DOI: 10.1534/genetics.103.025726

Population Genetic and Phylogenetic Evidence for Positive Selection on Regulatory Mutations at the *Factor VII* Locus in Humans

Matthew W. Hahn,*[,1] Matthew V. Rockman,* Nicole Soranzo,[†] David B. Goldstein[†] and Gregory A. Wray*

**Department of Biology, Duke University, Durham, North Carolina 27708 and [†]Department of Biology, University College London, London WC1E 6BT, United Kingdom*

Manuscript received December 11, 2003
Accepted for publication February 8, 2004

ABSTRACT

The abundance of *cis*-regulatory polymorphisms in humans suggests that many may have been important in human evolution, but evidence for their role is relatively rare. Four common polymorphisms in the 5′ promoter region of *factor VII* (*F7*), a coagulation factor, have been shown to affect its transcription and protein abundance both *in vitro* and *in vivo*. Three of these polymorphisms have low-frequency alleles that decrease expression of *F7* and may provide protection against myocardial infarction (heart attacks). The fourth polymorphism has a minor allele that increases the level of transcription. To look for evidence of natural selection on the *cis*-regulatory variants flanking *F7*, we genotyped three of the polymorphisms in six Old World populations for which we also have data from a group of putatively neutral SNPs. Our population genetic analysis shows evidence for selection within humans; surprisingly, the strongest evidence is due to a large increase in frequency of the high-expression variant in Singaporean Chinese. Further characterization of a Japanese population shows that at least part of the increase in frequency of the high-expression allele is found in other East Asian populations. In addition, to examine interspecific patterns of selection we sequenced the homologous 5′ noncoding region in chimpanzees, bonobos, a gorilla, an orangutan, and a baboon. Analysis of these data reveals an excess of fixed differences within transcription factor binding sites along the human lineage. Our results thus further support the hypothesis that regulatory mutations have been important in human evolution.

CHANGE in transcriptional regulation is a crucial contributor to phenotypic evolution (Raff and Kaufman 1983; Gerhart and Kirschner 1997; Carroll *et al.* 2001; Davidson 2001; Wray *et al.* 2003). Genetic changes in *cis*-regulatory sequences controlling transcription can be as important to gene function as any change in coding sequences; indeed, it has been hypothesized that the changes distinguishing humans from our closest relatives reside in *cis*-regulatory regions (King and Wilson 1975; Wilson 1975). The ubiquity of *cis*-regulatory polymorphisms that have significant phenotypic effects in humans (Rockman and Wray 2002) suggests that much current phenotypic variation, as well as past evolutionary change, may be due to regulatory mutations. Whether the evolution of regulatory mutations in humans has been the result of genetic drift or adaptive natural selection, however, is known in only a few cases (Hamblin and Di Rienzo 2000; Bamshad *et al.* 2002; Sabeti *et al.* 2002; Rockman *et al.* 2003). Here we examine the patterns of variation and change over time in the well-characterized *cis*-regulatory region at the *factor VII* (*F7*) locus to tease apart the action of multiple evolutionary forces.

Factor VII is a vitamin K-dependent protease essential for coagulation. When blood vessel injury occurs, tissue factor is released and binds to circulating F7, allowing proteolysis of F7 to its active form and the initiation of blood coagulation through the production of thrombin and fibrin clots. The cause of most myocardial infarctions is thrombosis, the presence of a blood clot within a blood vessel (Fuster *et al.* 1992). Because of this, F7 was long suspected of playing a role in death due to coronary heart disease and myocardial infarction; multiple studies have now shown that high levels of plasma F7 are a predictor of death due to heart disease (*e.g.*, Meade *et al.* 1986; Heinrich *et al.* 1994; Redondo *et al.* 1999; Girelli *et al.* 2000). Genetic studies of the *F7* locus have found five common polymorphisms that explain at least 40% of the variance in levels of both active and inactive plasma F7 among healthy adults (Marchetti *et al.* 1993; Bernardi *et al.* 1996; Humphries *et al.* 1996; Pollak *et al.* 1996; van 't Hooft *et al.* 1999). Of these, four are in the 5′ *cis*-regulatory region and one is an amino acid variant. The four regulatory variants are associated with transcript and protein abundance of

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY493422–AY493433.

[1]*Corresponding author:* Center for Population Biology, 2320 Storer Hall, University of California, Davis, CA 95616.
E-mail: mwhahn@ucdavis.edu

F7 in vivo, but because of linkage disequilibrium and epistasis the magnitude of the effect of each is not clear (Bernardi *et al.* 1996; Pollak *et al.* 1996; van 't Hooft *et al.* 1999; Di Castelnuovo *et al.* 2000; Girelli *et al.* 2000; Kudaravalli *et al.* 2002).

Three of the regulatory polymorphisms have minor alleles (defined by researchers on the basis of the relative frequencies in Europe) that individually decrease expression of *F7*: a single nucleotide polymorphism (SNP) at −401 (G/T) relative to the start of translation (based on the nomenclature of Pollak *et al.* 1996), an SNP at −122 (T/C), and a 10-bp insertion/deletion polymorphism at −324 (Bernardi *et al.* 1996; Pollak *et al.* 1996; van 't Hooft *et al.* 1999; Di Castelnuovo *et al.* 2000; Girelli *et al.* 2000). The indel polymorphism (dbSNP identifier rs5742910) has generally been ascribed to position −323. However, examination of the 5′ sequences generated by the Human Genome Project and by Pollak and colleagues (GenBank accession nos. NT_027140 and U40852, respectively) unambiguously puts the insertion allele between nucleotides −324 and −325. We therefore refer to this polymorphism as −324. The less-frequent allele of the common amino acid polymorphism at position 353 (R/Q) of the protein has also been shown to correlate with lower levels of plasma F7, but it is in strong linkage disequilibrium with the polymorphisms at −401, −324, and −122 (Bernardi *et al.* 1996; Humphries *et al.* 1996; Girelli *et al.* 2000). Lower levels of F7 may provide protection against myocardial infarction, with the frequencies of the low-expression alleles being found disproportionately in healthy individuals (Green *et al.* 1991; Iacoviello *et al.* 1998; Di Castelnuovo *et al.* 2000; Girelli *et al.* 2000). The fourth regulatory polymorphism, an SNP at −402 (G/A; dbSNP identifier rs762637), has a minor allele that leads to an increase in the level of transcription of *F7* (van 't Hooft *et al.* 1999). The alleles at sites −401 and −402 affect transcription through changes in transcription factor binding (van 't Hooft *et al.* 1999).

The association between allele frequencies and F7 activity has previously been examined in populations from Britain, Japan, India, Norway, Holland, France, Italy, and Greenland (Lane *et al.* 1992; Kario *et al.* 1995; Bernardi *et al.* 1997; de Maat *et al.* 1997). Populations differ significantly in the frequencies of the regulatory variants, even within Europe. Strong linkage disequilibrium among the regulatory indel (−324), a regulatory SNP (−401), and the amino acid variant was consistently observed. While extensive and careful study of these variants has shown that the regulatory polymorphisms all have effects on transcription, either independently or epistatically (*e.g.*, Kudaravalli *et al.* 2002), it is unknown whether natural selection acts on the polymorphisms independently.

