Natural Selection and Population History in the Human Angiotensinogen Gene (*AGT*): 736 Complete *AGT* Sequences in Chromosomes from Around the World

Toshiaki Nakajima,[1] Stephen Wooding,[4] Takuro Sakagami,[1] Mitsuru Emi,[5] Katsushi Tokunaga,[2] Gen Tamiya,[6] Tomoaki Ishigami,[7] Satoshi Umemura,[7] Batmunkh Munkhbat,[6,8] Feng Jin,[9] Jia Guan-jun,[10] Ikuo Hayasaka,[11] Takafumi Ishida,[3] Naruya Saitou,[12] Karel Pavelka,[13] Jean-Marc Lalouel,[4] Lynn B. Jorde,[4] and Ituro Inoue[1]

[1]Division of Genetic Diagnosis, The Institute of Medical Science, [2]Department of Human Genetics, Graduate School of Medicine, and [3]Unit of Human Biology and Genetics, School of Science, University of Tokyo, Tokyo; [4]Department of Human Genetics, University of Utah Health Sciences Center, Salt Lake City; [5]Department of Molecular Biology, Institute of Gerontology, Nippon Medical School, Kawasaki, Japan; [6]Department of Molecular Life Science, Tokai University School of Medicine, Isehara, Japan; [7]Internal Medicine, Yokohama City University, Yokohama; [8]Central Scientific Research Laboratory, National Institute of Medicine, Ulaanbaatar, Mongolia; [9]Institute of Genetics and Developmental Biology, Chinese Academy of Science, Beijing; [10]Red Cross Blood Center of Harbin, Harbin, China; [11]Kumamoto Primates Park, Sanwa Kagaku Kenkyusho Co. Ltd., Kumamoto, Japan; [12]Division of Population Genetics, National Institute of Genetics, Mishima, Japan; and [13]Institute of Rheumatology, Prague

Several lines of evidence suggest that patterns of genetic variability in the human angiotensinogen gene (*AGT*) contribute to phenotypic variability in human hypertension. The A(−6) promoter variant of *AGT* is associated with higher plasma angiotensinogen levels and increased risk of essential hypertension. The geographic distribution of the A(−6) variant leads to the intriguing hypothesis that the G(−6) promoter variant has been selectively advantageous outside Africa. To test these hypotheses, we investigated the roles of population history and natural selection in shaping patterns of genetic diversity in *AGT*, by sequencing the entire *AGT* gene (14,400 bp) in 736 chromosomes from Africa, Asia, and Europe. We found that the A(−6) variant is present at higher frequency in African populations than in non-African populations. Neutrality tests found no evidence of a departure from selective neutrality, when whole *AGT* sequences were compared. However, tests restricted to sites in the vicinity of the A(−6)G polymorphism found evidence of a selective sweep. Sliding-window analyses showed that evidence of the sweep is restricted to sites in tight linkage disequilibrium (LD) with the A(−6)G polymorphism. Further, haplotypes carrying the G(−6) variant showed elevated levels of LD, suggesting that they have risen recently to high frequency. Departures from neutral expectation in some but not all regions of *AGT* indicate that patterns of diversity in the gene cannot be accounted for solely by population history, which would affect all regions equally. Taken together, patterns of genetic diversity in *AGT* suggest that natural selection has generally favored the G(−6) variant over the A(−6) variant in non-African populations. However, important localized effects may also be present.

Introduction

The incidence of hypertension varies widely among populations with different geographic and ethnic origins. African Americans, for example, are at substantially greater risk for hypertension than are Americans of European descent (Burt et al. 1995). Some aspects of variation in hypertension susceptibility can be explained by variation in environmental factors like diet. However, genetic factors may also be involved (Lifton 1996). One hypothesis that explains population differences in hypertension susceptibility is the “sodium retention hypothesis,” which posits that tropical and temperate populations are adapted to different levels of sodium salt availability (Gleiberman 1973). Under this hypothesis, ancient human populations residing in hot, humid areas (i.e., the tropics) evolved the tendency to retain salt as an adaptation to low salt availability and the hazards of electrolyte imbalance. In contrast, populations in cooler, drier climates (i.e., the temperate zones) adapted to conditions of greater sodium availability and less sodium loss. Salt regulation is a key component of blood pressure homeostasis, and variable sodium sensitivity could explain the prevalence of hypertension in some populations. Thus, variation in the human genes underlying salt regulation might help to account for some of the regional differences in hypertension susceptibility.

Received August 4, 2003; accepted for publication February 25, 2004; electronically published April 9, 2004.

Address for correspondence and reprints: Dr. Toshiaki Nakajima, Division of Genetic Diagnosis, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. E-mail: tnakajim@ims.u-tokyo.ac.jp

0002-9297/2004/7405-0012$15.00

Salt regulation is controlled largely by the renin-angiotensin-aldosterone system (RAS), which controls rates of sodium excretion and reabsorption in the human kidney. The RAS is driven by a negative-feedback relationship between renin and blood pressure. The rate-limiting step in this system is the cleavage of angiotensinogen (AGT) by renin to produce angiotensin I (A-I) (Sealey and Laragh 1990). A-I is, in turn, processed by the angiotensin-converting enzyme to produce angiotensin II (A-II), which binds to the angiotensin receptor to increase salt retention by direct and indirect mechanisms. Polymorphisms in RAS genes have been associated with essential hypertension in several previous studies (Luft 2001). Among these, polymorphisms in *AGT* (MIM 106150) are regarded as some of the most promising candidates in the search for variants underlying human hypertension (Luft 2001).

It has been demonstrated that blood pressure increases after the injection of AGT (Menard et al. 1991) and that mice with multiple introduced copies of *AGT* show plasma levels correlated with the number of gene copies (Smithies and Kim 1994; Takahashi and Smithies 1999). In addition, two variants of *AGT,* A(−6) and T235, are in tight linkage disequilibrium (LD) and are associated with essential hypertension (EHT [MIM 145500]) and high plasma AGT (Jeunemaitre et al. 1992; Inoue et al. 1997). Further, *AGT* expression is affected by the A(−6)G polymorphism in vitro, with the A(−6) variant conferring 20% greater expression levels (Inoue et al. 1997). These variants are also associated with preeclampsia (Ward et al. 1993). Finally, the T235 variant has consistently been found at higher frequencies in African populations than in non-African populations (Corvol and Jeunemaitre 1997; Nakajima et al. 2002).

Both the functional effects of the A(−6)G polymorphism and the geographically disparate distribution of the G(−6)/M235 allele suggest the hypothesis that African and non-African populations are adapted to different salt retention requirements and that the G(−6)/M235 allele has increased in frequency as part of a selective sweep. However, differences in allele frequency are often found between African and non-African populations at genetic loci that are likely to be selectively neutral—a pattern accounted for by a founder effect in the populations that first left Africa (Reich et al. 2002; Tishkoff and Verrelli 2003; Watkins et al. 2003). Thus, differences in G(−6)/M23–allele frequency might be due to the simple, random assortment of alleles that occurred when human populations first left Africa rather than to evolutionary adaptation. These alternatives are distinguished by the extent of their effects. Although population history should affect all genomic regions equally, the effects of natural selection will probably be localized to specific regions.

To test the hypothesis that the G(−6) allele recently swept to high frequency and that this sweep occurred primarily in populations outside Africa, we studied patterns of DNA sequence variation in a 14,400-bp segment of *AGT* in a panel of 368 individuals from 16 populations in Africa, Asia, and Europe. Patterns of variation in these sequences were tested for the effects of a selective sweep in African and non-African populations across the entire gene, as well as within specific regions. The results of these tests indicate that the G(−6)/M235 allele has increased in frequency as the result of a selective sweep.

Methods

Population Samples

DNA samples were collected from four African populations (32 Pygmy, 7 Alur, 18 Nande, and 17 Hema), six Eurasian populations (24 living in Utah [CEPH], 24 Czech, 24 Druze, 24 Ashkenazi, 24 Palestinian, and 30 Indian/Pakistani), and six East Asian populations (24 Southern Chinese [Fujian], 24 Southern Chinese [Guangdong], 24 Northern Chinese [Heilongjian], 24 Korean, 24 Mongolian [Khalkh], and 24 Japanese), for a total of 368 people (736 chromosomes). DNA samples from Druze, Ashkenazi, and Palestinian subjects were obtained from the National Laboratory for the Genetics of Israeli Populations (Tel Aviv University). Nine Indian/Pakistani samples were obtained from the Coriell Institute for Medical Research. DNA samples from three ape species—one common chimpanzee (*Pan troglodytes verus*), one gorilla (*Gorilla gorilla*), and one orangutan (*Pongo pygmaeus*)—were also analyzed.

Identification and Genotyping of Nucleotide Variations

Overlapping primer sets covering the entire *AGT* gene (GenBank accession numbers NM_000029 and X15323) were designed on the basis of size and overlap of PCR amplicons (Nakajima et al. 2002). Genomic DNA was subjected to PCR amplification, followed by sequencing by use of the BigDye Terminator cycle. Polymorphisms were identified using the Sequencher program (Gene Code). Each polymorphism was confirmed by reamplifying and resequencing from the same or the opposite strand.

Fluorescence-labeled primers were designed to amplify the microsatellite region located downstream of exon 5 (Jeunemaitre et al. 1992). PCR amplicons were electrophoresed on an ABI 377 Prism automated DNA sequencer (Applied Biosystems). DNA fragment size was determined by GENESCAN software, version 3.1, and GENOTYPER software, version 2.5 (PE Biosystems).

