



SYMPOSIUM

Epigenetic Potential in Native and Introduced Populations of House Sparrows (*Passer domesticus*)

Haley E. Hanson,^{1,*} Bilal Koussayer,^{*} Holly J. Kilvitis,[†] Aaron W. Schrey,[‡] J. Dylan Maddox,^{§,¶,||} and Lynn B. Martin^{*}

^{*}Global and Planetary Health, University of South Florida, 3720 Spectrum Blvd, Suite 304, Tampa, FL 33620, USA;

[†]Department of Integrative Biology, University of South Florida, 4202 E. Fowler Ave, SCA110, Tampa, FL 33620, USA;

[‡]Department of Biology, Georgia Southern University, Armstrong Campus, 11935 Abercorn St, SC1010, Savannah, GA 31419, USA; [§]Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, IL 60605, USA; [¶]Laboratorio de Biotecnología y Bioenergética, Universidad Científica del Perú, Iquitos, Perú; ^{||}American Public University System, Environmental Sciences, Charles Town, WV 25414, USA

From the symposium “Epigenetic Variation in Endocrine Systems” presented at the annual meeting of the Society for Integrative and Comparative Biology January 3–7, 2020 at Austin, Texas

¹E-mail: haleyehanson@gmail.com

Synopsis Epigenetic potential, defined as the capacity for epigenetically-mediated phenotypic plasticity, may play an important role during range expansions. During range expansions, populations may encounter relatively novel challenges while experiencing lower genetic diversity. Phenotypic plasticity via epigenetic potential might be selectively advantageous at the time of initial introduction or during spread into new areas, enabling introduced organisms to cope rapidly with novel challenges. Here, we asked whether one form of epigenetic potential (i.e., the abundance of CpG sites) in three microbial surveillance genes: Toll-like receptors (TLRs) 1B (*TLR1B*), 2A (*TLR2A*), and 4 (*TLR4*) varied between native and introduced house sparrows (*Passer domesticus*). Using an opportunistic approach based on samples collected from sparrow populations around the world, we found that introduced birds had more CpG sites in *TLR2A* and *TLR4*, but not *TLR1B*, than native ones. Introduced birds also lost more CpG sites in *TLR1B*, gained more CpG sites in *TLR2A*, and lost fewer CpG sites in *TLR4* compared to native birds. These results were not driven by differences in genetic diversity or population genetic structure, and many CpG sites fell within predicted transcription factor binding sites (TFBS), with losses and gains of CpG sites altering predicted TFBS. Although we lacked statistical power to conduct the most rigorous possible analyses, these results suggest that epigenetic potential may play a role in house sparrow range expansions, but additional work will be critical to elucidating how epigenetic potential affects gene expression and hence phenotypic plasticity at the individual, population, and species levels.

Introduction

Phenotypic plasticity is an important mediator of range expansions, as populations moving into new areas must quickly adjust to local conditions and overcome bottlenecks and/or founder effects that affect overall genetic diversity (Lande 2015). Phenotypic plasticity is defined as the expression of different phenotypes from the same genotype across different environments (Pigliucci 2005). Plasticity can arise solely through epigenetic effects, that is, variation in factors that influence the phenotype but do not involve

changes in deoxyribonucleic acid (DNA) sequence (Pigliucci 2005). Epigenetic mechanisms, such as DNA methylation, work within the genome to produce a range of phenotypes. As such, the specific genomic elements on which these mechanisms work may themselves be selected to increase the capacity for phenotypic plasticity (Branciamore et al. 2010; Feinberg and Irizarry 2010). This latent capacity for phenotypic plasticity, termed epigenetic potential, may differ among individuals, populations, and species (Kilvitis et al. 2017).

Epigenetic potential can take several forms: (1) the mediators of epigenetic mechanisms (e.g., DNA methyltransferases, histone deacetylases, etc.) may differ in genetic sequence, leading to functional differences affecting gene expression or (2) the genomic elements upon which these mediators act can differ in number or location (Branciamore et al. 2010; Feinberg and Irizarry 2010; Kilvitis et al. 2017). For example, in vertebrates, DNA methylation generally occurs when a methyl group is added to a cytosine adjacent to a guanine, termed a CpG site (Meissner et al. 2008). DNA methylation can reduce or increase gene expression depending on the location of a CpG site within the genome (i.e., promoter, exon, and intron) and/or whether a CpG site falls within a distinct regulatory element such as a transcription factor binding site (TFBS; Zhu et al. 2016). Subsequently, the number of CpG sites within a gene represents its capacity to be methylated, and thus a portion of its capacity for epigenetically-mediated phenotypic plasticity (Kilvitis et al. 2017). When differences in such epigenetic potential occur in genes affecting fitness, natural selection can occur (Feinberg and Irizarry 2010; Flores and Amdam 2011; Flores et al. 2013). For individuals that rely on plasticity to cope with unpredictable conditions, such as during range expansions, epigenetic potential might be selectively advantageous, with high epigenetic potential favored at the time of initial introduction and/or in response to conditions experienced as the new range is colonized (Marsh et al. 2016; Kilvitis et al. 2017; Kilvitis et al. 2018).

One species in particular that has had exceptional success expanding its range is the house sparrow (*Passer domesticus*). This species has established a near-ubiquitous distribution after introductions on multiple continents and through natural and anthropogenic expansions of its native range (Hanson et al. 2020). One of the most recent house sparrow introductions occurred in Kenya, where the species was introduced to Mombasa in the 1950s from South Africa (Schrey et al. 2014). Despite the recency of the introduction, Kenyan house sparrow populations exhibit phenotypic patterns (e.g., exploratory behavior, propensity to eat novel foods, and regulation of their stress responses) suggesting that populations have already adjusted to prevailing conditions (Liebl and Martin 2012; Liebl and Martin 2014). This variation, however, occurs despite the fact that Kenyan populations have much lower genetic diversity than many populations from the native range, yet compared to native populations, they appear to have greater epigenetic diversity (Schrey et al. 2011;

Schrey et al. 2012). Moreover, in Kenya, populations with low genetic diversity have high epigenetic diversity, suggesting that populations might use epigenetic mechanisms to bolster phenotypic variation (Schrey et al. 2012; Liebl et al. 2013). However, in Australia where house sparrows were introduced much earlier than in Kenya, this trend is much weaker (Sheldon et al. 2018). The strength of the relationship may stem from multiple introduction events into Australia and/or the age of that introduction (Sheldon et al. 2018).

