Human Adaptive Evolution at *Myostatin* (*GDF8*), a Regulator of Muscle Growth

Matthew A. Saunders, Jeffrey M. Good, Elizabeth C. Lawrence, Robert E. Ferrell, Wen-Hsiung Li, and Michael W. Nachman

Myostatin (GDF8) is a negative regulator of muscle growth in mammals, and loss-of-function mutations are associated with increased skeletal-muscle mass in mice, cattle, and humans. Here, we show that positive natural selection has acted on human nucleotide variation at *GDF8,* since the observed ratio of nonsynonymous:synonymous changes among humans is significantly greater than expected under the neutral model and is strikingly different from patterns observed across mammalian orders. Furthermore, extended haplotypes around *GDF8* suggest that two amino acid variants have been subject to recent positive selection. Both mutations are rare among non-Africans yet are at frequencies of up to 31% in sub-Saharan Africans. These signatures of selection at the molecular level suggest that human variation at *GDF8* is associated with functional differences.

The genetic basis of muscle development and growth has been extensively studied in an effort to treat myopathies[1] and to understand individual variation in athletic performance.[2] Because musculature features might have provided a fitness advantage during human evolution, candidate genes related to musculature may have been targets of natural selection in humans.

The myostatin gene, also called "growth and differentiation factor 8" (*GDF8* [MIM 601788]) encodes a negative regulator of skeletal-muscle growth.[3] First described in the mouse, myostatin is expressed in different muscles throughout the body, both during early development and in adults. Mouse null mutants are significantly larger than wild-type animals, with 200%–300% more skeletal-muscle mass because of an increase in the number of myocytes (hyperplasy) and an increase in the size of muscle fibers (hypertrophy).[3] A similar phenotype is seen in some breeds of double-muscled cattle that also have *myostatin* mutations.[4,5] A loss-of-function mutation in the myostatin gene (a missplicing change in IVS1:G378A) has been associated with muscle hypertrophy in a human subject,[6] and myostatin expression levels have been shown to be inversely correlated with muscle mass in healthy and HIV-infected subjects.[7] These data suggest that myostatin acts in a similar fashion among all mammals. Here, we tested the hypothesis that patterns of human nucleotide variation at *GDF8* have been shaped by positive natural selection.

We resequenced the complete coding sequence of *GDF8,* including partial flanking intron sequences and the 5′ upstream *cis*-promoter region in human panels of 76 African Americans and 70 Europeans. DNA samples were collected from residents of Pittsburgh, after receipt of written informed consent. PCR amplification primers were designed to target the entire coding sequence for each *GDF8* exon (I–III) and the putative *cis*-promoter region for *GDF8,* located ~500 bp upstream of the ATG start codon.[8] DNA sequencing was performed on an ABI 3700 automated DNA sequencer with use of amplification oligonucleotides as primers. (Primers and PCR conditions are available on request). Sequences were inspected by eye and were aligned for each individual, with use of the program SEQUENCHER (Gene Codes). The sequence for each individual was submitted to GenBank under accession numbers DQ927046–DQ927191.

In the resequenced panels, we detected eight SNPs in the coding sequence, one indel polymorphism in intron 1, and two SNPs in the 5′ upstream *cis*-promoter region (fig. 1*A*). Surprisingly, nucleotide diversity in the coding sequence ($\pi = 0.038\%$) is four times higher than in the noncoding sequence ($\pi = 0.009\%$) (table A1). Five of the coding SNPs cause nonconservative amino acid replacement changes, whereas the three remaining coding SNPs cause silent changes (table 1).

Under the neutral model of molecular evolution, the ratio of replacement:silent changes is expected to be the same within and between species (McDonald-Kreitman test).[11] To test this null hypothesis for *GDF8,* we compared the polymorphism within humans with divergence between humans and other mammals (table B1). Between species, we observed many more silent than replacement mutations, as is typical for most genes. However, within humans, we observed more replacement than silent mutations, and this difference is statistically significant (table 2). This pattern is unusual for human genes and, together with observations described below, suggests that the high proportion of amino acid variation in humans is due to positive natural selection.

To further explore whether the patterns seen at *GDF8*

From the Department of Ecology and Evolutionary Biology, University of Arizona, Tucson (M.A.S.; J.M.G.; M.W.N.); Department of Human Genetics, University of Pittsburgh, Pittsburgh (E.C.L.; R.E.F.); and Department of Ecology and Evolution, University of Chicago, Chicago (M.A.S.; W.-H.L.)

Received July 6, 2006; accepted for publication September 15, 2006; electronically published October 10, 2006.

