Methods in Ecology and Evolution

British Ecological Society

Methods in Ecology and Evolution 2016, **7**, 60–69 doi: 10.1111/2041-210X.12435

EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing

Drew R. Schield, Matthew R. Walsh, Daren C. Card, Audra L. Andrew, Richard H. Adams and Todd A. Castoe*

Department of Biology, University of Texas at Arlington, 501 S. Nedderman Dr., Arlington, TX 76019, USA

Summary

1. Research addressing the role of epigenetics in a diversity of experimental and natural systems is rapidly accumulating. Diverse methods have been developed to study epigenetic states, including bisulphite sequencing and AFLP-based approaches. However, existing methods are sometimes difficult to apply to non-traditional model organisms that lack genomic resources (bisulphite sequencing), and can fail to be economical and readily scalable to diverse research questions because of reliance on traditional Sanger sequencing (AFLP approaches).

2. Here we develop a reduced-representation library-based approach that is scalable and economical to quantitatively compare patterns of genomewide methylation. This approach shares substantial similarity to the now widely used double digest restriction-site associated DNA sequencing-based method (ddRADseq), except that it utilizes a methylation-sensitive restriction enzyme. This method therefore identifies changes in the genomic methylation state of cytosine (to 5-methylcytosine; 5mC) by sampling loci (via next-generation sequencing) that are not methylated within a sample. We test this method to identify shifts in the epigenome of clonal water fleas (*Daphnia ambigua*) in response to exposure to fish predator cues, which are known to induce transgenerational changes in life-history traits.

3. We found evidence for differential transgenerational responses (inferred via significant shifts in the methylation state of sampled loci) to predator cues among our treatment groups and remarkably consistent responses within treatment groups. Our results demonstrate that this method is capable of producing highly repeatable results even without the use of a reference genome.

4. Applications of this general method are broad and diverse and include the analysis of epigenetic shifts in both experimental and natural study systems.

Key-words: CpG methylation, *Daphnia*, epigenetics, high-throughput methylation measurement, restriction-site associated DNA sequencing, transgenerational shift

Introduction

Interest in the role of epigenetic variation in a wide range of natural systems has grown considerably in recent years as the impacts of epigenetic modifications on gene expression and trait variation have been demonstrated in numerous systems (e.g. Jaenisch & Bird 2003; Jablonka & Raz 2009; Vandegehuchte *et al.* 2010; Hammoud *et al.* 2011; Yaish, Colasanti & Rothstein 2011). A motivation for exploring changes in the epigenetic landscape of a system is the potential to uncover the bases of ecologically and evolutionary relevant variation that cannot be explained by genetic differences and to understand how these sources of variation may interact with shifts in allele frequencies to drive trait variation. While heritable variation in the form of genetic sequence differences has long been understood, it is becoming apparent that other sources of variation, including epigenetic variation, may also underlie ecologically relevant traits (Richards *et al.* 2013; Schrey *et al.* 2013; Kilvitis *et al.* 2014). A number of molecular techniques have been developed to study epigenetic variation through the interrogation of patterns of genomic DNA methylation, including bisulphite whole-genome sequencing, ChIP-Seq targeting 5-methylcytosine (5mC) and methylation-sensitive AFLP.

Despite the availability of multiple techniques to study the epigenetic state of genomic DNA, existing approaches are not readily scalable and fail to effectively leverage next-generation sequencing-based technologies to quantify shifts in genomewide patterns of DNA methylation. Existing epigenetic analysis methods work by either binding or targeting 5mC sites, or by selectively cleaving or bypassing methylated cytosines. For example, methylation-sensitive AFLP (Reyna-Lopez, Simpson & Ruiz-Herrera 1997) is designed to survey genomewide methylation patterns by replacing standard AFLP restriction enzymes with methylation-sensitive enzymes. This approach is economical and has been used effectively in ecological epigenetic studies of non-model organisms (e.g. Kronforst *et al.* 2008; Herrera & Bazaga 2010; Massicotte, Whitelaw & Angers 2011). Several limitations of this method, however, make an alternative approach desirable (Schrey *et al.* 2013). For exam-

*Correspondence author. E-mail: todd.castoe@uta.edu

ple, methylation-sensitive AFLP (MS-AFLP) does not provide information about the exact genomic regions surveyed, or what regions may be adjacent to these surveyed loci. AFLP-based methods are also limited by the number of loci that can be simultaneously screened based on fragment length polymorphism. Additionally, MS-AFLP lacks the sensitivity to detect moderate, but significant, quantitative shifts in the methylation state at a given locus across cells in a sampled tissue, or across tissues in a sampled organism, which may vary in a continuous frequency-based fashion (Bell *et al*. 2012).

Other methods that have been developed to characterize genomewide methylation patterns involve targeted binding or conversion of 5mC at CpG sites. Bisulphite sequencing entails the treatment of genomic DNA with sodium bisulphite, which converts unmethylated cytosine to uracil and leaves 5mC unaltered, allowing the direct evaluation of the genome for epigenetic marks at nucleotide-level resolution. Bisulphite sequencing is also applicable at genomewide and locus-specific scales, making it useful across a broad range of research questions, but this approach is expensive due to its requirement to either sequence complete genomes or to amplify and sequence *a priori* targeted loci individually.

We sought to develop a highly scalable and economical method that leverages next-generation sequencing effectively for the detection of continuously variable methylation state differences. Here we develop and test a reduced-representation genomic library approach to survey the methylation state of loci across the genome that is analogous to double digest restriction-site associated DNA sequencing (ddRADseq; Peterson *et al*. 2012). We altered the general ddRADseq method to examine shifts in methylation, replacing one of the restriction enzymes in the typical ddRADseq method with a methylation-sensitive restriction enzyme (*HpaII*). This modification, which we call 'EpiRADseq', samples only loci that do not contain 5mC bases, allowing for the comparisons of the methylation state of loci between samples based on a comparison of the frequencies at which loci are sampled. The EpiRADseq method incorporates many of the positive attributes of RADseq sampling, including its scalability with regard to the number of loci sampled (via selection of library size and restriction enzyme choice) and the ability to map sequenced loci back to a reference genome or to proceed in a reference-independent fashion.

Here we demonstrate the EpiRADseq method using an experimental system consisting of water fleas (*Daphnia ambigua*) and their fish predators. *Daphnia* have long served as a model organism for ecological and evolutionary research (Stollewerk 2010; Miner *et al*. 2012) and are well known to respond phenotypically to the presence of predators by producing morphological defences (head and tail spines) and altering life-history traits despite reproducing asexually (Stibor 1992; Riessen 1999). We used a single clone of *Daphnia* that is known to respond to the threat of predator cues by altering life-history traits across generations (i.e. transgenerational plasticity; Walsh *et al*. 2015). Recent work showed that exposure to predator cues in *Daphnia* leads to phenotypic responses that are detectable (and significant) two generations following cue removal (Walsh *et al*. 2015). We reared this clone in the presence and absence of predator cues and then quantified genomewide shifts in methylation across generations. A distinct advantage of using *Daphnia* to explore transgenerational epigenetic responses is that they reproduce asexually in the laboratory (under favourable conditions). Significant differences in recovery frequencies of EpiRADseq loci may occur due to differences in methylation or differences in genotype (i.e. allele-specific dropout). In these asexual clonal lines, shifts in EpiRADseq signatures are unambiguously due to shifts in the epigenome (i.e. shifts in genomewide patterns of 5mC), because all samples have exactly the same genotype. In addition to our demonstration of the EpiRADseq approach using a clonal model system, we also discuss how this method could be applied to other systems, including sexually reproducing populations in which there is genotypic variation across samples and experiments.

