Title: Epigenetic Potential and DNA Methylation in an Ongoing House Sparrow (*Passer domesticus*) Range Expansion

Authors:

Haley E. Hanson,[a,1] Chengqi Wang,[a] Aaron W. Schrey,[b] Andrea L. Liebl,[c] Mark Ravinet,[d] Rays H. Y. Jiang,[a] Lynn B. Martin[a,1]

Author Affiliations:

[a] University of South Florida, Global and Planetary Health, Tampa, FL

[b] Georgia Southern University Armstrong Campus, Department of Biology, Savannah, GA

[c] University of South Dakota, Department of Biology, Vermillion, SD

[d] University of Nottingham, School of Life Sciences, Nottingham, UK

Corresponding Author:

[1] To whom correspondence may be addressed. University of South Florida, Global and Planetary Health, Tampa, FL. Email: haleyehanson@gmail.com or lbmartin@usf.edu

Running Title: Epigenetic Potential in House Sparrows

Keywords: range expansions, introduced species, CpG sites, phenotypic plasticity, invasion, non-native

MS type: Article

ABSTRACT:

During range expansions, organisms can use epigenetic mechanisms to adjust to conditions in novel areas by altering gene expression and enabling phenotypic plasticity. Here, we predicted that the number of CpG sites within the genome, one form of epigenetic potential, would be important for successful range expansions because DNA methylation can modulate gene expression, and consequently plasticity. We asked how the number of CpG sites and DNA methylation varied across five locations in the ~70 year-old Kenyan house sparrow (*Passer domesticus*) range expansion. We found that the number of CpG sites was highest towards the vanguard of the invasion and decreased towards the range core. Analysis suggests that this pattern may have been driven by selection, favoring birds with more CpG sites at the range edge. However, we cannot rule out other processes including non-random gene flow. Additionally, DNA methylation did not change across the range expansion, nor was it more variable. We hypothesize that as new areas are colonized, epigenetic potential may be selectively advantageous early but eventually be replaced by less plastic and perhaps genetically-canalized traits as populations adapt to local conditions. Although further work is needed on epigenetic potential, this form (CpG number) appears to be a promising mechanism to investigate as a driver of expansions via capacitated phenotypic plasticity in other natural and anthropogenic range expansions.

Introduction

Epigenetic modifications, such as DNA methylation, play a critical role in linking environmental variation to phenotypic variation by modifying how genes are expressed (Smith and Meissner 2013). In vertebrates, DNA methylation and other molecular epigenetic mechanisms are instrumental to cellular and tissue differentiation during development. Epigenetic modifications can also affect evolutionarily-relevant behavioral, morphological, and physiological plasticity (Feinberg 2007; Bock et al. 2012). Epigenetic variation, including DNA methylation patterns, predominates in particular genomic regions (e.g., CpG dinucleotides in vertebrates), so depending on genomic makeup, individuals might differ in the extent to which their genomes can be modified epigenetically (Feinberg and Irizarry 2010). Moreover, when genetic variation associated with epigenetic marks occurs in genes that affect fitness, natural selection may follow, leading to differences in epigenetic potential among populations (Feinberg and Irizarry 2010).

Epigenetic potential (i.e., genomic differences in the capacity for epigenetic mechanisms to mediate phenotypic variation) might become more common or rare in populations depending on the selective value of phenotypic plasticity in a given area (Kilvitis et al. 2017). For instance, during range expansions, individuals face relatively novel threats and opportunities that require rapid phenotypic responses. However, they also risk diminished organismal performance because of low population genetic diversity or founder effects, which could affect the predominance of gene variants associated with low fitness (Lee 2002). Thus, epigenetic potential may be selected for early in invasions and during expansions, but as populations adapt to colonized environments, the local value of epigenetic potential might wane as phenotypically plastic genotypes are outcompeted by genotypes with genetically canalized traits (Kilvitis et al. 2017). Alternatively, as epigenetic potential is underlain genetically, differences in epigenetic potential may arise via

mechanisms unrelated to selection. Population-level processes such as the dispersal of individuals across the invasion front could lead to differences in epigenetic potential for reasons unrelated to adavptive plasticity. Non-random dispesal of indivudals could lead to non-random gene flow contributing to patterns of epigenetic potential (Edelaar and Bolnick 2012). Another process contributing to differences in epigenetic potential could be the expression of transposable elements (TEs). During periods of stress, such as may be experienced during invasions, TEs may be activated, leading to their insertion across the genome and resulting in increased genetic variation and subsequently, higher epigenetic potential (Stapley et al. 2015; Marin et al. 2019). Here, we investigated epigenetic potential across an ongoing range expansion of an extremely successful introduced species, the house sparrow (*Passer domesticus*). We also explored genetic variation across the expansion to help parse the scenarios under which differences in epigenetic potential could have arisen. We predicted that epigenetic potential would be highest towards the advancing range edge where birds arrived most recently and decline towards the core of the population where birds were initially introduced.

House sparrows expanded out of the Middle East thousands of years ago as agriculture spread into Europe (Ravinet et al. 2018). Over the last 170 years, this species has achieved a near-global distribution, largely due to intentional or accidental movements by humans (Ravinet et al. 2018; Hanson et al. 2020*b*). As house sparrows spread globally, they have had to cope with a wide range of biotic and abiotic novelties. Concurrently, some introduced groups are expected to have faced founder effects and genetic bottlenecks, yet despite these challenges, house sparrows endure and often thrive in non-native areas, exhibiting extensive seemingly adaptive phenotypic variation across much of the globe (Johnston and Selander 1971; Blem 1973; Kendeigh 1976; Parkin and Cole 1985; Schrey et al. 2011). In one of their most recent range expansions in Kenya, house

sparrow trait variation, including the regulation of glucocorticoid hormones, immune genes, and several behaviors, track relative population age (Liebl and Martin 2012, 2013; Martin and Liebl 2014; Martin et al. 2014, 2017). This paradox of extensive trait variation when genetic variation is comparatively low (relative to native populations) might be resolved by phenotypic plasticity and/or epigenetic compensation, a result observed previously in this system (Schrey et al. 2012).

Epigenetic potential may take several forms, but here we investigate one form: the number of cytosine-phosphate-guanine sites, or CpG sites, in the genome, the genetic motifs upon which DNA can be methylated (Branciamore et al. 2010; Zhu et al. 2016; Kilvitis et al. 2017). In vertebrates, DNA methylation occurs when a methyl group is added to the fifth carbon position of a cytosine predominately within a CpG site (Smith and Meissner 2013). Depending on where DNA methylation occurs in the genome, it may suppress or enhance gene expression (Jones 2012). In principle, each CpG site represents an opportunity for DNA methylation to alter gene expression (Branciamore et al. 2010). DNA methylation can originate stochastically, due to underlying genetic variation (in *cis* or *trans*), or in response to particular environmental simuli (Richards 2006; Sepers et al. 2019). Indeed, DNA methylation can be induced or eliminated rapidly following stimulation from a range of factors (e.g., diet, transcription factor activity, stressors, etc.), or may remain fixed after induction during development (Richards 2006; Smith and Meissner 2013; Wu and Zhang 2014). In the context of range expansions or introductions, DNA methylation originating in response to external conditions during development and adulthood is expected to be particularly critical for allowing individuals to respond to unfamiliar environments. Regardless of the origin of DNA methylation, this type of epigenetic potential reflects the capacity for DNA methylation to occur.

Already, there is evidence that epigenetic potential of this form (i.e., CpG number) i) differs across house sparrow populations and ii) capacitates gene expression in one microbial surveillance gene, Toll-like receptor 4 (*TLR4*). First, epigenetic potential in the putative promoter region of *TLR4* was higher in introduced compared to native house sparrows from across the globe (Hanson et al. 2020*a*). There, the pattern was argued to reflect evidence for epigenetic potential fostering invasions because more plasticity in immune responses should be favorable in novel areas where many pathogens would also be novel to their hosts. Second, epigenetic potential in the gene promoter affected *TLR4* expression over time in house sparrows (Hanson et al. 2021). Specifically, birds with higher epigenetic potential expressed more *TLR4* in blood over the course of the experiment, and there was also an unexpected sex difference such that females with higher epigenetic potential exhibited greater reversibility in gene expression than those with low epigenetic potential. Epigenetic potential also had tissue-specific effects on *TLR4* expression, suggesting a further means by which epigenetic potential could underpin phenotypic plasticity. Altogether, these results suggest that epigenetic potential might enable more variation in gene expression among tissues and over time to provide a comparatively malleable response to parasites (Hanson et al. 2021).

Building off of these studies, here we asked whether epigenetic potential varied predictably across the ~70 year old Kenyan range expansion, by using a reduced representation library-based sequencing to compare CpG number and DNA methylation among house sparrows from five cities (Schield et al. 2016). These five cities spanned the extent of the ongoing invasion, from the location of intitial introduction into Kenya (range core), where house sparrows have been established for significant periods of time (~70 years), to the range edge at the time where birds arrived relatively recently (Martin et al. 2010; Liebl and Martin 2012; Schrey et al. 2014). We queried i) how the

number of CpG sites changed across the range expansion; ii) whether existing CpG sites were being lost or novel CpG sites gained; and iii.) whether these patterns potentially arose via selection or other processes. From our past research in the Kenyan house sparrow system, we predicted that epigenetic potential would be highest towards the range edge where the selective value of phenotypic plasticity should be high compared to the range core. We hypothesized that towards the range edge, positive section would be the most evident process, leading to less frequent losses of existing CpG sites. We also considered non-random gene flow and the activity of transposable elements as alternative mechanisms that could generate patterns of epigenetic potential. Non-random gene flow is not expected to generate any specific pattern of epigenetic potential across the range expansion, but if this process were occurring, we would expect to see high levels of population structure due to the dispersal of similar genotypes across the range expansion. If the activation of transposable elements was occurring, we would expect to see not only high levels of epigenetic potential at the range edge, but also high levels of genetic diversity. Moreover, we would also expect to see gains of novel CpG sites due to the insertion of new genetic material, rather than an increase in frequency of exisiting CpG sites. Lastly, from our epigenetic data, we described iv) global DNA methylation patterns amongst cities across the expansion. We did not make specific predictions about the directionality of DNA methylation across the range expansion as DNA methylation patterns can be highly context dependent, contingent on the gene or gene region where the CpG site is located, the tissue from which they are generated, and the environmental conditions to which individuals were exposed over their lives. We could not account for any of these factors here, as we have no information about prior experiences of these birds that might have affected gene expression patterns, nor specific expectations about methylation variability among and within genomic regions.

