Positive Darwinian Selection in Gene Evolution

Jianzhi Zhang

Jianzhi Zhang

Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA.

E-mail: jianzhi@umich.edu

About the Author

Jianzhi Zhang is a Professor of Ecology and Evolutionary Biology at the University of Michigan, Ann Arbor, Michigan, USA. He has a wide array of research interests in molecular and genomic evolution, including molecular basis of adaptation, evolution of duplicate genes, genetic basis of human origins, vertebrate sensory gene evolution, and evolutionary systems biology. His researches combine theoretical modeling, empirical data analysis, and experimental molecular biology. He has published over 100 research articles, reviews, and commentaries. Zhang obtained B.S. from Fudan University in 1992 and Ph.D. from Pennsylvania State University in 1998, both in genetics. He served as the Secretary of the Society for Molecular Biology and Evolution from 2007 to 2009, and is currently on the editorial boards of seven journals, including, for example, *PLoS Genetics*, *Genome Biology and Evolution*, and *Gene*.

Representative Articles

[1] Zhang J. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat. Genet.*,2006, 38:819-823.

[2] Wang X, Grus W E, Zhang J. Gene losses during human origins. *PLoS Biol.*, 2006, 4:366-377.

[3] Bakewell M A, Shi P, Zhang J. More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc. Natl. Acad. Sci. USA*, 2007, 104:7489-7494.

[4] Liao B Y, Zhang J. Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc. Natl. Acad. Sci. USA*, 2008, 105:6987-6992.

[5] He X, Qian W, Wang Z, Li Y, Zhang J. Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks. *Nat. Genet.*, 2010, 42: 272-276.

Abstract

When Charles Darwin proposed the theory of evolution by natural selection, he was concerned with phenotypic evolution. Because all phenotypes have their genetic basis, natural selection indirectly acts on genotypes through its action on phenotypes. In this article, I review the roles of positive Darwinian selection in the evolution of genes, focusing on the work conducted in my laboratory in the last decade. Using real examples, I show that different forms of positive selection can promote amino acid substitutions, cause parallel amino acid changes, increase the rate of insertion/deletion substitutions, accelerate gene loss, and enhance gene expression noise. Although positive selection probably does not account for the majority of changes in DNA sequence evolution, it can and did shape DNA sequence evolution in many different ways and is undoubtedly an important force in molecular evolution.

Key Words

Positive selection; molecular evolution

Introduction

Neo-Darwinism, the result of the modern synthesis of Darwin's theory of evolution by natural selection with Mendelian genetics, became the prevailing evolutionary theory in the 1950s. One of the key tenets of neo-Darwinism is that natural selection is the primary force driving evolutionary changes. This tenet was seriously challenged by the neutral theory of molecular evolution (Kimura 1983), the only conceptual revolution in evolutionary biology in the last 50 years (Zhang 2010). The neutral theory asserts that (i) most nucleotide differences between species result from fixations of neutral mutations by random genetic drift and (ii) most intraspecific polymorphisms are also neutral. Although the selectionist-neutralist debate has continued for over 40 years, there is still no agreement among evolutionary biologists about the relative importance of natural selection and genetic drift in molecular evolution. The genomic revolution in the last decade seems to have polarized the opposing views even further. After seeing the genomic data, some are now convinced that most nucleotide substitutions are neutral (Nei 2005), while others believe that most are adaptive and even think that the adaptation should now be used as the null hypothesis in explaining evolutionary observations (Hahn 2008). With the exception of extreme selectionists, most evolutionary geneticists, however, use neutrality as a null hypothesis in explaining evolution; positive selection is invoked only when neutrality is rejected. It is under this framework that much of the study of positive selection at the molecular level has been conducted in the last 20 years. The terms of "positive selection" and "negative selection" are sometimes confusing to non-specialists (Zhang 2008). Positive selection refers to the type of natural selection that promotes the spread of beneficial alleles, whereas negative selection or purifying selection refers to the type of natural selection that prohibits the spread of deleterious alleles. Both selectionists and neutralists agree about the prevalence of negative selection; it is the abundance of positive selection that they disagree about.

Since the first report of positive selection at the molecular level in 1988 (Hughes and Nei 1988), numerous cases have been described. Instead of counting positive selective cases or discussing the relative roles of positive selection and genetic drift in molecular evolution, I will focus on diverse roles of positive selection in gene evolution, by using real examples studied in my laboratory. I choose to use these examples not because they are more illustrious than others in the literature but rather because I am more familiar with them and thus am able to describe them more accurately. The detection of molecular-level positive selection usually requires statistical analysis of genetic data. Many statistical methods have been developed for this purpose, but I will not describe them in detail, because this subject has been reviewed several times in recent years (Nielsen 2005; Anisimova and Kosiol 2009). I will, however, briefly explain each method when it is used in the examples presented. In my view, statistical test of positive selection, while necessary, is not sufficient, for understanding mechanisms of molecular adaptation. It is the functions of the genes and the biology of the organisms that tell us the selective agent and thus the ultimate cause of adaptive evolution.

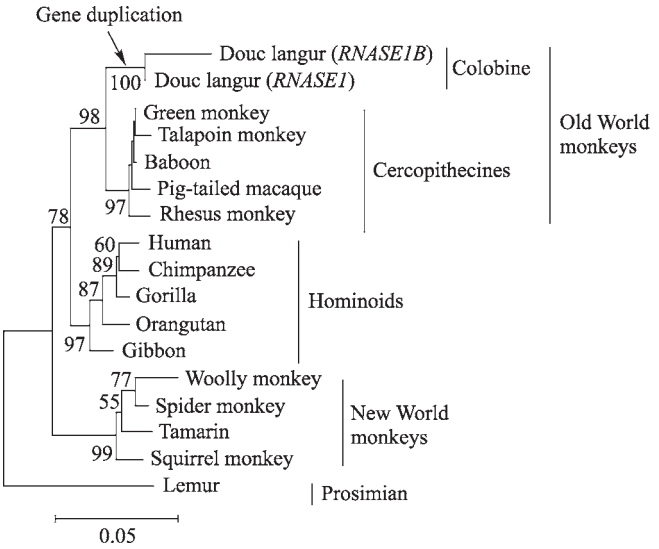
1. Positive Selection Favoring Amino Acid Replacements

Nucleotide substitutions in protein-coding DNA sequences can be divided into synonymous substitutions, which do not affect the encoded amino acids, and nonsynonymous substitutions, which affect the encoded amino acids (Nei and Kumar 2000). Synonymous substitutions are more or less neutral, because they do not affect the protein sequence. Thus, positive selection for nonsynonymous substitutions can be inferred if the rate of nonsynonymous substitution is significantly greater than that of synonymous substitution. Among the types of molecular-level positive selection reported in the literature, the one that favors nonsynonymous substitutions is most abundant. Below I describe one such case that occurred in the evolution of a duplicated pancreatic ribonuclease gene of a leaf-eating colobine monkey (Zhang *et al*. 2002b).

Colobines are a subfamily of Old World (OW)

monkeys that use leaves rather than fruits and insects as their primary food source; these leaves are fermented by symbiotic bacteria in the foregut. Similar to ruminants, colobines recover nutrients by breaking and digesting the bacteria with various enzymes, including pancreatic ribonuclease (RNASE1), which is secreted from the pancreas and transported into the small intestine to degrade RNAs. A substantially greater amount of ribonuclease has been found in the pancreas of foregut fermenting mammals (colobines and ruminants) than in other mammals (Barnard 1969; Beintema 1990). This is believed to be related to the fact that rapidly growing bacteria have the highest ratio of RNA-nitrogen to total nitrogen of all cells, and high concentrations of ribonuclease are needed to break down bacterial RNAs so that nitrogen can be recycled efficiently (Barnard 1969).

In contrast to the presence of only one *RNASE1* gene in each of the 16 non-colobine primate species studied, two *RNASE1* genes were found in the Asian colobine douc langur (*Pygathrix nemaeus*). A phylogentic analysis (Fig. 1) suggests that these two genes were generated by recent duplication postdating the separation of colobines from other OW monkeys. The branch lengths of the gene tree indicate that the nucleotide sequence of one daughter gene (*RNASE1*) has not changed since duplication while that of the other (*RNASE1B*) has accumulated many substitutions. To explore the evolutionary forces driving the accelerated evolution of *RNASE1B*, we compared the number of nucleotide substitutions per site at nonsynonymous sites in *RNASE1B* since its origin through gene duplication, and the corresponding number at synonymous and flanking non-coding sites. Based on Fisher's exact test (Zhang *et al*. 1997), we found that the former (0.0310) is significantly greater than the latter (0.0077; $P<0.002$). To investigate the nature of the amino acid substitutions favored by selection, we divided nonsynonymous substitutions into two groups: those altering the amino acid charge (radical substitutions) and those that leave charge unaltered (conservative substitutions). Earlier studies showed that, for



▲ **Fig. 1**

Phylogenetic relationships of *RNASE1* and *RNASE1B* genes of primates. Douc langur has both genes, whereas other species only have the *RNASE1* gene. Bootstrap percentages higher than 50 are shown on tree branches. Branch lengths are drawn to scale, indicating the number of nucleotide substitutions per site. This neighbor-joining tree was reconstructed with Kimura's two-parameter distances. Reprinted, with permission, from Zhang *et al*. (2002b).

most mammalian genes, the rate of radical substitution is lower than that of conservative substitution due to stronger purifying selection on the former (Zhang 2000). In *RNASE1B*, however, the opposite is found. The number of radical substitutions per site (0.067) since duplication is significantly greater than that (0.012) of conservative substitutions per site ($P<0.02$; Fisher's test). There are 9 amino acid substitutions in the mature peptide of RNASE1B, and 7 of them involve charge changes. Surprisingly, all these 7 charge-altering substitutions increase the negative charge of the protein. Apparently, the amino acid substitutions are nonrandom ($P<0.016$; randomization test), with negative charged residues being selectively favored.