Address for correspondence and reprints: Dr. Matthew A. Saunders, Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street, Chicago, IL 60637. E-mail: saunders@uchicago.edu

Am. J. Hum. Genet. 2006;79:1089–1097. 0002-9297/2006/7906-0011$15.00

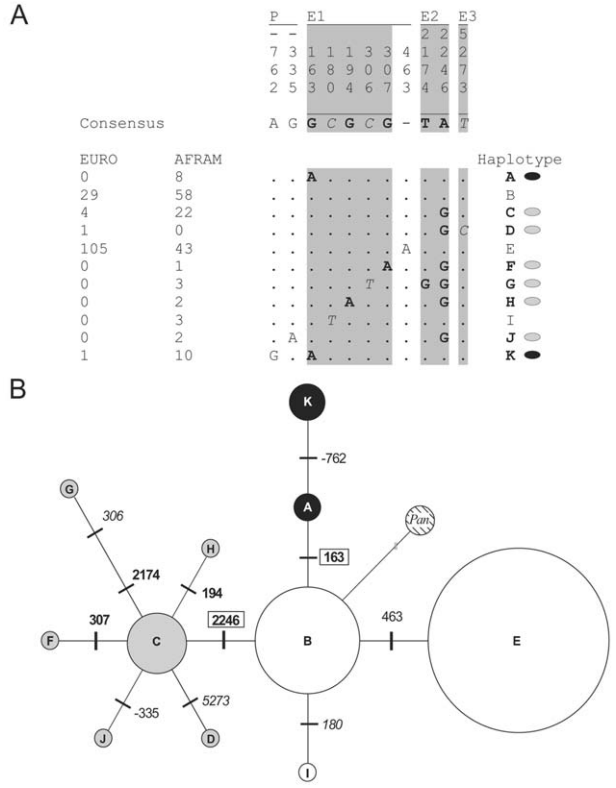

Figure 1. *A,* Sequences and number of occurrences of the inferred haplotypes (A–K) from *GDF8* in an African American panel (AFRAM) and a European panel (EURO). Surveyed regions are P (encompassing the putative *cis*-promoter region), E1, E2, and E3 (spanning exons 1, 2, and 3, respectively). Alignment positions are indicated at the top, relative (in base pairs) to the ATG coding start position. Polymorphisms in coding regions are shaded. The consensus sequence is denoted at the top, with sites that cause replacement polymorphisms shown in bold and silent polymorphisms shown in italics. A dot (·) indicates identity to consensus sequence, and a dash (—) indicates a deletion. Haplotypes bearing the intermediate frequency replacement changes Ala55Thr and Lys153Arg (nucleotide alignment positions 163 and 2246, respectively) are shown in bold. Haplotypes of haplogroup 55 and haplogroup 153 are indicated by black and gray ovals, respectively. *B,* Haplotype tree for *GDF8*. Each circle represents an inferred haplotype (A–K) in a size proportional to the haplotype frequency in the combined sample (AFRAM and EURO). Hatch marks represent single mutations labeled according to the alignment positions in panel A. Positions of replacement changes are indicated in boldface numbers. Replacement changes at amino acid sites 55 and 153 (alignment positions 163 and 2246, respectively) are boxed. Positions of silent changes are italicized. Haplotypes of haplogroup 55 and haplogroup 153 are indicated by black and gray circles, respectively. The haplotype inference method[9] is known to be inaccurate for rare ($q < 0.01$) frequency polymorphisms. Mutations 307 and 5273 are both at low frequency in the sample ($q = 0.0068$) and therefore may have been incorrectly assigned to a haplotype with mutation 2246. Alternative haplotype inferences for individual AFRAM_B11 would be (C,Y) (HapY: AG|GCGC**A**|—|TA|T) and, for individual EURO_B08, would be (C,Z) (HapZ: AG|GCGCG|A|TA|**C**) (see panel A and fig. A1).

in humans are unexpected, we compared *GDF8* sequences, obtained from public databases, among mammals and other vertebrates. We estimated the average ratio of replacement:silent substitutions per site (d_N:d_S) across 15 mammalian lineages, using maximum likelihood[12] (table B1). In the absence of functional constraint, the nonsynonymous substitution rate (d_N) is expected to be equal to the synonymous substitution rate (d_S), whereas d_N:d_S < 1 is indicative of purifying selection. The average d_N:d_S ratio across the mammalian phylogeny is 0.10, which suggests that *GDF8* has been under strong constraint throughout much of mammalian evolution. Furthermore, most pairwise interspecific comparisons with humans show even higher levels of constraint. For example, between mouse and human d_N:d_S = 0.05 for *GDF8,* well below the median genomewide value of 0.12.[13] Thus, the pattern of five replacement and three silent changes within humans stands out as being exceptional.

Next, we looked at the amino acid sites of each of the five human replacement polymorphisms across all vertebrate species with available *GDF8* sequences (a total of 20 species, including fish, birds, and mammals [table B1]). The five sites that are polymorphic in humans are remarkably conserved over evolutionary timescales. The ancestral amino acid states associated with the alleles Ala55Thr (G163A), Arg65His (G194A), and Asp103Asn (G307A) are conserved among all vertebrates; Lys153Arg (A2246G) is conserved among all taxa except fish (*Danio rerio*), and Met129Arg (T2174G) is conserved among all mammals except bovines. This high level of conservation for individual residues suggests that the mutations in humans have functional consequences.

To learn more about patterns of evolution at *GDF8*

Table 1. Summary of *GDF8* Polymorphisms in the Coding Region

Allele[a]	Exon	Codon Change[b]	Position[c]	Frequency AFRAM	Frequency EURO
A55T*	I	**G**CC→**A**CC	163	.118	.007
60I	*I*	*AT**C**→AT**T***	*180*	.020	.000
R65H	I	C**G**T→C**A**T	194	.013	.000
102S	*I*	*AG**C**→AG**T***	*306*	.020	.000
D103N	I	**G**AT→**A**AT	307	.007	.000
M129R	II	A**T**G→A**G**G	2174	.020	.000
K153R*	II	A**A**G→A**G**G	2246	.197	.036
E164K*	II	**G**AG→**A**AG	2278	NA	NA
P198A*	II	**C**CA→**G**CA	2380	NA	NA
I225T*	II	A**T**T→A**C**T	2462	NA	NA
354N	*III*	*AA**T**→AA**C***	*5273*	.000	.007

NOTE.—Alleles marked in italics are synonymous changes; all others are nonsynonymous changes. NA = not applicable.

