Review TRENDS in Genetics Vol.22 No.8 August 2006

Natural selection on gene expression

Yoav Gilad[1], Alicia Oshlack[2] and Scott A. Rifkin[3]

[1]Department of Human Genetics, University of Chicago, Chicago, Il 60637, USA
[2]Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria 3050, Australia
[3]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

Changes in genetic regulation contribute to adaptations in natural populations and influence susceptibility to human diseases. Despite their potential phenotypic importance, the selective pressures acting on regulatory processes in general and gene expression levels in particular are largely unknown. Studies in model organisms suggest that the expression levels of most genes evolve under stabilizing selection, although a few are consistent with adaptive evolution. However, it has been proposed that gene expression levels in primates evolve largely in the absence of selective constraints. In this article, we discuss the microarray-based observations that led to these disparate interpretations. We conclude that in both primates and model organisms, stabilizing selection is likely to be the dominant mode of gene expression evolution. An important implication is that mutations affecting gene expression will often be deleterious and might underlie many human diseases.

Introduction

Differences in gene regulation are likely to have an important role in phenotypic variation within and between species [1–3]. Accumulating evidence suggests that regulatory changes contribute to many adaptations in natural populations and influence the susceptibility to several human diseases [4,5]. Despite the potential phenotypic importance of regulatory variation, until recently little was known about the different selective pressures acting on regulatory patterns.

A better understanding of the forces influencing gene regulation is not only of interest in an evolutionary context but also promises to shed light on the contribution of regulatory region variation to human diseases [6]. To date, the main focus of disease susceptibility studies has been on coding regions, whereas mutations in regulatory sequences are relatively neglected [7]. If most regulatory changes are effectively neutral (i.e. do not affect fitness), mutations in regulatory regions are unlikely to contribute to disease and can be safely ignored. However, if most changes in gene regulation are selected against, it is reasonable to assume that these mutations underlie several disease phenotypes and should be studied more intensively.

One approach to studying the evolutionary forces that shape gene regulation is to start at the DNA level and examine how regulatory sequences evolve [6,8]. Using sequence variation data, population genetic tools can be used to infer the action of both purifying (negative) as well as directional (positive) selection [9,10]. However, unlike coding regions, regulatory sequences are difficult to identify [11,12] and therefore their sequence variation is difficult to characterize. Although the rates of evolution of protein sequences have clear implications for protein function [10], the link between sequence variation and gene regulation is less clear. Indeed, in certain cases, gene regulation can be conserved even when the regulatory sequences have changed [13–15].

An alternative approach is to start at the phenotypic level and analyze variation in patterns of gene expression. The challenge is then to use comparisons of variation within and between populations to distinguish between neutral changes in gene expression and patterns that are consistent with natural selection [16].

Inferring the mode of gene expression evolution

How can one distinguish between different modes of gene expression evolution? One approach is to search for departures from a neutral model. In the gene expression literature, this has often been taken to refer to examples in which mutations have no fitness consequences, so that their evolution is governed entirely by genetic drift. By contrast, Kimura's original formulation of the 'neutral theory', allows a subset of alleles to be too strongly deleterious to either segregate within a population or reach fixation. However, alleles that reach appreciable enough frequency within a population to be sampled or are fixed between species are selectively neutral [17,18]. Under Kimura's formulation of the neutral model, both the levels of polymorphism (i.e. differences within a population) and divergence (i.e. differences between populations) are simple functions of the mutation rate [19]. Therefore, the ratio of polymorphism to divergence should be the same across neutral loci, a prediction that forms the basis of a common test of neutrality using DNA sequences [20].

One alternative to Kimura's neutral theory, referred to as 'the nearly neutral theory', is that a large proportion of mutations are slightly deleterious and have selection coefficients on the order of the reciprocal of the effective population size [18,21]. Slightly deleterious mutations contribute to polymorphism seen within a sample but tend to be at low frequencies, and rarely reach fixation [22]. In this case, the ratio of polymorphism to divergence is expected to be greater than under the neutral theory. By contrast, if most mutations in a locus are beneficial (i.e. evolving under positive selection), they will be more likely

Corresponding authors: Gilad, Y. (gilad@uchicago.edu); Oshlack, A. (oshlack@wehi.edu.au); Rifkin, S.A. (srifkin@aya.yale.edu).
Available online 27 June 2006

 0168-9525/$ - see front matter doi:10.1016/j.tig.2006.06.002

to reach fixation than under the neutral or nearly neutral theories. Thus, the ratio of polymorphism to divergence will be less than expected under the neutral or the nearly neutral models.

