MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 2480–2487 doi: 10.1111/j.1365-294X.2012.05529.x

Protein elemental sparing and codon usage bias are correlated among bacteria

JASON G. BRAGG,*‡‡ ANTONIETTA QUIGG,†‡ JOHN A. RAVEN§ and ANDREAS WAGNER¶**††

*Department of Biological Sciences, Macquarie University, Sydney 2109, NSW, Australia, †Department of Marine Biology, Texas A&M University at Galveston, 200 Seawolf Parkway, Galveston, TX 77553, USA, ‡Department of Oceanography, Texas A&M University, 3146 TAMU, College Station, TX 77843, USA, §Division of Plant Sciences, The James Hutton Institute, University of Dundee at TJHI, Invergowrie, Dundee DD2 5DA, UK, ¶Department of Biochemistry, University of Zurich, Zurich CH-8057, Switzerland, **The Santa Fe Institute, Santa Fe, NM 87501, USA, ††The Swiss Institute of Bioinformatics, Basel, Switzerland

Abstract

Highly expressed proteins can exhibit relatively small material costs, in terms of the quantities of carbon (C), nitrogen (N) or sulphur (S) atoms they contain. This 'elemental sparing' probably reflects selection to reduce the quantities of potentially growth-limiting elements in abundant proteins, but the evolutionary mechanisms for adaptive elemental sparing are still poorly understood. Here, we predict that the extent of 'elemental sparing' in highly expressed proteins will vary among organisms, according to the effectiveness of selection in determining the fate of mutations. We test this hypothesis in bacteria by asking whether 'elemental sparing' is correlated with codon usage bias. Bacteria exhibit extraordinary variation in their life histories and demography and consequently in the effectiveness of selection in determining whether preferred codons are used in highly expressed genes. We find that C sparing and S sparing, but not N sparing, are significantly correlated with adaptive codon usage bias among 148 genera of bacteria, suggesting that selection for elemental sparing and codon bias are promoted by similar bacterial traits. Our study helps identify principles that determine how nutrient scarcity can shape the elemental composition of proteins.

Keywords: adaptation, bacteria, genomics/proteomics, molecular evolution

Received 15 April 2011; revision received 23 January 2012; accepted 1 February 2012

Introduction

Protein synthesis is central to life (Crick 1958), with consequences for competitiveness and fitness across broad scales of biological organization (Sterner & Elser 2002). It is also a very expensive process, in terms of both energy and materials. Protein degradation and resynthesis can comprise a substantial proportion of cellular maintenance costs (Raven *et al.* 2000; Quigg & Beardall 2003). For many organisms, a substantial fraction of all assimilated carbon (C), nitrogen (N) and sulphur (S) is used to make proteins. These costs of protein synthesis appear to evolve adaptively: it has been shown that highly expressed genes (Fauchon *et al.* 2002; Elser *et al.* 2006; Bragg & Wagner 2007; Li *et al.* 2009), genes that are induced during nutrient shortages (Mazel & Marlière 1989; Fauchon *et al.* 2002; Boer *et al.* 2003; Bragg & Wagner 2007; Gilbert & Fagan 2011) and genes that function in particular metabolic pathways (Van der Ploeg *et al.* 1996; Baudouin-Cornu *et al.* 2001; Acquisti *et al.* 2009) encode proteins that are relatively poor in elements that can be growth limiting (carbon, nitrogen or sulphur).

As evidence accumulates that protein material costs can evolve adaptively, it is important that we begin to ask how these adaptive signatures are produced by natural selection. If we can gain an understanding of these evolutionary mechanisms, we will be better able to make predictions about the environments, organisms and genes where adaptive biases in protein costs can evolve. The mechanisms of protein material cost

Correspondence: Jason G. Bragg, Fax: (+61) 2 62464864;
E-mail: jason.bragg@csiro.au
‡‡Present address: CSIRO Plant Industry, Canberra 2601, ACT, Australia.

evolution are difficult to investigate directly, because selection coefficients associated with mutations affecting protein costs are likely to be small, especially compared to mutations that affect protein function. One recent study considered hypothetical mutations leading to a change in the quantity of carbon, nitrogen and sulphur required to express proteins in the yeast *Saccharomyces cerevisiae*, and asked whether the changes in protein costs were large enough to be visible to natural selection (Bragg & Wagner 2009). It was predicted that in highly expressed genes, single amino acid substitutions could potentially have a large enough impact on cellular nutrient budgets, growth and fitness to be disfavoured by selection, if the corresponding element was growth limiting (Bragg & Wagner 2009).