To address the evolutionary history of functional variants affecting *F7* transcription we take both a phylogenetic and a population genetic approach. We first use sequences from the 5′ *cis*-regulatory region of *F7* from five chimpanzees, two bonobos, a gorilla, an orangutan, and a baboon to investigate the origin and interspecific patterns of selection on functional variants. We then present a population genetic survey of three regulatory variants with differing effects on F7 expression (−402, −401, and −324) from six locales: Italy, Cameroon, Ethiopia, China, India, and Papua New Guinea. Allele frequencies and population differentiation at each site were estimated, and comparison to 18 unlinked, putatively neutral SNPs in the same individuals allowed us to test for the action of natural selection. Our results demonstrate the action of adaptive natural selection in shaping the pattern of change over time in *F7* and in shaping patterns of variation in at least one of the regulatory polymorphisms. Thus, our results further support the hypothesis that regulatory mutations are important for human evolution.

MATERIALS AND METHODS

Subjects: DNA was obtained from 45 unrelated individuals in each of six focal populations: southern Italy, Cameroon, Ethiopia, China (Singaporean Chinese), India (Uttar Pradesh), and Papua New Guinea (Madang Coastal). In addition, we obtained DNA from 37 individuals from a population of Japanese. Human DNA samples were collected by the Goldstein lab with signed consent or were anonymous legacy collections provided to D. Goldstein by collaborators from other academic research universities. DNA from five chimpanzees (four *Pan troglodytes verus* and one *Pan troglodytes schweinfurthii*), two bonobos (*Pan paniscus*), and one baboon (*Papio papio*) was provided by A. Stone and D. Loisel. Gorilla (*Gorilla gorilla*) and orangutan (*Pongo pygmaeus*) DNA was purchased from Coriell Institute for Medical Research Cell Repositories.

Genotyping: A PCR-amplified 433-bp fragment of the 5′ regulatory region of *F7* was cloned and sequenced from all of the nonhuman primates using the primers 5′-ggctctgt ggctcacctaag-3′ and 5′-gactgacgggcaaagttctc-3′ (GenBank accession nos. AY493422–AY493433). Fragments were cloned into pGEM-T vectors (Promega, Madison, WI) and multiple clones were sequenced in each direction. All base-calls were confirmed in multiple identical clones, and the chimpanzee and bonobo products were also directly sequenced as a check on *Taq* error. For genotyping the six focal human populations we PCR amplified the same fragment using a 6-FAM labeled reverse primer. The −324 indel variant was genotyped by scoring the size fragment after running it out on an ABI 3700 capillary gel machine using the ABI Genescan/Genotyper software package. The −401 and −402 SNPs were genotyped by sequencing using only the forward primer. (The −401 polymorphism was submitted to dbSNP and given the identifier ss15737439.) The products were run out on the ABI 3700 and scored two independent times by the authors using Sequencher (Gene Codes, Ann Arbor, MI). For the Japanese population, all PCR products were sequenced in both directions with the forward and reverse primers.

Statistical analysis: All sequences were aligned in Clustal X (Thompson *et al.* 1994). Assignment of substitutions to branches of the primate phylogeny was done using HyPhy (S. Kosakovsky Pond and S. Muse; http://www.hyphy.org). Hardy-Weinberg equilibrium significance was tested by the random-permutation method implemented by the program GDA (P. Lewis and D. Zaykin; http://lewis.eeb.uconn.edu/lewishome/

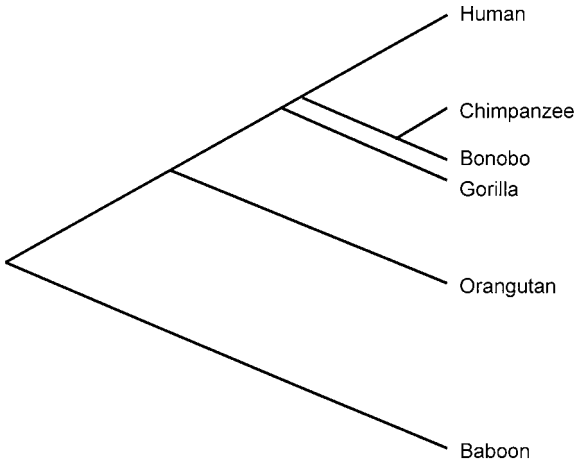

FIGURE 1.—Phylogeny of the great apes and baboon. Branch lengths are proportional to time (YI *et al.* 2002).

software.html). Linkage disequilibrium was assessed by Δ, a measure of composite genotypic disequilibrium that does not require a known phase of alleles (WEIR 1996), implemented in a perl program submitted to the Bioperl project (STAJICH *et al.* 2002). We corrected for multiple comparisons using the Dunn-Sidak method (SOKAL and ROHLF 1995, p. 239); this method is less conservative than a Bonferroni correction (SOKAL and ROHLF 1995). Wright's F_{ST} (WRIGHT 1951) was estimated using the diploid method of WEIR (1996, p. 178), also available through the Bioperl project (http://www.bioperl.org). All F_{ST} values <0 were set equal to 0 (negative values can occur when the true value is positive but very small). Significance of the difference between F_{ST} at *F7* −324, −401, and −402 and F_{ST} in 18 unlinked, neutral SNPs was estimated by comparison to a bootstrap-resampled distribution as follows: the difference between the single-locus F_{ST} of a randomly sampled neutral marker and the multilocus F_{ST} of 18 randomly resampled neutral markers was calculated 10,000 times. This resampling was performed for global and all pairwise F_{ST}'s. The observed difference between the functional variant F_{ST} value and the 18-locus estimate for the neutral loci was then compared to these distributions in a two-tailed test.

RESULTS

Phylogenetic patterns of F7 evolution: Analysis of a 433-bp alignment of the 5′ regulatory region (starting at −68 relative to the start of translation) in six primates revealed four fixed differences along the branch leading to humans (Figures 1 and 2). The level of divergence in the 5′ region along the human branch, as well as the substitution rate among all of these primates, is well within the average rate measured among 45 large, noncoding regions for this clade (YI *et al.* 2002). Three of the fixed differences in humans (positions −401, −220, and −108) are located in experimentally verified transcription factor binding sites (−220 and −108; POLLAK *et al.* 1996) or have been shown to have effects on transcription both *in vivo* and *in vitro* (−401; VAN 'T HOOFT *et al.* 1999). The fourth fixed difference, at −196, is not located within any known transcription factor binding site.

