Clusters of adaptive evolution in the human genome

Laura B. Scheinfeldt1,2, Shameek Biswas1, Jennifer Madeoy1, Caitlin F. Connelly1 and Joshua M. Akey1 *

1 Department of Genome Sciences, University of Washington, Seattle, WA, USA
2 Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA

INTRODUCTION

Interest in identifying regions of the human genome that have been subjected to recent positive selection has grown considerably since the availability of whole genome SNP and sequence data, resulting in large lists of candidate selection genes (Akey, 2009). Very little follow-up, however, has been conducted to explore the patterns of genetic variation at these loci in more detail and in geographically diverse populations. Recently, we described a detailed analysis of the evolutionary history of ALMS1 variation, which has a strong signature of selection from standing variation in European and Asian populations (Scheinfeldt et al., 2009). Here, we focus on a detailed analysis of a 3-Mb region encompassing ALMS1 that possesses patterns of variation consistent with the action of three independent selective events in human history.

In addition, we also evaluated whether the chromosome 2 cluster of positive selection was unique or if there were additional clusters of selection in the human genome. Our analysis of SNP data from the HapMap Phase II samples (International HapMap Consortium, 2005; Sabeti et al., 2007) indicates that there are indeed additional clusters of selective, and that these regions are unlikely to have arisen under a model of neutral evolution. Furthermore, several of the clusters we identified contain previously known candidate genes for selection; however, these regions have been interpreted as a single signature of selection across linked loci and possible independent selective events were not considered. Our work suggests that signatures of selection identified in genome-wide scans of selection are more complex than previously assumed, and a subset are comprised of multiple and independent selective targets. Thus, follow-up studies of genes and regions identified in genome-wide scans for positive selection are critical to foster a deeper understanding of the mechanistic basis of recent human evolutionary history.

MATERIALS AND METHODS

SAMPLES

We sequenced approximately 6 kb of ALMS1, approximately 4 kb of GC31, and approximately 2 kb of sequence in the regions between ALMS1 and SEC15L2 and between ALMS1 and GC31 in DNA samples from 91 individuals representing 6 human populations and 4 non-human primates that were obtained from the Coriell Institute for Medical Research Cell Repositories (Camden, NJ, USA). Coriell repository numbers for these samples are as follows: CEPH (n = 21: NA006990, NA007019, NA07348-9, NA10830-1, NA10842-5, NA10848, NA10850-4, NA10857-8, NA10860-1, NA17201), Han Chinese of L.A. (n = 21: NA17733–18749, NA17752–56), Middle East (n = 10: NA17041–50), Pygmy (n = 10: NA10469–73, NA10492–96), South Africa (n = 9: NA17341–49), South America (n = 10: NA17301–10), and South East Asia (n = 10: NA17081–90), gorilla (Gorilla gorilla; AG05251), bonobo (Pan paniscus; AG05253), chimpanzee (Pan troglodytes; AG06939), and orangutan (Pongo pygmaeus; AG12256). We acquired SEC15L2 sequence from Seattle SNPs. Genotype data from 210 unrelated individuals were obtained from the HapMap project (Release 22 NCBI Build 36; International HapMap Consortium, 2005), and genotype data from 947 unrelated individuals were obtained from the Human Genome Diversity Project–Centre.

Cluster of adaptive evolution in the human genome

Laura B. Scheinfeldt1,2, Shameek Biswas1, Jennifer Madeoy1, Caitlin F. Connelly1 and Joshua M. Akey1*

1 Department of Genome Sciences, University of Washington, Seattle, WA, USA
2 Department of Genetics, University of Pennsylvania, Philadelphia, PA, USA