More broadly, we do not yet know whether highly expressed proteins exhibit elemental sparing in most, many or few species. However, it is possible that many species will not exhibit elemental sparing, particularly if they have traits that make natural selection ineffective in determining the fate of mutations that change protein material costs. To illustrate this point, consider yeast cells, which in natural environments likely experience pulses in resource availability, leading to phases of nutrient-limited population growth. At such times, there is likely much scope for selection to act on mutations that affect the efficiency with which materials are used (Bragg & Wagner 2009). Also, yeast tend to have large effective population sizes (Tsai *et al.* 2008), meaning that small differences in growth rates and fitness can be acted upon by selection. These yeast traits likely make selection relatively effective in purging mutations that increase protein material costs. However, organisms with smaller effective population sizes, or organisms that rarely experience nutrient-limited population growth, might be less likely to exhibit adaptive elemental sparing in highly expressed proteins.

In this study, we aim to gain a better understanding of how protein costs evolve by studying variation among species in elemental sparing. In particular, we investigate protein costs in relation to adaptive biases in codon usage (Fig. 1). The genetic code is 'degenerate', meaning that many amino acids are encoded by more than one codon. Codons that encode the same amino acid are described as 'synonymous.' Codon usage describes the frequency with which different synonymous codons are used to encode amino acids (Grantham *et al.* 1980). In highly expressed proteins, selection can favour greater usage of 'preferred' codons (Ikemura 1981; Gouy & Gautier 1982), which tend to be translated more quickly or accurately than 'unpreferred' synonymous alternatives. The selection coefficient associated with a polymorphism between a preferred and unpreferred codon is likely to be small (Sharp *et al.* 2010).

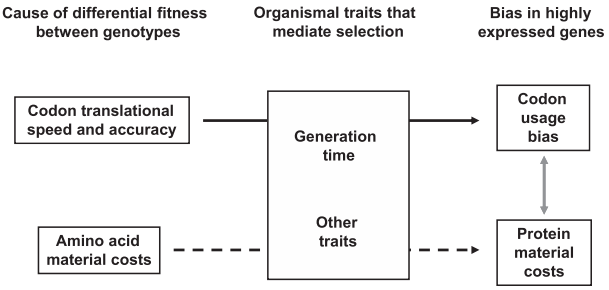

Fig. 1 Diagram illustrating relationships that are considered in this study. Codon usage bias in highly expressed genes is promoted by rapid generation times (solid black arrow; Vieira-Silva & Rocha 2010) and possibly by other traits. We hypothesize that protein elemental sparing might be promoted by a similar set of organismal traits (dotted black arrow) and that this potentially leads to an indirect association among organisms between the intensities of codon usage bias and 'elemental sparing' in the products of highly expressed genes.

Adaptive codon usage bias is therefore expected to be greater in organisms with large effective population sizes and tends to be greater in fast-growing organisms where selection for rapid translation is likely to be relatively intense (Rocha 2004; Subramanian 2008; Vieira-Silva & Rocha 2010).

Here, we study elemental sparing in highly expressed proteins across 148 bacterial genera and ask whether bacteria exhibiting adaptive codon usage bias in a set of highly expressed genes (Sharp *et al.* 2005) also exhibit 'elemental sparing' in the encoded proteins (Fig. 1). We find a greater tendency for carbon and sulphur sparing in bacteria that experience strong selection to use preferred codons. This study shows that adaptive elemental sparing influences the evolution of highly expressed proteins in diverse bacteria, and in ways that are likely mediated by the effectiveness of selection in different organisms. At a broader level, it also implies that adaptive elemental sparing occurs in many, but not all, bacteria.

Methods

Calculations of material costs and codon usage bias

We collected nucleotide and protein sequences from the KEGG (Kanehisa & Goto 2000) ftp site (ftp://ftp.genome.jp/pub/kegg/). For each protein-coding gene sequence, we calculated codon usage frequencies. For each protein, we calculated the C, N and S costs, as the numbers of C, N and S atoms (respectively) per amino acid.