The binding site spanning −233 to −215 is an estrogen response element that binds estrogen receptor (DI BITONDO *et al.* 2002); mutation at −220 from the invariant A to a C has been shown to severely diminish F7 expression *in vitro* (DI BITONDO *et al.* 2002). While the change along the human branch at site −220 is from G to the consensus A (Figure 2), this site lies in the middle of the required sequence for binding of estrogen receptor and therefore represents a possible gain-of-function mutation in humans. The binding site spanning −108 to −84 is a transcription factor Sp1 binding site (POLLAK *et al.* 1996). A rare, naturally occurring mutation at −94 in this site disrupts Sp1 binding and leads to severe F7 deficiency (CAREW *et al.* 1998). We do not know the effect of the observed change at position −108 in this binding site. The two alleles observed at position −401 are both derived, and therefore we have counted this as a fixed difference. The ancestral C nucleotide has been replaced by a G/T polymorphism; this does not appear to be a CpG-related mutation. Polymorphisms for which both alleles are different from the presumed ancestral variant for hominids are rare because of both a short divergence time since the split from chimpanzees and bonobos and low levels of polymorphism within our species (KAESSMANN *et al.* 1999). The two alleles have different effects on transcription (see DISCUSSION).

The ancestral states of the three other polymorphic regulatory variants in *F7* present some equally complex relationships between derived states and effect on expression. It is clear from the alignment of noncoding DNA that a C at nucleotide position −122 is the derived state in humans (Figure 2). This derived mutation is the allele associated with lower expression of F7. At position −402, the derived A allele causes increased expression in F7. And at position −324, an invariant 10-bp sequence in nonhuman primates is now the low-frequency insertion allele of the insertion/deletion polymorphism. The derived deletion allele is responsible for higher levels of F7 expression. Derived variants in humans are thus found to be both increasing and decreasing levels of F7 expression relative to the ancestral alleles.

Although there are only four fixed differences in humans, there appears to be an excess of fixations within transcription factor binding sites along the human lineage. To determine whether three of the four fixed differences falling within binding sites is more than that expected by chance, we counted the total number of nucleotides in this region that comprise transcription factor binding sites; our conservative estimate was 74 (experimentally defined: −108 to −84, −233 to −215, 10-bp indel; estimates: 10 bp for −402/−401 binding site, 10 bp for −122 binding site). Note that most binding sites are 6–10 bp long (FAIRALL and SCHWABE 2001). A ratio of the substitutions per site in replacement sites (K_a) to synonymous sites (K_s) in coding regions has been

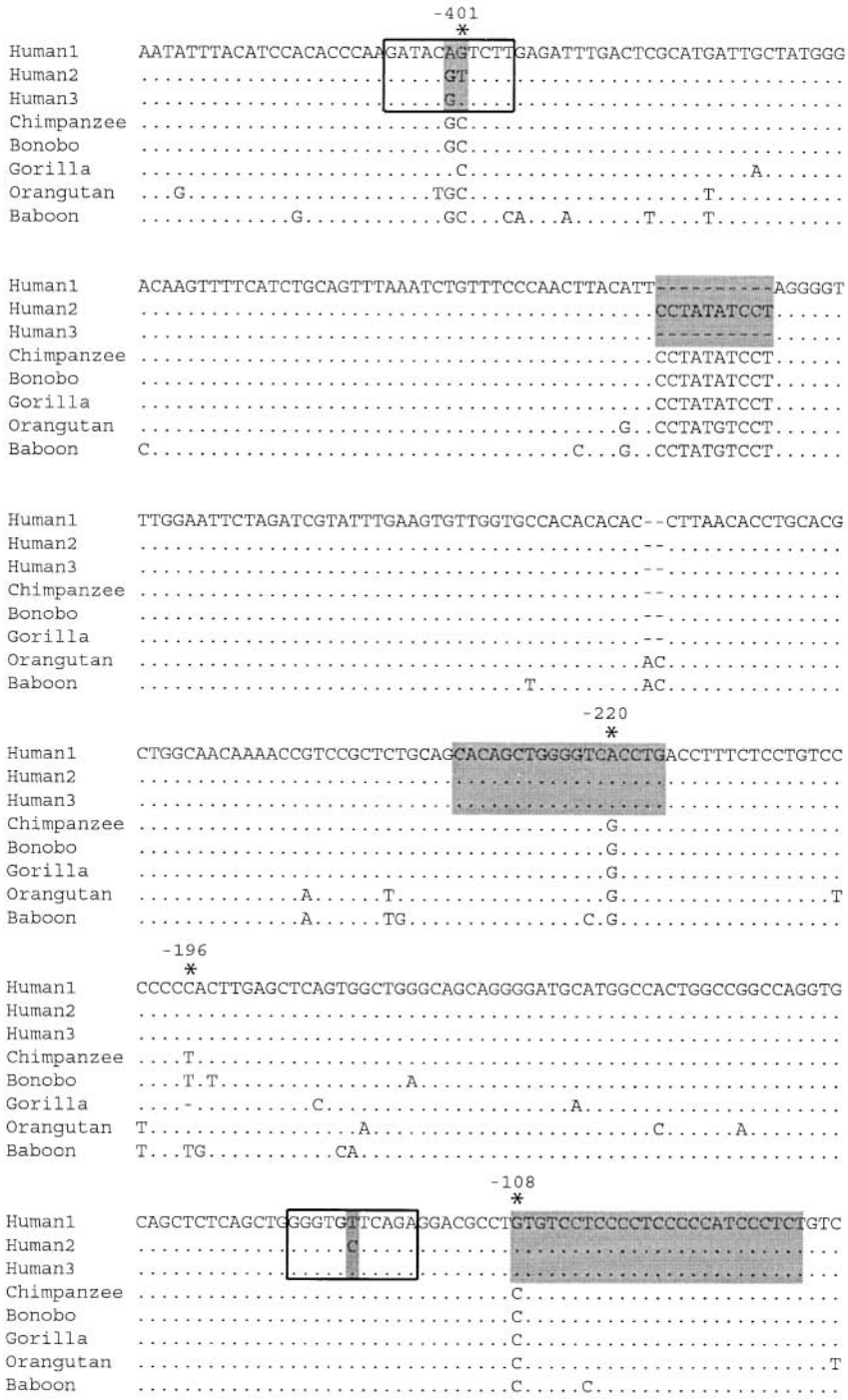

FIGURE 2.—Variation in the *F7* promoter region. Sequence alignment of part of the 5′ noncoding region of *F7* from −428 at the top left (based on numbering in humans), to −81 at the bottom right. The common human haplotypes (VAN 'T HOOFT *et al.* 1999) are listed separately. Experimentally verified transcription factor binding sites or nucleotides with effects on transcription are shaded, and presumed binding site nucleotides are boxed. Fixed differences along the human branch are noted with an asterisk and their nucleotide position.

used to quantify patterns of natural selection in proteins (KIMURA 1977). A K_a/K_s ratio <1 is consistent with a history of negative (purifying) selection, although it does not rule out positive selection, while a K_a/K_s ratio >1 indicates strong positive selection, although it does not mean that negative selection is not also acting (YANG and BIELAWSKI 2000). By analogy, we can measure the ratio of the substitutions per site in binding sites (K_b) to intervening, nonbinding nucleotide sites (K_i) in regulatory regions, with the same interpretation of results. Along the human branch of the primate tree, we find $K_b/K_i = 14.6$ [(3/74)/(1/359)]. Using a χ^2 statistic testing for an excess of fixed differences in binding sites relative to nonbinding sites gives a value of 9.2 ($P = 0.002$). This result suggests that positive selection may have been involved in the fixation of regulatory mutations in humans.

