

A Scan for Human-Specific Relaxation of Negative Selection Reveals Unexpected Polymorphism in Proteasome Genes

Mehmet Somel,^{*1} Melissa A. Wilson Sayres,¹ Gregory Jordan,² Emilia Huerta-Sanchez,¹ Matteo Fumagalli,¹ Anna Ferrer-Admetlla,¹ and Rasmus Nielsen¹

¹Department of Integrative Biology, University of California, Berkeley

²European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom

*Corresponding author: somel.mehmet@gmail.com.

Associate editor: John H. McDonald

Abstract

Environmental or genomic changes during evolution can relax negative selection pressure on specific loci, permitting high frequency polymorphisms at previously conserved sites. Here, we jointly analyze population genomic and comparative genomic data to search for functional processes showing relaxed negative selection specifically in the human lineage, whereas remaining evolutionarily conserved in other mammals. Consistent with previous studies, we find that olfactory receptor genes display such a signature of relaxation in humans. Intriguingly, proteasome genes also show a prominent signal of human-specific relaxation: multiple proteasome subunits, including four members of the catalytic core particle, contain high frequency nonsynonymous polymorphisms at sites conserved across mammals. Chimpanzee proteasome genes do not display a similar trend. Human proteasome genes also bear no evidence of recent positive or balancing selection. These results suggest human-specific relaxation of negative selection in proteasome subunits; the exact biological causes, however, remain unknown.

Key words: relaxation of constraints, human evolution, negative selection, olfactory transduction, proteasome.

Negative or purifying selection is selection acting against new deleterious mutations. A consequence of negative selection is the removal of new deleterious genetic variants from the population, resulting in evolutionary sequence conservation. The intensity of negative selection, or “evolutionary constraint,” can vary across the genome according to functional properties encoded by the respective loci (Dickerson 1971; Sunyaev et al. 2000; Siepel et al. 2005). Repetitive intergenic sequences far from coding regions likely evolve relatively freely because evolutionary changes here will not affect the fitness of the organism. By contrast, evolutionary changes in protein-coding genes, especially those involved in essential cellular functions, in development, or in the central nervous system, can readily impact organismal function and fitness; the evolution of such genes may be highly constrained. Such constraint can be detected in actual genomic data, both among and within species. For example, genes expressed across multiple tissues and/or functioning in the neural system have low nonsynonymous to synonymous mutation rates among mammals (dN/dS) (Kuma et al. 1995; Duret and Mouchiroud 2000; Mikkelsen et al. 2005; Nielsen et al. 2005). Within humans, genic regions contain fewer polymorphism than intergenic regions, and nonsynonymous single nucleotide polymorphisms (nsSNPs) are maintained at lower frequencies than synonymous SNPs (sSNPs) (Cargill et al. 1999; Sunyaev et al. 2000).

Levels of evolutionary constraint, reflected in genetic sequence conservation, are generally correlated across related species—the same genes and functions show high or low

constraint in hominids and rodents (Ohta 1993; Mikkelsen et al. 2005). Still, exceptions can arise through multiple mechanisms (Fay and Wu 2003). First, the strength of natural selection is expected to be a function of effective population size, such that lineages with relatively small population sizes may experience weaker negative selection throughout the genome. In populations experiencing reductions in population size, slightly deleterious mutations which were originally not tolerated may drift to high frequency, and even become fixed (Ohta and Kimura 1971). For example, humans and chimpanzees have higher genome-wide dN/dS levels compared with rodents, possibly due to the smaller effective population size in hominids allowing fixation of slightly deleterious alleles (Wu and Li 1985; Li and Tanimura 1987; Ohta 1993; Mikkelsen et al. 2005).

A decrease in negative selection pressure caused by reduced effective population size should affect a large number of genes across the genome. In turn, changes in a species’ environment or in its genome can relax negative selection on individual genes or classes of genes. One such change is gene duplication, which could reduce negative selection on the duplicated copies (Ohno 1972; Kondrashov et al. 2002). Changes in environment and subsistence mode can similarly cause relaxation of selection in specific genes. The best-known example in humans is olfactory receptors (Rouquier et al. 1998). This vast gene family appears under relaxed selection in New World monkeys, Old World monkeys, and hominoids, as observed in the high rates of pseudogenization and deletion of member genes; possibly due to a

transition from nocturnal to diurnal activity in these lineages (Gilad et al. 2003, 2004; Go and Niimura 2008; Kim et al. 2010; Matsui et al. 2010) (positive selection on olfactory genes in humans has also been reported [Williamson et al. 2007]). A recent study on relaxation of constraint in the human genome similarly found a high frequency of potentially damaging nonsynonymous polymorphism among olfactory receptors (Pierron et al. 2012). Interestingly, mutations causing color blindness are also common in humans, but not in chimpanzees, for reasons yet unclear (Terao et al. 2005).

Studying human-specific changes at evolutionarily conserved sites could provide insight into how the human species has diverged from its relatives. A number of studies have searched for conserved regions that have accumulated a high number of fixed substitutions on the human lineage; these are likely examples of adaptive evolution (Pollard, Salama, King, et al. 2006; Pollard, Salama, Lambert, et al. 2006; Prabhakar et al. 2006). Other studies have identified nonconserved regions under purifying selection in humans (Ward and Kellis 2012), found relaxed negative selection in recently evolved primate genes (Cai and Petrov 2010), and shown relaxed constraints among human and nonhuman primate olfactory genes using comparative genomic data (Gilad et al. 2003; Go and Niimura 2008) and using human polymorphism data (Pierron et al. 2012).

Here, we address the question of human-specific relaxation of negative selection at coding sites, combining genome resequencing data from humans, exome sequencing data from chimpanzees, and comparative genomic data.