We used codon frequency data for different genomes to calculate the S_C index of Sharp *et al.* (2005) (denoted S_C to avoid confusion with sulphur, S), which describes

the strength of selection associated with using preferred codons in a set of translation genes that tend to be highly expressed. This value, S_C, is an index of the intensity of codon usage bias in these highly expressed genes, relative to the rest of the genes in a genome. It is based on the frequency of usage of four 'wobble' pairs of synonymous codons, in which one codon (the preferred codon) is a perfect complement to the tRNA anticodon, and the other codon (unpreferred) is not, meaning that the preferred codon is likely to be recognized better by a single tRNA species. The four wobble pairs of synonymous codons are as follows: Asn codons AAU/AAC; Ile codons AUU/AUC; Phe codons UUU/UUC; and Tyr codons UAU/UAC. In each case, the codon ending in C is preferred. For a given genome, S_C was calculated for each wobble codon pair as

$$S_C = \ln\left[\frac{P_T}{1-P_T} \times \frac{1-P_A}{P_A}\right]$$

where P_T represents the average frequency of usage of the preferred codon in the translation genes and P_A represents the average frequency of usage of the preferred codon across all genes in the genome. The S_C value of the genome was calculated as the mean value for the four wobble codon pairs, weighted by the sum of preferred + unpreferred codons in the translation genes for the respective wobble pairs.

In addition to S_C, there are other statistics that describe codon usage bias (e.g. Sharp & Li 1987; Wright 1990). We chose the S_C index because several properties make it suitable for comparing diverse genomes. First, it takes into account mutational biases that vary among genomes and can lead to neutral variation in codon usage (Sharp *et al.* 2005). Second, the wobble codon pairs represent an unusual case where the same codon is likely to be preferred across many species (Sharp *et al.* 2005; see Hershberg & Petrov 2009; for a recent study of optimal codons among organisms). These properties mean S_C can be applied to all bacterial species and can be used to compare the intensity of codon usage bias in highly expressed genes among species that vary widely in their overall genomic composition (Sharp *et al.* 2005). (in Data S1, Supporting information, we confirm that S_C varies independently of base composition in our data set). See Sharp *et al.* (2005) for a detailed description of the properties of the S_C index, including possible effects of strand bias and other genome features. We calculated the S_C index using a set of 40 translation genes (Data S1, Supporting information) that are expected to be highly expressed (based on the list of Sharp *et al.* 2005; see Data S1, Supporting information) and that have predicted homologs in the 395 bacterial genomes that we analysed in the present study. For each of the genomes we analysed, we also calculated average protein C, N and S content (per amino acid; weighted by protein length) (i) for all encoded proteins and (ii) for the set of translation proteins.

We tested the correlations between the S_C index of codon usage bias and protein material costs using Kendall's τ rank correlation. In some cases, bacteria used in this study were closely related. This can lead to undesirable consequences in statistical analyses, including violation of the assumption of independence (see Harvey & Pagel 1991). We used two approaches to account for possible effects of phylogenetic nonindependence in these analyses. First, we aggregated the data set by calculating the mean values of the S_C index and protein material costs for all organisms in each genus (based on KEGG Taxonomy), and then performed analyses among genera. This is a simple approach to reduce the number of data points (from $n = 395$ genomes to $n = 148$ genera), and the number of degrees of freedom in statistical analyses, according to taxonomic information. Second, we tested the relationships by fitting generalized estimating equations (GEE) that explicitly incorporated information from a phylogenetic tree (Paradis & Claude 2002). Here, we present analyses among genera, because they are relatively simple and easily understood. In Data S1 (Supporting information) we show that the results are supported by the GEE analyses and by analyses that use all bacterial genomes as separate data points. Statistics were performed using R 2.12.2 (R Development Core Team, 2009).

We obtained generation times for 147 bacterial species from a study by Vieira-Silva & Rocha (2010), who compiled these values from the literature (see references cited in Vieira-Silva & Rocha 2010). We tested the relationships between protein material costs and generation times among 100 bacterial genera using Kendall's τ rank correlation.

Results

Material costs of ribosomal proteins

Among 148 bacterial genera, the mean C, N and S costs of translation proteins were correlated negatively with the S_C index of codon usage bias (C, $\tau = -0.18$, $P = 0.001$; N, $\tau = -0.12$, $P = 0.036$; S, $\tau = -0.21$, $P < 0.001$; Fig. 2). This means that bacteria with strong codon usage bias in the translation genes (large S_C values) also had low C, N and S costs in the proteins encoded by these genes.

Elemental sparing in highly expressed proteins

Next, we wanted to know whether the material costs of these translation proteins were *disproportionately* low