[a] Substitutions marked with an asterisk (*) denote polymorphisms identified previously.[10] Substitutions E164K, P198A, and I225T were not detected in our current samples.

[b] The polymorphic nucleotide for each codon is shown in bold.

[c] Position is denoted as distance (in bp) from the first nucleotide of the start codon (relative to the consensus human sequence).

Table 2. McDonald-Kreitman Tests of Neutrality

Variation Type and Species	No. of Mutations		Fisher's Exact Test (*P*)
	Silent	Replacement	
Polymorphism:			
Human	3	5	
Fixed differences:			
Macaque	8	1	<.05
Mouse	77	15	<.01
Rat	86	17	<.01
Dog	77	15	<.01

NOTE.—Results of a 2 × 2 Fisher's exact test of independence between the human polymorphism found in the current study and divergence (fixed differences) from macaque, mouse, rat, and dog *GDF8* sequences. Three additional replacement polymorphisms at *GDF8* have been reported in a different sample of humans,[10] which brings the total number of known *GDF8* replacement polymorphisms to 8 (table 1), in which case Fisher's exact test results are $P < .01$ and $P < .001$ relative to macaque and the other mammals, respectively.

in humans, we inferred haplotype phase for the diploid sequences.[9] The two intermediate frequency polymorphisms—Ala55Thr and Lys153Arg (nucleotide mutations 163 and 2246, respectively)—reside on separate haplotypes (fig. 1*A*). The mutation at site 163 is associated with two haplotypes: A and K (referred to as "haplogroup 55") (fig. 1A). Haplotype K bears a derived mutation in the *GDF8* promoter region (site −762) that is found exclusively in association with the nonsynonymous change at site 163. The mutation at site 2246 is associated with six different haplotypes (haplotypes C, D, F, G, H, and J; referred to as "haplogroup 153"). Interestingly, all the low-frequency replacement mutations appear to be in association with a haplotype bearing the 2246 mutation (fig. 1*B*). We estimated the age of the mutations 2246 and 163, using long-range linked-polymorphism data (see YRI HapMap data below). The age estimates suggest that both alleles are relatively young and arose within the past 10,000 years (appendix C).

Alleles that have experienced recent positive selection may bear a signature of unusually long-range linkage disequilibrium with surrounding SNPs.[15,16] To test for this pattern at *GDF8,* we examined the SNP data from the International HapMap Project[17] in the genomic region of *GDF8*. Both sites G163A and A2246G have been genotyped in the International HapMap Project (SNPs *rs1805085* and *rs1805086,* respectively), and the minor-allele frequency for both SNPs in the Yoruban (YRI) panel from Ibadan, Nigeria, is 22%. We retrieved phased haplotypes spanning 300 kb roughly centered on *GDF8* from the International HapMap Project (human genome build 16) for 60 individuals from the YRI panel. We similarly retrieved phased SNP data from the YRI panel at 10 additional anonymous genomic regions, each spanning ~300 kb from across chromosome 2. Nonoverlapping core haplotypes (restricted to a maximum size of 8 contiguous SNPs) were defined in

