

# Stoichiogenomics: the evolutionary ecology of macromolecular elemental composition

James J. Elser<sup>1</sup>, Claudia Acquisti<sup>1,2\*</sup> and Sudhir Kumar<sup>1,2</sup>

<sup>1</sup> School of Life Sciences, Arizona State University, Tempe, AZ 85287-4501, USA

<sup>2</sup> Center for Evolutionary Medicine and Informatics, Biodesign Institute, Arizona State University, Tempe, AZ 85287-5301, USA

**The new field of ‘stoichiogenomics’ integrates evolution, ecology and bioinformatics to reveal surprising patterns of the differential usage of key elements [e.g. nitrogen (N)] in proteins and nucleic acids. Because the canonical amino acids as well as nucleotides differ in element counts, natural selection owing to limited element supplies might bias monomer usage to reduce element costs. For example, proteins that respond to N limitation in microbes use a lower proportion of N-rich amino acids, whereas proteome- and transcriptome-wide element contents differ significantly for plants as compared with animals, probably because of the differential severity of element limitations. In this review, we show that with these findings, new directions for future investigations are emerging, particularly via the increasing availability of diverse metagenomic and metatranscriptomic data sets.**

## Natural selection, primary sequence evolution, and the role of key resources

During the past 60 years of DNA-centered research, a multitude of forces have been identified that shape the evolutionary landscape of molecules of inheritance (nucleic acids) and their downstream products (e.g. proteins and ribosomal RNA). These forces act on individual mutations in the blueprint of life (primarily DNA, except for a few viruses), which are subjected to the sieve of natural selection and to the whims of random genetic drift within populations. Most mutations at functionally important DNA positions experience negative selective pressures either directly or via the product that they encode [1–3]. Classically, this natural selection acting directly on DNA and protein sequences has been concerned with the requirements of biochemical function, structure, and/or the cellular milieu of DNA and protein sequences, including tRNA abundance and energy synthesis costs.

Historically, there has been a lack of appreciation and understanding of natural selection on the genome sequence of an organism exerted by its ecological context (i.e. the environment in which an organism lives). In some interpretations of the classical framework, particularly in the context of eukaryotic and multicellular biota, it is as

though the primary structure of the genome were insulated from ecological and environmental factors. However, eco-environmental natural selection can act directly on genomes because the availability and synthesis of the genome and protein building blocks (nucleotides and amino acids) are dictated by the presence of key macroelements [e.g. nitrogen (N)] in the immediate environment. Mutations that place a greater demand on limited elemental resources will experience purifying selection, both for ‘junk’ and the functionally important positions that they afflict. This selective pressure is expected to be low but its cumulative effects are potentially detectable if the limitation persists over time. They should be discernable by comparing genomes of species with and without resource-limited environments, because the fixation of natural variation will be under the control of purifying selection, albeit mild, and random genetic drift in the two genomes, respectively.

Some effects of environmental limitations on the primary structure of the genome (i.e. its linear DNA sequence) have been previously documented in prokaryotes and single-celled eukaryotes [4,5], perhaps because in those species each cell is directly exposed to the environment. However, these studies tend to emphasize the direct effects of the environment of the organism *per se* (e.g. direct physico-chemical impacts [6]) rather than the ecological context of its resource supplies. Indeed, how nutrient and energy supplies shape protein evolution in microbes and multicellular organisms had remained largely unexplored until recently.

Here, we consider resource-driven DNA and protein evolution within the broader framework of biological stoichiometry, which is the study of balance of energy and

## Glossary

**Biological stoichiometry:** the study of the balance of energy and multiple chemical elements in living systems.

**EST (expressed sequence tag):** a short sub-sequence of transcribed DNA. Generally, these are analyzed by cloning the associated mRNAs and, thus, can be rapidly assessed and quantified, enabling at least a relative comparison of expression levels of different genes.

**MUB (monomer usage bias):** a statistically disproportionate use of individual amino acids (or classes of amino acids) or nucleotides in a protein or nucleic acid sequence compared with some measure of random (equiproportional) use of that monomer.