One of the most consistent patterns of phenotypic variation observed in sparrows colonizing Kenya entails variation in the expression of toll-like receptors (TLRs); expression of *TLR2* and *TLR4* increased in populations located far from the introduction origin where conditions are probably more novel (Martin et al. 2014; Martin et al. 2015; Martin et al. 2017). TLRs reside on macrophages and other leukocytes and recognize pathogen-associated molecular patterns (PAMPs): *TLR2* recognizes peptidoglycan found on Gram-positive bacteria whereas *TLR4* recognizes lipopolysaccharide found on Gram-negative bacteria (Velová et al. 2018). Upon binding PAMPs, TLRs trigger cytokine production and transcription factors that instigate immune responses, leading either to control of infections, recruitment of the adaptive immune system, or both (Iwasaki and Medzhitov 2015). In this sense, TLRs are some of the first lines of surveillance against pathogens (Iwasaki and Medzhitov 2015). As the immune defenses they incite are broadly effective, these receptors are expected to be important for individuals invading new areas where novel and/or generalist parasites will be comparatively more frequent (Phillips et al. 2010; Marzal et al. 2011).

In this study, we asked whether epigenetic potential in the upstream regulatory regions of three TLRs, *TLR1B*, *TLR2A*, and *TLR4*, differed between native and introduced house sparrows (Supplementary Table S1). In a previous study of Kenyan house sparrows, levels of *TLR4* expression (at the individual level) were found to be inversely correlated with levels of methylation in one CpG site in its putative promoter (Kilvitis et al. 2019). Here, we queried whether introduced sparrows generally maintained more CpG sites in putative TLR promoters than native ones. We hypothesized that introduced birds would have higher epigenetic potential (i.e., more CpG sites within TLR promoters), as this could provide more phenotypic plasticity in gene expression, presumably more advantageous in new areas (Feinberg and Irizarry 2010; Kilvitis et al. 2017).

Table 1 Sample size, mean, and range of total, gains, and losses of CpG sites per gene by group (native vs. introduced)

Gene	Group	Sample size	Total		Gains		Losses	
			Mean	Range	Mean	Range	Mean	Range
TLR1B	Native	18	17.06	16–19	1.28	1–3	0.22	0–1
	Introduced	46	16.15	14–19	0.89	0–3	0.74	0–2
TLR2A	Native	32	2.47	0–4	0.78	0–2	0.31	0–2
	Introduced	77	3.01	0–6	1.18	0–4	0.30	0–2
TLR4	Native	37	7.54	6–9	0.03	0–1	0.49	0–2
	Introduced	92	7.90	6–14	0.13	0–6	0.23	0–2

We also hypothesized that the total number of CpG sites in introduced birds would be higher due to more gains and/or fewer losses of CpG sites than native birds. We anticipated this outcome despite the relatively lower genetic diversity expected of introduced populations due to genetic constraints, as found in previous studies (Schrey et al. 2011; Liebl et al. 2015). To better understand how these CpG sites might affect gene regulation, we also asked whether and how CpG numbers and location varied with respect to predicted TFBS. As transcription factors regulate gene expression and can be sensitive to DNA methylation, we expected that many CpG sites would fall within TFBS, and that gains or losses of CpG sites could both create and eliminate TFBS (Zhu et al. 2016).

To test our hypotheses, we quantified the number of CpG sites in ~500 base-pair (BP) regions upstream of the transcriptional start sites of each gene, then asked whether introduced birds had more CpG sites than native ones. As this study was opportunistic and relied on existing samples from sparrows collected from three native and eight introduced sites (Table 1), we could not probe how epigenetic potential in the three genes related within individuals nor how epigenetic potential varied among populations within native and introduced groups. To be as conservative as possible, we therefore simply compared total CpG sites, CpG site losses, and CpG site gains as aspects of epigenetic potential between birds assigned to native versus introduced groups. We also asked how any differences in epigenetic potential between these two groups were related to population genetic structure (i.e., genetic diversity) within populations, and we investigated whether CpG site locations overlapped predicted TFBS. However, our main goal was to test the simple possibility that introduced house sparrows maintain more epigenetic potential than native ones, in spite of the potentially extensive noise inherent to such a broad comparison.

Methods

Sample collection

We sequenced putative promoters of individual sparrows from three native (France, Germany, and Turkey) and eight introduced populations (Argentina [approximate age of introduction-1873], Brazil [1905], Florida [USA-1867], New Caledonia [1928], Kentucky [USA-1867], Kenya [1950], Senegal [1970], and South Africa [1900]; Table 1; see Supplementary Table S1 for sample sizes by population). For all populations, blood or tissue samples were stored in RNAlater (Qiagen) or Queen's lysis buffer until DNA extraction (Seutin et al. 1991). For DNA extraction protocol, see Supplementary Material S1.

Genetic and statistical analyses

Primers were designed to span a ~500 BP region upstream of the transcription start site (TSS) as this region may encompass the proximal promoters and/or other regulatory motifs, such as TFBS, which may be influenced by DNA methylation (Supplementary Table S2; Meissner et al. 2008; Kilvitis et al. 2019). For detailed primer design, PCR conditions, product cleanup, and sequencing protocol see Supplementary Material S1.

Chromatograms returned from sequencing were examined using Genome Compiler (Twist Bioscience). The number of CpG sites in each promoter was counted on each homologous chromosome separately to account for allelic variation. The frequency of homozygous CpG sites across all individuals (from any population) was used to determine whether a particular site had been lost or gained. A CpG site was determined to be gained if it was present in a focal individual but absent in >50% of all other individuals or lost if it was absent in a focal individual but present in >50% of all other individuals. Individuals were removed from analysis if all CpG sites or single nucleotide

polymorphisms (SNPs) could not be scored accurately (due to noise in chromatograms). For SNP calling procedure, see [Supplementary Material S1](#).

Each SNP was tested for linkage disequilibrium and deviation from Hardy–Weinberg equilibria using *GENEPOP* version 4.2 ([Rousset 2008](#)). The SNPs that were significantly linked were dropped and instead a single SNP was used to represent the linkage group in the following analysis. Observed (H_o) and expected heterozygosity (H_e) were estimated for each population for each gene in *GenAlEx* 6.5 ([Peakall and Smouse 2012](#)).