Table 1

Sequence Variations in AGT among 16 Population Samples

				African				Eurasian						East Asian						
																	Southern Chinese			
ID of Sequence Variation	Position	Singleton	Insertion/ Deletion	Pygmy	Alur	Nanda	Hema	Utah (CEPH)	Palestinian	Czech	Druze	Ashkenazi	Indian/ Pakistani	Japanese	Korean	Mongolian	Fujian	Guandong	Northern Chinese (Heilongjian)	Fst
1	−1178			.172	.214	.25	.088	.062	.104	.146	.208	.125	.25	.208	.229	.042	.167	.146	.083	.032
2	−1173		I/D	0	.143	.028	.029	0	0	0	0	0	0	0	0	0	0	0	0	.099
3	−1103	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
4	−1074			.828	.714	.75	.912	.937	.896	.854	.792	.875	.75	.792	.771	.958	.833	.854	.917	.038
5	−963	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
6	−837	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
7	−834	S		0	0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	.020
8	−829			0	0	0	0	.062	0	.042	0	.062	0	0	0	.021	0	0	0	.042
9	−812			.156	0	.167	0	0	.021	0	0	0	0	0	0	0	0	0	0	.134
10	−792			.828	.714	.75	.912	.937	.896	.854	.792	.875	.75	.792	.771	.958	.833	.854	.917	.038
11	−775			.281	.143	.278	.353	0	.187	.042	.146	.104	.117	.104	.104	.083	.146	.167	.125	.061
12	−604		I/D	.016	0	.056	0	0	0	0	0	0	0	0	0	0	0	0	0	.043
13	−532			.922	.857	.917	.941	.937	.917	.854	.792	.875	.75	.792	.771	.958	.833	.854	.917	.035
14	−411	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
15	−385	S		0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	0	.020
16	−260			0	0	0	0	.021	.021	0	0	0	0	0	0	0	0	0	0	.018
17	−245	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
18	−229	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
19	−217			.219	.214	.417	.088	.062	.125	.146	.208	.125	.25	.208	.229	.042	.167	.146	.083	.056
20	−189	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
21	−165			.094	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.072
22	−152			0	0	0	.029	0	.021	0	.062	.042	.067	.146	.021	.104	.021	.083	.042	.045
23	−20			.875	.929	.778	.794	.854	.896	.729	.792	.771	.75	.75	.896	.792	.75	.854	.854	.025
24	−15	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
25	−6			.062	.071	0	.294	.708	.562	.5	.396	.458	.25	.229	.125	.271	.187	.167	.375	.176
26	67			.109	.071	.028	.529	.708	.687	.521	.417	.458	.267	.25	.146	.312	.208	.208	.417	.185
27	172			.187	.357	.139	.059	.083	.042	.062	.062	.062	.1	.25	.521	.458	.375	.437	.292	.153
28	182	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	0	.020
29	266			0	0	0	0	.042	.021	0	0	0	0	0	0	0	0	0	0	.031
30	384			.297	.5	.389	.147	.062	.167	.146	.208	.125	.3	.229	.229	.042	.167	.167	.104	.082
31	409			.391	.571	.556	.147	.062	.167	.167	.208	.125	.3	.229	.229	.042	.167	.167	.104	.131
32	507			.172	.143	.028	.412	.708	.687	.5	.417	.583	.267	.25	.146	.312	.208	.208	.417	.169
33	511			0	0	0	0	0	0	0	0	0	0	0	0	.042	0	0	0	.039
34	515	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
35	577	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
36	654	S		0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	.020

37	676			.578	.857	.861	.382	.292	.312	.5	.542	.417	.7	.75	.854	.687	.792	.75	.542	.151
38	698			.578	.857	.861	.382	.292	.312	.5	.542	.417	.7	.75	.854	.687	.792	.75	.542	.151
39	770			.266	0	.278	.382	.146	.083	.437	.312	.229	.333	.271	.104	.229	.646	.187	.187	.116
40	892	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
41	934	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
42	934		I/D	.031	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.026
43	945			.062	0	.056	0	0	0	0	0	0	0	0	0	0	0	0	0	.052
44	992			.031	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
45	1029			.031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.029
46	1035			.109	0	.306	.176	.229	.146	.333	.333	.25	.4	.521	.625	.646	.625	.583	.437	.166
47	1101	S		0	0	0	0	0	0	0	0	0	0	0	.021	0	0	0	0	.020
48	1164			.547	.857	.667	.324	.146	.25	.208	.375	.187	.483	.646	.771	.562	.562	.687	.437	.178
49	1216	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
50	1257			.016	0	.222	0	0	.021	0	0	0	0	0	0	0	0	0	0	.180
51	1319	S		0	0	0	0	0	0	0	0	0	0	.021	0	0	0	0	0	.020
52	1321			0	0	0	0	0	0	0	0	0	0	.021	.021	.083	.042	0	.083	.051
53	1386			0	0	0	0	0	0	0	0	0	0	.083	.083	.083	.125	.021	.104	.066
54	1418			.031	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
55	1441	S		0	0	0	0	0	0	0	0	0	0	.021	0	0	0	0	0	.020
56	1490	S		0	0	0	0	0	0	0	0	0	0	0	0	.021	0	0	0	.020
57	1509		I/D	.031	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
58	1677		I/D	0	.286	.028	.059	0	0	0	0	0	0	0	0	0	0	0	0	.212
59	1687			0	0	0	0	0	0	0	0	0	0	0	0	0	0	.042	.021	.031
60	1713			.062	.143	.111	0	0	0	0	0	0	0	0	0	0	0	0	0	.098
61	1851			0	0	0	0	.021	.042	.104	.146	.021	.1	0	0	.042	0	0	0	.070
62	1949			0	0	0	0	0	0	0	0	0	0	.021	.021	0	0	0	0	.018
63	2077			.062	0	.083	0	.083	.062	.062	.042	.062	.1	.25	.458	.396	.354	.437	.292	.177
64	2180			.516	.214	.361	.353	0	.167	0	.146	.104	.15	0	.083	.083	0	.167	.125	.156
65	2237	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
66	2310			.016	0	.111	.147	.146	.083	.187	.187	.229	.217	.25	.104	.208	.25	.146	.146	.041
67	2325	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
68	2343			.125	.357	.083	.059	0	0	0	0	0	0	0	0	0	0	0	0	.215
69	2356	S		0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	.020
70	2394	S		0	0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	.020
71	2395			0	0	0	0	0	0	0	0	0	0	.021	0	0	.021	.062	.021	.034
72	2420			.031	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
73	2593			0	0	0	0	.042	.083	0	.042	0	0	0	0	0	0	0	0	.052
74	2622			0	0	.083	0	.083	.042	.062	.042	.062	.1	.229	.458	.396	.333	.354	.271	.174
75	2636	S		0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	0	.020
76	2762	S		0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	0	.020
77	2819		I/D	.031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.029
78	2842	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	.020
79	2951			.031	0	.139	0	0	.021	0	0	0	0	0	0	0	0	0	0	.098
80	3023			1	1	1	1	.979	1	1	1	.979	1	1	1	1	1	1	1	.018
81	3187			.031	0	.083	0	.083	.042	.062	.042	.062	.1	.229	.458	.396	.333	.375	.271	.171
82	3274	S		0	.071	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.067
83	3420	S		0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	.020
84	3517			.078	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.073
85	3542	S		0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	0	.020
86	3561			0	.071	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.053
87	3599			.047	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.044
88	3647	S		0	0	0	0	0	0	0	0	0	0	0	0	0	.021	0	0	.020
89	3679	S		0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	0	.020
90	3680	S		0	0	0	0	0	0	0	0	0	0	0	.021	0	0	0	0	.020
91	3721	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
92	3889			.016	0	.111	.088	.146	.062	.25	.146	.187	.183	.083	.062	.104	.146	.062	.104	.041
93	4000	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
94	4072			.047	.071	0	.294	.708	.562	.479	.375	.479	.25	.229	.125	.271	.187	.167	.375	.178
95	4219			0	.143	.028	.029	0	0	0	0	0	0	0	0	0	0	0	0	.099
96	4275			.078	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.073
97	4295			0	0	0	0	0	0	0	0	0	.05	.042	.042	.125	.125	.062	.208	.092

Table 1 (continued)

