Supplementary data
Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2008.10.011.

References

1 Marais, G. (2003) Biased gene conversion: implications for genome and sex evolution. *Trends Genet.* 19, 330–338
2 Galtier, N. *et al.* (2001) GC-content evolution in Mamm. *Genomes: the biased gene conversion hypothesis. Genetics* 159, 907–911
3 Webster, M.T. and Smith, N.G. (2004) Fixation biases affecting human SNPs. *Trends Genet.* 20, 122–126
4 Webster, M.T. *et al.* (2005) Male-driven biased gene conversion governs the evolution of base composition in human alu repeats. *Mol. Biol. Evol.* 22, 1468–1474
5 Duret, L. *et al.* (2006) A new perspective on isochore evolution. *Gene* 385, 71–74
6 Dreszer, T.R. *et al.* (2007) Biased clustered substitutions in the human genome: The footprints of male-driven biased gene conversion. *Genome Res.* 17, 1420–1430
7 Duret, L. and Arndt, P.F. (2008) The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genet.* 4, e1000071
8 Galtier, N. and Duret, L. (2007) Adaptation or biased gene conversion? Extending the null hypothesis of molecular evolution. *Trends Genet.* 23, 273–277
9 Pollard, K.S. *et al.* (2006) An RNA gene expressed during cortical development evolved rapidly in humans. *Nature* 443, 167–172
10 Ranwez, V. *et al.* (2007) OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol. Biol.* 7, 241
11 Bakewell, M.A. *et al.* (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7489–7494
12 Nagylaki, T. (1983) Evolution of a finite population under gene conversion. *Proc. Natl. Acad. Sci. U. S. A.* 80, 6278–6281
13 Yang, Z. (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol. Biol. Evol.* 15, 568–573
14 Winckler, W. *et al.* (2005) Comparison of fine-scale recombination rates in humans and chimpanzees. *Science* 308, 107–111
15 Myers, S. *et al.* (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science* 310, 321–324
16 Bernardi, G. *et al.* (1985) The mosaic genome of warm-blooded vertebrates. *Science* 228, 953–958
17 Bernardi, G. (2007) The neoselectionist theory of genome evolution. *Proc. Natl. Acad. Sci. U. S. A.* 104, 8385–8390
18 Otto, S.P. and Barton, N.H. (1997) The evolution of recombination: removing the limits to natural selection. *Genetics* 147, 879–906
19 Begun, D.J. and Aquadro, C.F. (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* 356, 519–520
20 Bachtrog, D. and Charlesworth, B. (2002) Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* 416, 323–326
21 Spencer, C.C. *et al.* (2006) The influence of recombination on human genetic diversity. *PLoS Genet.* 2, e148
22 Eyre-Walker, A. *et al.* (2006) The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173, 891–900

Genome Analysis

Protein material costs: single atoms can make an evolutionary difference

Jason G. Bragg[1,2] and Andreas Wagner[3,4,5]

[1] Biology Department, University of New Mexico, Albuquerque, NM 87131, USA
[2] Present address: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3] University of Zurich, Department of Biochemistry, Building Y27, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
[4] The Santa Fe Institute, Santa Fe, NM 87501, USA
[5] The Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland

The process of gene expression has material costs caused by the quantities of carbon, nitrogen, sulfur and phosphorus that are needed to make mRNAs and proteins. When any such chemical element is ecologically limiting, mutations increasing these costs can reduce growth. Here, we ask if such mutations are 'visible' to natural selection in the yeast *Saccharomyces cerevisiae*. We find that mutations causing small increases in expression and even single amino acid replacements can be subject to natural selection on the basis of their material costs.

Nutrient limitation can influence protein evolution

Over vast regions of the globe, elemental nutrients including nitrogen, phosphorus and carbon limit the growth of organisms and mediate competition between them. Specific elements often limit growth because they are needed to make important biomolecules [1]. For example, mRNAs contain carbon (C), nitrogen (N) and phosphorus (P) atoms and proteins contain carbon, nitrogen and sulfur (S) atoms. Protein composition is constrained by natural selection because proteins need to perform specific functions, but this is not the only compositional constraint. Natural selection can also influence protein material costs (e.g. see Refs [2–7]). For example, highly expressed proteins can contain fewer atoms of ecologically limiting elements (per amino acid) than other proteins in the same proteomes [4,6,7]. Currently, we lack a quantitative, mechanistic explanation for the adaptive evolution of protein material costs. That is, mutations might affect material costs by directly changing the number of atoms required to make a protein, or by changing gene expression levels. However, we do not know if the resulting changes in

Corresponding author: Wagner, A. (aw@bioc.unizh.ch).

a cellular nutrient budget are ever of a sufficient magnitude to be visible to natural selection.

