

Fixation of deleterious mutations at critical positions in human proteins

Sankar Subramanian

*Griffith School of Environment, Griffith University, 170 Kessels Road, Nathan Qld 4111,
Australia*

Keywords: human evolution, natural selection, deleterious mutations, fixation, WGA and population genetics theory

Running head: Fixation of deleterious mutations in human

Title length: 73 characters (including spaces)

Abstract length: 245 words

Total length of text: 26,662 characters (including spaces)

Total page requirement: 5.5 pages

Number of figures: 3

Number of tables: 2

Number of references: 27

Address for correspondence:

Dr. Sankar Subramanian
Griffith School of Environment
Griffith University
170 Kessels Road
Nathan QLD 4111
Australia

Phone: +61-7-3735 7495

Fax: +61-7-3735 7459

E-mail: s.subramanian@griffith.edu.au

Abstract

Deleterious mutations associated with human diseases are predominantly found in conserved positions and positions that are essential for the structure and/or function of proteins. However these mutations are purged from the human population over time and prevented from being fixed. Contrary to this belief here I show that high proportions of deleterious amino acid changing mutations are fixed at positions critical for the structure and/or function of proteins. Similarly a high rate of fixation of deleterious mutations was observed in slow-evolving amino acid positions of human proteins. The fraction of deleterious substitutions was found to be two times higher in relatively conserved amino acid positions than in highly variable positions. This study also found fixation of a much higher proportion of radical amino acid changes in primates compared to rodents and artiodactyls in slow-evolving positions. Previous studies observed a higher proportion of nonsynonymous substitutions in humans compared to other mammals, which was taken as indirect evidence for the fixation of deleterious mutations in humans. However the results of this investigation provide direct evidence for this prediction by suggesting that the excess nonsynonymous mutations fixed in humans are indeed deleterious in nature. Furthermore these results suggest that studies on disease associated mutations should consider that a significant fraction of such deleterious mutations have already been fixed in the human genome and thus the effects of new mutations at those amino acid positions may not necessarily be deleterious and might even result in reversion to benign phenotypes.

Introduction

The association of mutations with specific human genetic diseases has been known for a long time. Point mutations that change the encoded amino acids constitute more than half of the mutations (~55%) associated with diseases (Stenson et al. 2009). A number of studies have examined the patterns of the deleterious amino acid changing mutations in order to distinguish them from neutral population variations (Ng and Henikoff 2001; Sunyaev et al. 2001; Mooney and Klein 2002; Subramanian and Kumar 2006). Based on large-scale observations these studies predicted the deleterious potential of amino acid replacement mutations and suggested that the mutations occurring at positions that are conserved between human and other species and at positions that are necessary for the structure and/or function of the proteins are likely to be deleterious in nature (Ng and Henikoff 2001; Sunyaev et al. 2001; Mooney and Klein 2002; Subramanian and Kumar 2006). Furthermore the mutations that result in radical amino acid changes (involving dissimilar amino acids with respect to their biochemical properties) are more likely to be deleterious than those that result in conservative amino acid changes (involving amino acids with similar biochemical properties) (Sunyaev et al. 2001; Subramanian and Kumar 2006). These predictions are based on population genetic theories that suggest that deleterious mutations in a population are eliminated over time and are thus prevented from fixation (Kimura 1983). Therefore mutations that occur in positions that are critical for the structure and/or function of proteins are not fixed in human (and other) populations, which results in conservation of amino acids in critical positions of proteins in human and other species. Based on these assumptions several methods have been developed to predict the deleterious potential of amino acid changing mutations (Ng and Henikoff 2003; Bromberg, Yachdav, and Rost 2008; Adzhubei et al. 2010).

In summary the prediction of human disease-associated mutations is based on the assumption that deleterious mutations are not fixed in humans. However based on standard population genetic theories, fixation of deleterious mutations in the human genome is expected due to their small effective population size (Ohta 1972; Kimura 1983). Previous studies have shown that the ratio of divergence at nonsynonymous- to synonymous positions ($\omega=dN/dS$) obtained for humans or primates was significantly higher than that obtained for species with relatively large population sizes such as rodents or artiodactyls (Keightley and Eyre-Walker 2000; Mikkelsen et al. 2005). For instance comparative studies on primate and rodent genomes revealed that the genome-wide estimate of ω for the human-chimp pair (0.23) is much higher than that obtained for the mouse-rat comparison (0.13) (Mikkelsen et al. 2005). The higher ω of humans suggests fixation of an excess fraction of nonsynonymous mutations. These studies argued that this excess proportion of nonsynonymous substitutions in humans reflects the fixation of slightly deleterious amino acid replacement mutations in the human or primate lineage (Keightley and Eyre-Walker 2000; Mikkelsen et al. 2005). However it is not known whether the fraction of nonsynonymous mutations fixed in the human lineage is indeed deleterious in nature. Therefore direct evidence revealing the deleterious potential of amino acid substitutions in the human lineage is needed. Hence to examine this I have conducted a comparative genomic approach using the protein-coding genes from human and other vertebrate genomes. Since the methods that predict disease-associated mutations assume fixation of only neutral (or beneficial) mutations in humans, this study on the fixation of deleterious mutations has important implications for genome-wide disease association studies.