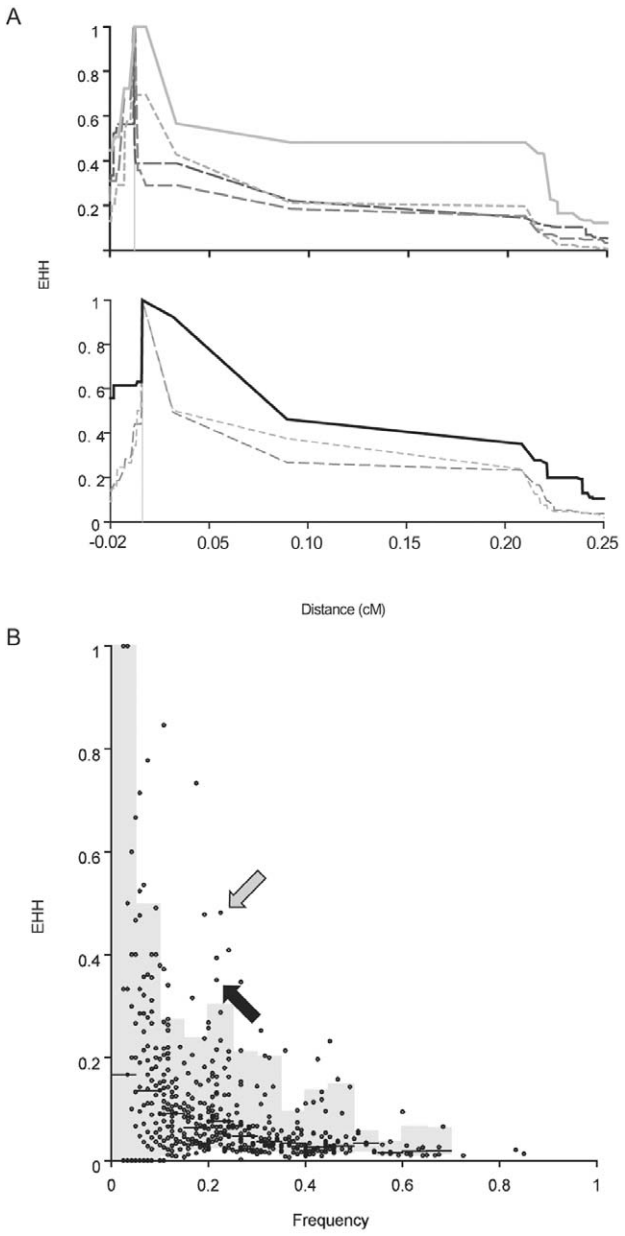

Figure 2. *A,* EHH at varying genetic distances from the core regions that include A2246G and G163A at *GDF8*. The EHH for each haplotype at a frequency >7% is displayed as follows. The core haplotypes of haplogroups 153 and 55 are indicated by a solid gray line and a solid black line, respectively. These extended haplotypes exhibit high levels of EHH at a distance of up to 0.2 cM from the core. All other core haplotypes are indicated by dashed lines. *B,* EHH relative to population frequency. EHH values were calculated for each core haplotype at a genetic distance of 0.2 cM. An empirical distribution of EHH values was produced for core haplotypes from 10 anonymous genomic regions across chromosome 2 in the HapMap YRI panel. These data were compared with core haplotypes from the genomic region spanning ~300 kb centered on *GDF8*. Values of EHH for the *GDF8* core haplotypes representing haplogroups 153 (*gray arrow*) and 55 (*black arrow*) are significantly higher than other core haplotypes within the same population frequency bin ($P = .01$ and $P = .03$, respectively).

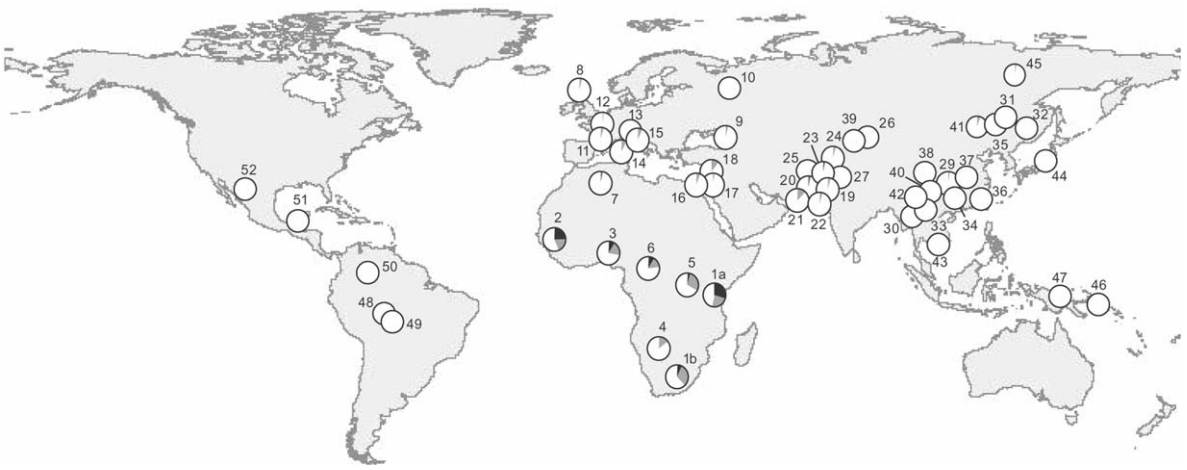

Figure 3. Population frequencies for *GDF8* replacement nucleotide polymorphisms G163A and A2246G (corresponding to haplogroups 55 and 153, respectively) in a worldwide panel (HGDP-CEPH Diversity Panel). For each population, the respective pie chart denotes the frequencies of haplogroups 55 (*dark shading*) and 153 (*light shading*). Population and haplogroup details are shown in table 3.

each genomic region, and extended haplotype homozygosity (EHH) was calculated at increasing genetic distances (measured in centimorgans).[15] We used the cores of the anonymous regions to generate an empirical distribution for the relationship between EHH and haplotype frequency. Statistical significance for departure of an EHH value within a frequency bin was determined for given *GDF8* haplotypes relative to the empirical distribution data. All long-range haplotype analyses were conducted using SWEEP software, according to standard documentation.[15] Using these data, we determined that the long-range haplotypes associated with haplogroups 153 and 55 exhibit a significant level of EHH relative to other core haplotypes up to ~0.2 cM away from *GDF8* (fig. 2*A*) and relative to an empirical distribution of core haplotypes from other genomic loci (fig. 2*B*) ($P = .01$ and $P = .03$ for haplogroups 153 and 55, respectively). It is noteworthy that a recent genomic scan for positive selection in the human genome that was based on a modification of the EHH statistic (i.e., the integrated haplotype score [IHS]) failed to identify long-range haplotypes in the genomic region of *GDF8* as extreme (top 1%) outliers.[14] However, this method (the IHS statistic) lacks power to detect selection on derived haplotypes at population frequencies < 0.5,[14] as is the case for *GDF8* and some other loci known to be under selection (e.g., *G6PD* [MIM 305900]).

