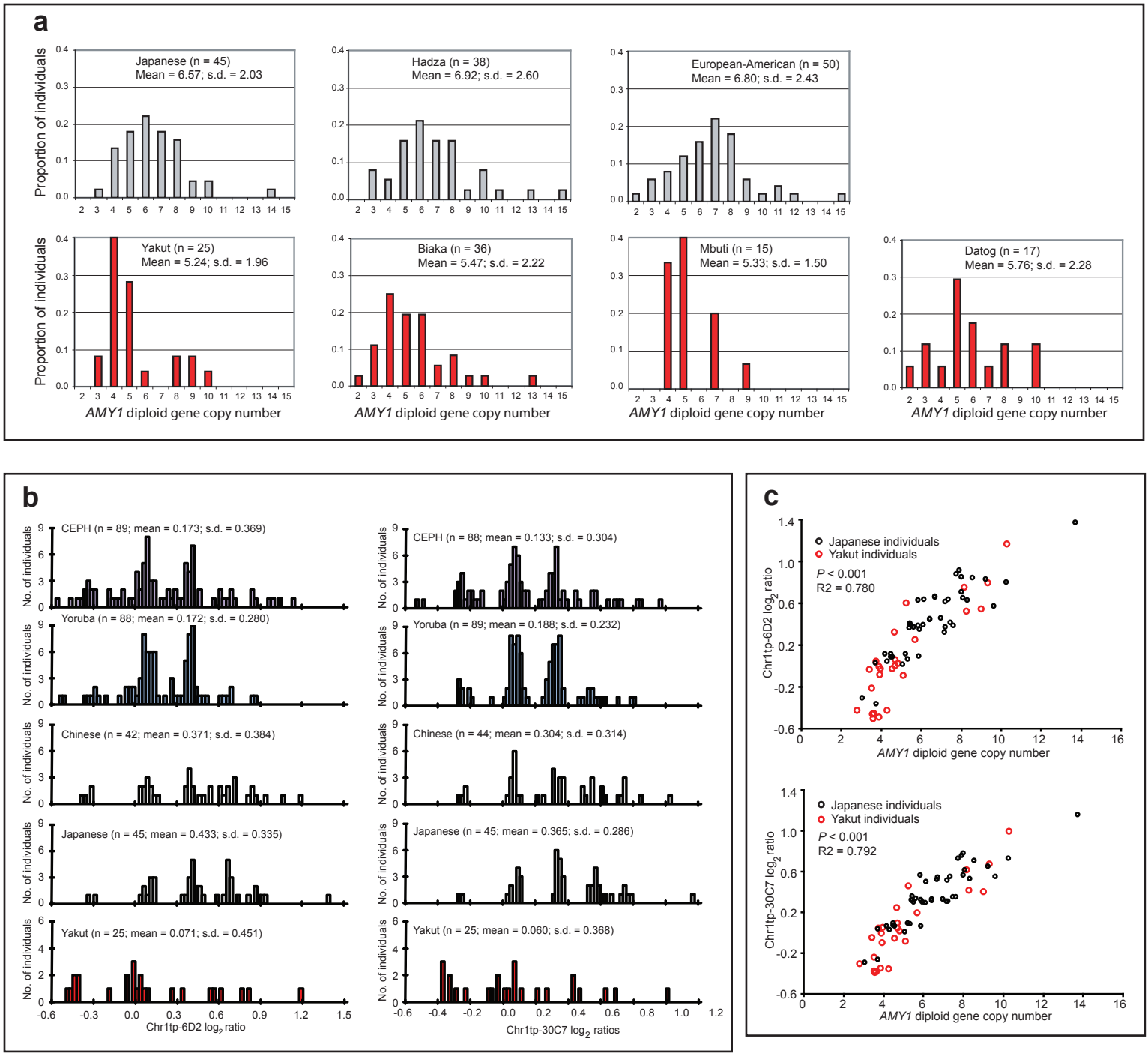
Supplementary information for:

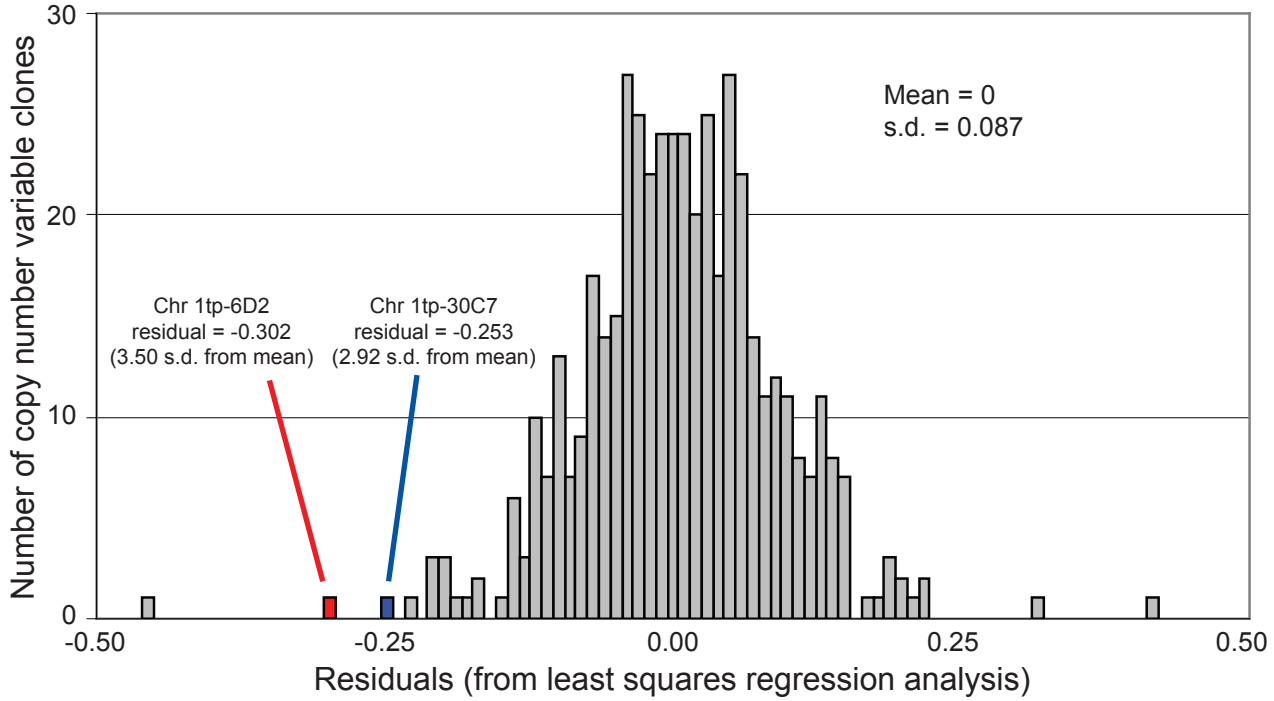
GH Perry, NJ Dominy, KG Claw, AS Lee, H Fiegler, R Redon, J Werner, FA Villanea, JL Mountain, R Misra, NP Carter, C Lee, and AC Stone