Materials and Methods

Genomic sequence data

Protein and cDNA sequence data of human (*Homo sapiens*), macaque (*Macaca mulatta*), cow (*Bos taurus*), pig (*Sus scrofa*), mouse (*Mus musculus*), rat (*Rattus norvegicus*), opossum (*Monodelphis domestica*), chicken (*Gallus gallus*), toad (*Xenopus tropicalis*) and zebra fish (*Danio rerio*) were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). A reciprocal BLAST (Altschul et al. 1997) hit approach was employed to obtain the genes from each species that are orthologous to human using the significance threshold described by Duret et.al. (Duret, Mouchiroud, and Gouy 1994). All orthologous protein sequences of each gene were aligned using CLUSTALW (Larkin et al. 2007) and the cDNA alignment for each gene was created using the protein sequence alignment as a guide. A genomic alignment of the 10 vertebrate species was generated by concatenating all gene alignments. All positions with alignment gap(s) were excluded. Population polymorphism data from the genomes of two Europeans, an Asian and an African was obtained from a previous study (Subramanian 2011).

Identification of critical amino acid sites and conserved domains

Information about the amino acid positions that are necessary for protein structure and function were obtained from the annotations in the GenBank reference protein sequence files. This information is based on the conserved domain database resource (CDD) (Marchler-Bauer et al. 2005). The amino acid positions that are known to affect the structural stability and/or function of proteins are listed under the subtitle “Sites” in the GenBank formatted files. In this study the corresponding codon positions are designated as “critical sites” and rest of the sites are referred as “non-critical sites”. Similarly the conserved domains of the protein sequences are annotated

as “Region” in the reference sequence file. In the present study these positions are referred as “conserved domains” and the remaining parts of the sequences are called “non-domains”.

Determination of the relative conservation of amino acid positions

To determine the rate of evolution of individual amino acid sites the method described by a previous study was used (Subramanian and Kumar 2006). Protein sequences from opossum, chicken, toad and zebra fish were taken from the genomic alignment of 10 species mentioned above. Using the *codeml* package of the software *PAML* (Yang 2007) site variability was estimated by employing a discrete gamma model to accommodate variation among sites and a *JTT* matrix was used to model substitutions among amino acids. This program estimated the rate of evolution for each amino acid position and these rate estimates ranged from 0.27-3.97. The amino acid sites of the genomic alignment were separated into four groups or Conservation Indices based on the evolutionary rates of the sites. The sites with rate estimates of 0.0-1.0, 1.01-2.0, 2.01-3.0 and 3.01-4.0 were assigned Conservation Indices 1 to 4 respectively. These Conservation Indices were then assigned to corresponding amino acid positions in the remaining six mammalian sequences in the genomic alignment.

Estimation of evolutionary divergence and conservative and radical replacement distance

Evolutionary divergences at synonymous and nonsynonymous positions along with the standard errors were estimated using the *codeml* package of *PAML* with the pair-wise option (Yang 2007). To estimate the proportions of radical and conservative amino acid changes per site (and SE), the software *Hon-new* was used (Zhang 2000). Generally conservative changes are those that occur between amino acids having similar biochemical properties and radical changes are those that occur between dissimilar amino acids. There are several ways amino acids could be grouped

based on the similarity in their biochemical properties. However a recent study examined this issue and showed that one particular grouping (classification A) performed better than others in capturing the magnitude of selection constraints on mammalian protein coding genes (Hanada, Shiu, and Li 2007). Therefore the current study used this amino acid classification along with transition/transversion ratios of 2.75, 2.7 and 2.1 for primates, artiodactyls and rodents respectively following a previous study (Rosenberg, Subramanian, and Kumar 2003).

Estimation of deleterious amino acid substitution in human

Population genetic theories suggest that the fixation of slightly deleterious mutations is determined by the product of effective population size (N_e) and selection coefficient (s). Therefore a higher fraction of deleterious mutations will be fixed in a species with a small population size compared to that fixed in a species with a large population size. Since the population size of humans or primates is known to be smaller than that of rodents or artiodactyls the former will accumulate more deleterious substitutions than the latter. The ratio of divergence at nonsynonymous- to synonymous positions ($\omega=dN/dS$) denotes the fraction of amino acid mutations fixed with respect to neutral mutations. Previous studies have shown a correlation between ω and species generation time and the latter was used as a proxy for population size (Ohta 1972; Keightley and Eyre-Walker 2000). It has been shown that the ω estimated for primates was much higher than that estimated for rodents and artiodactyls (Ohta 1993; Keightley and Eyre-Walker 2000). The difference in the ω ratios between species with small and large population sizes reflects the excess fraction of nonsynonymous mutation fixed in species with a small population size and theories predict this excess substitutions to be deleterious in nature.

