MOLECULAR ECOLOGY

Molecular Ecology (2015) doi: 10.1111/mec.13230

FROM THE COVER

Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations

ALEXANDER PLATT,[1] PAUL F. GUGGER, MATTEO PELLEGRINI and VICTORIA L. SORK
Ecology and Evolutionary Biology, University of California, 610 Charles E. Young Dr. E., Los Angeles, CA 90095, USA

Abstract

It has long been known that adaptive evolution can occur through genetic mutations in DNA sequence, but it is unclear whether adaptive evolution can occur through analogous epigenetic mechanisms, such as through DNA methylation. If epigenetic variation contributes directly to evolution, species under threat of disease, invasive competition, climate change or other stresses would have greater stores of variation from which to draw. We looked for evidence of natural selection acting on variably methylated DNA sites using population genomic analysis across three climatologically distinct populations of valley oaks. We found patterns of genetic and epigenetic differentiations that indicate local adaptation is operating on large portions of the oak genome. While CHG methyl polymorphisms are not playing a significant role and would make poor targets for natural selection, our findings suggest that CpG methyl polymorphisms as a whole are involved in local adaptation, either directly or through linkage to regions under selection.

Keywords: adaptation, angiosperms, bioinformatics/phyloinformatics, biotechnology, ecological genetics, natural selection and contemporary evolution

Received 12 December 2014; revision received 22 April 2015; accepted 23 April 2015

Introduction

Local adaptation is the process by which different individuals of the same species evolve towards different phenotypic optima conditional on the environment in which they find themselves. Local adaptation is further illustrative of how evolution works in response to climate change (Davis & Shaw 2001; Aitken *et al.* 2008; Sork *et al.* 2013). The fundamentals of adaptation have been known since Darwin: heritable variations conferring advantageous phenotypes increase in frequency. The actual biological mechanisms of this process, however, are incompletely understood. In particular, little is known about the nature of the heritable variation underlying adaptive traits. Studies of the genetic basis of local adaptation in plant genomes provide insight into the local architecture of adaptation. Frequently, such studies point to a combination of a small number of individual genetic loci contributing moderate fitness differences in combination with a great number of loci with minor effects (Savolainen *et al.* 2007; Wilczek *et al.* 2009; Fournier-Level *et al.* 2011; Hancock *et al.* 2011; Pyhäjärvi *et al.* 2013). Genetic variation, however, is not the only kind of heritable information that can contribute to plants' adaptation. Epigenetic processes may also shape the phenotypes that allow organisms to respond to their local environments (Richards *et al.* 2010). In particular, the presence or absence of methyl groups attached to cytosine nucleotides in DNA is both a partially heritable characteristic (although the degree and mechanism of heritability is not fully known; Jablonka & Raz 2009; Becker *et al.* 2011; Schmitz *et al.* 2011) with demonstrated potential for phenotypic effect (Soppe *et al.* 2000; Cortijo *et al.* 2014) and also contributes to cellular plasticity. Thus, DNA methylation is sufficient to act as the raw material for adaptive evolution (Richards *et al.* 2010; Anderson *et al.* 2011; Zhang *et al.* 2013). Furthermore, in plant species (which do not maintain a rigid distinction between germline and somatic cells), especially long-lived, sessile species producing large numbers of low-dispersal seeds, the ability to acquire

Correspondence: Victoria L. Sork, Fax: +1 (310) 825 9368; E-mail: vlsork@ucla.edu
[1]Present address: Center for Computational Genetics and Genomics, Temple University, 1900 N. 12th St., Philadelphia, PA, 19125, USA

de novo mutations may be especially advantageous (Franks & Hoffmann 2012; Herman *et al.* 2014). DNA methylation variation, with its higher spontaneous mutation rate (i.e. sites changing from methylated to unmethylated states or vice versa) than nucleotide substitution (Jablonka & Raz 2009; Becker *et al.* 2011), could facilitate this process.