*Correspondence:
Joshua M. Akey, Department of Genome Sciences, University of Washington, Foege Building, S303, Seattle, WA 98195-7730, USA.
e-mail: akeyj@u.washington.edu

http://pga.gs.washington.edu/
We designed sequencing primers from published human sequence (NM_015120) with primer3 for coding and non-coding regions of ALMS1 and GCS1 (primer sequences are available upon request). We used standard PCR-based sequencing reactions using Applied Biosystem's Big Dye sequencing protocol on an ABI 3130×1. Sequence data was assembled using Phred/Phrap (Ewing and Green, 1998; Ewing et al., 1998), and the alignments were inspected for accuracy with Consed (Gordon et al., 1998, 2001). Polymorphisms were identified with PolyPhred 4.0 (Bhangale and Green, 1998; Ewing et al., 1998), and the alignments were only accepted simulations in which the population specific FST values in the chromosome 2p13 region are 0.74, 0.68, and 0.92 in the CEU, ASN, and YRI samples, respectively. The lower stringency thresholds were chosen for computational efficiency (i.e., to increase the number of accepted replicates) and to be applicable to all of the regions described in Table 2, of which had slightly lower maximum population specific FST values compared to the chromosome 2p13 region. The ms command line argument for the model used in these simulations is:

```ms 122 1 -t 2000 -r 1000 5000000 -c 1.6 500 -l 3 38 42 42 -en 0.0005 1 0.24 -en 0.000875 2 0.077 -en 0.001 3 0.077 -en 0.001 2 0.0125 -en 0.001125 2 0.077 -en 0.005 3 2 -en 0.00475 3 0.00373 -en 0.004875 3 0.077 -en 0.0085 2 0.00294 -en 0.008625 2 0.077 -en 0.00875 2 1 -en 0.0075 1 0.03125 -en 0.007625 1 0.24 -en 0.0425 1 0.12.
```

**RESULTS**

We used SNP data from the HGDP–CEPH samples (Li et al., 2008) and CONTML from the Phylip package (Felsenstein, 1989, 2005) to construct phylogenies rooted with chimpanzee data (UCSC). We coded each SNP within the three genes (SEC15L2, ALMS1, GCS1) as allele frequencies within each continental group and used the default (gene frequency) mode of CONTML to construct phylogenetic trees.
Scheinfeldt et al. Clustered adaptive events

events through genome-wide analyses of the HapMap Phase II data.

PATTERNS OF $F_{ST}$ AND HETEROZYGOSITY REVEAL THREE DISTINCT REGIONS AT 2P13.3–2P13.1

Our previous analysis of population structure at ALMS1 revealed extreme levels of $F_{ST}$ between African and non-African HapMap samples (Scheinfeldt et al., 2009). Here, we expand this analysis to include a 3-Mb region encompassing ALMS1. We analyzed SNP data from HapMap Phase II data among the following HapMap samples: Yoruba (YRI) individuals from Ibadan, Nigeria (n = 60), CEPH (CEU) individuals with ancestry from northern and western Europe (n = 60), Japanese (JPT) individuals from Tokyo, Japan (n = 45), and Han Chinese (CHB) individuals from Beijing, China (n = 45). In all of the analyses, we combined the JPT and CHB individuals into a single Asian sample (ASN). As displayed in Figure 1, there are three peaks of high $F_{ST}$ in the region. The first peak (which encompasses SEC15L2) differentiates African and non-African HapMap samples, and displays extremely low heterozygosity in the ASN samples, consistent with a classic selective sweep. The second peak (which encompasses ALMS1) also differentiates African and non-African HapMap samples; however, there is only a modest decrease in heterozygosity, consistent with a model of selection acting on standing variation. And lastly, the third peak (which contains 19 refseq genes centered around GCS1) differentiates the CEU samples. Each of the peaks is separated by recombination hotspots suggesting individual evolutionary histories for each of the three peaks. For ease of presentation, we will refer to each of these peaks as region 1, 2, and 3 for the SEC15L2, ALMS1, and GCS1 peaks respectively below.

NEUTRALITY TESTS SUPPORT A MODEL OF THREE INDEPENDENT SELECTIVE EVENTS

Using sequence data from regions 1, 2, and 3 we performed three tests of positive selection on the genes central to each region (SEC15L2, ALMS1, GCS1): Tajima’s $D$ (Tajima, 1989), Fu and Li’s $D$ (Fu and Li, 1993), and Fay and Wu’s $H$ (Fay and Wu, 2000; Table 1). Standard site frequency spectrum statistics support a model of positive selection for SEC15L2 in the Asian American Seattle SNPs samples (Tajima’s $D$, Fu and Li’s $D$, and Fay and Wu’s $H$ tests, $p < 0.008$). Similarly, standard site frequency spectrum statistics support a model of positive selection at GCS1 in the

