Supplemental Tables

Continent	Country	Ethnic Group	Language Family	Subsistence Pattern	Number of Individuals	Source
Africa	Cameroon	Arabi Shuwa	Afroasiatic	Pastoralist	18	SAT
Africa	Cameroon	Banen	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	Bulu	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	Bangangte	Niger-Kordofanian	Agriculturalist	3	SAT
Africa	Cameroon	Batanga	Niger-Kordofanian	Agriculturalist	21	SAT
Africa	Cameroon	Bayangam	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	Djem	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	Duala	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	BgonkenxAbo	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	Evuzok	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	Ewondo	Niger-Kordofanian	Agriculturalist	2	SAT
Africa	Cameroon	Fang	Niger-Kordofanian	Agriculturalist	14	SAT
Africa	Cameroon	Fulani	Niger-Kordofanian	Pastoralist	48	SAT
Africa	Cameroon	Giziga	Afroasiatic	Agriculturalist	6	SAT
Africa	Cameroon	Kanuri	Nilo-Saharan	Agriculturalist	26	SAT
Africa	Cameroon	Kotoko	Afroasiatic	Agriculturalist	4	SAT
Africa	Cameroon	Mabea	Niger-Kordofanian	Agriculturalist (with fishing)	8	SAT
Africa	Cameroon	Mada	Afroasiatic	Agriculturalist	23	SAT
Africa	Cameroon	Mandara	Afroasiatic	Agriculturalist	14	SAT
Africa	Cameroon	Molkwo	Afroasiatic	Agriculturalist	1	SAT
Africa	Cameroon	Masa	Afroasiatic	Mixed Agriculturalist	4	SAT
Africa	Cameroon	Ngumba	Niger-Kordofanian	Agriculturalist	3	SAT
Africa	Cameroon	Nkunabembe	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	Baka Pygmy	Niger-Kordofanian	Hunter-gatherer	32	SAT

Continent	**Country**	**Ethnic Group**	**Language Family**	**Subsistence Pattern**	**Number of Individuals**	**Source**
Africa	Cameroon	Podokwo	Afroasiatic	Agriculturalist	5	SAT
Africa	Cameroon	Bakola Pygmy	Niger-Kordofanian	Hunter-gatherer	30	SAT
Africa	Cameroon	Rimaibe	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Cameroon	Tupuri	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Kenya	Burji	Afroasiatic	Mixed Agriculturalist	17	SAT
Africa	Kenya	Boni	Afroasiatic	Pastoralist	12	SAT
Africa	Kenya	Borana	Afroasiatic	Pastoralist	30	SAT
Africa	Kenya	Dikri Masaai	Nilo-Saharan	Pastoralist	2	SAT
Africa	Kenya	El Molo	Afroasiatic	Hunter-gatherer	12	SAT
Africa	Kenya	Gabra	Afroasiatic	Pastoralist	9	SAT
Africa	Kenya	Gabramigo	Afroasiatic	Pastoralist	2	SAT
Africa	Kenya	Gikuyu	Niger-Kordofanian	Agriculturalist	5	SAT
Africa	Kenya	Gurreh	Afroasiatic	Pastoralist	3	SAT
Africa	Kenya	Konso	Afroasiatic	Agriculturalist	12	SAT
Africa	Kenya	Ilchamus	Nilo-Saharan	Pastoralist	15	SAT
Africa	Kenya	Ilgwesi	Nilo-Saharan	Pastoralist	14	SAT
Africa	Kenya	Luo	Nilo-Saharan	Pastoralist	21	SAT
Africa	Kenya	Luhya	Niger-Kordofanian	Agriculturalist	1	SAT
Africa	Kenya	Maasai	Nilo-Saharan	Pastoralist	1	SAT
Africa	Kenya	Mumonyot	Nilo-Saharan	Pastoralist	10	SAT
Africa	Kenya	Ogiek	Nilo-Saharan	Hunter-gatherer	8	SAT
Africa	Kenya	Orma	Afroasiatic	Pastoralist	3	SAT
Africa	Kenya	Pokot	Nilo-Saharan	Pastoralist	10	SAT
Africa	Kenya	Pare	Niger-Kordofanian	Agriculturalist	20	SAT
Africa	Kenya	Samburu	Nilo-Saharan	Pastoralist	17	SAT
Africa	Kenya	Sengwer	Nilo-Saharan	Hunter-gatherer	8	SAT
Africa	Kenya	Sanye	Afroasiatic	Agriculturalist	13	SAT
Africa	Kenya	Taita	Niger-Kordofanian	Agriculturalist	3	SAT
Africa	Kenya	Tugen	Nilo-Saharan	Pastoralist	19	SAT