Interestingly, two replacement polymorphisms—G163A and A2246G (representing haplogroups 55 and 153, respectively)—are at relatively high frequency among African Americans (12% and 20%, respectively) and the YRI panel (22% each) but are at much lower frequencies among the Europeans sampled here (1% and 4%, respectively) (table 1) and in the HapMap panels from Europe (0% and 2%, respectively) and Asia (0% each). Furthermore, it is notable that >50% (39 of 76) of the African Americans and 75% (45 of 60) of the YRI individuals bear at least one of these replacement alleles (figs. 1*A* and A1). To have a more complete view of the frequencies of G163A and A2246G in human populations, we genotyped 1,040 individuals in a worldwide panel that included samples from Africa, Europe, Asia, and the Americas (Human Genome Diversity Project [HGDP]–CEPH Diversity Panel).[18] SNP genotyping of this global panel was performed using TaqMan Assays by Design (Applied Biosystems). A separate assay probe was designed for each of the polymorphisms (G163A and A2246G), and reactions were performed, following manufacturer's protocols, on an ABI 7500 Real-Time PCR machine. All homozygotes of the rare alleles and any ambiguous calls were confirmed by PCR and direct resequencing. Results of the global survey show that one or both of mutations G163A or A2246G are found in populations of sub-Saharan Africa at a frequency >14%, whereas, in non-African populations, their frequencies are typically absent and only rarely exceed 5% (fig. 3 and table 3).

The signature of positive selection at the molecular level in humans is often weak, in part because of the relatively small long-term effective population size and low levels of standing variation. As a result, recent studies have focused on methods to detect selection on the basis of subtle aspects of the data, such as the decay of long-range linkage disequilibrium around a target of selection.[14,15,19] In contrast, the results presented here for *GDF8* also reveal a strong signature of positive selection that is based on an excess of nonsynonymous polymorphism, as previously seen only for a few other genes in humans (e.g., *G6PD* and the major histocompatibility complex).[20,21] An earlier survey of human variation at *GDF8* revealed three additional low-frequency replacement polymorphisms (but no additional silent polymorphisms) that were not detected in our current resequenced panels,[10] which increases to

Table 3. Population and Haplogroup Details for Figure 3

Map Designation	Population	Country of Origin	Sample Size	Frequency (%) of Haplogroup	
				55	153
1a	Bantu NE	Kenya	12	29	25
1b	Bantu SW	South Africa	8	6	31
2	Mandenka	Senegal	23	24	20
3	Yoruba	Nigeria	24	8	19
4	San	Namidia	7	0	14
5	Mbuti Pygmies	Democratic Republic of Congo	14	4	29
6	Biaka Pygmies	Central African Republic	36	8	14
7	Mozabite	Algeria (Mzab)	29	2	2
8	Orcadian	Scotland (Orkney Islands)	15	0	3
9	Adygei	Russia (Caucasus)	17	0	3
10	Russian	Russia	25	0	0
11	French Basque	France	24	0	4
12	French	France	27	0	2
13	North Italian	Italy (Bergamo)	14	0	0
14	Sardinian	Italy	28	0	7
15	Tuscan	Italy	8	0	6
16	Bedouin	Israel (Negev)	48	0	5
17	Druze	Israel (Carmel)	47	0	1
18	Palestinian	Israel (Central)	49	1	9
19	Balochi	Pakistan	25	0	4
20	Brahui	Pakistan	25	2	2
21	Makrani	Pakistan	25	0	12
22	Sindhi	Pakistan	25	0	4
23	Pathan	Pakistan	24	2	0
24	Burusho	Pakistan	25	0	4
25	Hazara	Pakistan	22	0	0
26	Uygur	China	9	0	0
27	Kalash	Pakistan	25	0	0
29	Han	China	43	0	1
30	Dai	China	10	0	0
31	Daur	China	10	0	0
32	Hezhen	China	10	0	0
33	Lahu	China	10	0	0
34	Miaozu	China	10	0	0
35	Oroqen	China	10	0	0
36	She	China	10	0	0
37	Tujia	China	8	0	0
38	Tu	China	10	0	0
39	Xibo	China	9	0	0
40	Yizu	China	10	0	0
41	Mongolian	China	10	0	5
42	Naxi	China	10	0	0
43	Cambodian	Cambodia	10	0	0
44	Japanese	Japan	30	0	0
45	Yakut	Russia (Siberia)	24	0	2
46	Melanesian	Bougainville	22	0	0
47	Papuan	New Guinea	17	0	0
48	Karitiana	Brazil	24	0	0
49	Surui	Brazil	21	0	0
50	Colombian	Colombia	13	0	0
51	Maya	Mexico	24	0	0
52	Pima	Mexico	25	0	0

eight the total number of replacement changes at *GDF8* that have been reported to date (table 1) and contributes further to the observed excess of replacement polymorphisms among humans.

Although positive selection may increase the level of polymorphism at a locus, the excess of replacement changes seen at *GDF8* could, in principle, alternatively be explained either by a recent relaxation of selective constraint or by the presence of slightly deleterious variants among humans.[22,23] However, both of these explanations seem unlikely for *GDF8*. Relaxed selection at *GDF8* is improbable, in view of the overall strong conservation of this gene over deep evolutionary timescales and the major phenotypic effect associated with loss of function in humans.[6]

The two major replacement variants at *GDF8* are unlikely to be slightly deleterious, because these are at relatively high frequency (up to 31%) in sub-Saharan Africans. Moreover, the atypical long-range haplotype conservation associated with haplogroups 153 and 55 suggests that these variants have rapidly increased in frequency. Although demographic processes associated with population bottlenecks and expansions may create long-range haplotype patterns that mimic a signature of selection,[24] this is not likely to be the case at *myostatin,* since the signature of selection is seen in the ancestral African population. Together, these data argue that some form of recent diversifying selection has played a significant role in shaping patterns of variation at *GDF8*.