Here, we address this question in the yeast *Saccharomyces cerevisiae*. To do this, we estimate the C, N, S and P costs of mutations as a fraction of the total cellular budgets of these elements. We then relate these costs to a selection coefficient, s, and determine whether it exceeds a critical selection coefficient, s_c, above which we would expect the fate of a mutation to be dominated by natural selection. The goal is to determine whether mutations that are functionally neutral can affect material costs by a sufficient amount to be subject to natural selection, given necessary simplifying assumptions.

Material costs of mRNA and protein synthesis

We began by calculating the material costs of expressing protein-coding genes. These are the C, N and P costs of making mRNAs and the C, N and S costs of making proteins (see supplementary material online for details). We illustrate these calculations for carbon, but calculations for other elements proceed similarly. For each mRNA, we multiplied its abundance, R (number of copies; from Refs [8,9]), by the number of carbon atoms in a single copy of the mRNA, C_R, to obtain the total carbon cost of the mRNA (sequence information from *Saccharomyces* Genome Database; ftp://genome-ftp.stanford.edu/pub/yeast/). We then estimated the fractional carbon cost of the mRNA, F_R, as a dimensionless proportion of the total carbon content of the cell, using information on the biomass composition of yeast [10]. To estimate the fractional carbon cost of a protein, F_P, we carried out analogous calculations, using data for protein abundance (P; from Ref. [11]) and for the carbon content of the protein in atoms. We then estimated the carbon cost of expressing each protein-coding gene (F) as a dimensionless fraction of total cell carbon by adding the costs of the encoded mRNA (F_R) and protein (F_P). Among yeast genes, this fractional material cost F varies over approximately four orders of magnitude for C, N and S and over approximately three orders of magnitude for P (Figure 1a). Typically, the fractional carbon and nitrogen cost of a protein is greater than the cost of its corresponding mRNA (Figure S1).

Only small expression changes are neutral

If a mutation doubles the expression of a gene, the increase in material cost is equal to the fractional cost of expression, F. To estimate the effects of such mutations on growth, we developed a model of nutrient uptake and cellular biosynthesis by an exponentially growing population of yeast (see supplementary material online). The model shows that, under strong nutrient limitation, the mutant will have a growth rate of approximately $g_M = g(1 - F)$, where g_M is the growth rate of the mutant and g is the growth rate of the wild type. The selection coefficient, s, associated with the mutation is then given by $s = 1 - W$, where W is the ratio of the mutant growth rate to the wild-type growth rate ($W = g_M/g$). It follows that $s \approx 1 - (1 - F)$, or $s \approx F$. In the supplementary material online, we detail assumptions that were made in deriving this approximation and consider a range of additional factors, including the turnover
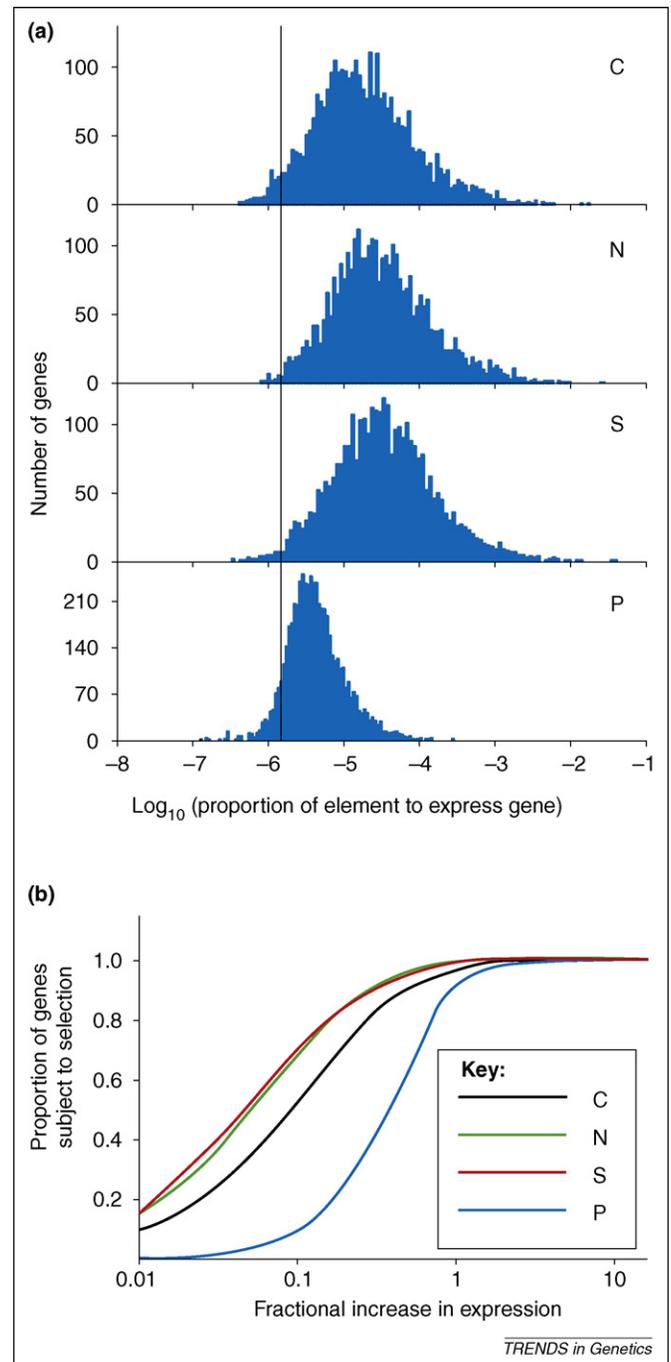
Figure 1. The material costs of gene expression. **(a)** The fraction of cellular carbon, nitrogen, sulfur and phosphorus bound in the expression of yeast genes. For carbon and nitrogen, values shown are the sum of mRNA and protein elemental content. The vertical line indicates the threshold value ($2s_c$) of material costs where a mutation is visible to selection during limitation by the nutrient. **(b)** The proportion of yeast genes (vertical axis) in which a fractional change in expression of a given magnitude (horizontal axis) would be visible to natural selection under limitation by the indicated nutrients, where $s = F/2$. The data indicate that small gene expression changes can have evolutionarily significant material costs for many yeast genes.