				African				Eurasian						East Asian						
																	Southern Chinese			
ID of Sequence Variation	Position	Singleton	Insertion/ Deletion	Pygmy	Alur	Nanda	Hema	Utah (CEPH)	Palestinian	Czech	Druze	Ashkenazi	Indian/ Pakistani	Japanese	Korean	Mongolian	Fujian	Guandong	Northern Chinese (Heilongjian)	Fst
98	4311			0	0	0	0	.062	0	.021	0	0	0	0	0	0	0	0	0	.047
99	4481			.016	0	.056	.029	0	.021	0	0	0	0	0	0	0	0	0	0	.031
100	4527			.016	0	.083	0	0	0	0	0	0	0	0	0	0	0	0	0	.066
101	4726	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
102	4873			.5	.643	.667	.353	.292	.25	.5	.458	.417	.6	.667	.771	.646	.667	.667	.5	.088
103	5034	S		0	.071	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.067
104	5063	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
105	5093			.078	.143	.056	.324	.708	.562	.479	.375	.479	.25	.229	.125	.271	.187	.167	.375	.153
106	5175			.125	0	.083	.029	0	0	0	0	0	0	0	0	0	0	0	0	.085
107	5229	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
108	5259			0	0	0	0	.021	.062	.042	0	.021	0	0	0	.021	0	0	0	.031
109	5342			0	0	0	0	0	0	0	0	0	0	0	0	0	.021	.021	.021	.017
110	5417			.016	.071	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.046
111	5433			.016	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.021
112	5469	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
113	5485			0	0	0	0	0	0	0	0	0	0	.021	0	0	.021	.062	.021	.034
114	5555			.125	.214	.028	.206	.708	.542	.479	.375	.479	.25	.229	.125	.271	.187	.167	.375	.147
115	5592			.125	.214	.056	.294	.708	.542	.479	.375	.479	.25	.229	.125	.271	.187	.167	.375	.138
116	5657			0	0	.028	0	0	0	0	0	0	.033	0	0	0	0	0	0	.027
117	5669			.375	.429	.389	.471	.146	.271	.25	.25	.292	.317	.187	.167	.187	.292	.25	.229	.042
118	5684	S		0	0	0	0	0	0	0	0	0	0	0	.021	0	0	0	0	.020
119	5744			.516	.5	.389	.559	.146	.292	.271	.354	.333	.367	.333	.187	.292	.312	.312	.271	.051
120	5754	S		0	0	0	0	0	0	0	0	0	0	.021	0	0	0	0	0	.020
121	5755	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	0	.020
122	5757	S	I/D	0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
123	5783			.031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.029
124	5835			0	0	0	0	0	.062	0	0	0	0	0	0	0	0	0	0	.058
125	5877			0	0	0	0	0	0	0	0	0	0	.021	0	0	0	0	.021	.018
126	5949	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
127	6002	S		0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	.020
128	6058	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
129	6065			.359	.357	.528	.118	.146	.167	.187	.271	.167	.383	.479	.687	.437	.437	.542	.333	.113
130	6151			.141	.214	.222	.059	.062	.104	.146	.229	.125	.25	.229	.208	.042	.146	.146	.042	.038
131	6184			.031	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.026
132	6207	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
133	6222			0	.071	.083	0	0	0	0	0	0	.033	0	0	0	0	0	0	.059
134	6232			.625	.643	.444	.882	.854	.833	.792	.729	.812	.617	.521	.312	.562	.521	.458	.687	.114
135	6265			.047	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.035
136	6285			.047	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.044
137	6308			.406	.429	.444	.294	.771	.667	.625	.521	.604	.55	.417	.333	.292	.333	.271	.458	.083
138	6319			.016	0	0	.059	0	0	0	0	0	0	0	0	0	0	0	0	.045
139	6359	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
140	6360			0	.214	.083	.118	0	.187	0	.062	.104	.033	0	0	0	0	0	0	.101
141	6419			.453	.643	.639	.529	.854	.896	.729	.729	.771	.683	.687	.812	.687	.687	.708	.75	.054
142	6427			.547	.357	.361	.471	.146	.104	.271	.271	.229	.317	.312	.187	.312	.312	.292	.25	.055
143	6441			0	0	0	.059	0	0	.021	.062	.042	.017	.125	.021	.104	.021	.062	.042	.040
144	6569			.125	0	.028	.029	0	0	0	0	0	0	0	0	0	0	0	0	.084
145	6597			0	0	.056	0	0	0	0	0	0	0	0	0	0	0	0	0	.053
146	6619	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
147	6626			0	0	0	0	.021	0	0	0	.021	0	0	0	0	0	0	0	.018
148	6745	S		0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	.020
149	7024			0	.071	.056	.088	0	.187	0	.062	.104	.033	0	0	0	0	0	0	.076

150	7108	S		0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	.020
151	7154	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	0	.020
152	7178	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
153	7359			.031	0	.083	0	0	0	0	0	0	.033	0	0	0	0	0	0	.052
154	7368			.016	0	.056	.029	.083	.042	.083	.042	.062	.1	.229	.5	.396	.333	.375	.292	.180
155	7388	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
156	7502			0	0	0	0	.021	0	0	0	.062	0	0	0	0	0	0	0	.047
157	7503	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
158	7537	S		0	0	0	0	0	0	0	0	0	0	0	.021	0	0	0	0	.020
159	7556			0	0	0	.059	0	0	0	0	0	0	0	0	0	0	0	0	.056
160	7676			.016	.071	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.056
161	7686	S		0	0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	.020
162	7692			.109	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.103
163	7797			0	.143	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.115
164	7886			.062	.143	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.107
165	7948	S		0	0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	.020
166	7974	S		0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	0	.020
167	8129	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
168	8178			0	0	0	0	.104	.062	.104	.104	0	.017	0	0	.021	0	0	0	.066
169	8248	S		0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	.020
170	8304			.75	.714	.861	.941	1	1	1	1	1	1	1	1	1	1	1	1	.191
171	8306			.531	.429	.528	.529	.854	.896	.771	.729	.771	.65	.646	.812	.687	.667	.667	.729	.072
172	8356			.156	.071	.056	.324	.792	.583	.583	.437	.542	.383	.417	.604	.646	.521	.521	.687	.175
173	8491	S		0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	.020
174	8532			0	0	.056	0	0	0	0	0	0	0	0	0	0	0	0	0	.053
175	8585	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
176	8716	S		0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	.020
177	8862			.125	0	.111	.324	.792	.583	.583	.437	.542	.383	.417	.604	.646	.521	.521	.687	.185
178	8918	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
179	8923			0	.214	.083	.088	.021	.187	0	.062	.104	.033	0	0	0	0	0	0	.096
180	9030	S	I/D	0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
181	9049			0	.071	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	.054
182	9164	S		0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	.020
183	9174	S		0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	.020
184	9196			0	0	0	0	0	0	0	0	0	0	0	0	.021	0	0	.021	.018
185	9359	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	.020
186	9372			.094	0	.028	.029	0	0	0	0	0	0	0	0	0	0	0	0	.060
187	9538	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
188	9561			0	0	0	.088	0	0	0	.021	0	0	0	0	0	0	0	0	.069
189	9562	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
190	9568			0	0	.056	0	0	0	0	0	0	0	0	0	0	0	0	0	.053
191	9596			.016	0	.056	.029	.062	.042	.083	.042	.062	.1	.229	.479	.396	.312	.375	.292	.176
192	9603	S	I/D	0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	.020
193	9668			.016	0	.056	.029	.062	.042	.083	.042	.062	.1	.229	.479	.396	.312	.375	.292	.176
194	9685			.031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.029
195	9738	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
196	9769			.016	0	.056	.029	.062	.042	.083	.042	.062	.1	.229	.479	.396	.312	.375	.292	.176
197	9811	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
198	9890			0	0	0	0	0	0	0	0	.042	0	0	0	0	0	0	0	.039
199	10044			0	0	0	0	.083	.125	.104	.083	.042	.017	0	0	.021	0	0	0	.062
200	10195	S		0	.071	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.067
201	10248	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
202	10382			.219	.286	.139	.059	0	0	0	0	0	0	0	0	0	0	0	0	.181
203	10409	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
204	10472	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
205	10544	S		0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	.020
206	10557			.031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.029
207	10561	S		0	0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	.020
208	10649	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
209	10672	S		0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	0	.020
210	10676	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	.020

Table 1 (continued)

				African				Eurasian						East Asian						
																	Southern Chinese			
ID of Sequence Variation	Position	Singleton	Insertion/ Deletion	Pygmy	Alur	Nanda	Hema	Utah (CEPH)	Palestinian	Czech	Druze	Ashkenazi	Indian/ Pakistani	Japanese	Korean	Mongolian	Fujian	Guandong	Northern Chinese (Heilongjian)	Fst
211	10714	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
212	10716	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	0	.020
213	10774	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
214	10835	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	.020
215	10845	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
216	10951	S		0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	.020
217	11018	S		0	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.027
218	11068	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
219	11108	S		.016	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.015
220	11136			.047	.071	.111	0	0	0	0	0	0	.033	0	0	0	0	0	0	.064
221	11145			0	.143	.056	0	0	0	0	0	0	0	0	0	0	0	0	0	.107
222	11169	S		0	0	0	0	0	0	0	0	0	.017	0	0	0	0	0	0	.016
223	11401			.094	0	.028	.029	0	0	0	0	0	0	0	0	0	0	0	0	.060
224	11534			.125	0	0	.294	.375	.417	.271	.187	.417	.233	.146	.042	.208	.208	.104	.312	.104
225	11535	S		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	.021	.020
226	11553		I/D	0	0	0	0	.021	0	.021	0	0	0	0	0	0	0	0	0	.018
227	11601			0	0	.056	0	0	0	0	0	0	0	0	0	0	0	0	0	.053
228	11607			.125	0	0	.294	.375	.417	.271	.187	.417	.233	.146	.042	.208	.208	.104	.312	.104
229	11750			0	0	0	0	0	.042	0	.062	0	.117	.125	.083	.104	.146	.187	.125	.068
230	11803	S		0	0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	.020
231	11972	S		0	0	0	0	0	0	0	0	.021	0	0	0	0	0	0	0	.020
232	12008			0	.071	0	.029	.021	0	.042	.062	0	0	0	0	0	0	0	0	.040
233	12057			.016	0	.056	.029	.042	.042	.062	.042	.062	.1	.229	.458	.396	.333	.375	.271	.178
234	12079			.047	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.035
235	12193			.016	0	.056	.029	.062	.042	.083	.042	.062	.1	.229	.458	.396	.333	.375	.271	.172
236	12349			0	0	0	0	0	0	0	0	0	0	.021	.042	0	0	.021	0	.026
237	12428			.016	0	.056	.029	.062	.042	.083	.042	.062	.1	.229	.458	.396	.333	.375	.271	.172
238	12457	S		0	0	0	0	0	.021	0	0	0	0	0	0	0	0	0	0	.020
239	12506	S		0	0	.028	0	0	0	0	0	0	0	0	0	0	0	0	0	.026
240	12651			0	0	0	0	0	0	.021	0	.021	0	0	0	0	0	0	0	.018
241	12685			.031	0	0	.029	0	0	0	0	0	0	0	0	0	0	0	0	.026
242	12686			.016	0	.083	.029	.062	.042	.062	.042	.062	.1	.229	.458	.396	.333	.375	.271	.171
243	12821			.141	0	.028	.294	.708	.542	.5	.396	.479	.283	.187	.125	.25	.208	.146	.396	.176
244	12999			.016	.071	.056	.029	0	0	0	0	0	0	0	0	0	0	0	0	.044
245	13073			.984	.929	.889	.971	.937	.958	.917	.958	.937	.9	.771	.521	.604	.729	.625	.729	.153

Statistical Analysis

The proportion of variation attributable to differences among continents was estimated using Wright's F_{ST} statistic. Haplotypes were inferred, and haplotype frequencies were estimated, using the expectation-maximization method of haplotype inference included in the Arlequin computer program (Schneider et al. 2000).