Based on this rationale using the ω ratios of primates (ω_{Primates}) and rodents (ω_{Rodents}) or artiodactyls ($\omega_{\text{Artiodactyls}}$), the proportion of deleterious amino acid replacement mutations fixed in primates (δ) can be estimated as:

$$\delta = 1 - \frac{\omega_{\text{Rodents/Artiodactyls}}}{\omega_{\text{Primates}}}$$

The estimate δ is the proportion of deleterious nonsynonymous substitutions out of all the amino acid replacement mutations fixed in humans.

Results and Discussion

Fixation of amino acid replacement mutations at critical positions of human proteins

First, I estimated the proportion of deleterious substitutions (δ) at positions that are critical for the structure and/or function of human proteins. For this purpose, 1457 orthologous protein-coding genes from human, macaque, cow, pig, mouse and rat were aligned and positions containing gaps were excluded. The amino acid positions that are known to be involved in structural stability of proteins and those required for proper function of proteins such as DNA/RNA binding sites, substrate/ligand/receptor binding sites and other active sites (critical sites) were identified from the annotations of human reference protein sequences and were separated from the remainder of the sites (non-critical sites). For primates (human-macaque), artiodactyls (cow-pig) and rodents (mouse-rat) pairs, divergences at synonymous and nonsynonymous sites were estimated separately for the critical sites and non-critical sites. Although divergence at synonymous sites are largely similar between critical and non-critical positions, nonsynonymous divergences is 2-4 times smaller at positions that are critical for proteins compared to those that are non-critical (Table 1). This suggests a much higher selection pressure on critical amino acid

positions than that on non-critical sites of human proteins. Importantly δ estimated for critical positions is significantly higher than that estimated for non-critical protein sites (Figure 1A). The δ estimates obtained using critical amino acid positions are 61% and 73% higher than those obtained using non-critical protein positions for the primate-rodent and primate-artiodactyls comparisons respectively ($P < 0.0001$ using a Z test). Similar highly significant results ($P < 0.0001$) were obtained using a bootstrap (1000 replications) procedure by re-sampling the 1457 genes. This removes any bias due to the presence of a few genes with very high nonsynonymous substitution rates. Some of the nonsynonymous changes observed in the human lineage might be due to the presence of replacement polymorphisms. Therefore I compared each nucleotide change with four human genomes belonging to two Europeans, an Asian and an African. This comparative analysis showed that in 98% of the cases the same nucleotide was observed in the reference human genome (used in this study) and in all four human genomes. Therefore the results of this study will not be influenced by the presence of population polymorphisms in the data.

A similar pattern was observed when the δ estimates were obtained using the conserved domains of 3399 human proteins. These domains in each protein were identified from the annotations given in reference protein sequences that are based on information from the conserved domain database (Marchler-Bauer et al. 2005). The δ obtained for the conserved protein domains is 32%-37% higher than that obtained for non-domain regions of proteins and these differences were highly significant ($P < 0.0001$) (Figure 1B). It is interesting to note that δ estimated using rodents is always higher than that estimated using artiodactyls (Table 1). Using generation time as a proxy for N_e it can be inferred that artiodactyls have relatively smaller N_e compared to

rodents as the generation time of the former is longer than the latter and these two measures have a negative relationship (Ohta 1972; Ohta 1993; Keightley and Eyre-Walker 2000). Therefore the rate of fixation of deleterious mutations in artiodactyls is expected to be higher than that of rodents as this rate is known to be inversely proportional to N_e . It is clear from Table 1 that ω of artiodactyls is higher than that of rodents, which results in a small δ for the primate-artiodactyls comparison than that for the primate-rodent comparison (Figure 1).

The measure δ of this study generally underestimates the fraction of deleterious substitutions because deleterious mutations will also be fixed (to a much lesser extent) in rodents or artiodactyls. However this relative estimate is adequate to examine the patterns of deleterious substitutions in the human genome, which is the focus of the present study. The measure δ is estimated under the assumption that the fraction of adaptive nonsynonymous substitutions in mammals is negligible. Since the fixation of slightly beneficial mutations is also determined by N_e s a much higher fraction of adaptive mutations is expected to be fixed in rodents than primates due to the large effective population sizes (N_e) of the former. Hence, assuming a much higher fraction of adaptive substitutions in mammals will only make the δ estimates of this study more conservative. Furthermore, the δ estimates are also based on the assumption of neutral evolution in synonymous sites. If this assumption is not valid then the observed dS is less than that expected under neutrality and thus ω ratios will be overestimated. However, the magnitude of selection in synonymous sites is higher for rodents (or artiodactyls) than primates due to the relatively high N_e of the former. Therefore the magnitude of the overestimation of ω will be

much higher for rodents (or artiodactyls) than primates, which will also make the δ estimates more conservative.