In the example of a quantitative phenotype, such as gene expression level, evolutionary constraint is likely to take the form of stabilizing selection, which maintains a constant mean and reduces the variance of the trait [16,23]. When a gene expression profile is under strong evolutionary constraint, mutations that lead to marked departures from this profile will be eliminated from the population. This selection pressure will maintain the same mean and a low variance of gene expression both within and between populations. By contrast, if departures from a certain gene expression profile are weakly deleterious (i.e. nearly neutral), alleles that lead to them can persist within populations but will only rarely reach fixation. Thus, the within population variance will increase, but the mean will remain similar. As a result, the between population variance will be less than that expected under the neutral theory [21]. Finally, the fixation of mutations that lead to a beneficial change in expression profiles will lead to a difference in the mean expression level between populations.

Thus, different models of selection on gene expression make distinct predictions. If changes in gene expression profiles do not affect the fitness of an individual, segregating variation will only be affected by stochastic processes. As a result, the gene expression evolution will reflect the mutational input. Under the assumption that there is no non-genetic component to expression variation, the increase in the variance between populations (fixed differences) will be proportional to the within population variance (segregating changes). Hence, in principle, neutral divergence can be predicted based on levels of segregating genetic variation, the effective population size and the number of generations separating populations [16,24]. If the observed divergence is less than this estimate, it suggests that stabilizing selection prevented segregating variation from reaching fixation. Conversely, if divergence is greater than the estimate based on segregating variation, disruptive (or differential) selection might have caused a shift in the mean level of the quantitative trait [25].

Testing these predictions with expression data is challenging. In particular, it requires the partition of the observed variation in a quantitative trait (such as gene expression) into its genetic and non-genetic (e.g. environmental and genetic–environment interaction) components. In model organisms, minimizing the difference in environment between samples helps to reduce the environmental variance. In other species, for example, primate species, it is nearly impossible to obtain accurate estimates of the purely genetic variance, making it difficult to reach reliable conclusions regarding the selection pressures acting on gene expression. Furthermore, stabilizing selection and neutrality are not exclusive processes. Traits can fluctuate without constraints within bounds set by stabilizing selection [26,27]. In the next sections, we discuss recent empirical work on measuring the impact of the different evolutionary forces potentially acting on gene expression.

Natural selection on gene expression

Several recent studies have focused on investigating the selection pressures acting on gene regulation. In the first study to investigate natural variation in gene expression, Oleksiak *et al.* [28] compared the specific mRNA abundances of heart ventricles of 18 individual post-reproductive males in three populations: two of *Funduls heteroclitus* (a saltwater fish) and one of its close relative, *Funduls grandis*. An ANOVA model was used to compare the variance within populations with the variance between populations in gene expression. Despite low migration rates between the two conspecific populations and across the species boundary, $<3\%$ of the 907 genes they surveyed varied significantly between populations, whereas an order of magnitude more genes varied significantly between individuals within populations. As a result, there was little evidence of population structure at the genome-wide expression level. In addition, patterns of variation between populations were inconsistent with the neutral prediction that phenotypic divergence should scale with genetic distance. Instead, gene expression profiles were more similar for the southern *F. heteroclitus* and *F. grandis* populations, suggesting that adaptation to different temperatures, rather than drift, drove the differentiation, a pattern that is consistent with sequence data.

