Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations | Graduate School

2016

Genome wide epigenetic analyses of Araptus attenuatus, a bark beetle

Chitra Seshadri
Virginia Commonwealth University

Follow this and additional works at: https://scholarscompass.vcu.edu/etd

 Part of the Biodiversity Commons, Desert Ecology Commons, Evolution Commons, Genetics Commons, Molecular Genetics Commons, and the Population Biology Commons

© The Author

Downloaded from
https://scholarscompass.vcu.edu/etd/4167

This Thesis is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

© Chitra Seshadri, April 2016

All Rights Reserved

Genome wide epigenetic analyses of *Araptus attenuatus*, a bark beetle

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science

at Virginia Commonwealth University

By

Chitra Seshadri

Bachelor of Science, University of Mumbai 2010

Advised by

Rodney Dyer, PhD

Department of Biology, Center for Environmental Studies

Virginia Commonwealth University

Virginia Commonwealth University,

Richmond, VA

April 2016

Acknowledgement

I thank my advisor — Dr. Rodney Dyer, for his unyielding support and guidance throughout the completion of this project. His optimism, energy, and, creativity, when tackling challenges constantly inspired me, and, showed me the kind of scientist I hope to become someday.

I also thank my committee members — Drs. Andrew Eckert, and Roy Sabo for their guidance through this process; their discussion and ideas in troubleshooting my experiments have been invaluable.

This project could not have happened without the data collection and previous work done in the region by Drs. Rodney Dyer and Ryan Garrick, Crystal Meadows, and many others. I am very grateful to Alex Waldrop, Alexandra Quinones, and Michael Kiflezghi for their efforts in the initial set of DNA extractions and quantifications that were performed.

I thank Dr. Brian Verrelli for not only allowing me to set up camp in his lab when ours was down under repairs, but also, for his valuable discussion and ideas on this research. I thank the previous and current members of the Dyer Lab, along with the Population Genetics lab group — especially Cathy Viverette, Lindsay Miles, Jameson Hinkle, and Anna Tucker for their numerous thesis edits, support, and friendship through these years.

I am grateful to Shira Shieken and others at the Global Education Office for all their help and accommodation in navigating through the endless stream of documents required for an international student. I thank the Center of Environmental Studies for this opportunity, along with Biology, for the great holiday parties, happy hours, and graduate school experience.

None of this would have happened for me if not for my husband — Kartiké Bhagat, who shared `this dream with me. I am thankful for his support through all these years; from every pep-talk, to, his complete and unwavering belief in my success.

Table of Contents

List of Tables

List of Figures

Abstract

GENOME WIDE EPIGENETIC ANALYSES OF *ARAPTUS ATTENUATUS*, A BARK BEETLE

By Chitra Seshadri, Bachelor of Science.

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2016.

Rodney Dyer, PhD
Department of Biology, Center for Environmental Studies

Phylogeographic studies have relied on surveying neutral genetic variation in natural populations as a way of gaining better insights into the evolutionary processes shaping present day population demography. Recent emphasis on understanding putative adaptive variation have brought to light the role of epigenetic variation in influencing phenotypes and the mechanisms underlying local adaptation. While much is known about how methylation acts at specific loci to influence known phenotypes, there is little information on the spatial genetic structure of genome-wide patterns of methylation and the extent to which it can extend our understanding of both neutral and putatively adaptive processes. This research examines spatial genetic structure using paired nucleotide and methylation genetic markers in the Sonoran bark beetle, *Araptus attenuatus*, for which we have a considerable knowledge about its neutral demographic history,

demography, and factors influencing ongoing genetic connectivity. Using the msAFLP approach, we attained 703 genetic markers. Of those, 297 were polymorphic in both nucleotide (SEQ) and methylation (METH) were assayed from 20 populations collected throughout the species range. Of the paired SEQ and METH locis, the METH were both more frequent (16% vs. 7%), maintained more diversity (Shannon I_{Meth} = 0.361 vs. I_{Seq}=0.272), and had more among-population genetic structure ($\Phi_{ST;\ Meth}$ = 0.035 vs. $\Phi_{ST;\ Seq}$= 0.008) than their paired SEQ loci. Interpopulation genetic distance in both SEQ and METH markers were highly correlated, with 16% of the METH loci having sufficient signal to reconstruct phylogeographic history. Allele frequency variation at five loci (two SEQ and three METH) showed significant relationships with at-site bioclimatic variables suggesting the need for subsequent analysis addressing non-neutral evolution. These results suggest that methylation can be as informative as nucleotide variation when examining spatial genetic structure for phylogeography, connectivity, and, identifying putatively adaptive genetic variance.

Keywords: Phylogeography, epigenetic variation, msAFLP, adaptation

Introduction

Environmental heterogeneity may influence the distribution of genetic variation across natural populations by exerting a selective pressure on various parts of the genome. The spatial distribution of this variation is, however, constrained by the species history, in that previous demographic perturbations determine the general distribution of total genetic variance. Once neutral history is accounted for though, the remaining spatial genetic structure can provide insight into putative adaptive mechanisms in natural populations (Gavrilets & Vose 2005; Nosil *et al.* 2009). While putatively adaptive genetic and phenotypic variation shaping natural populations have been well demonstrated, the biological mechanisms involved in such adaptive processes have not been fully elucidated. Recent studies have suggested that epigenetic mechanisms may be one additional route through which populations respond to environmental stressors; possibly conferring an increased adaptive potential on these groups (Bossdorf *et al.* 2010; Richards 2011).

Epigenetic factors and mechanisms influencing gene expression and heredity include DNA methylation, histone modification, small/micro RNAs, and other mechanisms that alter how DNA sequences are translated into functional gene products. To date, methylation of cytosine residues appears to be the most studied mechanism (Roberts & Gavery 2012). In eukaryotes, methylation is a chemical modification that involves the addition of a methyl group onto position 5 of a pyrimidine ring on cytosines (5mC), primarily within the cytosine-phosphate-guanine (CpG) dinucleotides. DNA methylation can affect various functions in gene expression (Razin & Riggs 1980); it can change the structure of the chromosome, or, if present in the promoter region

of the genome, restrict the access of transcription factors to the gene, effectively silencing the proximal gene (Klose & Bird 2006). In invertebrates, DNA methylation is thought to occur mostly in the coding regions of the genome (Roberts & Gavery 2012). DNA methylation, while varying widely between species, has been functionally linked to development, behavior, and phenotypic plasticity (Boyko *et al.* 2010; Day & Sweatt 2010; Law & Jacobsen 2010; Feng *et al.* 2010; Lyko & Maleszka 2011). For example, DNA methylation patterns in the honeybee brain – *Apis mellifera* have been associated with their life history; where workers progressing to different tasks (nursing the brood, to foraging outside the nest) with increasing age show varying patterns of methylated sites in their genome. In cases where these foragers revert back to nursing tasks, their methylation patterns have been shown to revert back to those characteristic of nest bees (Herb *et al.* 2012). While the role of epigenetics in adaptive evolution is still unclear, it has been proposed to occur by contributing to phenotypic plasticity, and thus providing a substrate upon which selection is able to act (Flores *et al.* 2013). As such, it is of interest to evaluate the utility of incorporating methylation markers in phylogeographic studies, and to see how it aids in understanding and identifying putative adaptive variation in natural populations (Petren *et al.* 2005).