Diet and the evolution of human amylase gene copy number variation

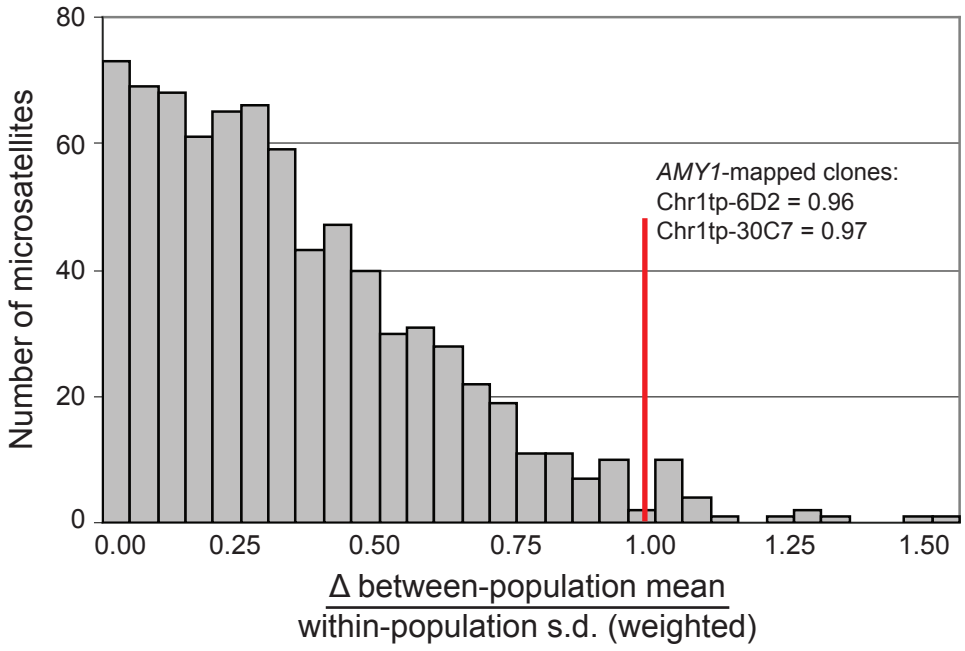
Supplementary Figure 1. qPCR and aCGH population-level *AMY1* copy number data. **(a)** qPCR-estimated *AMY1* diploid copy number frequency distributions for each of the populations in this study, including 3 populations with high-starch diets (gray bars) and 4 populations with traditionally low-starch diets (red bars). **(b)** Frequency distributions of aCGH relative intensity log2 ratios from *AMY1*-mapped clones Chr1tp-6D2 and Chr1tp-30C7 from aCGH experiments for the low-starch Yakut and the 4 high-starch HapMap populations. The Yakut mean log2 ratios for both clones are lower than those for every HapMap population. The greatest differences for the log2 ratio comparisons are with the Asian populations, who have the closest geographic similarity to the Yakut (Siberia). **(c)** qPCR-estimated diploid *AMY1* gene copy numbers for the Japanese and Yakut populations are significantly positively correlated with relatively intensity log2 ratios for *AMY1*-mapped clones Chr1tp-6D2 and Chr1tp-30C7 from aCGH experiments. The same reference DNA sample (NA10851) was used for all aCGH experiments.



Supplementary Figure 2. Genome-wide analysis of Japanese-Yakut copy number differentiation. We conducted a least squares regression analysis on the distribution of Japanese and Yakut mean log2 ratios from aCGH experiments for all autosomal clones on the WGTP array platform found to be copy number variable in at least one individual from both populations (see Fig. 4b). The log2 ratios are all relative to a single male reference individual (NA10851; European-American/ CEPH). If there is no difference in mean copy number between the two populations at a particular locus, then the population mean log2 ratios will be similar, even if the values are dissimilar from 0 (i.e., the reference individual may not have a copy number that reflects the general population mean at any particular locus). There is a positive relationship between Japanese and Yakut mean log2 ratios with slope ~1 ($P < 0.001$; R2 = 0.65; Yakut mean log2 = -0.0134 + 0.895 Japanese mean log2), as would be expected if the two populations have roughly similar levels of genome-wide differentiation with the population from which the reference individual was sampled. Copy number variants within clones with mean log2 ratios that significantly deviate from this distribution may have been subject to relatively unusual evolutionary pressures in one or both of the populations. Therefore, we evaluated the unusualness of the *AMY1*-mapped clones by comparing the residuals for these clones against the distribution of residuals from all copy number variable clones in the genome. This figure depicts the frequency distribution of the residuals (bins = 0.01) for the 474 autosomal copy number variable clones from the least squares regression analysis. The *AMY1*-mapped clones Chr1tp-6D2 (depicted in red) and Chr1tp-30C7 (depicted in blue) have residuals 3.5 s.d. ($P < 0.001$) and 2.9 s.d. ($P < 0.01$) from the mean, respectively.