Materials and methods

OVERVIEW OF EPIRADSEQ METHODOLOGY

As a genomic reduced-representation approach, EpiRADseq is not designed to exhaustively identify all sites in the genome that are differentially methylated across treatments/samples, but instead is designed to sample a relatively large set of these sites sufficient for testing the hypothesis that changes in patterns of genomic methylation occur between samples or experiments. By changing the non-methylation-sensitive restriction enzyme (e.g. from an enzyme that recognizes a 6-base sequence, to one that recognizes an 8-base sequence), or by changing the range of fragment sizes selected, the method is highly scalable to deliver the desired degree of genome sampling for a particular application (Fig. 1a). With EpiRADseq sampling, quantitative differences in the frequency of methylation at a particular sampled locus are determined by differences in the frequencies of reads obtained per locus for a given sample. Each sampled EpiRADseq locus thus represents a potentially methylated CCGG site, and both the presence/absence and the relative depth of reads per locus can be compared across samples or treatments (Fig. 1b). If a locus is methylated, no EpiRADseq reads will be obtained, whereas if a locus is unmethylated, EpiRADseq reads will be sampled at a level that is proportional to the occurrence of the locus being unmethylated (Fig. 1b).

EMPIRICAL EXPERIMENTAL DESIGN

This experiment used a single clone of *D. ambigua* obtained from Dodge Pond in Connecticut (Post *et al*. 2008) that is known to respond to the presence of predators by programming future generations for faster rates of development and the production of larger clutches of offspring (Walsh *et al*. 2015). We reared this clone in a common garden for two generations followed by two generations of experimental manipulation. The experimental details regarding the first two generations of common garden rearing closely follow previous work (Walsh *et al*. 2015) and are thus only briefly described here (see details in Supporting Information). We evaluated the influence of predator cues on genomewide methylation patterns using newly born third-generation laboratory raised individuals (Fig. 2). Our methylation approaches required 30 individuals per replicate. To generate three replicates, *Daphnia* were reared in jars at modest densities (10–12 individuals per container) in the presence of predator cues in experimental generation

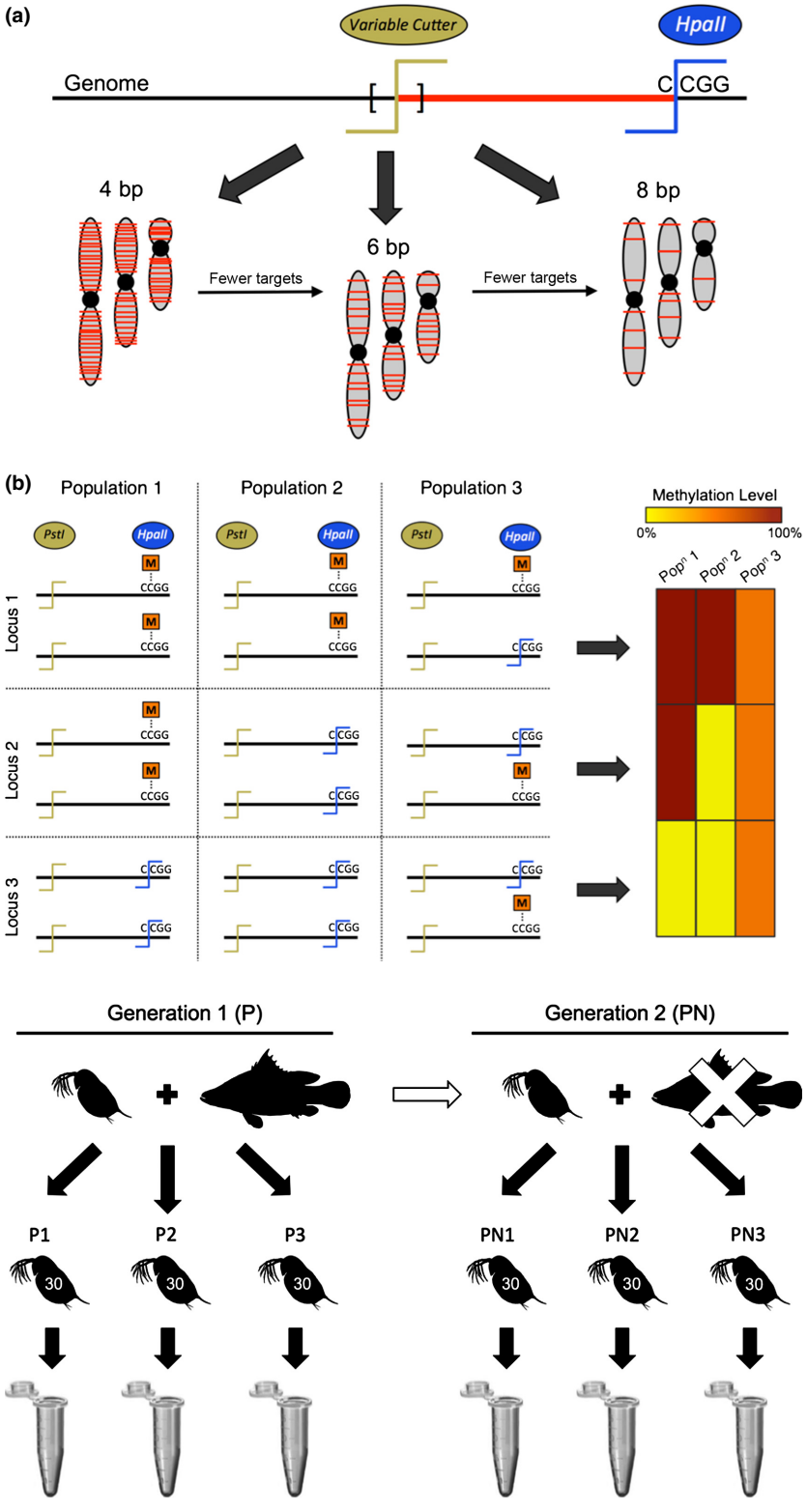

Fig. 1. Schematic overview of the EpiRADseq protocol. (a) Schematic showing EpiRADseq library restriction enzyme digestion, indicating how alternative restriction enzymes can be used to adjust the number of loci targeted. (b) Schematic example of how the EpiRADseq protocol provides information about the relative occurrence of 5mC at targeted loci.