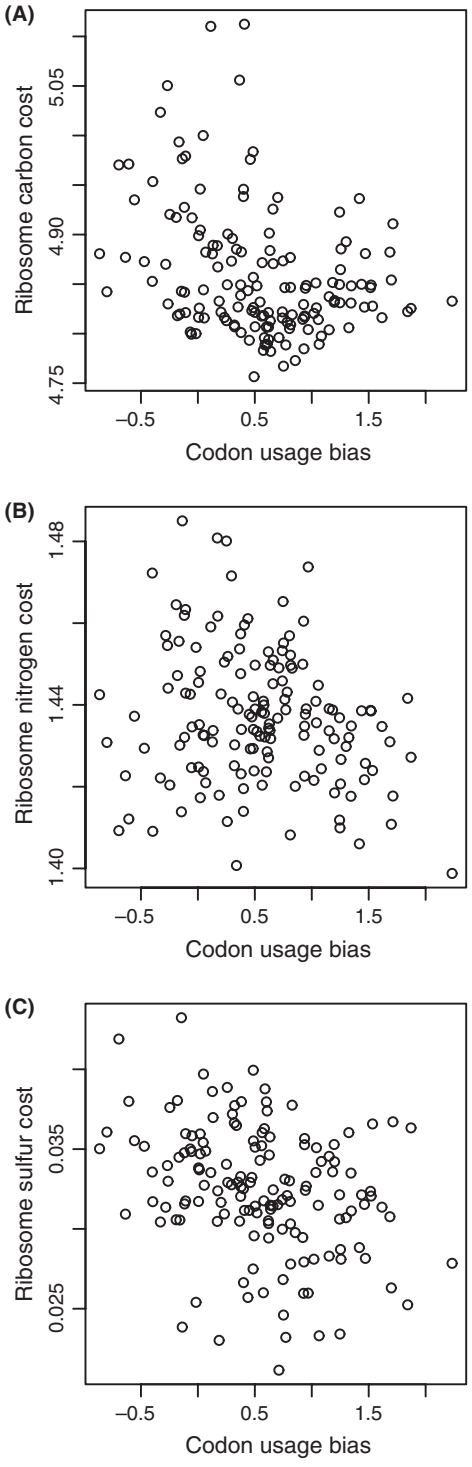

Fig. 2 The average number of (A) carbon, (B) nitrogen and (C) sulphur atoms per amino acid in 40 highly expressed proteins (vertical axis) plotted as a function of codon usage bias (the S_C index, horizontal axis) for 148 bacterial genera.

relative to the whole proteome, in organisms with strong codon usage bias in the translation genes (large S_C values). It was important to ask this question because the mean C, N and S costs of the translation proteins were associated positively with the mean C, N and S costs of whole proteomes ($n = 148$; for C, $\tau = 0.64$, $P < 0.001$; for N, $\tau = 0.46$, $P < 0.001$; for S, $\tau = 0.48$, $P < 0.001$). That is, we needed to test whether the material costs of whole proteomes tended to be smaller in organisms with strong codon usage bias (large values of the S_C index) or whether 'elemental sparing' was occurring specifically in translation proteins.

We asked this question using several related approaches. Each approach suggests that the C and S costs, but not the N costs, of the translation proteins tend to be disproportionately small (relative to whole proteomes) in organisms with strong codon usage bias in the translation genes (large S_C values). We began by considering the correlations between the material costs of whole proteomes and the S_C index of codon usage bias. The mean C and S costs of whole proteomes were not associated significantly with the S_C index of codon usage bias ($n = 148$; for C, $\tau = -0.06$, $P = 0.26$; for S, $\tau = -0.05$, $P = 0.36$). In contrast, the mean N costs of whole proteomes were correlated negatively with the S_C index of codon usage ($n = 148$; for N, $\tau = -0.22$, $P < 0.001$).

We next calculated residual deviations from the linear regressions of C, N and S costs of translation proteins on whole proteome C, N and S costs, respectively. These residuals describe the C, N and S costs of the translation proteins relative to the material costs of the whole proteome and thus are an index of 'elemental sparing' in the translation proteins. Large negative residuals would indicate that translation proteins had relatively small elemental costs (strong elemental sparing), and positive residuals would indicate that translation proteins had relatively large elemental costs. For carbon and sulphur, the residuals exhibited negative correlations with the S_C index of codon bias (C, $\tau = -0.29$, $P < 0.001$; S, $\tau = -0.25$, $P < 0.001$; Fig. 3). For N, the residuals were uncorrelated with codon bias ($\tau = 0.04$, $P = 0.46$; Fig. 3). These analyses suggest that bacteria with strong codon usage bias in translation proteins tend to exhibit C and S sparing in the same proteins, but not N sparing.

We confirmed these observations using GEE, in which the material cost of the translation proteins was modelled as a function of the S_C index, whole proteome material costs, and a matrix representation of phylogenetic relatedness among genomes (Paradis & Claude 2002; see Data S1, Supporting information).