Sequences were analyzed using Alibaba version 2.1, which predicts TFBS using a database of eukaryotic transcription factors ([Grabe 2002](#)). Here, we only included TFBS that overlapped with CpG sites and excluded any TFBS that are not known to occur in birds. All locations found to have a CpG were examined and the type and number of TFBS were recorded, if present. As with CpG sites, TFBS was also counted on each homologous chromosome. Some TFBS could not be accurately identified on homologous chromosomes and were eliminated from the analysis ([Supplementary Table S1](#)).

Genetic structure of populations was described with Discriminant Analysis of Principal Components (DAPC). Using the R package, *ade4*, we performed a Principal Component (PC) analysis on genetic data from all individuals for which data were available for all three genes ([Jombart 2008](#)). All PC values were retained and used in the DAPC. K-means clustering was run and the best-fit model was selected using Bayesian Information Criteria ([Jombart 2008](#)). This analysis was performed with population, population group (native and introduced), and source population (native, introduced from European sources- Kentucky, Florida, New Caledonia, Argentina, Brazil, and introduced from non-European sources- Senegal, Kenya, South Africa) as predefined groups for all loci combined ([Anderson 2006](#)).

We found that CpG count data were non-normally distributed using a Shapiro–Wilk test, so Mann–Whitney U tests were used to determine if the mean total number, gains, and losses of CpG sites differed between introduced and native birds for each gene in GraphPad Prism version 8.0 ([Mann and Whitney 1947](#); [Shapiro and Wilk 1965](#); [Prism 8.0 2018](#)). As this study was opportunistic and relied on previously collected samples, we attempted but were unable to conduct a more robust analysis. Instead, we took this conservative approach to assess whether native and introduced birds differed in epigenetic potential. To assess whether differences in

epigenetic potential were related to genetic diversity, population group (native or introduced), or genetic diversity in each group, we used generalized linear models (GLMs) with total number, gains, and losses of CpG sites each as dependent variables with H_o (as a proxy for genetic diversity), population group (native and introduced), and their interaction as fixed effects. GLMs were run in R using the function *glm* with a Poisson error distribution ([R Core Team 2018](#)). For *TLR4*, samples from South Africa were not included as H_o was 0, potentially due to small sample size ($n=3$) for that gene. Figures were made in GraphPad Prism version 8.0 and in R using packages *ggplot2* and *ggpubr* ([Ginestet 2011](#); [Kassambara 2018](#); Prism 8.0 2018).

Ethics statement

All procedures were approved by USF IACUC (W3877 and IS0000636).

Results

Total CpG sites

In *TLR1B*, CpG sites were identified at 14 unique genomic locations (28 possible CpG sites). For this gene, introduced birds had significantly fewer CpG sites than native birds (Mann–Whitney U-test, $P=0.0079$; [Fig. 1a](#), [Table 1](#)). In *TLR2A*, CpG sites were identified at three unique genomic locations (six possible CpG sites). Introduced birds had significantly more CpG sites than native birds in this gene (Mann–Whitney U-test, $P=0.034$; [Fig. 1b](#), [Table 1](#)). In *TLR4*, CpG sites were identified at eight unique genomic locations (16 possible CpG sites). Introduced birds had significantly more CpG sites than native populations in this gene (Mann–Whitney U-test, $P=0.020$; [Fig. 1c](#), [Table 1](#)).

Gains of CpG sites

In *TLR1B*, there was a trend for introduced birds to gain fewer CpG sites than native birds in *TLR1B* (Mann–Whitney U-test, $P=0.0615$; [Fig. 1a](#), [Table 1](#)). In *TLR2A*, introduced birds gained significantly more CpG sites than native birds (Mann–Whitney U-test, $P=0.036$; [Fig. 1b](#), [Table 1](#)). In *TLR4*, introduced birds did not gain more CpG sites than native birds (Mann–Whitney U-test, $P=0.411$; [Fig. 1c](#), [Table 1](#)).

Losses of CpG sites

In *TLR1B*, introduced birds lost significantly more CpG sites than native birds (Mann–Whitney U-test, $P=0.015$; [Fig. 1a](#), [Table 1](#)). In *TLR2A*, introduced birds did not lose more CpG sites than native birds

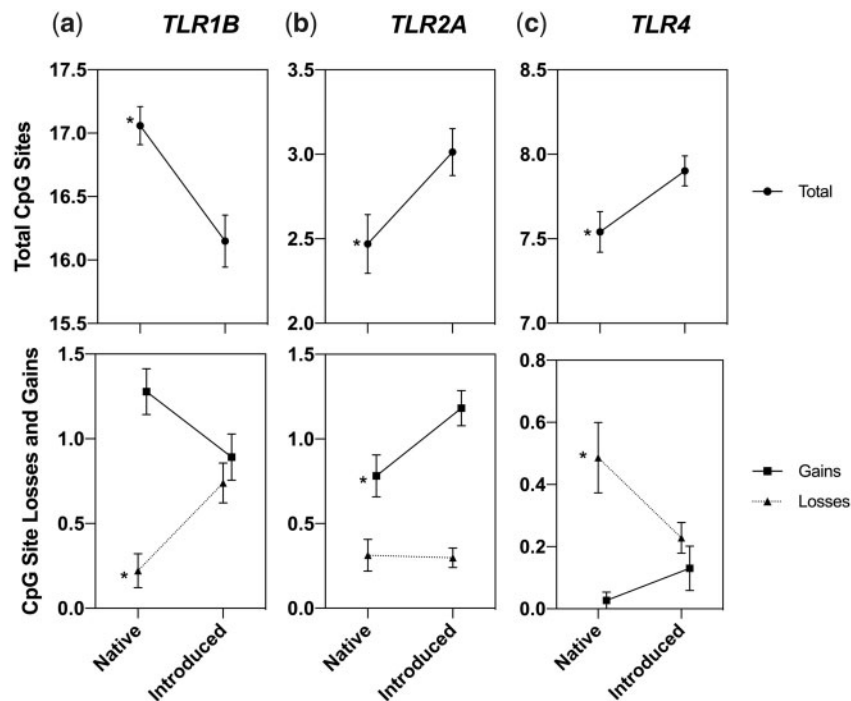


Fig. 1 Total number, gains, and losses of CpG sites for each gene and for native and introduced populations: (a) *TLR1B*, (b) *TLR2A*, and (c) *TLR4*. The top panel shows differences in the total number of CpG sites between native and introduced populations and the bottom panel shows differences between native and introduced populations in losses and gains of CpG sites. Asterisks represent significant differences between native and introduced populations. Note that the scales differ and CpG gains were plotted with horizontal jitter to minimize error bar overlap. Error bars are means $\pm$ SEM.

