Molecular population genetics of human *CYP3A* locus: signatures of positive selection and implications for evolutionary environmental medicine

Xiaoping Chen[1, 2, *], Haijian Wang[1, 3, 4, *], Gangqiao Zhou[1], Xiumei Zhang[1], Xiaojia Dong[1], Lianteng Zhi[1], Li Jin[3], and Fuchu He[1, 3]

[1]State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of radiation Medicine, Chinese National Human Genome Center at Beijing, Beijing 102206, China.

[2]Department of Pharmacology, School of Pharmaceutical Science, Central South University, Changsha, Hunan, China.

[3]Laboratory of Systems Biology, State Key Laboratory of Genetic Engineering and MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences and Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China.

[4]The Simons Center for Systems Biology, School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA.

[*]Both these authors contributed equally to this work.

Correspondence: Prof. Fuchu He, State Key Laboratory of Proteomics, Beijing Proteome Research Center, Beijing Institute of radiation Medicine, 27 Taiping Rd., Beijing 100850, P. R. China, Tel & fax: 86-10-80705155; E-mail: hefc@nic.bmi.ac.cn.

Supplemental Material, Table 1;

Supplemental Material, Figure 1.

Supplemental Material Table 1. Frequency distribution and derived status of polymorphisms identified at *CYP3A* locus.

Variation No.	Position[a]	Gene locus	Ancestral allele	Derived allele	Derived allele frequency in			Amino acid Substitution	Allele nomenclature[b]	F_{ST}[c]		
					Chinese	Caucasians	Africans			1	2	3
1	66596	*3A43*-3'UTR	T	C	0	0	0.02					
2	66884	*3A43*-3'UTR	G	A	0	0	0.15			0.13	0.13	
3	67140	*3A43*-3'UTR	A	C	0	0	0.06					
4	67265	*3A43*-Intron 12	A	G	0	0	0.02					
5	67347	*3A43*-Intron 12	A	G	0	0.04	0					
6	71432	*3A43*-Exon 11	G	A	0	0	0.02	L350L				
7	71433	*3A43*-Exon 11	G	A	0	0.10	0	A349A	*CYP3A43*1B*		0.09	0.09
8	71545	*3A43*-Intron 10	G	A	0.55	0.42	0.25			0.16		
9	73018	*3A43*-Intron 10	A	G	0	0	0.08			0.06	0.06	
10	73084	*3A43*-Exon 10	G	C	0.02	0.10	0.31	P340A	*CYP3A43*3*	0.26	0.10	
11	73172	*3A43*-Exon 10	A	G	0	0.04	0	T310 T				
12	73216	*3A43*-Exon 10	C	G	0	0.02	0	V296L				
13	73333	*3A43*-Intron 9	G	A	0	0	0.02					
14	73346	*3A43*-Intron 9	A	G	0	0	0.06					
15	73368	*3A43*-Intron 9	G	C or T	0	0	0.06					
16	83269	*3A43*-Intron 7	G	C	0.02	0.10	0.46			0.41	0.25	
17	83448	*3A43*-Exon 7	G	A	0.98	0.90	0.54	N183N		0.41	0.25	
18	83566	*3A43*-Intron 6	C	T	0.02	0.10	0.46			0.41	0.25	
19	83643	*3A43*-Intron 6	C	T	0.98	0.90	0.54			0.41	0.25	
20	88726	*3A43*-Intron 4	T	C	0.02	0.10	0.46			0.41	0.25	
21	93997	*3A43*-Intron 2	C	T	0	0.02	0					
22	96611	*3A43*-Exon 2	T	Del	0.02	0.08	0.15	Frame shift	*CYP3A43*2A*	0.09		
23	96643	*3A43*-Intron 1	A	G	0	0	0.06					
24	104903	*3A43*-Exon 1	A	G	0	0	0.04	A16A				
25	105031	*3A43*-5'UTR	G	A	0	0	0.13			0.11	0.11	
26	105074	*3A43*-5'Flank	C	T	0.03	0	0					

Supplemental Material Table 1 (continued)