The charge-altering substitutions drastically reduced the net charge of RNASE1B from 8.8 to 0.8 (at pH 7) and the isoelectric point from 9.1 to 7.3. Because RNA is negatively charged, the net charge of ribonuclease influences its interaction with the substrate and its catalytic performance (Sorrentino and Libonati 1997). We thus hypothesized that the charge-altering substitutions may have changed the optimal pH of RNASE1B in catalyzing the digestion of RNA. To test this hypothesis, we prepared recombinant proteins from the douc langur *RNASE1B* gene as well as the *RNASE1* genes of the human, rhesus monkey, and douc langur, and examined their ribonucleolytic activities at different pHs in a standard ribonuclease assay against yeast tRNA. We determined that the optimal pH for human RNASE1 is 7.4, a value that is within the pH range (7.4-8.0) measured in the small intestine of humans (Code 1968; Guyton and Hall 1996). The same optimal pH is observed for the RNASE1s of rhesus monkey and douc langur. Probably because of foregut fermentation and related changes in digestive physiology, the pH in the small intestine of colobine monkeys shifts to 6-7 (Kay and Davies 1994). Interestingly, the optimal pH for the douc langur RNASE1B is found to be 6.3. At pH 6.3, RNASE1B is about 6 times as active as RNASE1 in digesting RNA, and the difference in their activities is statistically significant ($P<0.001$, *t* test). These results suggest that the rapid amino acid substitutions in RNASE1B were driven by selection for enhanced ribonuclease activity at the relatively low pH environment of the colobine small intestine. Further studies suggested that the progenitor gene of the duplicates processed a second function in degrading double stranded RNA (dsRNA) and that this second function has been retained in RNASE1 but lost in RNASE1B after duplication. Thus, the conservation of colobine RNASE1 after duplication is likely due to the selective pressure to keep this second function. It is interesting to note that very similar patterns of gene duplication and functional divergence of RNASE1 also occurred independently in African colobines (Zhang 2006). Together, the studies of digestive ribonucleases of colobines demonstrate the important role of positive selection in the functional divergence of duplicate genes and in organismal adaptation to changing environments.

2. Positive Selection for Parallel Amino Acid Substitutions

Common selective pressures occurring in different evolutionary lineages may result in convergent or parallel changes, such as the independent acquisitions of wings by birds and bats. While convergent/parallel evolution is common at the morphological level, it is relatively rare at the protein sequence level and is thus considered to be a strong indication of molecular adaptation (Zhang and Kumar 1997). Here, a convergent change at an amino acid site refers to changes from different ancestral amino acids to the same descendant amino acid along independent evolutionary lineages. It is distinguished from a parallel change, in which amino acid changes along independent lineages have occurred from the same ancestral amino acid. In the above example of RNASE1 evolution, it was found that three parallel amino acid changes, which significantly exceeds the chance expectation, occurred in Asian and African colobine moneys after the independent duplications of *RNASE1* (Zhang 2006). Further, site-directed mutagenesis studies showed that these parallel changes are required for the shifting of the optimal pH of the ribonucleases (Zhang 2006). Thus, these parallel amino acid changes are likely the result of a common positive selection in Asian and African colobines

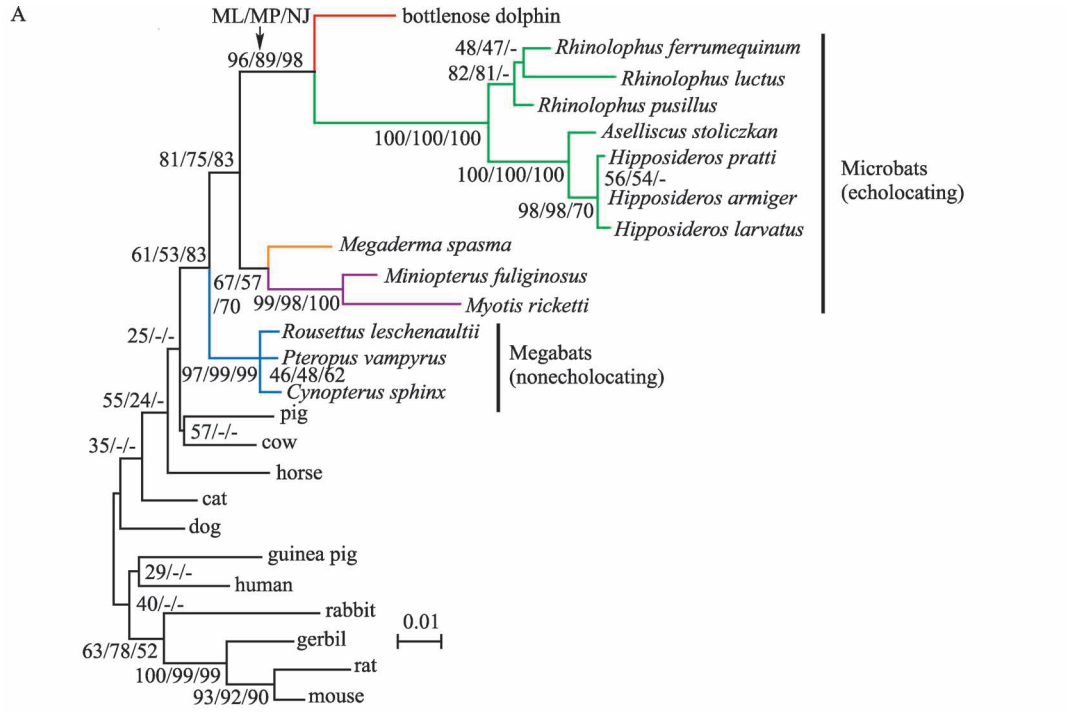
for higher catalytic efficiencies of digestive ribonucleases. Below I describe a case of more dramatic parallel protein sequence evolution that occurred in the hearing gene *Prestin* of echolocating bats and whales (Li *et al*. 2010).

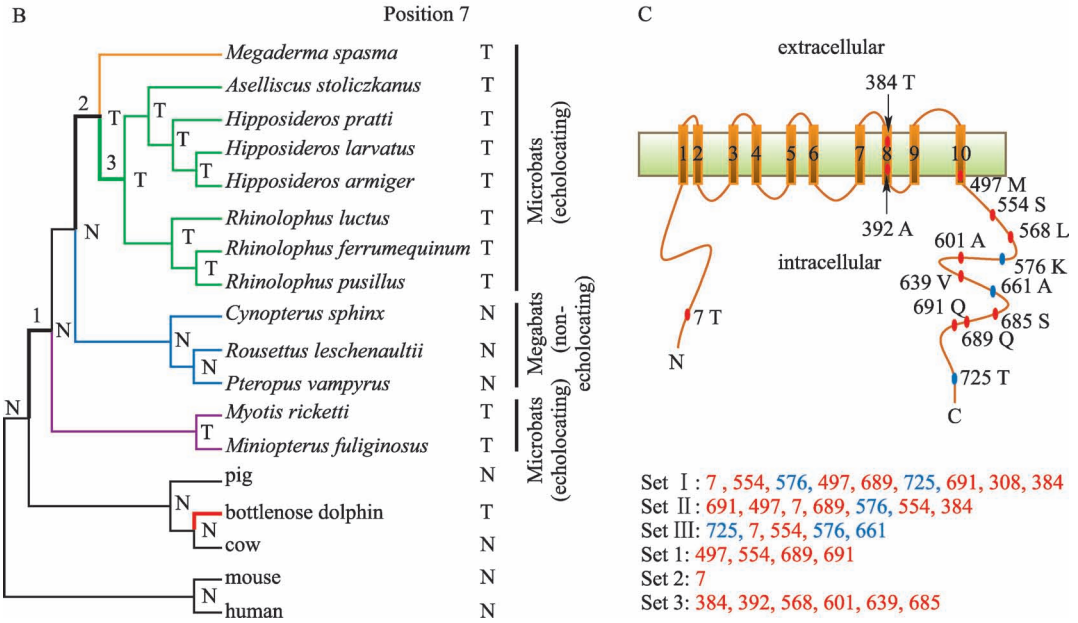
Echolocation, or biosonar, is a sensory mechanism to locate, range, and identify objects in an environment by emitting calls to the environment and listening to the echoes that return from the objects (Jones 2005). Only microbats and toothed whales have acquired sophisticated echolocation, which is indispensable for their orientation and foraging (Jones 2005). Mammalian prestin is a member of the SLC26 anion-transport family found primarily on the membrane of cochlear outer hair cells (OHCs). Prestin provides the electromobility of OHCs that is thought to be responsible for cochlear amplification, an active process that confers sensitivity and frequency selectivity to the mammalian auditory system (Dallos 2008). To examine the potential role of *Prestin* in echolocation, we reconstructed the prestin tree using its protein sequences from 25 mammals, including 10 (echolocating) microbats, three (nonecholocating) megabats, one toothed whale (bottlenose dolphin), and 11 other mammals. Surprisingly, the dolphin is placed within bats (specifically with the green-labeled families of Rhinolophidae and Hipposideridae in Fig. 2A), rather than with cow, its true closest relative represented in our data, and this unexpected grouping has a high bootstrap support in maximum likelihood (ML), maximum parsimony (MP), and neighbor-joining (NJ) trees (Fig. 2A). Furthermore, unlike the species tree where microbats are paraphyletic (Jones and Teeling 2006) (Fig. 2B), the prestin tree clusters the 10 microbats in exclusion of the three megabats with a moderate bootstrap support (Li *et al*. 2008), resulting in the misplacement of two purple-labeled microbats (Fig. 2AB). None of the other differences between the prestin tree and the species tree are statistically supported (Fig. 2A).

After excluding the possibilities of horizontal gene transfer, DNA contamination, gene paralogy, long-branch attraction, and biased amino acid frequencies, we established that the misplacement of dolphin in the prestin tree is most likely owing to the convergence of the prestin sequences of echolocating bats and whales, which probably resulted from a common selection for amino-acid-altering mutations that are beneficial to echolocation. Indeed, when only synonymous nucleotide substitutions are used, dolphin and cow are correctly grouped with 100% bootstrap support.

Two approaches were used to identify the amino acid sites causing the clustering of dolphin and bat prestins. To avoid confounding factors, we analyzed a subset of 18 species that includes only cetartiodactyls and bats, with human and mouse as outgroups (Fig. 2B). A minimum of 9 amino acid sites (referred to as set I in Fig. 2C) need to be removed to make the likelihood of the species tree higher than that of the prestin tree. Because the prestin tree misplaces both dolphin and two purple-labeled microbats (Fig. 2A), we further examined whether the two problems are caused by the same sites. By comparing the prestin tree without the two misplaced microbats and the corresponding species tree, we found that a minimum of 7 sites (referred to as set II in Fig. 2C) need to be removed to correct the phylogenetic position of dolphin. Similarly, we identified a minimum of 5 sites (referred to as set III in Fig. 2C) that need to be removed to rectify the positions of the two microbats. Sets II and III share 3 sites, significantly exceeding the random expectation ($P<0.002$, binomial test under the consideration that only variable sites may be included in any above set), suggesting that the misplacements of the dolphin and the two microbats are caused in part by the same sites.