(Mann–Whitney U-test, $P = 0.937$; Fig. 1b, Table 1). In *TLR4*, introduced birds lost significantly fewer CpG sites than native birds (Mann–Whitney U-test, $P = 0.029$; Fig. 1c, Table 1).

Influence of genetic diversity on CpG traits

In *TLR1B* and *TLR2A*, there was no significant relationship between genetic diversity (i.e., observed heterozygosity) and total, gains, or losses of CpG sites (Supplementary Table S3, Supplementary Fig. S1a and b). In *TLR4*, there was no significant relationship between genetic diversity and the total number of CpG sites ($p_{H_o \times group} = 0.977$; Supplementary Table S3, Supplementary Fig. S1c). Due to low variation, the model did not converge for the relationship between genetic diversity and gains of CpG sites in *TLR4*. In *TLR4*, low genetic diversity correlated with fewer losses in introduced birds but not in native birds ($p_{H_o \times group} = 0.025$; Supplementary Table S3, Supplementary Fig. S1c).

Population structure

Six clusters predicted by DAPC did not discriminate population group (native vs. introduced), population, nor population source (European or non-European); therefore, we did not include this

information into subsequent analyses (Supplementary Figs. S2–S10).

TFBSs

In *TLR1B*, we identified seven locations where predicted TFBSs overlapped with CpG sites and 13 types of TFBS (Table 2). TFBS shared by native and introduced birds included: estrogen receptor alpha ($ER\alpha$), nuclear factor 1 (NF1), cyclic AMP-dependent transcription factor ATF-1 (*ATF1*), CCAAT/enhancer binding protein-alpha (*C/EBP α*), Rev-Erba, specificity protein 1 (Sp1), retinoic acid receptor (*RAR α*), COUP-TFII (*NR2F2*), heat shock factor 1 (HSF1), upstream stimulatory factor 1 (USF1), and CCAAT/enhancer binding protein-beta (*C/EBP β*). One type of TFBS was exclusively predicted in introduced birds: glucocorticoid receptor (GR). $ER\alpha$ was the most commonly predicted TFBS in *TLR1B* with a relative frequency of 19.4% in native and 20.3% in introduced populations. In *TLR2A*, we identified one location where a predicted TFBS overlapped with a CpG site and one type of TFBS (*C/EBP α* ; Table 2). In *TLR4*, we discovered four locations where predicted TFBS overlapped with CpG sites and five types of TFBS (Table 2). Predicted TFBS shared by native and introduced birds included Sp1, pituitary-

Table 2 Predicted TFBSs which overlapped with CpG site locations in each gene

Gene	TFBS	Frequency		Relative frequency		Occurrences
		Native	Introduced	Native (%)	Introduced (%)	
<i>TLR1B</i>	ER $\alpha^{a,b}$	68	164	19.4	20.3	2
	C/EBP $\alpha^{a,c}$	63	140	18.0	17.4	2
	NF1 a,c	34	92	9.7	11.4	2
	ATF1 a	32	82	9.1	10.2	1
	Elf-1 a	31	78	8.9	9.7	1
	Rev-ErbA c	34	82	9.7	10.2	1
	Sp1 c	25	40	7.1	5.0	2
	RAR α^c	22	38	6.3	4.7	2
	COUP-TFII c	23	36	6.6	4.5	1
	HSF1 a	2	30	0.6	3.7	1
	USF1 a	13	9	3.7	1.1	1
	C/EBP β^a	3	8	0.9	1.0	1
	GR a	0	8	0.0	1.0	1
<i>TLR2A</i>	C/EBP α^c	42	111	100.0	100.0	1
<i>TLR4</i>	Sp1	149	368	96.8	94.6	2
	Pit-1a a	1	12	0.7	3.1	1
	C/EBP α^a	4	5	2.6	1.3	1
	Elf-1	0	2	0.0	0.5	1
	IKZF1	0	2	0.0	0.5	1

^aRepresents a predicted TFBS created or eliminated by a gain or a loss of a CpG site.

^bRepresents a predicted TFBS that persists if a CpG site is lost or gained.

^cRepresents a predicted TFBS that is created or eliminated due to a CpG and an additional SNP.

Relative frequency = (TFBS/total TFBS per gene) $\times$ 100. Occurrences refer to the number of genomic locations the predicted TFBS (overlapping with CpG sites) were present.

specific positive transcription factor 1 (Pit-1a), and C/EBP α . Introduced birds had two unique TFBS: ETS Like-1 protein (Elf-1) and IKAROS family zinc finger 1 (IKZF1). Sp1 was the most commonly predicted TFBS in *TLR4*, with a relative frequency of 96.8% in native and 94.6% in introduced sparrows.

Discussion

In this study, we revealed that introduced and native house sparrows differed in CpG number, one form of epigenetic potential, in putative promoter regions of TLRs 1B, 2A, and 4. We expected to find higher epigenetic potential in introduced birds across all three genes compared with native birds in order to maintain high levels of phenotypic plasticity when encountering novel challenges during introductions and range expansions. We found that introduced sparrows had more total CpG sites than native sparrows in *TLR2A* and *TLR4*, but fewer total CpG sites in *TLR1B* (Fig. 1). Native and introduced birds also differed in the extent that they lost and gained CpG sites (Fig. 1). In *TLR1B*, introduced birds gained

marginally fewer and lost significantly more CpG sites (Fig. 1a), in *TLR2A* introduced birds gained more CpG sites (Fig. 1b), and in *TLR4* introduced birds lost fewer CpG sites than native ones (Fig. 1c). Additionally, many CpG sites fell within TFBS, and losses and gains of CpG sites both created and eliminated TFBS (Table 2). Lastly, these results were not driven by differences in genetic structure nor genetic diversity of populations (Supplementary Table S3, Supplementary Figs. S1–S10). Below, we discuss how epigenetic potential may have played a role in house sparrow introductions and range expansions.