We used a population genetic approach to detect evidence of epigenetic variation contributing to local adaptation in natural tree populations. To this end, we sequenced bisulphite-treated reduced-representation 'genomes' (Feng *et al.* 2011) of 11 valley oaks (*Quercus lobata* Née) sampled from three proximal field sites in southern California with divergent elevations, levels of rainfall, and seasonality of rainfall (Fig. 1, Table S1, Supporting information) and intervening mountain ranges limiting among-site gene flow. The reduced-representation bisulphite sequencing generates 100 base pair sequence fragments from randomly selected loci across the genome and converts unmethylated cytosines to thymines. Similar methods have effectively detected variation in DNA methylation in species ranging from model plants (Stroud *et al.* 2014), major crops (Chodavarapu *et al.* 2012), insects (Li-Byarlay *et al.* 2013), and mammalian stem cells (Vincent *et al.* 2013). Thus, this approach can reveal polymorphisms consistent only with nucleotide variation [those observed as A/C, A/G, A/T, C/G or G/T polymorphisms, here referred to as SNPs, and those observed as C/T polymorphisms which represent both polymorphisms that are purely DNA methylation variation as well as sites polymorphic between methylated cytosines and (nonmethylated) thymine]. We estimate that 58% of these polymorphisms are purely methyl polymorphisms, and we hereafter refer to this category collectively as SMPs (for single methylation polymorphism, after Schmitz *et al.* 2011). To the extent that our SMPs contain a proportion of nucleotide polymorphisms, we have introduced a conservative bias that makes statistics related to SMPs more closely resemble those of SNPs. This bias may lead us to underestimate the true extent of epigenetic differentiation among populations.

As DNA methylation in plants is comprised of three separate systems (Law & Jacobsen 2010), it is important

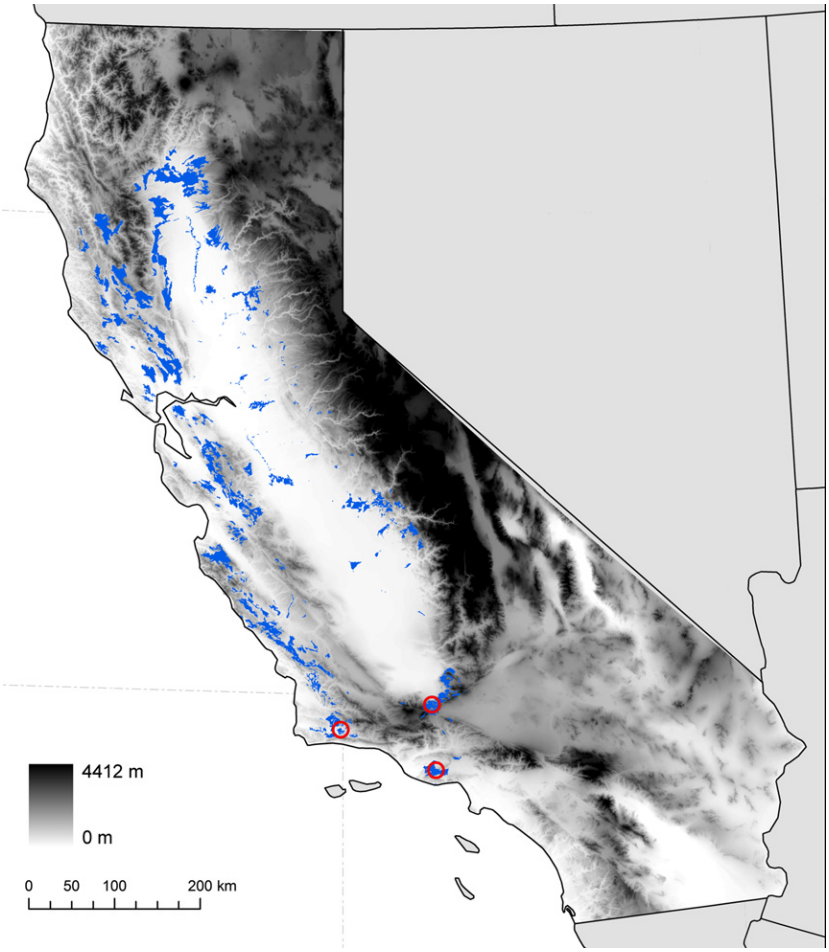

Fig. 1 Geographical range of *Quercus lobata* in blue with population samples indicated with red circles in southern edge.

to further categorize SMPs by their sequence context. First, polymorphisms consistent with CG methylation variation (CG-SMPs) represent methylation variation maintained by a homologue of mammalian DNA METHYLTRANSFERASE 1. In *Arabidopsis thaliana*, this type of methylated site is found within the coding sequences of one-third of genes, though does not result in silencing—rather, these are among the more highly expressed in the genome (Zhang *et al.* 2006; Schmitz *et al.* 2013). Additionally, the 5% of genes in which CG-SMPs occur within promoter regions exhibit elevated levels of tissue-specific expression differentiation (Schmitz *et al.* 2013). While the mechanisms through which CG methylation alters patterns of gene expression are not fully characterized, the fact that they are associated with gene regulation suggests they may serve as a useful tool of adaptive evolution. Second, polymorphisms consistent with CHG methylation variation (CHG-SMPs) represent methylation variation maintained by a plant-specific methyltransferase protein CHROMOMETHYLASE and a feedback loop involving histone 3 lysine 9 methylation (Law & Jacobsen 2010). This histone-directed maintenance is thought to be a more dynamic process than the replication-coupled maintenance of CG methylation. Thus, CHG methylation may be better suited to plastic responses than stably inherited adaptive ones. The third system is represented by sites consistent with CHH methylation and involves de novo methylation by DOMAINS REARRANGED METHYLTRANSFERASE 2. These sites are typically methylated at low levels in *Arabidopsis*, with fewer than 2% of such cytosines typically methylated (Law & Jacobsen 2010).