Materials and Methods

House sparrow sampling

House sparrows were captured via mist nets in five cities across southern Kenya in February through May of 2013. House sparrows were initially introduced to Mombasa in the 1950s, so as in other studies, we used distance (in km) from Mombasa as a proxy for time since introduction (see Liebl and Martin 2012; Schrey et al. 2014; Martin, Liebl, and Kilvitis 2015; Martin et al. 2014, 2010). This house sparrow range expansion seems to have occured northwestwards from Mombasa following the Mombasa highway (Schrey et al. 2014). House sparrows reached Nairobi between the late 1980s and mid 1990s and spread west since 2000, supporting the assumption that house sparrows have occupied different cities for differing amounts of time (Martin et al. 2010; Liebl and Martin 2012; Schrey et al. 2014). Genetic analyses indicate that populations are structured and hence behave as independent units, but they also indicate that admixture is still occurring among cities (Schrey et al. 2014).

Birds in the present study were captured from the following cities: Mombasa (0 km), Voi (160 km), Nairobi (500 km), Nakuru (650 km), and Kakamega (850 km). The distance did not differ for house sparrows captured within the same city (i.e. all house sparrows captured in Nairobi were considered 500 km away from Mombasa in these analyses, as all Nairobi sparrows were captured from the same specific location within Nairobi). Individuals were brought into captivity and housed at ambient conditions with *ad libitum* access to food and water for another study focused on neurogenesis. After five days of captivity, house sparrows were euthanized via isoflurane overdose and rapid decapitation. Whole brains were removed and stored in PBS with sodium azide at 4°C. Before DNA extraction, hippocampi were excised from whole brains and 0.1

g was used for DNA extraction. We used hippocampal samples here as they were collected from house sparrows for the aforementioned study. DNA was extracted in August 2017 using phenol/chloroform/isoamyl and stored at -20°C until sequencing (Green and Sambrook 2001).

Sequencing

Sequencing was performed on an Ion Torrent Personal Genome Machine (PGM) at Georgia Southern University Armstrong Campus' facility (Life Technologies). For library creation, we modified a standard GBS protocol for ddRAD and epiRAD sequencing on the Ion Torrent platform (Life Technologies) (ddRAD-seq; n= 64, epiRAD-seq; n = 53- sample sizes by city listed in Table 1) (Mascher et al. 2013; Schield et al. 2016). For ddRAD, we used enzymes *MspI* and *PstI* (all enzymes New England Biolabs, Ipswich, MA). For epiRAD, we used enzymes *HpaII* and *PstI*. *HpaII* and *MspI* cut DNA at the same sequence (CCGG), but *HpaII* is sensitive to methylation at the restriction site, allowing us to calculate population genetic statistics and compare the presence or absence of methylation among individuals (Schield et al. 2016). After restriction digestion, we ligated on barcodes and y-adaptors of the Ion Torrent IonXpress sequences. We conducted emulsion PCR following manufacturers protocols of the Ion PGM-Hi-Q-View OT2-200 kit on the Ion Express OneTouch2 platform. We then sequenced resultant fragments following manufacturers protocols of the Ion PGM-Hi-Q-View Sequencing 200 Kit using an Ion 316v2 BC Chip. This process generated two datasets; the genetic ddRAD data that were used to determine epigenetic potential (*MspI* with *PstI*), and the epigenetic epiRAD data that were used to measure DNA methylation among *HpaII* restriction sites (*HpaII* and *PstI*).

Genetic Data Quality Control and Analysis

The ddRADseq reads were demultiplexed within Torrent Suite™ version 4.4.3 (Life Technologies) and were returned in BAM format. The read lengths for each individual were extracted using *SAMtools* and imported into R (version 3.5.1) (Li et al. 2009; R: A language and environment for statistical computing 2018). For each of the five cities, a density plot showing the distribution of read lengths against number of reads retained was created (*Supplementary* Fig. S1). The average read length peaked at 75, and all reads fewer than 75 base pairs were removed (*Supplementary* Fig. S1). Reads were mapped back to the house sparrow genome using BWA using default parameters (Li and Durbin 2009; Elgvin et al. 2017). This genome belonged to an individual from an inbred, insular population of house sparrows in their native range (Elgvin et al. 2017). The output file from BWA was converted to bam format and sorted. The resulting bam files were used in the *STACKS* (Version 2.5.3) pipeline, starting with the function "gstacks" to identify variants, followed by "populations" to calculate population genetic statistics (Catchen et al. 2013). Here, we filtered for loci that occurred in a minimum of 60% of individuals using the parameter -r 0.6. The Variant Call Format (VCF) file was returned from "populations".

All CpG and GpC dinucleotides were identified within the mapped reads and the house sparrow genome. To identify gains or losses of a CpG sites, reads were compared with the house sparrow genome. SNPs occurring within a CpG (e.g. CpG -> CpA) or within a GpC (e.g. GpC -> ApC) were considered a loss of that dinucleotide (e.g. loss of an existing CpG site compared to the house sparrow genome). SNPs leading to the formation of a distinct CpG or GpC motif were considered gains (e.g. novel CpG sites compared to the house sparrow genome). In R, CpG sites were relativized to GpC dinucleotides (Fryxell and Moon 2005; Saxonov et al. 2006) using the equation:

$$\left(\sum_{X=A,T,G}\frac{N_{XG\to CG}}{N_{GX\to GC}+1}+\sum_{X=A,T,C}\frac{N_{CX\to CG}}{N_{XC\to GC}+1}\right)-\left(\sum_{X=A,T,G}\frac{N_{CG\to XG}}{N_{GC\to GX}+1}+\sum_{X=A,T,C}\frac{N_{CG\to CX}}{N_{GC\to XC}+1}\right)$$

Losses were relativized using GpC losses:

$$\left(\sum_{X=A,T,G}\frac{N_{XG\to CG}}{N_{GX\to GC}+1}+\sum_{X=A,T,C}\frac{N_{CX\to CG}}{N_{XC\to GC}+1}\right)$$

Gains were relativized using GpC gains:

$$\left(\sum_{X=A,T,G}\frac{N_{CG\to XG}}{N_{GC\to GX}+1}+\sum_{X=A,T,C}\frac{N_{CG\to CX}}{N_{GC\to XC}+1}\right)$$

In all equations, X represents the changing base pair, N represents the number of mutations, and → represents mutation direction. Linear mixed models were used to ask about the relationship between distance from Mombasa (as a continuous variable) and relativized CpG sites, gains of novel CpG sites, and losses of existing CpG sites using capture city as a random factor (Fig. 1). Linear mixed models were run in R using the package *lme4*. To determine whether CpG patterns were artifacts of history as birds moved among comparatively small populations, we used Discriminant Analysis of Principal Components (DAPC) to elucidate population structure using variants called by *STACKS* (Jombart 2008; Jombart et al. 2010; Catchen et al. 2013). Within the R package *adgenet*, the function "find.clusters" was used to predict the number of genetic clusters. The function "*dapc*" was then run to describe the relationship between clusters. Observed heterozygosity, expected heterozygosity, private allelic sites, and the inbreeding coefficient calculated by *STACKS* were compared to distance to Mombasa using Pearson correlation coefficients.

Tajima's *D* and Fay and Wu's *H* were calculated for both CpG sites and non-CpG sites for each city to determine whether selection or other mechanisms were driving spatial patterns in epigenetic potential. CpG sites here includes a CpG dinucleotide in which the C or the G position

contained a SNP across any individual. Note that other literature refers to non-CpGs as motifs other than CpG sites at which methylation may occur, but here we define non-CpG as any SNP not present in the C or G location of a CpG site or at the loci at which a CpG site was gained or lost. Negative values of Tajima's *D* suggest selection or population expansion, but cannot distinguish between these alternatives (Tajima 1989). However, by incorporating information from an outgroup, Fay and Wu's *H* can distinguish between selection or expansion and thus was also estimated for each city (Fay and Wu 2000). Tajima's *D* and Fay and Wu's *H* were calculated using both the R package *PopGenome* (Version 2.7.5) and *DnaSP* software (Version 6.12.04) using default parameters (Pfeifer et al. 2014; Rozas et al. 2017). Both methods returned the same results confirming the calculation accuracy. To perform calculations, we used the "populations" module of *STACKS* to output consensus sequences for each RAD locus as a FASTA file (converted from a VCF file). In order to calculate Fay and Wu's *H*, the Eurasian tree sparrow (*Passer montanus*) was used as an outgroup (Ravinet et al. 2018).