---

**FIGURE 1 | Patterns of population specific $F_{ST}$ and heterozygosity at chromosome 2 (p13.3–p13.1).** The location of each gene (going from left to right: SEC15L2, ALMS, GCS1) is marked by a gray rectangle and black arrow, and additional genes located in the region are shown as white rectangles. Previously inferred recombination hotspots are denoted by black rectangles. For each HapMap sample, population specific $F_{ST}$ and heterozygosity are plotted individually in orange and blue respectively.
Table 1 | Summary statistics of neutrality test statistics.

| Gene    | Sample              | $n^a$ | $S^b$ | $\mu^c$ | $\pi^d$ | $\theta_w^e$ | Tajima’s $D$ | Fu and Li’s $D$ | Fay and Wu’s $H$ |
|---------|---------------------|-------|-------|---------|---------|-------------|-------------|-----------------|-----------------|
| SEC15L2 | African American    | 42    | 99    | 35      | 3.76    | 4.97        | −0.86       | −0.99           | −5.36           |
|         | European American   | 42    | 47    | 9       | 2.49    | 2.36        | 0.23        | 0.37            | −10.31          |
|         | Hispanic American   | 40    | 52    | 13      | 2.06    | 2.64        | −0.78       | −0.06           | −11.50          |
|         | Asian American      | 44    | 27    | 26      | 0.30    | 1.34        | −2.62       | −5.49           | −12.95          |
| ALMS1   | CEPH                | 40    | 19    | 3       | 10.39   | 7.39        | 1.34        | 0.57            | −2              |
|         | Han                 | 42    | 9     | 3       | 3.32    | 3.46        | −0.12       | −0.54           | −0.96           |
|         | Middle East         | 20    | 15    | 2       | 9.41    | 6.99        | 1.28        | 0.84            | −2.53           |
|         | Pygmy               | 20    | 18    | 6       | 7.15    | 8.39        | −0.55       | −0.19           | −1.24           |
|         | South African       | 18    | 23    | 8       | 10.91   | 11.06       | −0.05       | −0.21           | 2.4             |
|         | South American      | 20    | 19    | 2       | 12.88   | 8.86        | 1.72        | 1.01            | 2.78            |
|         | Southeast Asian     | 20    | 7     | 3       | 2.03    | 3.26        | −1.24       | 0.05            | −0.39           |
| GCS1    | CEPH                | 40    | 4     | 1       | 1.36    | 2.32        | −1.03       | −2.12           | −1.88           |
|         | Han                 | 42    | 3     | 1       | 1.18    | 1.72        | −0.68       | −0.36           | −1.14           |
|         | Middle East         | 20    | 5     | 2       | 1.71    | 3.48        | −1.55       | −2.01           | −1.35           |
|         | Pygmy               | 20    | 7     | 1       | 3.23    | 4.87        | −1.16       | −1.16           | −1.63           |
|         | South African       | 18    | 4     | 3       | 1.91    | 2.87        | −1.05       | 0.23            | −1.17           |
|         | South American      | 20    | 4     | 0       | 2.89    | 2.78        | 0.11        | −0.76           | −0.05           |
|         | Southeast Asian     | 20    | 6     | 0       | 2.78    | 4.18        | −1.11       | −1.51           | −0.55           |

*aNumber of chromosomes.

*bNumber of segregating sites.

*cNumber of singletons.

*dNucleotide diversity per base pair × 10^{−4}.

*e$\theta_w$ per base pair × 10^{−4}.

*p < 0.05 in simulations with recombination and conditional on the number of segregating sites.

†p < 0.05 in simulations using previously inferred demographic parameters (Schaffner et al., 2005).

Values that remain significant after Bonferroni correction are highlighted in bold.

CEPH (Fu and Li’s $D$, and Fay and Wu’s $H$ tests, p < 0.05) and to a lesser extent in the Middle Eastern samples (Fu and Li’s $D$ test, p < 0.05). While previous work demonstrates no deviation from neutral expectations at ALMS1, additional analyses support a model of positive selection from standing variation on ALMS1 (Scheinfeldt et al., 2009). Furthermore, analysis of the sequence located between regions 1 and 2 as well as the sequence located between regions 2 and 3 show no significant deviations from neutral expectations.