We looked for two pieces of evidence to establish a role for epigenetics in local adaptation. First, if SMP involvement in local adaptation is a common phenomenon throughout the genome, we expect that SMPs will on average show greater population differentiation than SNPs. This would demonstrate that methylation patterns are being driven to be more population specific than could be explained by neutral demographic processes. This first piece of evidence is analogous to tests of local adaptation showing elevated morphological divergence with respect to genetic divergence (with methylation patterns playing the role of morphology). Second, if adaptation, as opposed to plasticity or tissue specificity, is driving population differences at SMPs, there should be a signature of natural of selection that leads all polymorphisms tightly linked to differentiated SMPs to also show levels of differentiation that are higher than background due to genetic hitchhiking. This signature would demonstrate that the excess SMP differentiation is due to selection on heritable variation at the loci identified by the SMPs and not merely induced by population-specific responses to environmental stimuli or tissue sampling. This type of evidence is typically impossible to generate for a traditional morphology-based test, but critical in understanding the role methylation plays in population divergence and distinguishes evolutionary processes from purely environmental ones.

We calculated population differentiation, F_{ST}, at each of the polymorphic loci as a measure of the extent to which each polymorphism is distributed nonrandomly across field sites. This index of population difference typically varies between 0 and 1.0, but for tree populations, the values are usually in the range of 0.08–0.20 for neutral loci (Loveless & Hamrick 1984; e.g. *Q. lobata* specieswide $F_{ST} = 0.12$ for microsatellite loci; Grivet *et al.* 2008). Two types of factors can drive differentiation between field sites and will produce different distributions of F_{ST} values. Purely neutral processes of demography and drift will drive intersite allele frequency differences on a genome-wide scale. The stochastic nature of these processes will produce heterogeneous levels of differentiation across the genome, but this interlocus variation will be independent of the biological function of the loci. Natural selection can also drive intersite allele frequency differences but does so on a locus-by-locus basis. Only polymorphisms that confer a specific fitness advantage that is conditional on the local environment, and those tightly linked to them, will experience these forces. Additionally, recurrent mutation acts to reduce population differentiation. As recurrent mutation is typically rare, this process can be largely ignored, although it may play a significant role affecting specific highly mutable polymorphisms. The relatively small number of sequenced individuals is appropriate for this study as the extremely large number of markers allows us to estimate parameters from distributions of F_{ST} values across large numbers of loci very precisely even though the estimate at any particular locus is poor.

Most SNPs observed within a species are likely to be selectively neutral. We expect the distribution of SNP F_{ST} values to largely reflect the demographic history of the sample, although individual SNPs contributing to or linked to other variants contributing to local adaptation may contribute to an inflated right hand tail of high F_{ST} values. For a class of polymorphisms to have a distribution of F_{ST} values with an even fatter tail would require those polymorphisms to be further enriched for loci contributing to local adaptation. Much like synonymous variants in tests that contrast synonymous and non-synonymous variation, strict neutrality at SNPs in general is not assumed or required to demonstrate selection acting on SMPs. Statistics drawn from the distribution of SNPs represent the combined effects of demography and any SNP-based adaptive pressures.

Deviations from these statistics when derived from SMPs represent additional SMP-related adaptive pressures. For a class of polymorphisms to have a distribution of F_{ST} values substantially closer to zero would require those polymorphisms to be experiencing a force such as rapid mutation that can eliminate the effects of genetic drift and demography.

Materials and methods

Study species

Quercus lobata Née (section *Quercus*) is a foundational species in oak savannas, oak woodlands and riparian forests across the foothills of the Sierra Nevada, Coast Ranges and Transverse Ranges that surround the Central Valley of California (Fig. 1), spread across diverse climatic and geographical zones from sea level to 1700 m. The genetic structure of valley oak differs significantly across populations throughout its range (Grivet *et al.* 2006, 2008) and is strongly associated with climate gradients (Sork *et al.* 2010; Gugger *et al.* 2013; Cokus *et al.* 2015). Three populations selected for this study were located at the southern end of the current species' range.