To ask whether Tajima's *D* and Fay and Wu's *H* changed across the range expansion, we used permutation tests. We randomized the sequence source (sampling city) in the extracted FASTA file. The shuffled file was then used to calculate Tajima's *D* and Fay and Wu's *H* for both CpG sites and non-CpG sites. This process was repeated 1000 times (Fig. 2*A*). Next, the Pearson correlation coefficients between the distance from Mombasa and i) Tajima's *D* and ii) Fay and Wu's *H* were calculated. This process was repeated 100 times (Fig. 2*B*). Next, the p-value, used to detect significant changes in either metric with distance from Mombasa, was calculated. If the Pearson correlation coefficient calculated on the actual (non-shuffled data) was negative (red line, Fig. 2*B*), the p-value was calculated as the fraction of the Pearson correlation coefficients from the shuffling procedure that fell below the actual value. Conversely, if the Pearson correlation

coefficient calculated on the actual (non-shuffled data) was positive, the p-value was calculated as the fraction of the Pearson correlation coefficients from the shuffling procedure that fell above the actual value. To avoid bias due to extremely low p-values, the entire procedure (from shuffling to p-value calculation) was repeated an additional 100x to calculate the p-value distribution (Fig. 2*C*).

DNA Methylation Quality Control and Analysis

The epiRADseq reads were demultiplexed within Torrent Suite™ and were returned in BAM format and were converted to SAM format using SAMtools (Li et al. 2009). The resulting SAM file was converted to FASTQ in the Linux environment. The files were filtered using *TRIMMOMATIC* with default options (Bolger et al. 2014). The files were mapped to the house sparrow genome using *BWA* (Li and Durbin 2009). The output file from *BWA* was converted to BAM format and sorted, then converted to a BED file using the function "bamToBed" in *BEDTools* (Quinlan and Hall 2010). Loci mapped by at least one read in all samples in 5 cities were extracted using "mergeBed" in *BEDTools* (Quinlan and Hall 2010). For these extracted loci, 'coverageBed' from *BedTools* was used to calculate the mapped read counts for every individual. Any individual with fewer than 1000 reads was removed from the analysis (*Supplementary* Fig. S2). Additionally, coverage filtering was performed by comparing the distribution of reads per million mapped reads (RPM) to quantile value to identify extremely high and low coverage loci, which were removed. Specifically, the top 3% of loci or any with less than 10x coverage were removed. A Wilcoxon test with p-value < 0.05 was used to detect loci that exhibited differential RPM between any two of the sampled cities. This step detected 4,518 differentially methylated loci. To investigate the difference among individuals for each locus, and to exclude the RPM fluctuations between different loci, Z-scores were used to normalize the RPM value for each locus

among all individuals. Given the number of individuals for each city (generally higher than 10), this resulted in more than 45,180 RPM values (4,518 different loci × number of individuals for each city) for each city. Therefore, we used the median level of normalized RPM to represent methylation level. The minimum level of normalized RPM was removed before calculating the median value for each city. The methylation level was represented as the negative value of calculated median RPM because the absence of a counted fragment indicated the presence of methylation. It is important to note that hippocampal samples were collected after house sparrows spent five days in captivity. While captive housing may have impacted methylation patterns, all birds were exposed to the same duration and conditions of captivity. We hoped to ask about DNA methylation data within genes relevant to range expansions, such as those related to memory and exploration of novel environments, but unfortunately the coarse nature of this sequencing approach made it so very few specific genes or gene regions could be identified (Kilvitis et al. 2017). Linear mixed models were used to ask about the relationship between the distance from Mombasa (as a continuous variable) and DNA methylation and the standard deviation of DNA methylation using capture city as a random factor. Linear mixed models were run in R using the package *lme4*.

Results

CpG Sites Across the Range Expansion

Following quality control, ddRADseq returned 452,465 reads and 1,205 single nucleotide polymorphisms (SNPs) across 257 unique loci. The number of CpG sites was highest towards the range edge and decreased towards the range core (Fig. 1; β= 0.0016 (± 0.0006), t= 2.752, p= 0.0059). We identified 118 SNPs causing a CpG gain across 28 unique loci, and 191 SNPs causing a CpG site loss across 36 unique loci. We did not detect a change in CpG site losses or gains across

the range expansion (CpG site losses: β= -0.0018 (± 0.0013), t= -1.364, p= 0.1725; CpG site gains: β= -0.0002 (± 0.0008), t= -0.2, p= 0.8415).

Patterns of Tajima's D *and Fay and Wu's* H

Tajima's *D* ranged from -1.446 to -0.023, and Fay and Wu's *H* ranged from -0.021 to -3.215 for CpG sites (n= 64) from birds collected across the range expansion (Table 1; Fig. 2). For non-CpG sites (n= 193), Tajima's *D* ranged from -0.938 to -0.366 and Fay and Wu's *H* ranged from -0.167 to 0.713 (Table 1, Fig 2). Tajima's *D* values for CpG sites significantly decreased towards the range core while no trend was detected for non-CpG sites (CpG sites: r= -0.97, p= 0.02; non-CpG sites: r= -0.43, p= 0.24; Fig. 2). For Fay and Wu's *H*, the correlation coefficient was also negative for CpG sites but not for non-CpG sites, but neither trend was statistically significant (CpG sites: r= -0.66, p= 0.16; non-CpG sites: r= 0.70, p= 0.16; Fig. 2). Due to a low read count and number of SNPs, individuals from Voi (160 km) were excluded from this analysis (See Materials and Methods; *Supplementary* Fig. S3).

DNA Methylation Across the Range Expansion

Of 14,659 loci, 4,518 loci were differentially methylation between pairs of sampled cities. We detected no significant change in DNA methylation or variation in DNA methylation across the range expansion (DNA methylation: β= -0.0003 (± 0.0002), t= -1.593, p= 0.1112; standard deviation of DNA methylation: β= 4.9e-05 (± 0.0002), t= 0.2429, p= 0.8081).

Population Statistics and Structure

Among the five Kenyan cities, observed heterozygosity ranged from 0.160 to 0.258, expected heterozygosity ranged from 0.125 to 0.187, number of private alleles ranged from 0 to 31, and the inbreeding coefficient ranged from -0.115 to -0.029 (Table 1). None of these variables were related to distance from Mombasa (*Supplementary* Fig. S4). The DAPC predicted two genetic clusters, yet both clusters contained at least one individual from every city. Collectively, there was little evidence that genetic diversity or founder effects explain geographic patterns in CpG number (*Supplementary* Figs S4 and S5).

Discussion

Epigenetic potential (i.e. relativized CpG sites) was highest towards the range edge and decreased towards the range core of the house sparrow invasion (Fig. 1). Losses of existing and gains of novel CpG sites were unrelated to position along the range expansion. Comparisons of Tajima's *D* and Fay and Wu's *H* indices among cities lends support to our hypothesis that positive selection is acting on CpG sites towards the range edge. However, we cannot definitively rule out other processes, including non-random gene flow or the activation of TEs (Fig. 2; Table 1). Both our inability to include Voi (the city located 160 km from Mombasa) due to an insufficient number of SNPs and the relatively small number of cities we had available to study overall require cautious interpretation of the data. Nevertheless, as there was no evidence that genetic artifacts associated with small population size (Table 1), nor that genetic differentiation among cities alone could explain these patterns (*Supplementary* Figs S4 and S5), we argue that the distribution of epigenetic potential along the invasion is likely a consequence of selection for phenotypic plasticity capacitated by epigenetic potential. Lastly, we found no pattern of DNA methylation or variation

in DNA methylation across the range expansion. Below we discuss the ramifications for these results for house sparrow invasions.

Epigenetic Potential Underlying Phenotypic Plasticity

Epigenetic potential was highest towards the expanding edge of the Kenyan house sparrow invasion and decreased towards the range core (Fig. 1). We propose that this particular pattern arose because epigenetic potential of this form may facilitate phenotypic plasticity via DNA methylation, a trait of great value in new contexts. As CpG sites in genomes represent the places at which methylation can impact gene expression, epigenetic potential could be an important mediator of phenotypic variation, and also can be subject to natural selection and other evolutionary processes (Feinberg and Irizarry 2010; Kilvitis et al. 2017). As an analogy, consider two stereo systems: the first has only one knob for volume whereas a second knobs for volume, bass, treble, and balance. Although both stereos produce sound, the second system allows finer tuning, matching better the sound quality to the environment in which it is being played. CpG sites are similar to stereo knobs, except that they are adjusted via DNA methylation; as the environment changes, knobs are turned, increasing or decreasing gene expression and hence adjusting phenotypic plasticity. The elegance of epigenetic potential as a gene regulatory trait is that no knob gets turned until a relevant environmental stimulus occurs. In other words, epigenetic potential may enable phenotypic variation to remain latent until environmental conditions release it.

In theory, the more CpG sites a genome has (i.e., the more epigenetic potential), the more gene expression may be tuned via DNA methylation to match the environment. This expectation seems reasonable for epigenetic potential in the putative promoter of *TLR4*, where individuals with high epigenetic potential had greater inducibility and reversibility (in females) of gene expression

(Hanson et al. 2021). There, epigenetic potential also seemed to imbue birds with tissue- and sex-specific gene expression, which may allow for additional flexibility in response to exposure to bacteria. It would be unreasonable to assume that there will always be a simple linear relationship between CpG sites and gene regulation, and indeed an enormous number of other factors play a role in gene regulation (Lelli et al. 2012). Our argument here is simply that, overall, epigenetic potential may represent one measurable form of capacitated phenotypic plasticity, despite its complexities. In support ot this idea, in two species of cnidarians, CpG site density was higher in the promoter regions of genes important for environmental adaptation compared to other functional classes of genes, which may facilitate for the regulation of these genes via DNA methylation (Marsh et al. 2016). Additionally, across taxa, the abundance of CpG sites near transcription start sites predicts levels of gene expression (Cheng et al. 2012; Yang et al. 2014). Further, in humans, the loss of CpG sites correlates with the loss of DNA methylation at that CpG site and sometimes in surrounding CpG sites (Zhi et al. 2013; Zhou et al. 2015). It is important to acknowledge we were not able to examine epigenetic potential across genes or gene regions in this study due to the coarse sequencing technique (see methods). Many studies have noted the different functional consequences of CpG sites between gene regions, which may lead to different evolutionary trajectories for CpG sites within these gene regions (Subramanian and Kumar 2003; Cohen et al. 2011; Jones 2012). Gene regions should be explored in relation to differences in epigenetic potential in future studies. Despite these complexities, we found a clear pattern of epigenetic potential across the range expansion with CpG sites being most abundant in birds towards the range edge which may imbue more phenotypic plasticity to birds there compared to birds at the range core where birds tend to have fewer CpG sites (Fig. 1; (Lande 2015)).