**UTR (untranscribed region):** segments of DNA that flank a gene on the ‘upstream’ (5’ UTR) or the ‘downstream’ end (3’ UTR).

**Stoichiogenomics:** the study of the elemental composition of macromolecules (protein, DNA, RNA, etc.) using genomics data and bioinformatics tools.

Corresponding author: Elser, J.J. (j.elser@asu.edu)

\* Current address: Institute for Evolution and Biodiversity, University of Muenster, 48149 Muenster, Germany.

multiple chemical elements in living systems (see [Glossary](#)) [7]. We evaluate the direct impact of organismal resource constraints on molecular evolution by examining evidence for the existence of element-based monomer usage biases (MUB) in fundamental biological molecules: proteins and nucleic acids. Such biases can arise because different monomers (amino acids and nucleotides) differ substantially in the number of key macroelements [e.g. N, sulfur (S) and carbon (C)] they contain. Often the environmental supplies of such elements limit the growth and reproduction of organisms. If such limitations are sufficiently strong or prolonged during the evolutionary history of a species, we predict that selection would work to reduce the use of these limiting elements as they will hamper production of major biopolymers. Thus, the over-riding theme of this review is an exploration of the stoichiometric signature of environmental resource limitation on the primary structure of life's most important molecules.

### Evidence for element-driven MUB in proteins

The evolutionary costs and benefits of producing a protein involve the interplay of the physiological and ecological domains of biology. Many biota (especially plants, algae and microbes) face direct limitation of their growth and reproduction by insufficient supplies of not only energy, but also key nutrient elements (such as C, N and S) needed to construct proteins and nucleic acids. Importantly, the bioenergetic costs of synthesis differ for the various amino acids [8] as do the number of C, N and S atoms needed to build them. This suggests that selection leading to usage bias of different monomers in proteins occurs based on not only the energetic costs of construction of different amino acids, but also their element costs. Furthermore, the ecological environment experienced by different biota will also have a role in determining the strength of selection for increased efficiency of resource use (energy or element) in different contexts. Finally, we expect this selection to act on some molecules more than others, such as those that are abundant in the cell and those that are used in different metabolic pathways or structures that are closely associated with conditions of nutrient limitation. Here, we discuss some emerging lines of evidence relevant to these hypotheses.

### Amino acid usage is element biased in microbes

Many initial analyses on biased amino acid usage involved microorganisms. For example, an early study [9] showed element-based MUB affecting use of S-containing cysteine as a means to avoid uptake of toxic molecules (selenate and chromate). Subsequently, evidence for element-based MUB associated with direct element limitations in the environment was obtained in proteome-wide studies of microorganisms [10] in which the element costs of enzymes involved in the acquisition and processing of a given chemical element (C, N, or S) were compared with the element costs of the overall proteome. The analysis revealed that the enzymes involved in the processing of a given chemical element have MUB that results in lower use of that element ([Figure 1a](#)). For example, the C-assimilatory enzymes in *Escherichia coli* and *Saccharomyces cerevisiae* are unusually low in C compared with the overall proteomes (and with the enzymes involved in processing other elements). Similarly,

N-assimilatory enzymes are low in N compared with the overall proteome in *S. cerevisiae*. These observations are interpreted to be adaptive responses to facilitate construction of the cellular components that are most needed under conditions of resource limitation of a given element.