Molecular methods

We prepared DNA from expanding leaf/flower bud tissue or mature leaves in 11 individuals from three sample sites (Table S1, Supporting information, Fig. 1) for reduced-representation bisulphite sequencing using a modified protocol based on Feng *et al.* (2011). Briefly, total genomic DNA was digested with *Msp*I and then end-repaired and adenylated with Klenow fragment (3′ to 5′ exo-). Unique Illumina TruSeq adapters were ligated to fragments in each library. Fragments of 200–500 bp were selected with AMPureXP beads. Libraries were treated with sodium bisulphite (EpiTect; Qiagen) to convert unmethylated cytosines to uracil, which are read as thymine during sequencing. Additionally, one sample was prepared without bisulphite treatment. The resulting 12 reduced-representation libraries were amplified by PCR with Illumina primers, pooled and single-end, 100-bp-sequenced in one lane of an Illumina HiSeq 2000, according to standard Illumina protocol.

Bioinformatics

From 66 million distinct 100-bp reads, 110 000 containing nucleotide ambiguities (Ns) were removed, as were 2.2 million single-copy reads. The remaining reads were assembled into 110 033 loci, 23 553 of which were variable and contained 110 279 polymorphisms that were called using the ustacks and cstacks routines of Stacks (Catchen *et al.* 2011). Unlike protocols that quantify the degree of methylation at a site as a continuous variable (Guo *et al.* 2013), this method does not rely on a high-quality reference genome for alignment and, following Schmitz *et al.* (2011, 2013), it treats methylation polymorphism like nucleotide polymorphism with a genome having zero, one or two methylated alleles at a locus. To the extent that degree of methylation can be variable, this is equivalent to discretizing levels of methylation to unmethylated, half methylated and fully methylated. Polymorphisms were further filtered for loci with >10× coverage and observations in at least three plants from more than one field site. Lastly, because rare alleles are more prone to inaccurate estimates of F_{ST} and unlikely to contribute strongly to local adaptation, we used only polymorphisms with minor allele frequency of at least 0.3 for F_{ST} analysis (analyses with other MAF filters produced qualitatively similar results; Table S2, Supporting information). All 7,991 remaining C/T polymorphisms were labelled as SMPs and all 10 388 others as SNPs. This conflates true C/T SNPs with SMPs. In the single sample that was sequenced without bisulphite treating, only 232 886 heterozygous sites were C/T polymorphisms out of 764 613 total heterozygous sites. Assuming the proportion of C/T nucleotide polymorphisms in our sample is in similar proportions, there should be approximately 3400 C/T nucleotide polymorphisms, leaving the SMP category containing approximately 58% methyl polymorphisms. SMPs were further categorized as CG-SMPs (4092) or CHG-SMPs (3638) conditional on sequence context. We found only 261 CHH-SMPs, which were omitted from further analysis. F_{ST} was calculated on a per-polymorphism basis as $(H_T - H_S)/H_T$, where H_T is the total sample heterozygosity and H_S is the average heterozygosity within each field site. F_{ST} values were then averaged across all polymorphisms of the same type within a locus to account for non-independence of linked loci.

Results and Discussion

Across all polymorphisms with a minor allele frequency of at least 0.3, the three populations exhibited a relatively high average F_{ST} of 0.19. Distributions of F_{ST}, however, vary considerably across classes of polymorphisms. At CHG-SMPs, much less population differentiation was exhibited (mean $F_{ST} = 0.08$), but excessive differentiation was found at CG-SMPs (mean $F_{ST} = 0.28$) and intermediate differentiation at SNPs (mean $F_{ST} = 0.18$). In fact, the pattern of distribution of SNP frequencies differed markedly between CHG-SMPs and the other two categories (Fig. 2), with CHG-SMPs