Population variation in human F7 regulatory mutations: Three of the functional regulatory polymorphisms at the *F7* locus were genotyped in humans: two SNPs (−402 and −401) and one indel (−324). The frequencies of the two alleles at each site were deter-

TABLE 1

Allele frequencies

Site (allele)	Cameroon	China	Ethiopia	India	Italy	Papua New Guinea
−324 (del)	0.65 (80)[a]	0.97 (88)	0.75 (88)	0.79 (86)	0.76 (88)	0.99 (88)
−401 (G)	0.56 (78)	0.90 (62)	0.74 (88)	0.70 (56)	0.75 (88)	0.99 (88)
−402 (A)	0.05 (78)	0.52 (62)	0.10 (88)	0.23 (56)	0.14 (88)	0.54 (88)

[a] The value in parentheses is the number of chromosomes sampled.

mined from six Old World human populations: southern Italy, Cameroon, Ethiopia, China (Singaporean Chinese), India (Uttar Pradesh), and Papua New Guinea (Madang Coastal). The frequencies of each allele in each of the populations are shown in Table 1. All populations were missing data for some individuals at all of the variants. The missing data ranged from 1 individual to 17 individuals of the 45 sampled in each population (Table 1). Tests for Hardy-Weinberg equilibrium in each population for each variant were nonsignificant in every case but the Ethiopian −402 data (Fisher's exact test, $P < 0.05$). The chromatograms from this population were rechecked for base-calling errors, but the significant disequilibrium remained. This result does not appear to affect any of our conclusions (see below). We found high levels of linkage disequilibrium between variants at −401 and −324 in every population ($P < 0.05$ in all comparisons; Table 3). This indicates that the low-expression, low-frequency variants were in positive linkage disequilibrium with each other. This pattern has been shown previously for *F7* in multiple human populations (*e.g.*, Bernardi *et al.* 1997; de Maat *et al.* 1997).

To measure the amount of population structure present in our data we calculated global and pairwise values for F_{ST} (Wright 1951) among all six populations. (F_{ST} can take on values between 0, no differentiation between populations, and 1, complete differentiation.) In the absence of selection, these values reflect divergence in allele frequencies due to genetic drift. To assess the possible role of natural selection in shaping the population differentiation in the functional SNPs, we compared F_{ST} at these variants to F_{ST} calculated for 18 SNPs genotyped in the same individuals and unlikely to be affected by selection (Rockman *et al.* 2003). These 18 SNPs were chosen to be >200 kb from any known gene and 50 kb from any mapped expressed sequence tag, unlinked (on different chromosomes), or at least 50 cM apart if on the same chromosome (on the basis of the UCSC April 2002 freeze, NCBI Build29). Alleles at each of the 18 unlinked loci are expected to have the same F_{ST}, but the stochastic nature of genetic drift creates a large variance in the distribution of values. We used the empirical distribution of the putatively neutral SNPs to identify variants with unusually high or low F_{ST}'s (*cf.* Bamshad *et al.* 2002; Fullerton *et al.* 2002; Hamblin *et al.* 2002). Significantly high F_{ST} values indicate positive selection driving changes in allele frequencies in individual populations, while significantly depressed values can indicate balancing selection maintaining allele frequencies between populations (Lewontin and Krakauer 1973). Statistical power to detect significantly depressed values in humans is low, as the average differentiation among continental populations is only ~0.12 (Akey *et al.* 2002).

Global F_{ST} values at sites −402, −401, and −324 were 0.228, 0.121, and 0.106, respectively; the mean of the neutral SNPs is 0.127. The value at position −402 is higher than that observed for 17 of the 18 neutral loci, but is not significantly different from the neutral distribution on the basis of bootstrap resampling of the difference between random single-locus and 18-locus F_{ST} values (see materials and methods). The global estimate of F_{ST} masks any population-specific processes by averaging over all populations. Therefore, we also calculated all the pairwise values for F_{ST} between populations for all variants (Table 2). Two consistent, significant patterns are seen. The first is that all of the functional variants show a significantly higher-than-expected F_{ST} in the China-India comparison. The neutral markers show extremely low levels of differentiation in this comparison, and thus even moderate F_{ST} values are far outside the neutral distribution. Previous studies of putatively neutral markers have also shown low levels of F_{ST} between East Asian and Indian populations (*e.g.*, Watkins *et al.* 2003). A second, striking pattern at position −402 is the significant pairwise comparisons between the Singaporean Chinese and Ethiopians, Indians, and Italians (Table 2). These significant values indicate that the derived A allele at −402 shows a greater-than-expected difference in frequency as compared to the neutral SNPs. Indeed, the F_{ST} values in these comparisons are higher than those measured for any of the neutral loci. Even though the Papua New Guineans show generally more differentiation from all human populations including the Chinese (Table 2), this comparison is not significant. Nevertheless, the large increase in frequency of the high-expression A allele at position −402 in Singaporean Chinese suggests the action of positive natural selection.

To ask whether the increase in frequency of the −402A allele in Singaporean Chinese represents a pop-

TABLE 2

Pairwise population differentiation (F_{ST})

	Population					
	Cameroon	China	Ethiopia	India	Italy	Papua New Guinea
Average of 18 neutral SNPs						
Cameroon		0.133	0.052	0.101	0.120	0.327
China			0.053	0.026	0.082	0.184
Ethiopia				0.018	0.041	0.251
India					0.033	0.222
Italy						0.302
Papua New Guinea						
−324						
Cameroon		0.276	0.012	0.037	0.019	0.327
China			0.165	0.125***	0.154	0.001
Ethiopia				0.000	0.000	0.213
India					0.000	0.173
Italy						0.203
Papua New Guinea						
−401						
Cameroon		0.234	0.053	0.021	0.063	0.418
China			0.069	0.112***	0.061	0.063
Ethiopia				0.000	0.000	0.225
India					0.000	0.315
Italy						0.215
Papua New Guinea						
−402						
Cameroon		0.430*	0.000	0.123	0.025	0.433
China			0.341***	0.142***	0.284***	0.000
Ethiopia				0.047	0.000	0.356
India					0.016	0.168
Italy						0.304
Papua New Guinea						

*$P < 0.1$; ***$P < 0.0001$.

ulation- or region-specific phenomenon, we also genotyped the *F7 cis*-regulatory region in a population of Japanese. Although we do not have data on the frequencies of the 18 neutral SNPs for this population, the frequency of the −402 alleles as well as alleles at −401, −324, and −122 provides us with data from another East Asian population. In a group of 37 individuals (74 chromosomes) sequenced, the −402A allele was present at a frequency of 36%. This value is midway between the frequency in the Singaporean Chinese (52%; Table 1) and the values in Cameroon, Ethiopia, India, and Italy (5–23%; Table 1). The frequencies of −401G (93%) and the −324 deletion (93%) in the Japanese are very similar to the values in the Singaporean Chinese (90 and 97%, respectively; Table 1). We were also able to score the C/T polymorphism at site −122 and found that the C was present at a frequency of 92%. The polymorphisms at −401, −324, and −122 were in almost perfect linkage disequilibrium among the Japanese: only one recombinant chromosome was detected. In contrast, nucleotide site −402 was not in linkage disequilibrium with any of these sites.