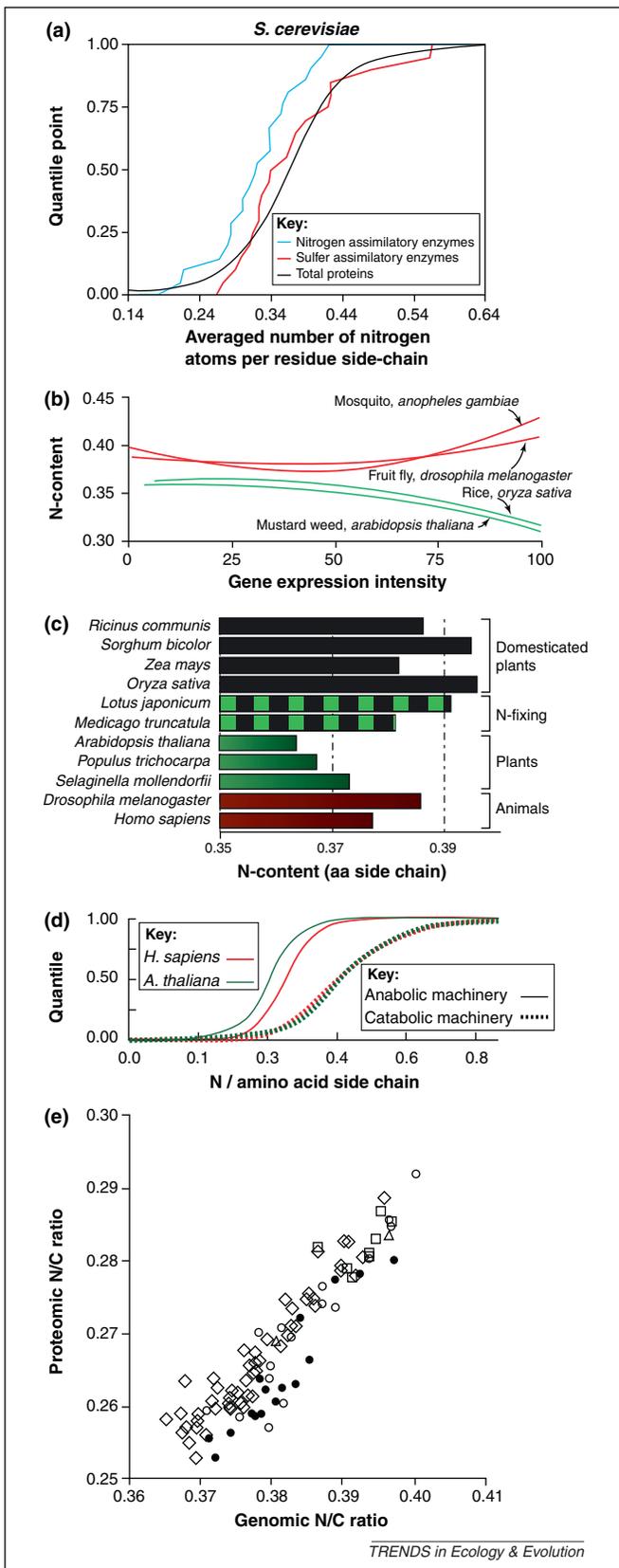
This conclusion is supported by recent work by J. Bragg and A. Wagner [11], who evaluated the visibility of element costs to natural selection for conservation of C, N, S and P in *Saccharomyces*. They did so by accounting for the element costs of changes in mRNA production and cellular protein levels associated with all the genes in the genome. They concluded that, under the respective forms of elemental limitation, mutations leading to a doubling in expression are visible to natural selection for > 90% of the genes in the yeast genome. That is, such mutations resulted in selection coefficients that were ten times larger than the selection coefficient for which natural selection and drift have a similar effect on the fate of a mutation.

Further work corroborating such findings comes from R.P. Carlson [12], who performed whole-metabolism pathway analysis in *E. coli* under conditions of nutrient sufficiency and nutrient (N) deficiency. In the whole-metabolism pathway analysis, both actual and potential alternative flux pathways leading among metabolic reactants are quantified and compared using available metabolic maps and evaluated for different growth conditions. He notes that the biochemical networks of *E. coli* are close to optimized for thermodynamic (energetic) efficiency under nutrient-sufficient conditions; however, under nutrient limitation, *E. coli* expresses pathways (such as the Entner-Doudoroff pathway) that have low proteome synthesis costs (reflecting the accumulated effects of differences in protein copy numbers, protein lengths and monomer N contents) and that operate inefficiently in energetic terms. More analyses of this kind seem promising, given emerging data from functional genomics.