In the second approach, we identified convergent/parallel amino acid substitutions that occurred between the dolphin branch (red in Fig. 2B) and any of the three bat branches marked 1, 2, and 3 in the established species tree in Fig. 2B, by comparing the inferred ancestral prestin sequences in all interior nodes of the species tree and the extant sequences. We considered these three bat branches because prestin function associated with echolocation in bats (especially Rhinolophidae and Hipposideridae) likely have





emerged in one or more of these branches. We identified no convergent site, but 4, 1, and 6 parallel sites between the red branch and branches 1, 2, and 3, respectively, and named these three sets of sites as set 1, 2, and 3, respectively (Fig. 2C). A statistical test (Zhang and Kumar 1997) shows that sets 1, 2, and 3 contain significantly more parallel sites than their respective random expectations ($P < 10^{-8}$, <0.0045, and $< 10^{-8}$, respectively), suggesting that a common selection, rather than chance, underlies the observed parallel substitutions. The total number of parallel substitutions observed here is the largest in all proteins reported to have undergone parallel sequence evolution (Zhang and Kumar 1997; Zhang 2006; Castoe *et al*. 2009). Six of the 7 sites in set II, which cause the misplacement of dolphin, experienced parallel changes (Fig. 2C), supporting our hypothesis that parallel amino acid substitutions is the reason for the grouping of dolphin and bats in the prestin tree.

After mapping the sites of sets I-III and sets 1-3 onto the structural model of prestin (Navaratnam *et al*. 2005) (Fig. 2C), we observed that all except three sites fall in the intracellular terminal regions, including one site in the N-terminus and 10 in the C-terminus, a pattern that is highly nonrandom (P<0.005, Fisher's exact test). Previous mutagenesis studies demonstrated that both N- and C-termini are used for voltage sensing (Navaratnam *et al*. 2005) and the N-terminus is also critical for homo-oligomerization of prestin (Navaratnam *et al*. 2005), which may influence the speed of conformational changes of prestin that is likely crucial for high-frequency acoustic sensitivities of echolocation. Although not all parallel changes identified may be necessary for echolocation, as some have reverted in some microbats or occurred also in some nonecholocating mammals, they are strong candidates for future experimental investigation. Of particular interest is position 7, which appears in sets I-III and experienced a parallel change from Asn to Thr in three branches (Fig. 2B). This site has a Thr in all echolocating mammals but an Asn in all non-echolocating mammals examined. We were able to amplify exon 1 of *Prestin* from the bowhead whale *Balaena mysticetus*, a nonecholocating whale, and determined that it has an Asn at position 7. Thus, the multiple Asn to Thr changes at position 7 were likely important for the multiple origins of echolocation.

Bats and whales vary greatly in echolocation (Au 2004). For example, bats use echolocation for ranges up to 3-4 meters, whereas whales use for ranges up to >100 meters (Au 2004). More importantly, the speed of sound in air is about one fifth that in water, making the information transfer during sonar transmission much slower for bats than for whales (Au 2004). Despite these gross differences, our findings suggest that the high-frequency acoustic sensitivities and selectivities of bat and whale echolocation appear to rely on a common molecular design of prestin. Because prestin function can be studied in knock-in mice and in cell lines (Dallos 2008), a functional analysis of the parallel amino acid substitutions identified here could shed light

◄ **Fig. 2**

Parallel evolution of prestins of echolocating bats and bottlenose dolphin. A. The maximum-likelihood (ML) tree reconstructed using the prestin protein sequences of 25 mammals, under the model of JTT-f with a gamma distribution of substitution rate variation among sites. Numbers on interior branches are bootstrap percentages from ML, maximum-parsimony (MP), and neighbor-joining (NJ) analyses. The bootstrap value from an analysis is indicted as "-" when the branch does not exist in that analysis. The scale bar shows 0.01 amino acid substitution per amino acid site. **B**. The species tree of 18 mammals. Tree branches are not drawn to scale. Parallel substitutions were examined between the red branch and the branches labeled 1, 2, and 3. The amino acid (N: Asn; T: Thr) at position 7 of prestin is shown for each interior and exterior node. **C**. Locations of evolutionarily interesting amino acid sites in the structural model of prestin with 10 transmembrane domains. Numbers associated with colored circles are the amino acid positions in the dolphin prestin sequence, with the residues observed in dolphin indicated. Sets I, II, and III are the most important sites responsible for the misplacement of both dolphin and the two purple-labeled microbats in panel A, misplacement of dolphin, and misplacement of the two microbats, respectively, and are listed in the order of their support of the gene tree relative to the species tree (from high to low). Sets 1, 2, and 3 are sites that have experienced parallel amino acid substitutions between the dolphin branch (red in panel B) and branches 1, 2, and 3, respectively. Parallel-evolution sites are shown in red, while the other sites are shown in blue. Reprinted, with permission, from Li *et al.* (2010).

on the structure-function relationship of prestin and the molecular underpinnings of the acoustic adaptations in echolocation. It could also help answer why the prestin of bottlenose dolphin is particularly similar to that of Rhinolophidae and Hipposideridae bats (Fig. 2A).

3. Positive Selection for Indels in Protein Sequences

While positive selection promoting amino acid replacements has been reported in many proteins, positive selection for insertions/deletions (indels) in protein sequences is rare. This is probably because many indel mutations disrupt the reading frame of a gene and thus are subject to strong negative selection, which makes the detection of positive selection difficult. Here, I describe a case of positive selection for indel substitutions that occurred in the evolution of the primate CATSPER1 gene (Podlaha and Zhang 2003).

CATSPER1 is a voltage-gated calcium ion channel that is exclusively found in the plasma membrane of the principal piece of the sperm tail (Ren *et al*. 2001). It is necessary for cAMP-induced Ca^{2+} influx, normal sperm motility, and penetration of the egg (Ren *et al*. 2001). Targeted disruption of the gene results in sperm immobility and male infertility in mice (Ren *et al*. 2001). The CATSPER1 protein contains an intracellular N-terminus region, 6 transmembrane domains, a pore-forming domain, and an intracellular C-terminus. In an alignment of the putative orthologous CATSPER1 sequences from the human and mouse, we noticed that the N-terminus region (mostly encoded by exon 1) contains multiple gaps, while the rest of the sequences are conserved. Such a high frequency of gaps is unusual for orthologous mammalian proteins, which prompted us to examine this region in detail. We determined the exon 1 sequence of CATSPER1 from 15 non-human primates representing major primate groups (common chimpanzee *Pan troglodytes*, pygmy chimpanzee *Pan paniscus*, gorilla *Gorilla gorilla*, orangutan *Pongo pygmaeus*, talapoin monkey *Miopithecus talapoin*, rhesus monkey *Macaca mulatta*, baboon *Papio hamadryas*, African green monkey *Cercopithecus aethiops*, colobus monkey *Colobus guereza*, woolly monkey *Lagothrix lagotricha*, owl monkey *Aotus trivirgatus*, squirrel monkey *Saimiri sciureus*, spider monkey *Ateles geoffroyi*, cotton-top tamarin *Saguinus oedipus*, and ring-tailed lemur *Lemur catta*), and compared them with the sequence from the human. As expected, all sequences have an open reading frame throughout the exon. But the sequence length of the exon varies among species, from 360 codons in the lemur to 443 codons in the orangutan. These sequences were conceptually translated and aligned by CLUSTAL X with the default parameters, and the DNA sequences were subsequently aligned following the protein alignment. A gene tree of the 16 sequences was reconstructed using the neighbor-joining method, which shows branching patterns that are largely consistent with the known species tree, indicating that the sequences obtained are orthologous to each other. Using the parsimony principle, we inferred events of indel substitutions in CATSPER1 exon 1 and mapped them onto the species tree. Note that parsimony makes our inference of the total number of indels conservative. Multiple parsimonious solutions are weighted equally. A total of 31 indel substitutions were found throughout the tree. To investigate the robustness of this result, we used a wide variety of penalty parameters in alignment. The resulting number of indels for the entire tree varied from 26 to 34. But, by our judgment, the alignment with 31 indels, which was obtained using the default parameters, appears most reasonable, and further analysis is based on this alignment. However, our conclusion is valid even when alignments with fewer indels are used.

To test whether the rate of indel substitutions in CATSPER1 exon 1 is significantly higher than the neutral expectation, it is necessary to first estimate the neutral rate of indel substitutions. For this, we used a recently published human-chimpanzee genomic comparison by Britten (Britten 2002). In this comparison, 1019 indels were found in an alignment of 779,142 nucleotides. Because only 1.1%-1.4% of the human genome contains protein-coding sequences, this alignment is largely comprised of noncoding sequences and may thus be regarded as neutrally

evolving regions. Because we will compare the indel rates between noncoding and coding regions, it is more relevant to compute the neutral substitution rate of indels with sizes of multiples of 3 nucleotides (3*n* indels), as only 3*n* indels are potentially non-deleterious when they occur in protein-coding regions. From the above human-chimpanzee genomic data, we estimated that the neutral substitution rate for 3*n* indels is $(1.92\pm0.14)\times10^{-11}$ per site per year, based on the assumption that the human and chimpanzee diverged 6.5 MY ago. In addition to the use of the Britten data, we also used the result from Silva and Kondrashov, who conducted a genomic comparison between human and baboon for 1,448,332 nucleotides (Silva and Kondrashov 2002). They identified 5883 indels, of which 1001 were 3*n* indels. Assuming that humans and baboons diverged 23 MY ago, we estimated from their data that the neutral substitution rate for 3*n* indels is $(1.50\pm0.05)\times10^{-11}$ per site per year.

With these estimates of genomic neutral substitution rates for 3*n* indels, we computed the expected number of 3*n* indels in CATSPER1 exon 1 under the assumption that all 3*n* indels are neutral. Using the estimate from the human-chimpanzee genomic comparison, the expected number of 3*n* indels in exon 1 between hominoids and Old World (OW) monkeys is 1.17 (Table 1). The observed average number of 3*n* indels in exon 1 between the 5 hominoids and 5 OW monkeys is 6, which is significantly greater than the expected value of 1.17 under neutral evolution ($P<0.001$, Poisson test, Table 1). Similarly, the comparison between hominoids and New World (NW) monkeys and that between OW and NW monkeys yielded the same conclusion (Table 1). Use of the neutral rate estimated from the human-baboon genomic data shows even stronger statistical significance (Table 1). Use of the alignment with the smallest number (i.e., 26) of gaps gave similar results. These comparisons strongly suggest that 3*n* indels are positively selected for in the evolution of primate CATSPER1 exon 1. In the above tests, we assumed that the rate of indel mutations at the CATSPER1 locus is similar to the genomic average. We verified this assumption by showing that the indel substitution rate in intron 1 of CASPTER1 is virtually identical to the rates estimated from the two genomic data sets.