Results and Discussion

We tested for human-specific relaxation of constraint among groups of functionally related genes using three criteria: 1) genes in a group should have low levels of mammalian protein sequence divergence but relatively high human nonsynonymous diversity, 2) that the group should have higher human nonsynonymous diversity levels than the genome average, 3) that the group should have low levels of chimpanzee nonsynonymous diversity relative to human diversity (fig. 1A).

We first catalogued over 64,000 nonsynonymous and over 54,000 synonymous SNPs ascertained in 54 unrelated individuals of diverse ancestry in the Complete Genomics genome-wide high coverage resequencing data set (Drmanac et al. 2010) (supplementary material, Supplementary Material online). The nonsynonymous minor allele frequency (nsMAF) distribution was highly skewed toward lower values relative to the synonymous minor allele frequency (sMAF) distribution (supplementary fig. S1A, Supplementary Material online), consistent with the role of negative selection preventing nonsynonymous mutations from reaching high frequency (Cargill et al. 1999). Median nsMAF values, calculated per gene, showed modest but significant correlation with dN/dS values per orthologous gene, calculated between mouse and macaque, or dog and elephant ($\rho \sim 0.20$, one-sided Spearman correlation test $P < 10^{-80}$, $n > 10,000$ genes) (supplementary fig. S1B and C, Supplementary Material online). This correlation likely

reflects similar levels of negative selection between and within species.

Using forward population genetic simulation (Hernandez 2008), we then tested whether 20 loci evolving under purely neutral regimes could be identified among 1,000 loci evolving under negative and positive selection (supplementary material, Supplementary Material online). There was modest (>40%) power to distinguish between neutral and negative selection, which increased to more than 60% using 40 loci (supplementary fig. S2, Supplementary Material online). We also found that MAF is a somewhat more robust measure for identifying neutrally evolving loci than derived allele frequency (DAF) if the alternative hypothesis is a mixture of positive and negative selection, as DAF is more strongly affected by positive selection (supplementary fig. S2, Supplementary Material online). These results suggest that median nsMAF is a useful statistic for investigating hypotheses regarding relaxation of constraints in the human genome.

Human nsMAF and mammalian dN/dS values showed correlation when summarized in 181 KEGG pathways (Kanehisa et al. 2008) (fig. 1B). Only a few functional groups stood out, with higher median nsMAF, despite low median dN/dS. These included not only various metabolic pathways, such as propanoate metabolism and glycosaminoglycan degradation, but also olfactory transduction, and the proteasome (supplementary table S1, Supplementary Material online). We then tested whether genes in a functional group have the same median ranks with respect to dN/dS, relative to their ranks with respect to nsMAF (fig. 1A and C) (Wilcoxon signed rank test). We further tested whether a gene set has the same median rank of nsMAF values compared with other genes (Wilcoxon rank sum test). Only two KEGG groups showed higher than expected polymorphism in both tests: olfactory transduction and the proteasome (Benjamini-Hochberg corrected $P < 0.05$, supplementary table S1, Supplementary Material online). Among the 373 genes in the olfactory transduction pathway, 323 contained at least one nsSNP (median nsMAF = 0.029). Among the 44 genes annotated within the proteasome category, 18 contained nsSNPs (median nsMAF = 0.044), 12 of these with modest to high frequency (>5%) alleles among both African and non-African individuals (fig. 1D). A randomization test across genes confirmed that nsMAF values in both groups are significantly skewed compared with other genes with nsSNPs ($P < 0.0001$; note that the olfactory receptors are found in clusters, which is ignored in this test; proteasome genes, however, tend to reside on different chromosomes and are not clustered).

We then asked whether these putative shifts in negative selection pressure can also be observed in our closest living relatives, the chimpanzees, or whether they might be human specific. For this, we used a published exome resequencing data set containing close to 25,000 nonsynonymous and 32,000 synonymous SNPs ascertained in 12 central chimpanzees, *Pan troglodytes troglodytes* (Hvilsom et al. 2012) (supplementary material, Supplementary Material online). Examining KEGG pathways for a difference in human nsMAF ranks relative to chimpanzee nsMAF ranks, we found olfactory