of mRNAs and proteins, recycling and the consequences of errors in our calculations or assumptions. If nutrient limitation is weaker, such that the cell can increase its uptake to partially compensate for the increased material costs, we expect a smaller reduction in the growth rate of the mutant and that $s < F$. Here, we concentrate on a value

of $s = F/2$ and present results for $s = F$ and $s = F/4$ in the supplementary material online.

For a diploid organism, natural selection and drift have a similar effect on the fate of a mutation when $s \approx 1/4N_e$, where N_e is the effective population size [12]. We assume that natural selection will dominate the fate of a mutation when the selection coefficient is ten-fold greater than this value (after Refs [13,14]), or when $s_c = 10/4N_e$. For yeast, we estimated $N_e = 3.41 \times 10^6$, using data on nucleotide diversity at synonymous sites [15] and the mutation rate per nucleotide [16]. This gives a critical selection coefficient of $s_c = 7.33 \times 10^{-7}$. This critical selection coefficient is adjusted to determine whether mutations are evolutionarily significant where $s = F/2$ by comparing values of F to $2 s_c$ (Figure 1a).

It is striking that, for most genes (>95%), a doubling of expression would be visible to selection during limitation by C, N or S (Figure 1). Because there is no phosphorus in proteins, the effect of a doubling of expression on phosphorus cost is usually less dramatic (Figure 1), although still visible to selection for 90% of genes under P-limitation.

Mutations can lead to changes in expression of varying magnitudes. This leads us to ask: how small can a change in gene expression be and still be visible to natural selection on the basis of material costs? To test this, we estimated the proportion of yeast genes where a given fractional change in expression would lead to an evolutionarily significant increase in material costs (Figure 1b). In general, we predict that small proportional increases in gene expression can have evolutionarily significant material costs in a substantial proportion of genes (Figure 1b). For example, where C, N or S is limiting, an increase of expression by 8% (a fractional increase of 0.08) would be visible to natural selection in >45% of genes. Small changes in expression would be evolutionarily significant in fewer genes on the basis of P costs (Figure 1b), although a 40% increase in gene expression would be visible to natural selection in >50% of genes during P-limitation.

