Natural selection at the brush–border: recent and ancient adaptations to carbohydrate diets in humans and other mammals

Chiara Pontremoli[1]*, Alessandra Mozzi[1], Diego Forni[1], Rachele Caglian[i1], Uberto Pozzoli[1], Giorgia Menozzi[1], Nereo Bresolin[1,2], Mario Clerici[3,4], Manuela Sironi[1]

1 Bioinformatics, Scientific Institute IRCCS E.MEDEA, 23842 Bosisio Parini, Italy. 2 Dino Ferrari Centre, Department of Physiopathology and Transplantation, University of Milan, Fondazione Ca' Granda IRCCS Ospedale Maggiore Policlinico, 20122 Milan, Italy. 3 Department of Physiopathology and Transplantation, University of Milan, 20090 Milan, Italy.4 Don C. Gnocchi Foundation ONLUS, IRCCS, 20148 Milan, Italy.
*e-mail: chiara.pontremoli@bp.lnf.it

Background

The ever-increasing availability of genome sequences from different organisms and from multiple individuals of the same species, together with resequencing data of ancient DNA samples, are opening the unprecedented opportunity to perform comprehensive evolutionary analyses of biological pathways. We exploited this wealth of information to investigate the evolutionary history of genes that encode intestinal brush-border proteins involved in carbohydrate metabolism. This decision was based on the well-accepted concept that the availability of food resources is a driver of pivotal importance in evolution in mammals and by the fact that one of the most important turning-points of human history, the introduction of agriculture, resulted in a dietary shift in terms of carbohydrate intake. In this respect, the availability of human DNA samples of pre-agricultural populations allows testing of specific hypotheses as to when adaptive alleles at genes involved in sugar metabolism arose.
The text-book examples of positive selection at the *LCT* locus in pastoralists, as well as the increase in amylase gene copy number in human populations that consume starch-rich diets were the starting point for our work, which focuses on 9 genes encoding apical brush-border proteins involved in carbohydrate digestion and absorption: *MGAM, SI, LCT, TREH, SLC2A2, SLC5A1, SLC2A5, TAS1R2, TAS1R3*.

Results

Most brush-border carbohydrate digestion/absorption genes evolve adaptively in mammals.

We performed an in-depth analysis of the evolutionary history of these genes in 40 mammalian species. We found evidences of adaptive evolution in these genes, with five of them targeted by positive selection. For the selected genes, episodic positive selection was also detected for several mammalian lineages; results indicated the presence of pervasive selection, which might reflect specific adaptations to flight (bats) and to specialized diets (insects, crustaceans).

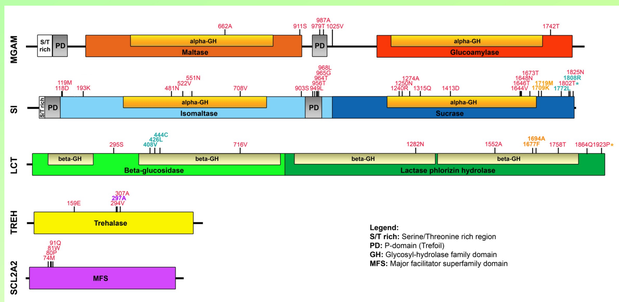

Figure 1. Domain representation of positively selected genes. Sites selected in whole phylogeny are in red; positively selected sites in the human, chimpanzee, and gorilla lineage are in cyan, orange, and violet, respectively. Asterisks denote lineage-specific sites that are also selected in whole phylogeny. Positions refer to the human sequence.

Table 1. Single branch analysis (significant branches only).