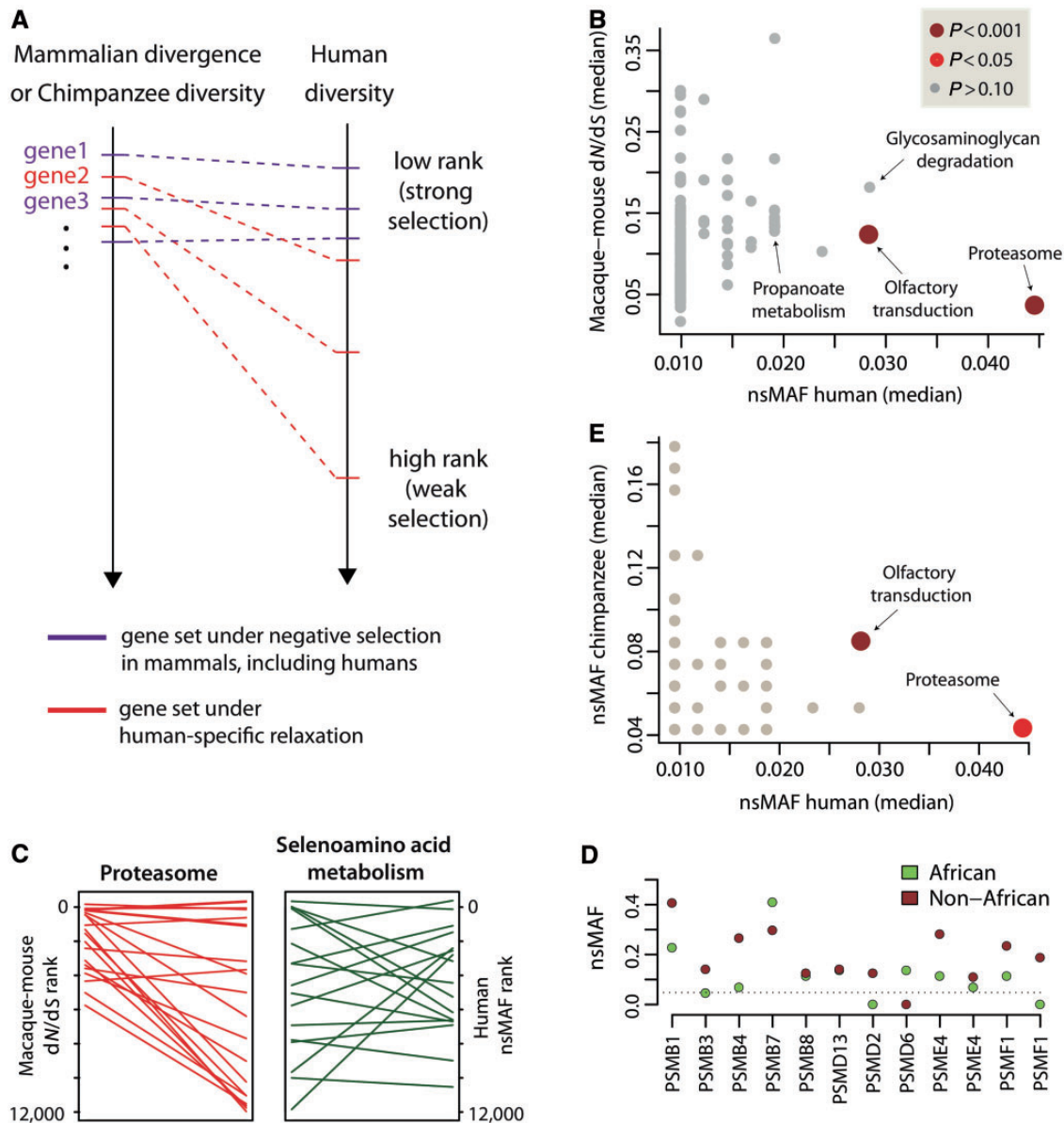


Fig. 1. (A) Representation depicting the tests for human-specific relaxed selection. Genes in a gene set are ranked according to functional human diversity (nsMAF), mammalian divergence (dN/dS), or chimpanzee diversity (nsMAF). The ranks are compared using a Wilcoxon signed rank test (supplementary material, Supplementary Material online). (B) Median human nsMAF and mouse-rhesus macaque dN/dS values among 181 KEGG pathways, each containing minimum five genes. Each dot represents one KEGG gene set. Groups that have significantly higher human diversity ranks relative to mammalian divergence ranks (as in A, Wilcoxon signed rank test), and also have significantly higher human diversity in member genes compared with the genome average (Wilcoxon rank sum test), are indicated in red. (C) Genes ($n = 12,405$) were ranked according to their relative mouse-rhesus macaque dN/dS (right) or human nsMAF (left) values, with higher ranks indicating weaker negative selection. Each line represents a gene and connects its dN/dS and nsMAF-based ranks. The red lines show proteasome and green lines show selenoamino acid metabolism genes. The two groups were chosen as examples representing significant shifts or no shifts toward human-specific relaxation, respectively. (D) nsMAF values at common proteasome nsSNPs (world-wide MAF > 0.05) in Africans and non-Africans. nsMAF in the non-African sample is higher, but not significantly (Wilcoxon signed rank test $P = 0.2$). (E) Median human nsMAF and chimpanzee nsMAF values among 182 KEGG pathways with minimum five genes, shown as in (A).

transduction and proteasome genes were among the top eight pathways exhibiting the most significant differences across 181 pathways (nominal Wilcoxon signed rank test $P < 0.05$; although the proteasome gene set was not significant after multiple testing correction) (fig. 1E; supplementary table S1, Supplementary Material online). By randomizing genes across KEGG groups and repeating the three tests

(i.e., comparing human nsMAF in each KEGG group with the genome average, comparing nsMAF per group with dN/dS, and with chimpanzee nsMAF) 10,000 times, we determined that finding two groups with significant results in all tests was unlikely ($P = 0.0015$).

We repeated the analysis using the mean instead of median nsMAF per gene, as well as the proportion of

common ($>5\%$ MAF) nonsynonymous SNPs per gene. Again, olfactory transduction and proteasome groups had significant results in all three tests (supplementary fig. S3A–D and table S2, Supplementary Material online). Next, we used the phase I version of the 1000 genomes data set, which contains a substantially larger set of individuals, although representing a smaller subset of worldwide human genetic variation (1000 Genomes Project Consortium 2012). Because our analyses focus on common polymorphism, we only included SNPs at MAF more than 0.005 in this data set ($\sim 54,000$ nsSNPs). Olfactory transduction and proteasome genes again showed a trend of higher human polymorphism relative to mammalian divergence or to chimpanzee polymorphism (although this trend was not statistically significant using mean nsMAF) (supplementary fig. S3F–H, Supplementary Material online).