We further investigated whether indels of certain lengths are particularly favored in CATSPER1 by comparing the (3*n*) indel-size distributions for the CATSPER1 data and the two genomic data sets used above. A significant distributional difference is detected between CATSPER1 and

Table 1 Comparison of the substitution rates of 3*n* indels from genomic date and from primate CATSPER1 exon 1 (adapted from Podlaha and Zhang 2003)

Comparisons	Divergence [a] (MY)	Indel rate (per site per 10^{11} year)			Number of indels in CATSPER1 exon 1			Probability [d]	
		Genomic data 1 [b]	Genomic data 2 [c]	From CATSPER 1	Expectation from data 1 [b]	Expectation from data 2 [c]	Observation	under date 1 [b]	under date 2 [c]
Hominoids vs. Old World monkeys	23×2	1.92	1.50	9.81	1.17	0.92	6	1.3×10^{-3}	3.9×10^{-4}
Hominoids vs. New World monkeys	35×2	1.92	1.50	10.3	1.79	1.40	9.6	1.1×10^{-4}	1.6×10^{-5}
Old World vs. New World monkeys	35×2	1.92	1.50	12.5	1.79	1.40	11.6	3.0×10^{-6}	2.8×10^{-7}

[a] The divergence times follow Glazko and Nei (2003), *Mol. Biol. Evol.*, 20:424-434.
[b] Britten (2002), *Proc. Natl. Acad. Sci. USA*, 99: 13633-13635.
[c] Silva and Kondrashov (2002), *Trends Genet.*, 18:544-547.
[d] The probabilities of the observation given the expectation are computed under the assumption that the number of indels follows a Poisson distribution.

either of the two genomic data sets ($P<10^{-19}$, χ^2 test). Longer indels are preferentially selected for in CATSPER1. For instance, the proportion of 3*n* indels with 15 or more nucleotides is 8-9% in the two genomic data, but 58% in the CATSPER1 data. However, even for indels of 3 nucleotides, the number of observed indels in CATSPER1 is about 2.5 times the expected number from the genomic data, and their difference is statistically significant ($P<0.02$). This suggests that both short and long indels are selectively favored in CATSPER1, with longer ones being under stronger positive selection.

Our observations provide strong evidence that indel substitutions, particularly those with greater sizes, are positively selected for in the evolution of primate CATSPER1. Why would indel substitutions be beneficial in CATSPER1 exon 1, which encodes the intracellular N-terminus region of the ion channel? To address this question, we turn to the structure and function of ion channels. Ion channels are transmembrane proteins that form pores through which ions can pass. A voltage-gated ion channel such as CATSPER1 is activated by depolarization (reduction in electric potential) of the cell membrane, which causes a conformational change of the channel and allows ions to pass through it. Within 1 millisecond of activation, the channel is inactivated and is impermeable to the ions, even though the membrane is still depolarized. The membrane must be repolarized or hyperpolarized to remove the channel from the inactive state and return it to the closed state where it is prepared for subsequent activation. Inactivation prevents the channel from remaining open, and is also responsible for the unidirectional propagation of action potential. The "ball-and-chain" model of ion channel inactivation, proposed by Bezanilla and Armstrong (Bezanilla and Armstrong 1977) and demonstrated by Aldrich and colleagues (Hoshi *et al*. 1990; Zagotta *et al*. 1990), offers a possible scenario where the length of the N-terminus region plays an important functional role. Specifically, Aldrich and colleagues showed that the N-terminus of a *Drosophila* voltage-gated potassium (K_V) channel named *Shaker* acts to inactivate the channel (Hoshi *et al*. 1990; Zagotta *et al*. 1990). Here, a "ball on a chain" structure is located at the N-terminus of the channel and acts as a tethered plug, which is able to physically block the intracellular end of the ion channel pore region and cause inactivation of the channel (Hoshi *et al*. 1990; Zagotta *et al*. 1990). The first ~20 residues of the N-terminus of *Shaker* channel form the intracellular "plug" and the next ~60 residues represent the "tether" (Hoshi *et al*. 1990; Zagotta *et al*. 1990). It was found that the length of this tethered plug controls the rate of channel inactivation (Hoshi *et al*. 1990). That is, lengthening or shortening of the tether resulted in slow or rapid channel inactivation, respectively. This is probably because a shorter tether restricts the space in which the "plug" wanders, making it easier for the "plug" to find the pore. Although this "ball-and-chain" model has only been demonstrated in K_V channels, it is possible that CATSPER1 has a similar mechanism of regulating its inactivation. In fact, structurally, CATSPER1 resembles K_V channels more than voltage-gated Ca (Ca_V) or Na (Na_V) channels, as CATSPER1 and K_V channels are each formed by 4 identical peptides, each having a single, 6-transmembrane-spanning repeat, whilst Ca_V and Na_V channels are made of a single peptide with 4 repeats of 6-transmembrane-spaning regions. The amino acid sequence of the pore-forming region, however, is more similar between CATSPER1 and Ca_V, presumably reflecting the identical ion selectivity. The hydropathy profile shows a greater similarity of CATSPER1 to K_V than to Ca_V or Na_V channels. If the "ball-and-chain" model indeed works in CATSPER1, the indels in the N-terminus region can potentially affect the inactivation rate of the channel, as in the *Drosophila* K_V channel *Shaker*. Because CATSPER1 determines sperm motility by regulating the cellular Ca^{2+} concentration (Ren *et al*. 2001), it is likely that the rate of channel inactivation influences sperm motility. As sperm motility is one of the most important factors in sperm competition, the exceptionally high rate of indel substitutions in CATSPER1 may be a signature of intense sperm competition.

To our knowledge, CATSPER1 represents the first case in which positive selection for indel substitutions is detected. A subsequent study revealed a similar evolutionary pattern of rodent

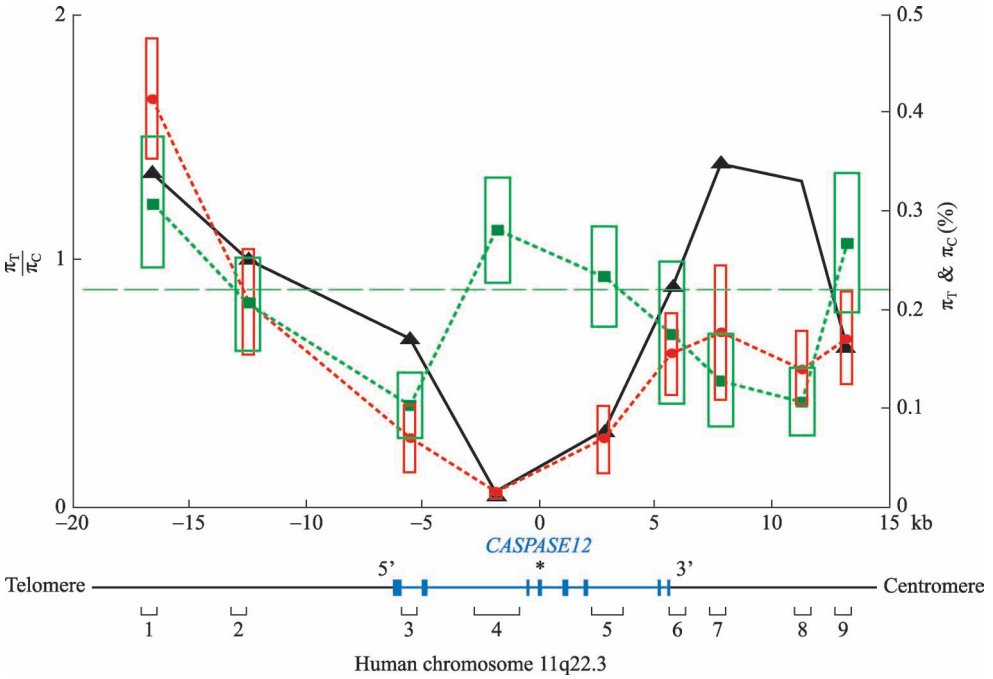
Catsper1 (Podlaha *et al*. 2005). These successes largely rely on the availability of genomic sequence data from closely related species, from which a neutral rate of indel substitution can be reliably estimated. From the present study, it seems that the statistical detection of selection on indels is relatively powerful. For instance, the number of expected 3*n* indels is about 1 for the CATSPER1 exon 1 sequences between hominoids and OW monkeys (Table 1). Under the assumption that the number of indels follows a Poisson distribution, an observation of 4 indels would lead to the rejection of the null hypothesis of neutral evolution at the 2% significance level. Because protein-length variation among orthologs and paralogs is quite common and indels are often seen in protein sequence alignments, we hypothesize that positive selection for indels is not rare. With the establishment of the basic methodology here and the estimation of neutral rates of indel substitutions from many more species, this hypothesis can be tested in the future.

4. Positive Selection for Null Alleles

While molecular-level positive selection mostly acts on functional genes, it may also act on pseudogenes if the inactivation of a gene becomes advantageous under certain environmental or genetic backgrounds. Below I describe such a case that occurred in the loss of the human CASPASE12 (CASP12) gene (Wang *et al*. 2006).

CASP12 belongs to the caspase family, which are *c*ysteinyl *asp*artate protein*ases* that play important roles in the processing of inflammatory cytokines and the initiation and execution of apoptosis (Alnemri *et al*. 1996; Lamkanfi *et al*. 2002). Human *CASP12* was identified as a pseudogene following the cloning of mouse *Casp12* (Fischer *et al*. 2002). Compared with other mammalian orthologs, human *CASP12* contains a premature stop codon due to a C to T nonsense mutation at nucleotide position 629 of exon 4 (Fischer *et al*. 2002; Saleh *et al*. 2004). This mutation leads to the production of truncated nonfunctional CASP12 in humans (Saleh *et al*. 2004). The null T allele is fixed in a sample of 347 non-Africans and has a frequency of 89% in 776 individuals of African descent (Saleh *et al*. 2004). Interestingly, the T allele is associated with a reduced incidence and mortality of severe sepsis (Saleh *et al*. 2004), suggesting that the loss of functional CASP12 is beneficial to present-day humans. To test whether the nearly complete fixation of the null allele at *CASP12* has been driven by positive selection, we looked for signals of recent (incomplete) selective sweeps by examining the intraspecific variation of putatively neutral regions surrounding the C/T polymorphism. The positive selection hypothesis predicts that the level of polymorphism in these regions is lower in the T allele than in the C allele, especially in the proximity of the C/T polymorphism, due to the hitchhiking effect (Maynard Smith and Haigh 1974). Furthermore, the frequency distribution of the neutral polymorphisms in the T allele should deviate from the neutral expectation, generating negative values of Tajima's *D* (Tajima 1989) and Fay and Wu's *H* (Fay and Wu 2000).