Possible Processes Contributing to the Distribution of Epigenetic Potential Across Kenya

Generally, we expect that epigenetic potential predominates in invasions, at range edges, or in other dynamic environments as it could allow for more and faster (within-generation) phenotypic change than through selection on fixed genetic variation. In invasions, there may be a minimum level of epigenetic potential required for any successful colonization or related event; initial colonizers may need to harbor relatively high levels of epigenetic potential or die before breeding. As populations age and adapt to the surrounding conditions, we expect that epigenetic potential could decline as the selective advantage of phenotypic plasticity decreases, allowing for fixed genetic variants to predominate. Many natural and anthropogenic introductions fail, and whereas the cause of some failures are obvious, some are not (Zenni and Nuñez 2013). Epigenetic potential might represent a general mechanism whereby some small populations can surmount genetic bottlenecks, the accumulation of rare lethal recessive alleles, or other phenomena associated with the founding of populations (Lee 2002; Taylor and Hastings 2005). Some modest level of epigenetic potential might provide just enough latent phenotypic plasticity in contexts where increases in genetic diversity via recombination is impossible. In this light, epigenetic potential might be less of an individual trait promoting adaptation via plasticity and more of a necessity for the viability of small or isolated populations.

Processes such as the activation of TEs or the mutation of methylated cytosines may contribute to the underlying genetic variation in invasive populations for which evolutionary processes may act upon and lead to differences in epigenetic potential (Bird 1980; Marin et al. 2019). Whereas we do not have data from the initial colonizers to understand the original distribution of standing variation in epigenetic potential, we hypothesize that as the Kenyan range expansion occurred, house sparrows with high epigenetic potential were more likely to survive

and reproduce in relatively novel areas, leading to positive selection on CpG sites, especially towards the range edge. As house sparrow populations ultimately adapt to local conditions, genetically-canalized responses may become more advantageous than plastic ones. Consistent with these hypotheses, CpG sites were more numerous towards the range edge (Fig. 1). We found no difference in the losses of existing or gains of novel CpG sites across the range expansion. Differing selection pressures for CpG sites across the expansion could be generating these patterns, as Tajima's *D* values for CpG sites decreased significantly towards the range edge, a trend not shared by non-CpG sites (Fig. 2). Although we did not detect a linear decrease in values of Fay and Wu's *H* for CpG sites or non-CpG sites across the range expansion, we did observe negative values of both Tajima's *D* and Fay and Wu's *H* at the range edge, indicating positive selection could be acting on CpG sites, at least in the city of Kakamega (Fig. 2, Table 1). Importantly, extensive admixture has been detected previously among these populations, so even though birds are intermixing among cities, the patterns of CpG sites across the expansion persist (Schrey et al. 2014). Further, population genetic statistics do not correlate with distance from Mombasa, indicating that the distribution of epigenetic potential among cities is not obviously an artifact of prior bottlenecks or founder effects (*Supplementary* Fig. S4). Given the conservative nature of our sequencing approach, the detection of any trends across Kenya (especially patterns of Tajima's *D*) lends further support to our interpretation; the coarse technique used here returned a largely random subset of the genome based on restriction sites, rather than a battery of sequences from genes related to traits that might facilitate range expansions. Altogether, there is support for selection acting on CpG sites already present in Kenyan house sparrows, leading to their higher frequency at the range edge (Figs. 1 and 2), but we remain cautious in concluding that selection

alone drove these trends. We were only able to use data from individuals from four cities for this analysis and our sample size within cities was fairly small (Table 1 and *Supplementary* Fig. S3).

Besides selection, population-level processes could be contributing to these patterns. One is the range expansion process itself, which might not be driven by random movements of individuals, or even disproportionate movements of individuals with high epigenetic potential, as our selection framework intimates (Bowler and Benton 2005; Shine et al. 2011). For example, some individuals might disperse to new areas more readily than others for reasons that have no relation to adaptive phenotypic plasticity or even epigenetic potential directly (Bowler and Benton 2005; Shine et al. 2011; Edelaar and Bolnick 2012). Subsequently, non-random dispersal may lead to non-random gene flow, producing the spatial patterns we observed (Edelaar and Bolnick 2012). Perhaps individuals with high epigenetic potential genotypes were more likely to disperse to novel habitats (advancing the range expansion) than those with lower epigenetic potential, but until we test directly the effects of epigenetic potential on fitness in popualtions of different age, we cannot distinguish non-random gene flow from the consequences of selection. In other words, there might not be a link between epigenetic potential and adaptive phenotypic plasticity; movements of birds among sites followed by subsequent selection for reasons that have nothing to do with adaptive plasticity could have generated the patterns we found. We disfavor this scenario, however, because if non-random dispersal was important, we would expect to see more population structure than we do here (Edelaar and Bolnick 2012). Previous studies revealed high levels of admixture across the Kenyan expansion, which may contribute to the lack of population structure detected in this study (*Supplementary* Fig. S5; Schrey et al. 2014). Non-random dispersal seems less likely than selection as an explanation of the observed patterns, but going forward, the best test will be to target genes

important to range expansions and investigate their patterns directly while also studying additional cities along many range expansions.

One other potential explanation for the observed patterns of epigenetic potential is the activation of TEs. TEs have been proposed as one mechanism by which the genetic paradox of invasions may be explained, as increased TE activity may increase genetic variation via their replication and insertion across the genome (Stapley et al. 2015; Marin et al. 2019). TE activity may be higher under periods of stress, such as that experienced during biological invasions (Dennenmoser et al. 2017; Goubert et al. 2017). If range edge individuals experience more stress than range core birds, TE activity may be be higher, leading to a higher rate of TE replication and insertion across the genome (Stapley et al. 2015; Marin et al. 2019). The insertion of TEs should increase genetic variation and subsequently epigenetic potential (Marin et al. 2019). We find this scenario unlikely, as we found no correlation between population genetic statistics across the range expansion, which would be expected with the addition of genetic variation via TEs (Table 1). Additionally, insertion of new genetic material should also lead to the gain of *novel* CpG sites. We found no evidence of Kenyan house sparrows gaining CpG sites differently across the range expansion. As we do not have data from the initial colonizers, we cannot completely exclude TEs playing a role in the early house sparrow invasion or in native populations generating genetic variation upon which evolutionary forces may act.

DNA Methylation Across Kenya

Across the Kenyan range expansion, we found no detectable pattern of DNA methylation or variation in DNA methylation. As DNA methylation is highly context dependent, with patterns varying among cell types, genes, and gene regions, the absence of a pattern is not surprising (Smith

and Meissner 2013; Lanata et al. 2018). These results are consistent with a previous study of Kenyan house sparrows, where no detectable methylation signature was found across the range expansion either (Liebl et al. 2013). In Australian house sparrows, which are also non-native, methylation differences were found across the sampled cities and these differences were attributed to local environmental conditions (Sheldon et al. 2018). We must caution interpretation and reiterate the descriptive nature of our epigenetic results even though we did not detect a pattern, as we could not account for many factors that could have affected DNA methylation, including the time that the house sparrows spent in captivity prior to sampling. Until we can account better for context (i.e., gene, type of stimulus), it will be difficult to interlink epigenetic potential, methylation, gene expression, phenotypic plasticity, and fitness, but this goal is a very important one (Smith and Meissner 2013).

Conclusions

Epigenetic potential represents the capacity for DNA methylation to occur within the genome of an individual and thus could in part affect the range of phenotypic plasticity achievable by an organism. While other mechanisms also influence phenotypic plasticity, epigenetic potential may be especially advantageous to initial colonizers and expanding populations, helping them overcome genetic, demographic, and environmental challenges associated with novel areas. Here, we revealed that epigenetic potential follows patterns we would expect; birds towards the range edge tended to have more CpG sites than those from the range core (Fig. 1). As CpG sites are genomic motifs, they can be inherited and positively selected, or be subject to other processes such as non-random gene flow (Branciamore et al. 2010). Selection seems to be acting on CpG sites at the range edge (Fig. 2, Table 1), but non-random gene flow could also be contributing to the

patterns we observed. We found no evidence that trends were artifacts of population structure, founder effects, or genetic bottlenecks. As we had individuals from relatively few cities to consider, we caution over-interpretation, but because we investigated distance from the city of initial introduction into Kenya in relation to CpG content, we think our result implicate epigenetic potential as consequential in this range expansion and perhaps others. Additional studies should be designed to better parse the contribution of selection, non-random gene flow, or other processes that may give rise to differences in epigenetic potential, especially as evolutionary drivers are likely to differ across ecological contexts. Importantly, here, we only considered one form of epigenetic potential (Kilvitis et al. 2017). Other forms may also be important and should be investigated in the future. Further, differences in this form of epigenetic potential across gene regions (e.g. gene bodies and promoters) should be examined. Factors direcetly related to range expansion, including predictability and variability in climate, altitude, parasite pressure, and more could also be impacting selection for CpG sites and patterns of DNA methylation. Consequently, future studies investigating additional range expansions as well as response to other environmental contexts will be vital to uncover whether epigenetic potential represents a type of adaptive plasticity. If so, epigenetic potential may be applicable across many fields besides invasion biology.