Continent	Country	Ethnic Group	Language Family	Subsistence Pattern	Number of Individuals	Source
Africa	Kenya	Turkana	Nilo-Saharan	Pastoralist	26	SAT
Africa	Kenya	Taveta	Niger-Kordofanian	Agriculturalist	5	SAT
Africa	Kenya	Wata	Afroasiatic	Hunter-gatherer	2	SAT
Africa	Kenya	Yaaku	Afroasiatic	Hunter-gatherer	17	SAT
Africa	Tanzania	Hadza	Khoesan	Hunter-gatherer	12	SAT
Africa	Tanzania	Sandawe	Khoesan	Hunter-gatherer	11	SAT
Middle East	Israel	Palestinian			20	CEPH
South Asia	Pakistan	Brahui			8	CEPH
Europe	Italy	Sardinian			4	CEPH
Europe	Spain	Spanish			16	AC
East Asia	China	Chinese			62	CEPH
East Asia	Japan	Japanese			9	CEPH
Central America	Mexico	Nahua			13	MS

Table S1. List of samples used in the present study. The number of individuals for each population included for sequencing is shown. For African samples, linguistic family and subsistence patterns are also noted for each population. Sources of samples are coded as follows: SAT (Sarah A. Tishkoff), CEPH-HGDP (Human Genome Diversity Panel), AC (Ángel Carracedo) and MS (Mark Shriver). African populations with different patterns of subsistence have very divergent diets. For example, agriculturalist populations in Africa cultivate a range of crops, such as millet, rice, cassava and bananas, which are the main staples of their diet (http://www.ethnologue.com/ ; Fedders and Salvadori 1998). Pastoralist populations, which practice animal husbandry, subsist largely on milk and blood as staple foods from cattle, sheep and/or camels with occasional meat to supplement their diet (http://www.ethnologue.com/ ; Fedders and Salvadori 1998). Hunter-gatherer diets consist primarily of meat with some supplementation with tubers in a few hunter-gatherer groups (http://www.ethnologue.com/ ; Fedders and Salvadori 1998).

Nucleotide Position	**Nucleotide Variation**	**Minor Allele Frequency**	**rs numbers**
8	A/G	0.004	
74	A/C	0.004	
76	G/C	0.001	
102	C/T	0.017	
147	C/T	0.002	
152	C/T	0.001	
173	A/G	0.001	
183	C/G	0.001	
308	A/G	0.001	
343	A/T	0.004	
385	G/A	0.001	
392	T/A	0.002	
393	A/T	0.038	rs75675140
422	T/G	0.078	rs61464348
428	A/G	0.021	
448	G/A	0.008	
524	A/G	0.001	
586	G/A	0.004	
656	C/T	0.005	
681	G/C	0.038	
767	G/T	0.004	
768	C/T	0.001	
775	G/A	0.001	
820	C/T	0.001	
839	C/T	0.001	
843	C/T	0.015	rs80134535

Nucleotide Position	**Nucleotide Variation**	**Minor Allele Frequency**	**rs numbers**
855	G/C	0.017	
907	G/A	0.008	
941	G/A	0.001	
1016	G/A	0.003	
1092	G/A	0.002	
1105	G/A	0.004	
1144	**C/G**	**0.476**	**rs713598**
1190	G/A	0.005	
1238	A/G	0.003	
1361	T/C	0.001	
1367	G/A	0.001	
1402	T/G	0.024	
1415	A/G	0.007	
1568	A/G	0.021	
1648	A/C	0.001	
1703	G/A	0.003	
1784	**C/T**	**0.311**	**rs1726866**
1801	C/T	0.001	
1808	T/G	0.001	
1819	C/T	0.013	
1821	C/T	0.001	
1822	G/T	0.001	
1885	**G/A**	**0.462**	**rs10246939**
1931	T/C	0.001	
2012	G/A	0.003	
2119	A/G	0.003	
2410	C/G	0.001	

Nucleotide Position	Nucleotide Variation	Minor Allele Frequency	rs numbers
2431	C/T	0.023	rs73740218
2508	T/C	0.001	
2614	T/G	0.004	
2687	T/C	0.001	
2872	C/T	0.040	rs73540580
2885	G/A	0.001	
2927	C/T	0.001	
2933	T/C	0.001	

Table S2. Minor Allele Frequency of SNPs in global populations**.** We identified 61 SNPs in the 2,975 bp region in African and non-African populations. Allelic variation, nucleotide position, and minor allele frequency are given for each marker; SNP markers with corresponding rs numbers in the dbSNP database are also indicated in this table. Markers 33, 43 and 49 in bold represent the common SNPs associated with PTC bitter taste sensitivity.