Since methylation can be induced *de novo*, there are at least three categories of potential markers that can be uncovered when compared to neutral sequence-based genetic variation. Methylation loci that are induced each generation, or across a handful of generations, in response to environmental, developmental, or other localized conditions are expected to be uncorrelated with both population-level and spatial genetic structure based upon sequence-based markers. Because they are induced repeatedly, these loci are, however, expected to be identifiable due to their correlation with quantifiable at-site conditions. A second category of markers are those whose

inheritance spans many generations, conferring assayable information about the recent demographic history of the species. Inheritance across many generations would create spatial structure in these methylation markers that is correlated with that estimated from neutral sequence-based markers, more commonly used in describing demographic history. The final group of markers are uncorrelated with both neutral genetic and local heterogeneity. It is possible that methylation may have a stochastic component or be due to processes other than demographic history or local adaptation.

Assaying both sequence and methylation based genetic variation can be performed using methylation-sensitive AFLP (hereafter msAFLP) as introduced by Hill *et al.* (1996) and Reyna-López *et al.* (1997). The msAFLP protocol produces both sequence and methylation variation, localized to a single nucleotide position in the genome, thereby controlling for intra-genomic heterogeneity. This technique, uses two restriction isoschizomers, *MSP*I and *HPA*II, both of which target the same genetic sequence (5'-CCGG-3'). Both enzymes cleave DNA when there is no methylation (denoted as +, + for *MSP*I and *HPA*II respectively), although they exhibit different expression patterns in the presence of methylation at one, or more of the cytosine residues. When a methyl group is attached to the internal cytosine (denoted as CmCGG), *HPA*II is blocked from cutting (+,-). This same pattern is also observed when a methyl group is attached to the external cytosine on one strand of DNA, a condition known as hemimethylation. The frequency of hemimethylation is low, and for the purposes of this work will be lumped in with markers indicating methylation at the internal cytosine location. If both cytosine residues are methylated (complete methylation, denoted as mCmCGG), both *MSP*I and *HPA*II are blocked from cutting (-,-). Some other combinations of methylation including at least one residue being methylated also produce this (-,-) pattern, though they are thought to occur at very

low frequencies, (see Fulneček and Kovařík 2014) and the bias associated with these low frequency events will be considered to be distributed randomly across the sampling populations and as such, would add noise to downstream analyses.

If methylation is induced by local environmental heterogeneity and is not entirely stochastic, then it has the potential to influence adaptive genetic variation (Schoville *et al.* 2012). Thus, correlation between genetic or epigenetic variation with environmental gradients can be interpreted as evidence supporting the signals of natural selection (Eckert *et al.* 2010). There are two general methods commonly used to identify putatively adaptive variation in non-model organisms based upon genome scans. First, the background level of genetic divergence among populations is determined in large part by the demographic history of the populations being examined. An outlier approach seeks to identify markers whose among population divergence is exceptionally high compared to this background level. Genomic regions that are in the vicinity of divergent selection should show higher differentiation due to linkage (Nosil *et al.* 2009). This method has been used repeatedly in many systems including intertidal snails (*Littorina saxatillis*) (Grahame & Wilding 2006), whitefish (*Coregonus clupeaformis*; Campbell & Bernatchez 2004), and the common frog (*Rana temporaria*; Bonin *et al.* 2006). Outlier loci have been shown to underestimate demographical effects, thus increasing the number of false positives, especially in the case of natural populations diverged during glacial periods (Excoffier *et al.* 2009; Garrick *et al.* 2013). To counter this potential problem, Rellstab *et al.* (2015) suggest that analyses be based upon environmental association in covariance while controlling for neutral genetic structure (Rellstab *et al.* 2015).

One such approach to do this is by the use of gradient analyses—the so-called spatial analysis method (SAM) by Joost (2007). This approach looks for systematic changes in allele frequencies along ecological gradients (Coop *et al.* 2010). The rationale here is that the if environmental gradients are at least correlated with fitness related components of the genome, variation in genetic markers should mimic gradients in ecological features. This approach does not require that the overall amount of differentiation is considerably larger than that created by demographic history, only that it changes systematically along with environmental gradients. It should be pointed out though, that both of these approaches only provide insights into genomic regions that may be linked to fitness related traits, and subsequent analyses and experimentation (Holderegger & Wagner 2008) are required to show they are actually adaptive.

The main objective in this study was to investigate the effectiveness of using epigenetic variation in informing demographic and evolutionary processes shaping natural populations. This study uses the Sonoran desert bark beetle, *Araptus attenuatus* Wood (Circulionidae), known only to inhabit the senescing stems of the succulent *Euphorbia lomelii* (Euphorbiaceae) in Baja California. The Sonoran bioregion is characterized by steep environmental gradients known to have promoted population subdivision during post Pleistocene range expansion (Garrick *et al.* 2009, 2013). While most species have exhibited a northward range expansion, this insect species has experienced a bi-directional range expansion, presumably due to a mid-peninsular warm desert refuge (Garrick *et al.* 2013). Previous studies have also demonstrated the influence of its host plants' demographic history to its own structure, the strength of which is attributed to the close relationship of the plant-insect pair. From a set of 20 populations, DNA sequence and

methylation states were assayed using the msAFLP technique (Fulneček & Kovařík 2014). The main questions were as follows:

1. *Is the amount and spatial distribution of genetic variation in methylation based markers congruent with what is observed in the paired sequence based markers?* Since each sequence and methylation locus is paired, targeting the exact same location in the genome, direct comparison of basic structure statistics is not confounded by marker location. The expectations for the relationship between methylation and sequence structure depend upon the processes inducing methylation. If methylation variation is entirely neutral and is inherited as a Mendelian trait across many generations then both sequence and methylation structure should share the same spatial-genetic structure. If methylation is induced on a recent timescale, is stochastic, or is responding to environmental conditions at a different temporal (e.g., developmental conditions) or spatial (perhaps regional conditions) scale, it would be expected to have less spatial genetic structure than sequence variation because demographic history is not being recapitulated. Finally, if methylation variation is greater than sequence variation then methylation loci may be conferring information about non-neutral processes that sequence variation is not capturing (e.g., hidden variance).
2. *Is there evidence in the spatial arrangement of genetic variance in either methylation or sequence-based markers for putatively adaptive genetic variance?* By adopting gradient-type analyses, both sets of markers were examined for systematic changes with at-site environmental conditions. While a stronger correlation of epigenetic loci to environmental factors is intuitively indicative as being more useful than sequence based variation in uncovering adaptation in natural populations, we should bear in mind the

complexities involved in the pathway connecting environmental variation, phenotypic plasticity, epigenetic, and genetic variation, and how it might affect inheritance underlying adaptation in a species. However, methylated loci may be a hidden repository of both neutral and putatively adaptive variation that we have yet to investigate fully. The potential for methylation based markers as a tool more broadly applicable in population and evolutionary genetic studies are discussed in light of the results of these questions.