Gene	Foreground branch (MA versus MA1)	−2ΔlnL	*p* value (corrected *p* value)[a]	MEME–BEB sites[b]
MGAM				
Region1	Microbat	15.66	7.58×10^{-5} (0.0001)	235R
Region2	Squirrel Monkey	21.1	4.37×10^{-6} (1.46×10^{-5})	1494E
	Macaque	39.14	3.94×10^{-10} (1.97×10^{-9})	1742T
	Olive Baboon	7.30	0.007 (1.38×10^{-2})	-
	Golden Hamster	58.00	2.62×10^{-14} (2.62×10^{-13})	1034P, 1688I
SI				
Region1	Elephant	40.63	1.84×10^{-10} (6.9×10^{-10})	606T
	Squirrel Monkey	20.84	4.98×10^{-6} (9.33×10^{-6})	810 R
	Macaque	14.17	0.0002 (2.51×10^{-4})	-
	Tree Shrew	52.54	4.21×10^{-13} (3.161×10^{-12})	38V, 1105R, 1119H
	Node *Mouse-Rat*	30.61	3.15×10^{-8} (6.76×10^{-8})	35T, 180F, 482C, 1477K
	Mouse	30.61	3.16×10^{-8} (6.76×10^{-8})	483I, 676G, 990Y, 1602A
	Node *Pika-Rabbit*	13.41	0.0002 (3.4×10^{-4})	92G
	Pig	41.52	1.16×10^{-10} (5.82×10^{-10})	297M, 298N, 299S
	Horse	18.32	1.86×10^{-5} (3.10×10^{-5})	180F, 1201T
	Microbat	32.75	1.05×10^{-8} (3.15×10^{-8})	359R, 1408N
	Megabat	52.54	4.21×10^{-13} (3.16×10^{-12})	560S, 1006A, 1201T, 1405T, 1449S
Region2	Squirrel	37.35	9.85×10^{-10} (8.86×10^{-9})	1735Y
	Horse	7.66	0.006 (1.69×10^{-2})	1747L
	Microbat	11.17	0.0008 (3.74×10^{-3})	1819E
TREH				
Region1	Microbat	33.67	6.51×10^{-9} (1.30×10^{-8})	26P, 28C, 287P
	Platypus	22.72	1.88×10^{-6} (1.88×10^{-6})	-
LCT				
	Elephant	15.1	0.0001 (3.06×10^{-4})	244D, 947A, 1711I
	Chimpanzee	18.27	1.92×10^{-5} (8.63×10^{-5})	1411H, 1442L
	Tree Shrew	5.26	0.02(3.28×10^{-2})	-
	Dolphin	6.28	0.01 (2.2×10^{-2})	-
	Cat	10.62	0.001 (2.51×10^{-3})	-
SLC2A2				
	Ferret	24.71	6.67×10^{-7}	-

NOTE: To identify specific branches with a proportion of sites evolving with ω>1, we used BS-REL. Branches identified using this approach were cross-validated with the branch-site likelihood ratio tests from codeml (the so-called modified model A and model MA1, "test 2"). MA and MA1 are branch-site models that assume four classes of sites: the MA model allows a proportion of codons to have ω≥ 1 on the foreground branches, whereas the MA1 model does not. 2ΔlnL is twice the difference of the natural logs of the maximum likelihood of the models being compared.
[a] *p* values were FDR-corrected for the number of tested branches (see also Figures S1 and S2)
[b] positively selected sites identified by both BEB and MEME

Several positively selected sites impinge on functional protein regions

Analysis of 3D structures and structural superimposition of homologous domains were used to infer the functional significance of positively selected sites. We found the corresponding residues of MGAM and SI to be targeted by selection: an interesting possibility is that some selected sites in SI and MGAM evolved to hone the folding, cellular trafficking, and membrane turnover of these enzymes. Moreover, in SI, missense mutations responsible for congenital sucrase-isomaltase deficiency (CSID) or identified in chronic lymphocytic leukemia patients (CLL) have been shown to alter the cellular processes mentioned above. We noted that mutations R91T (CLL, endoplasmic reticulum accumulation) and Q117R (CSID, mis-sorting to the basolateral membrane) immediately flank positively selected sites.

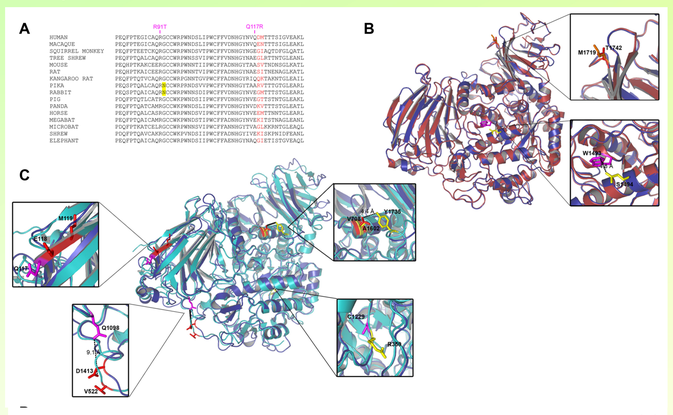

Figure 2. Parallel evolution at MGAM and SI, and lineage–specific selection. (A) Multiple alignment of SI aminoacids 78-130 for a few of representative mammalian species. The location of mutations R91T and Q117R is shown. (B and C) Superimposition of the structure of the sucrase domain with glucoamylase (MGAM, red) (B) and with isomaltase (SI, pale blue) (C). Enlargements highlight positively selected sites or residues subjected to pathological mutation located in the corresponding regions of the two different domains. Color codes are as follows: red, positively selected sites in the whole phylogeny; yellow, lineage-specific sites; orange and cyan, positively selected sites in the chimpanzee and human lineages, respectively; blue, aminoacids involved in ligand binding; violet: positively selected residues in gorilla. Human missense mutations affecting the protein sorting are reported in magenta.