Pairwise LD was estimated as $D = x_{ij} - p_i p_j$, where x_{ij} is the frequency of haplotype *ij*, and p_i and p_j are the frequencies of alleles *i* and *j* at loci A and B, respectively. A standardized LD coefficient, *r*, is obtained by $D/(p_i p_j q_i q_j)^{1/2}$, where q_i and q_j are the frequencies of the other alleles at loci A and B, respectively. Lewontin's coefficient is given by $D' = D/D_{\max}$, where $D_{\max} = \min(p_i p_j, q_i q_j)$ when $D < 0$, or $D_{\max} = \max(q_i p_j, p_i q_j)$ when $D > 0$ (Lewontin 1964).

Nucleotide diversity (π) and Watterson's estimator (θ_w) (Watterson 1975) were calculated using DnaSP version 3.50. (Rozas and Rozas 1999). θ_w was estimated from the total number of polymorphic sites (*S*). Tajima's neutrality test (Tajima 1989) and Fu and Li's neutrality tests (Fu and Li 1993) were also performed using DnaSP version 3.50. Tests of Fay and Wu's *H* statistic (Fay and Wu 2000) were performed using Fay's computer program.

Allele ages were estimated using two methods. First, allele ages were estimated using the formula $P = (1 - r)^G = e^{-rG}$, where *P* is the fraction of alleles in the ancestral state, *r* is the mutation rate of the repeat, and *G* is the human generation length in years (Kaplan et al. 1994; Tishkoff et al. 1996; Stephens et al. 1998). This method estimates age by examining the extent to which derived alleles have diverged from the ancestral state. Second, allele ages were estimated using allele frequencies with the formula $G = -4N_e[p\ln(p)/(1 - p)]$, which estimates allele age (*G*), given its frequency (*p*) and a constant effective population size (N_e) (Kimura and Ohta 1973). This formula operates under the assumption of neutrality and constant population size.

Genetic distances between populations were calculated using Nei's *D*. $D = -\ln(I)$, where $I = (p_i q_i)/(p_i^2 q_i^2)^{1/2}$, p_i and q_i are the allele frequencies of the *i*th allele in populations *p* and *q*, and *I* is summed over all variable sites (Nei 1987). Neighbor-joining trees were inferred using the "neighbor" program of PHYLIP (Felsenstein 1989).

To test whether evidence for selective effects was localized to the A(−6)G promoter polymorphism, we compared patterns of diversity in the 20 polymorphisms nearest the A(−6)G polymorphism with patterns of diversity across the gene as a whole, using Tajima's *D* statistic. Ordinarily, Tajima's *D* is tested by simulating populations that have remained constant in size and then comparing the simulated *D* values with observed *D* values. Wooding and Rogers (2002) suggested an alternative approach, which we used here. First, we inferred the demographic parameters that best explain diversity in the gene as a whole, using the methods of Wooding and Rogers (2002). Next, we simulated the distribution of Tajima's *D* under the inferred parameters. This approach is conservative, because diversity in the test region—near the A(−6)G variant—is included in the data used to generate the null distribution of *D*.

Demographic parameters were estimated for the AGT gene under a two-epoch "sudden growth" model. Under this model, demographic histories are described by two epochs within which population sizes are constant but between which sizes may vary. Thus, the model is described by three parameters: N_0, *t*, and N_1, where N_1 is ancient effective population size, *t* is the time of a sudden, instantaneous change in effective population size, and N_0 is the subsequent effective population size. This model, though simple, provides a reasonable approximation of more-complex models of population history (Rogers 1997).

Results

A total of 246 sequence variants, including 235 nt substitutions (roughly 1 nt substitution per 61 bp), was observed in the human sample (table 1). Of the derived variants, 100 were singletons (i.e., they were observed only once in the sample). All others were observed two or more times. Among the substitutions observed, transitions were more common than transversions (120 vs. 39), and the transition:transversion ratio was 3.08. Ten insertion/deletion variants and one dinucleotide repeat were also detected.

Fifty-four nucleotide positions were variable on all three continents. Ninety-one derived variants (including 43 singletons) were observed only among the African sequences. The Eurasian and East Asian populations contained 50 and 47 unique derived variants, respectively (fig. 1).

The two estimates of nucleotide diversify, π and θ_w, varied widely among populations (table 2). The African sample had the largest observed θ_w (16.06×10^{-4}), and N_e (14,820) values. It is interesting that the East Asian sample had both the largest observed π value, 14.44×10^{-4}, and the lowest observed θ_w value, 10.02×10^{-4}. Figure 2*A* shows how patterns of diversity varied across regions, on a site-by-site basis.

Genetic Distance and Allele Frequency

In a previous analysis of genetic variation in *AGT*, we found that the A(−6) and T235 variants are fixed in several nonhuman primate species (Inoue et al. 1997). Thus, A(−6) and T235 most probably represent the ancestral state of the human alleles, whereas G(−6) and M235

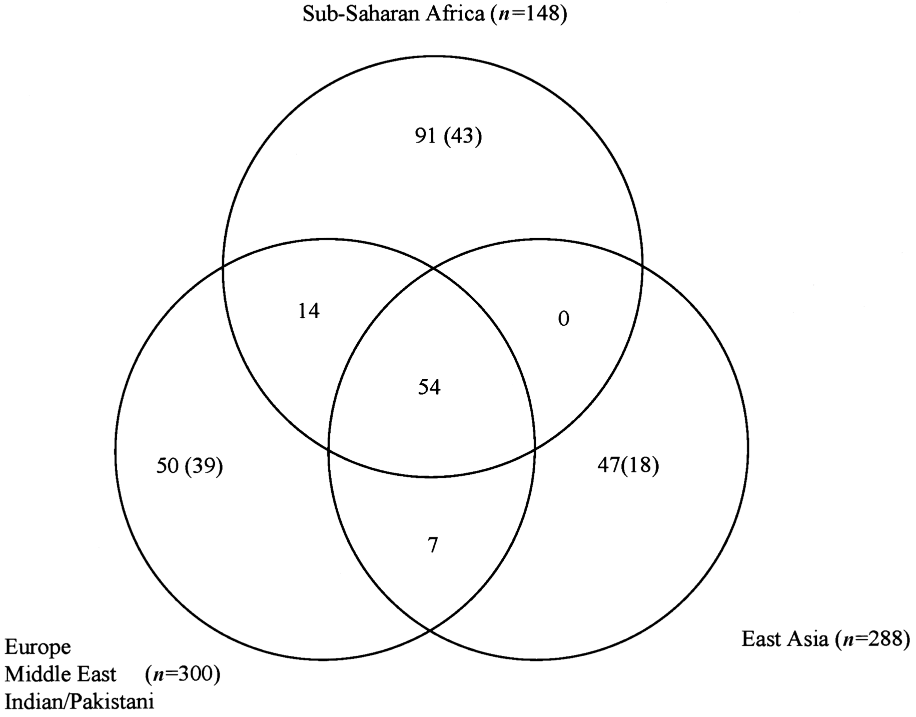

Figure 1 Distribution of sequence variations on three continents: Africa, Eurasia, and East Asia

represent the derived state. In our sample, the frequency of the derived M235 allele varied widely across populations, with higher frequencies outside of Africa. The frequency of M235 was highest in the Middle East and Europe—for instance, 0.396 in Druze, 0.458 in Ashkenazi, 0.562 in Palestinians, 0.500 in Czechs, and 0.708 in Utah—whereas its frequency in Africa was low (0.022), except in the Hema (0.294). Intermediate allele frequencies were seen in the Asian populations. The frequency of the G−6 allele showed an identical pattern because of its tight LD with M235, with low frequencies in African populations and high frequencies in non-African populations. G(−6)/M235 allele frequencies in the different populations are shown in figure 3.

LD

Twelve SNPs were in tight LD with A(−6)G ($r^2 \geqslant 0.5$). These SNPs were distributed across the entire gene, as shown in figure 2*B*. The mean pairwise r^2 value between these SNPs in the non-African sample was significantly higher than in the African sample (0.800 ± 0.177 vs. 0.444 ± 0.247; $P = .0004$). Pairwise LD values in 50 SNPs shared across the three continents showed further evidence of differences in the structure of LD between African and non-African samples. Whereas the D' statistic showed evidence of tight LD across *AGT* in both African and non-African populations, distributions of r^2 differed appreciably between these two groups, as shown in figure 4*A*. The African sample had far fewer SNP pairs with $r^2 \geqslant 0.5$ than did non-African samples. When pairwise r^2 values were plotted as a function of physical distance, the African samples were found to have lower levels of LD than any of the non-African samples (fig. 4*B*). The non-African populations had almost identical distributions of r^2, even though the neighbor-joining network of Nei's genetic distance showed that appreciable genetic distances are found between European and East Asian populations. The similarities in the distributions of r^2 value among non-Africans are consistent with the finding that they have similar haplotype structures in *AGT* but with different haplotype frequencies (Nakajima et al. 2002).

Relationship between SNP Haplotypes and CA Repeats

Correlations between common haplotypes that are based on six SNPs and CA-repeat alleles showed that the G(−6)/M235 allele was strongly associated with CA-repeat allele 197, as shown in table 3. A graphical display of nucleotide diversity across *AGT* showed that among the 30 observed G(−6)/M235 homozygotes, diversity levels were low (fig. 2*B*). The θ_w and π values of the 30 G(−6)/M235 homozygotes were also low (table 4), as was the variance of CA-repeat allele size. All three

Table 2

Nucleotide Diversities of *AGT* in 16 Populations

	African					Eurasian							East Asian						
														Southern Chinese					
Measurement	Pygmy (n = 64)	Alur (n = 14)	Nande (n = 36)	Hema (n = 34)	All African (n = 148)	Utah (CEPH) (n = 48)	Czech (n = 48)	Ashkenazi (n = 48)	Druze (n = 48)	Palestinian (n = 48)	Indian/ Pakistani (n = 60)	All Eurasian (n = 300)	Japanese (n = 48)	Guandong (n = 48)	Fujian (n = 48)	Northern Chinese (Heilongjian) (n = 48)	Mongolia (n = 48)	Korean (n = 48)	All East Asian (n = 288)
No. of sequence variations	106	58	106	88		73	70	73	73	74	78		67	63	68	71	68	66	
No. of singletons	11	3	16	13		7	6	6	7	3	10		3	1	4	5	1	4	
No. of insertions/deletions	4	2	5	4		1	1	0	1	0	0		0	0	0	0	0	0	
$\pi(\times 10^{-4})$	12.66	11.84	13.26	12.45	13.18	9.78	12.55	12.53	13.07	11.45	13.78	12.69	14.26	14.32	14.50	14.64	14.41	13.11	14.44
$\theta_W(\times 10^{-4})$	14.83	12.01	16.75	14.10	18.82	11.11	10.48	11.42	10.95	11.27	11.32	13.60	10.17	10.33	9.55	10.80	10.33	10.01	10.02
Ne	11,680	9,457	13,186	11,100	14,820	8,748	8,255	8,994	8,625	8,871	8,912	10,711	8,009	8,132	7,516	8,502	8,132	7,885	7,889
Tajima's D	−.505	−.064	−.774	−.435		−.424	.694	.343	.682	.057	.749		1.415	1.360	1.821	1.254	1.393	1.086	
Fu & Li's D*	−.845	.056	−.654	−1.120		.167	.713	.584	1.152	.810	.461		.806	1.092	1.151	.625	1.226	.923	
Fu & Li's F*	−.857	.026	−.827	−1.050		−.059	.843	.592	1.170	.637	.684		1.227	1.420	1.664[a]	1.021	1.535	1.172	

[a] Significantly different at the 5% level.