Proportions of deleterious nonsynonymous substitutions at conserved positions

In the previous analysis sites that are critical for protein structure/function were identified directly based on information from experimental studies. However nature helps us to identify these critical sites indirectly because mutations occurring in these sites are eliminated, which results in high similarity of the amino acids in these positions between species. Hence a multiple sequence alignment of distantly related species could be used to identify critical amino acid sites that are generally conserved across taxa. To explore this, protein sequences of a marsupial (opossum), a bird (chicken), an amphibian (Xenopus frog) and a fish (Zebra fish) were added to the six mammalian sequences. Multiple genomic alignment of 1978 genes from 10 species was created and positions with alignment gaps were excluded. The relative conservation or the evolutionary rate of each amino acid position was determined using the *codeml* program of the PAML software (Yang 2007). Based on the rate of evolution, amino acid positions were grouped into four categories or Conservation Indices (see methods). Under this classification, Conservation Index 1 consists of invariant (identical) as well as less variable or slow evolving amino acid positions and higher indices consist of positions with higher variability or faster rates of evolution. The Conservation Indices determined for each amino acid position of the non-mammal proteins were then assigned to corresponding positions in the mammalian proteins in the genomic alignment. This is under the assumption of similar selection pressures on the corresponding amino acid positions of mammals and non-mammals. The codons of amino acids belonging to each Conservation Index were then extracted from the six mammalian sequences

and pairwise dN and dS values were estimated for the human-macaque, cow-pig and mouse-rat pairs. This method of analysis makes the estimation of δ independent from the estimation of Conservation Index as the former was estimated using the six mammalian sequences and the four non-mammal vertebrates were used to estimate the latter.

The results of this analysis show a similarity of dS estimated for sites belonging to different Conservation Indices and this is true for all three pairs of mammals analyzed (Table 2). However the dN obtained for highly conserved sites are 8-12 times smaller than that obtained for highly variable positions, which suggests a high selection pressure on slow-evolving sites. Importantly, there is a negative relationship between the level of conservation of sites and the proportion of deleterious amino acid replacement substitutions in primates (Figure 2). δ values obtained for highly conserved amino acid positions (0.433 and 0.344 for primate-rodent and primate-artiodactyls comparisons respectively) are 2-3 times higher than that estimated for highly variable positions (0.214 and 0.119). Interestingly, the magnitude of this difference is much higher than the difference in δ estimates obtained for critical and non-critical protein sites (only 60%-70%, see Figure 1). This suggests that there could be a high fraction of sites necessary for protein structure and/or function that have not yet been experimentally identified. Due to this limitation these unidentified critical sites are currently grouped in the category “non-critical sites”. Therefore the difference observed between the critical and non-critical sites is an underestimate.

High fraction of radical amino acid substitutions in human

It is well known that the types of amino acid changes also affect the structure or function of a protein. Generally changes involving amino acids with dissimilar biochemical properties (radical changes) are more detrimental to protein structure or function than those involving similar amino acids (conservative changes). Zhang (Zhang 2000) developed a simple but elegant method to estimate the proportions of radical (dR) and conservative changes (dC). The ratio of radical- to conservative amino acid changes (dR/dC) captures the magnitude of selective constraints on proteins based on the type of amino acid substitutions alone (Hanada, Shiu, and Li 2007). This ratio was estimated for the protein-coding genes of the human-macaque, cow-pig and mouse-rat pairs. To determine conservative and radical amino acid changes I used the classification developed by Hanada et.al (Hanada, Shiu, and Li 2007), which was shown to perform better than other types of classifications in capturing selection pressure on mammalian genes. The dR/dC ratio estimated for primates was found to be 18% and 10% higher than the ratios obtained for rodents and artiodactyls respectively and the differences are highly significant ($P < 0.0001$, Z - test) (Figure 3A). The high dR/dC ratio of primates is suggestive of high fraction of radical amino acid substitutions. This points to the high rate of fixation of deleterious amino acid replacement mutations in humans compared to rodents or artiodactyls (Hughes and Friedman 2009). The population size effect is also evident from the fact that the ratio of radical- to conservative amino acid changes (dR/dC) estimated for artiodactyls is intermediate to the ratios obtained for primates and rodents (Figure 3A). Next, the dR/dC ratios were estimated for the same mammalian pairs but using only conserved or highly variable amino acid positions. As expected dR/dC ratios estimated for the conserved positions were much smaller than those obtained using highly variable sites, which reflects the relative intensity of selective constraints at these positions. However the dR/dC ratio estimated using conserved coding sites is 38% ($P <$

0.0001) and 28% ($P = 0.0022$) higher for the human-macaque pair compared to those obtained for the mouse-rat and the cow-pig pairs respectively (Figure 3B). In contrast, these ratios were similar among the mammalian pairs at highly variable amino acid positions as differences in the dR/dC ratios between primate and rodent ($P = 0.36$) and between primate and artiodactyls ($P = 0.56$) are not statistically significant (Figure 3B). These results suggest the fixation of a much higher proportion of radical amino acid mutations in conserved amino acid sites of human proteins. Since radical amino acid changes are more disruptive to protein structure/function the above observation clearly supports and confirms the fixation of a high fraction of deleterious mutations in humans.