Methods

Araptus attenuatus Wood (Curculionidae) is a Sonoran desert endemic bark beetle known only from the euphorb *Euphorbia lomelii* (Garrick *et al.* 2013). The insect and its host plant are distributed throughout peninsular Baja California and in at least two reticual populations in mainland Mexico in the states of Sonora and Sinaloa. Both host plant and beetle have been profoundly influenced by post-Pleistocene range expansion (Garrick *et al.* 2009). At present, three major clades have been identified within *Araptus.attenuatus*. The extent of mitochondrial DNA sequence divergence (8-12%) suggests potential cryptic speciation in this taxon. On peninsular Baja California, there are two divergent clades with regions in partial sympatry (Garrick *et al.* 2013; Figure 4); Clade B is widespread and diverse, while Clade C (Cape region) is predominantly restricted to the southern cape region. The third clade, Clade S (Sonora) is allopatric, found only on continental Sonora, although is postulated to have its ancestry in the Cape region due to both host plant and insect phylogeographic reconstructions (Garrick *et al.* 2009, 2013).

Individuals belonging to the larger clade (Clade B) were collected from 20 populations (Figure 2) spanning regions that were not overlapping with Clade C. Evidence from other studies in the Dyer laboratory suggests no introgression between peninsular clades (Garrick *et al.* 2013). From these populations, at least six individuals were sampled per population (Table 1) to maximize the number of populations from which to test. Each individual was collected from different plants to avoiding the confounding effects of sampling siblings. All samples were stored in 90% ethanol and kept in the Dyer laboratory until template DNA was extracted.

Genomic DNA was extracted from each individual using the Qiagen DNA Blood and Tissue extraction kit (Qiagen Inc., Natick Mass). To aid in cell lysis for this small beetle, samples were incubated overnight at 56 °C with proteinase-K. The elution step was performed in 50 ul of sterile DNase/RNase free water, instead of the recommended Qiagen elution buffer (buffer 'EB') as it was found to interfere with downstream analysis. A subset of the extractions was replicated for QC/QA to determine reproducibility of genetic markers in the AFLP protocol.

The msAFLP approach was used to contrast paired sequence and methylation markers sampled from across the genome. This approach is a modification of the standard AFLP protocol (Fulneček & Kovařík 2014), identifying a large number of loci without any prior information on the genome of the organism (Mueller & Wolfenbarger 1999). The Dyerlab protocol was adopted from Reyna-López *et al.* (1997) and Keyte *et al.* (2006). Template DNA was standardized in concentration so no more than 200 ng of genomic DNA was used. Samples of 20 ul were digested with a combination of *EcoRI* and either of the isoschizomers *HpaII* or *MspI*. These isoschizomers cut at the exact same nucleotide sequence (Figure 1) and in the absence of methylation, both cut target DNA sequences (+,+). When there is a methyl group attached to the internal cytosine residue (CmCGG), or in the rare case where the external residue is hemimethylated, *HpaII* is blocked from cleaving DNA (+,-). Both enzymes are blocked from cutting (-,-) when both cytosine residues are methylated (Keyte *et al.* 2006). Restriction digests were incubated at 37 °C for 3 h, followed by enzyme inactivation at 80 °C for 20 min. Both *EcoRI* and *HpaII*/*MspI* digestions were performed simultaneously. DNA primers (1.875 µmol) attached to the end of fragments cut with *EcoRI* and *HpaII*/*MspI* were ligated onto the fragments using T4 DNA ligase (New England Biolabs) at 16 °C overnight (Table 2).

Two sets of Polymerase Chain Reaction (PCR) amplifications were performed on the fragments. The pre-selective amplification has primers (see Table 2 again) matching the sequences ligated onto the fragments plus a single additional nucleotide. Amplifications consisted of 10 µl of digested-ligated product, 0.8 µM each of *Eco*RI + A primer, and *HPA*II/*Msp*I + C primer, 10X polymerase chain reaction (PCR) buffer, 1.5 µl 50 µM $MgCl_2$, 4 µl 10 mM deoxynucleoside triphosphates (dNTPs), and 0.5 µl 5U/µl Taq DNA polymerase, in a total volume of 50 µl. The reaction conditions were: 75 °C for 2 min, followed by 20 cycles of 94 °C for 50 s, 56 °C for 1 min, and 72 °C for 2 min, and a final extension for 30 min at 60 °C. Pre-selective PCR products were diluted 50 X in sterile water for the next step.

The Selective Amplification step was conducted using a 2X Type-It Microsatellite PCR kit, in a volume of 25 µl, using 10 µl of the diluted PCR product from the pre-selective amplification. This reaction was performed using 12.5 µl 2x Type-it Multiplex PCR Master mix, 2.5 µl 2µM of each primer – *Eco*RI + ACT, and *Msp*I + CGT (see Table 2 again). The *EcoRI* primer was pre-labeled with tetrachlorinated analogue of 6-carboxyfluorescein (6-FAM). The PCR parameters included a heat inactivation step at 95 °C for 5 min, followed by a 3-step cycling process, for 28 cycles – a denaturation step at 95 °C for 30 s, an annealing step at 63 °C for 90 s, and lastly an extension step 72 °C for 30 s. The samples were cleaned of unused primers and other oligonucleotides using the EXO-SAP IT kit (USB Co., Amersham).

Each sample was assayed for both a non-methylation sensitive (SEQ) marker profile using the *EcoRI+MspI* combination of enzymes and a methylation sensitive (METH) genetic marker profile using *EcoRI+HPAII*. In the absence of methylation, the presence or absence of DNA fragments will be identical in SEQ and METH profiles, whereas with methylation, either as a

hemimethylated cytosine or as a methylation site on the internal cytosine residue, SEQ and METH will produce different methylation profiles (see again Figure 1).

Fragments from both SEQ and METH profiles were identified using an ABI3730xl DNA Analyzer (Applied Biosystems, Inc.), using LIZ500 (orange) as the size standard. Individual trace files for each trace were analyzed using the 'Binner' package (version 0.1, Smith 2014) in R (version 3.0.1). Bin sizes for fragments were set to be in the range of 1-1.5bp in width. Profiles were generated automatically and then checked by hand. Duplicate samples were run for QA/QC and bands with poor repeatability (e.g., error rates in duplicate runs exceeding 5%) were dropped from the analyses (see again Table 1). Bins were defined based upon SEQ profiles and then METH profiles were called using the same classification scheme. Individual fragments were converted to AFLP genetic markers (absence/presence of fragment) for each locus using the gstudio library (Dyer, 2015).

Analyses of Relative Genetic Signal

Only loci with intermediate fragment frequencies between 0.05 – 0.95 were retained for subsequent analyses, since loci occurring at rates above 95% and below 5% may lead to spurious correlations and are not considered reliable (Pérez-Figueroa 2013). Analyses of methylation status were performed using the msap package (vers. 1.1.8 Pérez-Figueroa 2014) providing estimates of population level fragment frequencies for all methylation states (unmethylated, hemimethylated, internal cytosine methylation, and full methylation or absence of target). Overall fragment diversity for both SEQ and METH banding patterns were estimated using the Shannon index and statistical differences between the two were tested using a Wilcoxon rank

sum test (Lowry, Richard. "Concepts & Applications of Inferential Statistics". Retrieved 24 March 2011).