Rifkin *et al.* [29] took a more explicit quantitative genetic approach in a study of gene expression variation during *Drosophila* metamorphosis. They measured average levels of gene expression in four strains of cosmopolitan *Drosophila melanogaster* and one strain each of *D. simulans*, and *D. yakuba*, at the start of metamorphosis. At that time, appropriate estimates were not available to parameterize a null neutral model for gene expression [23]. Instead, to understand the selection pressures driving gene expression evolution, they classified genes into three groups based on comparisons of within-species variation with between-species variation. To identify genes with similar expression levels across all six species and strains (i.e. the entire clade), they tested whether expression values in the six samples were statistically indistinguishable from each other given the measurement error of the experiment. Using only the genes rejected by this test, they proceeded to identify genes with similar (i.e. statistically indistinguishable) expression levels between the *D. melanogaster* strains but different levels between species. Finally, having removed genes with little variation both within *D. melanogaster* and within the entire clade, they compared patterns of variation within species with the variation between species to identify genes for which a neutral model could not be rejected (Box 1). Based on this series of tests, they could not reject the overall low variation for 44% of the expressed genes, could not reject the species-specific gene expression patterns for 39% of the genes and could not reject a model consistent with neutrality for the remaining 17% of genes. They interpreted these results to indicate a dominant signature for stabilizing selection in gene expression evolution with smaller, but important, roles for directional selection and neutral evolution, respectively.

Box 1. Inferring the mode of gene expression evolution

The neutral theory of molecular evolution provides a model that describes the expected patterns of nucleotide diversity when mutations do not affect fitness and hence are not subjected to natural selection [18]. Analogous models have been developed for the evolution of quantitative traits that are solely caused by mutation and drift [16,23,43].

Lynch and Hill [23] developed a model to predict the within and between population variance in a quantitative trait under several breeding schemes, with additive and dominant mutations, different population sizes and recombination rates. They assumed small populations, so that a maximum of two alleles would be segregating at any locus at a given time. They concluded that, for all the conditions they investigated, the rate of increase in between-line variance per generation is $2V_m$, where V_m is the mutational variance (the increase in variance per generation in a trait that is solely a result of mutation) and the equilibrium level of within-population segregating genetic variance is between $2V_mN_e$ and $4V_mN_e$ (where N_e is the effective population size). The exact value depends on the breeding system, map distance between loci and degree of dominance of new mutations [23].

Mutational variance is a cumbersome parameter to measure even in model organisms and effectively impossible for many others. To circumvent this, Lynch [24] constructed a neutral null model for phenotypic evolution based on within and between population variance by making the assumptions (based on general estimates for several quantitative traits) that $10^{-4}V_e < V_m < 10^{-2}V_e$ and that $V_w/2 \leq V_e$ (where V_e is the environmental variance and V_w is the phenotypic variance within the population). In their reanalysis of gene expression data, Lemos *et al.* [30] used these assumptions to make the case that variation between species was much less than the neutral expectation. Rifkin *et al.*, [32] showed that V_m averaged $\sim 10^{-5}V_e$ for gene expression.

Recently, quantitative genetic models that feature epistasis were developed (e.g. Refs [44,45]). Because expression levels vary with genetic background, this work might lead to better expectations for the relationship between mutational, segregating and between population variance in gene expression.

In contrast to Rifkin *et al.*, Lemos *et al.* [30] explicitly tested a null neutral model of gene expression evolution by making two key assumptions about variance in gene expression. First, they used estimates of mutational variance in other quantitative traits as a measure of the mutational variance that might be affecting gene expression. Second, following Lynch [24], they assumed that environmental variance was half the within-population variance (i.e. that broad-sense heritability of gene expression patterns was at most 50%). Using these estimates, and based on the neutral model of Lynch and Hill [23], they calculated the minimal and maximal rates of gene expression diversification that would be consistent with neutrality (i.e. evolution without constraint). Then they proceeded to analyze published inter-species gene expression data sets from mice, *Drosophila* and apes. They estimated the rates of diversification in gene expression in each study by scaling between species variance by the product of within species variance and time. Gene expression diversification rates outside the neutral interval were interpreted as signatures of stabilizing selection (if the diversification rate were lower) or directional selection (if the diversification rate were greater).