Elemental sparing and life history

Finally, we asked whether elemental sparing in translation proteins is associated with generation time among

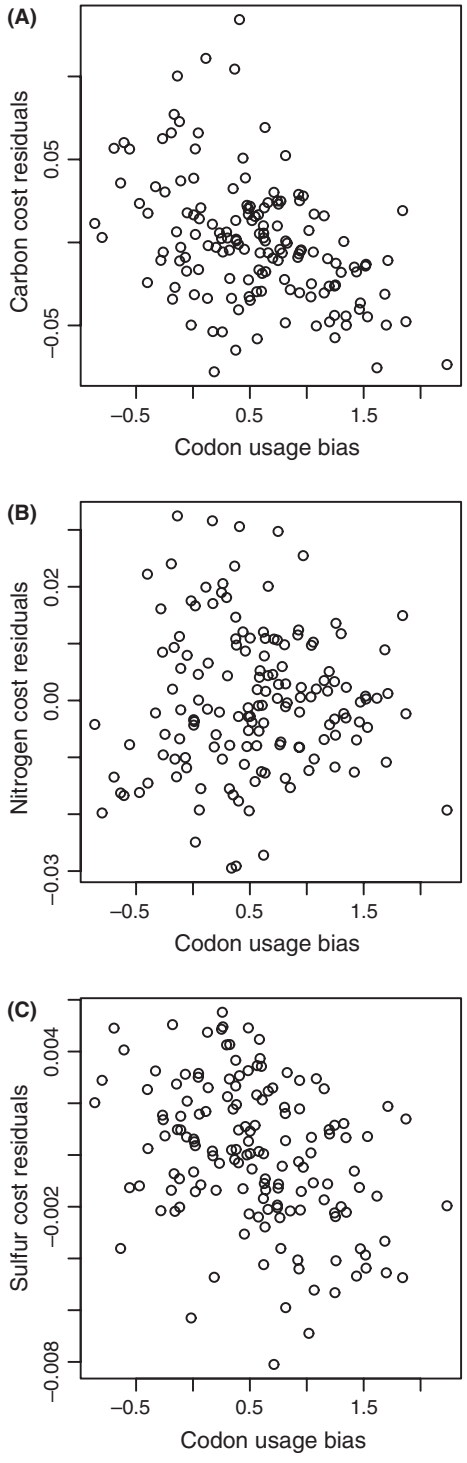

Fig. 3 Residuals describing the (A) carbon, (B) nitrogen and (C) sulphur content of highly expressed proteins relative to the rest of the proteome (vertical axis) as a function of codon usage bias (the S_C index, horizontal axis) for 148 bacterial genera.

bacteria. Generation time is an important life history trait that is associated negatively with the S_C index of codon usage (Vieira-Silva & Rocha 2010). This means that bacteria capable of rapid growth (short generation times) tend to exhibit stronger codon usage bias in translation proteins. Here, we tested the correlations between generation times and elemental sparing among 100 bacterial genera, again using residual deviations from linear regressions of translation protein material costs on whole proteome material costs as indices of elemental sparing. For C and S, these residuals were weakly and positively associated with generation time ($n = 100$; for C, $\tau = 0.24$, $P = 0.0004$; for S, $\tau = 0.23$, $P = 0.0008$; Fig. 4). This means that bacteria with shorter generation times (faster growth) tended to exhibit stronger C and S sparing or had translation proteins with C and S costs that were small relative to the C and S costs of their proteomes. There was no significant correlation between generation time and N cost residuals ($n = 100$, $\tau = 0.08$, $P = 0.23$; Fig. 4).

Discussion

In some organisms, the material costs of highly expressed proteins tend to be smaller than the material costs of other proteins. These differences likely arise because selection favours reduced quantities of ecologically limiting elements in highly expressed proteins (Fauchon *et al.* 2002; Elser *et al.* 2006; Bragg & Wagner 2007; Li *et al.* 2009). However, it has previously not been clear whether this 'elemental sparing' in highly expressed proteins is ubiquitous or rare among different kinds of bacteria, or whether the intensity of elemental sparing is related systematically to any other traits. Our study thus represents an important step forward, by showing that among diverse bacteria, carbon sparing and sulphur sparing in highly expressed translation proteins are associated with strong codon usage bias in the encoding genes. We suggest that these associations occur because material sparing and codon usage bias can each only evolve adaptively in organisms where selection is effective in determining the fate of mutations with small selection coefficients. This means that the associations between adaptive elemental sparing and codon usage bias might arise indirectly, if both are mediated by similar life history and population genetic traits (Fig. 1). This could include traits such as large effective population size, which promotes the effectiveness of selection. It might also include phases of nutrient-limited growth that are initiated by pulses of nutrient supply (Rocha 2004; Subramanian 2008; Vieira-Silva & Rocha 2010). During rapid growth, mutations that reduce the rapidity or accuracy of translation of highly expressed genes are likely to reduce fitness. This potentially includes mutations that change a preferred codon to an unpreferred codon. It potentially also includes mutations that increase the quantity of a