DISCUSSION

The estimated number of heterozygous functional *cis*-regulatory sites in humans is larger than the number of heterozygous amino acid sites (Rockman and Wray 2002). Regulatory polymorphisms are present at appreciable frequencies in all human populations, with an estimated 40% of all loci heterozygous for a functional regulatory mutation (Rockman and Wray 2002). With so much heritable variation available for selection to act upon, it behooves us to discover how natural selection has shaped and used *cis*-regulatory DNA in human evolution. The goal of our study was to uncover and characterize the patterns of selection, if any, in the *F7 cis*-regulatory region. Both our population genetic and phylogenetic sequence data showed patterns that dif-

TABLE 3

Linkage disequilibrium

Comparison	Cameroon	China	Ethiopia	India	Italy	Papua New Guinea
−324, −402	−0.01 (0.4)[a]	0.02 (1.0)	0.02 (0.4)	0.10 (8.1)	0.04 (3.4)	0.00 (0)
−324, −401	0.20 (25.4)*	0.04 (11.3)*	0.19 (40.5)*	0.19 (27.0)*	0.17 (40.2)*	0.01 (43.0)*
−401, −402	0.01 (0.1)	0.07 (4.7)	0.02 (0.6)	0.09 (7.0)	0.05 (3.7)	0.00 (0)

*$P < 0.05$ after Dunn-Sidak correction for multiple comparisons.
[a] First value is the composite genotypic disequilibria, Δ; χ^2 values are in parentheses.

fered from neutral expectations, with directional selection the most likely cause of these departures.

F7 is an important component of the blood-clotting pathway: levels of F7 protein in the bloodstream are strongly correlated with the risk of myocardial infarction. Expression is controlled by a well-characterized, TATA-less, *cis*-regulatory region 5′ of the start of transcription (Pollak *et al.* 1996). Four common regulatory polymorphisms have been shown to affect transcription and protein levels *in vivo*. Three of these polymorphisms have minor-frequency alleles that lower the level of F7 and, hence, lower the risk of heart attack, and one has a minor allele that raises the level of F7 (Bernardi *et al.* 1996; Pollak *et al.* 1996; van 't Hooft *et al.* 1999; Di Castelnuovo *et al.* 2000; Girelli *et al.* 2000). We studied the evolutionary history of these functional polymorphisms in humans and nonhuman primates.

Analysis of the 5′ *cis*-regulatory region among primates reveals the evolution of both high-expression and low-expression alleles in humans. At polymorphic position −122, the low-expression C allele is clearly derived in humans (Figure 2). At polymorphic positions −324 and −402, the high-expression alleles are derived. Position −401 segregates two derived alleles, one low expression and one high expression (the expression level of the ancestral allele is unknown). The sequencing of 10 chimpanzee and 4 bonobo chromosomes confirms our inferences of ancestral states of the four polymorphisms by showing that they are not segregating in these species. It is interesting to note that the low-expression derived allele at position −122 (C) is in positive linkage disequilibrium with the low-expression alleles at position −401 (T) and −324 (insertion; Table 3). Position −402 is not in linkage disequilibrium with any of these sites in our study, but was previously found to be in weak negative linkage disequilibrium with the low-expression alleles (van 't Hooft *et al.* 1999). The positive linkage disequilibrium between the low-expression variants found here and in previous studies (Bernardi *et al.* 1996; Humphries *et al.* 1996; Girelli *et al.* 2000) and the possible negative linkage disequilibrium between the high-expression allele at −402 and the low-expression alleles are important factors in studies that seek to diagnose the risk of heart attacks in patients with heart disease. The patterns of linkage between protective alleles in the seven human populations studied here indicate that only one or a few sites need to be genotyped to assess the complete haplotypic genotype of heart disease patients. This relatively low level of haplotype diversity is a promising development for large clinical studies of disease risk in multiple ethnic groups (Gabriel *et al.* 2002).

Combining information from previous biochemical studies of the *F7* polymorphisms with data on the ancestral states of these polymorphisms allows us to examine the evolution of transcription factor–DNA interactions. The derived, high-expression −402A allele has been shown to have a decreased binding affinity for an unknown nuclear protein (van 't Hooft *et al.* 1999). The low-frequency, low-expression −401T allele in the neighboring nucleotide position has been shown to have an *increased* binding affinity for the same unknown protein. This set of relationships between binding affinities and expression levels suggests that the unknown nuclear transcription factor acts as a repressor in the *F7* promoter: high levels of expression are associated with low binding affinities and low levels of expression are associated with high binding affinities. Therefore, we are able to infer the loss or severe lessening of a transcription factor-DNA interaction due to the derived −402A allele and an increased transcription factor-DNA interaction due to the −401T allele. Also, −402A shows a decreased binding affinity for a second, unknown nuclear protein, while −402T shows a decreased binding affinity for this protein relative to the more common G allele (van 't Hooft *et al.* 1999). These two SNPs therefore both change the binding affinities of multiple transcription factors that share the same or overlapping binding sites. Pleiotropic and epistatic interactions like these may be relatively common among regulatory mutations (Horan *et al.* 2003; Wray *et al.* 2003). No biochemical evidence is available for the transcription factor-DNA interactions at the other polymorphisms in this region.

There are four fixed differences along the human lineage in the 433 bp of 5′ noncoding sequence that we have collected. By analogy with the ratio of K_a to K_s, the ratio of nonsynonymous to synonymous substitutions per site, we can measure the ratio of the substitutions per site in binding sites (K_b) to intervening, nonbinding nucleotide sites (K_i) to show an excess of binding site substitutions along the human lineage (see

results). We believe that this is the first example of a case of positive selection in a *cis*-regulatory region detected by an analysis of K_b/K_i. A number of authors have previously used a modified McDonald-Kreitman test (McDonald and Kreitman 1991; Ludwig and Kreitman 1995) to demonstrate the action of positive selection in *cis*-regulatory regions (Jenkins *et al.* 1995; Jordan and McDonald 1998; Crawford *et al.* 1999). This modified test uses a comparison of the ratio of polymorphism to fixed differences of binding site and nonbinding site mutations (the original McDonald-Kreitman test compares nonsynonymous and synonymous mutations).

While the McDonald-Kreitman test requires a random sample of within-species haplotypes, the K_b/K_i ratio requires only single sequences from different species. Genotype data on previously ascertained polymorphisms, as we have collected, are not appropriate for use in a McDonald-Kreitman-type test. Indeed, it is exactly the previous ascertainment of functional regulatory polymorphisms that led us to study this region; however, no previous data on fixed differences existed. The previous ascertainment of functional polymorphisms in the *F7* *cis*-regulatory region raises the possibility that our study was biased toward common variants and thus that the differentiation among populations was biased at these sites. However, both the fact that overall differentiation among our functional polymorphisms was not different from the neutral differentiation and the fact that the high differentiation that we did find was in a population different from the ones used to ascertain the functional polymorphisms suggests that there was little or no effect of ascertainment bias. Our study of fixations along the human lineage should not have been affected by any ascertainment bias as we did not simply use one human haplotype in this comparison, but rather took full advantage of knowledge about polymorphism. It should also be noted that both types of tests of selection described here have been applied to promoter regions whose binding sites have been characterized in only one or a very few cell types. Sequences identified as nonbinding, therefore, may have some as yet unknown role in transcription factor binding. However, the proportion of binding site nucleotides identified in the *F7* *cis*-regulatory region (17%) agrees with the 10–20% range found in many other eukaryotic promoters (Wray *et al.* 2003).