Evolution of brush–border proteins in great apes.

We applied a population genetics-phylogenetics approach (gammaMap) to study the evolution of brush-border genes in the human, chimpanzee, and gorilla lineages: it jointly uses intra-specific variation and inter-specific diversity to estimate the distribution of selection coefficients (γ) along coding regions.
We observed a general preponderance of codons evolving under negative selection (γ < 0) in all genes and in all species. The most striking difference was observed for *SLC5A1*, which showed a preponderance of negative γ values in chimpanzee and to a lesser extent in gorilla, but not in our species, where an appreciable fraction of codons showed γ values higher than 5 .

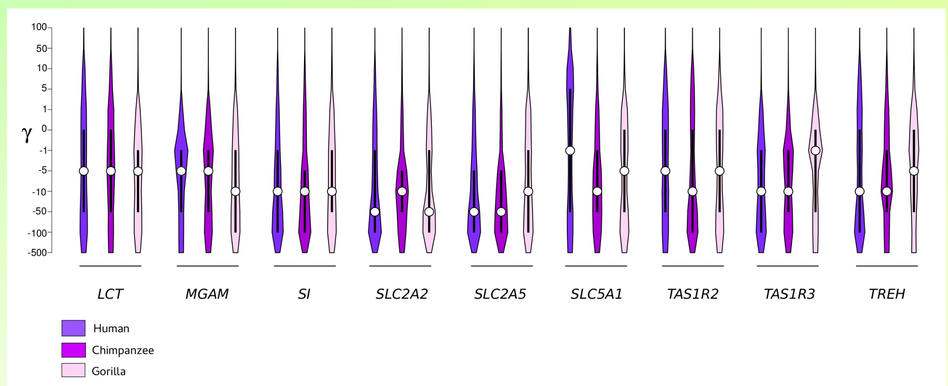

Figure 3. Violin plots. Violin plot of selection coefficients (median, white dot; interquartile range, black bar). Selection coefficients (γ) are classified as strongly beneficial (100, 50), moderately beneficial (10, 5), weakly beneficial (1), neutral (0), weakly deleterious (−1), moderately deleterious (−5, −10), strongly deleterious (−50, −100), and inviable (−500).

Pre–agricultural origin of most positively selected alleles

We investigated whether natural selection acted on these genes during the recent evolutionary history of human populations. We also exploited information from the Neandertal and Denisova high-coverage genomes, as well as from a hunter-gatherer Mesolithic European from Spain and a Paleolithic Siberian. We showed that five genes were targeted by positive selection in human populations, acting on non-coding variants within regulatory elements. Analysis of ancient DNA samples indicated that most derived alleles were already present in the Paleolithic, therefore predating the emergence of agriculture. Positively selected variants at *SLC2A5* (fructose transporter) were an exception and possibly spread following the domestication of specific fruit crops.

Table 2. Candidate targets of positive selection in human populations.