Lemos *et al.* [30] found that the vast majority genes in all data sets exhibited far less between species variation than expected under a neutral model. They interpreted this to be the result of stabilizing selection acting on within-species gene expression. Lemos *et al.* [30] estimated that even if the mutational input to gene expression were two orders of magnitude lower than they had assumed, levels of between population differentiations in gene expression would still be inconsistent with neutrality. Only in comparisons between mouse laboratory strains did an appreciable number of genes evolve in a manner consistent with neutrality.

Measuring the mutational input

As discussed earlier, mutational variance is a key parameter in quantitative genetic models. For model organisms, it can be estimated by measuring the variance of a phenotypic trait among a set of initially homogeneous lines maintained with minimally sized populations for many generations. Natural selection is at its weakest under such conditions, because there is rapid genetic drift in such small populations. In an extreme example, when a single, randomly chosen individual propagates each line, the only mutations that can be selected against are those that either kill the organism before reproduction or eliminate fertility completely. Instead, most mutations will be effectively neutral and will either quickly drift to fixation or be lost. As different lines fix different random mutations, the lines will drift apart. Variation between lines can then be used to estimate the mutational variance.

Using this scheme, Denver *et al.* [31] measured the rates of gene expression diversification within a set of four *Caenorhabditis elegans* mutation accumulation lines, where each generation was propagated from a single female. They compared the variance in gene expression among these lines with the variance in gene expression among five genotypically divergent natural isolates. They found the greatest ratios of intra-specific to mutational variances to be approximately one-tenth of the expectation in the absence of constraint – consistent with a strong role for stabilizing selection in limiting gene expression divergence in *C. elegans*.

In a similar study, Rifkin *et al.* [32] measured mutational variance for gene expression in 12 lines of *D. melanogaster* at the beginning of metamorphosis. They conservatively estimated the mutational variance for gene expression to be on the order of $10^{-5}V_e$ (V_e is the environmental variance). Based on this value, the neutral model of Lynch and Hill [23] predicts that gene expression differences among *D. melanogaster*, *D. simulans* and *D. yakuba* should be one or two orders of magnitude greater than the difference observed [29] (Figure 1). Consequently, Rifkin *et al.*, [32] also concluded that stabilizing selection places severe limits on gene expression divergence. For organisms where mutational variance has not been or cannot be estimated, $10^{-5}V_e$ would be a reasonable value to use to parameterize neutral models of gene expression evolution. However, for any particular gene this could be incorrect by several orders of magnitude [32].

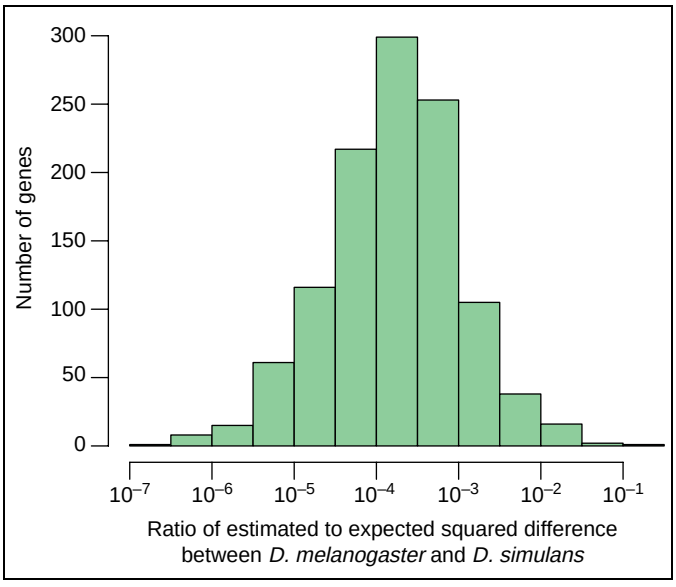

Figure 1. The estimated variation in gene expression levels is far less than the neutral expectation. Squared mean differences in gene expression levels between *Drosophila melanogaster* and *Drosophila simulans* were estimated from data in Refs [29,32] and corrected for any bias as a result of small and unequal sample sizes, assuming that the estimated within species variance from the *D. melanogaster* strains is a good estimate for both species. For the 1132 genes with positive estimates of divergence after the bias correction, we divided the estimate by the expected squared difference based on the mutational variance and a neutral model (Box 1).