To study the effects of natural selection, if any, in shaping the standing regulatory variation at *F7*, we genotyped three of the functional, phenotypically relevant promoter polymorphisms (−402, −401, and −324) in our six focal human populations. The amount of differentiation expected between populations in the absence of natural selection is a function of the time since divergence and the effective population size. As such, there is no one value of F_{ST} that indicates whether natural selection has acted between various populations without knowledge of the expected variation due to drift. To assess the role of natural selection in causing the differences in frequency between the human populations studied here, we compare the observed F_{ST} at the functional regulatory sites to a distribution of putatively neutral SNPs genotyped in the same individuals (Rockman *et al.* 2003). Comparison of loci of interest to an empirical neutral distribution of F_{ST}'s has been previously shown to be a powerful method for detecting selection on specific genes (Karl and Avise 1992; Taylor *et al.* 1995; Bamshad *et al.* 2002; Fullerton *et al.* 2002; Hamblin *et al.* 2002). If multiple polymorphisms within a single locus are not in linkage disequilibrium, then this method also offers the opportunity to detect selection on individual mutations contributing to differences in fitness.

A comparison of data from the three functional promoter polymorphisms examined here to the neutral distribution shows that the Singaporean Chinese population has diverged at position −402 much more than is expected under neutrality, consistent with a pattern of positive directional selection in this population (Table 2). Neither of the other two sites shows a signature of natural selection in any single population, although they both do show significantly high F_{ST}'s in the China-India comparison. Directional selection on the −402A allele is consistent with either an ongoing or an incomplete selective sweep in the Chinese population; it may be that an intermediate frequency is the evolutionary optimum for this polymorphism. Linkage disequilibrium between site −402 and other sites showing extreme patterns of differentiation could lead us to ascribe this selection to the incorrect polymorphism. Although the −401 and −324 polymorphisms show patterns of positive linkage disequilibrium with each other and several other polymorphisms, −402 does not show linkage disequilibrium with any known functional polymorphisms in *F7* in this study and only weak negative disequilibrium with −401 and −324 in a previous study (van 't Hooft *et al.* 1999). Because of this, we believe that the −402A allele is the mutation that contributes to differences in fitness and is therefore under positive selection.

Human populations have not diverged according to a strict branching process: migration between populations has almost certainly led to gene flow and the homogenization of genetic differences between populations (Goldstein and Chikhi 2002). Therefore, the effects of shared ancestry and migration will both cause populations to have similar allele frequencies. By looking at divergence in allele frequencies at many loci throughout the genome in the same individuals, we have attempted to control for phylogenetic and demographic effects. These effects are expected to act throughout the genome; our results show that the divergence in allele frequency at site −402 goes well beyond divergence at any of the neutral loci or even the other, closely linked functional polymorphisms. Our results using the neutral loci do not, however, tell us whether the increase in frequency of −402A found in Singaporean Chinese

is found in all East Asian populations or whether this pattern may actually be due to selection and shared ancestry with another East Asian population. To address this question we further characterized the allele frequencies of the functional *cis*-regulatory polymorphisms in a Japanese population. The frequencies of alleles −401G (93%) and the −324 deletion (93%) in the Japanese were almost exactly the same as those found in the Singaporean Chinese (90 and 97%; Table 1). But the frequency of −402A in the Chinese (52%) is still significantly higher than that found in the Japanese (36%); the Japanese are much closer in frequency to populations in Cameroon, Ethiopia, India, and Italy (Table 1). This result suggests two main scenarios for the slight increase in frequency of the −402A allele in the Japanese population. The first explanation is that there was a very large increase in frequency of this allele due to directional selection only in a Chinese population and that subsequently migration has been slowly homogenizing East Asian populations. The second main explanation is that there was some combination of genetic drift and directional selection in the ancestor of the Japanese and Chinese populations and that then selection continued only in the Chinese. Obviously, many other combinations of selection, demography, and phylogenetic history could explain the frequencies of the −402A allele; what we have argued here is simply that natural selection almost certainly had a role in East Asia.

Our study does not provide complete evidence for or against selection on the regulatory polymorphism at site −122 or the amino acid variant segregating at residue 353. In the Japanese population we have found that the polymorphism at −122 is in significant, strong positive linkage disequilibrium with sites −324 and −401, the two polymorphisms that we have found to be behaving neutrally; other studies have also consistently shown that sites −122 and residue 353 are both in linkage disequilibrium with these two polymorphisms (Lane *et al.* 1992; Kario *et al.* 1995; Bernardi *et al.* 1997; de Maat *et al.* 1997). Further studies comparing the amount of differentiation at residue 353 and nucleotide −122 to the neutral markers may definitively show whether or not selection has acted on them. While our results cannot exclude the possibility of either of these polymorphisms being under selection in an unknown population, neither does the lack of data contradict the conclusion that the *cis*-regulatory polymorphism at −402 is under positive selection.

Association studies and cell culture assays have both shown that the −402A allele confers increased transcriptional activity and increased levels of F7 in blood plasma (van 't Hooft *et al.* 1999). Results from long-term studies of heart disease, however, have shown that high levels of F7 are correlated with the probability of heart attack and death (*e.g.*, Meade *et al.* 1986; Heinrich *et al.* 1994; Redondo *et al.* 1999; Girelli *et al.* 2000). Given the above facts, is our finding of positive selection for an apparently deleterious allele unusual or even contradictory? We think that it is not. The allele frequencies in current populations are the result of drift and selection over many generations. Consequently, current patterns of population differentiation will be due to drift and selection acting across many generations of human history. Heart disease may be a relatively recent problem in human health, not common before 1900 (Fuster *et al.* 1992); it is also rare in other primates (Schmidt 1978). Heart disease also generally affects individuals after their reproductive years and so may not be an important component of fitness. It is therefore unlikely to have been a strong selective agent in our past. The "thrifty-genotype" hypothesis (Neel 1962) posits that humans are adapted to past conditions and that previously advantageous traits may now be associated with diseases such as diabetes. If alleles currently associated with disease are at frequencies that reflect past positive selection, then population genetic models must take this into account. The common disease/common variant hypothesis (Lander 1996) states that the genetic risk for common diseases will generally be associated with one or a few common alleles. The allelic spectra expected for disease-associated loci, however, have been predicted under the assumption of weak negative selection against disease alleles (Pritchard 2001; Reich and Lander 2001). Our results, as well as the thrifty-genotype hypothesis, would seem to violate these assumptions.