Fig. 2. Overview of the experimental design used to test the EpiRADseq method. *Daphnia* in experimental generation 1 (P) were exposed to predator cues. Experimental generation 2 (PN) *Daphnia* were clones of the P generation individuals and were not exposed to predator cues. In each generation, *Daphnia* were reared in three separate replicate populations from which 30 individuals were taken for downstream molecular methods.

1 ('P' generation hereafter, see description of kairomones in Supporting Information). All *Daphnia* were transferred to fresh media, algae and kairomones daily. Based upon our previous work (Walsh & Post 2011, 2012), we expected that 10 days were needed for *Daphnia* to release their second clutch. We thus initiated the second experimental generation after 10 days of exposure to predator cues. This experimental generation was again initiated by collecting newly born (<12 h old) individuals and rearing them under the same experimental conditions as the P generation. The key difference is that the second experimental generation was not exposed to predator cues ('PN' generation hereafter, to reflect the non-predator status in experimental generation 2). At day 10 of each generation (P and PN generations), all adults were collated and snap-frozen in liquid nitrogen for use in EpiRADseq experiments (*c.* 30 individuals per replicate and generation; Fig. 2).

EPIRADSEQ LIBRARY PREPARATION

We extracted DNA from snap-frozen samples, with each sample consisting of *c*. 30 *Daphnia* individuals per replicate, for a total of 6 samples (2 generations × 3 replicates per generation, *c*. 30 individuals per replicate). DNA was extracted using the Zymo Research Duet Kit and quantified using a Qubit fluorometer (Life Technologies, Grand Island, NY, USA). We used approximately 300 ng of DNA from each sample as starting material for preparation of each EpiRADseq library. The laboratory protocol for EpiRADseq is modelled closely after that for ddRADseq by Peterson *et al*. (2012), with minor modifications (including alternative enzymes and adapter sequences) described below and also in greater detail in the Supporting Information and Appendix S1. In brief, DNA samples were digested with the restriction enzymes *PstI* (CATCAG recognition site) and *HpaII* (CCGG recognition site) and purified using AMPure beads (Invitrogen, Carlsbad, CA, USA). Digested samples were then quantified using a Qubit fluorometer and restandardized to a common quantity of DNA. We then ligated double-stranded sequencing adapters with unique barcodes to each sample. After ligation, we pooled two groups of three replicates (P and PN groups) and size-selected for fragments within a range of 640–790 bp using a Blue Pippin Prep (Sage Science, Beverly, MA, USA). We chose this range based on an *in silico* digestion of the *Daphnia pulex* genome (Colbourne *et al*. 2011), targeting a sampling of *c*. 20 000 loci (Fig. S1, Supporting Information). Size-selected libraries were then amplified using primers with group-specific multiplexed indices. Amplification reactions were then purified and quantified using a DNA 7500 chip run on an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The two grouped libraries were then pooled together in an equimolar fashion, and the final library was sequenced on an Illumina MiSeq using 168-bp paired-end reads. Additional details pertaining to digestion, ligation, amplification and final pooling can be found in the Supporting Information.

EPIRADSEQ COMPUTATIONAL ANALYSIS

Raw Illumina reads were filtered for PCR clones with the STACKS v. 1.19 (Catchen *et al*. 2013) tool *clone_filter,* using 8-bp unique molecular identifiers (UMIs; see details in Appendix S1), and these UMIs were subsequently trimmed off. Reads were then demultiplexed using the STACKS v. 1.19 (Catchen *et al*. 2013) *process_radtags* tool, which identified and removed a 6-bp leading barcode sequence in each read and also removed reads lacking a restriction site. Reads were then quality-filtered and trimmed using TRIMMOMATIC v. 0.32 (Bolger, Lohse & Usadel 2014) with default settings. The genomes of multiple *Daphnia* are available (e.g. we used the *D. pulex* genome (Colbourne *et al*. 2011) for additional analyses detailed below); however, here we employed a genome-independent approach to initially map and quantify reads per locus. We took this genome-independent approach to demonstrate the utility of the EpiRADseq method for species without genomic resources. To do this, we created a 'pseudoreference genome' by clustering reads and assembling contigs using RAINBOW v. 2.0.2 (Chong, Ruan & Wu 2012) and CD-Hit v 4.6 (Li & Godzik 2006). We then used BWA v. 0.7.5 (Li & Durbin 2009) to map quality-filtered EpiRADseq data for each individual to these pseudoreference contigs and SAMTOOLS v. 0.1.19 (Li *et al*. 2009) to estimate the read depth at each mapped locus per individual. We used TMM normalization implemented in edgeR (Robinson, McCarthy & Smyth 2010) to standardize distributions and control for uneven sampling. To estimate significant changes in methylation between generations, we used pairwise exact tests of the negative binomial distribution in edgeR (Robinson, McCarthy & Smyth 2010), integrating tagwise dispersion for all comparisons. All loci with a Benjamini–Hochberg corrected FDR value of <0.05 were considered to exhibit a significant shift in methylation. For each of these loci, we calculated the relative frequency of methylation per locus per replicate. We generated heat maps of these values for loci identified as experiencing significant shifts between generations in R by calculating the Bray–Curtis dissimilarity matrix for the data set and used average linkage hierarchical clustering to calculate dendrograms that grouped loci by observed patterns of change. We also applied principle component analysis (PCA) to identify the degree to which patterns of EpiRADseq variation could differentiate among generations and individuals by comparing reads per locus across samples. For PCA analyses, raw EpiRADseq reads per locus for each replicate were normalized using a square root transformation, and PCA analysis was conducted in R using standard functions, and using singular value decomposition of expression matrices.

To understand the distribution and clustering of potential and observed EpiRADseq loci in the *Daphnia* genome, we compared the genomic distribution of three sets of sites: (i) all potential EpiRADseq loci, (ii) all EpiRADseq loci that were observed at least once in our experiments and (iii) all EpiRADseq loci identified as significantly changing in methylation between P and PN generations. To estimate all potential EpiRADseq loci, we conducted an *in silico* double digest using the *D. pulex* genome (Colbourne *et al*. 2011) that was designed to emulate our experimental design by searching for genomic loci that contained the recognition sites for the *PstI* and *HpaII* restriction enzymes and that matched our experimental fragment size selection range (550–700 bp; our experimental size selection minus the 90-bp adapters ligated to our empirical EpiRAD fragments). To compare the results of our *in silico* digest of the *D. pulex* genome with our observed EpiRAD loci, we mapped all pseudoreference loci (our empirically observed EpiRADseq loci) to the *D. pulex* genome using BWA (Li & Durbin 2009).