Epigenetic potential in TLRs and its putative role in range expansions

The enemy release hypothesis predicts that introduced hosts should have fewer enemies in new areas, but the enemies they most commonly encounter will be generalists (Torchin et al. 2003). This pattern seems to hold for house sparrows with regard to hemosporidians (Marzal et al. 2011), but no comparisons to our knowledge exist regarding the microbes regulated by TLRs. However, activation of TLRs

incites costly yet effective inflammatory responses that can mitigate infections from generalist parasites (Iwasaki and Medzhitov 2015). TLRs are also able to respond to infections rapidly (Lin et al. 2006; Vaure and Liu 2014). For example, in introduced house sparrows, *TLR4* expression from whole blood samples more than doubled in 4 h after LPS administration (Martin et al. 2011). Further, DNA methylation within *TLRs* may mediate response to infection (Hennessy and McKernan 2016). For example, in chickens, DNA methylation in the promoter of *TLR4* and within an exon of *TLR2A* correlates to susceptibility to *Salmonella enteritidis* infection (Gou et al. 2012). As conditions may change day to day, and over the lifetime of the individual, plasticity in expression and methylation of *TLRs* via epigenetic potential may help alleviate the trade-off between inflammatory responses and other costly physiological processes (Martin et al. 2017).

The functional roles of these three TLRs surely influence the level of epigenetic potential they exhibit. We expect that parasite pressure varies among native and introduced populations, leading to differing levels of expression and plasticity (Lafferty and Kuris 1999; Lazzaro and Little 2009). Previous studies have shown more *TLR2* and *TLR4* expression in Kenyan range-edge populations, and that DNA methylation affects *TLR4* expression (Martin et al. 2014; Martin et al. 2015; Martin et al. 2017; Kilvitis et al. 2019). Together, these findings suggest that these TLRs were important for the Kenyan house sparrow range expansions (Martin et al. 2014; Martin et al. 2015; Martin et al. 2017; Kilvitis et al. 2019). Past studies have not investigated *TLR1B*, thus we do not know its relevance to range expansions. In general, we know little about it in passerine immunology.

Possible consequences of gains, losses, and total number of CpG sites

Native and introduced birds maintained significantly different numbers of CpG sites in the three genes we investigated (Fig. 1), however, genes differed in the total CpG sites they maintained. *TLR1B* had the most total CpG sites of any gene, whereas the other two genes had substantially fewer sites (Table 1). The total number of CpG sites in a gene influences its expression, so subsequently, the CpG content of a gene might influence its level of plasticity attainable and hence fitness (Yang et al. 2014). Indeed, the total number of CpG sites may be indicative of the difference in capacity for DNA methylation to produce variable phenotypes (Feinberg and Irizarry 2010;

Flores and Amdam 2011; Kilvitis et al. 2017). For range expansions, we expect that the more CpG sites an individual have the more successful it will be during colonization because it can regulate gene expression more precisely (Yang et al. 2014). On the contrary, this might not be the only mechanism whereby epigenetic potential affects range expansions. As the number of CpG sites increases, the potential for plasticity could decrease, as each additional CpG site may have smaller effect size. In this light, whereas *TLR1B* might have had the most CpG sites, its capacity for plasticity gained through methylation at one CpG site may be comparatively less than from *TLR2A* or *TLR4*. We fully expect that methylation at some CpG sites will have larger effects on gene expression than others, and we are actively testing this hypothesis (Lioznova et al. 2019). Single CpG sites can be correlated to gene expression dependent on tissue or cell-type examined (Kitazawa and Kitazawa 2007; Lioznova et al. 2019; Luo et al. 2019). In future studies, it will be important to investigate epigenetic potential with a regard to the CpG content and identity and cellular context (Xin et al. 2011; Lioznova et al. 2019).

Whereas our data reveal losses and gains of several CpG sites in each promoter, the loss or gain of even a single CpG can impact gene expression (Polesskaya et al. 2006; Williams et al. 2014; Zhou et al. 2015). For example, in humans, a loss of a CpG site in the promoter of the glycine N-methyltransferase (*GNMT*) gene led to varied methylation status at that location; methylation status had a direct effect on *GNMT* transcription and metabolism of methionine (Williams et al. 2014). The ability to metabolize methionine is directly linked to the propensity to develop cardiovascular disease, cancer, dementia, and stroke in humans (Williams et al. 2014). Our data reveal examples of individuals from introduced populations losing as many as two or gaining as many as six CpG sites in one promoter (Table 1). Here, we could not measure methylation or gene expression, but the aforementioned studies have revealed that the subtle differences we discovered in promoter architecture may be sufficient to affect both. Other factors influencing levels of epigenetic potential may include the evolutionary history of the gene (i.e., gene duplication events), the underlying genetic architecture of the locus (i.e., high CpG content or not), and/or complex, epistatic interactions (Rodin et al. 2005; Huang et al. 2011; Velová et al. 2018).

Effect on TFBSs

One mechanism by which epigenetic potential can impact gene expression is through the binding of

transcription factors. Our ultimate goal was to describe whether CpG sites were impacting TFBS, as these are hotspots for the regulation of gene expression. Many CpG sites fell within predicted TFBS, and losses and gains of CpG sites can eliminate or create a TFBS, which can directly impact transcription (Table 2; Zemojtel et al. 2009; Wang et al. 2019). Importantly, methylation at CpG sites can create a TFBS independent of changes in sequence, and methylation at CpG sites within TFBS can (1) prevent transcription factors from binding or (2) change the affinity of transcription factors binding to that TFBS, both of which can impact the regulation of gene expression (Ung et al. 2014; Zhu et al. 2016). Further, upon binding to methylated sequences, transcription factors can also alter the DNA methylation state of both proximal CpG sites and those within the TFBS (Zhu et al. 2016).

The most commonly predicted TFBS (highest relative frequency) in *TLR1B* was ER α , which is sensitive to DNA methylation (Table 2; Ung et al. 2014). The binding of ER α to its TFBS can prevent methylation of the CpG within the TFBS (Ung et al. 2014). In *TLR2A*, the only predicted TFBS was CCAAT/enhancer binding protein- α (C/EBP α), which was eliminated if the CpG site was lost. C/EBP α can bind either methylated or unmethylated DNA, but binds to methylated sequences with a higher affinity (Table 2; Zhu et al. 2016). The CpG site in *TLR4* at which high levels of methylation correlated with low levels of expression (in a previous study) fell within Sp1, the most commonly predicted TFBS we observed across all sparrows (Table 2; Kilvitis et al. 2019). Sp1 is a transcriptional activator of *TLR2* and *TLR4*, and can bind to methylated or unmethylated DNA, but gene expression is reduced when DNA methylation is present within this TFBS or proximal sites (Furuta et al. 2008; Kim et al. 2016). Further, binding of Sp1 in the *TLR4* promoter may protect from methylation and/or lead to localized demethylation, potentially through competition with DNMT3a for binding (Brandeis et al. 1994; Furuta et al. 2008; Lo et al. 2017). Our data suggests that variation in CpG site number impacted the prediction of TFBS (e.g., a loss or a gain of a CpG site can correspond with the loss or gain of a predicted TFBS). Whereas additional data are needed, CpG sites within TFBS could provide direct routes for environmental conditions to impact gene expression.