Multilocus genetic divergence was estimated separately for SEQ and METH data using the AMOVA (Analysis of MOlecular VAriance) approach from Excoffier *et al.* (1992). Genetic divergence was estimated for both SEQ loci at METH sites, and, overall — including SEQ loci at non-methylated sites; in order to test for SEQ loci under selection for uniformity at METH regions. Significance of the test statistic, ϕ_{ST}, deviating from zero is estimated based upon 10000 permutations of the design matrix. In addition to the magnitude of structure in SEQ and METH fragment profiles, inter-population structure was also examined. If the same spatial signal is contained in both SEQ and METH profiles in bulk, then inter-population genetic distance should be significantly correlated. Multilocus Euclidean genetic distance among population centroids was estimated for both SEQ and METH genetic data sets using the *gstudio* package (Dyer 2015) and compared using a Mantel test (vegan package, version 2.3-1; Okasanen *et al.* 2015). Significance of the correlation coefficient was tested using 999 permutations of the design matrix.

The relative amount of genetic structure estimated in paired SEQ and METH loci were examined by estimating Weir and Cockerham's θ (1984). Similarity in paired estimates of standing genetic structure measured in both marker sets was estimated using Spearman's rank sum test. Finally, to identify the fraction of METH loci that are inherited over a period long enough to recapitulate phylogeographic history, population covariance at each METH locus was compared to multilocus covariance based upon all SEQ markers, using a Mantel test while correcting for multiple comparisons via a Bonferonni correction.

Putative Signals of Selection

Climate data were gathered from the WORLDCLIM Bioclim data at 30 arc-seconds resolution (1 km^2 resolution). We obtained data for monthly total precipitation, and monthly mean, minimum and maximum temperature, and 19 derived bioclimatic variables as listed on Table 4 (http://www.worldclim.org/).

A logistic regression of the presence of bands at both SEQ and METH loci on Bio-Climatic features was performed in R, to determine if variation in these loci covary with broad scale environmental gradients. These logistic regression models were generated for methylation and sequence based loci. All models were Bonferonni corrected for multiple comparisons and only those whose significance was less than $P = 0.0001$ were considered.

Results

The msAFLP analysis was performed on a total of 120 *A.attenuatus* individuals sampled from 20 *E. lomelii* populations (Figure 2). A total of 297 loci were identified that were both polymorphic and had variants that occurred at frequencies between 5% - 95%. Of these, 21 loci were polymorphic for SEQ variants though had no variation in methylation. Overall epigenetic diversity in methylation sensitive (I_{Meth}= 0.360) fragments was greater than that for genetic diversity (I_{Seq} = 0.275; Wilcoxon Rank Sum, W=3032, P = 0.0355). The frequency of methylation occurrence varied by methylation state (hemimethylated vs. internal cytosine methylation) and sampling locale (Table 3 - Methylation Frequency). At the level of the population, there is no correlation between hemimethylated and internal cytosine methylation frequencies (Pearson, t=-1.1323, df=18, P=0.272) suggesting that they may be treated as independent markers.

Analyses of Relative Genetic Signal

The level of among population differentiation varied between methylation and sequence fragment sets. The estimate of multilocus genetic divergence, ϕ_{ST}, was 4.5X greater in methylation loci than in sequence loci. In fact, among the populations sampled, methylation fragment profiles (representing nucleotide divergence) were not significantly different than zero (ϕ_{ST}= 0.007765, *P* = 0.277), whereas methylation sensitive (both hemimethylated and internal cytosine methylation) population divergence was ϕ_{ST}= 0.03465 (*P* =0.0016). We note that SEQ loci at methylated sites are possibly under selection for uniformity at METH regions, from the

comparison of the overall AMOVA scores of METH (ϕ_{ST}= 0.002382252) and SEQ loci (ϕ_{ST}=0.001713423), where METH loci are 1.5X more divergent than SEQ loci.

Inter-population divergence assayed as multilocus distance in both SEQ and METH (Figure 3) were highly correlated (Mantel; r = 0.8924, *P*=0.001), indicating that the multilocus covariance among populations in epigenetic variance can approximate that created from sequence-based markers. Pairwise genetic structure at individual loci ranged from -0.08 – 0.28 for MSPI loci and -0.19 – 0.33 for HPA loci and averaged 0.05 and 0.07, respectively. However, paired single-locus structure in both marker sets was not correlated (Spearman Rank; S=3890800, P = 0.468). In relation to multilocus covariance measured across populations estimated across all SEQ loci, single locus population covariance at 45 METH loci were found to be significantly correlated after correcting for multiple comparisons using a Bonferonni correction.

Putative Signals of Selection

Both sets of loci produced logistical regression models whose fit suggests a high degree of congruence with ecological gradients (Table 5). Under the most stringent criteria, there were 5 methylation loci related to systematic changes in temperature and latitude along the peninsula. One locus, 356.4, was found to have frequencies predicted by three different temperature variables. It is also noteworthy that from the 5 loci (both sequence and methylated) that were identified as responding to environmental features, only 2 of them were sequence marker, of which locus 306.7 was responsive to all the environmental variables tracked by sequence loci. The other sequence locus - locus 288.9 however was sensitive to only changes in latitude. These results suggest that methylation may contain at least as much biologically informative information suitable for investigating putatively adaptive genetic variance and local adaptation as

sequence based markers. Further investigation into the pathway involved in effecting these changes may shed light into the differences we have observed.

Taken individually, the magnitude of inter-population genetic structure present in SEQ and METH loci appear to diverge (Figure 4). Both marker types exhibit a skewed distribution of values with most loci exhibiting low levels of overall diversity and a few exhibiting increased divergence. Surprisingly, loci with high structure in SEQ loci are not the same loci with high structure in METH loci (e.g., few markers along the $\theta_{Seq} = \theta_{Meth}$).

Discussion

Using the msAFLP technique, genetic and epigenetic differentiation was compared, simultaneously, in populations of *A.attenuatus* its species range in Baja California. The data produced patterns of variation that support the hypothesis that methylation-based markers contain both neutral and putatively adaptive variation. They also seem to contain signal that we are unable to quantify using sequence-based loci alone providing more raw material on which to build phyolgeographic and evolutionary genetic studies. These results suggest that the addition of methylation-based markers may allow inferences regarding both neutral and adaptive genetic covariance rivaling the signal attained by sequence markers alone.

By scanning patterns of genome wide sequence based and epigenetic polymorphism for individuals in several populations, it was possible to identify genomic regions exhibiting increased divergence because of direct or indirect (through linkage) selection (Luikart *et al.* 2003). Identifying putatively adaptive loci underscored the potential importance of methylation polymorphism in phylogeographic studies. Surprisingly, when examining among-population structure measured at paired SEQ and METH loci, the ones with high divergence in one dataset were generally not the same outliers in the other dataset. If high divergence is an indicator of natural selection—a working hypothesis used in numerous evolutionary studies—then these results suggest that the amount of putatively adaptive variation found by using only SEQ markers may be a large underestimate of standing adaptive variation. It is clear from these data that markers of high METH divergence are attached to cytosine sequences that do not have correspondingly high divergence. The converse is also true in that there are SEQ loci that have high divergence while having polymorphic METH variation with low inter-population structure.