The molecular positions of polymorphic residues 55 and 153 within the human GDF8 peptide allow us to speculate about the phenotypic consequences of these variants. Both residues are found within the propeptide region (residues 1–266) of GDF8. As is characteristic of other members of the transforming growth factor beta (TGF-β) superfamily, the GDF8 precursor peptide is cleaved into an (N-terminus) propeptide and a (C-terminus) mature peptide. The active form of myostatin is a homodimer of the mature peptide, which binds to extracellular activin type II receptors (ACTRIIB [MIM 602730]) to induce intracellular activation of SMAD proteins.[25] Importantly, the propeptide of GDF8 binds to the mature homodimer to form a latent myostatin complex and thus regulates GDF8 activity by preventing the homodimer from binding to its target receptors.[26] Concordantly, overexpression of GDF8 propeptide in transgenic mice causes muscle hypertrophy and hyperplasia similar to that in GDF8-null mutants.[26,27] Moreover, intraperitoneal administration of myostatin propeptide to *Mdx* mice (models for Duchenne muscular dystrophy) has been shown to rescue some of the muscular pathophysiological effects found in this mutant.[28] Interestingly, residue 55 is within a major inhibitory domain of the GDF8 propeptide (residues 42–115)[29] and therefore may influence the regulatory properties of the propeptide. In general, any mutations that increase the binding affinity between the propeptide and the mature peptide could generate a relative deficiency of myostatin activity. One of the many possible adaptive implications of such an effect could be protection from muscle wasting in times of famine, a potentially recurrent phenomenon for early agricultural societies.[30]

The evidence of positive selection on *GDF8* in humans implies that some of the replacement changes cause phenotypic changes related to muscle development and/or growth. Laboratory studies that measured human adult muscle response after short-term physical training failed to detect associations between increased muscle mass and several common amino acid polymorphisms.[10] However, the potential phenotypic effect on preadult muscle development remains unknown. Also, natural selection may act on fitness differences that are subtle and not easily detected in laboratory settings. A separate evolutionary analysis has shown accelerated evolution at GDF8 along three bovid lineages in the propeptide region of GDF8, which demonstrates that *GDF8* has been a target of selection along another branch in mammalian evolution.[31] Although the phenotypic target of selection remains elusive in humans, transgenic mouse models and functional assays may facilitate the quantification of specific traits associated with the *GDF8* replacement changes identified here.

Acknowledgments

We thank B. A. Payseur, E. T. Wood, H. E. Hoekstra, and M. Slatkin, for helpful discussions. Two anonymous reviewers provided useful comments. This work was supported by a UNCF-Merck Science Initiative postdoctoral fellowship (to M.A.S.).

Appendix A

The figure is available in its entirety in the online edition of *The American Journal of Human Genetics*.

Figure A1. Table of polymorphism for diploid data for *GDF8*. The legend is available in its entirety in the online edition of *The American Journal of Human Genetics*.

Table A1. Summary Statistics of Nucleotide Variability for *GDF8*

Geographic Region and Sequence Type[a]	No. of Segregating Sites	No. of Changes		Nucleotide Diversity [%] (SD)		Tajima's *D*	Fu and Li's *D*	Divergence (%)[b]	
		Nonsynonymous	Synonymous	θ_π	θ_W			*Homo-Pan*	*Homo-Pongo*
Africa ($n = 76$):									
Coding	7	5	2	.061 (.007)	.111 (.048)	−.9963	.2388		
Noncoding	2	NA	NA	.015 (.004)	.036 (.026)	−.8283	.6547		
Total	9	5	2	.040 (.040)	.076 (.030)	−1.3124	.4935		
Europe ($n = 70$):									
Coding	3	2	1	.009 (.003)	.048 (.029)	−1.3887	−2.1189		
Noncoding	1	NA	NA	.001 (.001)	.018 (.018)	−.9906	−2.1156		
Total	4	2	1	.005 (.002)	.034 (.018)	−1.5858	−2.8351[c]		
Combined ($N = 146$)									
Coding	8	5	3	.038 (.004)	.114 (.046)	−1.4116	−.6140	.198	.717
Noncoding	2	NA	NA	.009 (.002)	.032 (.023)	−.9427	.6140	.813	2.060
Total	10	5	3	.024 (.003)	.076 (.028)	−1.5237	−.3207	.486	1.345

NOTE.—A total of 2,114 (1,125 coding; 989 noncoding) base pairs were analyzed. NA = not applicable.

[a] The noncoding sequence includes the surveyed region P (which includes the *GDF8* promoter region) and the intron sequence flanking the exons in the surveyed regions.

[b] Divergence estimates are based on the average of all pairwise comparisons between the human samples and respective outgroup samples (GenBank accession numbers DQ927199–DQ927203 and DQ927192–DQ927194).

[c] $P < .05$.