From a sample of 63 humans of African descent, we identified 4 C/C homozygotes and 43 T/T homozygotes. We sequenced the 4 C/C homozygotes and 4 randomly chosen T/T homozygotes in 9 noncoding regions of varying distances from the C/T polymorphism (Fig. 3). The sequenced regions vary in size from ~600 to 2,400 nucleotides. In total, 53 and 29 single nucleotide polymorphisms (SNPs) were identified from 8,925 nucleotide sites in C/C and T/T individuals, respectively. Although the T allele is much more prevalent than the C allele in the population, the T allele has a significantly lower number of SNPs per nucleotide than the C allele in the linked regions ($P<0.01$, Fisher's exact test). Nucleotide diversity per site (π) is also lower in the T alleles ($\pi_T = 0.00131 \pm 0.00019$) than in the C alleles ($\pi_C = 0.00218 \pm 0.00031$) ($P = 0.02$, two-tail Z test). More strikingly, although the variation of π_C across the 9 regions is more or less random, that of π_T exhibits a V shape, with the bottom of the valley located in region 4, which has its 3' end only 607 nucleotides from the C/T polymorphism (Fig. 3). When one moves ~10,000 nucleotides from this polymorphism, π_T rises to a level comparable to π_C. By sequencing 7 additional T/T individuals of African descent in regions 4, 5, and 6, we confirmed that the low



▲ **Fig. 3**

Intraspecific DNA sequence variation in noncoding regions linked with the human *CASPASE12* gene. *CASPASE12* is shown in blue, with the exons depicted by solid blue bars on the chromosome. The premature stop codon generated by the C→T nonsense mutation is shown by an asterisk in exon 4. The 9 noncoding regions sequenced are indicated below the chromosome. Exons, introns, the 9 noncoding regions, and spaces between regions are drawn to scale as indicated. Red circles (connected by the red dotted line) show nucleotide diversity per site among African T alleles (π_T) and the red boxes shows π_T ± one standard error of π_T. Green squares (connected by the green dotted line) show nucleotide diversity per site among African C alleles (π_C) and the green boxes shows π_C ± one standard error of π_C. The broken green line shows the mean π_C across the 9 noncoding regions sequenced. Black triangles (connected by the black solid line) show the ratio between π_T and π_C for each region. π_C is estimated from 8 alleles. π_T is estimated from 22 alleles for regions 4, 5, and 6, and from 8 alleles for the other regions. When only 8 alleles are used, π_T is 0.00018±0.00007, 0.00129±0.00071, and 0.00145±0.00057 for regions 4, 5, and 6, respectively. π_T is significantly lower than π_C in regions 4 and 5. Reprinted, with permission, from Wang *et al.* (2006).

π_T is not due to the small sample size. In region 4, where the greatest reduction in polymorphism is observed, only 1 SNP is found across the 2413 nucleotide positions among the 22 T alleles sequenced. By contrast, 19 SNPs were found in the same region among 8 C alleles examined. Region 4 was also sequenced in 6 non-Africans (all non-Africans are T/T homozygotes (Saleh *et al.* 2004)), but no SNP was detected and all non-African T alleles are identical to the predominant T allele from Africans. This indicates a common origin of African and non-African T alleles.

We conducted a formal test of the selective sweep hypothesis by using coalescent simulations. Our results showed that the high prevalence yet low polymorphism of the T allele cannot be explained by various demographic models considered. We also computed statistics D and H for regions 4 and 5 in the T allele, as these two regions have significantly lower π_T than π_C. Both statistics were significantly negative in region 5 ($D = -2.08$, $P<0.01$; $H = -4.71$, $P<0.025$), consistent with the expectations from a selective sweep. D (– 0.23, $P = 0.47$) and H (– 0.90, $P = 0.09$) were not significantly negative in region 4, probably because the number of SNPs is too small for the statistic tests to be powerful. It should be noted that the above tests are less rigorous than the coalescent simulations

because the tests are conducted on subsets of the genealogy (Evans *et al*. 2005). Taken together, our observations, especially the proximity of the π_T valley to the C/T polymorphism and the coalescent simulations, strongly suggest that the spread of the T allele among Africans and non-Africans has been driven by positive selection and that the selective advantage was directly conferred by the C→T nonsense mutation.

When did the pseudogenization of human *CASP12* start? We estimated that the pseudogenization started ~74 thousand years ago and that the 95% confidence interval of the time required for the T allele to reach today's frequency is 51 to 55 thousand years. Despite the potentially large errors, the two estimates were close, suggesting that the T allele might have been beneficial since its appearance. Because the T alleles of Africans and non-Africans share the same origin, the C→T nonsense mutation must predate the out-of-Africa migration of modern humans, which is believed to have occurred 40-60 thousand years ago (Cavalli-Sforza and Feldman 2003). Our dating suggests that the pseudogenization of *CASP12* began not long before this migration.

Our population genetic study provided strong evidence that the nearly complete fixation of a null allele at human *CASP12* has been driven by positive selection, possibly because it confers resistance to severe sepsis. *CASP12* is a functional gene in all mammals surveyed except humans (Saleh *et al*. 2004), suggesting that it is indispensable in a typical mammal. The functional human CASP12 acts as a dominant-negative regulator of essential cellular responses including the NF-κB and IL-1 pathways; it attenuates the inflammatory and innate immune response to endotoxins (Saleh *et al*. 2004). Because an appropriate level of immune response that is neither excessive nor insufficient is important to an organism, one can imagine that the immune suppression function of CASP12 becomes harmful when the immune system cannot fully respond to a challenge. It is likely that during human evolution alterations in our genetic and/or environmental background resulted in a malfunction of the immune response to endotoxins, which rendered the previously necessary function of CASP12 deleterious in humans and the null allele advantageous over the functional one. Identification of such genetic and/or environmental alterations will be valuable for understating human-specific immune functions.

Olson proposed in the "less-is-more" hypothesis that gene loss can sometimes play an active role in evolution (Olson 1999), with the premise that gene loss may provide opportunities for future adaptations. Our finding that gene loss itself can be adaptive supports and extends the "less-is-more" hypothesis. Although *CASP12* is the first demonstrated case of adaptive gene loss in humans, similar events may have occurred or are occurring at other loci. Given the high frequency of pseudogenization in genomic evolution, adaptive gene loss may not be rare and it would be interesting to explore this possibility in future studies.

5. Positive Selection for Elevated Gene Expression Noise

Just as positive selection can act on null alleles, it can also act on other seemingly harmful yet actually beneficial mutations. Here I describe our recent finding of positive selection for elevated gene expression noise in yeast (Zhang *et al*. 2009). Gene expression noise refers to the stochastic variation in the expression level of a gene among isogenic cells under the same condition. Here, *expression level* refers to the level of the protein product of the gene, as expression noise is usually measured at the level of protein. Gene expression noise has been measured in prokaryotes (Elowitz *et al*. 2002; Ozbudak *et al*. 2002; Rosenfeld *et al*. 2005), unicellular eukaryotes (Blake *et al*. 2003; Raser and O'Shea 2004), and mammalian cells (Ramsey *et al*. 2006). These and other studies showed that the level of expression noise varies substantially among genes, is determined genetically, and is selectable (Blake *et al*. 2006; Newman *et al*. 2006; Maheshri and O'Shea 2007; Ansel *et al*. 2008). Expression noise has both intrinsic and extrinsic sources (Orphanides and Reinberg 2002; Rao *et al*. 2002; Blake *et al*. 2003; Raser and O'Shea 2004; Kaern *et al*. 2005; Raser and

O'Shea 2005; Bar-Even *et al*. 2006; Newman *et al*. 2006; Volfson *et al*. 2006). Stochastic events in gene expression, including those in transcription initiation, mRNA degradation, translation initiation, and protein degradation, generate intrinsic noise (Raser and O'Shea 2005). Differences between cells, either in local environment or in the concentration or activity of any factor influencing gene expression, generate extrinsic noise (Raser and O'Shea 2005). We focus on intrinsic noise here, because only intrinsic noise is an intrinsic property of a gene.

Gene expression noise is often considered a two-edged sword. On the one hand, the noise could be deleterious because it ruins cellular homeostasis in metabolism and developmental programs, affects precise controls of biochemical processes in cells, and breaks the stoichiometric balances among members of protein complexes (Fraser *et al*. 2004; Batada *et al*. 2006; Lehner 2008). On the other hand, several benefits of expression noise have been suggested. In particular, it has been argued that stochastic noise is essential in cell fate determination (Colman-Lerner *et al*. 2005; Kaern *et al*. 2005; Losick and Desplan 2008) and thus is important in the development of multicellular organisms. In unicellular organisms, it has been shown both theoretically and experimentally that stochastic switching of expression level or high expression noise could be beneficial in the face of fluctuating environments or acute environmental stresses (Thattai and van Oudenaarden 2004; Blake *et al*. 2006; Acar *et al*. 2008). It is thus plausible that a certain fraction of genes in a genome have elevated expression noise driven by positive selection.

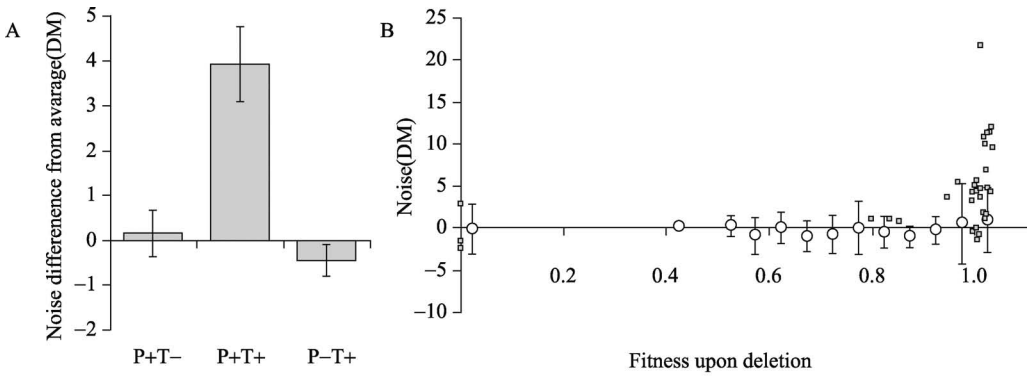
Newman and colleagues measured the expression noise for over 2000 genes of the budding yeast *Saccharomyces cerevisiae* in rich (YPD) medium (Newman *et al*. 2006). Because they controlled for several extrinsic factors, their noise estimates can be approximately regarded as intrinsic noise (Newman *et al*. 2006). The noise level is commonly measured by the coefficient of variation (*CV*), which is the standard deviation of the expression level divided by the mean. Newman and colleagues found a genome-wide pattern of lower *CV* for genes with higher mean expression. To control the influence of mean expression level on noise and allow among-gene comparison of noise levels, they used a new measure of noise named *DM*. For a given gene, *DM* is the difference of its *CV* from the median *CV* of those genes that have a similar mean expression as the focal gene (Newman *et al*. 2006).