Acknowledgements

We thank the Martin lab for feedback on the study design, analysis and manuscript drafts, Kelly Hanson, Mark Jacim, and Jaime Zolik for help in the lab, and Alex Mutati for help with field collections. This work was supported by the National Science Foundation (grant numbers 0920475, 1257773, 1656551, 1504662, 2027040), the USF College of Public Health, the USF

Genomics Hub, a USF Graduate Student Research Challenge Grant (to A.L.L), Sigma XI (G2016100191872782 to H.E.H.), the Porter Family Foundation grant (to H.E.H.), the American Ornithological Society Hesse Grant (to H.E.H), the American Museum of Natural History Frank M. Chapman Memorial Fund (to H.E.H), and National Science Foundation Research Coordination Network "g2p2pop" (grant number 1656063) Laboratory Exchange Grant (to H.E.H).

Statement of Authorship

H.E.H., A.W.S., and L.B.M. designed the experiment. A.L.L, H.E.H., and A.W.S. collected the data. C.W. analyzed the data. M.R. helped develop analytical methods. H.E.H., A.L.L, A.W.S., R.H.Y.J., M.R. and L.B.M., helped acquire funding. H.E.H. wrote the original draft. All authors contributed to revisions.

Data and Code Availability

Data can be accessed on Dryad at https://doi.org/10.5061/dryad.v41ns1rt2. Sequences can be accessed at NCBI accession numbers SRR18313456 - SRR18313519.

Literature Cited

Bird, A. P. 1980. DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Research 8:1499–1504.

Blem, C. R. 1973. Geographic Variation in the Bioenergetics of the House Sparrow. Ornithological Monographs 14:96–121.

Bock, C., I. Beerman, W. H. Lien, Z. D. Smith, H. Gu, P. Boyle, A. Gnirke, et al. 2012. DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells. Molecular Cell 47:633–647.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bowler, D. E., and T. G. Benton. 2005. Causes and consequences of animal dispersal strategies. Biological Reviews 205–225.

Branciamore, S., Z.-X. Chen, A. D. Riggs, and S. N. Rodin. 2010. CpG island clusters and pro-epigenetic selection for CpGs in protein-coding exons of HOX and other transcription factors. Proceedings of the National Academy of Sciences of the USA 107:15485–15490.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: An analysis tool set for population genomics. Molecular Ecology 22:3124–2140.

Cheng, C., R. Alexander, R. Min, J. Leng, K. Y. Yip, J. Rozowsky, K. K. Yan, et al. 2012. Understanding transcriptional regulation by integrative analysis of transcription factor binding data. Genome Research 22:1658–1667.

Cohen, N. M., E. Kenigsberg, and A. Tanay. 2011. Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell 145:773–786.

Dennenmoser, S., F. J. Sedlazeck, E. Iwaszkiewicz, X. Y. Li, J. Altmüller, and A. W. Nolte. 2017. Copy number increases of transposable elements and protein-coding genes in an invasive fish of hybrid origin. Molecular Ecology 26:4712–4724.

Edelaar, P., and D. I. Bolnick. 2012. Non-random gene flow: An underappreciated force in evolution and ecology. Trends in Ecology and Evolution 27:659–665.

Elgvin, T. O., C. N. Trier, O. K. Tørresen, I. J. Hagen, S. Lien, A. J. Nederbragt, M. Ravinet, et al. 2017. The genomic mosaicism of hybrid speciation. Science Advances 3.

Fay, J. C., and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413.

Feinberg, A. P. 2007. Phenotypic plasticity and the epigenetics of human disease. Nature 447:433–440.

Feinberg, A. P., and R. A. Irizarry. 2010. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. Proceedings of the National Academy of Sciences of the USA 107:1757–1764.

Fryxell, K. J., and W. J. Moon. 2005. CpG mutation rates in the human genome are highly dependent on local GC content. Molecular Biology and Evolution 22:650–658.

Goubert, C., H. Henri, G. Minard, C. Valiente Moro, P. Mavingui, C. Vieira, and M. Boulesteix. 2017. High-throughput sequencing of transposable element insertions suggests adaptive evolution of the invasive Asian tiger mosquito towards temperate environments. Molecular Ecology 26:3968–3981.

Green, M. R., and J. Sambrook. 2001. Molecular Cloning. A Laboratory Manual (Fourth.). Cold Spring Harbor Laboratory Press.

Hanson, Haley E et al. (2022), Data from: Epigenetic Potential and DNA Methylation in an Ongoing House Sparrow (Passer domesticus) Range Expansion, Dryad, Dataset, https://doi.org/10.5061/dryad.v41ns1rt2

Hanson, H. E., B. Koussayer, H. J. Kilvitis, A. W. Schrey, J. D. Maddox, and L. B. Martin. 2020*a*. Epigenetic Potential in Native and Introduced Populations of House Sparrows (*Passer domesticus*). Integrative and Comparative Biology.

Hanson, H. E., N. S. Mathews, M. E. Hauber, and L. B. Martin. 2020*b*. The Natural History of Model Organisms: The house sparrow in the service of basic and applied biology. eLife 9:e52803.

Hanson, H. E., C. Zimmer, B. Koussayer, A. W. Schrey, J. D. Maddox, and L. B. Martin. 2021. Epigenetic Potential Affects Immune Gene Expression in House Sparrows. Journal of Experimental Biology 224:jeb238451.

Johnston, R. F., and R. K. Selander. 1971. Evolution in the House Sparrow. II. Adaptive Differentiation in North American Populations. Evolution 25.

Jombart, T. 2008. Adegenet: a R package for multivariate analysis of genetic markers. Bioinformatics (Oxford, England) 24:1403–1405.

Jombart, T., S. Devillard, F. Balloux, D. Falush, M. Stephens, J. Pritchard, J. Pritchard, et al. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genetics 11:94.

Jones, P. A. 2012. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. Nature Reviews Genetics 13:484–492.

Kendeigh, S. C. 1976. Latitudinal Trends in the Metabolic Adjustments of the House Sparrow. Ecology 57:509–519.

Kilvitis, H. J., H. Hanson, A. W. Schrey, and L. B. Martin. 2017. Epigenetic potential as a mechanism of phenotypic plasticity in vertebrate range expansions. Integrative and Comparative Biology 57:385–395.

Lanata, C. M., S. A. Chung, and L. A. Criswell. 2018. DNA methylation 101: What is important to know about DNA methylation and its role in SLE risk and disease heterogeneity. Lupus Science and Medicine 5.

Lande, R. 2015. Evolution of phenotypic plasticity in colonizing species. Molecular Ecology 24:2038–2045.

Lee, C. E. 2002. Evolutionary genetics of invasive species. Trends in Ecology and Evolution 17:386–391.

Lelli, K. M., M. Slattery, and R. S. Mann. 2012. Disentangling the many layers of eukaryotic transcriptional regulation. Annual Review of Genetics 46:43–68.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2090.

Liebl, A. L., and L. B. Martin. 2012. Exploratory behaviour and stressor hyper responsiveness facilitate range expansion of an introduced songbird. Proceedings of the Royal Society B 279:4375–4381.

———. 2013. Stress hormone receptors change as range expansion progresses in house sparrows. Biology letters 9:20130181.

Liebl, A. L., A. W. Schrey, C. L. Richards, and L. B. Martin. 2013. Patterns of DNA methylation throughout a range expansion of an introduced songbird. Integrative and Comparative Biology 53:351–358.

Marin, P., J. Genitoni, D. Barloy, S. Maury, P. Gibert, C. K. Ghalambor, and C. Vieira. 2019. Biological invasion: The influence of the hidden side of the (epi)genome. Functional Ecology.

Marsh, A. G., K. D. Hoadley, and M. E. Warner. 2016. Distribution of CpG motifs in upstream gene domains in a reef coral and sea anemone: Implications for epigenetics in cnidarians. PLoS ONE 11:1–14.

Martin, L. B., J. L. Alam, T. Imboma, and A. L. Liebl. 2010. Variation in inflammation as a correlate of range expansion in Kenyan house sparrows. Oecologia 164:339–347.

Martin, L. B., C. A. . C. Coon, A. L. Liebl, and A. W. Schrey. 2014. Surveillance for microbes and range expansion in house sparrows. Proceedings of the Royal Society B 281:20132690.

Martin, L. B., H. J. Kilvitis, A. J. Brace, L. Cooper, M. F. Haussmann, A. Mutati, V. Fasanello, et al. 2017. Costs of immunity and their role in the range expansion of the house sparrow in Kenya. The Journal of Experimental Biology 220:2228–2235.

Martin, L. B., and A. L. Liebl. 2014. Physiological flexibility in an avian range expansion. General and Comparative Endocrinology 206:227–234.

Martin, L. B., A. L. Liebl, and H. J. Kilvitis. 2015. Covariation in stress and immune gene expression in a range expanding bird. General and Comparative Endocrinology 211:14–19.

Mascher, M., S. Wu, P. St. Amand, N. Stein, and J. Poland. 2013. Application of Genotyping-by-Sequencing on Semiconductor Sequencing Platforms: A Comparison of Genetic and Reference-Based Marker Ordering in Barley. PLoS ONE 8:e76925.

Parkin, D. T., and S. R. Cole. 1985. Genetic differentiation and rates of evolution in some introduced populations of the house sparrow, *passer domesticus* in australia and new zealand. Heredity 54:15–23.

Pfeifer, B., U. Wittelsbürger, S. E. Ramos-Onsins, and M. J. Lercher. 2014. PopGenome: An efficient swiss army knife for population genomic analyses in R. Molecular Biology and Evolution 31:1929–1936.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842.