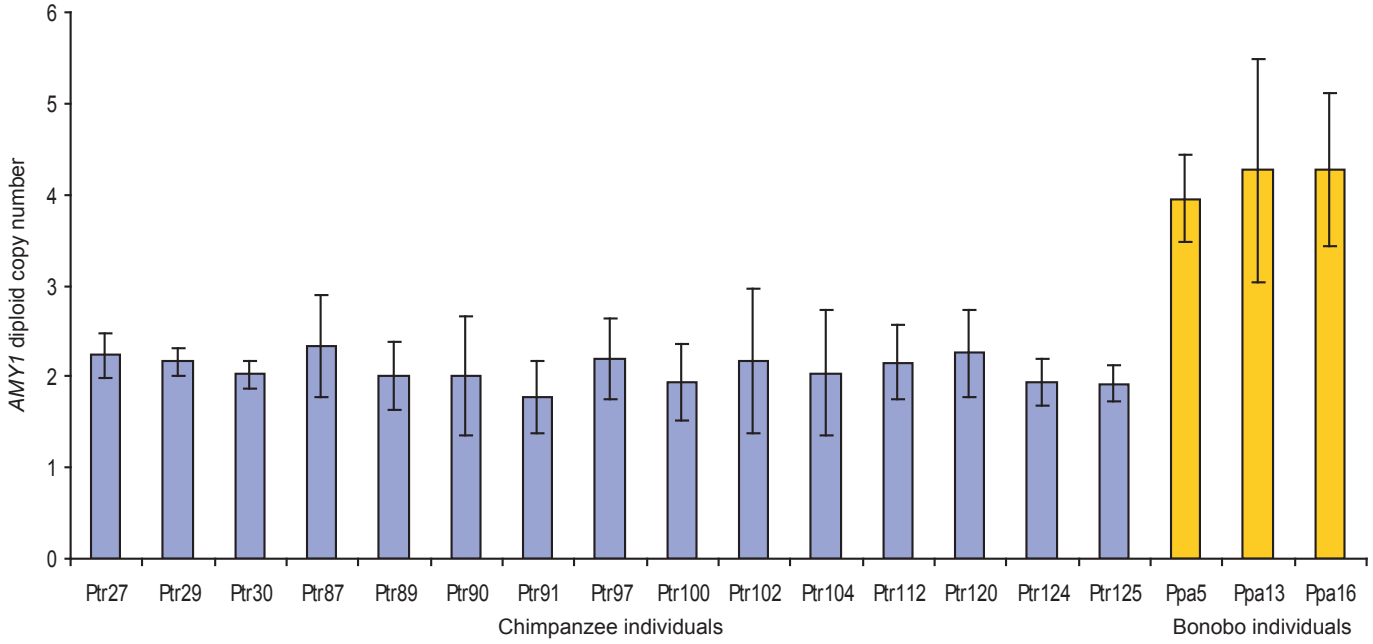


Supplementary Figure 3. Comparison of Japanese-Yakut population differentiation for *AMY1* and 783 genome-wide microsatellites. To make the microsatellite data (for which allele-specific data are available) and *AMY1* copy number data (for which we have only "combined" diploid information for each individual in our population samples, due to current technological limitations) more generally comparable, we combined the allele-specific microsatellite genotypes into one value for each individual for each microsatellite. For example, if one individual has alleles of 188 and 180 bp, the combined value = 368. These combined values were used in the analysis. For each microsatellite locus and the *AMY1*-mapped clones from the WGTP aCGH experiments, we computed the absolute difference in population mean values, scaled to within-population standard deviation (weighted by population size). Based on this statistic, the level of population differentiation at the *AMY1* locus exceeds that for >97% of the 783 microsatellites.

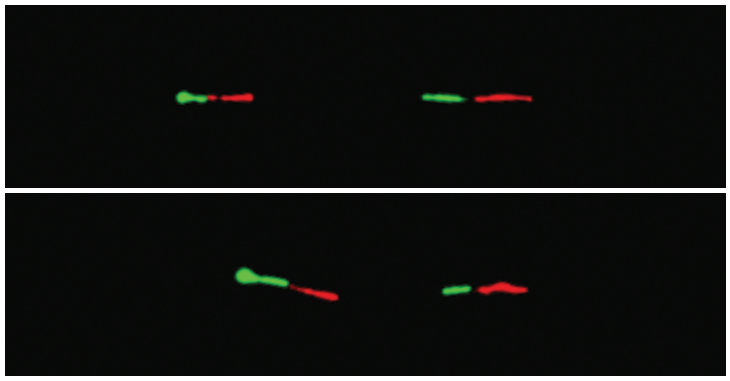


Supplementary Figure 4. *AMY1* copy number in chimpanzees and bonobos. **(a)** There is no evidence for *AMY1* copy number variation among chimpanzees; all 15 individuals examined with qPCR are estimated to have two diploid *AMY1* copies. All chimpanzee individuals were from the western chimpanzee subspecies (*Pan troglodytes verus*). Bonobos are estimated to have four diploid copies. Error bars indicate 2 s.d. **(b)** Fiber FISH confirms the qPCR estimate of four diploid AMY1 copies for bonobo individual Ppa16 (PR00251). The red probe (~10 kb) encompasses the entire *AMY1* gene, and the green probe (~8 kb) is directly upstream of the *AMY1* gene (see Methods).

a

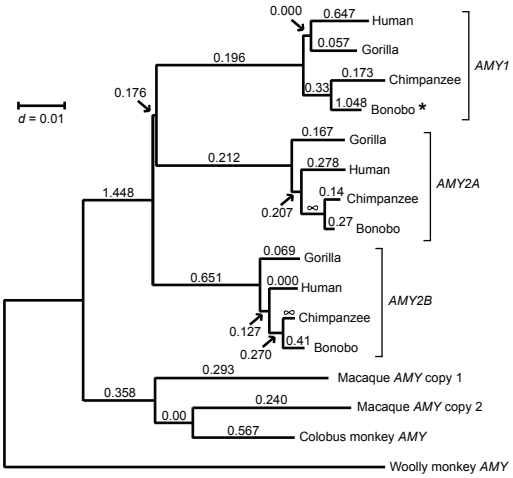


b



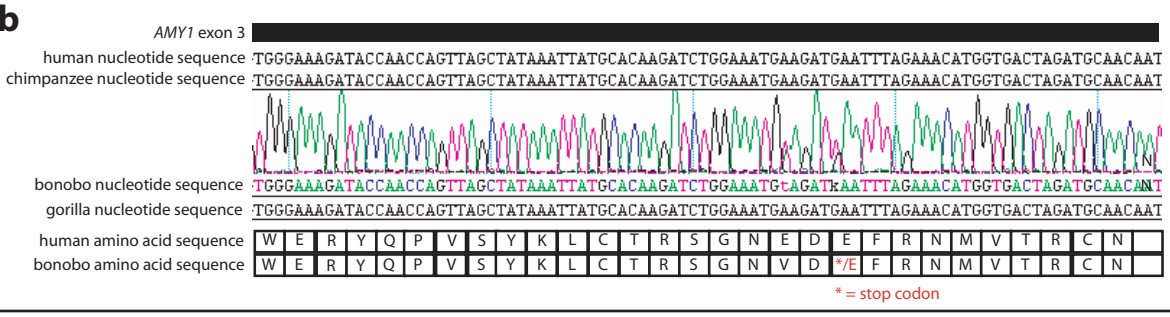
Supplementary Figure 5. Evolution of the primate amylase gene family. **(a)** Phylogenetic reconstruction using the Neighbor-Joining method for primate amylase gene sequences (introns and third positions of exons; 8,005 total sites) from human, chimpanzee, bonobo, gorilla, rhesus macaque, Angolan colobus monkey (*Colobus angolensis*), and common woolly monkey (*Lagothrix lagotricha*). Bootstrap support (10,000 replications) was 99% or 100% for all branches except for the human-gorilla *AMY1* grouping (92%) and the hominoid *AMY1-AMY2A* grouping (54%). Lineage-specific ratios of nonsynonymous substitutions (N) per site to synonymous substitutions (S) per site (dN/dS) estimated from coding regions are given above each branch. dN/dS = ∞ reflects nonzero N and zero S on that lineage (in both such cases, N = 1 and S = 0). For most lineages dN/dS < 1, suggesting an overall history of functional constraint for the primate amylase gene family. However, we identified two frameshift insertions (1 bp each) in exon 7 of bonobo *AMY1* (asterisk; see below), a possible indication of pseudogene status. For this branch dN/dS ~ 1 (N = 5.1, S = 1.4), which is consistent with neutral evolution. **(b)** Nucleotide sequences from human, chimpanzee, bonobo, and gorilla for a selected portion of *AMY1* exon 3. The bonobo sequence exhibits a G/T polymorphism (potentially representing a paralogous sequence variant; see Supplementary Fig. 4) in this exon which is inferred to result in a stop codon in the amino acid sequence for some bonobo *AMY1* transcripts. **(c)** Nucleotide sequences from human, chimpanzee, bonobo, and gorilla for a selected portion of *AMY1* intron 6 and exon 7. There are two single bp insertions within exon 7 of the bonobo sequence, which are inferred to disrupt (by frameshift) all downstream *AMY1* coding sequence.