Variation No.	Position[a]	Gene locus	Ancestral allele	Derived allele	Derived allele frequency in Chinese	Caucasians	Africans	Amino acid Substitution	Allele nomenclature[b]	F_{ST}[c] 1	2	3
27	105168	*3A43*-5'Flank	C	A	0	0	0.02					
28	105191	*3A43*-5'Flank	A	C	0	0	0.02					
29	105490	*3A43*-5'Flank	T	C	0.02	0.02	0					
30	105576	*3A43*-5'Flank	G	T	0	0	0.13			0.06		
31	105648	*3A43*-5'Flank	G	T	0	0	0.02					
32	105674	*3A43*-5'Flank	T	C	0	0	0.02					
33	105924	*3A43*-5'Flank	T	C	0.02	0	0					
34	105938	*3A43*-5'Flank	C	T	0	0.02	0.08			0.06		
35	106358	*3A43*-5'Flank	T	C	0.05	0	0					
36	106490	*3A43*-5'Flank	G	T	0	0	0.04					
37	106759	*3A43*-5'Flank	T	C	0	0	0.04					
38	106857	*3A43*-5'Flank	G	C	0	0.02	0					
39	107057	*3A43*-5'Flank	T	A	0.02	0.02	0.04					
40	116513	*3A43-3A4*-Interval	T	G	0.02	0.06	0.02					
41	116846	*3A43-3A4*-Interval	A	C	0	0.02	0					
42	116889	*3A43-3A4*-Interval	T	G	0.03	0	0					
43	122872	*3A43-3A4*-Interval	C	T	0.02	0	0					
44	123007	*3A43-3A4*-Interval	A	G	0	0.02	0.10			0.09		
45	123019	*3A43-3A4*-Interval	G	A	0	0.02	0					
46	123038	*3A43-3A4*-Interval	T	G	1.00	1.00	0.96					
47	123137	*3A43-3A4*-Interval	C	T	0.02	0	0					
48	123155	*3A43-3A4*-Interval	G	A	0	0.02	0					
49	123196	*3A43-3A4*-Interval	G	A	0.05	0	0					
50	123214	*3A43-3A4*-Interval	G	A	0	0	0.02					
51	123721	*3A43-3A4*-Interval	C	T	0.02	0	0					
52	123940	*3A43-3A4*-Interval	G	A	0	0.02	0					
53	129571	*3A43-3A4*-Interval	T	C	1.00	1.00	0.96					
54	130056	*3A43-3A4*-Interval	C	T	0.02	0	0					

Supplemental Material Table 1 (continued)

Variation No.	Position[a]	Gene locus	Ancestral allele	Derived allele	Derived allele frequency in Chinese	Caucasians	Africans	Amino acid Substitution	Allele nomenclature[b]	F_{ST}[c] 1	2	3
55	135960	*3A43-3A4*-Interval	A	G	0	0.02	0					
56	146945	*3A4*-5'Flank	C	T	0.03	0	0					
57	147080	*3A4*-5'Flank	C	T	0	0	0.04					
58	147108	*3A4*-5'Flank	C	T	0.03	0	0					
59	147154	*3A4*-5'Flank	A	T	0	0.04	0					
60	147742	*3A4*-5'Flank	C	T	0	0	0.15			0.13	0.13	
61	147767	*3A4*-5'Flank	T	C	1.00	0.96	0.44			0.55	0.48	
62	148155	*3A4*-5'Flank	Ins[d]	Del	1.00	1.00	0.83		*CYP3A4*15B*	0.15	0.15	
63	148173	*3A4*-5'Flank	G	A	0.02	0	0					
64	148323	*3A4*-5'Flank	G	C	0.02	0	0					
65	148607	*3A4*-5'Flank	G	A	1.00	0.96	0.58		*CYP3A4*1B*	0.40	0.32	
66	154914	*3A4*-Intron 2	T	C	0	0	0.13			0.11	0.11	
67	162638	*3A4*-Intron 4	G	A	0	0	0.02					
68	162802	*3A4*-Intron 4	A	T	0	0	0.02					
69	162869	*3A4*-Exon 5	A	G	0.02	0	0	I118V	*CYP3A4*4*			
70	163267	*3A4*-Exon 6	G	A	0	0	0.02	R162Q	*CYP3A4*15A*			
71	163355	*3A4*-Intron 6	T	G	0	0.04	0					
72	164626	*3A4*-Exon 7	C	T	0	0	0.04	I193I				
73	164751	*3A4*-Intron 7	G	T	1.00	0.96	0.52			0.47	0.39	
74	164781	*3A4*-Intron 7	T	C	0	0	0.10			0.09	0.09	
75	164835	*3A4*-Intron 7	T	A	0	0	0.04					
76	164975	*3A4*-Intron 7	C	T	0.67	0.73	0.54					
77	165611	*3A4*-Intron 7	T	C	0.88	0.90	0.31			0.49	0.51	
78	168996	*3A4*-Intron 9	C	A	0	0	0.02					
79	169081	*3A4*-Exon 10	C	T	0	0	0.02	A297A				
80	169228	*3A4*-Intron 10	A	G	0.81	0.92	0.31		*CYP3A4*1G*	0.39	0.55	
81	169263	*3A4*-Intron 10	C	T	0	0	0.08			0.06	0.06	
82	169307	*3A4*-Intron 10	G	C	0	0	0.10			0.09	0.09	