Appendix B

Table B1. Silent and Replacement Changes at *GDF8,* Relative to the *Homo sapiens* Consensus Sequence

Species	Common Name	Accession Number[a]	No. of Changes		d_N:d_S
			Silent	Replacement	
Homo sapiens	**Human**	**NM_005259**	**3**	**5**	**.663**
Pan troglodytes	Common chimpanzee	ENSPTRT00000023612[b]	1	1	.267
Pan paniscus	Pygmy chimpanzee	DQ927196[c]	1	1	.267
Gorilla gorilla	Gorilla	DQ927204[c]	1	1	.267
Pongo pygmeus	Orangutan	DQ927192[c]	6	1	.045
Papio hamadryas	Baboon	AF019619	7	2	.096
Macaca mulatta	Rhesus macaque	DQ927195[c]	4	1	.085
Macaca fascicularis	Crab-eating macaque	AY055750	8	1	.035
Mus musculus	Mouse	NM_010834	77	15	.051
Rattus norvegicus	Rat	NM_019151	86	17	.055
Canis familiaris	Dog	AY367768	77	15	.060
Vulpes vulpes	Red fox	AY647144	81	16	.058
Alopex lagopus	Arctic fox	AY606017	81	16	.058
Sus scrofa	Pig	AF019623	42	8	.056
Bos taurus	Cow	NM_001001525	69	23	.099
Bubalis bubalis	Water buffalo	AH013313	64	28	.130
Ovis aries	Sheep	AF019622	59	25	.128
Capra hircus	Goat	AY436347	60	27	.134
Equus caballus	Horse	AB033541	38	9	.072
Meleagris gallopavo	Turkey	AF019625	131	31	.050
Gallus gallus	Chicken	ENSGALT00000003669[b]	136	31	.049
Coturnix coturnix	Common quail	AF407340	130	33	.055
Coturnix chinensis	Painted Chinese quail	AF440864	130	33	.056
Danio rerio	Zebrafish	ENSDART00000001757[b]	233	159	.031

NOTE.—Bold text for the human data indicates polymorphisms.

[a] All are GenBank accession numbers unless otherwise noted.

[b] Ensembl accession number.

[c] Sequence generated for the current study.

Appendix C

Estimation of the Ages of Selected Alleles

We employed an approach described elsewhere[14] to roughly estimate the ages of mutations 163 and 2246 (representing alleles Ala55Thr and Lys153Arg, respectively). Long-range haplotypes, from the HapMap data, around *GDF8* in the YRI panel were used to calculate the EHH statistic (see the text and fig. 2*A*). By definition, EHH ≈ Pr (Homozygosity) = e^{-2rg}, where Pr(Homozygosity) is the probability that two chromosomes are homozygous at recombination distance r from the core, given identity by decent from a common ancestor g generations ago. With use of a generation time of 25 years for humans, $g = t/25$, where t is time in years.

For the core haplotype bearing mutation 163,

$$2r \approx 0.265\% \text{ and } \Pr[\text{Homozygosity}] = 0.35 ,$$

so

$$\text{t} = \frac{-(25)(100)\ln(0.35)}{0.265} \approx 10{,}000 \text{ years} ,$$

and, for the core haplotype-bearing mutation 2246,

$$2r \approx 0.246\% \text{ and } \Pr(\text{Homozygosity}) = 0.45 ,$$

so

$$t = \frac{-(25)(100)\ln(0.45)}{0.246} \approx 8{,}000 \text{ years} .$$

These estimates are crude, since the theoretical models for estimating ages of multiple alleles under selection in a population are not well developed. The current method assumes a starlike genealogy for the selected haplotypes; therefore, these results are likely underestimates of the true ages of the alleles.

Web Resources

Accession numbers and URLs for data presented herein are as follows:

Ensembl, http://www.ensembl.org/

GenBank, http://www.ncbi.nih.gov/Genbank/ (for *GDF8* [accession numbers DQ927046–DQ927191], *P. troglodytes* [accession numbers DQ927199–DQ927203], *H. sapiens* [accession number NM_005259], *P. paniscus* [accession number DQ927196], *G. gorilla* [accession number DQ927204], *P. pygmeus* [accession numbers DQ927192–DQ927194], *P. hamadryas* [accession number AF019619], *M. mulatta* [accession number AY055750], *M. musculus* [accession number NM_010834], *R. norvegicus* [accession number NM_019151], *C. familiaris* [accession number AY367768], *V. vulpes* [accession number AY647144], *A. lagopus* [accession number AY606017], *S. scrofa* [accession number AF019623], *B. taurus* [accession number NM_001001525], *B. bubalis* [accession number AH013313], *O. aries* [accession number AF019622], *C. hircus* [accession number AY436347], *E. caballus* [accession number AB033541], *M. gallopavo* [accession number AF019625], *C. coturnix* [accession number AF407340], and *C. chinensis* [accession number AF440864])

International HapMap Project, http://www.hapmap.org/

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for *GDF8, G6PD,* and ACTRIIB)