For each set of EpiRADseq loci, we used the annotation of the *D. pulex* genome to classify where in the genome each locus occurred: within or outside of genic regions (exons or introns), within exons, within introns and within regions 1 kb upstream or 1 kb downstream of annotated genic regions. For these analyses, the location of a locus was determined based on the genomic coordinates of the 3′ end of observed and *in silico* EpiRAD loci (the CCGG cut site), and upstream versus downstream directions were determined based on the strand the EpiRADseq locus mapped to. These *D. pulex* genomic coordinates were also used to calculate the distance between adjacent EpiRADseq loci along the genome assembly, and the frequency distributions of distances between loci were visualized using interpolated smoothing splines. Interpolated splines were generated in R using default parameters, except for a reduced knot value ($n = 7$), which maximizes smoothing around highly variable low-distance frequencies. To test for enrichment of EpiRADseq loci in particular types of genomic locations, we performed Fisher's exact tests to compare frequencies of loci in each category between EpiRADseq sets and used a Bonferroni correction to decrease the likelihood of false positives due to multiple comparisons.

Results

EPIRADSEQ NEXT-GENERATION SEQUENCING RESULTS

A total of *c*. 5.95 million reads were generated for the six sequenced samples. Approximately 30% of reads were identified as PCR clones and were removed, and a further 25% lacked an intact restriction site and were also discarded. A small fraction of additional reads were removed because they

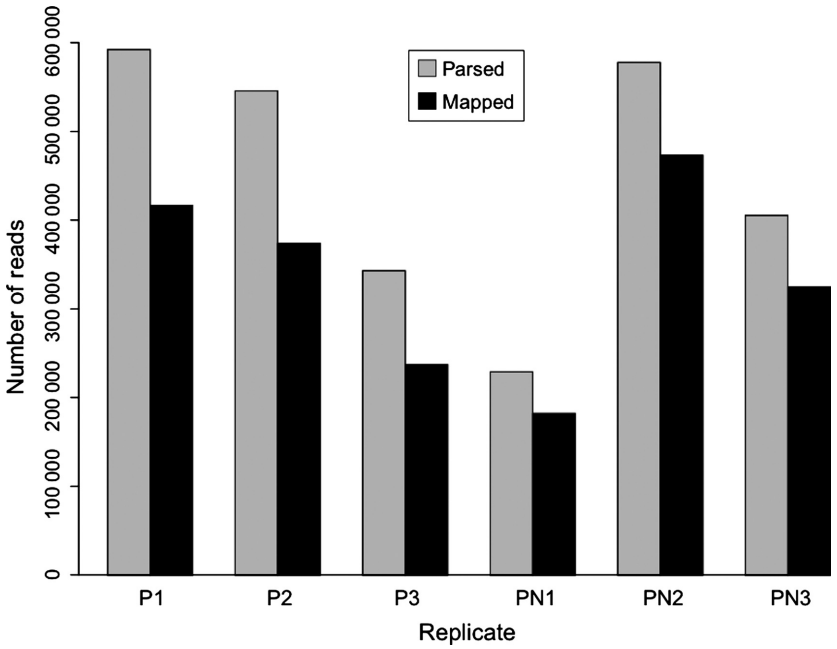

Fig. 3. Numbers of raw and mapped reads per sample after initial trimming and quality filtering.

failed quality thresholds (*c*. 0.4%). A pseudoreference genome constructed using the quality-filtered reads from all six samples (*c*. 2.72 million reads) yielded 23 134 pseudogenome contigs that represent our full set of observed EpiRAD loci. We generated frequency distributions in R to visualize the distributions of read depths among replicates for all observed EpiRAD loci, which show a trend of higher levels of coverage in loci identified as changing significantly between generations (Figs S2 and S3). The numbers of reads obtained per sample varied by approximately twofold (range = 228 890–592 294 reads; Fig. 3). The proportion of parsed reads within a sample that mapped back to our pseudoreference contigs was relatively consistent across all samples (mean = 75% ± 6% SD), though there was a higher proportion of mapped reads within the PN generation than in the P generation ($t = 12.44$, $P = 0.0002$; Fig. 3).

INFERENCE OF SIGNIFICANT SHIFTS IN METHYLATION STATE BETWEEN GENERATIONS

Our primary interest was to test for evidence that exposure to predator cues in the P generation led to transgenerational shifts in methylation between P and PN generations that were consistent across replicates within a generation. Of the 23 134 loci surveyed, 2002 loci showed significantly different normalized read depths (i.e. different relative frequencies) between P and PN generations, and therefore evidence for differential frequencies of observed methylation (Fig. 4a). Thus, the EpiRAD approach detected significant shifts in methylation at approximately 8.7% of all sampled loci.

To verify that our inferences of epigenetic shifts between P and PN generations were robust to biases from uneven sequence coverage across samples, we also conducted a parallel analysis in which we subsampled the same number of reads per sample down to the number obtained for the lowest coverage sample (sample PN1, with 228 157 reads) and repeated significance analyses. Although we expected this subsampling to reduce statistical power to detect significant shifts in EpiRAD loci, we also expected this subsampling to yield similar qualitative results. Indeed, this subsampled data set did yield qualitatively similar results (Fig. S4), including 949 significantly differentially methylated loci identified between P and PN generations. In both the full and subsampled data sets, differences between generations were remarkably consistent across individuals within a generation (Figs 4 and S4).

Principal component analysis conducted on the full EpiRADseq data set indicates that PC1 separates the generations (treatment effect) extremely well, accounting for 89.67% of the total variation in EpiRAD locus variation. The second principal component (PC2) primarily separated individuals within generations and accounted for relatively little variation compared to PC1 (3.57% of the total variance; Fig. 4b). Thus, our PCA demonstrates that variance in our data is primarily explained by treatment effects (between generations) and not by the variance among replicates within generations. Collectively, these results indicate that exposure to predator cues in the P generation leads to consistent changes in genomewide patterns of methylation in the PN generation.

GENOMIC DISTRIBUTIONS OF EPIRAD LOCI

All classes of EpiRAD loci that we analysed (*in silico*, observed, and significant observed) appear to be distributed in a clustered fashion when mapped to the *D. pulex* genome (Fig. 5). This pattern of clustering is clearly observed visually in the exemplary map of EpiRAD loci in the *D. pulex* genome at both 1 and 10 Mbp scales (Fig. 5a–b). In these example maps of EpiRADseq loci, it is notable that there are some instances where observed EpiRAD loci are not accompanied by a predicted EpiRAD locus; this is most likely due to sequence divergence in restriction cut sites between *D. pulex* (reference genome) and *D. ambigua* used in our experiments and the ability to map our empirical data to *D. pulex* even in