	Cameroon	**Cameroon**	**Cameroon**	**Cameroon**	**Cameroon**	**Kenya**	**Kenya**	**Kenya**	**Kenya**	**Tanzania**
Nucleotide Position	**AA** (2N=146)	**Fulani** (2N=96)	**NK** (2N=124)	**Pygmy** (2N=124)	**NS** (2N=52)	**AA** (2N=264)	**NK** (2N=68)	**NS** (2N=260)	**Luo** (2N=42)	**KS** (2N=46)
8	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000
74	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000
76	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
102	0.014	0.010	0.016	0.081	0.019	0.008	0.015	0.012	0.068	0.000
147	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.012	0.000	0.000
152	0.000	0.000	0.000	0.000	0.019	0.000	0.000	0.000	0.000	0.000
308	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000
343	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000
385	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000
392	0.000	0.000	0.024	0.000	0.000	0.000	0.000	0.000	0.000	0.000
393	0.021	0.010	0.016	0.081	0.019	0.083	0.061	0.038	0.068	0.000
422	0.062	0.021	0.097	0.081	0.038	0.129	0.076	0.108	0.295	0.043
428	0.007	0.000	0.000	0.000	0.000	0.076	0.045	0.027	0.000	0.000
448	0.007	0.000	0.024	0.008	0.000	0.011	0.000	0.015	0.000	0.000
586	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000
656	0.000	0.000	0.008	0.032	0.000	0.000	0.000	0.008	0.023	0.000
681	0.021	0.010	0.016	0.089	0.019	0.083	0.061	0.038	0.068	0.000
767	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000
768	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
775	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
820	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.022
843	0.027	0.021	0.024	0.032	0.038	0.004	0.045	0.004	0.023	0.022
855	0.014	0.010	0.016	0.081	0.019	0.008	0.015	0.012	0.068	0.000
907	0.007	0.000	0.024	0.008	0.000	0.011	0.000	0.015	0.000	0.000
941	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000
1016	0.000	0.000	0.024	0.008	0.000	0.000	0.000	0.000	0.000	0.000

	Cameroon	Cameroon	Cameroon	Cameroon	Cameroon	Kenya	Kenya	Kenya	Kenya	Tanzania
Nucleotide Position	**AA** (2N=146)	**Fulani** (2N=96)	**NK** (2N=124)	**Pygmy** (2N=124)	**NS** (2N=52)	**AA** (2N=264)	**NK** (2N=68)	**NS** (2N=260)	**Luo** (2N=42)	**KS** (2N=46)
1092	0.000	0.010	0.016	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1105	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000
1144	0.493	0.448	0.387	0.460	0.404	0.492	0.485	0.435	0.341	0.413
1190	0.007	0.010	0.024	0.000	0.019	0.000	0.000	0.004	0.000	0.000
1238	0.007	0.000	0.000	0.000	0.000	0.004	0.015	0.004	0.000	0.000
1361	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000
1367	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000
1402	0.021	0.010	0.000	0.008	0.038	0.072	0.015	0.019	0.000	0.043
1568	0.007	0.000	0.000	0.000	0.000	0.076	0.045	0.027	0.000	0.000
1648	0.007	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1703	0.000	0.000	0.000	0.032	0.000	0.004	0.000	0.000	0.000	0.000
1784	0.267	0.302	0.298	0.210	0.365	0.318	0.227	0.312	0.182	0.261
1801	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000
1808	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
1819	0.007	0.000	0.065	0.000	0.019	0.023	0.000	0.008	0.045	0.000
1822	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000
1885	0.486	0.448	0.395	0.460	0.423	0.470	0.470	0.412	0.341	0.413
1931	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000
2012	0.000	0.000	0.000	0.032	0.000	0.000	0.000	0.000	0.000	0.000
2119	0.000	0.010	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000
2410	0.000	0.000	0.000	0.016	0.000	0.000	0.000	0.000	0.000	0.000
2431	0.089	0.073	0.016	0.000	0.038	0.023	0.015	0.004	0.000	0.000
2508	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2614	0.000	0.000	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000
2687	0.000	0.000	0.000	0.008	0.000	0.004	0.000	0.000	0.000	0.000
2872	0.048	0.021	0.065	0.024	0.019	0.038	0.030	0.054	0.273	0.022
2885	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.023	0.000

	Cameroon	**Cameroon**	**Cameroon**	**Cameroon**	**Cameroon**	**Kenya**	**Kenya**	**Kenya**	**Kenya**	**Tanzania**
Nucleotide Position	**AA** (2N=146)	**Fulani** (2N=96)	**NK** (2N=114)	**Pygmy** (2N=124)	**NS** (2N=52)	**AA** (2N=264)	**NK** (2N=68)	**NS** (2N=260)	**Luo** (2N=42)	**KS** (2N=46)
2927	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table S3. Minor Allele Frequency of SNPs in linguistically and/or culturally distinct African populations. Niger-Kordofanian (NK), Nilo-Saharan (NS), Afroasiatic (AA), and Khoesan (KS) represent groups of populations composed of individuals that share genetic and linguistic similarities. Genetically distinct populations within language families were classified as separate groups (for example, Fulani Niger-Kordofanian-speakers, Pygmy Niger-Kordofanian-speakers, and Luo Nilo-Saharan-speakers)(Tishkoff et al. 2009). 2N is the number of chromosomes analyzed in each linguistically and/or culturally defined group.