SWEEP, http://www.broad.mit.edu/mpg/sweep/download.html

References

1. Gordon ES, Dressman HAG, Hoffman EP (2005) The genetics of muscle atrophy and growth: the impact and implications of polymorphisms in animals and humans. Int J Biochem Cell Biol 37:2064–2074
2. Thomis MAI, Huygens W, Heuninckx S, Chagnon M, Maes HHM, Claessens AL, Vlietinck R, Bouchard C, Beunen GP (2004) Exploration of myostatin polymorphisms and the angiotensin-converting enzyme insertion/deletion genotype in responses of human muscle to strength training. Eur J Appl Physiol 92:267–274
3. McPherron AC, Lawler AM, Lee SJ (1997) Regulation of skeletal muscle mass in mice by a new TGF-β superfamily member. Nature 387:83–90
4. McPherron AC, Lee SJ (1997) Double muscling in cattle due to mutations in the myostatin gene. Proc Natl Acad Sci USA 94:12457–12461
5. Kambadur R, Sharma M, Smith TPL, Bass JJ (1997) Mutations in *myostatin* (*GDF8*) in double-muscled Belgian blue and Piedmontese cattle. Genome Res 7:910–916
6. Schuelke M, Wagner KR, Stolz LE, Hubner C, Riebel T, Komen W, Braun T, Tobin JF, Lee SJ (2004) Myostatin mutation associated with gross muscle hypertrophy in a child. N Engl J Med 350:2682–2688
7. Gonzalez-Cadavid NF, Taylor WE, Yarasheski K, Sinha-Hikim I, Ma K, Ezzat S, Shen RQ, Lalani R, Asa S, Mamita M, Nair G, Arver S, Bhasin S (1998) Organization of the human myostatin gene and expression in healthy men and HIV-infected men with muscle wasting. Proc Natl Acad Sci USA 95:14938–14943
8. Ma K, Mallidis C, Artaza J, Taylor W, Gonzalez-Cadavid N, Bhasin S (2001) Characterization of 5′-regulatory region of human myostatin gene: regulation by dexamethasone in vitro. Am J Physiol Endocrinol Metab 281:E1128–E1136
9. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989
10. Ferrell RE, Conte V, Lawrence EC, Roth SM, Hagberg JM, Hurley BF (1999) Frequent sequence variation in the human myostatin (GDF8) gene as a marker for analysis of muscle-related phenotypes. Genomics 62:203–207
11. McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *adh* locus in *Drosophila*. Nature 351:652–654
12. Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556
13. Mouse Genome Sequencing Consortium (2002) Initial sequencing and comparative analysis of the mouse genome. Nature 420:520–562
14. Voight BF, Kudaravalli S, Wen XQ, Pritchard JK (2006) A map of recent positive selection in the human genome. PloS Biol 4:446–458

15. Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatkowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837
16. Saunders MA, Slatkin M, Garner C, Hammer MF, Nachman MW (2005) The extent of linkage disequilibrium caused by selection on *G6PD* in humans. Genetics 171:1219–1229
17. International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320
18. Cavalli-Sforza LL (2005) The Human Genome Diversity Project: past, present and future. Nat Rev Genet 6:333–340
19. Toomajian C, Ajioka RS, Jorde LB, Kushner JP, Kreitman M (2003) A method for detecting recent selection in the human genome from allele age estimates. Genetics 165:287–297
20. Verrelli BC, McDonald JH, Argyropoulos G, Destrol-Bisol G, Froment A, Drousiotou A, Lefranc G, Helal AN, Loiselet J, Tishkoff SA (2002) Evidence for balancing selection from nucleotide sequence analyses of human *G6PD*. Am J Hum Genet 71:1112–1128
21. Hughes AL, Nei M (1989) Nucleotide substitution at major histocompatibility complex class-II loci: evidence for overdominant selection. Proc Natl Acad Sci USA 86:958–962
22. Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, Civello D, Adams MD, Cargill M, Clark AG (2005) Natural selection on protein-coding genes in the human genome. Nature 437:1153–1157
23. Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87
24. Currat M, Excoffier L, Maddison W, Otto SP, Ray N, Whitlock MC, Yeaman S (2006) Comment on "Ongoing adaptive evolution of *ASPM,* a brain size determinant in *Homo sapiens*" and "*Microcephalin,* a gene regulating brain size, continues to evolve adaptively in humans." Science 313:172
25. Yingling JM, Blanchard KL, Sawyer JS (2004) Development of TGF-β signaling inhibitors for cancer therapy. Nat Rev Drug Discov 3:1011–1022
26. Lee SJ, McPherron AC (2001) Regulation of myostatin activity and muscle growth. Proc Natl Acad Sci USA 98:9306–9311
27. Yang JZ, Ratovitski T, Brady JP, Solomon MB, Wells KD, Wall RJ (2001) Expression of myostatin pro domain results in muscular transgenic mice. Mol Reprod Dev 60:351–361
28. Bogdanovich S, Perkins KJ, Krag TOB, Whittemore SA, Khurana TS (2005) Myostatin propeptide-mediated amelioration of dystrophic pathophysiology. FASEB J 19:543–549
29. Jiang MS, Liang LF, Wang SS, Ratovitski T, Holmstrom J, Barker C, Stotish R (2004) Characterization and identification of the inhibitory domain of GDF-8 propeptide. Biochem Biophys Res Commun 315:525–531
30. Wells JCK (2006) The evolution of human fatness and susceptibility to obesity: an ethological approach. Biol Rev 81: 183–205
31. Tellgren A, Berglund AC, Savolainen P, Janis CM, Liberles DA (2004) Myostatin rapid sequence evolution in ruminants predates domestication. Mol Phylogenet Evol 33:782–790