Gene expression in apes

Understanding phenotypic evolution in primates is more difficult than in model organisms because key experiments often cannot be performed to distinguish between competing hypotheses or to estimate important parameters. Moreover, material is often scarce, leading to largely unknown and uncontrolled environmental variance between samples. These limitations are particularly problematic for dynamic, environmentally sensitive traits like gene expression. Nevertheless, several studies have used microarrays to study the evolution of gene expression in primates, with somewhat conflicting conclusions [30,33–38].

The early studies that compared gene expression within and between primates concluded that gene expression patterns in the human brain, but not in the liver, were consistent with adaptive evolution [33,34]. By contrast, a meta-analysis of the same data [33] suggested that a greater number of genes showed significantly changed expression in the human liver compared with those in brain [39]. Furthermore, based on a comparison of variance within species with variance between species, Hsieh *et al.* [39] suggested that most changes in gene expression (in liver and brain) are not adaptive.

Using a comparison of variance within and between species, Khaitovich *et al.* [38] measured gene expression in brain samples from six humans, three chimpanzees, one orang-utan and one rhesus macaque. They observed a linear relationship between the mean squared difference in expression levels and divergence time between species, and a correlation in gene expression variation within and between species. Under certain assumptions, these observations are consistent with lack of evolutionary constraint on gene expression and this is how they have been interpreted (Box 2). However, because Khaitovich *et al.* [38] hybridized non-human primate RNA to microarrays containing only human DNA, their findings might have been confounded by the hybridization of mismatched sequences [34,37,40]. Sequence mismatches attenuate hybridization [40,41], and the greater the divergence between species, the larger the effect on hybridization. Because global normalization can only correct for the mean attenuation of the signal caused by sequence mismatches (Box 3), the apparent increase of expression divergence

Box 2. Challenges in modeling the evolution of gene expression levels

The prediction of a linear trend of diversity with time arises from neutral evolutionary models with a stochastic basis [46,47]. However, these models might not be the most appropriate descriptions of gene expression evolution. In particular, boundaries in gene expression exist at both ends of the spectrum owing to both biological and technical factors. At the low end, expression cannot go below zero and detection will only become significant above the background at greater expression levels (the magnitude of which is dependent on the physical and chemical aspects of the assay in addition to capturing the data accurately [48]). At the high end, energetic costs of transcription and physical limitations of the transcriptional machinery might put a limit on gene expression levels. In addition, saturation of RNA binding to microarray probes limits the levels of expression that can be detected [49]. Therefore, the unconstrained limits in neutral models are probably not realistic when considering gene expression measures, particularly if the mutational input is on the scale of the boundaries in gene expression. Placing boundaries on expression levels effectively reduces the range of possible differences observed between species particularly for transcripts at either high or low concentrations [32]. It will have a greater effect for more-divergent species, because differences have had more time to accumulate. In addition, different genes are expected to have distinct boundaries, leading to variation in divergence across specific genes [26,27]. Therefore, to study the evolution of gene expression rigorously, neutral models need to be constructed that include the effect of the boundaries on gene expression levels [50].

Box 3. Normalization cannot correct sequence mismatches

Normalization of microarray data is crucial to remove systematic non-biological effects in any expression study. Interspecies comparisons are particularly susceptible to normalization strategies because systematic effects between species dominate hybridization and are difficult to distinguish from biological variation in transcript abundance. Normalization between arrays is usually achieved by shifting and rescaling the probe intensities to correct for overall variation that might be introduced during preparation of the target RNA, the manufacture of the arrays or the processing of the arrays (e.g. scanning) [51,52]. In its simplest form, normalization is performed by scaling the average expression level of each array to the same value. However, the simple normalization does not adjust for examples where there are non-linear relationships between arrays. More-complex methods such as quantile normalization have been successfully introduced to account for non-linear effects [51].

Normalization methods transform probe intensities (e.g. so that the intensity distribution of each array is the same) or log-ratios (e.g. to balance dye effects in two color arrays). However, normalization preserves the order of gene intensities within an array relative to each other. This implies that, for interspecies comparisons using a single-species microarray, normalization corrects for average reductions in hybridization efficiencies as a result of sequence mismatches but the reordering of gene intensities that occurs as a result of differing levels of hybridization affinities between probes and RNA from different species will be maintained.