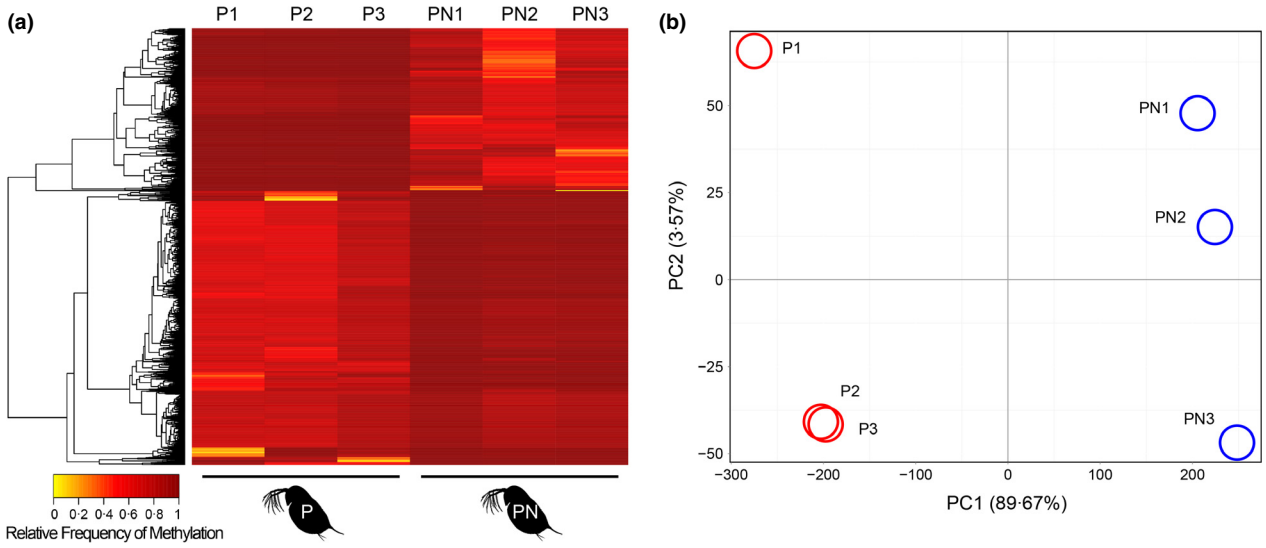

Fig. 4. Comparisons of EpiRADseq data between experimental *Daphnia* generations. (a) Heatmap of pairwise comparison of EpiRADseq loci sampled between experimental generations, depicting the relative frequency of methylation for 2002 loci identified as significantly differentially methylated between generations. Yellow indicates loci with low methylation (and high EpiRADseq read depth), orange indicates intermediate relative methylation frequency, and dark reds indicate high methylation (and low EpiRADseq read depth). Locus-specific EpiRADseq profiles are clustered by similarity, as indicated by tree on left. (b) Principal component analysis of variation in EpiRADseq reads per locus obtained across all replicates from each generation. The amount of total variance in the EpiRADseq-based methylation signature that each principal component (PC) explains is shown on each respective PC axis. Red circles represent P generation replicates, and blue circles represent PN generation replicates.

instances where the cut sites were not present in *D. pulex* (Figs 5a–b). The distributions of distances between adjacent EpiRAD loci in the *D. pulex* genome also show clear evidence of clustering of sites, with high proportions of loci occurring within 100 bp or less of another EpiRAD locus (Fig. 5c). In these comparisons (Fig. 5c), clustering is indicated specifically by the excess of shorter distances compared to the average distances between loci, which were 1971.8, 2542.9 and 37 320.0 bp for *in silico*, observed and significantly observed EpiRAD loci, respectively.

In silico EpiRAD analysis using the *D. pulex* genome identified a total of 29 396 potentially assayable loci using *HpaII* and *PstI* restriction enzymes. Our empirical EpiRAD data set contained 23 134 loci, 15 240 of which unambiguously mapped back to the *D. pulex* genome. Of the 2002 observed loci with significant shifts in methylation, 1181 mapped back to the *D. pulex* genome. We found that 54.02% of *in silico* loci mapped to non-genic regions, with 9.05% and 7.72% of loci falling in regions 1 kb up- or downstream of genic regions, respectively (Fig. 6). Comparisons between observed and significant EpiRAD loci suggest that these two sets of loci have very similar (and non-significantly differentiated) patterns of distribution with respect to genic and non-genic regions, exonic and intronic regions, and regions upstream and downstream of genes (Fig. 6). However, comparisons between *in silico* vs. observed or significant EpiRAD loci indicate that *in silico* loci tended to be significantly more frequent in non-genic regions and downstream of genes (Fisher's exact test P-values < 0.001; Fig. 6). To be observed, EpiRAD loci must be at least partially unmethylated in at least one sample. We therefore interpret the differences between *in silico* and observed distributions as a difference between loci that remain strictly methylated (and thus unsampled) and loci that may be at least partially unmethylated in some samples (in the observed data set). Based on this inference, our results suggest that sites that are constitutively methylated tend to occur more frequently in non-genic and downstream regions, whereas loci that are differentially methylated to some extent are more frequently observed in genic regions (Fig. 6).

Discussion

Daphnia are known to respond to the presence of predator cues by growing morphological defence structures (e.g. spines) and increasing developmental rates in the following generation (Miner *et al.* 2012; Walsh *et al.* 2015). Here, we used the EpiRADseq approach to test whether we could detect shifts in the *Daphnia* epigenome in successive generations that may coincide (and perhaps underlie) the transmission of these signals to the next generation. We chose to use clonally reproducing *Daphnia* in our empirical tests of the EpiRAD approach to rule out the confounding effects that genotypic variation may have on interpretation of EpiRADseq data because all *Daphnia* throughout the experiments were asexual clones of one another and thus all share the same genotype. This lack of genotypic variation between generations and among individuals allows the exclusion of the possibility that allele-specific or lineage-specific dropout (due to genetic variation in restriction sites; Arnold *et al.* 2013) is a cause of variation in EpiRAD locus recovery. Therefore, within this system (and other experiments where genotype is held constant), differences in the relative frequency of locus-specific EpiRAD coverage can be attributed solely to differential methylation of the *HpaII* restriction site, although we discuss below how the EpiRAD approach