a

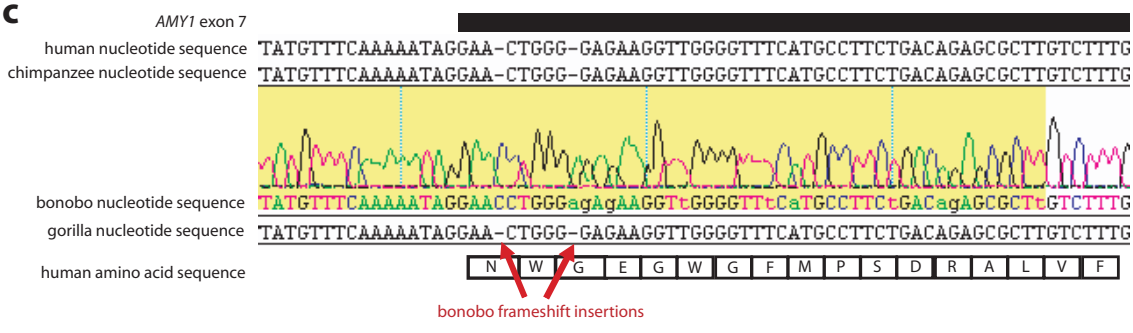


We used a long-range PCR amplification strategy to isolate full-length individual amylase genes from a chimpanzee (PR00496), bonobo (PR00367), gorilla (PR00573), Angolan colobus monkey (PR00099), and woolly monkey (PR00525). All samples were obtained from the Integrated Primate Biomaterials and Information Resource. We then used nested PCR to amplify 2-3 kb fragments, which were purified with Shrimp Alkaline Phosphatase and Exonuclease I (USB Corp.), cycle sequenced with BigDye Terminator Cycle Sequencing Kit version 3.1 (Applied Biosystems), cleaned with isopropanol, and sequenced using an Applied Biosystems 3730 capillary sequencer. PCR and sequencing primers and conditions are provided in Supplementary Table 3 online. Amylase gene sequences were also obtained from the human (hg18) and rhesus macaque (rheMac2) reference genome sequences. Sequences were aligned using ClustalW in MEGA3.1 (S. Kumar et al. 2004, Brief Bioinform 5, 150-63). The Neighbor-Joining method (implemented in MEGA3.1) was used for phylogenetic reconstruction, considering only intron and exon 3rd position sites. Lineage-specific dN/dS ratios were estimated by maximum likelihood using PAML (Z. Yang 1997, Comput Appl Biosci 13, 555-6). The sequences generated for this study have been deposited in GenBank under the accession numbers EU022320-EU022330.

b



c



Supplementary Table 1. qPCR, protein quantification, and aCGH data for high- and low-starch population samples.

Population	Sample	Starch level (high/low)	Diploid *AMY1* copies (qPCR)	Standard Deviaion	Diploid *AMY1* copies (integer)	AMY1 protein mg/mL	Chr1tp-6D2 log2 ratio	Chr1tp-30C7 log2 ratio
European-American	EUR001	High	7.30	0.50	7	2.83		
European-American	EUR002	High	4.26	0.49	4	1.65		
European-American	EUR003	High	7.10	0.93	7	3.85		
European-American	EUR004	High	4.91	0.21	5	1.09		
European-American	EUR005	High	8.15	0.78	8	1.63		
European-American	EUR006	High	11.73	0.55	12	5.17		
European-American	EUR007	High	6.50	0.38	6	3.24		
European-American	EUR008	High	8.44	0.70	8	2.80		
European-American	EUR009	High	5.73	0.39	6	3.30		
European-American	EUR010	High	8.36	0.93	8	4.28		
European-American	EUR011	High	7.63	0.45	8	2.91		
European-American	EUR012	High	6.89	0.51	7	2.89		
European-American	EUR013	High	11.20	0.80	11	3.76		
European-American	EUR014	High	6.18	0.31	6	2.65		
European-American	EUR015	High	7.94	1.12	8	1.70		
European-American	EUR016	High	5.56	0.76	6	3.20		
European-American	EUR017	High	8.53	0.75	9	2.96		
European-American	EUR018	High	9.67	0.56	10	4.87		
European-American	EUR019	High	7.46	1.00	7	4.00		
European-American	EUR020	High	3.41	0.56	3	0.93		
European-American	EUR021	High	2.21	0.50	2	0.22		
European-American	EUR022	High	5.27	0.54	5	1.65		
European-American	EUR023	High	9.14	0.64	9	2.72		
European-American	EUR024	High	7.64	0.22	8	2.46		
European-American	EUR025	High	5.87	0.36	6	1.35		
European-American	EUR026	High	2.61	0.20	3	0.22		
European-American	EUR027	High	5.23	0.59	5	1.37		
European-American	EUR028	High	4.58	0.49	5	2.33		
European-American	EUR029	High	6.62	1.02	7	3.37		
European-American	EUR030	High	5.38	0.59	5	2.24		
European-American	EUR031	High	8.55	0.64	9	3.72		
European-American	EUR032	High	4.66	0.78	5	3.85		
European-American	EUR033	High	8.19	0.97	8	2.50		
European-American	EUR034	High	6.56	0.62	7	5.67		
European-American	EUR035	High	6.26	0.87	6	4.61		
European-American	EUR036	High	7.16	0.73	7	3.41		
European-American	EUR037	High	6.30	1.24	6	5.09		
European-American	EUR038	High	4.30	0.80	4	2.15		

Supplementary Table 1. qPCR, protein quantification, and aCGH data for high- and low-starch population samples.

European-American	EUR039	High	5.56	0.54	6	4.33		
European-American	EUR040	High	3.15	0.44	3	3.13		
European-American	EUR041	High	3.79	0.16	4	4.24		
European-American	EUR042	High	6.96	0.72	7	4.33		
European-American	EUR043	High	14.66	2.04	15	6.04		
European-American	EUR044	High	7.85	0.78	8	1.89		
European-American	EUR045	High	7.57	0.65	8	3.48		
European-American	EUR046	High	7.11	0.72	7	2.43		
European-American	EUR047	High	7.14	0.94	7	2.74		
European-American	EUR048	High	4.22	0.39	4	1.83		
European-American	EUR049	High	10.70	0.75	11	4.48		
European-American	EUR050	High	7.01	0.55	7	3.41		
			Summary:	**n =**	**50**			
				Mean:	**6.80**			
				SD:	**2.43**			
Population	**Sample**	**Starch level (high/low)**	**Diploid *AMY1* copies (qPCR)**	**Standard Deviaion**	**Diploid *AMY1* copies (integer)**	**AMY1 protein mg/mL**	**Chr1tp-6D2 log2 ratio**	**Chr1tp-30C7 log2 ratio**
Hadza	BAR10	High	10.33	0.81	10			
Hadza	BAR11	High	5.94	0.47	6			
Hadza	BAR12	High	7.05	0.78	7			
Hadza	BAR2	High	15.43	2.89	15			
Hadza	BAR3	High	8.43	0.31	8			
Hadza	BAR4	High	9.90	1.10	10			
Hadza	BAR6	High	2.64	0.48	3			
Hadza	BAR7	High	5.06	0.81	5			
Hadza	BAR8	High	6.46	0.54	6			
Hadza	BAR9	High	5.48	1.23	5			
Hadza	END 9	High	6.38	2.81	6			
Hadza	END13	High	5.72	0.45	6			
Hadza	END14	High	7.74	1.29	8			
Hadza	END15	High	9.46	0.67	9			
Hadza	END16	High	6.04	0.37	6			
Hadza	END17	High	3.90	1.30	4			
Hadza	END18	High	7.49	0.63	7			
Hadza	END19	High	8.36	1.11	8			
Hadza	END2	High	4.83	0.60	5			
Hadza	END20	High	4.17	0.64	4			
Hadza	END23	High	7.97	1.03	8			
Hadza	END6	High	10.10	1.40	10			

Supplementary Table 1. qPCR, protein quantification, and aCGH data for high- and low-starch population samples.