Nucleotide Position	17	93	106	145	191	239	362	368	403	416	569	649	704	785	802	809	820	822	823	886	932	
Amino Acid Position	6	31	36	49	64	80	121	123	135	139	190	217	235	262	268	270	274	274	275	296	311	
Variants	G A	G A	G A	C G	G A	A G	T C	G A	T G	A G	A G	A C	G A	C T	C T	T G	C T	C T	G A	G A	T C	**Haplotype Frequency**
H1 (AAI)	R	L	V	A	R	H	L	R	W	K	N	T	R	A	P	L	R	R	D	I	I	0.1090
H2 (AAV)	R	L	V	A	R	H	L	R	W	K	N	T	R	A	P	L	R	R	D	V	I	0.0135
H3 (AAI)	R	L	V	A	R	H	L	R	W	K	N	T	R	A	P	L	C	R	D	I	I	0.0135
H4 (AVI)	R	L	V	A	R	H	L	R	W	K	N	T	R	V	P	L	R	R	D	I	I	0.3001
H5 (AVI)	R	L	V	A	R	H	L	R	W	K	N	T	R	V	P	L	R	R	D	I	T	0.0007
H6 (AVI)	R	L	V	A	R	H	L	R	W	K	N	T	R	V	P	L	R	R	D	I	I	0.0007
H7 (AVI)	R	L	V	A	R	H	L	R	W	K	N	T	R	V	P	R	R	R	D	I	I	0.0007
H8 (AVI)	R	L	V	A	R	H	L	R	W	K	N	P	R	V	P	L	R	R	D	I	I	0.0013
H9 (AAI)	R	L	V	A	R	H	L	R	G	K	N	T	R	A	P	L	R	R	D	I	I	0.0236
H10 (AAI)	R	L	V	A	R	H	P	R	W	K	N	T	R	A	P	L	R	R	D	I	I	0.0007
H11 (AAI)	R	L	V	A	R	R	L	R	W	K	N	T	R	A	P	L	R	R	D	I	I	0.0020
H12 (AAI)	R	L	V	A	R	R	L	H	W	K	N	T	R	A	P	L	R	R	N	I	I	0.0007
H13 (AVI)	R	L	V	A	Q	H	L	R	W	K	N	T	R	V	P	L	R	R	D	I	I	0.0047
H14 (PAV)	R	L	V	P	R	H	L	R	W	K	N	T	R	A	P	L	R	R	D	V	I	0.4872
H15 (PAV)	R	L	V	P	R	H	L	R	W	K	N	T	R	A	S	L	R	R	D	V	I	0.0013
H16 (PAV)	R	L	V	P	R	H	L	R	W	K	N	T	H	A	P	L	R	R	D	V	I	0.0034
H17 (PAV)	R	L	V	P	R	H	L	R	W	K	S	T	R	A	P	L	R	R	D	V	I	0.0209
H18 (PAV)	R	L	V	P	R	H	L	R	W	R	N	T	R	A	P	L	R	R	D	V	I	0.0074
H19 (PAV)	R	L	I	P	R	H	L	R	W	K	N	T	R	A	P	L	R	R	D	V	I	0.0040
H20 (AAI)	R	L	V	A	R	H	L	R	W	K	N	T	R	A	P	L	R	R	D	I	I	0.0020
H21 (AVI)	H	L	V	A	R	H	L	R	W	K	N	T	R	V	P	L	R	R	D	I	I	0.0027

Table S4. Haplotype and nucleotide variation in the *TAS2R38* coding region. The first row in this table depicts variable nucleotide positions in the coding region of *TAS2R38* and the second row denotes the corresponding amino acid position. The third row indicates nucleotide variation with the nucleotides in black representing the major allele and the nucleotide in red representing the minor allele. Listed below the nucleotide polymorphisms are rows of haplotypes with their corresponding haplotype frequencies summarized at the end of each row. The letters within each haplotype represent amino acids. Amino acids in red are nonsynonymous amino acid substitutions corresponding to the nonsynonymous nucleotide variants near the top of each column. Amino acids in green represent synonymous amino acid substitutions. Columns shaded in yellow indicate the three common variable nucleotide/amino acid sites associated with PTC sensitivity.

Population	Domain	*2N*	Sites	*S*	π	θ	Tajima's *D*	
Cameroon								
Afro-Asiatic	Coding	146	1002	10	0.00152	0.0018	-0.37	
	5' Non-coding		999	10	0.00036	0.0018	-1.94	*
	3' Non-coding		974	2	0.00026	0.0004	-0.42	
	Total (2975bp)		2975	22	0.00072	0.0013	-1.29	
Fulani (Niger-Kordofanian)	Coding	96	1002	8	0.00153	0.0016	-0.04	
	5' Non-coding		999	7	0.00019	0.0014	-2.04	*
	3' Non-coding		974	3	0.00020	0.0006	-1.19	
	Total (2975bp)		2975	18	0.00064	0.0012	-1.30	
Pygmy (Niger-Kordofanian)	Coding	124	1002	7	0.00152	0.0013	0.39	
	5' Non-coding		999	15	0.00142	0.0028	-1.32	
	3' Non-coding		974	5	0.00026	0.0010	-1.50	
	Total (2975bp)		2975	27	0.00107	0.0017	-1.06	
Niger-Kordofanian	Coding	124	1002	7	0.00162	0.0013	0.53	
	5' Non-coding		999	11	0.00057	0.0021	-1.86	*
	3' Non-coding		974	3	0.00019	0.0006	-1.18	
	Total (2975bp)		2975	21	0.00080	0.0013	-1.14	
Nilo-Saharan	Coding	52	1002	6	0.00161	0.0016	0.54	
	5' Non-coding		999	7	0.00034	0.0011	-2.04	*
	3' Non-coding		974	2	0.00012	0.0005	-1.31	
	Total (2975bp)		2975	15	0.00070	0.0011	-1.15	