R: A language and environment for statistical computing. 2018.R: A Language and Environment for Statistical Computing.

Ravinet, M., T. O. Elgvin, C. Trier, M. Aliabadian, A. Gavrilov, and G.-P. Sætre. 2018. Signatures of human-commensalism in the house sparrow genome. Proceedings of the Royal Society B 285:20181246.

Richards, E. J. 2006. Revisiting Soft Inheritance. Nature Review Genetics 7:395–402.

Rozas, J., A. Ferrer-Mata, J. C. Sanchez-DelBarrio, S. Guirao-Rico, P. Librado, S. E. Ramos-Onsins, and A. Sanchez-Gracia. 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Molecular Biology and Evolution 34:3299–3302.

Saxonov, S., P. Berg, and D. L. Brutlag. 2006. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the USA 103:1412–1417.

Schield, D. R., M. R. Walsh, D. C. Card, A. L. Andrew, R. H. Adams, and T. A. Castoe. 2016. EpiRADseq: Scalable analysis of genomewide patterns of methylation using next-generation sequencing. Methods in Ecology and Evolution 7:60–69.

Schrey, A. W., C. A. C. Coon, M. T. Grispo, M. Awad, T. Imboma, E. D. McCoy, H. R. Mushinsky, et al. 2012. Epigenetic Variation May Compensate for Decreased Genetic

Variation with Introductions: A Case Study Using House Sparrows (*Passer domesticus*) on Two Continents. Genetics Research International 2012:1–7.

Schrey, A. W., M. Grispo, M. Awad, M. B. Cook, E. D. McCoy, H. R. Mushinsky, T. Albayrak, et al. 2011. Broad-scale latitudinal patterns of genetic diversity among native European and introduced house sparrow (*Passer domesticus*) populations. Molecular Ecology 20:1133–1143.

Schrey, A. W., A. L. Liebl, C. L. Richards, and L. B. Martin. 2014. Range expansion of house sparrows (*Passer domesticus*) in Kenya: Evidence of genetic admixture and human-mediated dispersal. Journal of Heredity 105:60–69.

Sepers, B., K. van den Heuvel, M. Lindner, H. Viitaniemi, A. Husby, and K. van Oers. 2019. Avian ecological epigenetics: pitfalls and promises. Journal of Ornithology 160:1183–1203.

Sheldon, E. L., A. W. Schrey, S. C. Andrew, A. K. Ragsdale, and S. C. Griffith. 2018. Epigenetic and genetic variation among three separate introductions of the house sparrow (*Passer domesticus*) into Australia. Royal Society Open Science 5.

Shine, R., G. P. Brown, and B. L. Phillips. 2011. An evolutionary process that assembles phenotypes through space rather than through time. Proceedings of the National Academy of Sciences of the USA 108:5708–5711.

Smith, Z. D., and A. Meissner. 2013. DNA methylation: Roles in mammalian development. Nature Reviews Genetics 14:204–220.

Stapley, J., A. W. Santure, and S. R. Dennis. 2015. Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Molecular Ecology 24:2241–2252.

Subramanian, S., and S. Kumar. 2003. Neutral substitutions occur at a faster rate in exons than in noncoding DNA in primate genomes. Genome Research 13:838–844.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Taylor, C. M., and A. Hastings. 2005. Allee effects in biological invasions. Ecology Letters 8:895–908.

Wu, H., and Y. Zhang. 2014. Reversing DNA methylation: Mechanisms, genomics, and biological functions. Cell 156:45–68.

Yang, H., D. Li, and C. Cheng. 2014. Relating gene expression evolution with CpG content changes. BMC Genomics 15:1–10.

Zenni, R. D., and M. A. Nuñez. 2013. The elephant in the room: The role of failed invasions in understanding invasion biology. Oikos 122:801–815.

Zhi, D., S. Aslibekyan, M. R. Irvin, S. A. Claas, I. B. Borecki, J. M. Ordovas, D. M. Absher, et al. 2013. SNPs located at CpG sites modulate genome-epigenome interaction. Epigenetics 8:802–806.

Zhou, D., Z. Li, D. Yu, L. Wan, Y. Zhu, M. Lai, and D. Zhang. 2015. Polymorphisms involving gain or loss of CpG sites are significantly enriched in trait-associated SNPs. Oncotarget 6:39995–40004.

Zhu, H., G. Wang, and J. Qian. 2016. Transcription factors as readers and effectors of DNA methylation. Nature Reviews Genetics 17:551–565.

Tables

Table 1: House sparrow sampling locations across Kenya, sample sizes, and estimates of observed heterozygosity (H_o), expected heterozygosity (H_e), private allelic sites, the inbreeding coefficient (F_{is}), Tajima's *D* (for CpG and non-CpG sites), and Fay and Wu's *H* (for CpG and non-CpG sites).

City	Distance from Mombasa (km)	Sample Size		Observed Heterozygosity (Ho)	Expected Heterozygosity (He)	Private Sites	Inbreeding Coefficient (Fis)	Tajima's *D*		Fay and Wu's *H*	
		ddRADseq	epiRADseq					CpG Sites	Non-CpG Sites	CpG Sites	Non-CpG Sites
Mombasa	0	13	11	0.258	0.187	31	-0.105	-0.023	-0.336	-0.598	-0.167
Voi	160	14	11	0.196	0.136	0	-0.115	NA	NA	NA	NA
Nairobi	500	12	12	0.160	0.125	13	-0.029	-0.548	-0.938	-0.021	-0.002
Nakuru	650	12	10	0.193	0.133	19	-0.108	-1.163	-0.740	-1.186	0.713
Kakamega	850	13	9	0.198	0.148	11	-0.072	-1.446	-0.519	-3.215	0.296

Figure Legends

Figure 1: Epigenetic potential is highest towards the range edge in the Kenyan house sparrow range expansion and declines towards the range core (β= 0.0016 (± 0.0006), t= 2.752, p= 0.0059). CpG data were relativized to GpC sites. Confidence intervals are plotted around the regression line.

Figure 2: Selection for CpG sites, but not non-CpG sites, seems to occur at the range edge in Kenyan house sparrows. (A) Tajima's *D* significantly declines across the range expansion for CpG sites but not for non-CpG sites whereas Fay and Wu's H tends to decrease across the range expansion for CpG sites but not non-CpG sites, but neither trend (for Fay and Wu's H) is statistically significant. Tajima's *D* and Fay and Wu's *H* were calculated for both CpG sites (green) and non-CpG sites (blue). To contrast observed versus expected distributions of these values, both Tajima's *D* and Fay and Wu's *H* were calculated using a permutation test in which sampling city was randomized. The results of these permutations are shown as grey lines. (B) A Pearson correlation analysis was then performed between Tajima's *D or* Fay & Wu's *H* values and the distance to Mombasa (our proxy for population age) for the actual data (red line) and each permutation (individual points). Jitter was used to visualize individual points. (C) The fraction of Pearson correlation values calculated using permuted data that fell above or below (see methods) the true value was used to calculate p-values. Each p-value is calculated based on 100 permutations of the data. This procedure was repeated an additional 100 times to generate the distribution of p-values seen here. Note that Voi (160 km) was excluded from these analyses as it lacked the number of SNPs necessary for its inclusion.