These analyses indicate that mutations in regulatory regions causing even small changes in gene expression levels could be subject to natural selection on the basis of their material costs, complementing previous observations for energetic costs of expression [17,18]. They also indicate that gene duplications can be selected against owing to increased material costs. If we consider the specific case of a newly arisen duplicate that is expressed at a similar level to its parent gene, the resulting increase in material cost would be equivalent to a doubling in expression and, for most yeast genes, we would expect selection against the duplication, particularly if C, N or S were limiting. Material costs might, therefore, have a role in determining the fate of some gene duplications, along with advantages associated with increased dosages of gene products [19,20], and other non-adaptive evolutionary processes [21].

Selection for a single atom change in protein elemental content

Next, we estimated the fractional material cost of amino acid substitutions that directly change the cost per

molecule of making a protein. First, we considered substitutions that increased the material cost of a protein by one atom. If P is the abundance of a protein, then such substitutions will lead to an increase in cellular material content of P atoms. We converted this cost to a fraction of a cellular nutrient content, as we did previously for the costs of gene expression.

Such single atom costs range over ~5 orders of magnitude among yeast genes and they are substantially greater for sulfur than for nitrogen and carbon (Figure 2a). This is because most cellular sulfur occurs in proteins and because there are typically fewer S atoms in proteins than N or C atoms (for yeast proteins, the median carbon, nitrogen and sulfur content is 1952, 525 and 13 atoms, respectively). Therefore, adding one sulfur atom to a protein has a relatively large impact on total cellular elemental content. During limitation by sulfur, an additional sulfur atom
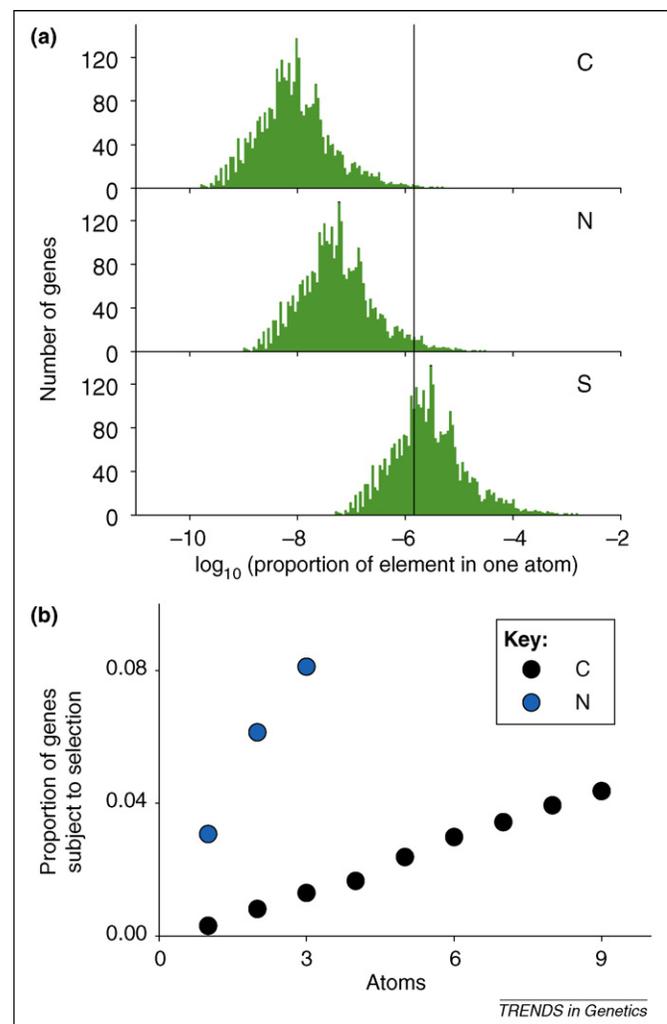
Figure 2. The material costs of amino acid substitutions. **(a)** The fraction of cellular carbon, nitrogen and sulfur associated with a single atom in yeast proteins. The vertical line indicates the threshold value ($2 s_c$) of material costs where a mutation is visible to selection during limitation by the nutrient. **(b)** The proportion of genes (vertical axis) whereby adding different numbers of carbon or nitrogen atoms (horizontal axis) to the protein product results in an evolutionarily significant increase in material costs during limitation by carbon or nitrogen, respectively. The data indicate that, for a substantial number of genes, single amino acid substitutions can increase material costs sufficiently to be visible to natural selection.

would be visible to natural selection for a majority (66%) of proteins. Conversely, during limitation by nitrogen or carbon, an additional nitrogen or carbon atom would be visible to natural selection in only 3% and 0.3% of proteins, respectively.