Genetic variation and other potential influences on population differences

In this study, introduced house sparrows maintained more CpG sites in both *TLR2A* and *TLR4* promoters

than native birds (Fig. 1). While introduced populations often experience reductions in genetic diversity due to bottlenecks/founder effects at the time of introduction, we expect that the initial founders of these introduced populations had high epigenetic potential in order to cope with challenges via phenotypic plasticity (Schrey et al. 2011; Schrey et al. 2012; Liebl et al. 2015; Kilvitis et al. 2017). In other words, there may be a minimum threshold level of epigenetic potential, where only individuals with sufficient epigenetic potential survive the initial introduction or range expansion. Over time, as populations persist in these areas and become accustomed to the surrounding environment, phenotypic plasticity may be costly and impose lags or mismatches with the environment, compared to genetically-fixed phenotypic responses (Bonduriansky et al. 2012; Snell-Rood 2012). Consequently, selection could act to maintain epigenetic potential during the initial phase of the introduction and subsequent range expansion, but winnow epigenetic potential as canalized genetic variants become more common.

Our data indicate that neither genetic diversity nor population genetic structure underpin the differences in epigenetic potential we detected (Supplementary Table S3, Supplementary Figs S1–S4). Similar outcomes could hold true for other genes or be limited to genes such as *TLR2A* and *TLR4* because of their presumed advantage during colonization. Indeed, overall epigenetic potential might be over-represented with genes particularly integral to function and fitness in new areas. CpG sites are strongly preserved in genes important in development in both the coding regions and in CpG islands, putatively because of their functionality for gene transcription (Branciamore et al. 2010; Branciamore et al. 2015). In further support, a study of several cnidarians revealed that CpG site content was highest in regions proximal to the TSS of genes found to be important to respond to environmental stressors such as temperature and salinity (Marsh et al. 2016). Additional research will be needed to elucidate whether epigenetic potential exhibits similar trends across the genome or whether it occurs predominantly in particular genes or gene regions.

Conclusion

Despite the statistical limitations associated with our opportunistic dataset, our data suggest that epigenetic potential in some TLRs could be playing a role in house sparrow introductions. Because these sites are directly heritable, we expect that they might

play important evolutionary roles during range expansions by acting as capacitors of phenotypic plasticity, as well as genomic locations that respond to methylation via mutation due to deamination (Feinberg and Irizarry 2010; Flores and Amdam 2011; Kilvitis et al. 2017). Importantly, this work only implicates CpG site number variation in specific genes; the relationship among the number of CpG sites, TFBS, DNA methylation, gene expression, and fitness require direct investigations.

Acknowledgments

We thank Jaime Zolik, Margaret Dixon, Angelo Pata, Miranda Solly, Alexandra Sauser, and Kimberly Mungal for their help identifying SNP and CpG locations, and the Martin lab for their many comments and suggestions. We thank our international collaborators David Westneat, Nicolas Barré, Marcos Lima, Pablo Cuervo, Paula Presti, Andrea Previtali, Gabriele Sorci, Bruno Faivre, Staffan Bensch, Helena Westerdaal, Tamer Albayrak, and Colleen Downs for samples.

Funding

We thank the National Science Foundation (0920475, 1257773, 1656551), and the Pritzker Laboratory for Molecular Systematics and Evolution, operated with support from the Pritzker Foundation. H.E.H. acknowledges support from Sigma XI (G2016100191872782), the Porter Family Foundation, the American Ornithological Society Hesse Grant, and the American Museum of Natural History Frank M. Chapman Memorial Fund. J.D.M. was supported, in part, by El Fondo Nacional de Desarrollo Científico, Tecnológico y de Innovación Tecnológica (44-2019-FONDECYT) during a portion of this research. Finally, we thank SICB divisions DAB, DCE, DCPB, DEDB, DEDE, and DNNSB, and the Company of Biologists (Scientific Meeting Grant-EA87) and the Society for Experimental Biology (Scientific Meeting Sponsorship) for their gracious support of the symposium “Epigenetic Variation in Endocrine Systems.”

Authors' contributions

H.E.H., H.J.K., A.W.S., and L.B.M. designed the study. J.D.M. conducted sequencing. B.K. and H.E.H. conducted transcription factor binding site analysis. H.E.H. prepared samples, conducted PCR, completed data analysis, and wrote the first draft of the manuscript. All authors contributed to subsequent drafts of the manuscript.

Data accessibility

Data are available at <https://doi.org/10.6084/m9.figshare.12389894.v1>.

Supplementary data

Supplementary data available at *ICB* online.