Population	Domain	$2N$	Sites	S	π	θ	Tajima's D	
All Cameroonians	Coding	542	1002	15	0.00157	0.0022	-0.65	
	5' Non-coding		999	21	0.00062	0.0031	-1.97	*
	3' Non-coding		974	9	0.00022	0.0013	-1.17	#
	Total (2975bp)		2975	45	0.00081	0.0022	-1.73	#
Kenya								
Afro-Asiatic	Coding	264	1002	8	0.00177	0.0013	0.77	
	5' Non-coding		999	12	0.00075	0.0019	-1.41	
	3' Non-coding		974	3	0.00013	0.0005	-1.14	
	Total (2975bp)		2975	23	0.00088	0.0012	-0.75	
Nilo-Saharan	Coding	260	1002	13	0.00157	0.0021	-0.62	
	5' Non-coding		999	11	0.00052	0.0017	-1.61	#
	3' Non-coding		974	3	0.00012	0.0005	-1.17	
	Total (2975bp)		2975	27	0.00074	0.0014	-1.34	
Niger Kordofanian	Coding	68	1002	6	0.00152	0.0013	0.49	
	5' Non-coding		999	7	0.00058	0.0014	-1.47	
	3' Non-coding		974	2	0.00009	0.0004	-1.32	
	Total (2975bp)		2975	15	0.00073	0.0010	-0.87	
Luo (Nilo Saharan)	Coding	42	1002	4	0.00131	0.0009	1.00	
	5' Non-coding		999	7	0.00099	0.0015	-0.97	
	3' Non-coding		974	2	0.00046	0.0005	-0.03	
	Total (2975bp)		2975	13	0.00092	0.0010	-0.19	
All Kenyans	Coding	634	1002	14	0.00163	0.0020	-0.40	
	5' Non-coding		999	14	0.00066	0.0020	-1.46	
	3' Non-coding		974	5	0.00015	0.0007	-1.33	
	Total (2975bp)		2975	33	0.00082	0.0016	-1.23	

Population	**Domain**	***2N***	**Sites**	***S***	π	θ	**Tajima's *D***	
Tanzania								
Khoesan-speakers (Hadza/Sandawe)	Coding	46	1002	4	0.00147	0.0009	1.42	
	5' Non-coding		999	3	0.00017	0.0007	-1.58	#
	3' Non-coding		974	1	0.00004	0.0002	-1.11	
	Total (2975bp)		2975	8	0.00057	0.0006	-0.20	
All Africans	Coding	1222	1002	19	0.00160	0.0025	-0.80	
	5' Non-coding		999	25	0.00064	0.0033	-1.92	*
	3' Non-coding		974	10	0.00018	0.0013	-1.69	#
	Total (2975bp)		2975	54	0.00082	0.0024	-1.73	*
Non-Africans								
	Coding	264	1002	5	0.00155	0.0008	1.71	
	5' Non-coding		999	5	0.00005	0.0008	-1.75	#
	3' Non-coding		974	3	0.00004	0.0005	-1.42	
	Total (2975bp)		2975	13	0.00055	0.0007	-0.38	

Table S5. Nucleotide diversity statistics and tests of neutrality**.** Tajima's D was calculated for Africans and for non-Africans in the coding and non-coding regions of the *TAS2R38 gene.* 2N is the number of chromosomes analyzed in each population. S is number of segregating sites. π is the average pairwise nucleotide differences per site. θ is Waterson's estimate of heterozygosity per site. Significance was calculated with coalescence simulations assuming constant population size. Significant p-values ($p < 0.05$) are indicated by * while marginally significant p-values ($0.05<P<0.01$) are indicated by #.

(A)