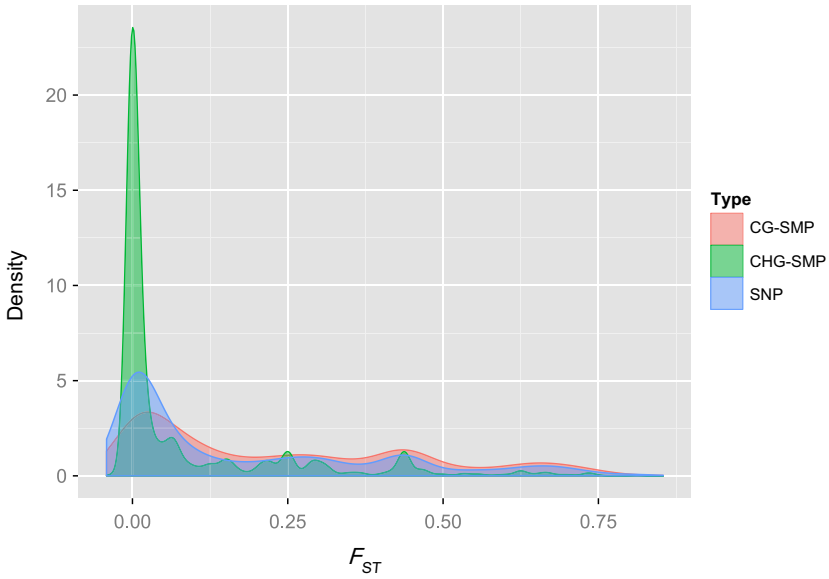

Fig. 2 Density of F_{ST} values by polymorphism type across three subpopulations of *Quercus lobata* in southern California for two types of methylated DNA polymorphisms and single nucleotide polymorphisms. Thicker right hand tails indicate greater population differentiation.

clustering around zero and the other categories showing a much broader range of values up to $F_{ST} = 0.80$.

Even under strict neutrality, F_{ST} values are expected to fluctuate stochastically along the genome. To effectively assess the extent to which the overall distributions of F_{ST} values are shaped by the genomic distribution of different types of polymorphisms as opposed to different adaptive properties, we directly compared F_{ST} values of polymorphisms that occur within the same 100-bp locus. We then classified all 100-bp loci based on which types of polymorphisms they contain (Fig. S1, Supporting information). A comparison of the distributions of F_{ST} values of SNPs and CG-SMPs in specific subsets of loci defined by the varieties of polymorphisms they contain (Fig. 3) reveals that SNPs found in loci without SMPs exhibited less differentiation (mean $F_{ST} = 0.16$) than those in loci containing CG-SMPs. Considering only loci containing both SNPs and CG-SMPs, all polymorphisms exhibited the same elevated levels of differentiation seen in CG-SMPs located in loci without SNPs (mean $F_{ST} = 0.28$). Furthermore, as shown in Fig. S2 (Supporting information), on a locus-by-locus basis, SNP differentiation and CG-SMP differentiation are significantly correlated (Pearson correlation 0.46 ± 0.06).

The SNPs we sequenced can therefore be considered as coming from two classes of loci: those that contain CG-SMPs and those that do not. In the most conservative case, the lower baseline level of differentiation at polymorphisms in SNP-only loci could be ascribed to purely neutral properties of demographic history and genetic drift. These same neutral forces of demography and drift, however, must apply to the entire genome. Therefore, the existence of an entire class of loci, marked by the presence of CG-SMPs and in which both CG-SMPs and SNPs show a level of population differentiation elevated above this baseline, must be experiencing diverging selection.

The local adaptation that CG-SMPs are associated with is not simply the result of a handful of loci experiencing strong diverging selection, dragging CG-SMP markers to different allele frequencies. The entire distribution of F_{ST} values at thousands of loci across the genome is being affected. Without an assembled reference genome, it is impossible to identify exactly where the loci with elevated F_{ST} values are. Ongoing genomic studies show rapid decay of linkage disequilibrium within a few hundred bases (V. L. Sork, unpublished data), which is consistent with the quick decay of disequilibrium observed in other tree species (Brown *et al.* 2004; Krutovsky & Neale 2005). Therefore, even in the worst-case scenario where all sequenced 100-bp loci are contiguously located on a single chromosome, only a tiny fraction could possibly be effectively linked to each other. The vast majority of them must be segregating independently and requiring a separate locus of selection to alter its population-specific allele frequency.

There are four (nonmutually exclusive) scenarios that could generate this pattern: (1) local adaptation is selecting for phenotypes induced by heritable epialleles. In this scenario, the methyl cytosine is acting analogously to a fifth nucleotide, both encoding a phenotype and providing its own template for replication. (2) Local adaptation is selecting for phenotypes induced by methyl states that are encoded for by tightly linked nucleotide variation. In this scenario, cytosine methylation is acting as a locally adaptive phenotype controlled by a nearby polymorphic nucleotide (that may or may