References

- Anderson TR. 2006. Biology of the ubiquitous house sparrow: from genes to populations. New York: Oxford University Press.
- Bonduriansky R, Crean AJ, Day T. 2012. The implications of nongenetic inheritance for evolution in changing environments. *Evol Appl* 5:192–201.
- Branciamore S, Chen ZX, Riggs AD, Rodin SN. 2010. CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors. *Proc Natl Acad Sci U S A* 107:15485–90.
- Branciamore S, Rodin AS, Gogoshin G, Riggs AD. 2015. Epigenetics and evolution: transposons and the stochastic epigenetic modification model. *AIMS Genet* 2:148–62.
- Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Names A, Temper V, Razin A, Cedar H. 1994. Spl elements protect a CpG island from de novo methylation. *Nature* 371:435–8.
- Feinberg AP, Irizarry RA. 2010. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. *Proc Natl Acad Sci U S A* 107:1757–64.
- Flores KB, Amdam GV. 2011. Deciphering a methylome: what can we read into patterns of DNA methylation? *J Exp Biol* 214:3155–63.
- Flores KB, Wolschin F, Amdam GV. 2013. The role of methylation of DNA in environmental adaptation. *Integr Comp Biol* 53:359–72.
- Furuta T, Shuto T, Shimasaki S, Ohira Y, Suico MA, Gruenert DC, Kai H. 2008. DNA demethylation-dependent enhancement of toll-like receptor-2 gene expression in cystic fibrosis epithelial cells involves SP1-activated transcription. *BMC Mol Biol* 9:1–15.
- Genome Compiler, Twist Bioscience (<https://designer.genomecompiler.com/>).
- Ginestet C. 2011. ggplot2: elegant graphics for data analysis. *J R Stat Soc Ser A* 174:245–6.
- Gou Z, Liu R, Zhao G, Zheng M, Li P, Wang H, Zhu Y, Chen J, Wen J. 2012. Epigenetic modification of TLRs in leukocytes is associated with increased susceptibility to *Salmonella enteritidis* in chickens. *PLoS One* 7:e33627–10. doi : 10.1371/journal.pone.0033627.
- Grabe N. 2002. AliBaba2: context specific identification of transcription factor binding sites. *Silico Biol* 2:S1–15.
- Hanson HE, Zolik JE, Martin LB. 2020. House sparrow (*Passer domesticus* Linnaeus, 1758). Chapter 11. In: Downs CT, Lorinda AH, editors. *Invasive birds: global trends and impacts*. Wallingford: CABI.
- Hennessy C, McKernan DP. 2016. Epigenetics and innate immunity: the “unTolld” story. *Immunol Cell Biol* 94:631–9.
- Huang Y, Temperley ND, Ren L, Smith J, Li N, Burt DW. 2011. Molecular evolution of the vertebrate TLR1 gene family - a complex history of gene duplication, gene

- conversion, positive selection and co-evolution. *BMC Evol Biol* 11:149.
- Iwasaki A, Medzhitov R. 2015. Control of adaptive immunity by the innate immune system. *Nat Immunol* 16:343–53.
- Jombart T. 2008. ADEGENET: a R package for multivariate analysis of genetic markers. *Bioinformatics* 24:1403–5.
- Kassambara A. 2018. ggpubr: “ggplot2” based publication ready plots. R Packag. R package version 0.1.8 (<https://cran.r-project.org/web/packages/ggpubr/ggpubr.pdf>).
- Kilvitis HJ, Ardia DR, Thiam M, Martin LB. 2018. Corticosterone is correlated to mediators of neural plasticity and epigenetic potential in the hippocampus of Senegalese house sparrows (*Passer domesticus*). *Gen Comp Endocrinol* 269:177–83.
- Kilvitis HJ, Martin LB, Hanson HE, Schrey AW. 2017. Epigenetic potential as a mechanism of phenotypic plasticity in vertebrate range expansions. *Integr Comp Biol* 57:385–95.
- Kilvitis HJ, Schrey AW, Ragsdale AK, Berrio A, Phelps SM, Martin LB. 2019. DNA methylation predicts immune gene expression in introduced house sparrows *Passer domesticus*. *J Avian Biol* 50 (doi:10.1111/jav.01965).
- Kim TW, Lee SJ, Oh BM, Lee H, Uhm TG, Min JK, Park YJ, Yoon SR, Kim BY, Kim JW, et al. 2016. Epigenetic modification of TLR4 promotes activation of NF- κ B by regulating methyl-CpG-binding domain protein 2 and Sp1 in gastric cancer. *Oncotarget* 7:4195–209.
- Kitazawa R, Kitazawa S. 2007. Methylation status of a single CpG locus 3 bases upstream of TATA-box of receptor activator of nuclear factor- κ B ligand (RANKL) gene promoter modulates cell- and tissue-specific RANKL expression and osteoclastogenesis. *Mol Endocrinol* 21:148–58.
- Lafferty KD, Kuris AM. 1999. How environmental stress affects the impacts of parasites. *Limnol Oceanogr* 44:925–31.
- Lande R. 2015. Evolution of phenotypic plasticity in colonizing species. *Mol Ecol* 24:2038–45.
- Lazzaro BP, Little TJ. 2009. Immunity in a variable world. *Philos Trans R Soc Lond B Biol Sci* 364:15–26.
- Liebl AL, Martin LB. 2012. Exploratory behaviour and stressor hyper responsiveness facilitate range expansion of an introduced songbird. *Proc Biol Sci* 279:4375–81.
- Liebl AL, Martin LB. 2014. Living on the edge: range edge birds consume novel foods sooner than established ones. *Behav Ecol* 25:1089–96.
- Liebl AL, Schrey AW, Andrew SC, Sheldon EL, Griffith SC. 2015. Invasion genetics: lessons from a ubiquitous bird, the house sparrow *Passer domesticus*. *Curr Zool* 61:465–76.
- Liebl AL, Schrey AW, Richards CL, Martin LB. 2013. Patterns of DNA methylation throughout a range expansion of an introduced songbird. *Integr Comp Biol* 53:351–8.
- Lin FY, Chen YH, Tasi JS, Chen JW, Yang TL, Wang HJ, Li CY, Chen YL, Lin SJ. 2006. Endotoxin induces toll-like receptor 4 expression in vascular smooth muscle cells via NADPH oxidase activation and mitogen-activated protein kinase signaling pathways. *Arterioscler Thromb Vasc Biol* 26:2630–7.
- Lioznova AV, Khamis AM, Artemov AV, Besedina E, Ramensky V, Bajic VB, Kulakovskiy IV, Medvedeva YA. 2019. CpG traffic lights are markers of regulatory regions in human genome. *BMC Genom* 20:1–12.
- Lo W, Chuang J, Ko C, Chou S, Chen R, Chang K, Hung J, Su W, Chang W, Hsu T. 2017. P05.06 upregulation of CYP17A1 by Sp1-mediated DNA demethylation confers temozolomide resistance through DHEA-mediated protection in glioma. *Neuro Oncol* 19:iii48.
- Luo X, Yang C, Wei Y. 2019. Detection of cell-type-specific risk-CpG sites in epigenome-wide association studies. *Nat Commun* 10:3113.
- Mann HB, Whitney DR. 1947. On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 18:50–60.
- Marsh AG, Hoadley KD, Warner ME. 2016. Distribution of CpG motifs in upstream gene domains in a reef coral and sea anemone: implications for epigenetics in cnidarians. *PLoS One* 11:e0150840–14.
- Martin LB, Coon CAC, Liebl AL, Schrey AW. 2014. Surveillance for microbes and range expansion in house sparrows. *Proc Biol Sci* 281:20132690.
- Martin LB, Kidd L, Liebl AL, Coon C. 2011. Captivity induces hyper-inflammation in the house sparrow (*Passer domesticus*). *J Exp Biol* 214:2579–85.
- Martin LB, Kilvitis HJ, Brace AJ, Cooper L, Haussmann MF, Mutati A, Fasanello V, O’Brien S, Ardia DR. 2017. Costs of immunity and their role in the range expansion of the house sparrow in Kenya. *J Exp Biol* 220:2228–35.
- Martin LB, Liebl AL, Kilvitis HJ. 2015. Covariation in stress and immune gene expression in a range expanding bird. *Gen Comp Endocrinol* 211:14–9.
- Marzal A, Ricklefs RE, Valkiūnas G, Albayrak T, Arriero E, Bonneaud C, Cziráj G, Ewen J, Hellgren O, Hořáková D, et al. 2011. Diversity, loss, and gain of malaria parasites in a globally invasive bird. *PLoS One* 6:e21905–8.
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, et al. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454:766–70.
- Peakall R, Smouse PE. 2012. GenALEX 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–9.
- Phillips BL, Brown GP, Shine R. 2010. Life-history evolution in range-shifting populations. *Ecology* 91:1617–2439.
- Pigliucci M. 2005. Evolution of phenotypic plasticity: where are we going now? *Trends Ecol Evol* 20:481–6.
- Polesskaya OO, Aston C, Sokolov BP. 2006. Allele C-specific methylation of the 5-HT2A receptor gene: evidence for correlation with its expression and expression of DNA methylase DNMT1 oxana. *J Neurosci Res* 83:362–73.
- Prism 8.0, GraphPad Software. 2018. La Jolla (CA) (www.graphpad.com).
- R Core Team. 2018. R: a language and environment for statistical computing (<http://www.r-project.org/>).
- Rodin SN, Parkhomchuk DV, Riggs AD. 2005. Epigenetic changes and repositioning determine the evolutionary fate of duplicated genes. *Biochem* 70:559–67.
- Rousset F. 2008. GENEPOP’007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Mol Ecol Resour* 8:103–6.
- Schrey AW, Coon CAC, Grispo MT, Awad M, Imboma T, McCoy ED, Mushinsky HR, Richards CL, Martin LB. 2012. Epigenetic variation may compensate for decreased genetic variation with introductions: a case study using house