Conclusions

Previous studies observed a much higher fraction of nonsynonymous substitutions in primates compared to other mammals and this pattern was taken as the indirect evidence for the theoretical prediction of fixation of deleterious mutations in primates owing to their small population size (Ohta 1972; Ohta 1993; Keightley and Eyre-Walker 2000; Mikkelsen et al. 2005). The present study provides direct evidence for this prediction by showing that nonsynonymous mutation fixed in humans are indeed deleterious in nature. This is based on two observations: a) in primates a much higher fraction of amino acid replacement mutations is fixed at amino acid positions crucial for protein structure/function compared to rodents or artiodactyls, b) substitutions in these positions are more radical (with respect to the biochemical properties of amino acids) in primates compared to those of other mammals. It is well known that mutations at critical positions are more disruptive than those at non-critical sites and that radical amino acid changes are more detrimental to protein structure or function than conservative amino acid

changes. Therefore fixation of such deleterious mutations is largely determined by genetic drift in primates as selection is inefficient in eliminating these mutations due to small population sizes. It should be noted that the compensatory mutations reported before (Kondrashov, Sunyaev, and Kondrashov 2002) are a different class of mutations as the deleterious effects of these mutations are compensated by other mutations and thus they are not deleterious any more.

A number of methods have been developed to identify the mutations associated with human diseases and to distinguish them from neutral population polymorphisms (Ramensky, Bork, and Sunyaev 2002; Ng and Henikoff 2003; Bromberg, Yachdav, and Rost 2008). These methods are widely used in genome-wide association studies using human exomes (Boyko et al. 2008; Ng et al. 2008; Choi et al. 2009; Cooper et al. 2010; Ng et al. 2010). All these methods use information about protein structure and function, conservation of amino acids based on multiple sequence alignment and type of amino acid changes. These methods predict that if a mutation occurs at a critical position in a protein and/or at a conserved site and/or the amino acid change is radical then it is highly likely to be associated with a disease (Ng and Henikoff 2003; Bromberg, Yachdav, and Rost 2008; Adzhubei et al. 2010). However this study showed that a high fraction of such deleterious mutations is actually fixed in humans. This result is supported by the findings of a recent study, which showed that amino acid replacement mutations fixed in humans since our divergence from Neanderthal were largely radical and predominantly found at conserved amino acid positions (Burbano et al. 2010). Therefore the newly arising nonsynonymous mutations (in human populations) at these positions might be neutral or even beneficial due to possible reversion to the ancestral phenotype. Such new amino acid replacement mutations might spread through the population despite the fact that they occurred at

critical or conserved protein sites. This study suggests that the methods that predict the disease-causing potential of mutations should consider the possibility that a fraction of fixed human mutations could also be deleterious and that new mutations in those positions might actually result in benign phenotypes.

Acknowledgments

The author is grateful to David Lambert and thanks Leon Huynen and two anonymous reviewers for critical comments.

References

- Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev. 2010. A method and server for predicting damaging missense mutations. *Nature Methods* **7**:248-249.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**:3389-3402.
- Boyko, A. R., S. H. Williamson, A. R. Indap et al. (14 co-authors). 2008. Assessing the evolutionary impact of amino acid mutations in the human genome. *Plos Genetics* **4**:e1000083.
- Bromberg, Y., G. Yachdav, and B. Rost. 2008. SNAP predicts effect of mutations on protein function. *Bioinformatics* **24**:2397-2398.
- Burbano, H. A., E. Hodges, R. E. Green et al. (20 co-authors). 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* **328**:723-725.
- Choi, M., U. I. Scholl, W. Ji et al. (14 co-authors). 2009. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A* **106**:19096-19101.
- Cooper, G. M., D. L. Goode, S. B. Ng, A. Sidow, M. J. Bamshad, J. Shendure, and D. A. Nickerson. 2010. Single-nucleotide evolutionary constraint scores highlight disease-causing mutations. *Nature Methods* **7**:250-251.
- Duret, L., D. Mouchiroud, and M. Gouy. 1994. HOVERGEN: a database of homologous vertebrate genes. *Nucleic Acids Research* **22**:2360-2365.
- Hanada, K., S. H. Shiu, and W. H. Li. 2007. The nonsynonymous/synonymous substitution rate ratio versus the radical/conservative replacement rate ratio in the evolution of mammalian genes. *Mol Biol Evol* **24**:2235-2241.
- Hughes, A. L., and R. Friedman. 2009. More radical amino acid replacements in primates than in rodents: Support for the evolutionary role of effective population size. *Gene* **440**:50-56.
- Keightley, P. D., and A. Eyre-Walker. 2000. Deleterious mutations and the evolution of sex. *Science* **290**:331-333.
- Kimura, M. 1983. *The neutral theory of molecular evolution*. Cambridge University press, Cambridge.