### Amino acid usage is also element biased in multicellular organisms

This work on protein elemental composition was extended to higher organisms by J.J. Elser and colleagues [13], focusing especially on N and comparing photoautotrophic organisms ('plants') and metazoan species. This analysis was motivated by an overall expectation that plants are likely to have experienced longer and more severe N limitation compared with animals. Thus, plants should show signs of N-targeted MUB. Consistent with this idea, across nine 'plant' species (two with complete genome sequences) relative to nine animal species (eight with whole-genome sequences), plant proteins used an average of 7.1% less N in their side chains compared with animal proteins. Although plant and animal proteins potentially differ in other important ways (such as average sequence length) and N-rich amino acids differ in many ways other than in their N content, many potential confounding factors were systematically evaluated and rejected [13]. Consequently, the observed difference has been interpreted to be a deep imprint of N limitation at the level of protein primary sequences.

To test these stoichiometric interpretations, Elser and colleagues examined the element content of proteins as a



**Figure 1.** Examples of element-based MUB in prokaryotes and eukaryotes. **(a)** N-processing enzymes (blue line) in yeast (*Saccharomyces cerevisiae*) are unusually low in N compared with the average protein in the proteome (black line) and with the average protein involved in sulfur-processing (red line) [10]. This suggests a form of 'element sparing' in molecules that are most important during limitation by a given element. **(b)** Protein N content declines with gene expression intensity in plants (green line) but not in animals (red line) [13], an outcome that is consistent with the fact that plants are likely to experience more frequent and sustained direct N limitation than are animals. **(c)** Proteome N content in domesticated plants (black

bars) and in plants associated with N-fixing symbionts (green-and-black bars) is similar to that found in animals (red bars) and higher than that found in undomesticated plants (green bars) [15]. This suggests that release from selection pressures for N conservation has reduced the MUB for N conservation in crop plants. Error bars (standard errors) are very small (max: ~0.001) and thus were not included, as they are not discernible on the plot. **(d)** The proteins involved in anabolic machinery (spliceosome and ribosomes; solid black line) have a higher N content than do proteins associated with catabolic machinery (proteasome, lysosome and vacuole; dashed black line) for *Arabidopsis thaliana* (green line) and *Homo sapiens* (red line) [16], a result that is consistent with an interpretation of N sparing during nutrient limitation when catabolic processes dominate. **(e)** In microorganisms, the N:C ratio of the proteome of a taxon is positively correlated with the N:C ratio of its genome [17], suggesting that selection for N conservation was operating during early evolution of the canonical genetic code. Key: filled circles, Archaea; open symbols, Bacteria (circle, free-living only; diamond, capable of living as animal symbionts; square, capable of living as plant symbionts; triangle, capable of living as animal and plant symbionts).

### MUB is weakened in crop plants

Domestication of crops, along with fertilization of cultivated soils, represents an ideal natural experiment to investigate the role of environmental N availability in shaping the element composition of proteins. The N conservation hypothesis predicts that the massive N enrichment by fertilization of cultivated soils is tantamount to removing the selection pressure exerted by N limitation in crop species. A similar effect is also expected for plants in symbiosis with N-fixing bacteria (e.g. legumes). Indeed, there are significant signs of decreased N conservation (thus, higher N content) in the proteomes of crops and legumes compared with undomesticated plants [15]. An analysis of the complete proteomes of four crops, two legumes and three undomesticated plants, has shown that the N content of amino acid residues is >7% higher in crops and legumes than in undomesticated plants (Figure 1c). Consistent with the hypothesis of N limitation, these observations suggest that the global N content of the plant proteomes is mainly driven by the importance of N limitation in the evolutionary history of a species, rather than merely a reflection of a phylogenetic group. For example, the dicots *Medicago truncatula* and *Lotus japonicum* have an N content that is more similar to the monocot crop species *Zea mays* and to the undomesticated dicot *Arabidopsis thaliana* (Figure 1c). This observation constitutes an example of convergent evolution in proteome composition brought on by the relaxation of the resource limitations in independent lineages.