A closer examination of the chimpanzee polymorphism at olfactory transduction-related genes revealed that these genes also had higher-than-average nsMAF in chimpanzees (fig. 2A and B). Relaxed negative selection on olfactory genes

may therefore be a shared trend between humans and chimpanzees, albeit more pronounced in humans. This is consistent with the tendency of the olfactory receptor family toward pseudogenization in primates, likely reflecting decreased reliance on olfactory perception in this lineage (Gilad et al. 2003, 2004; Kim et al. 2010; Matsui et al. 2010; Pierron et al. 2012), as well as positive or balancing selection (Williamson et al. 2007).

In contrast, there is no indication of higher nsMAF values among chimpanzee proteasome genes (fig. 2D), despite what was observed in humans (fig. 2C). Our chimpanzee data are limited to only 12 individuals and therefore the variability in chimpanzee MAF estimates are high. Nevertheless, we find no evidence of relaxation of selection in chimpanzees. For example, the nsMAF spectrum in chimpanzee Wnt signaling pathway genes, which regulate major developmental processes and are thus expected to remain under strong negative selection in both humans and chimpanzees, is comparable with the distribution of nsMAF in chimpanzee proteasome genes (fig. 2D and F). This provides another

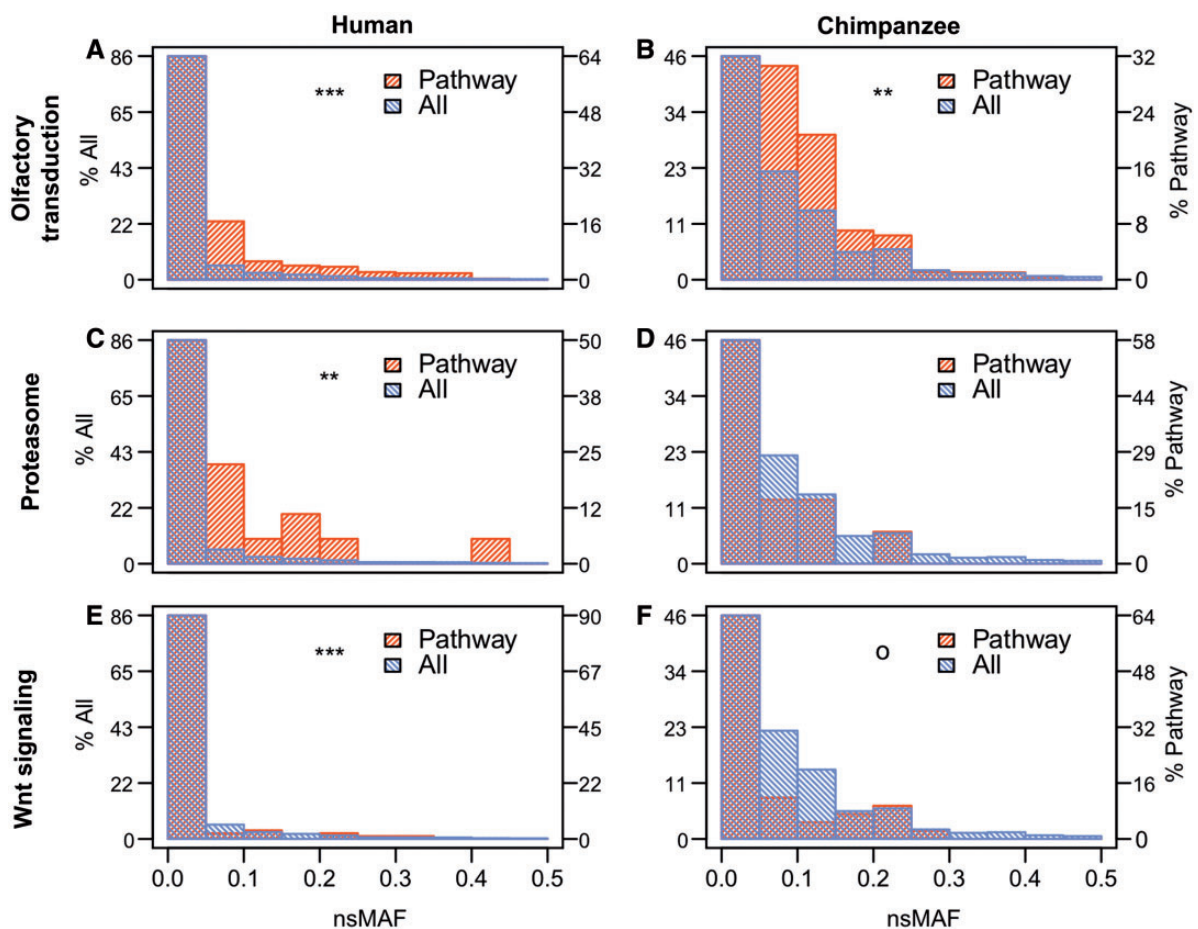


Fig. 2. Distribution of median MAF for nsSNPs and sSNPs in human (left panels) and chimpanzee (right panels), across genes in three KEGG pathways: olfactory transduction ($n = 323$), proteasome ($n = 18$), and Wnt signaling ($n = 87$), compared with the all genes annotated in KEGG and with nsSNPs in human and chimpanzee data sets ($n = 3,741$). Wnt signaling was chosen here as an example category that is expected to be under strong negative selection in both human and chimpanzee. The left and right y axes show the percentage of genes falling in a MAF quantile across all genes or genes in a KEGG pathway, respectively. We use the same subset of genes with SNPs detected in both human and chimpanzee. Note that the chimpanzee site frequency spectrum is less skewed to the left than that of human, due to the fact that we use 12 chimpanzee and 54 human subjects. The asterisks show significance measured in a two-sided Wilcoxon rank sum test. $^{\circ}P < 0.10$; $**P < 0.01$; $***P < 0.001$.

line of evidence that the proteasome genes are likely evolving under negative selection in chimpanzees.