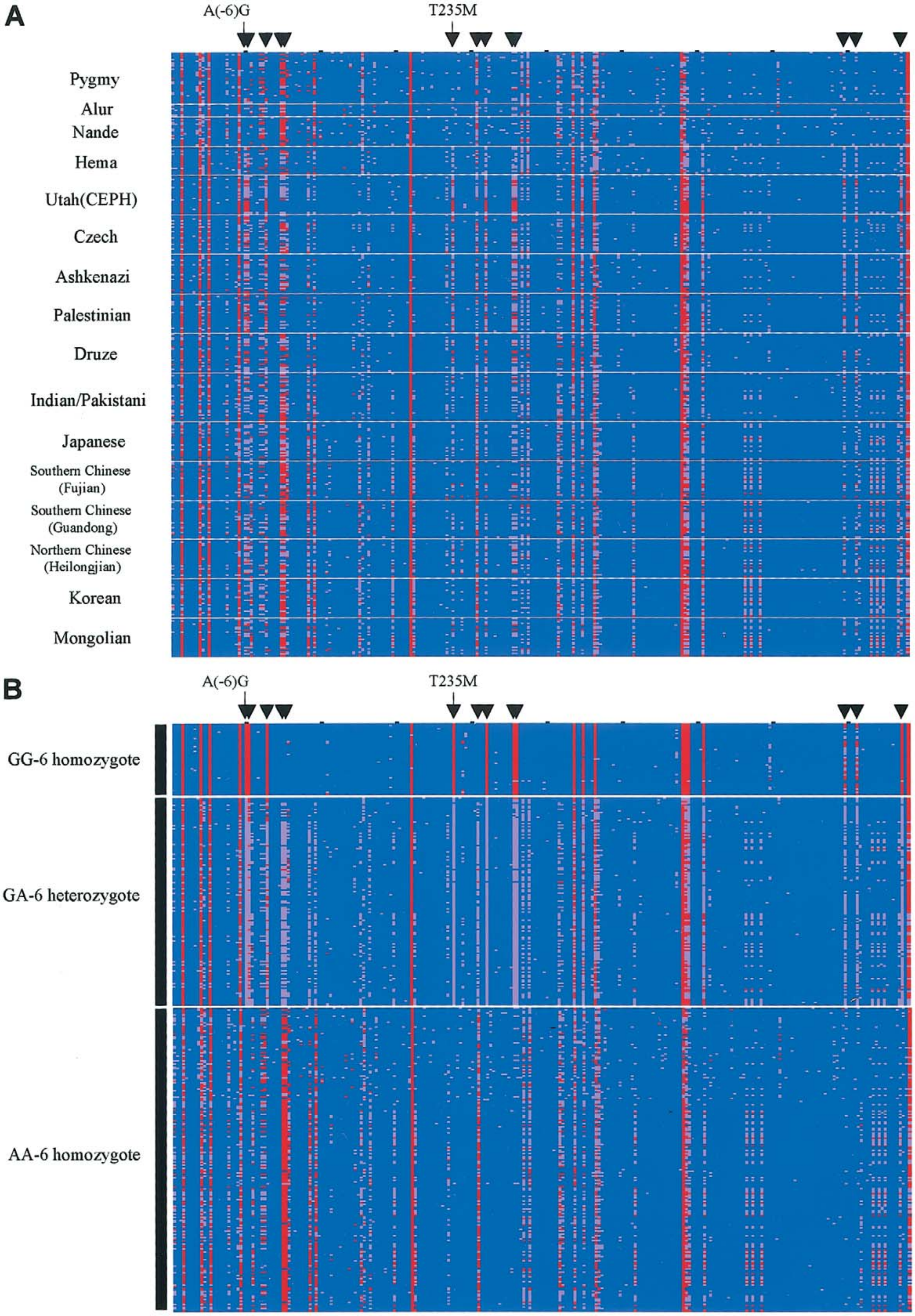

Figure 2 *A,* Summary of sequence variation in *AGT.* Variable sites are color coded. Homozygotes for the derived allele are red, heterozygotes are purple, and homozygotes for the ancestral allele are blue. The ancestral allele at each locus was estimated on the basis of primate (*Pan troglodytes verus, Gorilla gorilla,* and *Pongo pygmaeus*) sequences. *B,* Sequence variation in *AGT* genotypes homozygous for G(−6), heterozygous, and homozygous for A(−6). SNPs in tight LD with A(−6)G ($r^2 > 0.5$) are indicated by arrows and arrow heads.

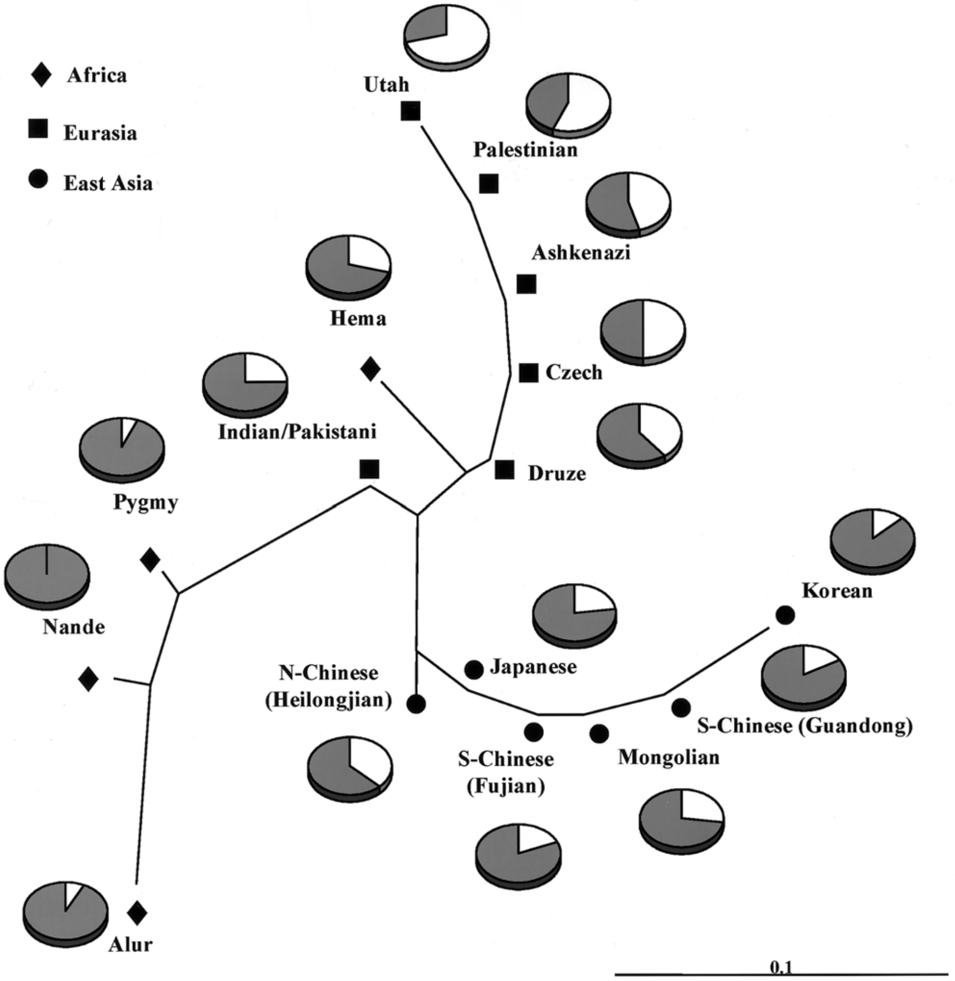

Figure 3 Neighbor-joining tree based on Nei's distance among seven populations. The shaded portion of each circle indicates the frequency of the A(−6)/T235 allele.

of these patterns are consistent with the high level of LD among haplotypes carrying the G(−6)/M235 variants.

Tests for Selective Neutrality

When complete sequences from the combined human samples were analyzed, Tajima's *D* and Fu and Li's *D** and *F** did not deviate significantly from expectations under neutrality, except in the Southern Chinese (Fujian) sample (table 2). However, prior evidence that the A(−6)G promoter polymorphism can have important phenotypic effects suggested that natural selection might have operated on that site. When restricted to include only polymorphisms near the A(−6)G polymorphism (i.e., the 10 polymorphisms on either side of it), all three of the test statistics rejected the hypothesis of neutrality in most populations, with observed values being significantly greater than expected (fig. 5).

Similarly, when whole *AGT* sequences were analyzed, Fay and Wu's *H* test failed to reject the hypothesis of evolutionary neutrality in the combined sample—or in Africans or non-Africans separately—for any assumed recombination rate (data not shown). However, this test rejected the hypothesis of neutrality in sites near the A(−6)G polymorphism in most populations, with observed values being lower than expected.