Supplemental Material Table 1 (continued)

Variation No.	Position[a]	Gene locus	Ancestral allele	Derived allele	Derived allele frequency in Chinese	Caucasians	Africans	Amino acid Substitution	Allele nomenclature[b]	F_{ST}[c] 1	2	3
83	171039	*3A4*-Intron 11	C	T	0	0.02	0					
84	172079	*3A4*-Intron 11	T	C	1.00	0.98	0.85			0.13	0.18	
85	174956	*3A4*-3' UTR	T	G	0	0.02	0					
86	175065	*3A4*-3' UTR	A	T	0	0	0.04					
87	175102	*3A4*-3' UTR	T	G	0.02	0	0					
88	179919	*3AP2*-5'Flank	C	T	0	0	0.04					
89	180278	*3AP2*-5'Flank	A	G	0	0	0.04					
90	180392	*3AP2*-5'Flank	A	G	0.80	1.00	0.92				0.06	0.18
91	180823	*3AP2*-Intron 1	T	G	0.26	0.06	0.40				0.26	0.11
92	180867	*3AP2*-Intron 1	C	T	0	0.02	0.15			0.13	0.08	
93	181112	*3AP2*-Intron 1	G	T	0	0	0.04					
94	181116	*3AP2*-Intron 1	C	T	0	0	0.04					
95	181172	*3AP2*-Intron 1	T	A	0	0	0.02					
96	185418	*3AP2*-Intron 2	C	T	0	0	0.02					
97	186360	*3AP2*-Intron 2	A	G	0.05	0.04	0.04					
98	186372	*3AP2*-Intron 2	C	T	0	0	0.02					
99	186444	*3AP2*-Intron 2	C	T	0	0	0.02					
100	186731	*3AP2*-Intron 2	G	A	0	0	0.02					
101	186823	*3AP2*-Intron 2	T	C	0	0	0.04					
102	190645	*3AP2*-3A7 Interval	C	T	0.02	0.06	0.10					
103	190714	*3AP2*-3A7 Interval	T	A	0.02	0	0					
104	190918	*3AP2*-3A7 Interval	T	A	0.22	0.04	0.35				0.25	0.12
105	190944	*3AP2*-3A7 Interval	C	T	0	0	0.02					
106	191109	*3AP2*-3A7 Interval	G	C	0.02	0.02	0					
107	191370	*3AP2*-3A7 Interval	C	A	1.00	1.00	0.98					
108	191430	*3AP2*-3A7 Interval	A	G	0	0	0.02					
109	191479	*3AP2*-3A7 Interval	G	A	0	0	0.06					
110	191495	*3AP2*-3A7 Interval	C	T	1.00	1.00	0.98					

Supplemental Material Table 1 (continued)

Variation No.	Position[a]	Gene locus	Ancestral allele	Derived allele	Derived allele frequency in Chinese	Caucasians	Africans	Amino acid Substitution	Allele nomenclature[b]	F_{ST}[c] 1	2	3
111	191713	*3AP2*-3A7 Interval	G	A	0	0	0.02					
112	191735	*3AP2*-3A7 Interval	A	T	0	0	0.04					
113	196202	*3A7*-5'Flank	C	T	0	0	0.02					
114	196634	*3A7*-5'Flank	C	G	0	0	0.02					
115	196780	*3A7*-5'Flank	A	G	0	0	0.02					
116	202141	*3A7*-Intron 2	G	T	0	0	0.06					
117	202168	*3A7*-Intron 2	C	G	0.02	0	0					
118	210763	*3A7*-Exon 3	T	C	0	0.02	0	V71A				
119	210896	*3A7*-Intron 3	G	T	0	0	0.02					
120	211137	*3A7*-Intron 3	A	T	0.02	0	0					
121	211193	*3A7*-Intron 3	A	G	0.22	0.04	0.14					0.12
122	215147	*3A7*-Intron 4	A	T or C	0.02	0	0					
123	215240	*3A7*-Intron 4	C	T	0	0	0.02					
124	215375	*3A7*-Intron 4	T	A	0	0	0.02					
125	215413	*3A7*-Intron 4	G	A	0.03	0	0					
126	215526	*3A7*-Exon 5	G	A	0.02	0	0	G140A				
127	215703	*3A7*-Intron 5	T	C	0.79	1.00	0.96			0.10		0.19
128	215962	*3A7*-Intron 6	T	C	0.17	0	0			0.16	0.16	
129	217213	*3A7*-Exon 7	A	G	1.00	1.00	0.96	N192S				
130	217283	*3A7*-Exon 7	A	C	0	0	0.02	P215P				
131	217436	*3A7*-Intron 7	T	C	0	0.02	0					
132	218580	*3A7*-Intron 8	G	A	0	0	0.02					
133	219626	*3A7*-Intron 9	T	C	0.02	0	0					
134	219724	*3A7*-Intron 9	G	A	0.03	0.04	0.04					
135	221937	*3A7*-Intron 9	G	A	0	0	0.02					
136	233094	*3AP1*-Exon 1	G	T	0	0	0.10			0.09	0.09	
137	233141	*3AP1*-Intron 1	A	G	0.07	0.13	0.15					
138	251565	*3A5*-5'Flank	C	T	0.81	1.00	0.58			0.10	0.40	0.18