**DISTINCT PATTERNS OF WORLDWIDE VARIATION AT EACH PEAK**

The geographic distribution of genetic variation across the $SEC15L2$, ALMS1, and GCS1 regions shows considerable heterogeneity. As shown in Figures 2–4, the East Asian samples show the most dramatic changes in $SEC15L2$ and ALMS1 derived allele frequencies compared with other non-African samples. However, as we previously noted (Scheinfeldt et al., 2009), the geographic pattern of variation for ALMS1 in the American samples is peculiar and consistent with recent selection in East Asia roughly 15 kya, while the pattern of variation at $SEC15L2$ is more consistent with an older time of selection as both the American and Asian samples demonstrate high derived allele frequencies. The worldwide pattern of allele frequency variation at GCS1 is more difficult to reconcile with a simple model of selection in European samples, but is clearly distinct from the pattern at $SEC15L2$ and ALMS1. To better quantify patterns of variation shown in the allele frequency maps, we performed a phylogenetic analysis of HGDP allele frequency data. Specifically, we used CONTML (Felsenstein, 1989, 2005) to construct phylogenies for each of the three genes using SNP data from the HGDP–CEPH samples (Li et al., 2008). As shown in Figure 5, the continental groups cluster differently in each phylogeny. The $SEC15L2$ tree displays East Asia and America clustering together at the farthest distance from the chimpanzee outgroup. The ALMS1 tree shows East Asia and Oceania clustering together at the farthest distance from the chimpanzee outgroup. Finally, the GCS1 tree shows Europe and the Middle East clustering together at the farthest distance from the chimpanzee outgroup.

**GENOME-WIDE SCAN IDENTIFIES ADDITIONAL CLUSTERS OF POSITIVE SELECTION**

To next tested whether the patterns of genetic variation at 2p13.3–2p13.1 were unique or if other regions of the genome exhibited similar evidence for clustering of independent selective events. Using the population specific $F_{ST}$ thresholds (98.7th percentile) of the 2p13.3–2p13.1 region, we asked how many other 5 Mb regions of the HapMap Phase II data (International HapMap...
Consortium, 2005) possess highly differentiated population specific $F_{ST}$ values for all three samples. Our scan (Table 2) identified 19 additional regions that met these criteria, suggesting that additional clusters of independent substrates of positive selection exist in the human genome. As expected, gene density is significantly higher ($p = 0.024$, Mann–Whitney test) in windows that exhibit evidence of independent signals of selection, which likely reflects the greater mutational target size of gene dense windows for selection to act on. Moreover, we tested whether the recombination rate was different between windows with and without evidence of clustered signals of selection and found no significant difference ($p = 0.338$; Mann–Whitney test). This result is consistent with the observation that although there is considerable heterogeneity of fine-scale recombination rates in humans, rates over Mb intervals are much more uniform (Meyers et al., 2005).
FIGURE 4 | Distribution of GCS1 alleles in 52 populations. Haplogroup frequencies are indicated with pie charts. The ancestral allele is shown in white, and the derived allele is shown in black.

FIGURE 5 | Phylogenetic relationships among seven continental groups. Maximum likelihood trees for each of the three genes are shown rooted with chimpanzee as the outgroup. The length of the scale bar corresponds to 1% of sites differing between haplotypes.
Table 2 | Summary of additional candidate regions for adaptive hotspots.

| Chr | Start (Mb) | Stop (Mb) | Notable candidates for positive selection in region |
|-----|------------|----------|--------------------------------------------------|
| 1   | 35         | 40       | 1,2,3,4,5                                         |
| 1   | 50         | 55       | 1,2,4,6                                          |
| 1   | 170        | 175      | 1,2,3,4                                          |
| 2   | 70         | 75       | ALMS1,1,2,4                                       |
| 2   | 95         | 100      | ZAP70²                                            |
| 2   | 105        | 110      | EDAR¹,5, SULT1C²                                  |
| 3   | 190        | 195      | 1,2,4,6                                          |
| 5   | 30         | 35       | SLC45A2¹,5, MATP²                                 |
| 6   | 125        | 130      | 1,2,3,4,6                                        |
| 7   | 105        | 110      | 1,2,3,4                                          |
| 11  | 60         | 65       | VPS37C²                                          |
| 14  | 55         | 60       | RTN²                                             |
| 15  | 40         | 45       | 1,2,3,6                                          |
| 15  | 60         | 65       | HERC⁶                                            |
| 17  | 50         | 55       | RAD51C²                                          |
| 17  | 60         | 65       | 1,2,3,4,6                                        |
| 20  | 20         | 25       | XRN2²                                            |