Allele frequencies at five loci showed significant relationships with at-site bioclimatic variables; three methylated loci, and two sequence loci were identified as being significantly correlated to environmental features ($P<0.0001$). However, one of the sequence based loci changed with latitude, while the other locus, 306.7, was significantly associated with latitudinal and longitudinal changes, and also with temperature measures during the wettest quarter, and precipitation seasonality. Along with the epigenetic loci; which exhibited a correlation with average temperature changes and latitudinal changes, these associations suggest subsequent analysis for local adaptation.

The use of methylation markers is something that has recently started to receive attention in population and landscape genetic studies. Before they become more prominent, it is perhaps prudent to discuss some of the complications associated with their use. By far, the most problematic issues identified is that they have been shown to underestimate genome-wide levels of methylation due to a variety of limitations brought about by the restriction enzymes used. Firstly, the enzymes used herein recognize only the CCGG site, leaving methylation nested within other sequence motifs undetected. In the absence of cuts in using both SEQ and METH profiles, this approach is unable to differentiate between full methylation (e.g., a methyl group on both cytosines) and the absence of sequence-based recognition sites (Fulneček & Kovařík 2014). However, at present, there is no indication that the bias introduced by these faults would result in systematic problems estimating among-population structure and/or correlations between the presence of bands and environmental gradients.

Moreover, despite these limitations of this method, it is suitable for rapid assays producing a large set of putative fragments amenable for analyses. In this study, 276 (92.9%) methylated

sites (hemimethylated or methylated at the inner cytosine) were detected, of which 226 were polymorphic (82%). It is noteworthy that although a higher level of polymorphism was detected using non-methylated loci, overall epigenetic diversity at the population level was significantly higher (Wilcoxon Rank Sum, W=3032, P = 0.0355) at methylation-based loci. Taken in total, among-population multilocus structure showed a high degree of congruence with sequence-based markers. While inter-population covariance was significantly correlated, the magnitude of standing genetic structure measured among populations was greater in the METH loci than estimated among SEQ loci.

However, despite these complications, these results suggest that methylation-based markers should be considered as an additional repository of genetic (both neutral and putatively adaptive) variation that has not been fully appreciated in population-level and evolutionary genetic studies. A pertinent follow-up question would be with respect to the autonomy of epigenetic variation. In line with that, we recommend the following steps for further analysis (Bossdorf *et al.* 2008): Performing a deeper sequencing technique to analyzing the extent and structure of epigenetic variation with and among this species. SMP - Single Methylation Polymorphisms allow for the differentiation between the internal cytosine methylation and other patterns that are not assayable using this technique (e.g., Platt *et al.* 2015).

While these data provide some inferences in the partitioning of methylation based loci into categories that recapitulate neutral history, align with at-site conditions, or deviate with both, the use of neutral population covariance structure significantly aided in differentiating among the first two categories. In general, methylation studies are not conducted in the context of

population history and instead focus on specific gene products and resulting phenotypes. Further understanding about the distinctions between these categories are needed. For example, it would be beneficial to set up breeding experiments in investigating not only the fidelity of methylation variation, but also, how they might might affect phenotypic variation in the identified ecologically relevant traits. Given the life history traits of our study species and its coassociation with its host plant, these kinds of studies may uncover not only individual evolutionary trajectories but also elucidate coevoluationary processes in both taxa. At the very least, population and landscape geneticists alike now have an additional data set on which to develop and test their hypotheses.

Bibliography

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). *Molecular Biology and Evolution*, **23**, 773–783.

Bossdorf O, Arcuri D, Richards CL, Pigliucci M (2010) Experimental alteration of DNA methylation affects the phenotypic plasticity of ecologically relevant traits in Arabidopsis thaliana. *Evolutionary Ecology*, **24**, 541–553.

Bossdorf O, Richards CL, Pigliucci M (2008) Epigenetics for ecologists. *Ecology letters*, **11**, 106–15.

Boyko A, Blevins T, Yao Y *et al.* (2010) Transgenerational Adaptation of Arabidopsis to Stress Requires DNA Methylation and the Function of Dicer-Like Proteins. *PLoS ONE*, **5**, e9514.

Campbell D, Bernatchez L (2004) Generic Scan Using AFLP Markers as a Means to Assess the Role of Directional Selection in the Divergence of Sympatric Whitefish Ecotypes. *Molecular Biology and Evolution*, **21**, 945–956.

Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using Environmental Correlations to Identify Loci Underlying Local Adaptation. *Genetics*, **185**, 1411–1423.

Day JJ, Sweatt JD (2010) DNA methylation and memory formation. *Nature neuroscience*, **13**, 1319–23.

Eckert AJ, van Heerwaarden J, Wegrzyn JL *et al.* (2010) Patterns of Population Structure and Environmental Associations to Aridity Across the Range of Loblolly Pine (Pinus taeda L., Pinaceae). *Genetics*, **185**, 969–982.

Excoffier L, Foll M, Petit RJ (2009) Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 481–501.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Feng S, Jacobsen SE, Reik W (2010) Epigenetic Reprogramming in Plant and Animal Development. *Science*, **330**, 622–627.

Flores KB, Wolschin F, Amdam G V (2013) The role of methylation of DNA in environmental adaptation. *Integrative and comparative biology*, **53**, 359–72.

Fulneček J, Kovařík A (2014) How to interpret Methylation Sensitive Amplified Polymorphism (MSAP) profiles? *BMC Genetics*, **15**, 2.

Garrick RC, Nason JD, Fernández-Manjarrés JF, Dyer RJ (2013) Ecological coassociations influence species' responses to past climatic change: An example from a Sonoran Desert bark beetle. *Molecular Ecology*, **22**, 3345–3361.

Garrick RC, Nason JD, Meadows C a., Dyer RJ (2009) Not just vicariance: Phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula. *Molecular Ecology*, **18**, 1916–1931.

Gavrilets S, Vose A (2005) Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 18040–18045.

Grahame JW, Wilding CS (2006) Adaptation To a Steep Environmental Gradient and an Associated Barrier To Gene Exchange in Littorina Saxatilis. *Evolution*, **60**, 268–278.

Herb BR, Wolschin F, Hansen KD *et al.* (2012) Reversible switching between epigenetic states in honeybee behavioral subcastes. *Nature Neuroscience*, **15**, 1371–1373.

Hill M, Witseboer H, Zabeau M *et al.* (1996) AFLP fingerprinting as a tool for studying genetic relationships in Lactuca spp. *Theoretical and Applied Genetics*, **93**, 1202–1210.

Holderegger R, Wagner HH (2008) Landscape Genetics. *BioScience*, **58**, 199.

Joost S, Bonin A, Bruford MW *et al.* (2007) A spatial analysis method (SAM) to detect candidate loci for selection: Towards a landscape genomics approach to adaptation. *Molecular Ecology*, **16**, 3955–3969.