- Kondrashov, A. S., S. Sunyaev, and F. A. Kondrashov. 2002. Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci U S A* **99**:14878-14883.
- Larkin, M. A., G. Blackshields, N. P. Brown et al. (13 co-authors). 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**:2947-2948.
- Marchler-Bauer, A., J. B. Anderson, P. F. Cherukuri et al. (24 co-authors). 2005. CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Research* **33**:D192-196.
- Mikkelsen, T. S., L. W. Hillier, E. E. Eichler et al. (67 co-authors). 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**:69-87.
- Mooney, S. D., and T. E. Klein. 2002. The functional importance of disease-associated mutation. *BMC Bioinformatics* **3**:24.
- Ng, P. C., and S. Henikoff. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**:3812-3814.
- Ng, P. C., and S. Henikoff. 2001. Predicting deleterious amino acid substitutions. *Genome Res* **11**:863-874.
- Ng, P. C., S. Levy, J. Huang, T. B. Stockwell, B. P. Walenz, K. Li, N. Axelrod, D. A. Busam, R. L. Strausberg, and J. C. Venter. 2008. Genetic variation in an individual human exome. *Plos Genetics* **4**:e1000160.
- Ng, S. B., K. J. Buckingham, C. Lee et al. (12 co-authors). 2010. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**:30-35.
- Ohta, T. 1972. Evolutionary Rate of Cistrons and DNA Divergence. *Journal of Molecular Evolution* **1**:150-157.
- Ohta, T. 1993. An examination of the generation-time effect on molecular evolution. *Proc Natl Acad Sci U S A* **90**:10676-10680.
- Ramensky, V., P. Bork, and S. Sunyaev. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Research* **30**:3894-3900.
- Rosenberg, M. S., S. Subramanian, and S. Kumar. 2003. Patterns of transitional mutation biases within and among mammalian genomes. *Mol Biol Evol* **20**:988-993.
- Stenson, P. D., M. Mort, E. V. Ball, K. Howells, A. D. Phillips, N. S. Thomas, and D. N. Cooper. 2009. The Human Gene Mutation Database: 2008 update. *Genome Med* **1**:13.

- Subramanian, S. 2011. High proportions of deleterious polymorphisms in constrained human genes. *Mol Biol Evol* **28**:49-52.
- Subramanian, S., and S. Kumar. 2006. Evolutionary anatomies of positions and types of disease-associated and neutral amino acid mutations in the human genome. *BMC Genomics* **7**:306.
- Sunyaev, S., V. Ramensky, I. Koch, W. Lathe, 3rd, A. S. Kondrashov, and P. Bork. 2001. Prediction of deleterious human alleles. *Hum Mol Genet* **10**:591-597.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**:1586-1591.
- Zhang, J. 2000. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *Journal of Molecular Evolution* **50**:56-68.

Table 1. Evolutionary divergences at critical codon positions and conserved protein domains.

Comparisons	dN (SE)	dS (SE)	$\omega = dN/dS$
<i>Critical sites</i> (Codons:107490)			
Human-Macaque	0.005 (0.00024)	0.07 (0.00163)	0.068 (0.0038)
Cow-Pig	0.011 (0.00038)	0.264 (0.00375)	0.042 (0.0016)
Mouse-Rat	0.006 (0.00028)	0.187 (0.00297)	0.032 (0.0016)
<i>Non-critical sites</i> (Codons:1922658)			
Human-Macaque	0.012 (0.00009)	0.062 (0.00035)	0.185 (0.0018)
Cow-Pig	0.037 (0.00017)	0.254 (0.00083)	0.144 (0.0008)
Mouse-Rat	0.022 (0.00013)	0.178 (0.00066)	0.124 (0.0009)
<i>Conserved domains</i> (Codons:2504157)			
Human-Macaque	0.010 (0.00007)	0.065 (0.00032)	0.149 (0.0013)
Cow-Pig	0.029 (0.00013)	0.262 (0.00075)	0.111 (0.0006)
Mouse-Rat	0.017 (0.00010)	0.187 (0.00060)	0.091 (0.0006)
<i>Non-domains</i> (Codons:1829994)			
Human-Macaque	0.015 (0.00011)	0.060 (0.00035)	0.245 (0.0023)
Cow-Pig	0.050 (0.00021)	0.255 (0.00085)	0.198 (0.0011)
Mouse-Rat	0.030 (0.00016)	0.173 (0.00066)	0.175 (0.0011)

Table 2. Divergence estimates at codon positions belonging to different Conservation Indices.