### Catabolic proteins are especially low in nitrogen

The role of N limitation in shaping molecular evolution was further investigated by studying the elemental composition of the metabolic apparatus that is involved in the

cellular response to nutritional stress [16]. By examining the elemental composition of major functional classes of proteins involved in anabolic and catabolic processes in multicellular eukaryotic model organisms, it was found that ecophysiological selection for nutrient conservation specifically targets the cellular components of the catabolic machinery, which are highly expressed in response to nutrient limitation (Figure 1d). Thus, it appears that the RNA component of the anabolic machineries (i.e. ribosome and spliceosome) underpins the stoichiometric differences between the two apparatus. Given that N-rich amino acids tend to have a positive charge, the charge distribution of nucleotides requires a high N content in proteins that have a close physical interaction with nucleic acids. This purely functional point of view indicates high nutritional costs of the processes (such as transcription, translation and genome duplication) that operate during fast growth and nutrient sufficiency. It also reinforces the idea that resource availability and the optimization of nutrient allocation are important factors shaping the molecular architecture of cellular structures.

#### Element biases are deeply rooted in the genetic code

At least some of the causal basis of element-driven MUB appears to lie deep in the history of life, as evidenced by a positive association between the N:C ratios of amino acids and the N:C ratios of their corresponding nucleic acid codons in the canonical genetic code [17]. This result is suggested to reflect the intrinsic correlations of hydrophobicity of particular amino acids and the hydrophobicity of their corresponding codons. Strong correlations have also been documented of whole-proteome N:C ratios with whole-genome N:C ratio (correlation = 0.90) across 94 species of fully sequenced microorganisms (Figure 1e). These relationships reflect the underlying stoichiometric correlation between amino acids and codons along with the cross-species variation in genomic GC content and its known association with amino acid frequencies. As a result, average proteomic and genomic N use was positively correlated with GC content, whereas genomic and proteomic C use was negatively correlated with GC content.

These findings illustrate that, among microbes at least, there is a deep stoichiometric association of C and N use in the cellular machinery of life. However, the N-based MUB observed in plant proteomes [13] exists above and beyond any such effect of GC-content. It was noted [13] that the GC content of *A. thaliana* introns is 32%, whereas that of *Drosophila melanogaster* introns is 37%. Both of these differ from mouse (GC content = 45%), but the estimated N content of fruit fly and mouse proteomes are nearly identical (0.387 and 0.381, respectively). Thus, it is improbable that GC content differences are responsible for the observed differences in proteomic N content between plants and animals [13].

#### From amino acid to nucleotide usage biases

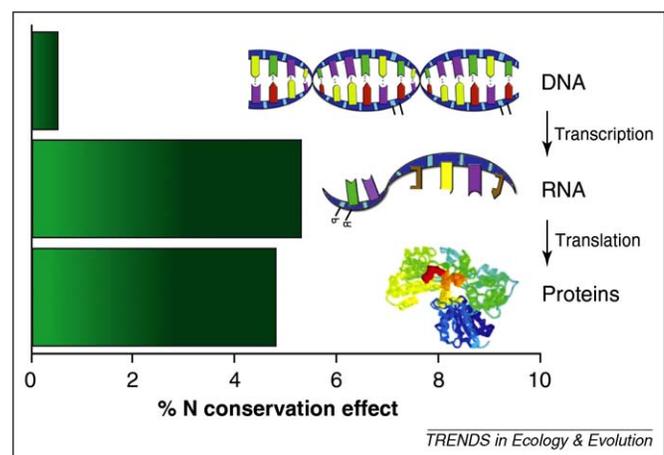
Nucleotide biosynthesis originates with amino acids, so element limitations that affect processing of amino acids should also affect the construction of intronic and other non-coding DNA that rely on these precursors [18]. Indeed,

genome-scale investigations of animals and plants have established a link between ecological resource limitations and the element composition of functionally important parts of the genome.