Keyte AL, Percifield R, Liu B, Wendel JF (2006) Infraspecific DNA methylation polymorphism in cotton (Gossypium hirsutum L.). *The Journal of heredity*, **97**, 444–50.

Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. *Trends in Biochemical Sciences*, **31**, 89–97.

Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nature Reviews Genetics*, **11**, 204–220.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Lyko F, Maleszka R (2011) Insects as innovative models for functional studies of DNA methylation. *Trends in Genetics*, **27**, 127–131.

Mueller, Wolfenbarger (1999) AFLP genotyping and fingerprinting. *Trends in ecology & evolution*, **14**, 389–394.

Nosil P, Funk DJ, Oritz-Barrientoz D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Pérez-Figueroa A (2013) msap : a tool for the statistical analysis of methylation-sensitive amplified polymorphism data. *Molecular Ecology Resources*, **13**, 522–527.

Petren K, Grant PR, Grant BR, Keller LF (2005) Comparative landscape genetics and the adaptive radiation of Darwin's finches: The role of peripheral isolation. *Molecular Ecology*, **14**, 2943–2957.

Platt A, Gugger PF, Pellegrini M, Sork VL (2015) Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations. *Molecular Ecology*, **24**, 3823–3830.

Razin A, Riggs AD (1980) DNA Methylation and Gene Function. *Science*, **210210**, 604–610.

Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holderegger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, (in press).

Reyna-López GE, Simpson J, Ruiz-Herrera J (1997) Differences in DNA methylation patterns are detectable during the dimorphic transition of fungi by amplification of restriction polymorphisms. *Molecular & general genetics : MGG*, **253**, 703–10.

Richards EJ (2011) Natural epigenetic variation in plant species: a view from the field. *Current opinion in plant biology*, **14**, 204–9.

Roberts SB, Gavery MR (2012) Is there a relationship between DNA methylation and phenotypic plasticity in invertebrates? *Frontiers in Physiology*, **2**, 1–5.

Schoville SD, Bonin A, François O *et al.* (2012) Adaptive Genetic Variation on the Landscape: Methods and Cases. *Annual Review of Ecology, Evolution, and Systematics*, **43**, 23–43.

Tables & Figures

Table 1: Sample identification numbers from 20 mainland populations found within the larger Baja California clade (Clade B). Individual identification numbers represent the population number followed by a decimal and the plant number within the population. Samples with asterisks indicate individuals resampled and re-run to be used to verify repeatability of fragment profiles.

Population	Individuals							
12	12.1	12.2	12.3	12.8	12.7	12.6	12.5	
51	51.17*	51.18*	51.19*	51.20*	51.13*	51.15		
58	58.15	58.18	58.17	58.12	58.13	58.19	58.2	
64	64.14	64.8	64.7	64.5				
88	88.15	88.17	88.11	88.12	88.13	88.14		
89	89.31	89.35	89.33	89.32	89.37	89.4	89.34	
93	93.11	93.19	93.15	93.18	93.16	93.14	93.17	
153	153.13	153.15	153.17	153.19	153.11*	153.12		
159	159.12	159.18	159.13	159.14	159.11	159.17		
160	160.11	160.12	160.14	160.15	160.17	160.19	160.2	
161	161.1*	161.6	161.4*	161.2*	161.1	161.3	161.9*	
162	162.2	162.3	162.4	162.5	162.7	162.1	162.11	162.12
165	165.4	165.2	165.1	165.3	165.9	165.1	165.11	
168	168.1	168.5	168.8	168.2	168.3	168.9	168.6	
169	169.3	169.7	169.6	169.8	169.1	169.2	169.4	
171	171.2	171.3	171.4	171.5	171.6	171.7	171.1	
173	173.1	173.4	173.3	173.2	173.6	173.5		
174	174.18	174.17	174.16	174.15	174.14	174.13		
175	175.1	175.2	175.3	175.4	175.5			
177	177.1	177.4	177.5	177.7	177.2	177.3A	177.6	177.8

Table 2: Adaptors and Primers used in this study (Overhanging nucleotides are shown in italics)

Adaptors	
*Eco*RI – adaptorI	5'*CTC*GTAGACTGCGTACC
*Eco*RI – adaptorII	5'*AATT*GGTACGCAGTCTAC
*HPA*II/*Msp*I – adaptorI	5'*GACG*ATGAGTCCTGAG
*HPA*II/*Msp*I – adaptorII	5'*CG*CTCAGGACTCAT
Preselective primers	
*Eco*RI + A	5'GACTGCGTACCAATTCA
*HPA*II/*Msp*I + C	5'GACGATGAGTCCTGAGCGGC
Selective primers	
*Eco*RI + ACT	*Eco*RI + A + CT
*HPA*II/*Msp*I + CGT	*HPA*II/*Msp*I + C + GT

Table 3: Population-level nucleotide and methylation frequencies.

	Methylation Levels			
Population	HPA+/MSP+ (Unmethylated)	HPA+/MSP- (Hemimethylated)	HPA-/MSP+ (Internal cytosine methylation)	HPA-/MSP- (Full methylation or absence of target)
12	0.05745	0.15217	0.14079	0.64959
51	0.04469	0.05797	0.10145	0.79589
58	0.05487	0.14545	0.14803	0.65166
64	0.09239	0.21467	0.11685	0.57609
88	0.05495	0.1413	0.17874	0.625
89	0.07609	0.14674	0.17572	0.60145
93	0.08696	0.1413	0.22826	0.54348
153	0.08756	0.20894	0.10447	0.59903
159	0.05918	0.15157	0.17814	0.61111
160	0.0854	0.1972	0.1382	0.5792
161	0.0471	0.1029	0.1341	0.7159
162	0.08832	0.22192	0.1481	0.54167
165	0.0854	0.1962	0.1641	0.5543
168	0.06039	0.1564	0.12198	0.66123
169	0.07867	0.1853	0.19151	0.54451
171	0.08799	0.17805	0.18892	0.54503
173	0.1123	0.1902	0.2144	0.4831
174	0.0558	0.06667	0.26449	0.61304
175	0.09783	0.18841	0.21232	0.50145
177	0.06386	0.22328	0.08469	0.62817
Mean	0.07386	0.163332	0.161763	0.601045

Table 4: WorldClim bioclimatic feature variables at 30 arc-seconds resolution extracted from WORLDCLIM.

Code	Bioclimatic Feature
BIO1	Annual Mean Temperature
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp))
BIO3	Isothermality (BIO2/BIO7) (* 100)
BIO4	Temperature Seasonality (standard deviation *100)
BIO5	Max Temperature of Warmest Month
BIO6	Min Temperature of Coldest Month
BIO7	Temperature Annual Range (BIO5-BIO6)
BIO8	Mean Temperature of Wettest Quarter
BIO9	Mean Temperature of Driest Quarter
BIO10	Mean Temperature of Warmest Quarter
BIO11	Mean Temperature of Coldest Quarter
BIO12	Annual Precipitation
BIO13	Precipitation of Wettest Month
BIO14	Precipitation of Driest Month
BIO15	Precipitation Seasonality (Coefficient of Variation)
BIO16	Precipitation of Wettest Quarter
BIO17	Precipitation of Driest Quarter
BIO18	Precipitation of Warmest Quarter
BIO19	Precipitation of Coldest Quarter

Table 5: Loci sensitive to the environmental factors as identified using more stringent regression models ($P < 0.0001$), are listed. Some loci are sensitive to more than one environmental factor. Loci in red are METH loci, and those in black are SEQ loci.