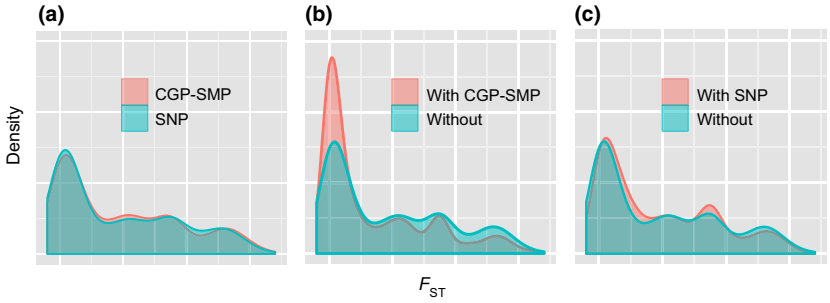

Fig. 3 Density of F_{ST} values in subsets of polymorphism data. Panel (a) compares the distribution of CG-SMP and SNP F_{ST} values at loci containing both, (b) compares SNP F_{ST} values at loci containing CG-SMPs to those at loci that do not, and (c) compares CG-SMP F_{ST} values at loci containing SNPs to those at loci that do not.

not have been sequenced). While biologically distinct from scenario 1, evolutionarily they are equivalent. Without recombination between the methylated base and the nucleotide polymorphism, they will always be co-inherited, and selection and drift will have the same effect on both the genetic allele and epiallele. (3) CpG methylation is clustered in regions of the genome predisposed to harbouring or acquiring the kind of genetic variation on which local adaptation may act. If a significant fraction of genes are all contributing to local adaptation, and CpG methylation levels are highest in gene bodies and promoter regions, CG-SMPs would be overrepresented at diverging loci. In this scenario, the epigenomic architecture of the organism is organized to link methyl variation to the most adaptively fertile regions of the genome. (4) Altered methyl states are biological responses to other genetic phenomena associated with population differentiation. Insertion of transposable elements or other genetic disruptions may both influence phenotype and alter the methylation of linked loci. In this scenario, methylation variation is induced by linked genetic variation and the association between it and the adaptive phenotype is an indirect one.

To the extent that CG-SMPs are themselves the targets of natural selection, epigenetic variation, in addition to genetic variation, must be considered in determining the evolutionary potential of a species. Even populations of isogenic clones may retain some ability for heritable adaptive change, as has been proposed in experimental populations of *Arabidopsis* (Suter & Widmer 2013). Furthermore, in species without adaptive standing variation, the waiting time for a beneficial de novo mutation to occur and give rise to a hard sweep can be reduced if the potential targets of mutation are expanded to include CG methylation sites. To the extent that the pattern of CG-SMPs is driven by genetic hitchhiking, CG-methylated sites must be preferentially located near phenotypically active genetic elements. Any correlation between the phenotypic effect of a mutation in CpG methylation and the phenotypic change induced by the driving genetic variant will create an opportunity for selection to act on future CG-SMPs created by gain or loss of methylation.

In contrast, CHG-SMPs show depressed differentiation regardless of whether they were found alone (mean $F_{ST} = 0.08$), in loci with SNPs (mean $F_{ST} = 0.09$) or in loci with CG-SMPs (mean $F_{ST} = 0.11$). The correlation between F_{ST} measured at SNPs and that measured at CHG-SMPs in loci containing both (Pearson correlation 0.27 ± 0.07) is considerably lower than the correlation of differentiation between SNPs and CG-SMPs. As these are all highly polymorphic loci, purifying selection cannot drive this pattern. Where alleles are being passed from generation to generation in a stable manner, tightly linked polymorphisms will all be drawn from similarly shaped genealogies reflecting shared underlying evolutionary histories. That CHG-SMPs were consistent in their lack of differentiation, regardless of the differentiation reflected in other types of polymorphisms found in the same loci, suggests that these polymorphisms are highly mutable and that two individuals sharing a CHG-SMP allele will often be identical by state but not identical by descent. This further implies that CHG-SMPs are not merely uninvolved in the local adaptation being seen in these plants, but may be largely unavailable for natural selection. This is not to say that CHG-SMPs cannot influence ecologically relevant phenotypes, but rather that selection for particular CHG methylation states will produce little change in the frequencies of those states.

None of these results regarding distributions of F_{ST} values can be predominantly driven by differential environmentally induced epigenetic variation (i.e. environmental plasticity of DNA methylation), variation due to population-specific differences in tissue and sample date (Law & Jacobsen 2010; Vining *et al.* 2012) or selection on a small number of loci acting as master regulators of CpG methylation. While all of

these factors have the potential to inflate F_{ST} at SMPs, none of them can generate an increase in population differentiation of SNPs located near CG-SMPs. All of these phenomena take place within a single organism and within a single generation and thus will not contribute to inherited differentiation. The variation introduced by these effects is therefore independent of genealogy and cannot influence the divergence at other linked polymorphisms. None of them therefore can explain the elevated level of divergence seen at SNPs found at loci containing CG-SMPs. For the pattern of increased population differentiation to be present both at CG-SMPs and at SNPs found at the same locus requires that both types of polymorphism are reflecting the same underlying local genealogy and the same inherited process. Epigenetic variation controlled by environmental conditions, developmental timing and tissue specificity should not reflect any genealogy at all. Epigenetic variation controlled by an unlinked genetic regulator would reflect the genealogy of a different locus than that of the linked nucleotide variation. Natural selection acting on heritable variation at or tightly linked to a great number of CG-SMPs is the only process that will produce elevated F_{ST} values both at CG-SMPs genomewide and at SNPs linked to CG-SMPs.