Because there is good evidence that the expression noise is lessened by natural selection for genes important to cell growth (Fraser *et al*. 2004; Batada and Hurst 2007; Lehner 2008), we need to control for the "importance" of a gene when evaluating whether it is noisier than the expectation. The importance of a gene to yeast cell growth can be measured by the reduction in growth rate (i.e., fitness) in YPD upon deletion of the gene from the genome. Fortunately, such data exist for virtually every yeast gene (Giaever *et al*. 2002; Steinmetz *et al*. 2002). We separate all genes with expression noise data into 21 bins of different importance levels, with the fitness of the deletion strains being in the ranges of <0.05, 0.05-0.10, 0.10-0.15, …, 0.95-1.00, and >1.00, respectively. The last bin is not empty because the fitness value of a gene-deletion strain was originally measured relative to the mean of all viable gene-deletion strains, rather than to the wild-type strain (Steinmetz *et al*. 2002). To test whether the noise level of genes belonging to a given GO category exceeds the expectation, we randomly draw genes (with replacement) from the genome-wide expression noise data to form a gene set that has the same number of genes in each of the 21 bins as the focal GO has. We repeat this process 20,000 times and calculate the proportion of times when the mean noise level of the GO is lower than that of the randomly constructed gene set. If this probability is lower than 5%, we regard the GO to be significantly noisier than expected. Because we examine numerous GO categories, we further control for multiple testing using a 5% false discovery rate (Storey and Tibshirani 2003). That is, only GOs with a Q-value < 0.05 are considered as truly significant. To ensure that there is sufficient statistical power to detect elevated noise of a GO, only those GOs with at least 30 genes were examined.

GO categories are organized into three groups: biological process, cellular component, and molecular function (Ashburner *et al*. 2000). The three groups characterize different aspects of a gene's function and are thus examined separately in our analysis. We found that in terms of biological process, 18 GOs related to metabolism and transport show significantly higher-than-expected noise. In terms of cellular component, 5 GOs related to organelles (particularly mitochondrion) have high noise. In terms of molecular function, 4 GOs related to catalytic activity and transporter activity have high noise. The high expression noise of proteins localized to the mitochondrion (and other low-copy-number organelles) was noted before and was thought to be caused by unequal partitioning of mitochondria (and other organelles) during mitosis (Newman *et al*. 2006). Further, the high noise of enzymes is likely due to their special insensitivity to dosage, because it is well known that, in a metabolic pathway, even a considerable change in the concentration of an enzyme has a minimal effect on the flux of the pathway (Kacser and Burns 1981). This phenomenon arises from the kinetic connection via the shared substrates/products of adjacent biochemical reactions such that the effect of changing the catalytic activity in one reaction tends to be buffered by the response to this of the other reactions (Kacser and Burns 1981). To be conservative, we removed all mitochondrial proteins and all enzymes, and re-tested each GO. This time, we identified plasma membrane as the only cellular-component GO category and transporter activity as the only molecular-function GO category that show significantly higher-than-expected noise. No biological-process GO category is significantly noisier than expected. We further excluded all known factors that could potentially lead to a relaxation of purifying selection against noise and result in unexpected high expression noise of plasma membrane genes and transporter genes.

We noticed that plasma-membrane transporters are significantly noisier than expected after the control for gene importance and the removal of enzymes and mitochondrial proteins ($P = 3.3\times10^{-6}$; two-tail Z-test), whereas plasma-membrane proteins that are non-transporters ($P = 0.77$) and transporters that are not localized to the plasma membrane ($P = 0.21$) are not significantly different from the expectation (Fig. 4A). A careful examination shows that the majority of plasma-membrane transporters (79%) belong to the last bin of gene importance (i.e., fitness of the gene-deletion strain >1.00) (Fig. 4B). For this bin, the genomic average noise level is $DM = 0.87 \pm 0.16$, only slightly, although significantly, greater than the mean noise (-0.10 ± 0.18) of the first bin (i.e., fitness <0.05), suggesting that the effect of negative selection in reducing the expression noise of important genes is overall relatively small (Fig. 4B). By contrast, the mean noise of the plasma-membrane transporters in the last bin is $DM = 5.62 \pm 1.00$, suggesting that the effect of positive selection in elevating expression noise can be substantial (Fig. 4B). Again, the above comparison is based on the dataset after the removal of enzymes and mitochondrial proteins. Fig. 4C lists the 20 noisiest plasma-membrane transporters. These proteins transport a diverse array of chemicals, such as amino acids, glucose, ions, thiamine, polyamine, oligopeptides, and nucleotides, across the cell membrane. They are involved in the uptake of nutrients and ions, excretion of end products of metabolism and deleterious substances, and communication between cells and the environment.

Why would high noise be beneficial to plasma-membrane transporters? It is likely that the optimal expression level of each transporter depends on environmental factors such as the nutrients available to the cell. Under-expression of a transporter may limit the nutrient uptake rate and hence limit the cell's Darwinian fitness. On the other hand, over-expression of a transporter could also be disadvantageous for two reasons. First, over-expression has a fitness cost due to the waste of energy in transcription and translation (Wagner 2005; Stoebel *et al*. 2008). Second and more importantly, presence of unwanted transporters could reduce the metabolic efficiency and hence the fitness. For example, imagine that two carbon sources C_1 (e.g., maltose) and C_2 (e.g., lactose) are both present in the medium, but C_1 is energetically more efficient than C_2 for the cell to use. If the total number



C

Gene name	Noise (DM)	Average noise	Functional annotation(from SGD)
AGP1	21.6	0.9	Low-affinity amino acid permease
THI7	12.1	1.0	Plasma membrane transporter responsible for the uptake of thiamine
TPO4	11.5	0.9	Polyamine transport protein
HXT2	11.3	1.0	High-affinity glucose transporter of the major facilitator superfamily
TPO1	10.9	0.9	Polyamine transporter
ZRT2	10.0	0.9	Low-affinity zinc transporter of the plasma membrane
TPO2	9.5	1.1	Polyamine transport protein
OPT1	6.9	0.9	Proton-coupled oligopeptide transporter of the plasma membrane
FUI1	5.4	0.8	High affinity uridine permease
QDR2	5.6	1.0	Multidrug transporter of the major facilitator superfamily
TPO3	5.0	1.0	Polyamine transport protein specific for spermine
HXT3	4.7	0.9	Low affinity glucose transporter
AQR1	4.8	1.0	Plasma membrane multidrug transporter
MUP1	4.7	1.0	High affinity methionine permease
FTR1	3.7	0.0	High affinity iron permease
FET4	4.5	1.0	Low-affinity Fe(II)transporter
FCY2	4.2	0.9	Purine-cytosine permease
SAM3	4.2	1.0	High-affinity S-adenosylmethionine permease
VHT1	2.8	–0.1	High-affinity plasma membrane H+-biotin(vitamin H)symporter
DIP5	3.7	0.9	Dicarboxylic amino acid permease

▲ **Fig. 4**

Higher-than-expected expression noise of plasma-membrane transporters in yeast. (A) Plasma-membrane transporters (P+T+) are significantly noisier than the neutral expectation. By contrast, non-transporter plasma membrane proteins (P+T-) and non-plasma-membrane transporters (P-T+) are not noisier than the expectation. The expectation is computed by the mean *DM* of all genes in the genome with the same level of gene importance (after the removal of enzymes and proteins localized to mitochondrion). Error bars represent one standard error. (B) The noise levels of plasma-membrane transporters, in comparison to those of all genes in the genome (after the removal of enzymes and proteins localized to mitochondrion). Genes are divided into 21 bins based on the fitness of the gene-deletion yeast strains. The mean and standard deviation of the noise level for each bin is shown by an open circle and error bars, respectively. No circle is shown if a bin contains no gene, and no error bar is shown if a bin contains only one gene. Plasma-membrane transporters are shown by grey squares. (C) Twenty noisiest plasma-membrane transporters in yeast. The expected noise level is computed by the mean *DM* of all genes in the genome with the same level of gene importance (after the removal of enzymes and proteins localized to mitochondrion). Reprinted, with permission, from Zhang *et al*. (2009).

of carbon-source molecules that the cell can catabolize per unit time is limited, it would be better for the cell to use C_1 rather than C_2. Thus, over-expression of the transporter for C_2 will reduce the number of carbon-source molecules catabolized by the cell per unit time and thus be deleterious. Certainly, many transporter genes are under transcriptional regulation such that the transporter concentrations differ under different environments. However, changes of expression by gene regulation take time and are energetically costly (Perez-Ortin *et al*. 2007). More importantly, the cell does not have regulatory responses to all possible environmental changes. Thus, high

expression noise of transporters allows at least some cells to have high fitness in an unpredictable environment. Below we show mathematically that, under certain conditions, genotypes with high expression noise can have greater Darwinian fitness than those with low noise.

Let us consider two genotypes A and B. The only difference between them is that A has a higher level of expression noise than B for gene *X*. The mean expression level (m) of *X* is identical between the two genotypes. The distribution of the expression noise (e) for gene *X* is described by probability density functions $g_A(e)$ and $g_B(e)$ for the two genotypes, respectively. Genome-wide expression noise data showed that e generally follows a normal distribution (Bar-Even *et al*. 2006; Newman *et al*. 2006). Let us assume that a population consisting of A and B cells experiences an environmental change such that the mean expression level of *X* becomes suboptimal. Let $f(x)=f(m+e)$ be the fitness of the cell that has an expression level of X equal to x. So, the fitness of genotype A, or the mean fitness of A cells, equals