Supplemental Material Table 1 (continued)

Variation No.	Position[a]	Gene locus	Ancestral allele	Derived allele	Derived allele frequency in			Amino acid Substitution	Allele nomenclature[b]	F_{ST}[c]		
					Chinese	Caucasians	Africans			1	2	3
139	251825	*3A5*-5'Flank	C	T	0	0	0.06					
140	251839	*3A5*-5'Flank	T	A	0	0	0.04					
141	251874	*3A5*-5'Flank	C	T	0	0	0.04					
142	252179	*3A5*-5'Flank	A	C	0	0.02	0.06					
143	252387	*3A5*-5'Flank	T	A	0.21	0	0.08			0.07	0.06	0.22
144	252631	*3A5*-5'Flank	C	A	0	0.02	0.06					
145	253096	*3A5*-5'UTR	G	A	0	0.02	0		*CYP3A5**1B*			
146	253108	*3A5*-5'UTR	C	T	0.02	0.06	0.02		*CYP3A5**1C*			
147	256811	*3A5*-Intron 1	C	T	0.02	0	0					
148	256880	*3A5*-Exon 2	C	T	0	0	0.02	R28C	*CYP3A5**8*			
149	258590	*3A5*-Intron 3	T	C	0	0	0.02					
150	260167	*3A5*-Intron 3	A	G	0.78	0.96	0.29	Splicing defect	*CYP3A5**3*	0.37	0.64	0.12
151	261081	*3A5*-Intron 4	T	A	0	0.06	0					
152	261215	*3A5*-Intron 3	C	T	0	0	0.02					
153	261309	*3A5*-Intron 4	A	G	0.03	0.02	0.04					
154	266133	*3A5*-Intron 5	T	C	0.02	0	0	Splicing defect	*CYP3A5**5*			
155	266749	*3A5*-Intron 6	T	C	0	0.02	0					
156	269060	*3A5*-Exon 8	T	G	0	0	0.02	F248C				
157	270299	*3A5*-Intron 9	C	T	0	0.02	0					
158	270300	*3A5*-Intron 9	G	A	0	0	0.02					
159	270344	*3A5*-Intron 9	G	T	0.21	0	0.08		*CYP3A5**1E*		0.06	0.19
160	270403	*3A5*-Intron 9	T	C	0	0	0.02					
161	272346	*3A5*-Intron 9	T	A	0	0.02	0					
162	280231	*3A5*-Intron 10	A	G	0.02	0.13	0				0.11	0.07
163	280388	*3A5*-Exon 11	A	G	0	0.02	0	I371V				
164	283332	*3A5*-Intron 12	G	T	0	0	0.02					

Supplemental Material Table 1 (continued)

Variation No.	Position[a]	Gene locus	Ancestral allele	Derived allele	Derived allele frequency in			Amino acid Substitution	Allele nomenclature[b]	F_{ST}[c]		
					Chinese	Caucasians	Africans			1	2	3
165	283577	*3A5*-Intron 12	G	A	0	0.02	0					
166	284732	*3A5*-Exon 13	T	C	0.02	0	0	I488T	*CYP3A5*3F*			
167	284792	*3A5*-3'UTR	C	T	0.71	0.98	0.54		*CYP3A5*1D*		0.40	0.23

[a] Positions of the polymorphic site relative to the reference sequence GeneBank accession number NG_000004.2;
[b] CYP450 allele nomenclature see website http://www.imm.ki.se/CYPalleles/;
[c] Only significant F_{ST} values were listed, 1 = Chinese *vs* Africans, 2 = Caucasians *vs* Africans, 3 = Chinese *vs* Caucasians;
[d] Insertion of nine nucleotide (ATGGAGTGA) between 148155-148156 in the reference sequence.