We do not know the selective forces responsible for the large increase in frequency of the high-expression allele at site −402. However, it is interesting to note that the other high-expression alleles show no signatures of natural selection. Simple assays of F7 levels in the bloodstream or transcription levels in only one or a few cell types may be missing a host of complex regulatory interactions or even whole transcriptional domains that might explain differences between "high-expression" alleles. Indeed, Wilcox *et al.* (2003) found that *F7* was expressed in multiple unanticipated tissues and therefore may have a role in multiple unknown cellular functions. Further research into the specific effects of individual *F7* alleles in multiple genetic backgrounds and in multiple cell lines is needed. Population genetic studies in humans showing positive directional selection within a population (this study; Sabeti *et al.* 2002), balancing selection between disease and protective states within and among populations (Hamblin and Di Rienzo 2000; Bamshad *et al.* 2002; Rockman *et al.* 2003), epistatic interactions between variants (Rockman and Wray 2002; Horan *et al.* 2003), and varying selective effects among polymorphisms (this study) argue against a simplistic view of *cis*-regulatory population dynamics. Transcriptional regulation is complex, idiosyncratic, and context dependent (Wray *et al.* 2003); understand-

ing the evolutionary forces shaping regulatory variation will be commensurately difficult.

We thank M. Rausher, C. Cunningham, J. Willis, and M. Uyenoyama for comments and discussion. Stephen Schaeffer and two anonymous reviewers also provided constructive comments. We acknowledge support of grants from the National Science Foundation (NSF; M.W.H., M.V.R., and G.A.W.), the National Aeronautics and Space Administration (G.A.W.), and the Leverhulme Trust (D.B.G. and N.S.). D.B.G. is a Royal Society/Wolfson Research Merit Award holder. M.W.H. is an NSF Integrative Informatics postdoctoral fellow.

LITERATURE CITED

Akey, J. M., G. Zhang, K. Zhang, L. Jin and M. D. Shriver, 2002 Interrogating a high-density SNP map for signatures of natural selection. Genome Res. **12:** 1805–1814.

Bamshad, M. J., S. Mummidi, E. Gonzalez, S. S. Ahuja, D. M. Dunn *et al.*, 2002 A strong signature of balancing selection in the 5′ *cis*-regulatory region of CCR5. Proc. Natl. Acad. Sci. USA **99:** 10539–10544.

Bernardi, F., G. Marchetti, M. Pinotti, P. Arcieri, C. Baroncini *et al.*, 1996 Factor VII gene polymorphisms contribute about one third of the factor VII level variation in plasma. Arterioscler. Thromb. Vasc. Biol. **16:** 72–76.

Bernardi, F., P. Arcieri, R. M. Bertina, F. Chiarotti, J. Corral *et al.*, 1997 Contribution of factor VII genotype to activated FVII levels: differences in genotype frequencies between northern and southern European populations. Arterioscler. Thromb. Vasc. Biol. **17:** 2548–2553.

Carew, J. A., E. S. Pollak, K. A. High and K. A. Bauer, 1998 Severe factor VII deficiency due to a mutation disrupting an Sp1 binding site in the factor VII promoter. Blood **92:** 1639–1645.

Carroll, S. B., J. K. Grenier and S. D. Weatherbee, 2001 *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Blackwell Science, Malden, MA.

Crawford, D. L., J. A. Segal and J. L. Barnett, 1999 Evolutionary analysis of TATA-less proximal promoter function. Mol. Biol. Evol. **16:** 194–207.

Davidson, E. H., 2001 *Genomic Regulatory Systems: Development and Evolution*. Academic Press, San Diego.

De Maat, M. P. M., F. Green, P. De Knijff, J. Jespersen and C. Kluft, 1997 Factor VII polymorphisms in populations with different risks of cardiovascular disease. Arterioscler. Thromb. Vasc. Biol. **17:** 1918–1923.

Di Bitondo, R., A. J. Hall, I. R. Peake, L. Iacoviello and P. R. Winship, 2002 Oestrogenic repression of human coagulation factor VII expression mediated through an oestrogen response element sequence motif in the promoter region. Hum. Mol. Genet. **11:** 723–731.

Di Castelnuovo, A., A. D'Orazio, C. Amore, A. Falanga, M. B. Donati *et al.*, 2000 The decanucleotide insertion/deletion polymorphism in the promoter region of the coagulation factor VII gene and the risk of familial myocardial infarction. Thromb. Res. **98:** 9–17.

Fairall, L., and J. W. R. Schwabe, 2001 DNA binding by transcription factors, pp. 65–84 in *Transcription Factors*, edited by J. Locker. Academic Press, San Diego.

Fullerton, S. M., A. Bartoszewicz, G. Ybazeta, Y. Horikawa, G. I. Bell *et al.*, 2002 Geographic and haplotype structure of candidate type 2 diabetes-susceptibility variants at the *calpain-10* locus. Am. J. Hum. Genet. **70:** 1096–1106.

Fuster, V., L. Badimon, J. J. Badimon and J. H. Chesebro, 1992 Mechanisms of disease: the pathogenesis of coronary artery disease and the acute coronary syndromes, I. N. Engl. J. Med. **326:** 242–250.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy *et al.*, 2002 The structure of haplotype blocks in the human genome. Science **296:** 2225–2229.

Gerhart, J., and M. Kirschner, 1997 *Cells, Embryos, and Evolution: Toward a Cellular and Developmental Understanding of Phenotypic Variation and Evolutionary Adaptability*. Blackwell Science, Malden, MA.

Girelli, D., C. Russo, P. Ferraresi, O. Olivieri, M. Pinotti *et al.*, 2000 Polymorphisms in the factor VII gene and the risk of myocardial infarction in patients with coronary artery disease. N. Engl. J. Med. **343:** 774–780.

Goldstein, D. B., and L. Chikhi, 2002 Human migrations and population structure: what we know and why it matters. Annu. Rev. Genomics Hum. Genet. **3:** 129–152.

Green, F., C. Kelleher, H. Wilkes, A. Temple, T. Meade *et al.*, 1991 A common genetic polymorphism associated with lower coagulation factor VII levels in healthy individuals. Arterioscler. Thromb. **11:** 540–546.

Hamblin, M. T., and A. Di Rienzo, 2000 Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. Am. J. Hum. Genet. **66:** 1669–1679.

Hamblin, M. T., E. E. Thompson and A. Di Rienzo, 2002 Complex signatures of natural selection at the Duffy blood group locus. Am. J. Hum. Genet. **70:** 369–383.

Heinrich, J., L. Balleisen, H. Schulte, G. Assmann and J. Vandeloo, 1994 Fibrinogen and factor VII in the prediction of coronary risk: results from the PROCAM study in healthy men. Arterioscler. Thromb. **14:** 54–59.

Horan, M., D. S. Millar, J. Hedderich, G. Lewis, V. Newsway *et al.*, 2003 Human growth hormone 1 (*GH1*) gene expression: complex haplotype-dependent influence of polymorphic variation in the proximal promoter and locus control region. Hum. Mutat. **21:** 408–423.