			West Central Africa (Cameroon)			
Haplotypes	**Afroasiatic (2N=146)**	**Nilo-Saharan (2N=52)**	**Niger-Kordofanian (2N=124)**	**Fulani (2N=96)**	**Pygmy (2N=124)**	**Total number of haplotypes (2N=542)**
1 (**AAI**)	27 (0.185)	8 (0.154)	28 (0.226)	12 (0.125)	30 (0.242)	105 (0.194)
2 (**AAV**)	1 (0.007)	1 (0.0192)	1 (0.008)	0	0	3 (0.006)
3 (**AAI**)	1 (0.007)	1 (0.0192)	8 (0.065)	0	0	10 (0.018)
4 (**AVI**)	37 (0.253)	18 (0.346)	31 (0.250)	26 (0.271)	25 (0.202)	137 (0.253)
5 (**AVI**)	0	0	0	0	0	0
6 (**AVI**)	0	0	0	0	0	0
7 (**AVI**)	0	0	0	1 (0.010)	0	1 (0.002)
8 (**AVI**)	1 (0.007)	0	0	1 (0.010)	0	2 (0.004)
9 (**AAI**)	3 (0.021)	2 (0.039)	0	1 (0.010)	1 (0.008)	7 (0.013)
10 (**AAI**)	0	0	0	0	0	0
11 (**AAI**)	1 (0.007)	0	0	0	0	1 (0.002)
12 (**AAI**)	0	0	0	0	0	0
13 (**AVI**)	1 (0.007)	1 (0.0192)	3 (0.024)	1 (0.010)	0	6 (0.011)
14 (**PAV**)	72 (0.493)	21 (0.403)	48 (0.387)	53 (0.552)	57 (0.460)	251 (0.463)
15 (**PAV**)	1 (0.007)	0	0	0	0	1 (0.002)
16 (**PAV**)	0	0	0	0	4 (0.032)	4 (0.007)
17 (**PAV**)	1 (0.007)	0	0	0	0	1 (0.002)
18 (**PAV**)	0	0	0	0	0	0
19 (**PAV**)	0	0	0	0	6 (0.048)	6 (0.011)
20 (**AAI**)	0	0	2 (0.016)	1 (0.010)	0	3 (0.006)
21 (**AVI**)	0	0	3 (0.024)	0	1 (0.008)	4 (0.007)

(B)

			East Africa (Kenya and Tanzania)			
Haplotypes	**Afroasiatic (2N=264)**	**Nilo-Saharan (2N=260)**	**Niger-Kordofanian (2N=68)**	**Luo (2N=42)**	**Hadza/Sandawe (Khoesan) (2N=46)**	**Total number of haplotypes (2N=680)**
H1 (**AAI**)	14 (0.053)	17 (0.065)	15 (0.221)	4 (0.095)	5 (0.109)	55 (0.081)
H2 (**AAV**)	10 (0.0379)	6 (0.023)	1 (0.015)	0	0	17 (0.025)
H3 (**AAI**)	6 (0.023)	2 (0.008)	0	2 (0.048)	0	10 (0.0147)
H4 (**AVI**)	84 (0.318)	79 (0.304)	15 (0.221)	8 (0.190)	12 (0.261)	198 (0.291)
H5 (**AVI**)	0	1 (0.004)	0	0	0	1 (0.001)
H6 (**AVI**)	0	0	0	0	0	0
H7 (**AVI**)	0	0	0	0	0	0
H8 (**AVI**)	0	0	0	0	0	0
H9 (**AAI**)	19 (0.072)	5 (0.019)	1 (0.015)	0	2 (0.043)	27 (0.040)
H10 (**AAI**)	0	1 (0.004)	0	0	0	1 (0.001)
H11 (**AAI**)	1 (0.004)	0	1 (0.015)	0	0	2 (0.003)
H12 (**AAI**)	0	1 (0.004)	0	0	0	1 (0.001)
H13 (**AVI**)	0	1 (0.004)	0	0	0	1 (0.001)
H14 (**PAV**)	109 (0.413)	139 (0.535)	32 (0.471)	28 (0.667)	27 (0.587)	335 (0.493)
H15 (**PAV**)	0	1 (0.004)	0	0	0	1 (0.001)
H16 (**PAV**)	1 (0.004)	0	0	0	0	1 (0.001)
H17 (**PAV**)	20 (0.076)	7 (0.027)	3 (0.044)	0	0	30 (0.044)
H18 (**PAV**)	0	0	0	0	0	0
H19 (**PAV**)	0	0	0	0	0	0
H20 (**AAI**)	0	0	0	0	0	0
H21 (**AVI**)	0	0	0	0	0	0

(C)

			NON-AFRICANS			
Haplotypes	**Americas (2N=26)**	**Europe (2N=40)**	**Middle East (2N=40)**	**Pakistan (2N=16)**	**East Asia (2N=142)**	**Total number of haplotypes (2N=264)**
H1 (**AAI**)	0	1 (0.025)	2 (0.05)	0	0	3 (0.011)
H2 (**AAV**)	0	0	0	0	0	0
H3 (**AAI**)	0	0	0	0	0	0
H4 (**AVI**)	4 (0.154)	28 (0.70)	18 (0.45)	8 (0.5)	53 (0.373)	111 (0.420)
H5 (**AVI**)	0	0	0	0	0	0
H6 (**AVI**)	0	0	0	1 (0.0625)	0	1 (0.004)
H7 (**AVI**)	0	0	0	0	0	0
H8 (**AVI**)	0	0	0	0	0	0
H9 (**AAI**)	0	0	0	0	0	0
H10 (**AAI**)	0	0	0	0	0	0
H11 (**AAI**)	0	0	0	0	0	0
H12 (**AAI**)	0	0	0	0	0	0
H13 (**AVI**)	0	0	0	0	0	0
H14 (**PAV**)	19 (0.731)	11 (0.275)	20 (0.50)	7 (0.4375)	81 (0.570)	138 (0.523)
H15 (**PAV**)	0	0	0	0	0	0
H16 (**PAV**)	0	0	0	0	0	0
H17 (**PAV**)	0	0	0	0	0	0
H18 (**PAV**)	3 (0.115)	0	0	0	8 (0.056)	11 (0.042)
H19 (**PAV**)	0	0	0	0	0	0
H20 (**AAI**)	0	0	0	0	0	0
H21 (**AVI**)	0	0	0	0	0	0