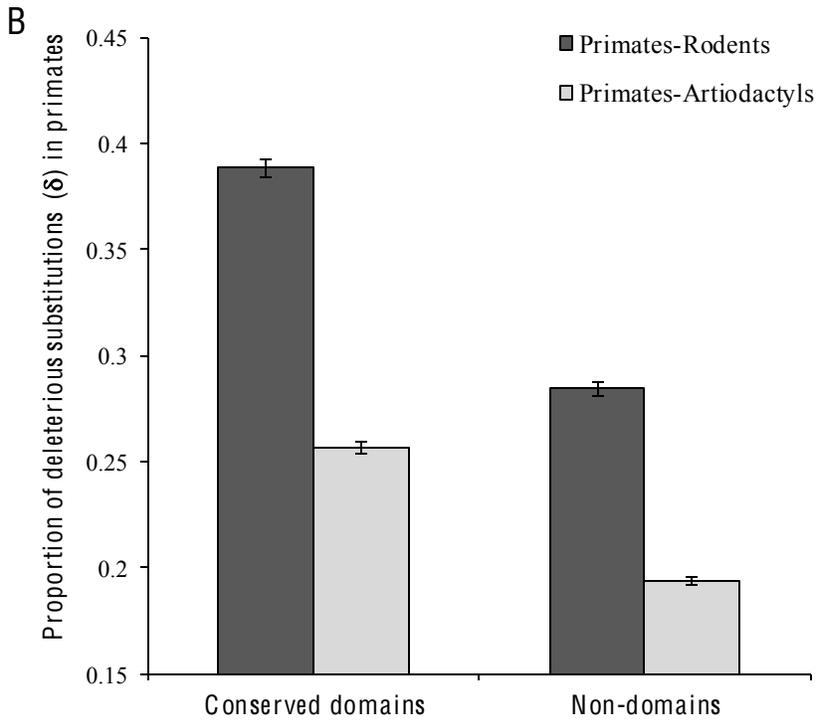
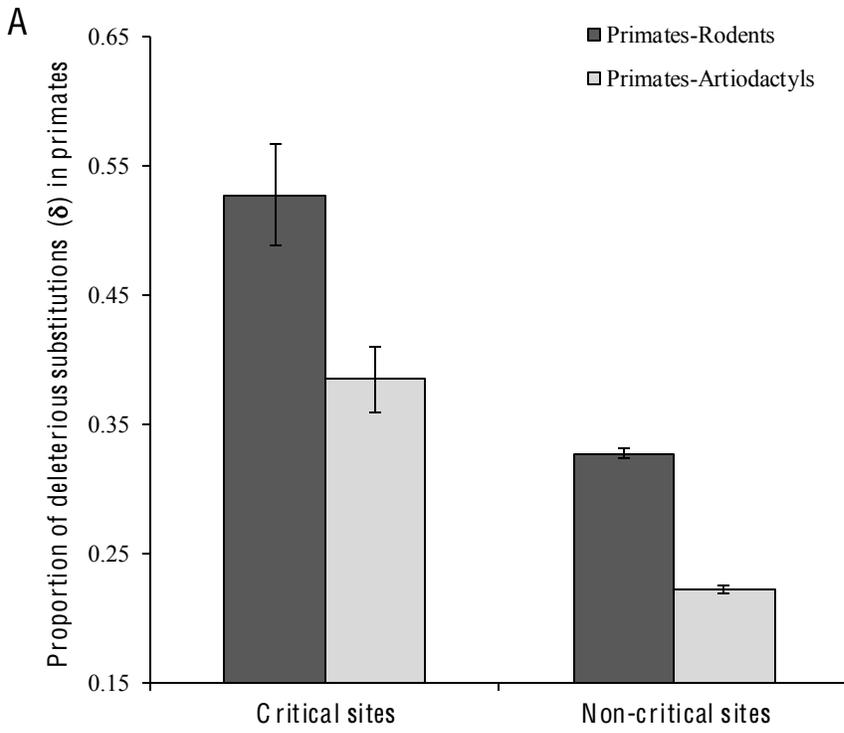
Comparisons	dN (SE)	dS (SE)	$\omega = dN/dS$
<i>Cons. Index 1</i> (Codons: 1443477)			
Human-Macaque	0.004 (0.00006)	0.061 (0.00040)	0.063 (0.0011)
Cow-Pig	0.010 (0.00010)	0.252 (0.00095)	0.041 (0.0004)
Mouse-Rat	0.006 (0.00008)	0.182 (0.00077)	0.036 (0.0005)
<i>Cons. Index 2</i> (Codons: 478005)			
Human-Macaque	0.012 (0.00019)	0.058 (0.00068)	0.211 (0.0041)
Cow-Pig	0.040 (0.00036)	0.257 (0.00172)	0.154 (0.0017)
Mouse-Rat	0.024 (0.00028)	0.180 (0.00135)	0.134 (0.0019)
<i>Cons. Index 3</i> (Codons: 258423)			
Human-Macaque	0.023 (0.00037)	0.063 (0.00096)	0.373 (0.0082)
Cow-Pig	0.080 (0.00071)	0.270 (0.00237)	0.296 (0.0037)
Mouse-Rat	0.049 (0.00055)	0.181 (0.00181)	0.270 (0.0041)
<i>Cons. Index 4</i> (Codons: 115869)			
Human-Macaque	0.032 (0.00065)	0.065 (0.00146)	0.493 (0.0149)
Cow-Pig	0.119 (0.00133)	0.274 (0.00357)	0.434 (0.0074)
Mouse-Rat	0.073 (0.00101)	0.188 (0.00275)	0.387 (0.0078)