$$F_A=\int_{-\infty}^{\infty} f(m+e)g_A(e)de.$$

Similarly, the fitness of genotype B equals

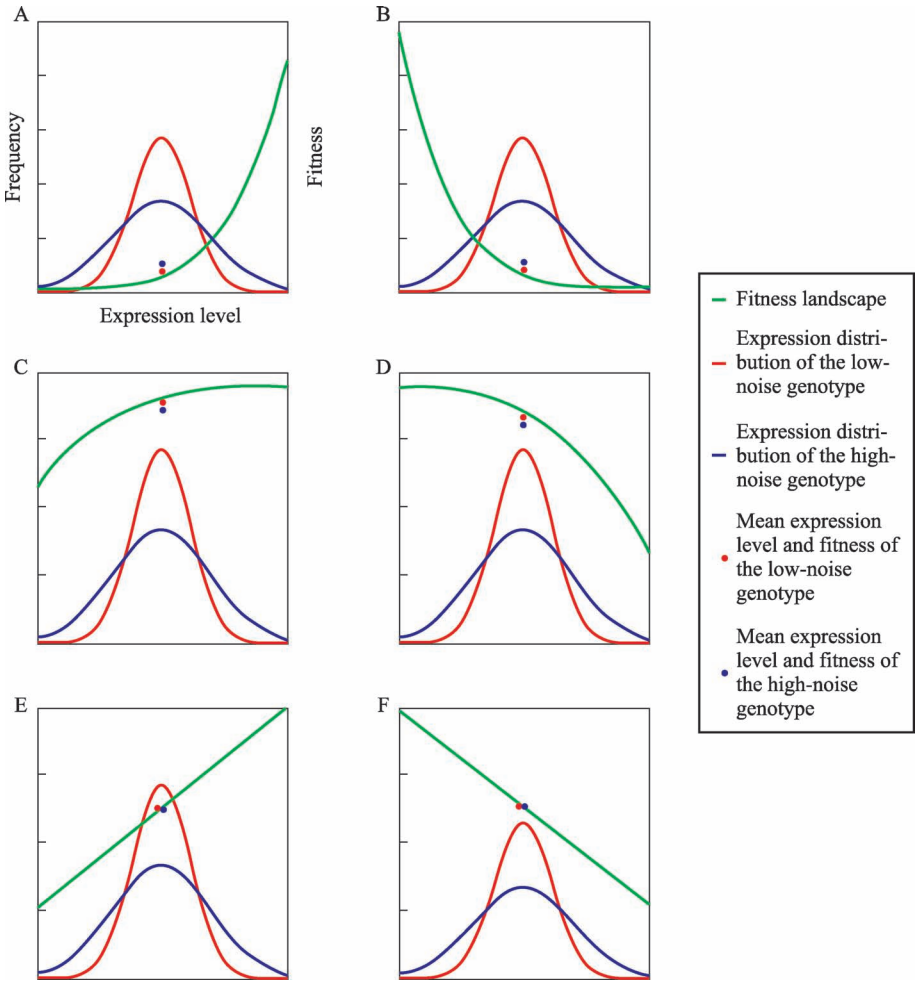
$$F_B=\int_{-\infty}^{\infty} f(m+e)g_B(e)de.$$

It can be shown that (i) when $f(x)$ is a convex function (i.e., the second derivative of $f(x)$ is positive), $F_A > F_B$; (ii) when $f(x)$ is a concave function, $F_A < F_B$; and (iii) when $f(x)$ is linear, $F_A = F_B$ (Fig. 5). Because $f(x)$ may not be concave or convex for all possible values of x, what matters is whether $f(x)$ is concave or convex for the range of x realized in the majority (e.g., 95% or 99%) of A and B cells. Note that in our model, the optimal expression level can be either higher or lower than m (Fig. 5). Although the shape of $f(x)$ is generally unknown, it is reasonable to assume that at least for many genes if not most genes, it is bell-shaped with the optimal expression level in the center (Kacser and Burns 1981; Hartl *et al*. 1985; Bedford and Hartl 2009). In such cases, $f(x)$ is concave when x is close to the optimal expression level, but convex when x is far from the optimal. Thus, big environmental changes tend to generate conditions under which high noise is beneficial. Note that although we compared mean fitness values of cells with two different genotypes, there is no involvement of group selection in our model. When $f(x)$ is convex, in a population fixed with the wild-type, a mutant with a higher level of noise is expected to increase its frequency in the population because its fitness is greater than that of the wild-type. Given the large effective population of yeast, F_A-F_B can easily reach a level that is detectable by natural selection.

We expect that all unicellular organisms that face unpredictable and frequent environmental changes would show a similar pattern of elevated expression noise in those genes whose expression levels are often suboptimal, and it will be interesting to test this prediction in the future when genome-wide expression noise data become available for additional species. The power and versatility of natural selection in seizing and utilizing even seemingly harmful biological properties such as the stochasticity in gene expression to enhance organismal fitness is a wonderful tribute to the theory of evolution by means of natural selection.

6. Final Remarks

In this review, I described a few cases of adaptive gene evolution in which positive selection acted on diverse types of mutations, including missense mutations, nonsense mutations, indel mutations, and presumably regulatory mutations that affect gene expression noise. Positive selection at the molecular level is not limited to the few types described here. For example, positive selection can also lead to stable retention of multiple allelic forms of a gene for a long evolutionary time (Cho *et al*. 2006) and biased expansion/shrinkage of gene families in different evolutionary lineages (Zhang *et al*. 2000; Shi and Zhang 2007). Identification of positive selection in gene evolution helps understand the factors causing the unusual evolutionary patterns (e.g., parallel substitution) of these genes, which in turn sheds light on the mechanisms underlying the evolution of the phenotypes (e.g., echolocation) that are controlled by these genes. If positive selection is prevalent in phenotypic evolution, it must also play an important role in



▲ Fig. 5

Fitness landscape affects the relative fitness of high-noise and low-noise genotypes. In each panel, the green curve shows $f(x)$, the fitness of the cell with the expression level of gene X equal to x. The blue and red curves show the frequency distributions of the expression levels (x) of the high-noise and low-noise genotypes, respectively. The blue and red dots are the mean fitness of the high-noise and low-noise genotypes, respectively. When $f(x)$ is convex, the mean fitness of the high-noise genotype is greater than that of the low-noise genotype, no matter whether the optimal expression level is higher (**A**) or lower (**B**) than the mean expression levels of the two genotypes. When $f(x)$ is concave, the fitness of the high-noise genotype is smaller than that of the low-noise genotype, no matter whether the optimal expression level is higher (**C**) or lower (**D**) than the mean expression levels of the two genotypes. When $f(x)$ is linear, the fitness of the high-noise genotype equals that of the low-noise genotype, no matter whether the optimal expression level is higher (**E**) or lower (**F**) than the mean expression levels of the two genotypes. Reprinted, with permission, from Zhang *et al*. (2009).

gene evolution, because phenotypic changes are ultimately caused by genetic changes. Although it is possible that most nucleotide substitutions between species are due to random fixations of neutral mutations, the role of positive selection in molecular evolution cannot be ignored, because positive selection is likely responsible for the nucleotide substitutions that underlie some of the most important phenotypic adaptations, such as the brain-size expansion and the emergence of speech/language in human origins (Zhang *et al*. 2002a; Zhang 2003).

Acknowledgements

I thank the students, postdoctoral fellows, and collaborators who have contributed to the studies reviewed here. I also thank Manyuan Long and Hongya Gu for organizing the Darwin 200 China conference, in which part of this review was presented. Researches in my lab have been supported by the University of Michigan and the US National Institutes of Health.

References

[1] Acar M, Mettetal J T, van Oudenaarden A. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet*, 2008, 40:471-5.

[2] Alnemri E S, Livingston D J, Nicholson D W, Salvesen G, Thornberry N A, Wong W W, Yuan J. Human ICE/CED-3 protease nomenclature. *Cell*, 1996, 87:171.

[3] Anisimova M, Kosiol C. Investigating protein-coding sequence evolution with probabilistic codon substitution models. *Mol Biol Evol*, 2009, 26:255-71.

[4] Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, Francois J, Yvert G. Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genet*, 2008, 4:e1000049.

[5] Ashburner M, Ball C A, Blake J A, Botstein D, Butler H, Cherry J M, Davis A P, Dolinski K, Dwight S S, Eppig J T, Harris M A, Hill D P, Issel-Tarver L, Kasarskis A, Lewis S, Matese J C, Richardson J E, Ringwald M, Rubin G M, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 2000, 25:25-9.

[6] Au W W L. A comparison of the sonar capabilities of bats and dolphins.In: Thomas J A, *et al*.,eds. Echolocation in bats and dolphins. Chicago: The University of Chicago Press, 2004, xiii-xxvii.

[7] Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, Barkai N. Noise in protein expression scales with natural protein abundance. *Nat Genet*, 2006, 38:636-43.

[8] Barnard E A. Biological function of pancreatic ribonuclease. *Nature*, 1969, 221:340-4.

[9] Batada N N,Hurst L D. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat Genet*, 2007, 39:945-9.

[10] Batada N N, Reguly T, Breitkreutz A, Boucher L, Breitkreutz B J, Hurst L D, Tyers M. Stratus not altocumulus: a new view of the yeast protein interaction network. *PLoS Biol*, 2006, 4:e317.

[11] Bedford T, Hartl D L. Optimization of gene expression by natural selection. *Proc Natl Acad Sci USA*, 2009, 106: 1133–8.

[12] Beintema J J. The primary structure of langur, *Presbytis entellus*, pancreatic ribonuclease: adaptive features in digestive enzymes in mammals. *Mol Biol Evol*, 1990, 7:470-7.

[13] Bezanilla F, Armstrong C M. Inactivation of the sodium channel. I. Sodium current experiments. *J Gen Physiol*, 1977, 70:549-66.

[14] Blake W J, Balazsi G, Kohanski M A, Isaacs F J, Murphy K F, Kuang Y, Cantor C R, Walt D R, Collins J J. Phenotypic consequences of promoter-mediated transcriptional noise. *Molecular Cell*, 2006, 24:853-865.

[15] Blake W J, M K A, Cantor C R, Collins J J. Noise in eukaryotic gene expression. *Nature*, 2003, 422:633-7.

[16] Britten R J. Divergence between samples of chimpanzee and human DNA sequences is 5%, counting indels. *Proc Natl Acad Sci USA*, 2002, 99:13633-5.

[17] Castoe T A, de Koning A P, Kim H M, Gu W, Noonan B P, Naylor G, Jiang Z J, Parkinson C L, Pollock D D. Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci USA*, 2009, 106:8986-91.

[18] Cavalli-Sforza L L, Feldman M W. The application of molecular genetic approaches to the study of human evolution. *Nat Genet*, 2003, 33 Suppl:266-75.

[19] Cho S, Huang Z Y, Green D R, Smith D R, Zhang J. Evolution of the complementary sex-determination gene of honey bees: balancing selection and trans-species polymorphisms. *Genome Res*, 2006, 16:1366-75.

[20] Code C F. Handbook of Physiology. American Physiological Association, Washington DC, 1968.

[21] Colman-Lerner A, Gordon A, Serra E, Chin T, Resnekov O, Endy D, Pesce C G, Brent R. Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 2005, 437:699-706.

[22] Dallos P. Cochlear amplification, outer hair cells and prestin. *Curr Opin Neurobiol*, 2008, 18:370-6.

[23] Elowitz M B, Levine A J, Siggia E D, Swain P S. Stochastic gene expression in a single cell. *Science*, 2002, 297:1183-6.

[24] Evans P D, Gilbert S L, Mekel-Bobrov N, Vallender E J, Anderson J R, Vaez-Azizi L M, Tishkoff S A, Hudson R R, Lahn B T. Microcephalin, a gene regulating brain size, continues to evolve adaptively in humans. *Science*, 2005, 309:1717-20.