Gene	SNP ID	Derived allele	DAF[a]			DIND rank (population)	F_{ST} rank (comparison)	Ancient samples				Notes
			YRI	CEU	CHBJPT			Nea[b]	Den[c]	Paleo[d]	Meso[e]	
SI	rs41273563	C	0.32	0.89	0.87	0.97 (YRI)	0.96 (YRI+CHBJPT)	T/T	T/T	n.a.	n.a.	
	rs11919067	C	0.98	1	1	>0.999 (YRI)	-	T/T	T/T	C	n.a.	modern-human-specific site
	rs112446029	A	0.98	1	1	0.99 (YRI)	-	G/G	G/G	A	n.a.	modern-human-specific site
	rs6788812	G	0.32	0.89	0.87	0.98 (YRI), 0.98 (CHBJPT)	0.95 (YRI+CEU)	A/A	A/A	n.a.	G	
	rs9917722	G	0.85	1	1	-	-	G/G	C/C	G	n.a.	modern-human-specific site; identified by gammaMap
TREH	rs527619	A	0.52	0.42	0.27	0.98 (YRI), 0.97(CEU); 0.99 (CHBJPT)	-	G/G	G/G	n.a.	G	
SLC5A1	rs117628874	T	1	0.94	1	0.95 (CEU)	-	C/T	T/T	n.a.	T	modern-human-specific site
	rs74399071	G	1	0.94	1	0.96 (CEU)	-	T/G	G/G	n.a.	G	modern-human-specific site
	rs79022443	T	1	0.94	1	0.97 (CEU)	-	C/C	T/T	T	T	
	rs78578916	G	1	0.94	1	0.98 (CEU)	-	A/G	G/G	G	G	modern-human-specific site
	rs2899174	T	1	0.94	1	0.95 (CEU)	-	C/T	T/T	T	T	modern-human-specific site
	rs17683430	G	1	0.94	1	-	-	A/G	A/A	G	A/G	modern-human-specific site; identified by gammaMap
SLC2A2	rs117720640	G	0.64	0.86	0.96	0.99 (CHBJPT)	0.95 (CHBJPT/YRI)	A/A	A/A	G	G	In LD with rs11920090 and rs10513686 (GWAS)
	rs1905504	T	0.64	0.86	0.96	>0.999 (CHBJPT)	0.95 (CHBJPT/YRI)	C/C	C/C	n.a.	T	In LD with rs11920090 and rs10513686 (GWAS)
	rs7635100	G	0.64	0.86	0.96	>0.999 (CHBJPT)	0.95 (CHBJPT/YRI)	T/T	T/T	n.a.	n.a.	In LD with rs11920090 and rs10513686 (GWAS)
	rs6780208	A	0.64	0.86	0.96	>0.999 (CHBJPT)	0.95 (CHBJPT/YRI)	T/T	T/T	A	A	In LD with rs11920090 and rs10513686 (GWAS)
	rs7649712	G	0.90	0.98	1	0.98 (YRI)	0.99 (YRI/CEU); 0.99 (YRI/CHBJPT)	G/G	G/G	G	G	
	rs75513459	G	0.90	0.98	1	0.97 (YRI)	0.99 (YRI/CEU); 0.99 (YRI/CHBJPT)	G/G	G/G	G	n.a.	
	rs79438006	T	0.90	0.98	1	0.96 (YRI)	0.99 (YRI/CEU); 0.99 (YRI/CHBJPT)	T/T	T/T	T	T	
	rs75975268	C	0.90	0.98	1	0.97 (YRI)	0.99 (YRI/CEU); 0.99 (YRI/CHBJPT)	C/C	C/C	C	C	
	rs74828611	A	0.90	0.98	1	0.98 (YRI)	0.99 (YRI/CEU); 0.99 (YRI/CHBJPT)	A/A	A/A	n.a.	A	
SLC5A2	rs875996	A	0	0.18	0.16	0.97 (CEU)	0.99 (YRI/CEU)	G/G	G/G	G	G	In LD with rs113568511 (eQTL)
	rs34605482	T	0	0.18	0.17	0.95 (CEU)	0.99 (YRI/CEU)	C/C	C/C	n.a.	C	In LD with rs113568511 (eQTL)

[a] Derived allele frequency;
[b] Altai Neandertal individual;
[c] Denisova individual;
[d] Upper Paleolithic Siberian individual;
[e] Mesolithic hunter-gatherer individual.
Note: the genotype is reported for ancient samples; we report only one allele if the coverage was not sufficient for inferring the correct genotype.

Conclusions

Our results indicated pervasive selection in mammals and human populations for genes involved in sugar absorption/digestion at brush-border that might reflect specific adaptation to specialized diets. Data herein suggest that the dietary shift in terms of carbohydrate intake, occurred by the introduction of agriculture, determined no major selective event at carbohydrate metabolism genes in humans. Whether the onset of selection occurred before the Paleolithic or these alleles segregated as neutral standing variation in these early populations remains to be evaluated, possibly through the sequencing of additional ancient samples.
Finally, some of the selective events we identified open many interesting research avenues. As an example, we detected positive selection at *SLC2A5* (fructose transporter) in Europeans. Reminiscent of lactose intolerance, some degree of fructose intolerance, associated with gastrointestinal symptoms, is quite common and fructose absorption is reduced by the presence of sorbitol. Thus, selection at *SLC2A5* might have been driven by the domestication and widespread consumption in temperate areas of fruit crops (e.g. apples and pears) that contain excess fructose plus sorbitol. It will be extremely interesting to test whether the positively selected variant we identified herein (and which is in LD with an eQTL) modulates fructose absorptive capacity.