#### Element-based nucleotide biases also occur, especially in plants

The number of N atoms per monomer ranges from one to four in amino acids and from two to five in nucleotides, whereas phosphorus (P) is a fundamental component of nucleotides but not of amino acids. Therefore, an efficient logic of nutrient allocation predicts that nutrient limitations should influence not only the molecular composition of proteins, but also that of genes and transcripts, favoring monomer usage biases that conserve limiting elements in a coordinated manner from genomes to transcriptomes to proteomes. However, biases affecting N use are expected to be smallest in DNA because among-nucleotide variation is buffered by the complementarity rules of the double helix (A and T entail seven N atoms, whereas G and C entail eight N atoms). Single-stranded RNA shows a larger palette of possible N content per nucleotide with purines (adenine and guanine, five N atoms each) containing more N than do pyrimidines (cytosine has three N atoms and uracil has two N atoms). Furthermore, RNA generally contributes five to ten times more than DNA to cellular biomass [19]. Therefore, an efficient strategy of N and P conservation should primarily affect the transcribed portions of the genome.

Recent work [15] has shown that this is indeed the case, reporting that the local composition of introns in plants promotes overall N conservation via nucleotide biases in the transcribed strand (Figure 2). Thus, chronic nutrient limitation in plants has influenced the elemental composition of the transcriptome. Moreover, consistent with the predictions of natural selection for N conservation and consistent with previous observations of MUB in proteins, crop transcriptomes show higher N content than do those of 'wild' plants, reflecting relaxation of natural



**Figure 2.** Nitrogen conservation effect size (as a percentage) in the vascular plant *Arabidopsis thaliana* compared with humans *Homo sapiens* and the fruit fly *Drosophila melanogaster* in DNA, RNA and proteins [15]. In this analysis, the overall monomer N content for each molecule in *Arabidopsis* was compared with the monomer N content for that molecule averaged for both human and fruit fly. As expected, the apparent magnitude of N conservation in plants versus animals is weak for the genome (owing to base-pairing rules) but is strong for both the transcriptome and for the proteome.

selection for N conservation in response to the extensive use of N-rich fertilizers and other effects of domestication [15].