Figure 1

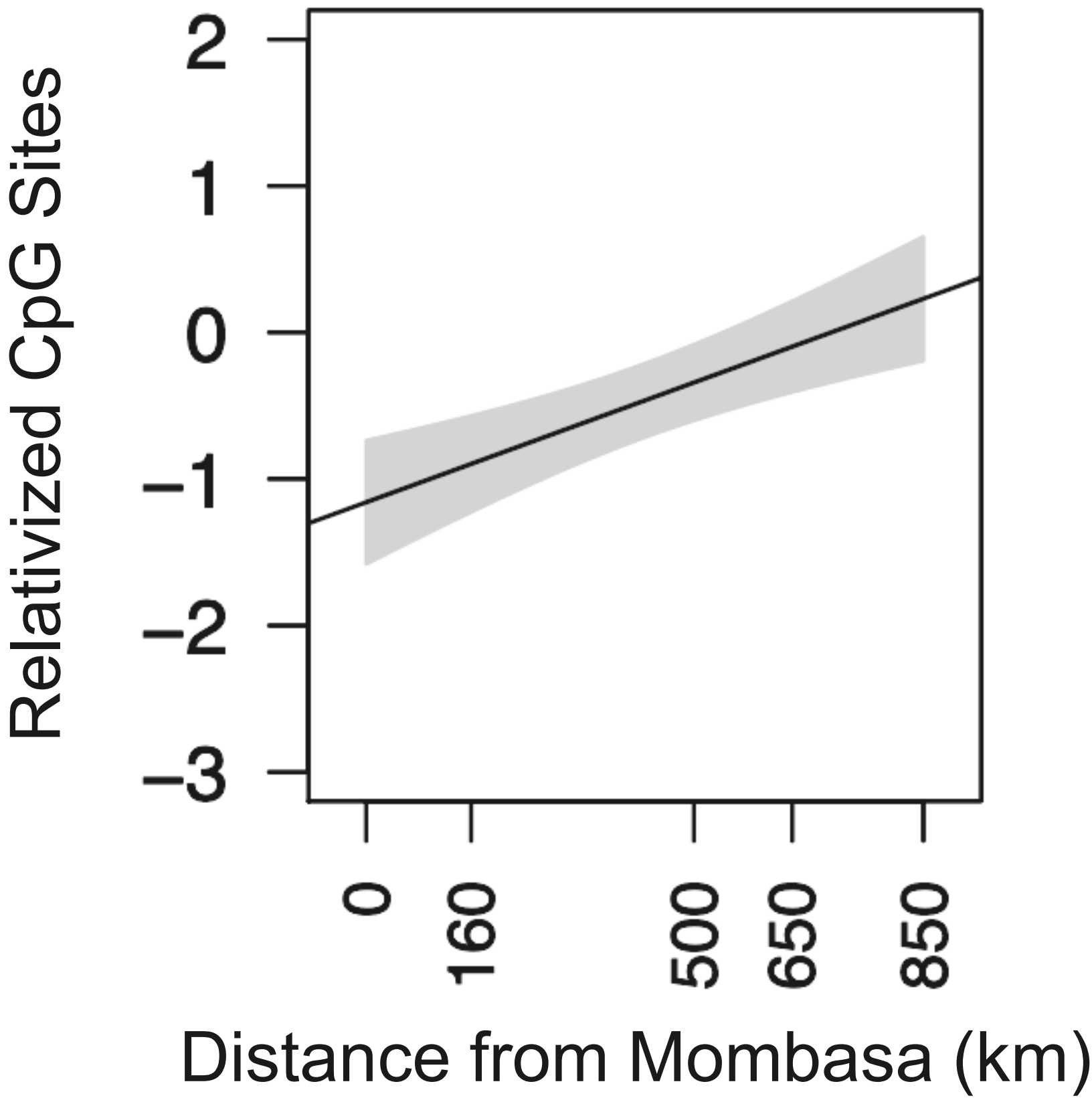
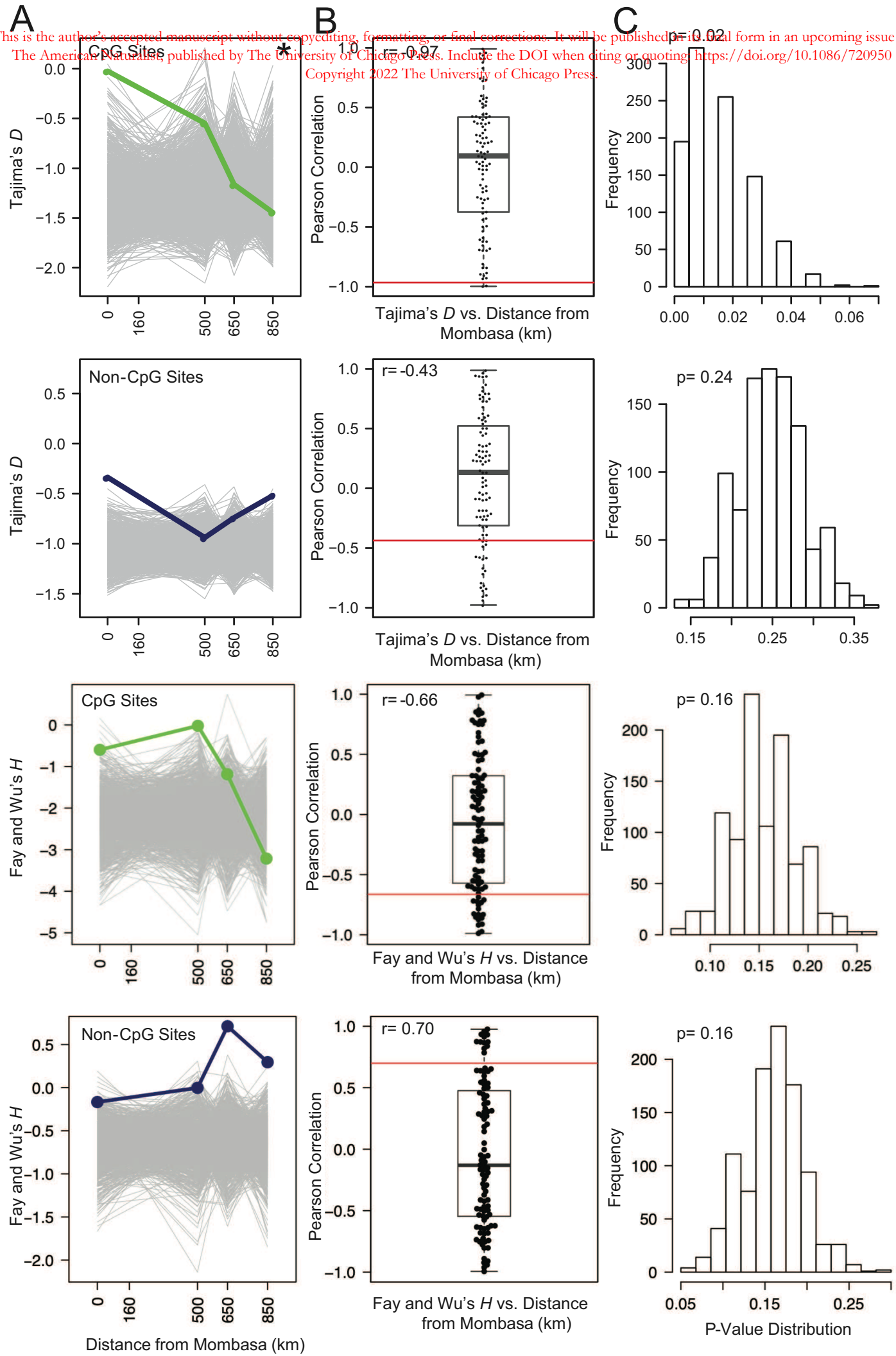


Figure 2



Supplementary Material for

Epigenetic Potential and DNA Methylation in an Ongoing House Sparrow (*Passer domesticus*) Range Expansion

Authors:

Haley E. Hanson,[a,1] Chengqi Wang,[a] Aaron W. Schrey,[b] Andrea L. Liebl,[c] Mark Ravinet,[d] Rays H. Y. Jiang,[a] Lynn B. Martin[a,1]

Author Affiliations:

[a] University of South Florida, Global and Planetary Health, Tampa, FL

[b] Georgia Southern University Armstrong Campus, Department of Biology, Savannah, GA

[c] University of South Dakota, Department of Biology, Vermillion, SD

[d] University of Nottingham, School of Life Sciences, Nottingham, UK

Corresponding Author:

[1] To whom correspondence may be addressed. University of South Florida, Global and Planetary Health, Tampa, FL. Email: haleyehanson@gmail.com or lbmartin@usf.edu

This PDF file includes:

Figures S1 to S5

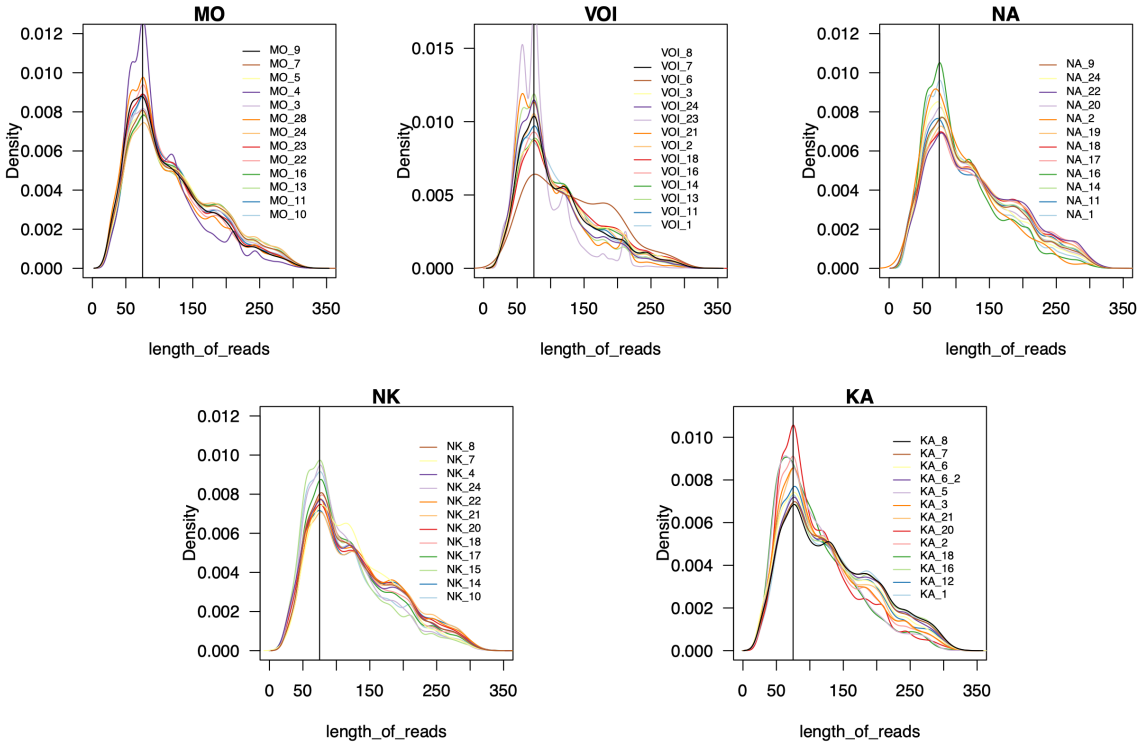


Fig. S1. Density by read length plots. Plot shows the distribution of RAD-Seq reads length in different populations. Sampling sites are abbreviated Mombasa (MO- 0 km), Voi (160 km), Nairobi (NA- 500 km), Nakuru (NK- 650 km), and Kakamega (KA- 850 km). The vertical line represents the read length equal 75bp.