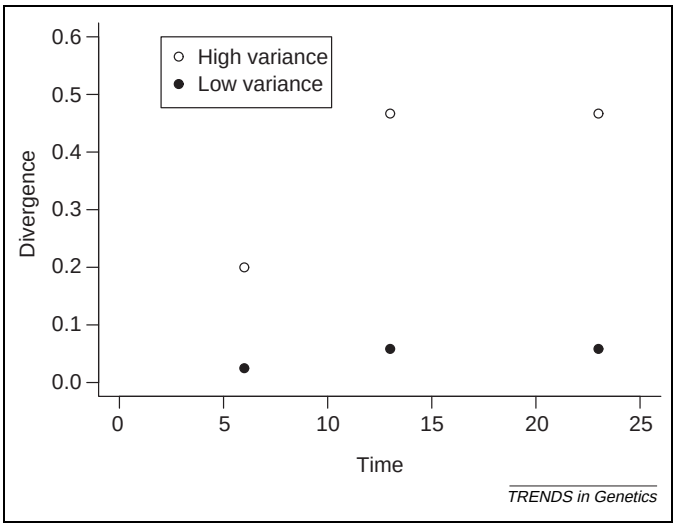

Figure 2. An illustration of the bias in divergence estimates. Here the divergence, calculated as the squared difference between two means, is plotted against evolutionary time. However, in this example the true divergence between species is zero. The differences between species come from bias associated with calculating the squared difference between samples of different size (six humans, three chimpanzees, one orang-utan and one rhesus macaque), following the sampling scheme used in Ref. [38]. Two examples are shown, one with high within species variance (open circles) and one with low within species variance (black circles).

with time can be driven by an increasing variance of the sequence mismatch effect between species. In addition, regardless of the evolutionary scenario, the squared difference between the means of two samples will be biased by an amount that is proportional to the sampling variance. Hence, their progressively smaller sample sizes will lead to estimates of squared differences between species that are increasingly biased upwards, even in the absence of divergence in expression levels (Figure 2).

To address these issues empirically, Gilad *et al.* [42] used multi-species arrays to correct for the effects of sequence mismatches between species and a balanced design to equalize the sampling variance for each species. They did not find a linear trend of divergence with time for primate livers (Figure 3). Instead, for most genes, there was little evidence for a change in expression levels across the four primates, consistent with widespread stabilizing selection. Similarly, Khaitovich *et al.* [35] found that expression divergence between human and chimpanzee was lower for genes that are expressed in multiple tissues compared with genes that are expressed in only one or a few tissues. They interpreted this to be evidence that negative (stabilizing) selection has a role in the evolution of gene expression.

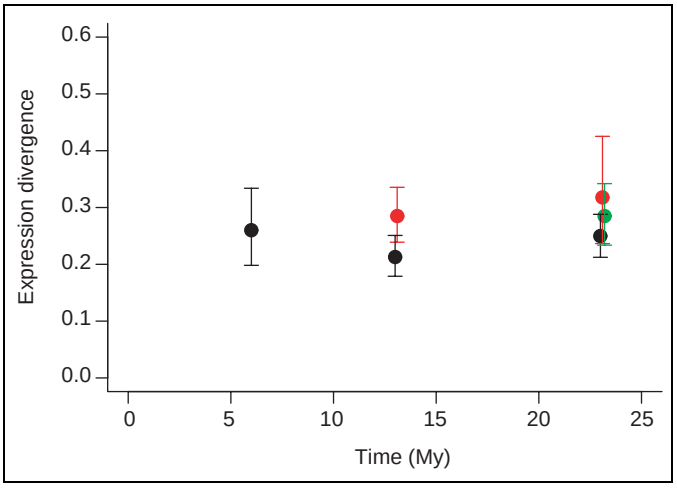

Figure 3. Gene expression divergence between primates. The mean squared expression difference among human, chimpanzee, orang-utan and rhesus macaque is plotted against divergence time (from data of Ref. [42]). The data points in black represent the comparisons involving humans; those in red represent the comparison involving chimpanzees; those in green represent the comparison of orang-utan with rhesus macaque. The vertical error bars indicate 95% confidence intervals calculated by 10 000 bootstraps over genes. Abbreviation: My, million years.