Table S6. PTC haplotype counts in diverse populations from West Central Africa (Cameroon)(A), East Africa (Kenya and Tanzania) and non-Africans (C). The number of occurrences of each haplotype and corresponding frequencies in parentheses are given for Afroasiatic-, Nilo-Saharan-, Niger-Kordofanian-, Luo Nilo-Saharan- , Khoesan- speaking and non-African populations. Total haplotype counts and corresponding frequencies in parentheses are summarized in the last column of Table 4. Shaded rows indicate the common haplotypes associated with PTC sensitivity.

Regression Statistics	Raw Data	Significance (ANOVA)	Sex and Age Adjusted Data	Significance (ANOVA)
R squared	0.355	**$p < 0.001$**	0.339	**$p < 0.001$**
Adjusted R square	0.353		0.338	
Standard Error	1.973		2.049	
Observations	463		463	

Table S7. Linear Regression Statistics for both raw PTC data and age-sex corrected PTC data. An ANOVA was performed to test the significance of the fit of the regression model to the data for 463 African individuals. Significant p-values for both regression analyses using the raw PTC data and the age/sex corrected PTC data are shown here in bold.

(A)

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
PTC_Phenotype	463	8.21	2.420	1	13
GENOTYPE	463	3.85	2.717	1	10

Test Statistics[b,c]

			PTC_ Phenotype
Chi-Square			163.266
df			8
Asymp. Sig.			.000
Monte Carlo Sig.	Sig.		.000[a]
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000

a. Based on 10000 sampled tables with starting seed 2000000.

b. Kruskal Wallis Test

c. Grouping Variable: GENOTYPE

(B)

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
Corrected_ Phenotype_Score	463	7.9589	2.51818	-.23	13.65
GENOTYPE	463	3.85	2.717	1	10

Test Statistics[b,c]

			Corrected_ Phenotype_ Score
Chi-Square			138.577
df			8
Asymp. Sig.			.000
Monte Carlo Sig.	Sig.		.000[a]
	99% Confidence Interval	Lower Bound	.000
		Upper Bound	.000

a. Based on 10000 sampled tables with starting seed 299883525.

b. Kruskal Wallis Test

c. Grouping Variable: GENOTYPE

Table S8. Kruskal Wallis non-parameteric test for genotype/phenotype association**.** Association between genotype and phenotype was also assessed using the nonparametric Kruskal Waliis test in SPSS. In this analysis, raw PTC scores (A) were entered as the Test Variable and genotypes were entered as the grouping variable. Similarly, PTC scores adjusted by age and sex (B) were entered in a separate analysis as the Test Variable, while genotypes were entered as the grouping variable. Significant p-values (p<0.001) are indicated by the superscript a.

Model	R	R square	Adjusted R square	Standard Error of the Estimate
1	0.587	0.345	0.342	1.963

Model	Sum of Squares	df	Mean Square	F	Significance
Regression	932.626	2	466.313	120.97	0.000
Residual	1773.206	461	3.855		
Total	2705.832	463			

		Unstandardized Coefficients		Standardized Coefficients		
Model		B	Standard Error	Beta	t	Significance
1	(Constant)	7.423	0.138		53.593	0.000**
	Additive Effect	2.089	0.138	0.576	15.085	0.000**
	Dominance Effect	1.105	0.185	0.227	5.961	0.000**

Table S9. Analysis of PAV Dominance**.** The PTC scores adjusted by age and sex were entered into the regression model as the dependent variable. The recoded additive and dominance coefficients were entered as independent variables. Significant p-values ($p<0.001$) are indicated by **.

Feature Key	Positions	Length (number of amino acids)	Description	Number of nonsynonymous variants	Nucleotide Position (Amino Acid Position)
Topological domain	1-17	17	Extracellular	2	17(6);23(8)
Transmembrane	18-38	21	TM1	1	106(36)
Topological domain	39-55	17	Cytoplasmic	1	145(49)*
Transmembrane	56-76	21	TM2	1	**191(64)**
Topological domain	77-94	18	Extracellular	1	**239(80)**
Transmembrane	95-115	21	TM3	0	
Topological domain	116-142	27	Cytoplasmic	4	**362(121)**;369(123);**403(135)**;416(139)
Transmembrane	143-163	21	TM4	0	
Topological domain	164-190	27	Extracellular	0	569(190)
Transmembrane	191-211	21	TM5	0	
Topological domain	212-251	40	Cytoplasmic	2	649(217);704(235)
Transmembrane	252-272	21	TM6	3	785(262)*;802(268);809(270)
Topological domain	273-276	4	Extracellular	2	820(274);823(275)
Transmembrane	277-297	21	TM7	1	886(296)*
Topological domain	298-333	36	Cytoplasmic	1	932(311)

Table S10 The distribution of nonsynonymous SNP variation within the protein structure of the *TAS2R38* receptor. *TAS2R38* encodes a transmembrane (TM) protein that recognizes bitter tasting ligands outside the cell (Extracellular) and activates signal transduction inside the cell (Cytoplasmic). This table indicates the amino acid domains of the *TAS2R38* receptor protein in relation to inner and outer surfaces of the cell membrane. The number of identified nonsynonymous variation together with their nucleotide position and amino acid position (in parentheses) are given. The positions with asterisks indicate the common variants associated with PTC sensitivity, and the numbers in bold are the novel rare amino acid variants that were shown to modify PTC sensitivity based on statistical analyses and the Polyphen-2 software (Ramensky, Bork, and Sunyaev 2002).