To fully understand the specific mechanisms through which this epigenetic variation is producing environmentally dependent fitness effects, it is useful to test individual hypotheses about particular phenotypic traits. Nonetheless, purely molecular studies reveal valuable information about the broad genetic and epigenetic architecture of local adaptation. That local adaptation in these oaks is a genomewide phenomenon adds to growing evidence that a predominantly neutral model of genome evolution may be less applicable than previously thought (Lawrie *et al*. 2013; Long *et al*. 2013). This observation is also consistent with an observed architecture of local adaptation where small numbers of loci contribute large adaptive effects but are joined by very large numbers of loci with individually small contributions (Savolainen *et al*. 2007; Wilczek *et al*. 2009; Fournier-Level *et al*. 2011; Hancock *et al*. 2011; Pyhäjärvi *et al*. 2013). Even if small, these broadly distributed loci show distinct signatures of local adaptation in divergent CpG methylation patterns and linked nucleotide polymorphisms. That CHG-SMPs fail to show population divergence even in the face of limited gene flow and ongoing adaptive differentiation indicates that they are not sufficiently stable to play a sizeable role in adaptive evolution. That CG-SMPs are strongly associated with local adaptation in natural populations demonstrates an intimate connection between variable CpG methylation and adaptive evolution.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.

Anderson JT, Willis JH, Mitchell-Olds T (2011) Evolutionary genetics of plant adaptation. *Trends in Genetics*, **27**, 258–266.

Becker C, Hagmann J, Müller J *et al*. (2011) Spontaneous epigenetic variation in the *Arabidopsis thalianaI* methylome. *Nature*, **480**, 245–249.

Brown GR, Gill GP, Kuntz RJ, Langley CH, Neale DB (2004) Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences*, **101**, 15255–15260.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Chodavarapu RK, Feng S, Ding B *et al*. (2012) Transcriptome and methylome interactions in rice hybrids. *Proceedings of the National Academy of Sciences*, **109**, 12040–12045.

Cokus SJ, Gugger PF, Sork VL (2015) Evolutionary insights from *de novo* transcriptome assembly and SNP discovery in California white oaks. *BMC Genomics*, in press.

Cortijo S, Wardenaar R, Colomé-Tatché M *et al*. (2014) Mapping the epigenetic basis of complex traits. *Science*, **343**, 1145–1148.

Davis MB, Shaw RG (2001) Range shifts and adaptive responses to quaternary climate change. *Science*, **292**, 673–679.

Feng S, Rubbi L, Jacobsen SE, Pellegrini M (2011) High-throughput next generation sequencing: methods and applications. In: *Methods in Molecular Biology*, pp. 223–238. Springer, New York.

Fournier-Level A, Korte A, Cooper MD, Nordborg M, Schmitt J, Wilczek AM (2011) A map of local adaptation in *Arabidopsis thaliana*. *Science*, **334**, 86–89.

Franks SJ, Hoffmann AA (2012) Genetics of climate change adaptation. *Annual Review of Genetics*, **46**, 185–208.

Grivet D, Deguilloux M, Petit RJ, Sork VL (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Molecular Ecology*, **15**, 4085–4093.

Grivet D, Sork VL, Westfall RD, Davis FW (2008) Conserving the evolutionary potential of California valley oak (*Quercus lobate* Née): a multivariate genetic approach to conservation planning. *Molecular Ecology*, **17**, 139–156.

Gugger PF, Ikefami M, Sork VL (2013) Influence of late quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology*, **22**, 3598–3612.

Guo W, Fiziev P, Yan W *et al*. (2013) Bs-seeker2: a versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*, **14**, 774.

Hancock AM, Brachi B, Faure N *et al*. (2011) Adaptation to climate across the *Arabidopsis thaliana* genome. *Science*, **334**, 83–86.

Herman JJ, Spencer HG, Donohue K, Sultan SE (2014) How stable 'should' epigenetic modifications be? Insights from adaptive plasticity and bet hedging. *Evolution*, **68**, 632–643.

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

Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics*, **171**, 2029–2041.