Indeed, like the spliceosome, the proteasome is a major component of eukaryotic cells. This is the key machinery for degrading proteins marked by ubiquitin, which can be damaged or misfolded proteins, signaling proteins involved in cell cycle and apoptosis, or antigens (Driscoll and Goldberg 1990; Richter-Ruoff and Wolf 1993; Chondrogianni et al. 2003; see Ciechanover 2005 and Tanaka 2009 for reviews). The main structure contains a 20S catalytic core, consisting of two sets of seven alpha and seven beta subunits, and two sets of 19S regulatory particles, containing 18 distinct subunits each (fig. 3A and B). Each subunit is coded by one gene. Combined with additional peptides, these structures can further give rise to the immunoproteasome or the thymoproteasome (Tanaka 2009). Notably, among proteasome subunits, those with nsSNPs have slightly higher dN/dS levels relative to those without nsSNPs (supplementary fig. S4, Supplementary Material online). Still, all proteasome gene sets have significantly lower mammalian dN/dS values

compared with the rest of the genome, indicating strong conservation (Wilcoxon rank sum test $P < 0.05$).

Because high frequency coding polymorphism among a conserved gene set is unexpected, we asked whether the proteasome SNPs, or their high allele frequencies, may be caused by artifacts relating to the computational processing of the data. To do this, we first compared the frequencies of the 12 common ($>5\%$ nsMAF) proteasome SNPs identified in the Complete Genomics data set, with nsMAF calculated from the phase I version of the 1000 genomes data set (1000 Genomes Project Consortium 2012) (supplementary material, Supplementary Material online). This revealed high consistency between data sets in comparisons with 14 populations ($0.38 < \rho < 0.95$, $P < 0.05$ in 11 comparisons) (supplementary fig. S5, Supplementary Material online). However, because the samples included in the two data sets overlap, we used a second, independent exome sequencing data set of 200 Danish individuals (Li et al. 2010) (supplementary material, Supplementary Material online). Importantly, this data set is derived from primary tissue

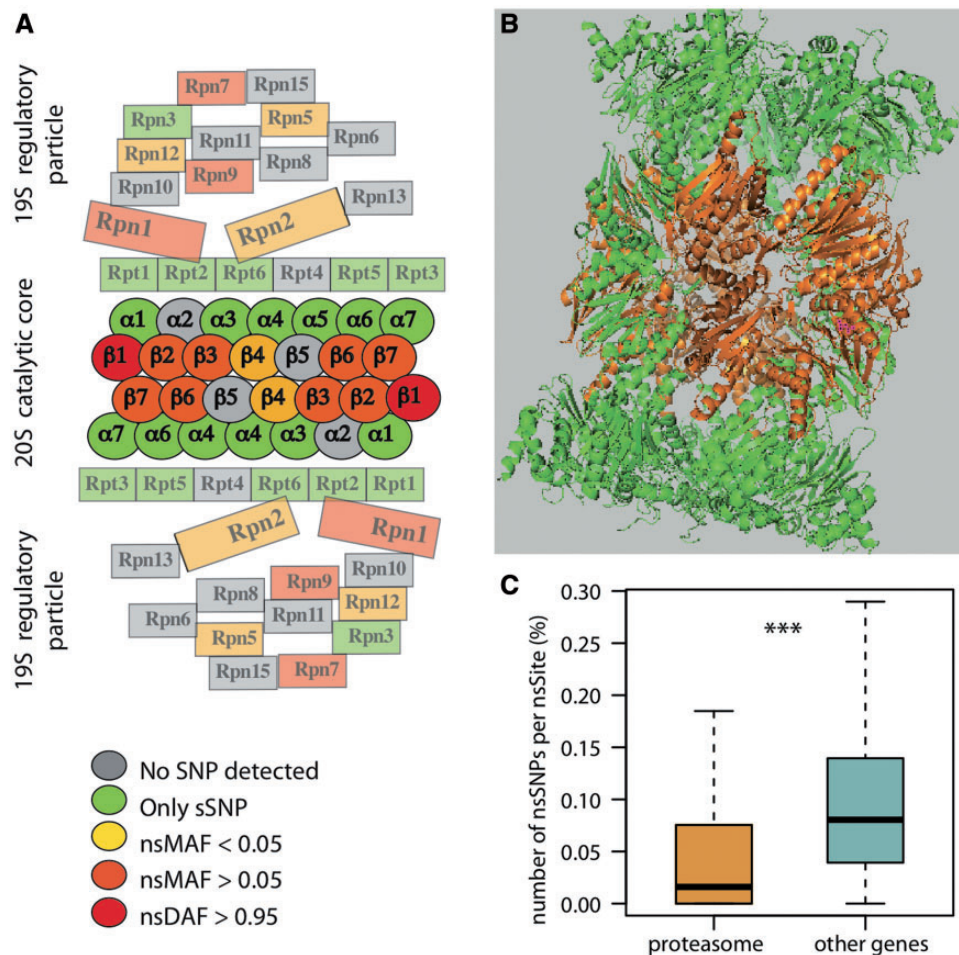


Fig. 3. (A) Schematic representation of the proteasome 20S core and 19S regulatory particles based on (Tanaka 2009). Peptides are colored with respect to their polymorphism characteristics. Note that the $\beta 1$ subunit coded by PSMB6 contains a SNP that is close to fixation in humans, with derived allele frequency (DAF) = 0.98. (B) Representation of bovine 20S core complex, with nsSNP containing chains shown in orange. (C) Box plot showing the percentage of number of nonsynonymous mutations per nonsynonymous site among proteasome genes compared with all genes with at least one mutation ($n = 17,294$) in the Complete Genomics data set. We assign zero to any gene where only synonymous SNPs are detected ($n = 2,191$). Note that among 33 proteasome genes, 15 contain only sSNPs but no nsSNP. Outliers are not shown. *** $P < 0.001$.