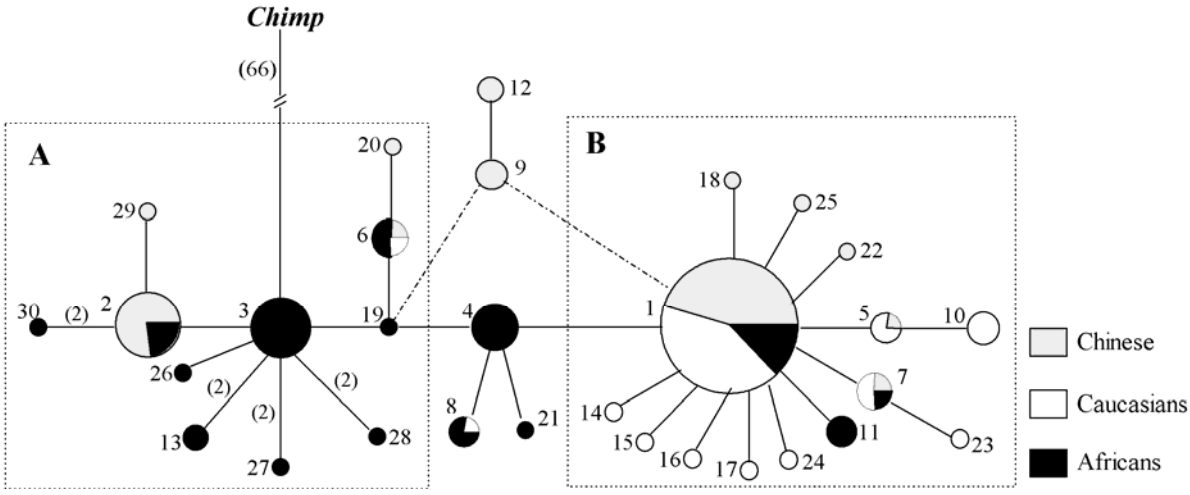
Supplemental material, Figure 1

Haplotype in *CYP3A5* and minimum-spanning tree (MST) of the haplotypes. (a) Haplotypes and their estimated frequencies in each population, the ancestral haplotype in Chimpanzee was also indicated. (b) MST of the haplotypes in *CYP3A5*. The size of each node is proportional to haplotype frequency in the overall population, and the frequency of each haplotype within each subpopulation is indicated by various shades within each node. Branches represent one nucleotide substitution, excepted where noted in the parentheses. Dashed branches indicate alternative topologies of equal length. Haplotype clusters are shown in the left and right dashed boxes, respectively

a

Haplotype	25156	25185	25189	25187	25224	25227	25236	25239	25301	25318	25366	25385	25400	26107	26115	26119	26163	26193	26204	26209	27203	27204	27204	27204	28023	28038	28038	28037	28037	28432	28479	28492	Chinese	Caucasians	Africans
Hap 1	A	C	T	C	A	T	C	G	C	C	C	T	G	T	C	A	T	T	T	C	G	G	T	G	A	A	G	G	T	T			0.603	0.667	0.208
Hap 2	G	.	.	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	C			0.172		0.063
Hap 3	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C					0.229
Hap 4	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.					0.146
Hap 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.			0.017	0.063	
Hap 6	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	C			0.017	0.021	0.042
Hap 7	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.			0.017	0.042	0.021
Hap 8	.	.	.	.	C	.	A	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.				0.021	0.063
Hap 9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C			0.052		
Hap10	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.				0.063	
Hap11	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.					0.063
Hap12	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C			0.034		
Hap13	G	.	A	T	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C					0.042
Hap14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.				0.021	
Hap15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.				0.021	
Hap16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.				0.021	
Hap17	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.				0.021	
Hap18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.			0.017		
Hap19	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C					0.021
Hap20	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	C			0.017		
Hap21	.	.	.	.	.	.	.	.	.	.	.	.	A	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.					0.021
Hap22	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.			0.017		
Hap23	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.				0.021	
Hap24	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.				0.021	
Hap25	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.			0.017		
Hap26	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	C					0.021
Hap27	G	.	.	.	.	.	.	.	.	.	C	.	A	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	C					0.021
Hap28	G	.	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.					0.021
Hap29	G	.	.	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.			0.017		
Hap30	G	.	.	.	.	A	.	.	.	.	.	.	A	.	.	.	.	.	.	G	.	.	T	C	.	.	.	.	.	C					0.021
Chimp	G	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C					

b



Supplemental material, Figure 1