., Ball, H. C., Carter, G. G., Cooper, L. N., Dechmann, D. K. N., Devanna, P., Fasel, N. J., Galazyuk, A. V., Günther, L., Hurme, E., Jones, G., Knörnschild, M., Lattenkamp, E. Z., … Horvath, S. (2021). DNA methylation predicts age and provides insight into exceptional longevity of bats. *Nature Communications*, *12*(1), 1615. https://doi.org/10.1038/s41467-021-21900-2

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

Williams, R. J. (2003). Restriction endonucleases: Classification, properties, and applications. *Molecular Biotechnology*, *23*(3), 225–243. https://doi.org/10.1385/mb:23:3:225

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology*, *13*(6), e1005510. https://doi.org/10.1371/journal.pcbi.1005510

Wingett, S. W., & Andrews, S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Research*, *7*, 1338. https://doi.org/10.12688/f1000research.15931.2

Wreczycka, K., Gosdschan, A., Yusuf, D., Grüning, B., Assenov, Y., & Akalin, A. (2017). Strategies for analyzing bisulfite sequencing data. *Journal of Biotechnology*, *261*, 105–115. https://doi.org/10.1016/j.jbiotec.2017.08.007

Wulfridge, P., Langmead, B., Feinberg, A. P., & Hansen, K. D. (2019). Analyzing whole genome bisulfite sequencing data from highly divergent genotypes. *Nucleic Acids Research*, *47*(19), e117. https://doi.org/10.1093/nar/gkz674

Xi, Y., & Li, W. (2009). BSMAP: Whole genome bisulfite sequence MAPping program. *BMC Bioinformatics*, *10*(1), 232. https://doi.org/10.1186/1471-2105-10-232

Xu, H., Luo, X., Qian, J., Pang, X., Song, J., Qian, G., Chen, J., & Chen, S. (2012). FastUniq: A fast de novo duplicates removal tool for paired short reads. *PLoS One*, *7*(12), e52249. https://doi.org/10.1371/journal.pone.0052249

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P. K., Kivioja, T., Dave, K., Zhong, F., Nitta, K. R., Taipale, M., Popov, A., Ginno, P. A., Domcke, S., Yan, J., Schübeler, D., Vinson, C., & Taipale, J. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science*, *356*(6337), eaaj2239. https://doi.org/10.1126/science.aaj2239

Zabet, N. R., Catoni, M., Prischi, F., & Paszkowski, J. (2017). Cytosine methylation at CpCpG sites triggers accumulation of non-CpG methylation in gene bodies. *Nucleic Acids Research*, *45*(7), 3777–3784. https://doi.org/10.1093/nar/gkw1330

Zemach, A., McDaniel, I. E., Silva, P., & Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science*, *328*(5980), 916–919. https://doi.org/10.1126/science.1186366

Zhang, Y., Baheti, S., & Sun, Z. (2018). Statistical method evaluation for differentially methylated CpGs in base resolution next-generation DNA sequencing data. *Briefings in Bioinformatics*, *19*(3), 374–386. https://doi.org/10.1093/bib/bbw133

Zhou, J., Zhao, M., Sun, Z., Wu, F., Liu, Y., Liu, X., He, Z., He, Q., & He, Q. (2020). BCREval: A computational method to estimate the bisulfite conversion ratio in WGBS. *BMC Bioinformatics*, *21*(1), 38. https://doi.org/10.1186/s12859-019-3334-z

Zhou, L. I., Ng, H. K., Drautz-Moses, D. I., Schuster, S. C., Beck, S., Kim, C., Chambers, J. C., & Loh, M. (2019). Systematic evaluation of library preparation methods and sequencing platforms for high-throughput whole genome bisulfite sequencing. *Scientific Reports*, *9*(1), 10383. https://doi.org/10.1038/s41598-019-46875-5

Ziller, M. J., Müller, F., Liao, J., Zhang, Y., Gu, H., Bock, C., & Meissner, A. (2011). Genomic Distribution and Inter-Sample Variation of Non-CpG Methylation across Human Cell Types. *PLOS Genetics*, *7*(12), e1002389. https://doi.org/10.1371/journal.pgen.1002389

Ziller, M. J., Hansen, K. D., Meissner, A., & Aryee, M. J. (2015). Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nature Methods*, *12*(3), 230–232. https://doi.org/10.1038/nmeth.3152

Zimmer, C., Larriva, M., Boogert, N. J., & Spencer, K. A. (2017). Transgenerational transmission of a stress-coping phenotype programmed by early-life stress in the Japanese quail. *Scientific Reports*, *7*(1), 46125. https://doi.org/10.1038/srep46125

How to cite this article: Laine, V. N., Sepers, B., Lindner, M., Gawehns, F., Ruuskanen, S., & van Oers, K. (2022). An ecologist's guide for studying DNA methylation variation in wild vertebrates. *Molecular Ecology Resources*, *00*, 1–21. https://doi.org/10.1111/1755-0998.13624