#### Nutrients can also be conserved by shortening polymer length

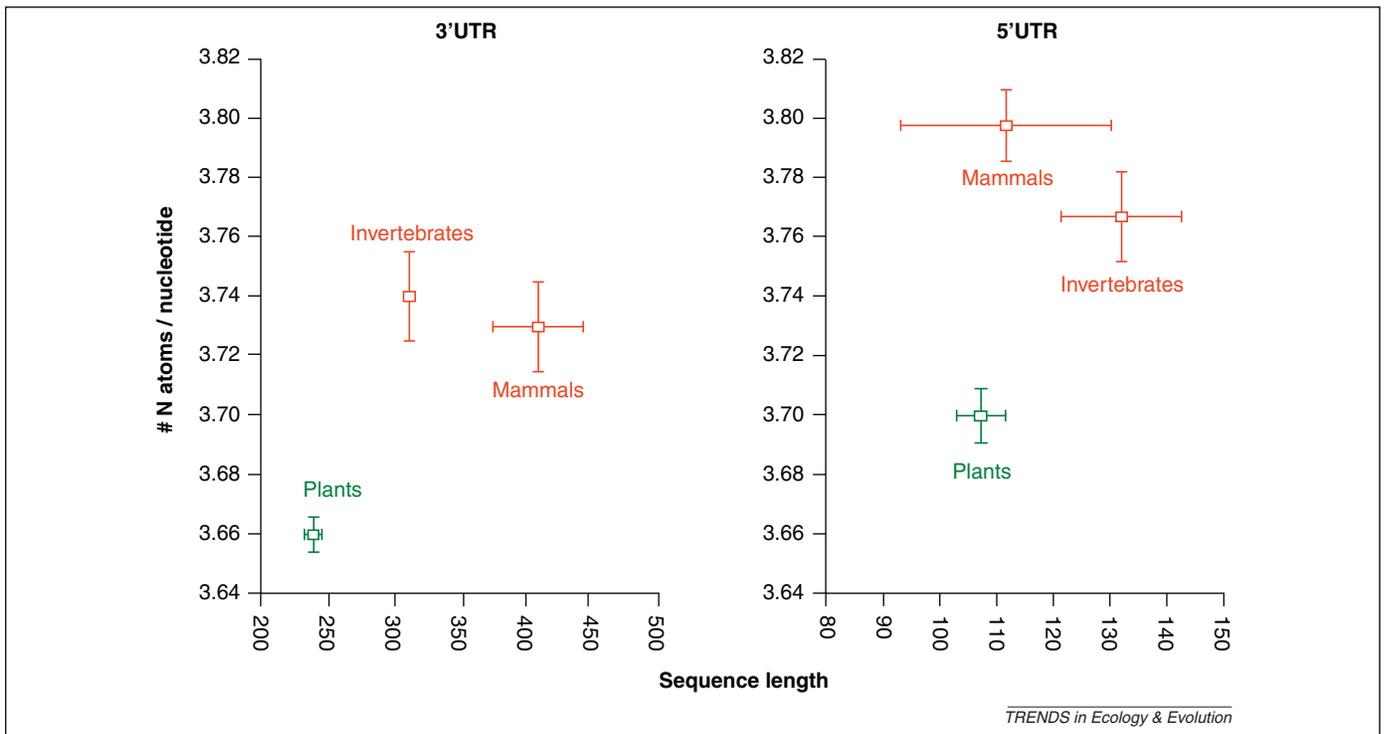
Given that the production of every nucleotide has N and P costs, a regime of nutrient conservation should affect not only the nucleotide composition, but also the total number of nucleotides used in transcripts. This predicts that, in autotrophs, purifying selection will prevent lengthening of transcribed regions, whereas it might enable a substantial expansion of non-transcribed regions because these do not incur the materials costs associated with transcript production. Using existing genomic data, it is clear that, as predicted, *A. thaliana* has a consistent signature of N and P conservation in functionally diverse transcribed non-coding sequences [introns and untranslated regions (UTRs; 3' and 5')] owing to biased nucleotide composition and reduced transcript length (unpublished results). Although the length of introns is known to correlate with genome size, the average intron length in plants is significantly shorter than in animals when organisms with similar genome size are compared. For example, *A. thaliana* and *D. melanogaster* have comparable genome sizes, but the average intron length is over one order of magnitude lower in *A. thaliana*. Despite their different functional requirements, the same signal is detected in UTRS: both 5' and 3' UTRS exons have a lower N content per nucleotide and are shorter in plants than in animals (Figure 3; *t*-test  $p < 0.0001$ ). These results further reinforce the view that N and P conserva-

tion represents a ubiquitous force shaping non-coding regions of the transcriptome.

#### Future directions

The field of stoichiogenomics is in its infancy. Work to date has focused on a relatively narrow range of questions within the still-limited number of species for which comprehensive genomic data with excellent gene annotation are available. This situation is changing rapidly owing to the revolutionary advances in DNA sequencing technology that have been accelerating the rate of genome sequencing for different species and individuals within species. Furthermore, these sequence-based techniques can now quantify gene and protein expression more precisely and generate longer sequences from diverse environmental samples (e.g. metagenomics).

These advances will facilitate population, ecological and taxonomic investigations of reciprocal interactions between environmental resource supplies and macromolecular evolution. Given that selection for element conservation is most likely to be seen in molecules that are at high abundance in organisms, future studies should combine sequence information with functional data that link macromolecular element composition to the probable abundance of those molecules in the cell. However, achieving a comprehensive understanding of the role of nutrient limitation in affecting macromolecular evolution is likely to be challenging, given the difficulty in assessing the role of nutrient limitation in the evolutionary history of many biota. Nevertheless, we highlight some possibilities here, although this list is now limited only by biologists' imaginations.



**Figure 3.** Nitrogen conservation in plants as illustrated by a lower average N content per nucleotide and shorter sequence length (exons) of UTRs [(a) 3' UTR and (b) 5' UTR] in plants versus animals. The mean and standard errors are plotted per organisms for a set of invertebrate, mammal and plant species. Sequence data were obtained from the UTRdb (<http://utrdb.ba.itb.cnr.it/>) and species with information from at least 50 genes were selected; number of species for 3' and 5' UTRs, respectively, were as follows: mammals (13 and 16), invertebrates (61 and 43) and plants (81 and 68).