Mutation #	Age of Mutation (TMRCA in years)	Standard Deviation of TMRCA
1	9,187	20,403
2	26,010	43,908
3	28,397	25,533
4 (49)	**1,312476**	**242,211**
5	11,931	23,266
6	26,726	47,726
7	73,140	99,509
8	25,652	54,169
9	49,158	65,385
10	5,918	17,181
11	14,675	21,715
12 (262)	**336,471**	**89,845**
13	4,689	16,584
14	52,254	6,5623.8
15 (296)	**1,035,663**	**267,268**

Table S11. Inferred ages of mutations in the *TAS2R38* coding region for global populations. The mutations with numbers in parentheses are the amino acid positions of the common variants associated with PTC sensitivity. Mutation numbers in Table correspond to the mutation numbers (in small print) on the gene tree in Figure 4.

Sample	Fixed	Polymorphic	Fisher's Exact Test
West Central Africa (Cameroon) (2N=542)			
Synonymous	3	2	
Nonsynonymous	3	18	p = 0.062 #
East Africa (Kenya and Tanzania) (2N=680)			
Synonymous	3	1	
Nonsynonymous	3	18	**p = 0.031 ***
All Africans (2N=1222)			
Synonymous	3	2	
Nonsynonymous	3	22	**p = 0.041 ***
Non-Africans (2N=264)			
Synonymous	3	1	
Nonsynonymous	3	8	p = 0.560

Table S12. McDonald-Kreitman test of neutrality for the coding region. The 2 X 2 contingency table contains the number of synonymous and nonsynonymous polymorphisms within a species, and the number of synonymous and nonsynonymous changes that are fixed between species. Significance was determined using Fisher's Exact Test. Statistically significant ($P<0.05$) p-values are indicated by *, while marginally significant ($0.05<p<0.1$) p-values are indicated by #

	Observed Statistic	Population Growth Scenarios						
Sample	Tajima's D	A	B	C	D	E	F	G
Pooled Africans	-0.8024	0.232	0.147	**0.014 ***	**0.027 ***	**0.016 ***	NA	NA
Non-Africans	1.7114	NA	NA	NA	NA	NA	0.064	**0.014 ***

Table S13. Observed Tajima's D statistic for the coding region of *TAS2R38* and corresponding p-values for different scenarios of population growth. Using the ms program (Hudson 1990), seven demographic scenarios were simulated: the first five scenarios (A-D) are specific to African populations and the last two scenarios (E-G) are specific to non-African populations: (A) represents a 10-fold increase in population size in Africans (from an initial population size of 10,000 individuals) starting at 70 kya until the present (B) represents a 15-fold increase in population size of Africans (from an initial population size of 10,000 individuals) starting at 70 kya until the present (C) represents a 20-fold increase in population size of Africans (from an initial population size of 10,000 individuals) starting at 70 kya until the present, (D) represents a 40-fold increase in population size (from an initial population size of 10,000 individuals) starting at 70 kya until the present, and (E), (F) and (G), respectively, represent a population bottleneck (from an initial population size of 10,000 individuals and decreasing to 3,000 individuals) at 60,000 years ago, corresponding to the time modern humans migrated out of Africa, followed by a 10-fold (E) , 20-fold (F) and 40-fold (G) increase in population size starting at 50 kya until the present. A p-value of 0.025 or less indicates that the observed statistic is significantly greater than expected consistent with a scenario of balancing selection and is denoted by *

Supplemental References

Ethnologue Languages of the world, http://www.ethnologue.com/.

Fedders, A., and C. Salvadori. 1998. Peoples and Cultures of Kenya. Transafrica, Nairobi.

Hudson, R. R. 1990. Gene Genealogies and the Coalescent Process. Oxford University Press, Oxford.

Ramensky, V., P. Bork, and S. Sunyaev. 2002. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894-3900.

Tishkoff, S. A., F. A. Reed, F. R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, J. B. Hirbo, A. A. Awomoyi, J. M. Bodo, O. Doumbo, M. Ibrahim, A. T. Juma, M. J. Kotze, G. Lema, J. H. Moore, H. Mortensen, T. B. Nyambo, S. A. Omar, K. Powell, G. S. Pretorius, M. W. Smith, M. A. Thera, C. Wambebe, J. L. Weber, and S. M. Williams. 2009. The genetic structure and history of Africans and African Americans. Science 324:1035-1044.