The rejection of the neutrality hypothesis in sites near the A(−6)G polymorphism, but not across the gene as a whole, suggested that patterns of variation in these sites differ from those found in other regions of the gene. A sliding-window analysis showed that the values of all three of these statistics varied widely across the sequenced regions, as shown in figure 5. When windows were adjusted to include 20 SNPs, high *D, D*,* and *F** values were observed in the vicinity of sequence variants in disequilibrium with the A(−6)G polymorphism in the Druze, Ashkenazi, and Czech samples (fig. 5). Similar

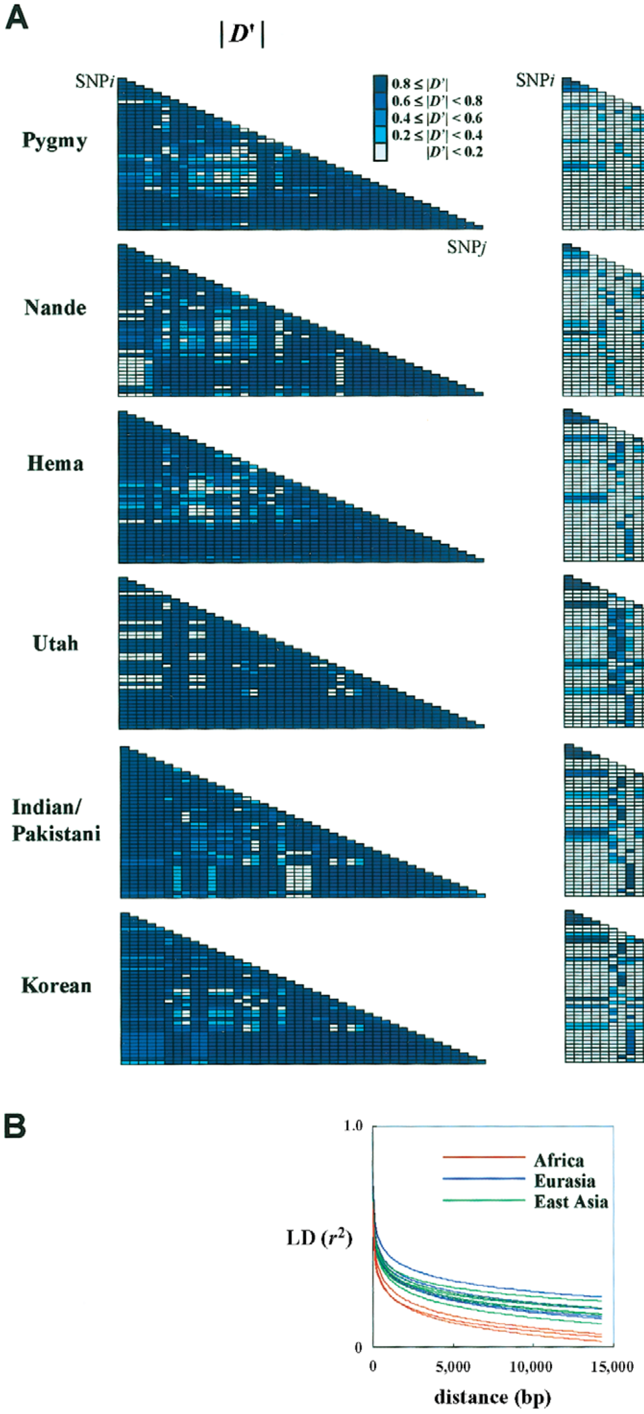

Figure 4 *A,* Structure of LD in three African populations (Pygmy, Nande, and Hema) and three non-African populations (Utah, Indian/Pakistani, and Korean). Each point represents a disequilibrium coefficient, $|D'|$ and r^2, calculated pairwise among 50 SNPs. Pairs in LD ($D' \geqslant 0.75$, $r^2 \geqslant 0.5$) are shaded. Arrowheads indicate SNPs that were not observed in each population sample. *B,* Pairwise r^2 values in 15 population samples were plotted as a function of physical distance.

Table 3

Relationships between Estimated SNP Haplotype and CA-Repeat Allele in Non-African Populations

Haplotype[a]	No. of Chromosomes	Average Size of CA Repeats	Variance (*F* Test)	Heterozygosity
A-A-G-C-A-M(G[-6]/M235)	202	196.5	7.34 (3.06×10^{-22})	.61
A-C-$\overline{\text{A}}$-C-A-T(C-20/T235)	113	193.6	4.60 (3.24×10^{-22})	.52
G-$\overline{\text{A}}$-A-C-A-T(G-1178/T235)	88	201.0	8.60 (3.94×10^{-9})	.66
$\overline{\text{A}}$-A-A-T-A-T(T172 × T235)	131	203.5	28.24 (.51)	.75
A-A-A-$\overline{\text{C}}$-G-T(G676/T235)	31	196.0	25.30 (.99)	.72
Others	23	…	…	…
All CA-repeat alleles	588	198.1	25.92	.85

[a] Haplotypes based on five sequence variants—A(−1178G)-A(−20)C-A(−6)G-C172T-A676G-T235M—were shown.

positive deviations of *D*, *D**, and *F** in the vicinity of A(−6)G were also observed in the other populations. These results suggested that diversity levels near the A(−6)G polymorphism are greater than in other regions of the gene. However, such differences might be attributable to random variation along the gene. For example, peaks in the sliding-window analysis could arise due to correlations among sites near one another, which probably have correlated gene genealogies. Further, it is possible that the departure from neutrality observed near the A(−6)G polymorphism is really due to demographic factors, which can leave signatures of genetic variation that are nearly identical to those of selection (Bamshad and Wooding 2003). To address these problems, we used Tajima's *D* to compare patterns of diversity near the A(−6)G polymorphism with patterns of diversity across the gene as a whole.

Under the sudden-growth model, diversity patterns in the gene as a whole implied a 60-fold increase in human population sizes 250,000 years ago, under the assumption that the ancient human population had an effective size of 10,000. These values are similar to those inferred for numerous other genes (Harpending and Rogers 2000), although the inferred expansion time is somewhat earlier than most estimates, possibly because of the effects of selection in limited regions of the gene. Simulations of Tajima's *D* statistic under these parameters showed that *D* values near the A(−6)G variant were significantly greater than expected ($P < .01$). Thus, not only were patterns of variation near A(−6)G greater than expected by chance under the assumption of constant population size, but they were also greater than expected given the population history implied by the gene as a whole.

Allele Ages

The analyses of allele age gave dramatically different results, depending on the method used. The analysis of STR allele sizes—with the formula $P = (1-r)^G e^{-rG}$, as described in the "Methods" section—implied a recent origin of the G(−6)/M235 allele. Under the assumption that the dinucleotide-repeat mutation rate is between 10^{-4} and 10^{-3} per generation, our data implied an origin of the G(−6)/M235 allele 22,500–44,500 years ago. This result is congruent with a previous estimate that was based on patterns of diversity in 176 European and 154 Japanese chromosomes, which implied an origin 20,500–40,000 years ago (Nakajima et al. 2002). In contrast, an analysis of SNP allele frequencies—with the formula $G = -4N_e[p \ln(p)/(1-p)]$, as described in the "Methods" section—implied a more ancient origin. Under the unrealistic assumption of selective neutrality, a nucleotide-substitution rate of 1.59×10^{-9} per site per year, and an effective population size of 10,000, our data implied an origin of the G(−6)/M235 allele 495,000 years ago. The disparity between these results indicates either that very few STR mutations took place on the G(−6)/M235 allele background while that allele frequency rose slowly to high frequency or that the allele frequency rose to high frequency quickly.

Discussion

The key role of AGT in the renin-angiotensin system suggests a number of ways in which variation in the gene could have fitness consequences. Two are most obvious. First, patterns of variation in *AGT* affecting expression levels could have significant phenotypic effects. Variations in the *AGT* promoter are known to affect *AGT*-promoter activity in vitro. Second, patterns of variation affecting the interaction of AGT with renin, the enzyme that converts AGT into A-I, could affect the rate at which A-I, an intermediate product of the RAS, is produced. Reduced availability of A-I limits the rate at which A-II is produced, thereby influencing blood pressure phenotypes, with possible consequences for fitness. The varying ways in which selection could influence patterns of variability in *AGT* suggest that signatures of natural selection might vary across the gene.

The implications of patterns of genetic variation across *AGT* as a whole were unambiguous: none of the con-

Table 4

Nucleotide Diversity and CA-Repeat Alleles in G(−6) Homozygotes

		Nucleotide Diversity		CA-Repeat Allele		
	No. of Chromosomes	π	θ_W	Average Size	Variance of Allele Size (F Test)	Heterozygosity
G(−6) Homozygotes	90	2.10×10^{-4}	5.45×10^{-4}	196.2	6.37(2.50×10^{-13})	.62
Non-African	588	14.30×10^{-4}	15.18×10^{-4}	198.1	25.92	.85
All (African and Non-African)	736	14.46×10^{-4}	22.74×10^{-4}	197.9	27.30	.87

ventional tests for evolutionary neutrality—Tajima's D, Fu and Li's D^* and F^*, or Fay and Wu's H—gave significant results. Thus, patterns of diversity across the region are consistent with the hypothesis of selective neutrality and constant human population size. A problem with these analyses, however, is that *AGT* is large, and natural selection (if present) is likely to be acting only in limited regions of the gene. Because genetic recombination can allow different regions of a gene to evolve semi-independently, it is more appropriate to examine the gene on a region-by-region basis than as a single unit. In particular, the phenotypic association of the A(−6)G promoter variant suggests that selective effects might be localized to the promoter region.