A single amino acid substitution can add up to three nitrogen atoms (e.g. Gly→Arg) or up to nine carbon atoms (Gly→Trp) to a protein. We, therefore, consider the selective consequences of adding different numbers of atoms to yeast proteins (Figure 2b). The addition of three N atoms would be visible to natural selection in 8.1% of proteins during nitrogen limitation and the addition of nine C atoms would be visible to natural selection in 4.4% of proteins during carbon limitation.

These calculations help explain how highly expressed proteins could become depleted in their content of important elemental nutrients [4,6,7]. They show that natural selection could oppose increases in the elemental content of highly expressed genes, even if those increases occur in small steps of one amino acid substitution at a time. In particular, previous studies show that the effect of natural selection on the sulfur content of proteins can be quite dramatic (e.g. see Refs [2–4,22,23]). This might seem curious, because sulfur is perhaps less often limiting in nature than the other nutrients we have considered. However, we show that if sulfur limitation is encountered, the potential for natural selection to act on protein sulfur content is high.

Concluding remarks

Nutrient limitation shapes living systems at essentially all scales of their organization [1], down to the quantities of different atoms used in making biopolymers (e.g. see Refs [4,6,7]). Here, we show how natural selection can operate on the material costs of gene expression, by showing that mutations increasing the quantities of C, N or S costs of expression can have large enough impacts on cellular material content to be opposed by selection. To do this, we used a framework linking ecological nutrient limitation and population genetics, to predict the selective consequences of changes in material costs of gene expression. In the future, we hope that this framework will be extended to different organisms to understand more broadly how material costs shape the evolution of proteins and genomes.

Supplementary data

Supplementary data associated with this article can be found at doi:10.1016/j.tig.2008.10.007.

References

1 Sterner, R.W. and Elser, J.J. (2002) *Ecological Stoichiometry*. Princeton University Press
2 Mazel, D. and Marlière, P. (1989) Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. *Nature* 341, 245–248
3 Baudouin-Cornu, P. *et al.* (2001) Molecular evolution of protein atomic composition. *Science* 293, 297–300
4 Fauchon, M. *et al.* (2002) Sulfur sparing in the yeast proteome in response to sulfur demand. *Mol. Cell* 9, 713–723
5 Panina, E.M. *et al.* (2003) Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis and rearrangement of ribosomal proteins. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9912–9917
6 Elser, J.J. *et al.* (2006) Signatures of ecological resource availability in the animal and plant proteomes. *Mol. Biol. Evol.* 23, 1946–1951
7 Bragg, J.G. and Wagner, A. (2007) Protein carbon content evolves in response to carbon availability and may influence the fate of duplicated genes. *Proc. R. Soc. Lond. B. Biol. Sci.* 274, 1063–1070
8 Holstege, F.C.P. *et al.* (1998) Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728
9 Arava, Y. *et al.* (2003) Genome-wide analysis of mRNA translation profiles in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 3889–3894
10 Förster, J. *et al.* (2003) Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 13, 244–253
11 Ghaemmaghami, S. *et al.* (2003) Global analysis of protein expression in yeast. *Nature* 425, 737–741
12 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press
13 Hartl, D.L. *et al.* (1985) Limits of adaptation: the evolution of selective neutrality. *Genetics* 111, 655–674
14 Petrov, D.A. and Hartl, D.L. (2000) Pseudogene evolution and natural selection for a compact genome. *J. Hered.* 91, 221–227
15 Johnson, L.J. *et al.* (2004) Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics* 166, 43–52
16 Drake, J.W. *et al.* (1998) Rates of spontaneous mutation. *Genetics* 148, 1667–1686
17 Akashi, H. and Gojobori, T. (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3695–3700
18 Wagner, A. (2005) Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* 22, 1365–1374
19 Papp, B. *et al.* (2003) Evolution of *cis*-regulatory elements in duplicated genes of yeast. *Trends Genet.* 19, 417–422
20 Vitkup, D. *et al.* (2006) Influence of metabolic network structure and function on enzyme evolution. *Genome Biol.* 7, R39
21 Force, A. *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545
22 Pardee, A.B. (1966) Purification and properties of a sulfate-binding protein from *Salmonella typhimurium*. *J. Biol. Chem.* 241, 5886–5892
23 Scheibel, T. *et al.* (1997) *S. cerevisiae* and sulfur: a unique way to deal with the environment. *FASEB J.* 11, 917–921