instead of cell lines, which also allows us the potential to exclude in vitro proteasome mutations. Ten proteasome nsSNPs from the Complete Genomics data set were detected in the Danish data set. All 10 had $MAF > 0.05$ in both data sets and the allele frequencies across the SNPs were highly correlated ($\rho = 0.71$, $P < 0.02$) (supplementary fig. S6A, Supplementary Material online). We then asked whether high allele frequencies may be influenced by copy number variation. Using sequencing depth in the Danish data set as proxy for copy number, we found no indication that proteasome sites had higher sequencing depth than other sites (supplementary fig. S6B and C, Supplementary Material online). Known paralogs are not likely to cause the signal, as human proteasome genes have no close paralogs, and proteasome genes with and without distant paralogs showed no difference in nsMAF (supplementary fig. S6D, Supplementary Material online). Also, we find no indication for higher synonymous polymorphism among proteasome genes (supplementary fig. S7, Supplementary Material online). Furthermore, we used a stringent cutoff for Hardy–Weinberg equilibrium testing to filter out SNPs with an excess of heterozygotes that could also be caused by paralogs (supplementary material, Supplementary Material online). In conclusion, the pattern of high frequency nonsynonymous polymorphisms in human proteasome genes is replicable between data sets and is unlikely to be artifactual.

We then considered the possibility that the identified high frequency proteasome nsSNPs may be affecting the least functional parts of the respective proteins. We thus tested whether sites with common proteasome nsSNPs ($MAF > 0.05$) show conservation across mammals. Six proteasome nsSNPs could be aligned to 13 high-quality mammalian genomes, and we found no substitutions among these species at these sites (supplementary material, Supplementary Material online). In contrast, in 39% of common nsSNPs in other genes we find at least one substitution (one-sided binomial test $P = 0.051$). These results imply that some of the proteasome subunit sites bearing common nonsynonymous polymorphism in humans are not evolving neutrally in other mammals, and may have functional effects.

We then examined the potential functional role of common proteasome nsSNP alleles. With respect to physiological function, no genome-wide disease association has been reported for the common proteasome SNPs. With respect to protein structure, PolyPhen (v2) (Adzhubei et al. 2010) predicted one of the 12 common mutations as “possibly-damaging,” and SIFT (Kumar et al. 2009) predicted two as “deleterious” (supplementary table S3, Supplementary Material online). These were in the *PSMB3* and *PSMB4* genes, encoding subunits of the 20S proteasome core complex (fig. 3A). The *PSMB4* mutation is within putative conserved domains according to BLASTP, and the *PSMB3* mutation is within an active site of the protein. Taken together with the conservation tendency across mammals, this information suggests that some proteasome gene-related SNPs in humans may affect functional sites.

Intuitively, only about one-half of proteasome genes as annotated in KEGG ($n = 18$) carry nsSNPs in the Complete

Genomics data set; the others ($n = 15$) only contain synonymous SNPs. This is significantly higher than the genome average (expected = 13%, observed = 45%, hypergeometric test $P < 0.001$). Likewise, the median density of nonsynonymous mutations per nonsynonymous site is five times lower among proteasome genes than the genome average (Wilcoxon rank sum test $P < 0.001$) (fig. 3C). Surveying the distribution of the peptides coded by nsSNP-bearing genes within the proteasome complex, we further noticed a peculiar pattern: within the 20S catalytic core, only beta subunits carried human nsSNPs, with four genes with common nsSNPs (fig. 3A and B). Yet another beta subunit, *PSMB6*, also carries a high frequency (>95%) derived nsSNP. In contrast, the seven alpha subunits only contained sSNPs. Thus, the human proteasome, as a whole, appears under negative selection, despite common nsSNPs in one layer of its catalytic core.

What could allow nonsynonymous mutations in the core human proteasome to reach such high frequencies? A number of scenarios are conceivable: balancing selection specific to humans, ongoing positive selection, compensatory mutations following a fixation event in humans, or relaxed constraints. Given the multiple roles of this complex, including immune response, stress response, and aging (Chondrogianni et al. 2003; reviewed in Ciechanover 2005; Tanaka 2009), positive or balancing selection on proteasome genes, due to human-specific changes in longevity or immune response is not implausible. To identify any signature of past or ongoing selection, we first surveyed human-chimpanzee differences in core proteasome genes. Aligning human, chimpanzee, and rhesus macaque beta subunit protein sequences revealed no human-specific substitution, only two high-frequency derived SNPs. Thus, these genes do not show evidence for adaptive change in their coding sequences in the human lineage. Proteasome genes also showed no indication of recent positive selection in European–American or in Yoruban individuals (supplementary fig. S8, Supplementary Material online), either as high haplotype homozygosity measured by iHS (Voight et al. 2006), or as extremely low or high Tajima’s D values indicating an excess of rare alleles or an excess of intermediate alleles, respectively (Tajima 1989). Neither did proteasome genes have high scores in a recently developed composite likelihood-based statistic for balancing selection (DeGiorgio M, Lohmueller KE, Nielsen R, in review). In fact, while proteasome genes had negative Tajima’s D values consistent with negative selection and/or human population growth, those proteasome subunits carrying common nsSNPs tended to have Tajima’s D values closer to zero than other proteasome genes, implying weaker negative selection. Notably, common nsMAF (>5%) SNPs do not show conspicuous differentiation between African and non-African populations (fig. 1D). We also find four common proteasome nsSNPs in the Denisovan genome (Meyer et al. 2012) (supplementary table S3, Supplementary Material online), which suggests that a possible shift in selection pressure appeared at least 800,000 years ago. Finally, using a primate gene expression data set including different tissues (Brawand et al. 2011), we found no

consistent differentiation in gene expression between humans and chimpanzees among nsSNP containing proteasome genes (data not shown).