[25] Fay J C, Wu C I. Hitchhiking under positive Darwinian selection. *Genetics*, 2000, 155:1405-13.

[26] Fischer H, Koenig U, Eckhart L, Tschachler E. Human caspase 12 has acquired deleterious mutations. *Biochem Biophys Res Commun*, 2002, 293:722-6.

[27] Fraser H B, Hirsh A E, Giaever G, Kumm J, Eisen M B. Noise minimization in eukaryotic gene expression. *PLoS Biol*, 2004, 2:e137.
[28] Giaever G, Chu A M, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A ,Anderson K, Andre B, Arkin A P, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A, Entian K D, Flaherty P, Foury F, Garfinkel D J, Gerstein M, Gotte D, Guldener U, Hegemann J H, Hempel S, Herman Z, Jaramillo D F, Kelly D E, Kelly S L, Kotter P, LaBonte D, Lamb D C, Lan N, Liang H, Liao H, Liu L, Luo C, Lussier M, Mao R, Menard P, Ooi S L, Revuelta J L, Roberts C J, Rose M, Ross-Macdonald P, Scherens B, Schimmack G, Shafer B, Shoemaker D D, Sookhai-Mahadeo S, Storms R K, Strathern J N, Valle G, Voet M, Volckaert G, Wang C Y, Ward T R, Wilhelmy J, Winzeler E A, Yang Y, Yen G, Youngman E, Yu K, Bussey H, Boeke J D, Snyder M, Philippsen P, Davis R W, Johnston M. Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, 2002, 418:387-91.
[29] Guyton A.C, Hall J E. Textbook of Medical Physiology. 9th ed. London: Saunders Co. , 1996.
[30] Hahn M W. Toward a selection theory of molecular evolution. *Evolution*, 2008, 62:255-65.
[31] Hartl D L, Dykhuizen D E, Dean A M. Limits of adaptation: the evolution of selective neutrality. *Genetics*, 1985, 111:655-74.
[32] Hoshi T, Zagotta W N, Aldrich R W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science*, 1990, 250:533-8.
[33] Hughes A L, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature*, 1988, 335:167-70.
[34] Jones G. Echolocation. *Curr Biol*, 2005, 15:R484-8.
[35] Jones G, Teeling E C. The evolution of echolocation in bats. *Trends Ecol Evol*, 2006,21:149-56.
[36] Kacser H, Burns J A. The molecular basis of dominance. *Genetics*, 1981, 97:639-66.
[37] Kaern M, Elston T C, Blake W J, Collins J.J. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet*, 2005, 6:451-64.
[38] Kay R N B, Davies A G. Digestive physiology. In: Davies A G, Oates J F, eds. In Colobinem Monkeys: Their Ecology, Behaviour and Evolution. Cambridge: Cambridge University Press, 1994.
[39] Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge University Press, 1983.
[40] Lamkanfi M, Declercq W, Kalai M, Saelens X, Vandenabeele P. Alice in caspase land. A phylogenetic analysis of caspases from worm to man. *Cell Death Differ*, 2002, 9:358-61.
[41] Lehner B. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Mol Syst Biol* , 2008, 4:170.
[42] Li G, Wang J, Rossiter S J, Jones G, Cotton J A, Zhang S. The hearing gene Prestin reunites echolocating bats. *Proc Natl Acad Sci USA*, 2008, 105:13959-64.
[43] Li Y, Liu Z, Shi P, Zhang J. The hearing gene *Prestin* unites echolocating bats and whales. *Curr. Biol.*, 2010, 20:R55-R56.
[44] Losick R, Desplan C. Stochasticity and cell fate. *Science*, 2008, 320:65-8.
[45] Maheshri N, O'Shea E K. Living with noisy genes: How cells function reliably with inherent variability in gene expression. *Annual Review of Biophysics and Biomolecular Structure*, 2007, 36:413-434.
[46] Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res*, 1974, 23:23-35.
[47] Navaratnam D, Bai J P, Samaranayake H, Santos-Sacchi J. N-terminal-mediated homomultimerization of prestin, the outer hair cell motor protein. *Biophys* J, 2005, 89:3345-52.
[48] Nei M. Selectionism and neutralism in molecular evolution. *Mol Biol Evol* , 2005, 22:2318-42.
[49] Nei M, Kumar S. Molecular Evolution and Phylogenetics. New York: Oxford University Press, 2000.
[50] Newman J R, Ghaemmaghami S, Ihmels J, Breslow D K, Noble M, DeRisi J L, Weissman J S. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature*, 2006, 441:840-6.
[51] Nielsen R. Molecular signatures of natural selection. *Annu Rev Genet*, 2005, 39:197-218.
[52] Olson M V. When less is more: gene loss as an engine of evolutionary change. *Am J Hum Genet*, 1999, 64:18-23.
[53] Orphanides G, Reinberg D. A unified theory of gene expression. *Cell*, 2002, 108:439-51.
[54] Ozbudak E M, Thattai M, Kurtser I, Grossman A D, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nat Genet*, 2002, 31:69-73.
[55] Perez-Ortin J E, Alepuz P M, Moreno J. Genomics and gene transcription kinetics in yeast. *Trends Genet*, 2007, 23:250-7.
[56] Podlaha O, Webb D M, Tucker P K, Zhang J. Positive selection for indel substitutions in the rodent sperm protein catsper1. *Mol Biol Evol*, 2005, 22:1845-52.
[57] Podlaha O, Zhang J. Positive selection on protein-length in the evolution of a primate sperm ion channel. *Proc Natl Acad Sci USA*, 2003, 100:12241-6.
[58] Ramsey S, Ozinsky A, Clark A, Smith K D, de Atauri P, Thorsson V, Orrell D, Bolouri H. Transcriptional noise and cellular heterogeneity in mammalian macrophages. *Philos Trans R Soc Lond B Biol Sci*, 2006, 361:495-506.
[59] Rao C V, Wolf D M, Arkin A P. Control, exploitation and tolerance of intracellular noise. *Nature*, 2002, 420:231-7.
[60] Raser J M, O'Shea E K. Control of stochasticity in eukaryotic gene expression. *Science*, 2004, 304:1811-4.
[61] Raser J M, O'Shea E K. Noise in gene expression: origins, consequences, and control. *Science*, 2005, 309:2010-3.
[62] Ren D, Navarro B, Perez G, Jackson A C, Hsu S, Shi Q, Tilly J L, Clapham D E. A sperm ion channel required for sperm motility and male fertility. *Nature*, 2001, 413:603-9.
[63] Rosenfeld N, Young J W, Alon U, Swain P S, Elowitz M B. Gene regulation at the single-cell level. *Science*, 2005, 307:1962-5.
[64] Saleh M, Vaillancourt J P, Graham R K, Huyck M, Srinivasula S M, Alnemri E S, Steinberg M H, Nolan V, Baldwin C T, Hotchkiss R.S, Buchman T G, Zehnbauer B A, Hayden M R, Farrer L A, Roy S, Nicholson D W. Differential modulation of endotoxin responsiveness by human caspase-12 polymorphisms. *Nature*, 2004, 429:75-9.

[65] Shi P, Zhang J. Comparative genomic analysis identifies an evolutionary shift of vomeronasal receptor gene repertoires in the vertebrate transition from water to land. *Genome Res*, 2007, 17:166-74.
[66] Silva J C, Kondrashov A S. Patterns in spontaneous mutation revealed by human-baboon sequence comparison. *Trends Genet*, 2002, 18:544-7.
[67] Sorrentino S, Libonati M. Structure-function relationships in human ribonucleases: main distinctive features of the major RNase types. *FEBS Lett*, 1997, 404:1-5.
[68] Steinmetz L M, Scharfe C, Deutschbauer A M, Mokranjac D, Herman Z S, Jones T, Chu A M, Giaever G, Prokisch H, Oefner P J, Davis R W. Systematic screen for human disease genes in yeast. *Nat Genet*, 2002, 31:400-4.
[69] Stoebel D M, Dean A M, Dykhuizen D E. The cost of expression of Escherichia coli lac operon proteins is in the process, not in the products. *Genetics*, 2008, 178:1653-60.
[70] Storey J D, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*, 2003,100:9440-5.
[71] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 1989, 123:585-95.
[72] Thattai M, van Oudenaarden A. Stochastic gene expression in fluctuating environments. *Genetics*, 2004, 167:523-30.
[73] Volfson D, Marciniak J, Blake W J, Ostroff N, Tsimring L S, Hasty J. Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 2006, 439:861-4.
[74] Wagner A. Energy constraints on the evolution of gene expression. *Mol Biol Evol*, 2005, 22:1365-74.
[75] Wang X, Grus W E, Zhang J. Gene losses during human origins. *PLoS Biol*, 2006, 4:e52.
[76] Zagotta W N, Hoshi T, Aldrich R W. Restoration of inactivation in mutants of Shaker potassium channels by a peptide derived from ShB. *Science*, 1990, 250:568-71.
[77] Zhang J. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes. *J Mol Evol*, 2000, 50:56-68.
[78] Zhang J. Evolution of the human ASPM gene, a major determinant of brain size. *Genetics*, 2003, 165:2063-70.
[79] Zhang J. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. *Nat Genet*, 2006, 38:819-23.
[80] Zhang J. Positive selection, not negative selection, in the pseudogenization of rcsA in Yersinia pestis. *Proc Natl Acad Sci USA*, 2008, 105:E69.
[81] Zhang J. Evolutionary genetics: Progresses and challenges. In: Bell M A, *et al*. eds. Evolution Since Darwin: The First 150 Years. Sunderland: Sinauer, 2010.
[82] Zhang J, Dyer K D, Rosenberg H F. Evolution of the rodent eosinophil-associated RNase gene family by rapid gene sorting and positive selection. *Proc Natl Acad Sci USA*, 2000, 97:4701-6.
[83] Zhang J, Kumar S. Detection of convergent and parallel evolution at the amino acid sequence level. *Mol Biol Evol*, 1997, 14:527-36.
[84] Zhang J, Kumar S, Nei M. Small-sample tests of episodic adaptive evolution: a case study of primate lysozymes. *Mol Biol Evol*, 1997, 14:1335-8.
[85] Zhang J, Webb D M, Podlaha O. Accelerated protein evolution and origins of human-specific features: Foxp2 as an example. *Genetics*, 2002a, 162:1825-35.
[86] Zhang J, Zhang Y P, Rosenberg H F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nat Genet*, 2002b, 30:411-5.
[87] Zhang Z, Qian W, Zhang J. Positive selection for elevated gene expression noise in yeast. *Mol Syst Biol*, 2009, 5:299.