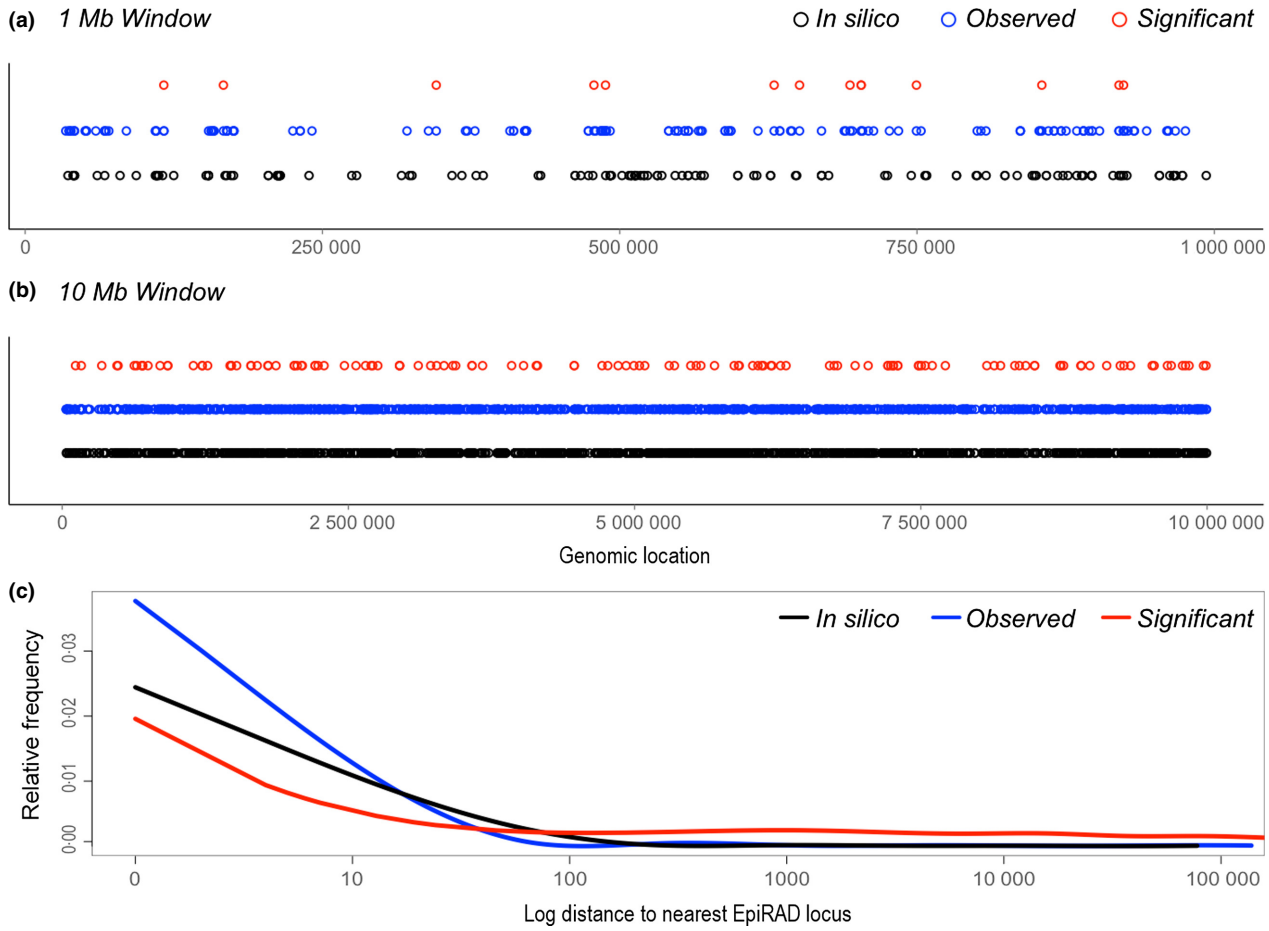

Fig. 5. Genomic distributions and clustering of EpiRAD loci. (a) Genomic locations of all assayable EpiRAD loci (n = 29 396) inferred from an *in silico* digest of the *D. pulex* genome (black), and all experimentally observed (n = 15 240) and experimentally significant (n = 1181) EpiRAD loci from our *D. ambigua* data set mapped to the *D. pulex* genome (blue and red, respectively). *D. pulex* scaffolds were sorted from longest (left) to shortest (right) and viewed in 1 and 10 Mb windows. (b) Interpolated splines of relative frequencies of distances between adjacent EpiRAD loci for *in silico*, observed and significant locus sets.

can be extended to account for genetic variation within an experiment. Our empirical analysis using EpiRADseq to quantify differences in the epigenetic state of *Daphnia* upon exposure to predator cues (P), and in the generation after exposure to such cues (PN), provides the first clear evidence of consistent shifts in the methylation state of a large number of loci between generations associated with transgenerational inheritance in *Daphnia* (Fig. 4).

Previous studies have shown that the distribution of CpG dinucleotides in the genomes of a diversity of organisms is non-random (e.g. Tweedie *et al.* 1997; Regev, Lamb & Jablonka 1998; Martienssen & Colot 2001; Zilberman *et al.* 2007). In particular, invertebrate genomes are a mosaic landscape of methylation, characterized by genomic 'islands' and 'deserts' of potentially methylated CpG sites (Bird, Taggart & Smith 1979; Tweedie *et al.* 1997). Our analysis of observed and predicted EpiRAD loci in the *D. pulex* genome is consistent with these previous findings and further highlights the non-random distribution of these sites in the genome. From the distribution patterns of both our predicted and observed sets of EpiRAD loci, it is clear that these sites form clusters and

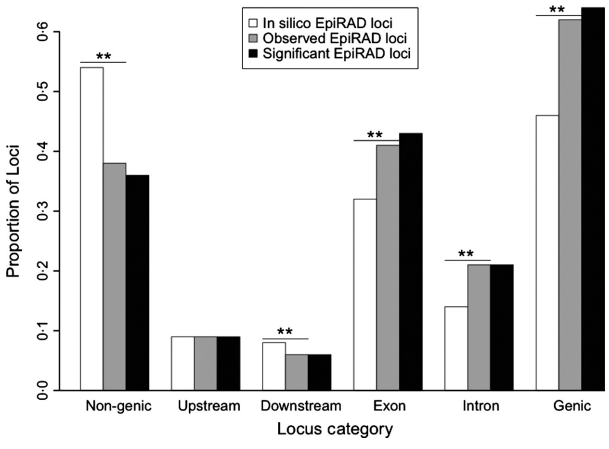

Fig. 6. Relative frequencies of occurrences of mapped EpiRAD loci in various categories of annotated genomic regions throughout the *D. pulex* genome. Double asterisks indicate significance (Fisher's exact test $P < 0.001$) of differences in frequencies between EpiRADseq sets in particular categories of genomic regions, with horizontal lines indicating the pair of frequencies being compared.

are enriched in particular genomic regions (Figs 5–6). In the case of our empirical EpiRAD loci, sites that we predicted but did not observe are likely to be constitutively methylated across generations and thus not sampled by the EpiRADseq method. In contrast, observed loci are only sometimes methylated and to varying degrees across generations and were assayable using EpiRADseq. Our comparison of the distributions of predicted versus observed EpiRADseq loci suggests that EpiRAD loci that are observed, and thus variably methylated within our experiment, tended to occur at significantly greater frequencies in genic regions (both exons and introns) and were less frequent in regions downstream of genes (Fig. 6). Thus, while the distribution of potentially sampled sites in the genome is non-random (Bird 1985; Suzuki & Bird 2008), our sampling of variably methylated sites in the genome is also non-random with regard to the proximity to genes. This observed enrichment towards genic regions in observed EpiRADseq loci in our *Daphnia* experiment suggests that the loci that change significantly between generations are reasonably likely to impact gene expression and/or splicing of transcripts.