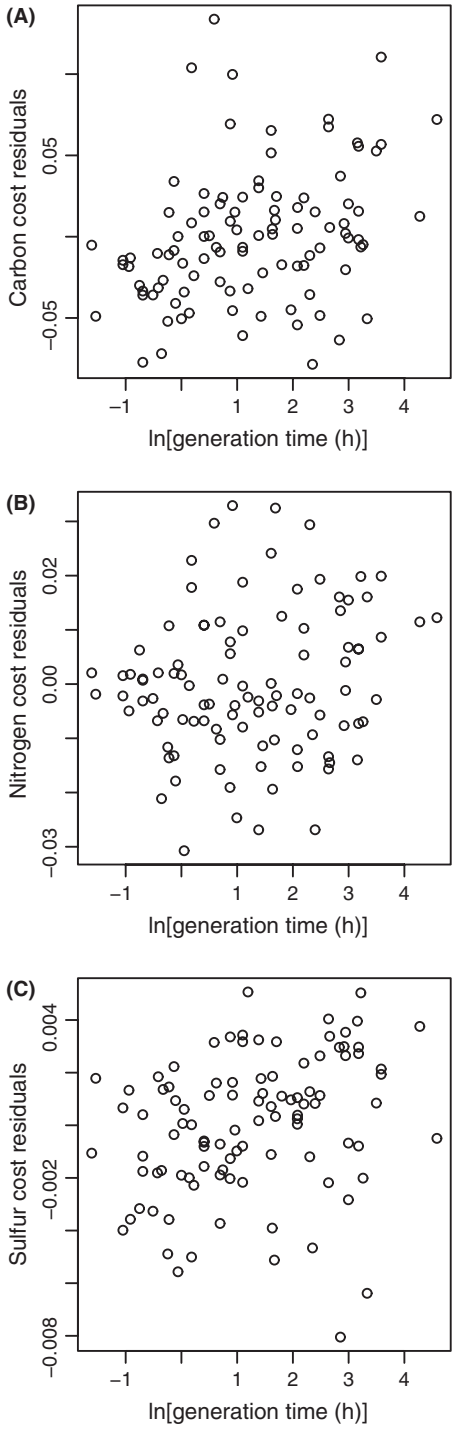

Fig. 4 Residuals describing the (A) carbon, (B) nitrogen and (C) sulphur content of highly expressed proteins relative to the rest of the proteome (vertical axis) as a function of generation time (in hours, $\log_e$-transformed) (horizontal axis) for 100 bacterial genera.

growth-limiting element that is needed to make a highly expressed protein (Bragg & Wagner 2009). We find limited support for the notion that adaptive codon usage bias and elemental sparing are related to common life history traits by showing that carbon sparing and sulphur sparing are weakly correlated with generation time, a trait that is known to be associated with codon usage bias (Rocha 2004; Subramanian 2008; Vieira-Silva & Rocha 2010). In sum, our observations suggest that adaptive elemental sparing in highly expressed proteins (i) occurs in some, but not all, bacteria and (ii) might be promoted by bacterial traits that also promote the adaptive evolution of codon usage biases. In future, we hope that a better understanding of the association between elemental sparing and codon usage will lead to a better mechanistic understanding of elemental sparing, possibly through the application of theoretical tools that have been used to study codon usage bias (Bulmer 1991; Sharp *et al.* 2005).

Selection for a mutant version of a protein that contains a smaller quantity of an element, but is otherwise similar to the wild-type version, is only possible if the element is limiting to growth (Bragg & Wagner 2009). Sulphur is rarely considered a limiting nutrient in the environment (see reviews: Giordano *et al.* 2008; Zhao *et al.* 2008), but our observations are consistent with a number of previous studies which have shown that protein sulphur content can evolve in response to sulphur scarcity. First, sulphur sparing has been observed in proteins that are induced during sulphur starvation (Mazel & Marlière 1989; Baudouin-Cornu *et al.* 2001), and also following cellular events that require substantial reallocation of acquired sulphur (*e.g.* Fauchon *et al.* 2002). These observations demonstrate that sulphur limitation can exert a powerful selective force on protein evolution and that sulphur stress can occur even when sulphur is plentiful in the environment. Second, amino acid substitutions that change the sulphur content of a highly expressed protein by one atom can have a relatively large impact on the total sulphur budget of a cell (relative to substitutions affecting one carbon or nitrogen atom), meaning that if sulphur limitation occurs, there is substantial scope for selection to act on mutations that change protein sulphur costs (Bragg & Wagner 2009).