Hadza	END7	High	7.71	0.70	8			
Hadza	GOR11	High	11.17	1.07	11			
Hadza	GOR12	High	12.82	1.13	13			
Hadza	GOR23	High	7.20	0.69	7			
Hadza	GOR24	High	6.57	1.19	7			
Hadza	GOR27	High	8.03	1.46	8			
Hadza	GOR29	High	6.77	0.34	7			
Hadza	GOR30	High	5.66	0.98	6			
Hadza	GOR31	High	5.72	0.74	6			
Hadza	GOR32	High	4.78	0.59	5			
Hadza	GOR33	High	5.50	0.63	5			
Hadza	GOR34	High	6.09	0.86	6			
Hadza	GOR35	High	2.90	0.17	3			
Hadza	GOR37	High	4.59	0.33	5			
Hadza	GOR38	High	3.16	0.42	3			
Hadza	GOR8	High	6.74	0.29	7			
			Summary:	**n =**	**38**			
				Mean:	**6.92**			
				SD:	**2.60**			
Population	**Sample**	**Starch level (high/low)**	**Diploid *AMY1* copies (qPCR)**	**Standard Deviaion**	**Diploid *AMY1* copies (integer)**	**AMY1 protein mg/mL**	**Chr1tp-6D2 log2 ratio**	**Chr1tp-30C7 log2 ratio**
Japanese	GM18940	High	8.09	0.63	8		0.653	0.615
Japanese	GM18942	High	7.76	0.30	8		0.878	0.7315
Japanese	GM18943	High	9.62	0.18	10		0.5735	0.55
Japanese	GM18944	High	7.92	0.24	8		0.911	0.7585
Japanese	GM18945	High	3.73	0.48	4		-0.364	-0.2645
Japanese	GM18947	High	10.26	0.95	10		0.7975	0.73
Japanese	GM18948	High	8.31	0.48	8		0.628	0.532
Japanese	GM18949	High	7.22	0.54	7		0.3705	0.31
Japanese	GM18951	High	5.34	0.27	5		0.061	0.089
Japanese	GM18952	High	7.62	0.26	8		0.387	0.3485
Japanese	GM18953	High	6.44	0.30	6		0.4495	0.329
Japanese	GM18956	High	5.88	0.31	6		0.094	0.063
Japanese	GM18959	High	4.56	0.45	5		0.0885	0.0625
Japanese	GM18960	High	5.43	0.35	5		0.407	0.3175
Japanese	GM18961	High	4.18	0.19	4		0.1135	0.0655
Japanese	GM18964	High	7.47	0.61	7		0.412	0.3475
Japanese	GM18965	High	8.56	0.38	9		0.8405	0.7105
Japanese	GM18966	High	5.23	0.88	5		0.111	0.0925

Supplementary Table 1. qPCR, protein quantification, and aCGH data for high- and low-starch population samples.

Japanese	GM18967	High	4.50	0.31	4		0.1115	0.074
Japanese	GM18968	High	5.86	0.29	6		0.383	0.3215
Japanese	GM18969	High	3.07	0.14	3		-0.31	-0.2915
Japanese	GM18970	High	4.48	0.38	4		0.0835	0.0905
Japanese	GM18971	High	7.16	0.22	7		0.325	0.307
Japanese	GM18972	High	13.73	0.93	14		1.3685	1.16
Japanese	GM18973	High	3.72	0.52	4		0.0255	0.0385
Japanese	GM18974	High	5.40	0.70	5		0.3615	0.3245
Japanese	GM18975	High	8.00	0.30	8		0.853	0.7795
Japanese	GM18976	High	5.64	0.62	6		0.3745	0.337
Japanese	GM18978	High	5.45	0.48	5		0.391	0.358
Japanese	GM18980	High	7.21	0.47	7		0.613	0.5175
Japanese	GM18981	High	4.30	0.57	4		0.042	0.0255
Japanese	GM18987	High					0.6455	0.5455
Japanese	GM18990	High	5.94	0.33	6		0.348	0.301
Japanese	GM18991	High	5.85	0.29	6		0.3915	0.2915
Japanese	GM18992	High	6.44	1.30	6		0.4415	0.317
Japanese	GM18994	High	6.70	0.43	7		0.6645	0.543
Japanese	GM18995	High	5.84	0.52	6		0.6285	0.561
Japanese	GM18996	High	5.38	0.87	5			
Japanese	GM18997	High	6.70	1.34	7		0.6555	0.5235
Japanese	GM18998	High	7.34	0.66	7		0.636	0.551
Japanese	GM18999	High	5.08	0.56	5		0.0155	0.0075
Japanese	GM19000	High	5.51	0.52	6		0.3845	0.302
Japanese	GM19003	High	6.97	0.83	7		0.4555	0.3315
Japanese	GM19005	High	9.23	1.05	9		0.8265	0.651
Japanese	GM19007	High	6.14	0.40	6		0.635	0.5035
Japanese	GM19012	High	8.00	0.38	8		0.707	0.561
			Summary:	**n =**	**45**			
				Mean:	**6.47**			
				SD:	**2.03**			
Population	**Sample**	**Starch level (high/low)**	**Diploid *AMY1* copies (qPCR)**	**Standard Deviaion**	**Diploid *AMY1* copies (integer)**	**AMY1 protein mg/mL**	**Chr1tp-6D2 log2 ratio**	**Chr1tp-30C7 log2 ratio**
Biaka	GM10469	Low	7.63	0.81	8			
Biaka	GM10470	Low	4.04	0.13	4			
Biaka	GM10472	Low	6.11	0.17	6			
Biaka	GM10473	Low	4.11	0.26	4			
Biaka	HGDP00448	Low	12.69	1.07	13			
Biaka	HGDP00451	Low	5.14	0.38	5			

Supplementary Table 1. qPCR, protein quantification, and aCGH data for high- and low-starch population samples.