Concluding remarks

The effective population size of primates is smaller than that of many model organisms, rendering selection less effective. Hence, as long as the selection coefficients associated with gene expression changes are small, one might expect that the expression levels of many primate genes evolve solely under the influence of genetic drift. However, the current evidence in primates is consistent with widespread stabilizing selection on gene expression, as found in model organisms. This finding suggests that changes in gene expression are frequently deleterious and thus many mutations affecting gene expression might contribute to disease susceptibility. An important challenge is to identify the subset of genes in which regulation is most tightly constrained, and the set of genes in which regulation has evolved under adaptive pressures [6].

Acknowledgements

We thank G.K. Smyth for help with statistical analysis, T.P. Speed, P. De Candia, R.R. Hudson, M. Przeworski and K.P. White for helpful discussions and comments on the manuscript. Y.G. is supported by DDRCC grant number P30 DK42086. A.O. is supported by NHMRC Program Grant 406657. S.A.R. is supported by NIH grant number 5 R01 GM068465-02 to D. Hartl.

References

1 King, M.C. and Wilson, A.C. (1975) Evolution at two levels in humans and chimpanzees. *Science* 188, 107–116
2 Jin, W. *et al*. (2001) The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*. *Nat. Genet.* 29, 389–395
3 Carroll, S.B. (2003) Genetics and the making of *Homo sapiens*. *Nature* 422, 849–857
4 Kleinjan, D.A. and van Heyningen, V. (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* 76, 8–32
5 Taron, M. *et al*. (2004) BRCA1 mRNA expression levels as an indicator of chemoresistance in lung cancer. *Hum. Mol. Genet.* 13, 2443–2449
6 Rockman, M.V. *et al*. (2005) Ancient and recent positive selection transformed opioid cis-regulation in humans. *PLoS Biol.* 3, e387
7 Botstein, D. and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* 33, (Suppl.) 228–237.
8 Keightley, P.D. *et al*. (2005) Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol.* 3, e42
9 Hellmann, I. *et al*. (2003) Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res.* 13, 831–837
10 Bustamante, C.D. *et al*. (2005) Natural selection on protein coding genes in the human genome. *Nature* 437, 1153–1157
11 Moses, A.M. *et al*. (2004) MONKEY: identifying conserved transcription-factor binding sites in multiple alignments using a binding site-specific evolutionary model. *Genome Biol.* DOI:10.1186/gb-2004-5-12-r98 (http://genomebiology.com/2004/5/12/R98)
12 Gasch, A.P. *et al*. (2004) Conservation and evolution of cis-regulatory systems in ascomycete fungi. *PLoS Biol.* 2, e398
13 Dickinson, W.J. (1988) On the architecture of regulatory systems: evolutionary insights and implications. *BioEssays* 8, 204–208