1 Identified in Kimura et al. (2007). PLoS ONE 14, e2688.
2 Identified in Wang et al. (2008). PNAS 103, 135–140.
3 Identified in International HapMap Consortium (2005). Nature 449, 851–861.
4 Identified in Wang et al. (2006). PNAS 103, 4786–4791.
5 Identified in Enard et al. (2002). Nature 418, 869–872.
6 Identified in Zhang et al. (2002). Genetics 162, 1825–1835.

To more rigorously evaluate the evidence that clustering of population specific \( F_{ST} \) in each HapMap sample is unusual under neutrality, we performed additional coalescent simulations (using the calibrated model of human demography from Schaffner et al., 2005) that takes into account the way in which these regions were ascertained. Specifically, we initially identified the chromosome 2p13.3–2p13.1 region by observing a high population specific \( F_{ST} \) value in the ASN sample and then asked if population specific \( F_{ST} \) values in the other two HapMap samples were unusually large. Thus, to recapitulate this process we used a rejection sampling algorithm to generate simulated regions that were 5 Mb in length (with recombination), and only accepted regions where one sample possessed a large population specific \( F_{ST} \) (see Materials and Methods). Next, we estimated the proportion of accepted replicates that had large population specific \( F_{ST} \) values in all samples. In practice, we used thresholds that were less stringent than that observed in the empirical data (see Materials and Methods) for computational efficiency. Out of 2 × 10⁴ simulations, 5,846 were accepted and even at this reduced level of stringency none exhibited large population specific \( F_{ST} \) values in all three samples, resulting in a conservative \( p \)-value of <0.0002. Thus, the observation of finding clusters of highly differentiated population specific \( F_{ST} \) values in each sample is very unusual under neutrality.

**DISCUSSION**

What emerges from this analysis is a striking incidence of multiple, independent, and regionally restricted signals of positive selection in a 3-MB region on chromosome 2. Interestingly, we also identified 19 additional regions that possess similar patterns of genetic variation (Table 2) and thus may represent additional clusters of independent selective events. Included in this list are regions containing EDAR (see also Figure 6), SLC45A2, and FOXP2, all previously reported as strong candidates for recent positive selection (Enard et al., 2002; Zhang et al., 2002; Carlson et al., 2005; Kelley et al., 2006; Kimura et al., 2007; Sabeti et al., 2007). These previous analyses presented the candidates as single signals of positive selection; however, our analysis suggests that there were multiple events contributing to the signals identified through genome-wide scans for selection. For example, as displayed in Figure 6, all three HapMap samples display peaks in \( F_{ST} \) that are coincident with EDAR; however, the ASN and YRI samples each exhibit additional peaks upstream and downstream of EDAR, and these signals are separated by recombination hotspots. Thus, while previous discussion of the region has focused on EDAR (Kimura et al., 2007; Sabeti et al., 2007), our analysis indicates that this 5 Mb region contains additional substrates of positive selection.

It is interesting to consider why adaptive genetic variation might be clustered in some regions of the human genome. One hypothesis is that these regions simply possessed multiple adaptive mutations that selection was free to independently act on because of the local recombinational landscape. This idea is consistent with the observation that gene density is significantly higher in the 20 regions shown in Table 2, and for at least two of these regions (Figures 1 and 6) recombination hotspots occur between the three distinct patterns of population differentiation.

A second, non-mutually exclusive hypothesis is that clusters of independent adaptive alleles could be responding to the same selective pressure. To explore this idea, we investigated biological relationships among the genes in the 2p13.3–2p13.1 region. Interestingly, SNPs from both ALMS1 (rs7598660) and GCS1 (rs6758593) have previously been implicated in human association studies of insulin levels (Saxena et al., 2007; Scheinfeldt et al., 2009) and Type I diabetes (Wellcome Trust Case Control Consortium, 2007), respectively. Additionally, in mice SEC15L2 is present in a module of genes with strain-specific gene expression indicative of a role in liver metabolism (Keller et al., 2008). It is intriguing that all three loci have putative roles in metabolic phenotypes, and it is possible that a single selective pressure underlies the independent response to selection observed at these three loci; however, additional work is necessary to elucidate the exact function of these proteins and characterize the ways in which functional variation in the region affects phenotypic variation.