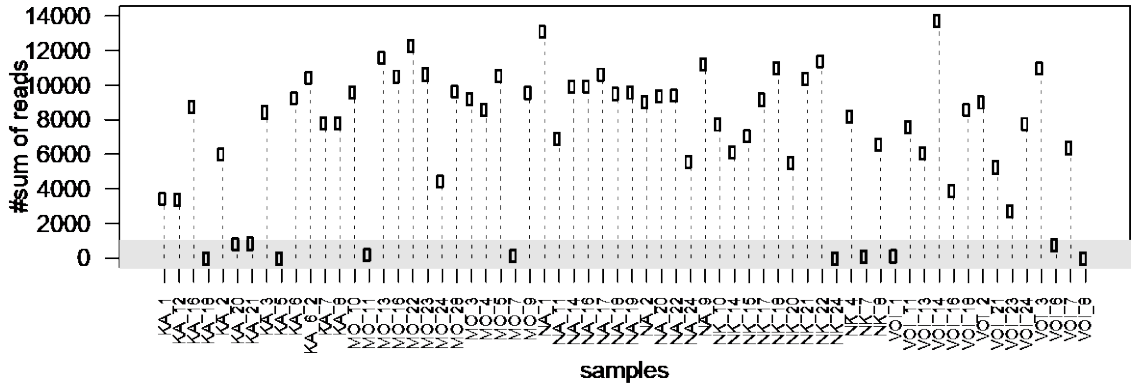


Fig. S2. epiRADseq reads by individual. Gray bar represents 1000 reads threshold. Samples with fewer than 1000 reads were not used in the epiRADseq analysis.

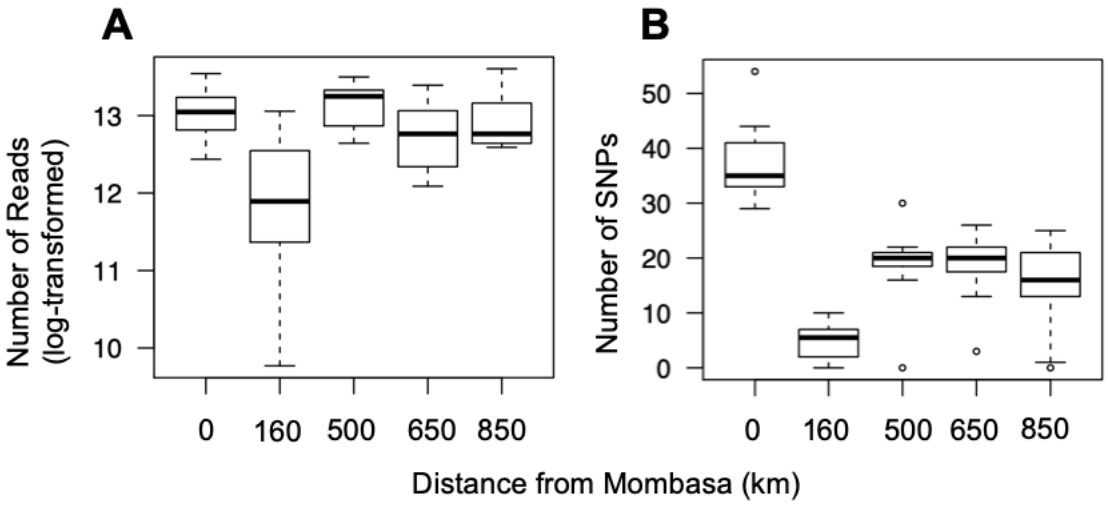


Fig. S3. Average number reads and SNPs by sampling site. Voi (160 km) has a low number of reads (A) and SNPs (B) compared to other sampling sites.

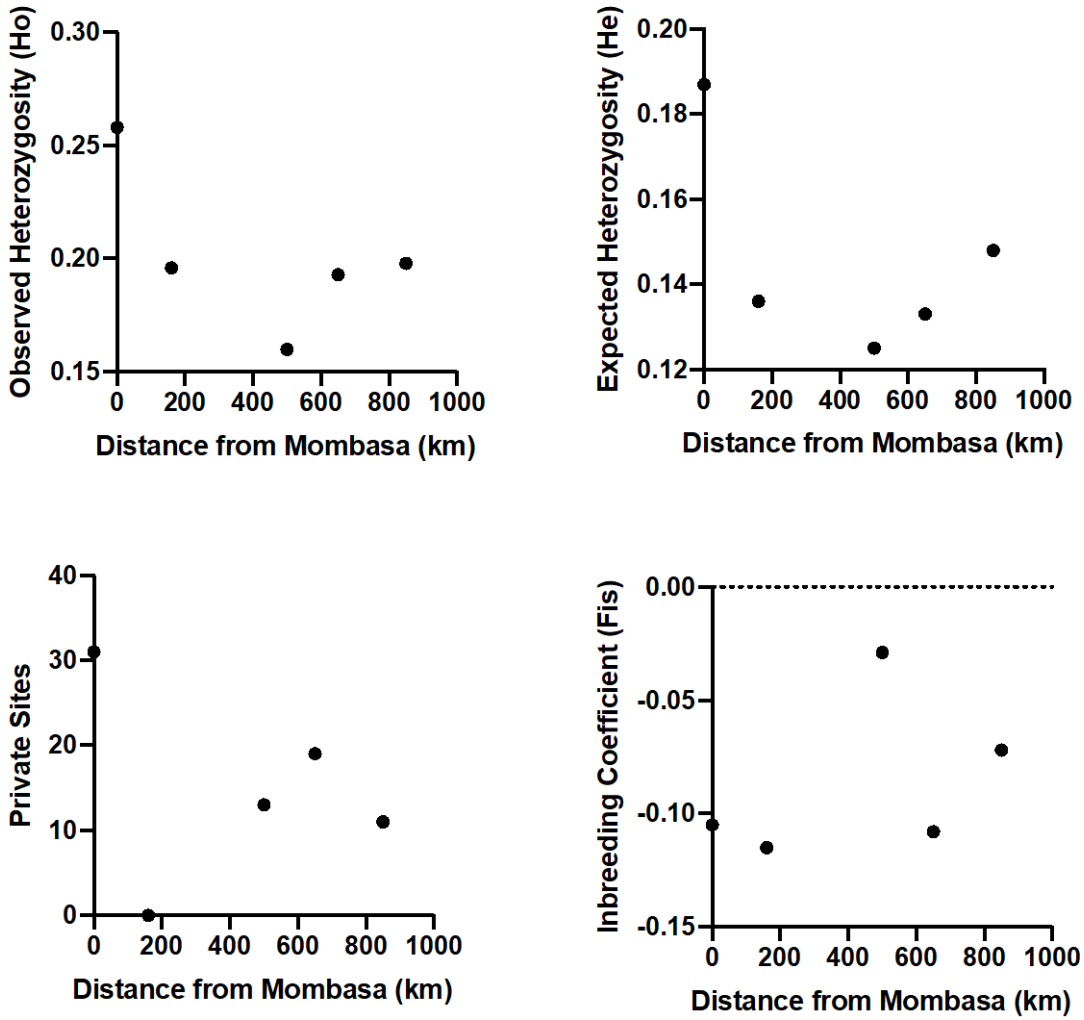


Fig. S4. Estimates of genetic diversity compared to distance from Mombasa (km). No estimate correlated with distance to Mombasa. (A) Observed heterozygosity (r= -0.5868, p= 0.2983), (B) Expected Heterozygosity (r= -0.5391, p= 0.3485); (C) Private allelic sites (r= -0.2374, p= 0.7006); (D) Inbreeding Coefficient (Fis) (r= 0.4205, p= 0.4809).

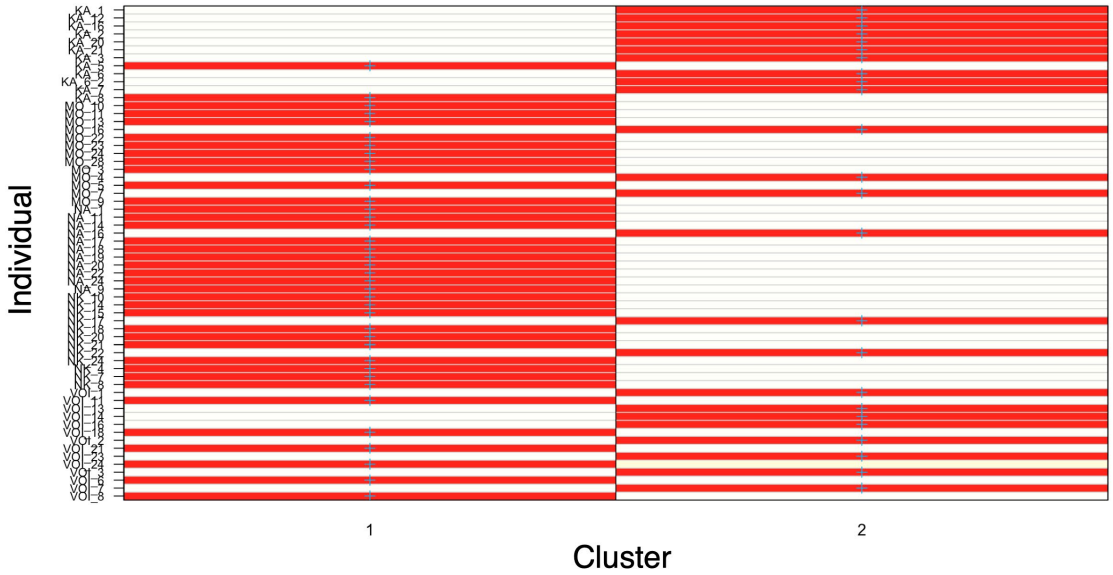


Fig. S5. Discriminant analysis of principal components. The assigned membership for each sample to one of the two predicted clusters. At least one individual from each sampling site was assigned to each cluster. In the individual IDs, sampling sites are abbreviated Mombasa (MO- 0 km), Voi (160 km), Nairobi (NA- 500 km), Nakuru (NK- 650 km), and Kakamega (KA- 850 km).