To summarize, our genome-wide scan for human-specific relaxation of constraint identified two candidate functions: olfactory transduction and proteasomal degradation. The former is well-studied, but the latter is novel and unexpected. Our results do not imply an overall relaxation of constraint among proteasome subunits; rather, common nonsynonymous SNPs appear clustered in the complex, and include sites under constraint in other mammals. The trend appears specific to the hominin lineage. Still, given the small sample size of the chimpanzee data set used here, nonsynonymous polymorphism in proteasome genes require further investigation among great apes.

The reason for high frequency nonsynonymous polymorphism in human proteasome genes remains unclear. We find no indication for positive or balancing selection on the genes in question. One possibility is that a general trend of relaxation of constraint in humans (Wu and Li 1985; Ohta 1993; Mikkelsen et al. 2005; Kosiol et al. 2008) affected these previously conserved sites by chance, although the significant GO classification makes this explanation unlikely. Alternatively, one may speculate that changes in human evolution that affected functions regulated by the proteasome-ubiquitin pathway, such as increased longevity (Li and de Magalhães 2011), or increased intake of dietary protein (Finch and Stanford 2004; Babbitt et al. 2011), may have altered selective pressure on particular proteasome subunits. For example, the ubiquitin-proteasome system is a major contributor of protein turnover in muscle, which is directly affected by essential amino acid intake (Tawa et al. 1992; Wakshlag et al. 2003; Combaret et al. 2005). All known human diets are more protein rich than that of chimpanzees (Finch and Stanford 2004). It is thus conceivable that increased consumption of essential amino acids reduced selection pressure on the human proteasome's efficacy for maintaining protein turnover, allowing for slightly deleterious mutations at previously conserved sites. More studies of the functional effects of common proteasome mutations humans may help to clarify this enigmatic point in the future.

Supplementary Material

Supplementary figures S1–S8 and tables S1–S3 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

Acknowledgments

The authors thank Michael DeGiorgio and Kirk Lohmueller for kindly sharing their results, Rori Rohlf, members of the Rasmus Nielsen group at UC Berkeley, Philipp Khaitovich, Tang Kun, Hui Yang, and two anonymous reviewers for helpful suggestions. This work was supported by the European Molecular Biology Organization grants ALTF-229-2011 to M.F. and ALTF-1475-2010 to M.S., Miller Institute for Basic Research in Science grant to M.A.W.S., and

National Institutes of Health grants 35R01HG003229-07 to R.N. and 3R01HG003229-08S2 to E.H.-S.

References

- 1000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491:56–65.
- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nat Methods* 7:248–249.
- Babbitt CC, Warner LR, Fedrigo O, Wall CE, Wray GA. 2011. Genomic signatures of diet-related shifts during human origins. *Proc Biol Sci* 278:961–969.
- Brawand D, Soumillon M, Necsulea A, et al. (18 co-authors). 2011. The evolution of gene expression levels in mammalian organs. *Nature* 478:343–348.
- Cai JJ, Petrov DA. 2010. Relaxed purifying selection and possibly high rate of adaptation in primate lineage-specific genes. *Genome Biol Evol* 2: 393–409.
- Cargill M, Altshuler D, Ireland J, et al. (18 co-authors). 1999. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22:231–238.
- Chondrogianni N, Stratford FLL, Trougakos IP, Friguet B, Rivett AJ, Gonos ES. 2003. Central role of the proteasome in senescence and survival of human fibroblasts: induction of a senescence-like phenotype upon its inhibition and resistance to stress upon its activation. *J Biol Chem* 278:28026–28037.
- Ciechanover A. 2005. Proteolysis: from the lysosome to ubiquitin and the proteasome. *Nat Rev Mol Cell Biol* 6:79–87.
- Combaret L, Dardevet D, Rieu I, Pouch M-N, Béchet D, Taillandier D, Grizard J, Attaix D. 2005. A leucine-supplemented diet restores the defective postprandial inhibition of proteasome-dependent proteolysis in aged rat skeletal muscle. *J Physiol (Lond)* 569:489–499.
- Dickerson R. 1971. The structure of cytochrome c and the rates of molecular evolution. *J Mol Evol* 1:26–45.
- Driscoll J, Goldberg AL. 1990. The proteasome (multicatalytic protease) is a component of the 1500-kDa proteolytic complex which degrades ubiquitin-conjugated proteins. *J Biol Chem* 265: 4789–4792.
- Drmanac R, Sparks AB, Callow MJ, et al. (65 co-authors). 2010. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* 327:78–81.
- Duret L, Mouchiroud D. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. *Mol Biol Evol* 17:68–74.
- Fay J, Wu C. 2003. Sequence divergence, functional constraint, and selection in protein evolution. *Annu Rev Genomics Hum Genet* 4: 213–235.
- Finch CE, Stanford CB. 2004. Meat-adaptive genes and the evolution of slower aging in humans. *Quarter Rev Biol* 79:3–50.
- Gilad Y, Man O, Paabo S, Lancet D. 2003. Human specific loss of olfactory receptor genes. *Proc Natl Acad Sci U S A* 100:3324–3327.
- Gilad Y, Wiebe V, Przeworski M, Lancet D, Paabo S. 2004. Loss of olfactory receptor genes coincides with the acquisition of full trichromatic vision in primates. *PLoS Biol* 2:e5.
- Go Y, Niimura Y. 2008. Similar numbers but different repertoires of olfactory receptor genes in humans and chimpanzees. *Mol Biol Evol* 25:1897–1907.
- Hernandez RD. 2008. A flexible forward simulator for populations subject to selection and demography. *Bioinformatics* 24:2786–2787.
- Hvilsom C, Qian Y, Bataillon T, et al. (17 co-authors). 2012. Extensive X-linked adaptive evolution in central chimpanzees. *Proc Natl Acad Sci U S A* 109:2054–2059.
- Kanehisa M, Araki M, Goto S, et al. (11 co-authors). 2008. KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36: D480–D484.
- Kim HL, Igawa T, Kawashima A, Satta Y, Takahata N. 2010. Divergence, demography and gene loss along the human lineage. *Philos Trans R Soc Lond B Biol Sci* 365:2451–2457.

- Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biol.* 3:RESEARCH0008.
- Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. 2008. Patterns of positive selection in six mammalian genomes. *PLoS Genet.* 4:e1000144.
- Kuma K, Iwabe N, Miyata T. 1995. Functional constraints against variations on molecules from the tissue level: slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families. *Mol Biol Evol.* 12:123–130.
- Kumar P, Henikoff S, Ng PC. 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4:1073–1081.
- Li WH, Tanimura M. 1987. The molecular clock runs more slowly in man than in apes and monkeys. *Nature* 326:93–96.
- Li Y, de Magalhães JP. 2011. Accelerated protein evolution analysis reveals genes and pathways associated with the evolution of mammalian longevity. *Age (Dordr)* 35:301–314.
- Li Y, Vinckenbosch N, Tian G, et al. (38 co-authors). 2010. Resequencing of 200 human exomes identifies an excess of low-frequency non-synonymous coding variants. *Nat Genet.* 42:969–972.
- Matsui A, Go Y, Niimura Y. 2010. Degeneration of olfactory receptor gene repertoires in primates: no direct link to full trichromatic vision. *Mol Biol Evol.* 27:1192–1200.
- Meyer M, Kircher M, Gansauge M-T, et al. (34 co-authors). 2012. A high-coverage genome sequence from an archaic denisovan individual. *Science* 338:222–226.
- Mikkelsen T, Hillier L, Eichler E, et al. (67 co-authors). 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87.
- Nielsen R, Bustamante C, Clark AG, et al. (13 co-authors). 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* 3:e170.
- Ohno S. 1972. An argument for the genetic simplicity of man and other mammals. *J Hum Evol.* 1:651–662.
- Ohta T. 1993. An examination of the generation-time effect on molecular evolution. *Proc Natl Acad Sci U S A.* 90:10676–10680.
- Ohta T, Kimura M. 1971. On the constancy of the evolutionary rate of cistrons. *J Mol Evol.* 1:18–25.
- Pierron D, Cortés NG, Letellier T, Grossman LI. 2012. Current relaxation of selection on the human genome: tolerance of deleterious mutations on olfactory receptors. *Mol Phylogenet Evol.* 66:558–564.
- Pollard K, Salama S, King B, et al. (13 co-authors). 2006. Forces shaping the fastest evolving regions in the human genome. *PLoS Genet.* 2:e168.
- Pollard K, Salama S, Lambert N, et al. (16 co-authors). 2006. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443:167–172.
- Prabhakar S, Noonan J, Paabo S, Rubin E. 2006. Accelerated evolution of conserved noncoding sequences in humans. *Science* 314:786.
- Richter-Ruoff B, Wolf DH. 1993. Proteasome and cell cycle. Evidence for a regulatory role of the protease on mitotic cyclins in yeast. *FEBS Lett.* 336:34–36.
- Rouquier S, Taviaux S, Trask BJ, Brand-Arpon V, van den Engh G, Demaille J, Giorgi D. 1998. Distribution of olfactory receptor genes in the human genome. *Nat Genet.* 18:243–250.
- Siepel A, Bejerano G, Pedersen JS, et al. (16 co-authors). 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15:1034–1050.
- Sunyaev SR, Lathe WC, Ramensky VE, Bork P. 2000. SNP frequencies in human genes an excess of rare alleles and differing modes of selection. *Trends Genet.* 16:335–337.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tanaka K. 2009. The proteasome: overview of structure and functions. *Proc Jpn Acad Ser B Phys Biol Sci.* 85:12–36.
- Tawa NE, Kettelhut IC, Goldberg AL. 1992. Dietary protein deficiency reduces lysosomal and nonlysosomal ATP-dependent proteolysis in muscle. *Am J Physiol.* 263:E326–E334.
- Terao K, Mikami A, Saito A, et al. (15 co-authors). 2005. Identification of a protanomalous chimpanzee by molecular genetic and electroretinogram analyses. *Vision Res.* 45:1225–1235.
- Voight B, Kudravalli S, Wen X, Pritchard J. 2006. A map of recent positive selection in the human genome. *PLoS Biol.* 4:e72.
- Wakshlag JJ, Barr SC, Ordway GA, Kallfelz FA, Flaherty CE, Christensen BW, Shepard LA, Nydam DV, Davenport GM. 2003. Effect of dietary protein on lean body wasting in dogs: correlation between loss of lean mass and markers of proteasome-dependent proteolysis. *J Anim Physiol Anim Nutr (Berl).* 87:408–420.
- Ward LD, Kellis M. 2012. Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* 337:1675–1678.
- Williamson S, Hubisz M, Clark A, Payseur B, Bustamante C, Nielsen R. 2007. Localizing recent adaptive evolution in the human genome. *PLoS Genet.* 3:e90.
- Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci U S A.* 82:1741–1745.