Because the EpiRADseq method is based on methylation-sensitive enzymes that do not cut the genome (and thus do not deliver a sequence read) when a cut site is methylated, the quantities of reads that map to a particular locus provide evidence for the degree of methylation. Thus, greater numbers of reads (i.e. read depth) at a locus indicate low methylation at that locus. An important aspect of this approach is that loci are not scored as homo- or hemi-methylated, but are rather sampled along a continuous scale averaged over all sampled cells and DNA strands. For example, in sampling of 30 whole *Daphnia* individuals, we would expect to observe the same signature of methylation at a site if it is methylated in 100% of cells in 50% of individuals (and 100% unmethylated in the other 50% of individuals) as we would if the same locus were methylated in 50% of all cells in all individuals. Based on these attributes of the data obtained by EpiRADseq, these data share many general characteristics with RNAseq data, including that both methods are based on sampling loci from a distribution of loci with varying frequencies. Thus, EpiRADseq data can be readily analysed using numerous existing approaches and programs otherwise intended for analysis of RNAseq data.

The EpiRADseq method relies on the recognition of restriction cut sites, and differences in genotype will affect the outcome of EpiRADseq result if genotypic variation is not accounted for. Thus, without any modification or additional extensions of the current method, the EpiRADseq approach is immediately applicable to research questions that involve the measuring of epigenetic differences among genetically identical samples or individuals, or closely related individuals that are unlikely to have confounding genetic differences. Such applications of EpiRADseq include the monitoring of epigenetic states throughout an organism's development and measuring the epigenetic effects of environmental factors including toxins and other natural environmental cues that may impact an organism's phenotype. Beyond research questions in evolutionary biology and ecology, the EpiRADseq approach may also be applied to addressing the role of epigenetics on signal transduction and ageing, for studying epigenetic shifts in cell culture upon exposure to various treatments, or in other instances where the genotype of the sample would remain constant, such as in the serial sampling of tissues that can be taken in a non-lethal fashion.

There are straightforward modifications of the EpiRADseq approach in which differences in genotypes may be accounted for directly, with the addition of standard ddRADseq data using the enzyme *MspI* (which cuts at CCGG sites like *HpaII,* but is insensitive to methylation). In such instances, analysis of the standard ddRADseq data (using *MspI*) could be used to identify the presence/absence of particular loci among individuals (due to differences in genotypes), and this information could be used to interpret whether differences in EpiRAD locus coverage were based on the allelic presence/absence, or on the methylation state of a locus. For example, a simplistic approach to account for variation in genotype among individuals would be to limit EpiRADseq analysis to include only loci for which all individuals are homozygous for the presence of the *MspI* recognition site based on ddRADseq, thereby removing the effect of genotypic bias in the interpretation of EpiRADseq results.

Given that the EpiRADseq approach can be readily extended to systems that include genetic variation (e.g. sexual populations), the approach is highly applicable to a diversity of research questions in ecology and evolutionary biology that involve epigenetic shifts. Research focusing on major ontogenetic, developmental or physiological shifts, where massive fluctuations in gene expression are likely integrated with major changes in genomic methylation, would also benefit from a thorough appraisal of genome-wide methylation patterns using EpiRADseq. For example, EpiRADseq would be valuable for studying epigenetic shifts that accompany, and possibly direct, metamorphosis in insects (Shinoda & Itoyama 2003) and amphibians (Denver 1997), or ontogenetic shifts traits such as venom composition in snakes (Saviola *et al.* 2015) and hormone-driven sexual differentiation in birds (Balthazart & Ball 1995) and lizards (Cox, Stenquist & Calsbeek 2009). Major regenerative phenomena, including limb and tail regeneration in amphibians (Monaghan *et al.* 2007), lizards (Hutchins *et al.* 2014) and invertebrates (Konstantinides & Averof 2014), also represent intriguing model systems in which to study epigenetic regulation using EpiRADseq. In addition to data on shifts in patterns of the epigenome, the coupling of information on epigenomic fluctuations inferred using EpiRADseq with gene expression data holds exciting potential for developing new insight into the processes directing genomic change that results in phenotypic shifts.

Acknowledgements

We would like to thank several anonymous reviewers for their constructive feedback and helpful suggestions. Support was provided from start-up funds from the

University of Texas at Arlington, and a UTA REP grant to TAC and MRW. We thank Jill Castoe for sequencing assistance.

Author contributions

Todd A. Castoe, Matthew R. Walsh and Drew R. Schield conceived and designed the method and experiments. MRW reared sample specimens under experimental conditions and prepared samples for downstream molecular methods. Drew R. Schield performed molecular laboratory work and prepared samples for sequencing. Daren C. Card, Audra L. Andrew, Richard H. Adams, Drew R. Schield and Todd A. Castoe designed sequence data analysis pipelines and performed analyses. Drew R. Schield, Matthew R. Walsh and Todd A. Castoe wrote the manuscript, and all authors edited the final manuscript.

Data accessibility

Raw Illumina reads for all samples are available through the NCBI Sequence Read Archive (SRR1640572 for the P treatment and SRR1640573 for the PN treatment). Each read file (organized by generation) contains three multiplexed individuals that can be parsed by unique barcodes (see Appendix S1). These same raw Illumina reads, the pseudoreference genome, individual sample mapping files (in BAM format) and individual sample raw methylation count data, plus the raw reads, mapping files and methylation counts for the subsampled analysis and the coordinates used to test for genomic enrichment, are available via Dryad (Schield *et al.* 2015; http://dx.doi.org/10.5061/dryad.hs200). The Dryad repository also includes the commands, scripts and barcode information necessary to reanalyse the raw data.

References

Arnold, B., Corbett-Detig, R.B., Hartl, D. & Bomblies, K. (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179–3190.

Balthazart, J. & Ball, G.F. (1995) Sexual differentiation of brain and behavior in birds. *Trends in Endocrinology & Metabolism*, **6**, 21–29.

Bell, J.T., Tsai, P.C., Yang, T.P., Pidsley, R., Nisbet, J., Glass, D. *et al.* (2012) Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genetics*, **8**, 189–200.

Bird, A.P. (1985) CpG-rich islands and the function of DNA methylation. *Nature*, **321**, 209–213.

Bird, A.P., Taggart, M.H. & Smith, B.A. (1979) Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell*, **17**, 889–901.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.

Chong, Z., Ruan, J. & Wu, C.I. (2012) Rainbow: an integrated tool for efficient clustering and assembling RAD-seq reads. *Bioinformatics*, **28**, 2732–2737.

Colbourne, J.K., Pfrender, M.E., Gilbert, D., Thomas, W.K., Tucker, A., Oakley, T.H. *et al.* (2011) The ecoresponsive genome of *Daphnia pulex*. *Science*, **331**, 555–561.

Cox, R.M., Stenquist, D.S. & Calsbeek, R. (2009) Testosterone, growth and the evolution of sexual size dimorphism. *Journal of Evolutionary Biology*, **22**, 1586–1598.

Denver, R.J. (1997) Proximate mechanisms of phenotypic plasticity in amphibian metamorphosis. *American Zoologist*, **37**, 172–184.

Hammoud, S.S., Nix, D.A., Hammoud, A.O., Gibson, M., Cairns, B.R. & Carrell, D.T. (2011) Genome-wide analysis identifies changes in histone retention and epigenetic modifications at developmental and imprinted gene loci in the sperm of infertile men. *Human Reproduction*, **26**, 2558–2569.