Humphries, S., A. Temple, A. Lane, F. Green, J. Cooper *et al.*, 1996 Low plasma levels of factor VIIc and antigen are more strongly associated with the 10 base pair promoter (−323) insertion than the glutamine 353 variant. Thromb. Haemostasis **75:** 567–572.

Iacoviello, L., F. Zito, A. Di Castelnuovo, M. De Maat, C. Kluft *et al.*, 1998 Contribution of factor VII, fibrinogen and fibrinolytic components to the risk of ischaemic cardiovascular disease: their genetic determinants. Fibrinol. Proteol. **12:** 259–276.

Jenkins, D. L., C. A. Ortori and J. F. Y. Brookfield, 1995 A test for adaptive change in DNA sequences controlling transcription. Proc. R. Soc. Lond. Ser. B Biol. Sci. **261:** 203–207.

Jordan, I. K., and J. F. McDonald, 1998 Interelement selection in the regulatory region of the *copia* retrotransposon. J. Mol. Evol. **47:** 670–676.

Kaessmann, H., V. Wiebe and S. Paabo, 1999 Extensive nuclear DNA sequence diversity among chimpanzees. Science **286:** 1159–1162.

Kario, K., N. Narita, T. Matsuo, K. Kayaba, A. Tsutsumi *et al.*, 1995 Genetic determinants of plasma factor VII activity in the Japanese. Thromb. Haemostasis **73:** 617–622.

Karl, S. A., and J. C. Avise, 1992 Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science **256:** 100–102.

Kimura, M., 1977 Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature **267:** 275–276.

King, M. C., and A. C. Wilson, 1975 Evolution at two levels in humans and chimpanzees. Science **188:** 107–116.

Kudaravalli, R., T. Tidd, M. Pinotti, A. Ratti, R. Santacroce *et al.*, 2002 Polymorphic changes in the 5′ flanking region of factor VII have a combined effect on promoter strength. Thromb. Haemostasis **88:** 763–767.

Lander, E. S., 1996 The new genomics: global views of biology. Science **274:** 536–539.

Lane, A., J. K. Cruickshank, J. Mitchell, A. Henderson, S. Humphries *et al.*, 1992 Genetic and environmental determinants of factor VII coagulant activity in ethnic groups at differing risk of coronary heart disease. Atherosclerosis **94:** 43–50.

Lewontin, R. C., and J. Krakauer, 1973 Distribution of gene-frequency as a test of the theory of the selective neutralism of polymorphisms. Genetics **74:** 175–195.

Ludwig, M. Z., and M. Kreitman, 1995 Evolutionary dynamics of the enhancer region of even-skipped in *Drosophila*. Mol. Biol. Evol. **12:** 1002–1011.

Marchetti, G., P. Patracchini, M. Papacchini, M. Ferrati and F. Bernardi, 1993 A polymorphism in the 5′ region of coagulation factor VII gene (F7) caused by an inserted decanucleotide. Hum. Genet. **90:** 575–576.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature **351:** 652–654.

Meade, T. W., M. Brozovic, R. R. Chakrabarti, A. P. Haines, J. D. Imeson *et al.*, 1986 Haemostatic function and ischaemic heart disease: principal results of the Northwick Park Heart Study. Lancet **2:** 533–537.

Neel, J. V., 1962 Diabetes mellitus: A "thrifty" genotype rendered detrimental by "progress"? Am. J. Hum. Genet. **14:** 353–362.

Pollak, E. S., H. L. Hung, W. Godin, G. C. Overton and K. A. High, 1996 Functional characterization of the human factor VII 5′-flanking region. J. Biol. Chem. **271:** 1738–1747.

Pritchard, J. K., 2001 Are rare variants responsible for susceptibility to complex diseases? Am. J. Hum. Genet. **69:** 124–137.

Raff, R. A., and T. C. Kaufman, 1983 *Embryos, Genes, and Evolution: The Developmental-Genetic Basis of Evolutionary Change*. Macmillan Publishing, New York.

Redondo, M., H. H. Watzke, B. Stucki, I. Sulzer, F. D. Biasiutti *et al.*, 1999 Coagulation factors II, V, VII, and X, prothrombin gene 20210G → A transition, and factor V Leiden in coronary artery disease: high factor V clotting activity is an independent risk factor for myocardial infarction. Arterioscler. Thromb. Vasc. Biol. **19:** 1020–1025.

Reich, D. E., and E. S. Lander, 2001 On the allelic spectrum of human disease. Trends Genet. **17:** 502–510.

Rockman, M. V., and G. A. Wray, 2002 Abundant raw material for *cis*-regulatory evolution in humans. Mol. Biol. Evol. **19:** 1991–2004.

Rockman, M. V., M. W. Hahn, N. Soranzo, D. B. Goldstein and G. A. Wray, 2003 Positive selection on a human-specific transcription factor binding site regulating *IL4* expression. Curr. Biol. **13:** 2118–2123.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter *et al.*, 2002 Detecting recent positive selection in the human genome from haplotype structure. Nature **419:** 832–837.

Schmidt, R. E., 1978 Systematic pathology of chimpanzees. J. Med. Primatol. **7:** 274–318.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry*. W. H. Freeman, New York.

Stajich, J. E., D. Block, K. Boulez, S. E. Brenner, S. A. Chervitz *et al.*, 2002 The bioperl toolkit: perl modules for the life sciences. Genome Res. **12:** 1611–1618.

Taylor, M. F. J., Y. Shen and M. E. Kreitman, 1995 A population genetic test of selection at the molecular level. Science **270:** 1497–1499.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

van 't Hooft, F. M., A. Silveira, P. Tornvall, A. Iliadou, E. Ehrenborg *et al.*, 1999 Two common functional polymorphisms in the promoter region of the coagulation factor VII gene determining plasma factor VII activity and mass concentration. Blood **93:** 3432–3441.

Watkins, W. S., A. R. Rogers, C. T. Ostler, S. Wooding, M. J. Bamshad *et al.*, 2003 Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. Genome Res. **13:** 1607–1618.

Weir, B. S., 1996 *Genetic Data Analysis II*. Sinauer Associates, Sunderland, MA.

Wilcox, J. N., S. Noguchi and J. Casanova, 2003 Extrahepatic synthesis of factor VII in human atherosclerotic vessels. Arterioscler. Thromb. Vasc. Biol. **23:** 136–141.

Wilson, A. C., 1975 Evolutionary importance of gene regulation. Stadler Genet. Symp. **7:** 117–134.

Wray, G. A., M. W. Hahn, E. Abouheif, J. P. Balhoff, M. Pizer *et al.*, 2003 The evolution of transcriptional regulation in eukaryotes. Mol. Biol. Evol. **20:** 1377–1419.

Wright, S., 1951 The genetical structure of populations. Ann. Eugen. **15:** 323–354.

Yang, Z. H., and J. P. Bielawski, 2000 Statistical methods for detecting molecular evolution. Trends Ecol. Evol. **15:** 496–503.

Yi, S. J., D. L. Ellsworth and W. H. Li, 2002 Slow molecular clocks in Old World monkeys, apes, and humans. Mol. Biol. Evol. **19:** 2191–2198.

Communicating editor: S. W. Schaeffer