Moreover, it is important to note that even with a uniformly distributed advantageous mutation rate across the genome, clustering of independent selective events may occur depending on many parameters such as the time selective alleles arose, the mode of selection, and the timeframe over which a signature of selection persists. In this manuscript, we simply focused on how unusual
clusters of independent selective events are under neutrality. In the future, it would also be of interest to evaluate how often clustering occurs in models incorporating selection, and what particular parameter values lead to clustering of independent selective events.

Our data clearly supports a model of non-neutral evolution at the chromosome 2p13 locus, as well as the additional 19 regions that exhibit patterns of variation similar to or more extreme than this region. However, some caution is warranted in the interpretation of multiple independent selective events because the dynamics of selection acting in the milieu of a complex demographic process could conceivably generate unexpected and difficult to predict patterns of genetic variation within and between populations. Although additional theoretical and simulation studies on the interaction of selection and demography over a range of selective models and demographic processes is important, the simplest explanation for the data presented in this paper is independent selective events in these 20 regions. Indeed, the observation of recombination hotspots coincident with changes in patterns of population differentiation (see Figures 1 and 6) and the incompatibility of highly differentiation clusters of population specific $F_{ST}$ values in each HapMap samples under neutrality strongly suggests multiple and independent selective events.

In summary, while many recent scans of positive selection have resulted in extensive lists of candidate regions (Kelley et al., 2006; Voight et al., 2006; Wang et al., 2006; Zhang et al., 2006; Kimura et al., 2007; Sabeti et al., 2007; Tang et al., 2007), very little follow-up analysis has been reported. Here, we have focused on a region of chromosome 2p13 that contains three independent substrates of recent positive selection, and we have shown that additional clusters of independent selective events likely exist in the human genome. Our results demonstrate the importance of careful follow-up work to genome-wide scans for selection and offers a novel perspective on the organization of adaptive genetic variation in humans.
ACKNOWLEDGMENTS

We thank members of the Joshua M. Akey lab for helpful discussions and comments on the manuscript. Funding: This work was supported by a research grant [1R01GM076036-01A1] from the National Institute of Health to Jennifer M adeoye; a Sloan Fellowship in Computational Biology to Jennifer M adeoye; and by a National Human Genome Research Institute Interdisciplinary Training in Genomics Sciences grant [HG00035] to Laura B. Scheinfeld.
Scheinfeldt, L. B., Biswas, S., Madeoy, J., Connelly, C. F., Schadt, E. E., and Akey, J. M. (2009). Population genomics analysis of ALMS1 in humans reveals a surprisingly complex evolutionary history. *Mol. Biol. Evol.* 26, 1357–1367.

Shriver, M. D., Kennedy, G. C., Parra, E. J., Lawson, H. A., Sonpar, V., Huang, J., Akey, J. M., and Jones, K. W. (2004). The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. *Hum. Genomics* 1, 274–286.

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

Wang, E. T., Kodama, G., Baldi, P., and Moyzis, R. K. (2006). Global landscape of recent inferred Darwinian selection for Homo sapiens. *Proc. Natl. Acad. Sci. U.S.A.* 103, 135–140.

Wellcome Trust Case Control Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678.

Zhang, J., Webb, D. M., and Pollaha, O. (2002). Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics* 162, 1825–1835.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 May 2011; accepted: 19 July 2011; published online: 09 September 2011.

Citation: Scheinfeldt LB, Biswas S, Madeoy J, Connelly CF and Akey JM (2011) Clusters of adaptive evolution in the human genome. *Front. Genet.* 2:50. doi: 10.3389/fgene.2011.00050

This article was submitted to Frontiers in Evolutionary and Population Genetics, a specialty of Frontiers in Genetics. Copyright © 2011 Scheinfeldt, Biswas, Madeoy, Connelly and Akey. This is an open-access article subject to a non-exclusive license between the authors and Frontiers Media SA, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and other Frontiers conditions are complied with.