Among the proteins encoded in a single genome, protein carbon costs can be strongly correlated with the energetic costs of amino acid biosynthesis (Quigg & Beardall 2003; Bragg & Wagner 2007; Li *et al.* 2009). This means that protein carbon costs might act as a surrogate of protein energy costs, which also tend to be smaller in highly expressed proteins than in whole proteomes in some organisms (Akashi & Gojobori 2002; Heizer *et al.* 2006). Here, we have avoided calculating and analysing the energetic costs of amino acids explicitly, because energetic costs of biosynthesis are strongly dependent on metabolic pathways used by different organisms, including variations in the nitrogen source,

and vary greatly among the bacteria represented in our study.

We did not find an association between codon usage bias and nitrogen sparing in translation proteins. The nitrogen content of translation proteins and whole proteomes were each correlated negatively with codon usage bias, such that translation proteins did not have disproportionately low nitrogen content relative to whole proteomes in organisms with strong selection for codon usage bias. Previous studies of microbes (Bragg & Wagner 2007; Li *et al.* 2009) and plants (Elser *et al.* 2006) have found that highly expressed proteins tend to have smaller nitrogen costs than other proteins. However, it has also been shown that ribosomal proteins tend to be rich in nitrogen (Acquisti *et al.* 2009; Gilbert & Fagan 2011), possibly due to constraints of ribosome function and the biochemical properties of nitrogen-rich amino acids (such as Arg and His, Acquisti *et al.* 2009). It is possible that these functional constraints reduce the scope for adaptive sparing of nitrogen in ribosomal proteins.

Evolutionary pressures for nutrient investment (carbon, nitrogen, sulphur and phosphorus) probably fundamentally constrained the evolution of the composition of cells and molecules (Quigg *et al.* 2003, 2011). A growing body of evidence suggests that the costs of making proteins can evolve adaptively. However, we currently have few unifying principles to help us understand how these costs evolve, and in which organisms selection on protein material costs is likely to occur. In future, we hope it will be possible to establish a mechanistic understanding of how protein costs evolve, with a basis in population genetics, ecology and functional genomics. The present study represents a step towards this goal, by showing that the tendency for adaptive evolution of protein material costs is related to the strength of codon usage bias among bacteria and by suggesting that similar features of organisms promote these traits.

Acknowledgements

This research grew from discussions at Working Group 33 of the Australian Research Council Network for Vegetation Function. Dundee University is a Scottish Registered Charity, No. SC015096.

References

Acquisti CA, Kumar S, Elser JJ (2009) Signatures of nitrogen limitation in the elemental composition of the proteins involved in the metabolic apparatus. *Proceedings of the Royal Society B*, **276**, 2605–2610.

Akashi H, Gojobori T (2002) Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilus*. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 3695–3700.

Baudouin-Cornu P, Surdin-Kerjan Y, Marlière P, Thomas D (2001) Molecular evolution of protein atomic composition. *Science*, **293**, 297–300.

Boer VM, de Winde JH, Pronk JT, Piper MDW (2003) The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus or sulfur. *Journal of Biological Chemistry*, **278**, 3265–3274.

Bragg JG, Wagner A (2007) Protein carbon content evolves in response to carbon availability and may influence the fate of duplicated genes. *Proceedings of the Royal Society B*, **274**, 1063–1070.

Bragg JG, Wagner A (2009) Protein material costs: single atoms can make an evolutionary difference. *Trends in Genetics*, **25**, 5–8.

Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics*, **129**, 897–907.

Crick FHC (1958) On protein synthesis. *The Symposia of the Society for Experimental Biology*, **12**, 138–163.

Elser JJ, Fagan WF, Subramanian S, Kumar S (2006) Signatures of ecological resource availability in the animal and plant proteomes. *Molecular Biology and Evolution*, **23**, 1946–1951.

Fauchon M, Lagniel G, Aude JC *et al.* (2002) Sulfur sparing in the yeast proteome in response to sulfur demand. *Molecular Cell*, **9**, 713–723.

Gilbert JDJ, Fagan WF (2011) Contrasting mechanisms of proteomic nitrogen thrift in *Prochlorococcus*. *Molecular Ecology*, **20**, 92–104.

Giordano M, Norici A, Ratti S, Raven JA (2008) Role of sulfur in algae: acquisition, accumulation, metabolism, ecology and evolution. In: *Sulfur Metabolism in Phototrophic Organisms* (eds Hell R, Dahl C, Knaff D, Leustek T), pp. 397–415. Springer, Dordrecht.