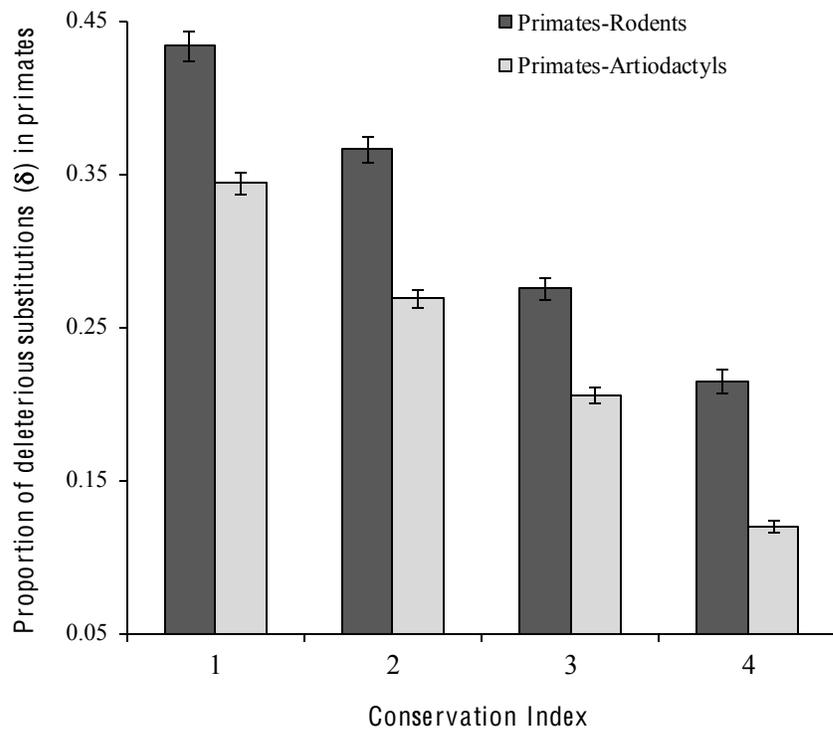
Figure Legends

Figure 1. A. The proportion of deleterious mutations fixed in primates (δ) was estimated using codon positions that are crucial for the stability of the structure and/or proper function of the proteins and using the rest of the codon positions (non-critical sites). **B.** The codons of conserved protein domains and those of non-domains were used to estimate δ in primates. δ was estimated by comparing ω of primates with that of rodents or artiodactyls. Differences in δ estimates between critical sites and non-critical sites as well as between conserved domain sites and non-domain sites are highly significant ($P < 0.0001$) for both primates-rodents and primates-artiodactyls comparisons. Error bars denote the standard error of the mean.

Figure 2. Fraction of deleterious substitutions in codon positions evolving under different rates. Codons of human protein coding genes were separated into four categories based on their relative levels of conservation (Conservation Index). The proportion of deleterious replacement mutations (δ) in each class of sites was estimated. Differences in the δ estimated using codons belonging to any two Conservation Indices are highly significant ($P < 0.0001$). Error bars denote the standard error of the mean.

Figure 3. A. Ratio of radical- to conservative amino acid changes (dR/dC) estimated for the human-macaque, cow-pig and mouse-rat pairs. Differences in the dR/dC ratios are highly significant ($P < 0.0001$). **B.** dR/dC estimated for the above mentioned pairs only using conserved or highly variable amino acid positions. dR/dC ratio estimated using conserved sites is significantly higher for primates than that obtained for rodents ($P < 0.0001$) and artiodactyls ($P = 0.0022$). However ratios obtained using highly variable sites are not significantly different between primates and rodents ($P = 0.36$) or between primates and artiodactyls ($P = 0.36$). Error bars denote the standard error of the mean.





Subramanian

Figure 2