Biaka	HGDP00452	Low	3.43	0.21	3			
Biaka	HGDP00453	Low	1.81	0.12	2			
Biaka	HGDP00454	Low	5.58	0.17	6			
Biaka	HGDP00455	Low	6.07	0.38	6			
Biaka	HGDP00457	Low	4.72	0.24	5			
Biaka	HGDP00458	Low	3.91	0.22	4			
Biaka	HGDP00459	Low	3.89	0.51	4			
Biaka	HGDP00460	Low	8.03	0.48	8			
Biaka	HGDP00461	Low	10.03	0.86	10			
Biaka	HGDP00464	Low	4.28	0.42	4			
Biaka	HGDP00465	Low	4.63	0.19	5			
Biaka	HGDP00466	Low	2.69	0.14	3			
Biaka	HGDP00469	Low	5.25	0.33	5			
Biaka	HGDP00470	Low	4.36	0.35	4			
Biaka	HGDP00473	Low	3.26	0.16	3			
Biaka	HGDP00475	Low	8.92	1.58	9			
Biaka	HGDP00477	Low	5.53	0.43	6			
Biaka	HGDP00479	Low	6.39	0.39	6			
Biaka	HGDP00981	Low	4.88	0.22	5			
Biaka	HGDP00985	Low	8.07	0.47	8			
Biaka	HGDP00986	Low	6.19	0.72	6			
Biaka	HGDP01084	Low	6.88	0.68	7			
Biaka	HGDP01085	Low	4.10	0.33	4			
Biaka	HGDP01086	Low	6.14	0.32	5			
Biaka	HGDP01088	Low	7.15	0.52	7			
Biaka	HGDP01089	Low	2.94	0.13	3			
Biaka	HGDP01090	Low	5.94	0.36	6			
Biaka	HGDP01091	Low	4.25	0.28	4			
Biaka	HGDP01093	Low	3.55	0.35	4			
Biaka	HGDP01094	Low	5.09	0.25	5			
			Summary:	**n =**	**36**			
				Mean:	**5.47**			
				SD:	**2.22**			
Population	**Sample**	**Starch level (high/low)**	**Diploid *AMY1* copies (qPCR)**	**Standard Deviaion**	**Diploid *AMY1* copies (integer)**	**AMY1 protein mg/mL**	**Chr1tp-6D2 log2 ratio**	**Chr1tp-30C7 log2 ratio**
Mbuti	HGDP00449	Low	4.69	0.51	5			
Mbuti	HGDP00450	Low	4.82	0.20	5			
Mbuti	HGDP00456	Low	6.77	0.34	7			
Mbuti	HGDP00462	Low	8.66	0.59	9			

Supplementary Table 1. qPCR, protein quantification, and aCGH data for high- and low-starch population samples.

Mbuti	HGDP00463	Low	5.35	0.34	5			
Mbuti	HGDP00467	Low	4.48	0.18	4			
Mbuti	HGDP00468	Low	6.79	0.46	7			
Mbuti	HGDP00471	Low	6.90	0.60	7			
Mbuti	HGDP00474	Low	3.79	0.44	4			
Mbuti	HGDP00476	Low	5.37	0.23	5			
Mbuti	HGDP00478	Low	3.58	0.23	4			
Mbuti	HGDP00982	Low	3.78	0.48	4			
Mbuti	HGDP00983	Low	5.10	0.58	5			
Mbuti	HGDP00984	Low	4.50	0.12	5			
Mbuti	HGDP01081	Low	4.10	0.38	4			
			Summary:	**n =**	**15**			
				Mean:	**5.33**			
				SD:	**1.50**			
Population	**Sample**	**Starch level (high/low)**	**Diploid *AMY1* copies (qPCR)**	**Standard Deviaion**	**Diploid *AMY1* copies (integer)**	**AMY1 protein mg/mL**	**Chr1tp-6D2 log2 ratio**	**Chr1tp-30C7 log2 ratio**
Yakut	HGDP00945	Low	3.97	0.32	4		-0.03	0.0485
Yakut	HGDP00946	Low	5.26	0.43	5		0.5995	0.4565
Yakut	HGDP00947	Low	3.43	0.15	3		-0.0345	-0.052
Yakut	HGDP00948	Low	3.55	0.18	4		-0.216	-0.2445
Yakut	HGDP00949	Low	9.33	0.59	9		0.7935	0.668
Yakut	HGDP00950	Low	2.80	0.21	3		-0.427	-0.309
Yakut	HGDP00951	Low	8.17	1.20	8		0.7475	0.6135
Yakut	HGDP00952	Low	3.75	0.42	4		0.046	0.0455
Yakut	HGDP00953	Low	3.56	0.46	4		-0.465	-0.381
Yakut	HGDP00954	Low	4.55	0.86	5		-0.0305	-0.055
Yakut	HGDP00955	Low	4.69	0.81	5		0.3195	0.2425
Yakut	HGDP00956	Low	3.95	0.37	4		-0.0865	-0.102
Yakut	HGDP00957	Low	4.84	0.58	5		0.0215	0.0135
Yakut	HGDP00958	Low	9.02	0.68	9		0.5405	0.402
Yakut	HGDP00959	Low	10.30	1.24	10		1.167	0.991
Yakut	HGDP00960	Low	3.88	0.21	4		-0.493	-0.3515
Yakut	HGDP00961	Low	4.29	0.73	4		-0.429	-0.358
Yakut	HGDP00962	Low	3.91	0.69	4		-0.0085	-0.0055
Yakut	HGDP00963	Low	4.71	0.35	5		0.058	0.092
Yakut	HGDP00964	Low	3.59	0.25	4		-0.508	-0.392
Yakut	HGDP00965	Low	3.66	0.64	4		-0.458	-0.388
Yakut	HGDP00966	Low	8.29	1.17	8		0.522	0.413
Yakut	HGDP00967	Low	4.70	0.45	5		0.0005	0.0445

Supplementary Table 1. qPCR, protein quantification, and aCGH data for high- and low-starch population samples.

Yakut	HGDP00968	Low	5.69	0.48	6		0.2475	0.196
Yakut	HGDP00969	Low	5.12	0.27	5		-0.09	-0.0885
			Summary:	**n =**	**25**			
				Mean:	**5.24**			
				SD:	**1.96**			
Population	**Sample**	**Starch level (high/low)**	**Diploid *AMY1* copies (qPCR)**	**Standard Deviaion**	**Diploid *AMY1* copies (integer)**	**AMY1 protein mg/mL**	**Chr1tp-6D2 log2 ratio**	**Chr1tp-30C7 log2 ratio**
Datog	GOR20	Low	2.40	0.30	2			
Datog	GOR49	Low	10.21	0.69	10			
Datog	GOR51	Low	6.20	0.50	6			
Datog	GOR52	Low	4.69	0.54	5			
Datog	GOR53	Low	4.59	0.24	5			
Datog	GOR54	Low	2.72	0.41	3			
Datog	GOR55	Low	4.81	0.33	5			
Datog	GOR56	Low	6.08	1.01	6			
Datog	GOR57	Low	3.49	0.61	3			
Datog	GOR59	Low	4.71	0.53	5			
Datog	GOR6	Low	8.00	0.80	8			
Datog	GOR62	Low	4.25	0.97	4			
Datog	GOR63	Low	4.68	0.31	5			
Datog	GOR64	Low	7.64	0.92	8			
Datog	GOR67	Low	9.53	0.43	10			
Datog	GOR71	Low	7.17	0.60	7			
Datog	GOR72	Low	6.19	0.55	6			
			Summary:	**n =**	**17**			
				Mean:	**5.76**			
				SD:	**2.28**			

Supplementary Table 2. Subsistence classification, dietary description, and traditional starch intake of sampled populations.