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

Lawrie DS, Messer PW, Hershberg R, Petrov DA (2013) Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genetics*, **9**, e1003527.

Li-Byarlay H, Li Y, Stroud H *et al.* (2013) RNA interference knockdown of DNA methyl-transferase 3 affects gene alternative splicing in the honey bee. *Proceedings of the National Academy of Sciences*, **110**, 12750–12755.

Long Q, Rabanal FA, Meng D *et al.* (2013) Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. *Nature Genetics*, **45**, 884–890.

Loveless MD, Hamrick JL (1984) Ecological determinants of genetic structure in plant populations. *Annual Review of Ecology and Systematics*, **15**, 65–95.

Pyhäjärvi T, Hufford MB, Mezmouk S, Ross-Ibarra J (2013) Complex patterns of local adaptation in *teosinte*. *Genome Biology and Evolution*, **5**, 1594–1609.

Richards CL, Bossdorf O, Pigliucci M (2010) What role does heritable epigenetic variation play in phenotypic evolution? *BioScience*, **60**, 232–237.

Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, **38**, 595–619.

Schmitz RJ, Schultz MD, Lewsey MG *et al.* (2011) Transgenerational epigenetic instability is a source of novel methylation variants. *Science*, **334**, 369–373.

Schmitz RJ, Schultz MD, Urich MA *et al.* (2013) Patterns of population epigenomic diversity. *Nature*, **495**, 193–198.

Soppe WJ, Jacobsen SE, Alonso-Blanco C *et al.* (2000) The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Molecular Cell*, **6**, 791–802.

Sork VL, Davis FW, Westfall R *et al.* (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.

Sork VL, Aitken S, Dyer R, Eckert A, Legendre P, Neale DB (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.

Stroud H, Do T, Du J *et al.* (2014) Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nature Structural & Molecular Biology*, **21**, 64–72.

Suter L, Widmer A (2013) Environmental heat and salt stress induce transgenerational phenotypic changes in *Arabidopsis thaliana*. *PLoS One*, **8**, e60364.

Vincent JJ, Huang Y, Chen PY *et al.* (2013) Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell*, **12**, 470–478.

Vining KJ, Pomranaing KR, Wilhelm LG *et al.* (2012) Dynamic DNA cytosine methylation in the *Populus trichocarpa* genome: tissue-level variation and relationship to gene expression. *BMC Genomics*, **13**, 27.

Wilczek AM, Roe JL, Knapp MC *et al.* (2009) Effects of genetic perturbation on seasonal life history plasticity. *Science*, **323**, 930–934.

Zhang YY, Fischer M, Colot V, Bossdorf O (2013) Epigenetic variation creates potential for evolution of plant phenotypic plasticity. *New Phytologist*, **197**, 314–322.

Zhang X, Yazaki J, Sundaresan A *et al.* (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell*, **126**, 1189–1201.

All authors were involved in the design of the study and manuscript preparation. P.F.G. led field and lab work. A.P. led data analysis and writing with contributions from the other authors.

Data accessibility

Sequencing data are available through NCBI BioProject PRJNA269991. They include 100-bp, single-end insert reads and Illumina adapter reads in QSEQ format for twelve multiplexed samples (equimolar) that were run on an Illumina HiSeq 2000 v3 at the Broad Stem Cell Research Center, UCLA, on 04 September 2012. Eleven samples were treated with sodium bisulphite to enable assessment of cytosine methylation. Sequences have been demultiplexed with individual accessions given in Table S1 (Supporting information). One sample (SRR1727825; same individual as SRR1727824) was left untreated as a reference/control. Further documentation of Stacks files and individual polymorphisms is archived at Dryad doi:10.5061/dryad.0f839.

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Violin plots of F_{ST} values for polymorphisms in seven categories of loci: those containing (1) only SNPs (8443 polymorphisms), (2) only CG-SMPs (3145 polymorphisms), (3) only CHG-SMPs (2368 polymorphisms), (4) both SNPs and CG-SMPs (1018 and 661 polymorphisms respectively), (5) both CG-SMPs and CHG-SMPs (160 and 186 polymorphisms respectively), (6) both SNPs and CHG-SMPs (744 and 922 polymorphisms respectively), and (7) SNPs, CG-SMPs, and CHG-SMPs (183, 126, and 162 polymorphisms respectively).

Fig. S2 Average F_{ST} values for SNPs and SMPs in each locus containing SNPs and either CG-SMPs or CHG-SMPs but not both.

Table S1 Sampling information.

Table S2 Mean F_{ST} values (standard error) of contigs stratified by minimum minor allele frequency threshold and contig type.