Herrera, C.M. & Bazaga, P. (2010) Epigenetic differentiation and relationship to adaptive genetic divergence in discrete populations of the violet *Viola cazorlensis*. *New Phytologist*, **187**, 867–876.

Hutchins, E.D., Markov, G.J., Eckalbar, W.L., George, R.M., King, J.M., Tokuyama, M.A. *et al.* (2014) Transcriptomic analysis of tail regeneration in the lizard *Anolis carolinensis* reveals activation of conserved vertebrate developmental and repair mechanisms. *PLoS ONE*, **9**, e105004.

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

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

Kilvitis, H.J., Alvarez, M., Foust, C.M., Schrey, A.W., Robertson, M. & Richards, C.L. (2014) Ecological epigenetics. *Ecological Genomics: Ecology and the Evolution of Genes and Genomes*, **781**, 191–210.

Konstantinides, N. & Averof, M. (2014) A common cellular basis for muscle regeneration in arthropods and vertebrates. *Science*, **343**, 788–791.

Kronforst, M.R., Gilley, D.C., Strassmann, J.E. & Queller, D.C. (2008) DNA methylation is widespread across social Hymenoptera. *Current Biology*, **18**, R287–R288.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li, W.Z. & Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. & Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Martienssen, R.A. & Colot, V. (2001) DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science*, **293**, 1070–1074.

Massicotte, R., Whitelaw, E. & Angers, B. (2011) DNA methylation: a source of random variation in natural populations. *Epigenetics*, **6**, 422–428.

Miner, B.E., De Meester, L., Pfrender, M.E., Lampert, W. & Hairston, N.G. Jr (2012) Linking genes to communities and ecosystems: *Daphnia* as an ecogenomic model. *Proceedings of the Royal Society B: Biological Sciences*, **279**, 1873–1882.

Monaghan, J.R., Walker, J.A., Page, R.B., Putta, S., Beachy, C.K. & Voss, S.R. (2007) Early gene expression during natural spinal cord regeneration in the salamander *Ambystoma mexicanum*. *Journal of Neurochemistry*, **101**, 27–40.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S. & Hoekstra, H.E. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.

Post, D.M., Palkovacs, E.P., Schielke, E.G. & Dodson, S.I. (2008) Intraspecific variation in a predator affects community structure and cascading trophic interactions. *Ecology*, **89**, 2019–2032.

Regev, A., Lamb, M.J. & Jablonka, E. (1998) The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Molecular Biology & Evolution*, **15**, 880–891.

Reyna-Lopez, G.E., Simpson, J. & Ruiz-Herrera, J. (1997) Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Molecular & General Genetics*, **253**, 703–710.

Richards, C.L., Boruta, M., Bossdorf, O., Coon, C.A.C., Foust, C.M., Hughes, A.R. *et al.* (2013) Epigenetic mechanisms of phenotypic plasticity. *Integrative and Comparative Biology*, **53**, E179–E179.

Riessen, H.P. (1999) Predator-induced life history shifts in *Daphnia*: a synthesis of studies using meta-analysis. *Canadian Journal of Fisheries & Aquatic Sciences*, **56**, 2487–2494.

Robinson, M.D., McCarthy, D.J. & Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.

Saviola, A.J., Pla, D., Sanz, L., Castoe, T.A., Calvete, J.J. & Mackessy, S.P. (2015) Comparative venomics of the Prairie Rattlesnake (*Crotalus viridis viridis*) from Colorado: identification of a novel pattern of ontogenetic changes in venom composition and assessment of the immunoreactivity of the commercial antivenom CroFab(R). *Journal of Proteomics*, **121**, 28–43.

Schield, D.R., Walsh, M.R., Card, D.C., Andrew, A.L., Adams, R.H. & Castoe, T.A. (2015) Data from: EpiRADseq: scalable analysis of genome-wide patterns of methylation using next-generation sequencing. *Dryad Data Repository*, http://dx.doi.org/10.5061/dryad.hs200.

Schrey, A., Alvarez, M., Foust, C., Kilvitis, H., Liebl, A., Martin, L.B., Richards, C. & Robertson, M. (2013) Ecological epigenetics: beyond MS-AFLP. *Integrative and Comparative Biology*, **53**, E191–E191.

Shinoda, T. & Itoyama, K. (2003) Juvenile hormone acid methyltransferase: a key regulatory enzyme for insect metamorphosis. *Proceedings of the National Academy of Sciences USA*, **100**, 11986–11991.

Stibor, H. (1992) Predator induced life history shifts in a freshwater cladoceran. *Oecologia*, **92**, 162–165.

Stollewerk, A. (2010) The water flea *Daphnia* – a 'new' model system for ecology and evolution? *Journal of Biology*, **9**, 21.

Suzuki, M.M. & Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics*, **9**, 465–476.

Tweedie, S., Charlton, J., Clark, V. & Bird, A. (1997) Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Molecular and Cellular Biology*, **17**, 1469–1475.

Vandegehuchte, M.B., Lemiere, F., Vanhaecke, L., Vanden Berghe, W. & Janssen, C.R. (2010) Direct and transgenerational impact on *Daphnia magna* of chemicals with a known effect on DNA methylation. *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology*, **151**, 278–285.

Walsh, M.R. & Post, D.M. (2011) Interpopulation variation in a fish predator drives evolutionary divergence in prey in lakes. *Proceedings of the Royal Society B: Biological Sciences*, **278**, 2628–2637.

Walsh, M.R. & Post, D.M. (2012) The impact of intraspecific variation in a fish predator on the evolution of phenotypic plasticity and investment in sex in Daphnia ambigua. *Journal of Evolutionary Biology*, **25**, 80–89.

Walsh, M.R., Cooley, F., Biles, K. & Munch, S.B. (2015) Predator-induced phenotypic plasticity within- and across-generations: a challenge for theory? *Proceedings of the Royal Society B: Biological Sciences*, **282**, 20142205.

Yaish, M.W., Colasanti, J. & Rothstein, S.J. (2011) The role of epigenetic processes in controlling flowering time in plants exposed to stress. *Journal of Experimental Botany*, **62**, 3727–3735.

Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. & Henikoff, S. (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nature Genetics*, **39**, 61–69.

Received 8 April 2015; accepted 18 June 2015
Handling Editor: Michael Bunce

Supporting Information

Additional Supporting Information may be found in the online version of this article.

Data S1. Supplementary Materials and Methods.

Appendix S1. Adapter design.

Fig. S1. Estimated fragment size distribution after *in silico* digestion of the *D. pulex* genome with *HpaII* and *PstI* restriction enzymes.

Fig. S2. Frequency distributions of the average read depth for loci across all replicates in (a) all observed EpiRAD loci, and (b) all observed EpiRAD loci with significant shifts in methylation across experimental generations.

Fig. S3. Frequency distributions of the read depth for loci in our observed epiRAD sampling for each individual replicate.

Fig. S4. Comparisons of experimental generations after subsampling of raw reads.