Region/Population	Subsistence	General Dietary Description	Traditional Starch Intake
Central and East Africa			
Datog[1,2]	Pastoralism	Meat, blood and dairy products from cattle, sheep, and goats; (cultivated foods acquired through trade, such as maize[a])	Low, low-moderate[b]
Hadza[3,4]	Hunter-gathering	Meat; tubers; baobab seeds; some high-carbohydrate products, such as fruit and honey; (cultivated foods acquired through trade, such as maize and millet)	High; tubers presently represent ca. 40% of daily Kcal for women; the starch content of tubers ranges from 20.0 – 26.0 g/100 g
Mbuti[5]	Hunter-gathering	Meat; seeds; honey; (cultivated foods acquired through trade, such as cassava, plantains, yams, and sweet potatoes[a])	Low
Biaka[6]	Hunter-gathering	Meat and protein from caterpillars; fruit; seeds; honey; (cultivated foods acquired through trade, such as cassava[a])	Low
Asia			
Japanese[7]	Industrial agriculture	High-carbohydrate products, such as rice; meat, fish; fats and oils; fruit, vegetables	High; ca. 140 g day^{-1}
Yakut[8,9]	Pastoralism/Fishing	Meat and dairy products from fishing, hunting, and nomadic herding[a]	Low
Europe			
European-American[7,10,11]	Industrial agriculture	High-carbohydrate products, such as cereal grains; meat, poultry; dairy products; fats and oils; fruit, vegetables	High; consumption in Europe = ca. 99-220 g day^{-1}; consumption in the U.S.A. = ca. 73-112 g day^{-1}

[a]The starch intake of traditional low-starch populations has increased substantially in the past century due to the ingestion of cultivated foods obtained from trade with small-scale agricultural settlements.

[b]The Datog now store some of the maize and grain that they acquire in trade, and therefore appear to have the highest levels of starch consumption of the low-starch populations in our study. It is unclear how long this practice has continued, though it is likely that traditional levels of starch consumption were considerably lower.

Supplementary Table 2. Subsistence classification, dietary description, and traditional starch intake of sampled populations.

References

1. Sieff, D.E. Herding strategies of the Datoga pastoralists of Tanzania: Is household labor a limiting factor? *Hum. Ecol.* **25**, 519-544 (1997).
2. Sellen, D.W. Nutritional consequences of wealth differentials in East African pastoralists: The case of the Datoga of Northern Tanzania. *Hum. Ecol*. **31**, 529-570 (2003).
3. Marlowe, F.W. A critical period for provisioning by Hadza men: Implications for pair bonding. *Evol. Hum. Behav*. **24**, 217-229 (2003).
4. Schoeninger, M.J., Bunn, H.T., Murray, S.S. & Marlett, J.A. Composition of tubers used by Hadza foragers of Tanzania. *J. Food Comp. Anal.* **14**, 15-25 (2001).
5. Hart, T.B. & Hart, J.A. The ecological basis of hunter-gatherer subsistence in African rain forests: The Mbuti of eastern Zaire. *Hum. Ecol.* **14**, 29-55 (1986).
6. Bahuchet, S. Food Supply Uncertainty among the Aka Pygmies. in *Coping with Uncertainty in Food Supply* (eds. de Garine, I. & Harrison, G.A.) 118-149 (Clarendon Press, Oxford, 1988).
7. Cassidy, A., Bingham, S.A. & Cummings, J.H. Starch intake and colorectal cancer risk: An international comparison. *Br. J. Cancer* **69**, 937-942 (1994).
8. Tokarev, S.A. & Gurvich, I.S. The Yakuts. in *Peoples of Siberia* (eds. Levin, M.G. & Potapov, L.P.) 243-304 (University of Chicago Press, Chicago, 1956).
9. Sorensen, M.V., Snodgrass, J.J., Leonard, W.R., Tarskaia, A., Ivanov, K.I., Krivoshapkin, V.G. & Spitsyn, V.A. Health consequences of postsocialist transition: Dietary and lifestyle determinants of plasma lipids in Yakutia. *Amer. J. Hum. Biol*. **17**, 576-592 (2002).
10. Page, L. & Friend, B. The changing United States diet. *BioScience* **28**, 192-197 (1978).
11. Block, G., Dresser, C.M., Hartman, A.M. & Carroll, M.D. Nutrient sources in the American diet: Quantitative data from the NHANES II survey: II. Macronutrients and fats. *Am. J. Epidemiol*. **122**, 27-40 (1985).

Supplementary Table 3. PCR primers and conditions.

Primer Description	Name[1]	5' Sequence 3'	Primer used for:												
			P. troglodytes			*P. paniscus*			*G. gorilla*			*C. angolensis*	*L. lagotricha*	qPCR	Fiber FISH
			AMY1	*AMY2A*	*AMY2B*	*AMY1*	*AMY2A*	*AMY2B*	*AMY1*	*AMY2A*	*AMY2B*				
qPCR primers															
AMY1-specific	AMY1 qF	GGAAGAATATAGATGCCAACCC												x	
	AMY1 qR	TCAGCTGAACTGGATCATTTGT													
TP53 (reference/calibrator)	TP53 F	CCCTTCCCAGAAAACCTACC												x	
	TP53 R	CAGGCATTGAAGTCTCATGG													
Long Range Amplification Primers[2]															
long range amplification - AMY1	AMY1 F	CCTAGCCTGTTTTTGCAATTTTCTCT	x			x			x						x
	AMY1 R	TAAAATTTGGCTTTCCTTCCCTTGTA													
long range amplification - AMY2A	AMY2A F	GCTGCATGTCTTTGGTTGTACTTCCT		x			x			x					
	AMY2A R	CGTTGCCTATGGAAAAAGGAATGTTT													
long range amplification - AMY2B	AMY2B F	ATCACACCTTCTACAAGGGACTGCAC			x			x			x				
	AMY2B R	TGGGAAGAAAACCCAGACTACAAGGT													
long range amplification - colobus	AMY *Can* F	ATGAAGATCAAGATGATCGCACCC										x			
	AMY *Can* R	TAAAAAGCATACAACCGGGCTAACTT													
long range amplification - upstream	AMYup F	TAAGCCTTGGGAAAGAAGTTGTCC													x
	AMYup R	GCCCTTCCCAGCCTCTAGATAAAT													
Nested PCR Re-Amplification Primers[3]															
re-amplification - gene region 1	-674 F	CTTGGTCACTTCATGGCTAAA	x	x	x	x	x	x	x	x	x	x			x
	1961 R	GGGGATGCTCACATTCTACTA													
re-amplification - gene region 1	133 F	TGATATTGCTCTTGAATGTGAGC											x		
	1845 R	GTGGGCAAATGTATATTGATTAAA													
re-amplification - gene region 2	1789 F	GGACAGAGGTTAACAAGTTTGAC	x	x		x			x		x		x		x
	4022 R	CTTCTCCCCAGTTCCTATTTTT													
re-amplification - gene region 2	1580 F	AAGTAGGGACTTTCCAGCAGTC			x		x	x		x		x			
	4022 R	CTTCTCCCCAGTTCCTATTTTT													
re-amplification - gene region 3	3691 F	ATCCTTCTGGAGTGCCTCTAA	x	x	x	x	x	x	x	x	x	x	x		x
	6083 R	TTATCCTCTCCTTTTCATTCCA													
re-amplification - gene region 4	5871 F	GACCTGAGGAGTTCAAGACCTA	x	x		x			x						x
	8214 R	AACACACTTAGCTATTGCTTTGC													
re-amplification - gene region 4	5647 F	CAAATGAGGAAACTGAGACACAGA			x		x	x		x	x	x			
	8223 R	AAGAAAAGAAACACACTTGGCTA													
re-amplification - gene region 4	5939 F	AAATAAAAAGTAGCTGCCTGTGG											x		
	7977 R	GACCAGTTTGCAAAGTTAAAGAA													
re-amplification - upstream region 1	Up F1	CCTTTTTAAGGGCTCACAACAC													x
	Up R1	TTTTTCTCCAGGTGACATTGTG													
re-amplification - upstream region 2	Up F2	CACAATGTCACCTGGAGAAAAA													x
	Up R2	GTTAAGAGCCACCATTTCCTTG													
re-amplification - upstream region 3	Up F3	TCGATAGTCTTTGGTCCTTGGT													x
	Up R3	TCTTTAGTGCTGTGTGCCCTTA													
Sequencing Primers															
Sequencing - gene region 1	114 R	CGCTCACATTCAAGAGCAATA										x			
Sequencing - gene region 1	376 R	CTTGAAAATATTAACAGCACA	x	x	x	x	x	x	x	x	x				
Sequencing - gene region 1	527 F	TTGTAGGTCTCTCCACCAAA											x		
Sequencing - gene region 1	556 R	GGTATCTTTCCCACCAAGGTCT				x									
Sequencing - gene region 1	562 R	TGGTTGGTATCTTTCCCACCA											x		
Sequencing - gene region 1	751 F	TTTGGAGCGGAAAGTTTTCCA										x	x		
Sequencing - gene region 1	955 R	AGCAAATATTGTAACTGGAATTA	x	x	x	x	x	x	x	x	x	x	x		
Sequencing - gene region 1	1483 F	GGTTCGTATTTATGTGGATGC											x		