Environmental Variables	288.9	306.7	241.2	356.4	397.3
BIO15		$R^2 = 0.055$ ***P* = 3.50x10^{-5}**			
BIO2				$R^2 = 0.078$ ***P* = 5.49x10^{-5}**	
BIO6				$R^2 = -0.073$ ***P* = 1.38x10^{-5}**	
BIO7				$R^2 = 0.074$ ***P* = 4.39x10^{-5}**	
BIO8		$R^2 = 0.013$ ***P* = 8.43x10^{-5}**			
Latitude	$R^2 = -0.698$ ***P* = 9.38x10^{-5}**	$R^2 = -0.851$ ***P* = 9.60x10^{-6}**	$R^2 = -0.772$ ***P* = 7.30x10^{-5}**		$R^2 = -0.839$ ***P* = 2.45x10^{-5}**
Longitude		$R^2 = 0.885$ ***P* = 8.18x10^{-5}**			

	HpaII+	*HpaII-*
MspI+	Unmethylated 5’---C▼CGG---3’ 3’---GGC▲C---5’	Internal Cytosine Methylation 5’---C▼^{m}CGG---3’ 3’---GGC^{m}▲C---5’
MspI-	Hemimethylated 5’---C▼^{m}CGG---3’ 3’---GGC▲C ---5’	Full Methylation 5’---^{m}C▼^{m}CGG---3’ 3’---GGC^{m}▲C^{m} ---5’

Figure 1: This figure outlines the inferences that may be drawn based on the absence (-), or, presence (+) of the *MSP*I, and/or, *HPA*II bands. The methylation status at each cut-site, along with a depiction of the enzymatic action at that cut site is also detailed.

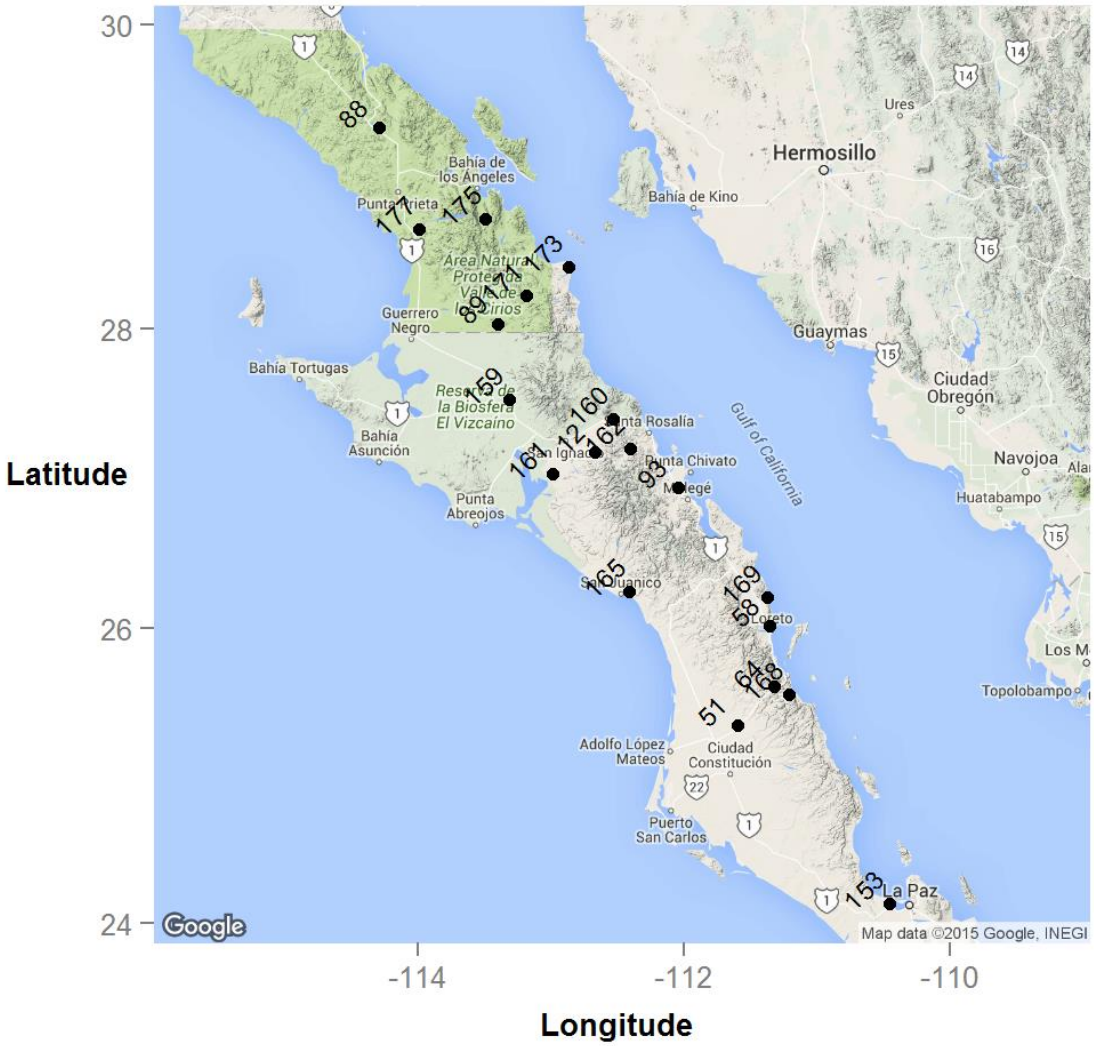

Figure 2: Spatial locations of the 20 Peninsular populations in Baja California, Mexico.