Gouy M, Gautier C (1982) Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Research*, **10**, 7055–7074.

Grantham R, Gautier C, Gouy M, Mercier R, Pavé A (1980) Codon catalog usage and the genome hypothesis. *Nucleic Acids Research*, **8**, 49–62.

Harvey PH, Pagel MD (1991) *The Comparative Method in Evolutionary Biology*. Oxford University Press, Oxford, UK.

Heizer EM Jr, Raiford DW, Raymer ML *et al.* (2006) Amino acid cost and codon-usage biases in 6 prokaryotic genomes: a whole-genome analysis. *Molecular Biology and Evolution*, **23**, 1670–1680.

Hershberg R, Petrov DA (2009) General rules for optimal codon choice. *PLoS Genetics*, **5**, e1000556.

Ikemura T (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *Journal of Molecular Biology*, **151**, 389–409.

Kanehisa M, Goto S (2000) KEGG: Kyoto encyclopaedia of genes and genomes. *Nucleic Acids Research*, **28**, 27–30.

Li N, Lv J, Niu D-K (2009) Low contents of carbon and nitrogen in highly abundant proteins: evidence of selection for the economy of atomic composition. *Journal of Molecular Evolution*, **68**, 248–255.

Mazel D, Marlière P (1989) Adaptive eradication of methionine and cysteine from cyanobacterial light-harvesting proteins. *Nature*, **341**, 245–248.

Paradis E, Claude J (2002) Analysis of comparative data using generalized estimating equations. *Journal of Theoretical Biology*, **218**, 175–185.

Quigg A, Beardall J (2003) Protein turnover in relation to maintenance metabolism at low photon flux in two marine microalgae. *Plant Cell and Environment*, **26**, 693–703.

Quigg A, Finkel ZV, Irwin AJ *et al*. (2003) The evolutionary inheritance of elemental stoichiometry in marine phytoplankton. *Nature*, **425**, 291–294.

Quigg A, Irwin AJ, Finkel ZV (2011) Evolutionary inheritance of elemental stoichiometry in phytoplankton. *Proceedings of the Royal Society B*, **278**, 526–534.

R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria ISBN 3-900051-07-0, URL http://www.R-project.org.

Raven JA, Kübler JE, Beardall J (2000) Put out the light, and then put out the light. *Journal of the Marine Biology Association of the United Kingdom*, **80**, 1–27.

Rocha EPC (2004) Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Research*, **14**, 2279–2286.

Sharp PM, Li W-H (1987) The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, **15**, 1281–1295.

Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE (2005) Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Research*, **33**, 1141–1153.

Sharp PM, Emery LR, Zeng K (2010) Forces that influence the evolution of codon bias. *Philosophical Transactions of the Royal Society B*, **365**, 1203–1212.

Sterner RW, Elser JJ (2002) *Ecological Stoichiometry*. Princeton University Press, Princeton, New Jersey.

Subramanian S (2008) Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes. *Genetics*, **178**, 2429–2432.

Tsai IJ, Bensasson D, Burt A, Koufopanou V (2008) Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle. *Proceedings of the National Academy of Sciences of the United States of America*, **2008**, 4957–4962.

Van der Ploeg JR, Weiss MA, Saller E *et al*. (1996) Identification of sulfate starvation-regulated genes in *Escherichia coli*: a gene cluster involved in the utilization of taurine as a sulfur source. *Journal of Bacteriology*, **178**, 5438–5446.

Vieira-Silva S, Rocha EPC (2010) The systematic imprint of growth and its uses in ecological (meta)genomics. *PLoS Genetics*, **6**, 1000808.

Wright F (1990) The 'effective number of codons' used in a gene. *Gene*, **87**, 23–29.

Zhao F-J, Tausz M, De Kok LJ (2008) Role of sulfur for plant production in agricultural and natural ecosystems. In: *Sulfur Metabolism in Phototrophic Organisms* (eds Hell R, Dahl C, Knaff D, Leustek T), pp. 417–435. Springer, Dordrecht.

Data accessibility

Genome sequences: http://www.genome.jp/kegg/, accessed via ftp://ftp.genome.jp/pub/kegg/ 27 July 2008. Bacterial 16S ribosomal RNA gene alignment: http://www.arb-silva.de, accessed 30 August 2011. Growth rates: Vieira-Silva and Rocha EPC (2010), doi:10.1371/journal.pgen.1000808, accessed 3 July 2010.

Supporting information

Additional supporting information may be found in the online version of this article.

Data S1 Supplementary information and calculations.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.