Neutrality tests restricted to the 10 sites on either side of the A(−6)G variant showed that Tajima's D and Fu and Li's D^* and F^* had values significantly greater than expected under neutrality, whereas Fay and Wu's H had values significantly less than expected under neutrality, in most populations. Significantly positive D, D^*, and F^* values are often interpreted as evidence for balancing selection, population subdivision, or decrease in population size—factors that result in a relative overabundance of derived variants with intermediate frequencies. However, all three of these tests assume a complete absence of genetic recombination. Previous analyses have found evidence of recombination in *AGT* (Nakajima et al. 2002), suggesting that tests that take recombination into account are more appropriate for the analysis of *AGT.* One such test is the H test, derived by Fay and Wu (2000), who showed that positive natural selection can result in a relative overabundance of intermediate- and high-frequency variants when recombination is present. The application of the H test to the promoter region of *AGT* yielded statistically significant results in most non-African populations but in only one African population, the Hema (table 5; fig. 6). Thus, the only test that takes recombination into account rejects the hypothesis of neutrality in non-African populations. Patterns of diversity in the *AGT* promoter are more consistent with the hypothesis of a selective sweep.

Sliding-window analyses across *AGT,* shown in figure 5, bring within-gene patterns of variation into focus—values of D, D^*, and F^* statistics fluctuate dramatically across the length of the gene. Such fluctuation is expected because of stochastic variation in population genetic processes. A comparison of figures 2*B* and 5 shows that positive values are found around sites associated with A(−6)G. The strongest deviations from neutrality are observed in populations outside Africa, but one population inside Africa, the Hema, showed patterns of variation similar to those of non-Africans (fig. 5). Further, the values of Tajima's D in the vicinity of the A(−6)G and T235M polymorphisms are greater than expected by chance, given the demographic parameters implied by diversity patterns across the gene as a whole ($P < .01$), but these deviations are somewhat smaller in Africa than elsewhere. Thus, patterns of variation within *AGT* suggest that signatures of natural selection involving the A(−6)G and T235M polymorphisms may differ between African and non-African populations.

If a selective sweep outside of Africa did preferentially favor the G(−6)/M235 allele, then two patterns of LD should be present. First, because selective sweeps cause a rapid increase in allele frequency, levels of LD around a selected variant are often high (Sabeti et al. 2002; Toomajian and Kreitman 2002). Thus, in the case of *AGT,* levels of LD should be elevated among haplotypes carrying the G(−6)/M235 variants but not among haplotypes carrying the A(−6)/T235 variants. Second, because such a selective sweep would involve primarily non-African populations, levels of LD should be relatively higher in the non-African sample than in the African sample. Figures 2*B* and 4 show that both patterns are observed here. A more powerful test of this hypothesis would be the long-range LD test devised by Sabeti et al. (2002). We predict that this test, which detects changes in LD across large genomic regions, would find appreciable decreases in LD with increasing distance from *AGT.*

An alternative explanation for African/Eurasian differences in LD is that the population histories of the two regions are different. Population bottlenecks followed by growth can result in relatively high LD. However, regional differences in population growth would be expected to affect *AGT* as a whole. Here, the differences are pronounced around the A(−6)G and T235M polymorphisms. Thus, the hypothesis that regional differences in population growth caused elevated LD fails to explain within-gene patterns of variation.

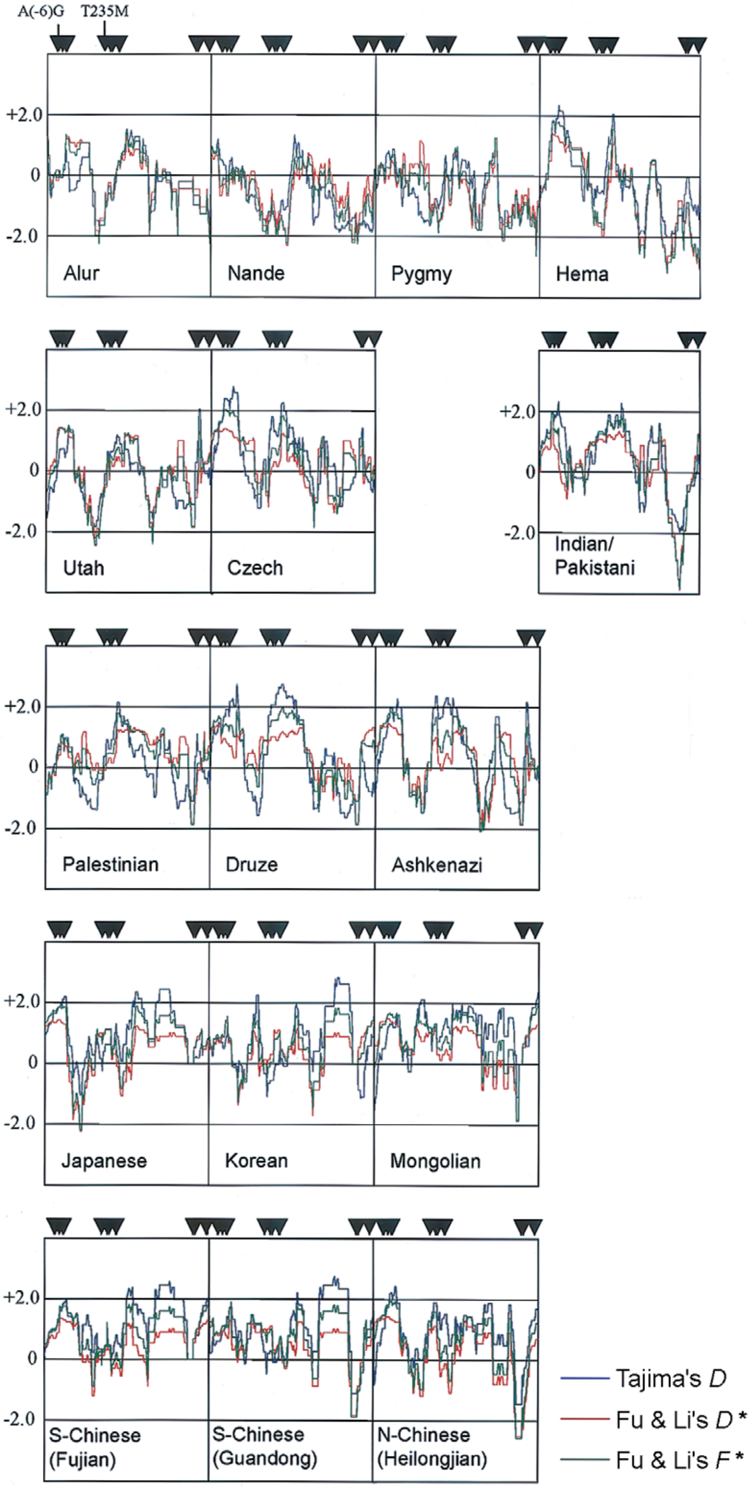

Figure 5 Sliding-window test of neutrality in *AGT*. Each window contains 20 SNPs, with a step size of 1 SNP. Arrows indicate SNPs in tight LD with T235M.

Table 5

Fay and Wu's *H* Test for the Promoter Region of AGT

Population (No. of Chromosomes)	θ_π	θ_H	*P*
Utah (48)	1.478	7.926	.008[a]
Czech (48)	2.349	6.161	.029[b]
Europe (Utah and Czech) (96)	1.942	6.921	.016
Ashkenazi (48)	2.418	6.433	.033[b]
Druze (48)	2.884	5.669	.049[b]
Palestinian (48)	2.081	7.409	.012[b]
Middle East (144)	2.468	6.371	.037[b]
Indian/Pakistani (60)	3.072	4.995	.102
Japanese (48)	2.873	5.340	.067
Southern Chinese: Fujian (48)	2.409	5.634	.033[b]
Southern Chinese: Guandong (48)	2.250	6.176	.028[b]
Northern Chinese: Heilongjian (48)	1.818	6.990	.014[b]
Korean (48)	2.450	5.550	.045
Mongolian (48)	1.373	3.350	.082
Pygmy (64)	2.679	6.464	.053
Alur (14)	2.681	5.934	.054
Nande (36)	3.087	5.998	.065
Hema (34)	2.307	7.027	.034[b]
African (148)	2.745	6.194	.066

[a] Significantly different at the 1% level.

[b] Significantly different at the 5% level.

Finally, under the hypothesis that a selective sweep recently raised the frequency of the G(−6)/M235 allele in non-African populations, the G(−6)/M235 allele should be relatively young. The variance of allele size in CA repeats associated with the G(−6)/M235 allele is much lower than that associated with other alleles, suggesting that the G(−6)/M235 allele has risen to high frequency recently. Two other common SNP haplotypes, C(−20)/T235 and G(−1178)/T235, were also relatively young. However, these haplotypes distribute equally among African and non-African populations—F_{ST} for A(−20)C and A(−1178)G are 0.025 and 0.032, respectively. Results of our analysis of patterns of CA-repeat variability suggest that the age of the G(−6)/M235 allele is 20,000–40,000 years. This age is of some interest, because it postdates most estimates of the onset of large-scale population expansion in humans (Harpending and Rogers 2000). Further, the age of the G(−6)/M235 allele, as estimated from CA-repeat variability, is substantially lower than the estimate based on the frequency of the G(−6)/M235 allele. Under the assumption that selective processes have not had significant effects on allele frequencies and that the G(−6)/M235 allele has risen to high frequency because of genetic drift alone, the age is 495,000 years. The disparity of these two estimates of allele age is consistent with the hypothesis that the high frequencies of G(−6)/M235 allele outside of Africa are not the result of genetic drift but are the result of selective forces driving a new, favored variant toward fixation.

Signatures of positive natural selection have been found in numerous human genes. To date, most evidence for selective sweeps has been found in smaller genomic regions and has thus been easier to detect using simpler statistical methods. For example, Wooding et al. (2002) found evidence for a selective sweep in a region 5′ of the *CYP1A2* gene, but the signature was not localized to a small part of the sequenced region. Studies across larger genomic regions have also found evidence for positive natural selection in genes such as *G6PD,* CD40 ligand, and *HFE* (Sabeti et al. 2002; Toomajian and Kreitman 2002). These studies have exploited long-range patterns of LD to identify very recent selective sweeps. Our results emphasize the value of examining intragenic variation in signatures of natural selection and reiterate the need to develop more sophisticated statistical tools for studying genetic variation on intermediate scales.