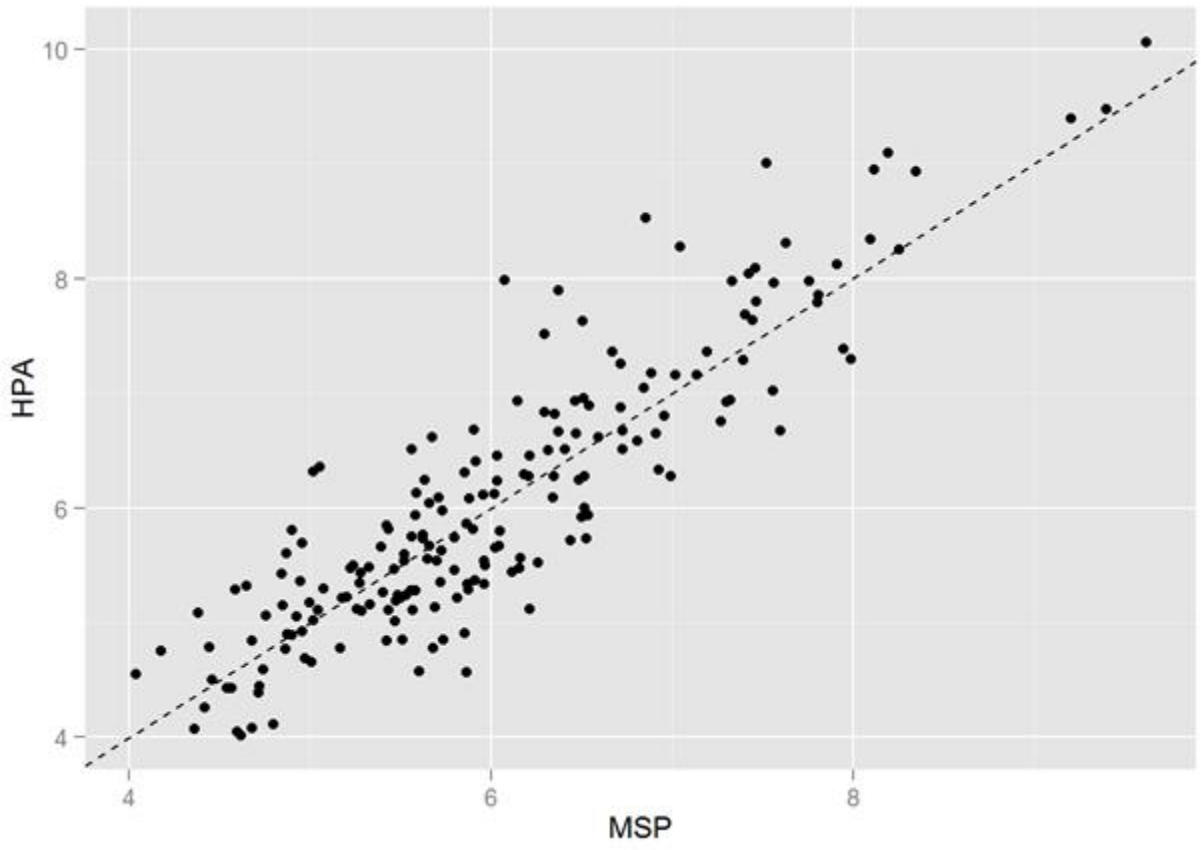

Figure 3: Scatterplot of inter-population MSP/HPA values showing correlation between sequence markers (MSP) and methylation markers (HPA). Significance of correlation was tested using a Mantel test.

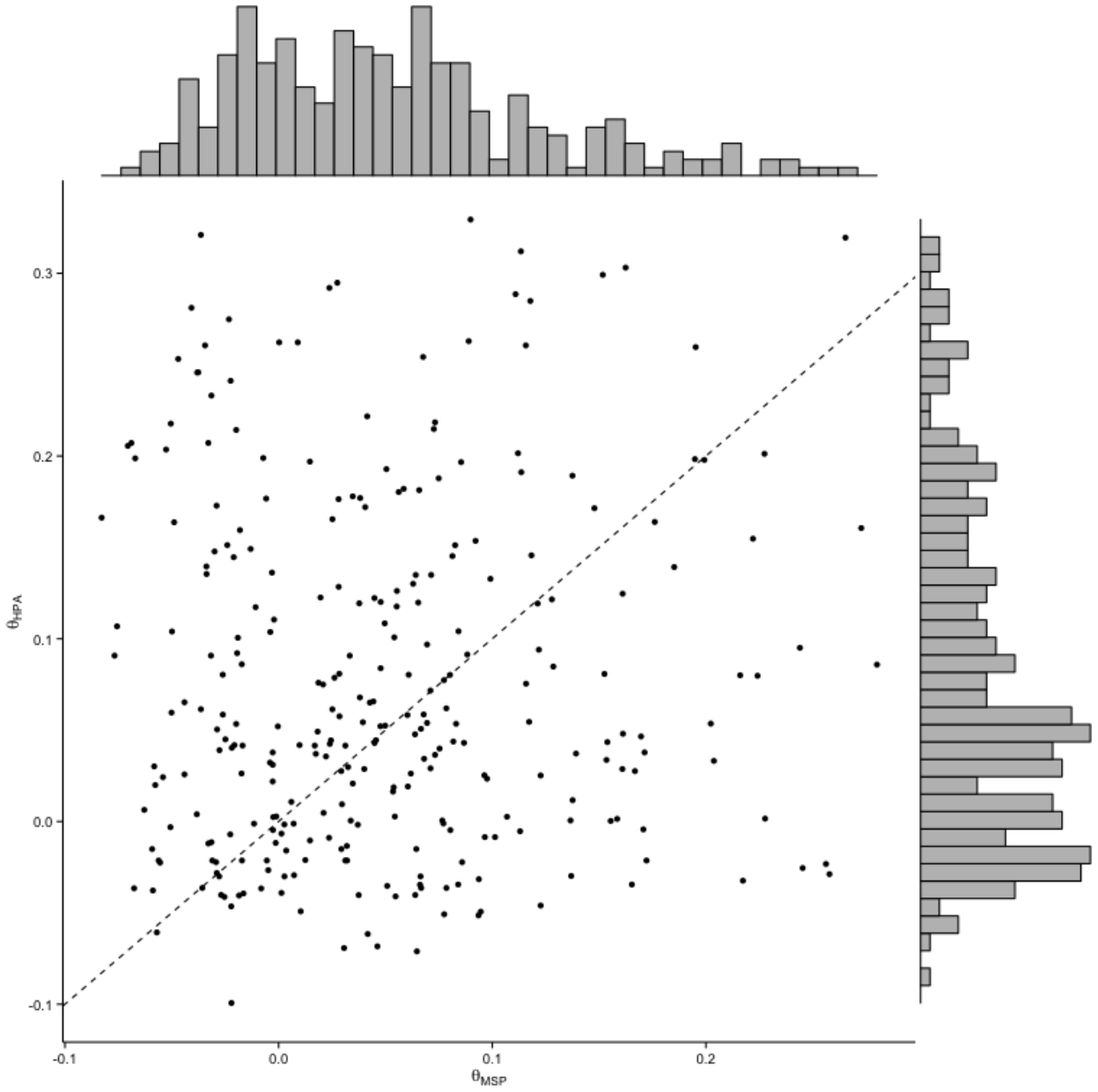

Figure 4: Among population genetic structure assayed via Wier & Cockerham's (1984) for both SEQ and METH genetic markers with marginal distributions for each marker type individually. The dashed line represents where both marker types have the same degree of divergence.

.

Vita

Chitra Seshadri was born in New Delhi (India) on May 11, 1987 and completed her A Levels from University of Cambridge in 2006. She earned a Bachelor of Science in Microbiology from University of Mumbai in 2010, and Master of Science in Microbiology – Part 1, in 2011. While attending college, she interned in the pathology department at King Edward Memorial Hospital and Seth G.S. Medical College. She enrolled in the Center for Environmental Studies at Virginia Commonwealth University in Fall 2012. After graduating, she will continue to pursue her research interest in epigenetic mechanisms, with the aim of elucidating their role in various cellular functions.