14 Ludwig, M.Z. *et al*. (2000) Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564–567
15 Ludwig, M.Z. *et al*. (2005) Functional evolution of a cis-regulatory module. *PLoS Biol.* 3, e93
16 Lande, R. (1976) Natural-selection and random genetic drift in phenotypic evolution. *Evolution* 30, 314–334
17 Kimura, M. (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet. Res.* 11, 247–269
18 Kimura, M. (1983) The neutral mutation-random drift hypothesis as an evolutionary paradigm. In *The Neutral Theory*, pp. 34–55, Cambridge University Press
19 Watterson, G.A. (1975) On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* 7, 256–276
20 Hudson, R.R. *et al*. (1987) A test of neutral molecular evolution based on nucleotide data. *Genetics* 116, 153–159
21 Ohta, T. (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. *J. Mol. Evol.* 40, 56–63
22 Akashi, H. (1999) Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. *Genetics* 151, 221–238
23 Lynch, M. and Hill, W.G. (1986) Phenotypic evolution by neutral mutation. *Evolution* 40, 915–935
24 Lynch, M. (1990) The rate of morphological evolution in mammals from the standpoint of the neutral expectation. *Am. Nat.* 136, 727–741
25 Smith, J.M. (1979) The effects of normalizing and disruptive selection on genes for recombination. *Genet. Res.* 33, 121–128
26 Whitehead, A. and Crawford, D.L. (2006) Variation within and among species in gene expression: Raw material for evolution. *Mol. Ecol.* 15, 1197–2211
27 Whitehead, A. and Crawford, D.L. (2006) Neutral and adaptive variation in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 5425–4530
28 Oleksiak, M.F. *et al*. (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* 32, 261–266
29 Rifkin, S.A. *et al*. (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* 33, 138–144
30 Lemos, B. *et al*. (2005) Rates of divergence in gene expression profiles of primates, mice, and flies: stablizing selection and variability among functional categories. *Evolution* 59, 126–137
31 Denver, D.R. *et al*. (2005) The transcriptional consequences of mutation and natural selection in *Caenorhabditis elegans*. *Nat. Genet.* 37, 544–548
32 Rifkin, S.A. *et al*. (2005) A mutation accumulation assay reveals a broad capacity for rapid evolution of gene expression. *Nature* 438, 220–223
33 Enard, W. *et al*. (2002) Intra- and interspecific variation in primate gene expression patterns. *Science* 296, 340–343
34 Caceres, M. *et al*. (2003) Elevated gene expression levels distinguish human from non-human primate brains. *Proc. Natl. Acad. Sci. U. S. A.* 100, 13030–13035
35 Khaitovich, P. *et al*. (2005) Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science* 309, 1850–1854
36 Khaitovich, P. *et al*. (2004) Regional patterns of gene expression in human and chimpanzee brains. *Genome Res.* 14, 1462–1473
37 Karaman, M.W. *et al*. (2003) Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. *Genome Res.* 13, 1619–1630
38 Khaitovich, P. *et al*. (2004) A neutral model of transcriptome evolution. *PLoS Biol.* 2, E132
39 Hsieh, W.P. *et al*. (2003) Mixed-model reanalysis of primate data suggests tissue and species biases in oligonucleotide-based gene expression profiles. *Genetics* 165, 747–757
40 Gilad, Y. *et al*. (2005) Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. *Genome Res.* 15, 674–680
41 Sartor, M.A. *et al*. (2006) A new method to remove hybridization bias for interspecies comparison of global gene expression profiles uncovers an association between mRNA sequence divergence and differential gene expression in *Xenopus Nucleic Acids Res.*. 34, 185–200
42 Gilad, Y. *et al*. (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* 440, 242–245
43 Chakraborty, R. and Nei, M. (1982) Genetic differentiation of quantitative characters between populations or species.1. Mutation and random genetic drift. *Genet. Res.* 39, 303–314
44 Hansen, T.F. and Wagner, G.P. (2001) Epistasis and the mutation load: a measurement-theoretical approach. *Genetics* 158, 477–485
45 Hill, W.G. *et al*. (2005) Prediction of effects of genetic drift on variance components under a general model of epistasis *Theor. Popul. Biol.* DOI:10.1016/j.tpb.2005.10.001 (http://www.sciencedirect.com/science/journal/00405809)
46 Felsenstein, J. (2003) *Inferring phylogenies*, Sinauer Associates
47 Khaitovich, P. *et al*. (2005) Toward a neutral evolutionary model of gene expression. *Genetics* 170, 929–939
48 Yang, Y.H. *et al*. (2001) Normalization for cDNA microarray. In *Proceedings of SPIE. Microarrays: Optical Technologies and Informatics* (Vol. 4266) (Bittner, M.L. *et al*. eds), pp. 141–152, Society for Optical Engineering
49 Hekstra, D. *et al*. (2003) Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays. *Nucleic Acids Res.* 31, 1962–1968
50 Tachida, H. and Cockerham, C.C. (1989) A building block model for quantitative genetics. *Genetics* 121, 839–844
51 Bolstad, B.M. *et al*. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193