Patterns of variation in *AGT* illustrate a number of problems that are likely to be encountered in tests for natural selection in larger gene regions. Whereas diversity patterns in *AGT* as a whole are not consistent with a major role for natural selection, these patterns vary dramatically within the gene, and evidence for natural selection is present only in localized areas. Although earlier studies of natural selection in human genes were often able to make the assumption of complete LD across a sequence of interest, modern technologies provide us with sequences so long that recombination is a significant—and potentially problematic—factor. Here, we found evidence for natural selection that was limited to specific portions of a gene that were of interest a priori (i.e., the promoter region and the region surrounding the T235M polymorphism), and we were able to use relatively simple statistical tools to identify some ob-

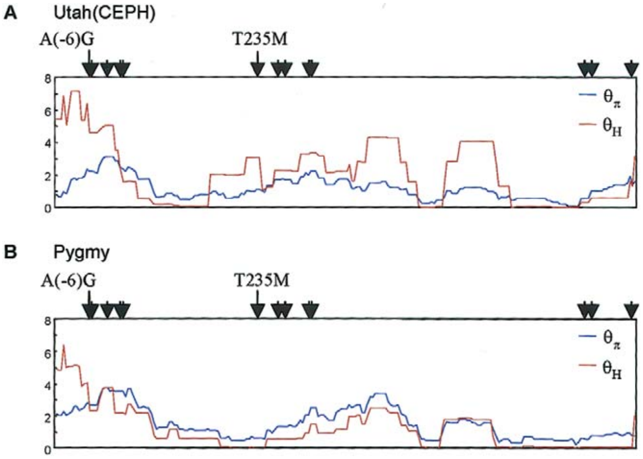

Figure 6 Sliding-window plots of θ_π and θ_H in African samples (*A*) and Utah samples (*B*). Each window contains 20 SNPs, with a step size of 1 SNP. The sites of SNPs in tight LD with A(−6)G ($r^2 > 0.5$) are indicated by arrows.

vious patterns. Future studies of natural selection that use sliding-window analysis may need to address the statistical problem of distinguishing true signals of selection from those arising spuriously by multiple comparisons, especially when the genetic signatures left by natural selection are subtle.

A key implication of our data is that, although patterns of genetic diversity in *AGT* provide evidence for a selective sweep by the G(−6)/M235 allele, this sweep did not affect all populations equally. On average, the G(−6)/M235 allele is found at higher frequencies outside Africa, and non-African samples show stronger evidence for a selective sweep. But high frequencies of this allele occur in at least one African sample (Hema), and low frequencies occur in several non-African samples (e.g., Korea). Thus, the geographic distribution of *AGT* alleles cannot be explained by a simple African/non-African distinction. The pattern observed here (fig. 3) differs from population relationships inferred from multiple neutral alleles, in which the Hema cluster with other African populations (Watkins et al. 2003). One explanation for this finding is that natural selection has acted to favor or disfavor the G(−6)/M235 allele on a finer geographic scale than has previously been recognized. Such selection might be driven by local variation in salt availability, but it could reflect other unknown factors.

The implication of natural selection as a source of geographic variation in *AGT*-allele frequencies raises basic questions about patterns of genetic diversity in populations that are, so far, poorly sampled. If natural selection has favored the A(−6)/T235 variant in sub-Saharan Africa, it might also have favored the A(−6)/T235 allele in other regions of the world with similar environments. Tropical Southeast Asia, Australasia, and Central and South America, for instance, all include large indigenous populations that have inhabited humid tropical environments for tens of thousands of years. Are allele frequencies of A(−6)/T235 higher in those populations than in others? If natural selection has affected A(−6)G/T235M-allele frequencies on smaller geographic scales than the one we examined here, phenotypes might vary similarly. Such patterns could have implications for the diagnosis and treatment of hypertension in geographically and ethnically defined populations around the world.

Acknowledgments

We are grateful for comments from Dr. Andreas Rohrwasser. This work was supported by the Future Program Grant of The Japan Society for the Promotion of Science (I.I. and M.E.); by the Ministry of Public Health and Welfare Research on Human Genome and Grant-in-Aid for Scientific Research on Medical Genome Science from the Ministry of Education, Culture, Sports, Science and Technology of Japan; and by National Institutes of Health grant HL070048.

Electronic-Database Information

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

Arlequin, http://lgb.unige.ch/arlequin/ (for software for population genetic data analysis)

GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for the sequence of *AGT* [accession number NM_000029 and X15323])

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

References

Bamshad MJ, Wooding S (2003) Signatures of natural selection in the human genome. Nat Rev Genet 4:99–111

Burt VL, Whelton P, Roccella EG, Brown C, Cutler JA, Higgins M, Horan MJ, Labarthe D (1995) Prevalence of hypertension in the US adult population: results from the Third National Health and Nutrition Examination Survey 1988–1991. Hypertension 25:305–313

Corvol P, Jeunemaitre X (1997) Molecular genetics of human hypertension: role of angiotensinogen. Endocr Rev 18:662–677

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Felsenstein J (1989) PHYLIP: Phylogeny Inference Package (version 3.2). Cladistics 5:164–166

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Gleiberman L (1973) Blood pressure and dietary salt in human populations. Ecol Food Nutr 2:143–156

Harpending H, Rogers A (2000) Genetic perspectives on human origins and differentiation. Annu Rev Genomics Hum Genet 1:361–385

Inoue I, Nakajima T, Williams CS, Quackenbush J, Puryear R, Powers M, Cheng T, Ludwig EH, Sharma AM, Hata A, Jeunemaitre X, Lalouel JM (1997) A nucleotide substitution in the promoter of human angiotensinogen is associated with essential hypertension and affects basal transcription in vitro. J Clin Invest 99:1786–1797

Jeunemaitre X, Soubrier F, Kotelevstev YV, Lifton RP, Williams CS, Charru A, Hunt SC, Hopkins PN, Williams RR, Lalouel JM, Corvol P (1992) Molecular basis of human hypertension: role of angiotensinogen. Cell 71:169–180

Kaplan NL, Lewis PO, Weir BS (1994) Age of the ΔF508 cystic fibrosis mutation. Nat Genet 8:216–217

Kimura M, Ohta T (1973) The age of a neutral mutant persisting in a finite population. Genetics 75:199–212

Lewontin RC (1964) The interaction of selection and linkage. I. General considerations, heterotic models. Genetics 49:49–67

Lifton RP (1996) Molecular genetics of human blood pressure variation. Science 272:676–680

Luft FC (2001) Molecular genetics of salt-sensitivity and hypertension. Drug Metab Dispos 29:500–504

Menard J, El Amrani A-I, Savoie F, Bouhnik J (1991) Angiotensinogen: an attractive and underrated participant in hypertension and inflammation. Hypertension 18:705–706

Nakajima T, Jorde LB, Ishigami T, Umemura S, Emi M, Lalouel J-M, Inoue I (2002) Nucleotide diversity and haplotype structure of the human angiotensinogen gene in two populations. Am J Hum Genet 70:108–123

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet 32:135–142

Rogers AR (1997) Population structure and modern human origins. In: Donnelly PJ, Tavare S (eds) Progress in population genetics and human evolution. Springer-Verlag, New York, pp 55–79

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner DF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, Ackerman HC, Campbell SJ, Altshuler D, Cooper R, Kwiatowski D, Ward R, Lander ES (2002) Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832–837

Schneider S, Kueffer J-M, Roesslie D, Excoffier L (2000) Arlequin: a software for population genetic data analysis. University of Geneva, Geneva

Sealey JE, Laragh JH (1990) The rennin-angiotensin-aldosterone system for normal regulation of blood pressure and sodium and potassium homeostasis. In: Laragh JH, Brenner BM (eds) Hypertension: pathophysiology, diagnosis and management. Raven Press, New York, pp 1287–1317

Smithies O, Kim HS (1994) Targeted gene duplication and disruption for analyzing quantitative genetic traits in mice. Proc Nat Acad Sci USA 91:3612–3615

Stephens JC, Reich DE, Goldstein DB, Shin HD, Smith MW, Carrington M, Winkler C, et al (1998) Dating the origin of the *CCR5-Δ32* AIDS-resistance allele by the coalescence of haplotypes. Am J Hum Genet 62:1507–1515

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

Takahashi N, Smithies O (1999) Gene targeting approaches to analyzing hypertension. J Am Soc Nephrol 10:1598–1605

Tishkoff SA, Dietzsch E, Speed W, Pakstis AJ, Kidd JR, Cheung K, Bonne-Tamir B, Santachiara-Benerecetti AS, Moral P, Krings M (1996) Global patterns of linkage disequilibrium at the CD4 locus and modern human origins. Science 271: 1380–1387

Tishkoff SA, Verrelli BC (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340

Toomajian C, Kreitman M (2002) Sequence variation and haplotype structure at the human HFE locus. Genetics 161:1609–1623

Ward K, Hata A, Jeunemaitre X, Helin C, Nelson L, Namikawa C, Farrington PF, Ogasawara M, Suzumori K, Tomoda S, Berrebi S, Sasaki M, Corvol P, Lifton, RP, Lalouel JM (1993) A molecular variant of angiotensinogen associated with preeclampsia. Nat Genet 4:59–61

Watkins WS, Rogers AR, Ostler CT, Wooding S, Bamshad MJ, Brassington AM, Carroll ML, Nguyen SV, Walker JA, Prasad BVR, Reddy PG, Das PK, Batzer MA, Jorde LB (2003) Genetic variation among world populations: inferences from 100 *Alu* insertion polymorphisms. Genome Res 17:1607–1618

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7: 256–276

Wooding S, Rogers AR (2002) The matrix coalescent and an application to single-nucleotide polymorphisms. Genetics 161:1641–1650

Wooding SP, Watkins WS, Bamshad MJ, Dunn DM, Weiss RB, Jorde LB (2002) DNA sequence variation in a 3.7-kb noncoding sequence 5′ of the *CYP1A2* gene: implications for human population history and natural selection. Am J Hum Genet 71:528–542