Supplementary Table 3. PCR primers and conditions.

Primer Description	Name[1]	5' Sequence 3'	Primer used for:											
			P. troglodytes			*P. paniscus*			*G. gorilla*			*C. angolensis*	*L. lagotricha*	qPCR	Fiber FISH
			AMY1	*AMY2A*	*AMY2B*	*AMY1*	*AMY2A*	*AMY2B*	*AMY1*	*AMY2A*	*AMY2B*				
Sequencing - gene region 1	1612 R	TCTCCACTTCCAGTTTTACATT	x	x	x	x	x	x	x	x	x	x	x		
Sequencing - gene region 2	2183 F	TGTGCGTTCCAAGATTGCCGA											x		
Sequencing - gene region 2	2210 F	ATCTCATTGACATTGGTGTTG									x	x	x		
Sequencing - gene region 2	2318 R	TCCTGGTAAATGAAAGGTTTAC									x	x	x		
Sequencing - gene region 2	2509 F	ATATCTAATTCTTTATCACA	x	x	x	x	x	x	x	x	x	x			
Sequencing - gene region 2	3208 F	ATGGAGAGAAGATGTCTTAC	x	x	x	x	x	x	x	x	x	x	x		
Sequencing - gene region 2	3357 R	ACAAAAGATGATTGATTAGAG							x	x					
Sequencing - gene region 3	3936 F	AGAAAGAATTTAATCTTCAG	x	x	x	x	x	x	x	x	x				
Sequencing - gene region 2	3979 R	TCTTGAGACAGACACTCTAAA											x		
Sequencing - gene region 3	4248 R	TCCAACTGCCATTTTGTACAG				x									
Sequencing - gene region 3	4284 F	TTGGATTTATGCTTGCTCATC											x		
Sequencing - gene region 3	4323 F	AATGTCAAGCTACCGTTGGC	x	x	x	x	x	x	x	x	x	x			
Sequencing - gene region 3	4419 R	AATAAATAAATACTTAATAGTTG				x									
Sequencing - gene region 3	4465 R	AAAATCAGAGCCCTGTATCAGG										x			
Sequencing - gene region 3	4488 R	CTTTAATATAACCTGCATTAC	x	x	x		x	x	x	x	x		x		
Sequencing - gene region 3	4795 R	TAACTCAGCTGAACTGGATCA				x							x		
Sequencing - gene region 3	5005 F	TATTTGGAAAGCTAGTAGAAG	x	x	x	x	x	x	x	x	x	x	x		
Sequencing - gene region 3	5089 R	AAAAAGCTAGCTAGATATCTG			x	x		x			x				
Sequencing - gene region 3	5120 R	GTAGTGGACTTCTCAAACTGAA											x		
Sequencing - gene region 3	5228 F	GACAACTTCTTATAATTTCTAC				x									
Sequencing - gene region 3	5380 R	AGTAACCCTAGGAAGTTTATC				x									
Sequencing - gene region 3	5584 F	AAGGCATTTTCACATATATTAC							x						
Sequencing - gene region 3	5868 R	TGCTAGCTTGCCCAGATAGG				x			x						
Sequencing - gene region 3	5940 R	GGAACTGTTGGTGTGACACAA											x		
Sequencing - gene region 4	6453 F	TGGCGCCAAATAAGGTGAGAAT	x									x			
Sequencing - gene region 4	6619 R	CAAAAGCCACTTGGTTGCTC	x	x	x	x	x	x	x	x	x	x	x		
Sequencing - gene region 4	7158 F	TCTCTGTCTCCTTGTGACAAA	x	x	x	x	x	x	x	x	x	x	x		
Sequencing - gene region 4	7246 R	CAGATGAAAGGCCAATAATGAG	x	x	x	x	x	x	x	x	x	x	x		
Sequencing - gene region 4	7847 F	GTTACTTTTGGTCCTAGAAAG	x			x			x						
Sequencing - gene region 4	7918 R	TAAACAGCATAACTTTATG	x			x			x						
Sequencing - gene region 4	7921 F	CTGCCTAGAGTTCTGCAGCATC											x		

[1]Numbers correspond to primer position with respect to the "A" of start codon ATG in*AMY1* from the human hg18 reference sequence

[2]For long range amplification we used TripleMaster Taq (Eppendorf) with the Tuning Buffer and the following PCR conditions: 93° for 3 min, 40x of (93° for 15 sec, 64° for 30 sec, 68° for 14 min)

[3]For nested PCR amplification we used HotMaster Taq (Eppendorf) with these conditions: 93° for 3 min, 35x of (93° for 30 sec, 59° for 30 sec, 70° for 4 min)
Nested PCRs were performed in 50 ul reactions using 5 ul of a 1:100 dilution (PCR product: H2O) of the unpurified long-range PCR product