- sparrows (*Passer domesticus*) on two continents. *Genet Res Int* 2012:1–7.
- Schrey AW, Grispo M, Awad M, Cook MB, McCoy ED, Mushinsky HR, Albayrak T, Bensch S, Burke T, Butler LK, et al. 2011. Broad-scale latitudinal patterns of genetic diversity among native European and introduced house sparrow (*Passer domesticus*) populations. *Mol Ecol* 20:1133–43.
- Schrey AW, Liebl AL, Richards CL, Martin LB. 2014. Range expansion of house sparrows (*Passer domesticus*) in Kenya: evidence of genetic admixture and human-mediated dispersal. *J Hered* 105:60–9.
- Seutin G, White BN, Boag PT. 1991. Preservation of avian blood and tissue samples for DNA analyses. *Can J Zool* 69:82–90.
- Shapiro SS, Wilk MB. 1965. An analysis of variance test for normality (complete samples). *Biometrika* 52:591–611.
- Sheldon EL, Schrey A, Andrew SC, Ragsdale A, Griffith SC. 2018. Epigenetic and genetic variation among three separate introductions of the house sparrow (*Passer domesticus*) into Australia. *R Soc Open Sci* 5:172185.
- Snell-Rood EC. 2012. Selective processes in development: implications for the costs and benefits of phenotypic plasticity. *Integr Comp Biol* 52:31–42.
- Torchin ME, Lafferty KD, Dobson AP, McKenzie VJ, Kuris AM. 2003. Introduced species and their missing parasites. *Nature* 421:628–30.
- Ung M, Ma X, Johnson KC, Christensen BC, Cheng C. 2014. Effect of estrogen receptor α binding on functional DNA methylation in breast cancer. *Epigenetics* 9:523–32.
- Vaure C, Liu Y. 2014. A comparative review of toll-like receptor 4 expression and functionality in different animal species. *Front Immunol* 5:1–15.
- Velová H, Gutowska-Ding MW, Burt DW, Vinkler M. 2018. Toll-like receptor evolution in birds: gene duplication, pseudogenization, and diversifying selection. *Mol Biol Evol* 35:2170–84.
- Wang H, Lou D, Wang Z. 2019. Crosstalk of genetic variants, allele-specific DNA methylation, and environmental factors for complex disease risk. *Front Genet* 10:1–15.
- Williams SR, Yang Q, Chen F, Liu X, Keene KL, Jacques P, Chen WM, Weinstein G, Hsu FC, Beiser A, et al. 2014. Genome-wide meta-analysis of homocysteine and methionine metabolism identifies five one carbon metabolism loci and a novel association of ALDH1L1 with ischemic stroke. *PLoS Genet* 10:e1004214.
- Xin Y, O'Donnell AH, Ge Y, Chanrion B, Milekic M, Rosoklija G, Stankov A, Arango V, Dwork AJ, Gingrich JA, et al. 2011. Role of CpG context and content in evolutionary signatures of brain DNA methylation. *Epigenetics* 6:1308–18.
- Yang MQ, Yoshigoe K, Yang W, Tong W, Qin X, Dunker AK, Chen Z, Arbania HR, Liu JS, Niemierko A, et al. 2014. Relating gene expression evolution with CpG content changes. *BMC Genom* 15:1–10.
- Zemojtel T, Kielbasa SM, Arndt PF, Chung HR, Vingron M. 2009. Methylation and deamination of CpGs generate p53-binding sites on a genomic scale. *Trends Genet* 25:63–6.
- Zhou D, Li Z, Yu D, Wan L, Zhu Y, Lai M, Zhang D. 2015. Polymorphisms involving gain or loss of CpG sites are significantly enriched in trait-associated SNPs. *Oncotarget* 6:39995–40004.
- Zhu H, Wang G, Qian J. 2016. Transcription factors as readers and effectors of DNA methylation. *Nat Rev Genet* 17:551–65.