### *Metagenomics is providing a major arena for new stoichiogenomic studies*

Owing to the revolution in high throughput sequencing, scientists increasingly have access to large volumes of genomic data (both DNA and RNA) for naturally occurring biota [20], especially microorganisms [21]. Often these data are obtained in conjunction with monitoring of key environmental variables [22], such as temperature and chemical conditions [e.g. pH, salinity and nutrient (N and P) supplies]. These data offer an unprecedented opportunity to connect patterns of genomic evolution to key ecological drivers and their consequences. For example, do the proteomes and transcriptomes of the microbes that dominate under low-productivity conditions in the ocean bear signs of MUB that conserve energy and key nutrient resources? The biochemical versatility of the microbes that occupy low-nutrient environments is already demonstrated by some remarkable features, as in the ability of various open-ocean cyanobacteria (*Synechococcus*, *Prochlorococcus*, *Crocospaera* and *Trichodesmium*) to substitute completely sulfolipids for phospholipids under P-limited growth conditions [23]. Is such biochemical fine tuning also reflected in the proteomes and transcriptomes of such species?

Under resource-limiting conditions, one would expect features such as preferential use of amino acids with low bioenergetic construction costs and with low N contents, perhaps together with shorter transcripts and shorter polymers used in various structures (e.g. rRNAs and ribosomal proteins). Consistent with this, recent analyses have shown that differences in nutrient availability in sea-surface versus deep-sea environments affect the average proteome elemental composition in strains of *Prochlorococcus* adapted to different depths, as strains isolated from deep-water layers with low light (low photosynthesis) and high N concentrations had proteomic MUB in favor of low-C and high-N amino acids, whereas the opposite was observed for strains from surface waters (where light is abundant and N is low) [24].

Also informative would be similar studies of soil microbes performed over chronosequences during which the overall and relative availabilities of N and P change. One can also imagine similar stoichiogenomic-inspired studies of the gut microbiome in animals differing in dietary habits (e.g. carnivores versus herbivores).

### *Key comparisons will shed light on major evolutionary selection events*

Some major evolutionary events are attractive as targets for future stoichiogenomic studies. For example, it is well known that domestication produced a suite of correlated genetic changes in the plant taxa adopted by humans for food production [25]. Given that part of the domestication process involved modification of soil fertility (e.g. rotation to fresh soils and manuring in the distant past; provision of chemical fertilizers in the recent past), it is probable that crop plants have experienced relaxed selection for nutrient conservation in proteomes and transcriptomes. Indeed, as discussed above, there is some support for this prediction [15] but data for key comparisons are missing. For example, it would be best to compare domesticated species with their ancestral taxa or closest living relatives, such that,

for example, 'red rice' (*Oryza rufipogon*) could be compared with 'white rice' (*Oryza sativa*) or teosinte compared with maize and corn. These comparisons would enable researchers to evaluate the effects on MUB of such relaxed selection from nutrient limitation combined with selection for high production.

### *New bioinformatics resources are emerging*

As is clear from above, understanding how the autecology and environment of an organism shape the evolution of genes, proteins and genomes is of fundamental importance in modern biology. Rapid development of DNA and protein sequence databases, knowledge of the basic biological function of individual proteins, and the availability of information characterizing relative gene-expression intensities provide unique and exciting opportunities to construct ecogenomic hypotheses and test their predictions.

Biologists in general are also becoming interested in understanding how environmental constraints (such as nutrient limitation) shape protein, DNA and RNA sequences. Such investigations require the analysis of element composition of proteins and genomic segments across many species, which necessitates the availability of bioinformatics knowledge bases at the interface of ecology, evolution and genomics. As an example, the online resource <http://www.graspdb.net> was recently developed for 12 drosophilid species that are evolutionarily closely related but ecologically divergent taxa [26]. This resource enables biologists to look for element-driven MUB in the genomes of drosophilids having divergent diets or evolutionary histories in strongly contrasting habitats. In general, however, there is a need for flexible and scalable systems that will contain an integrated ecological dataset (characterizing key life-history traits of species) and genomic data consisting of protein and DNA compositions. Such resources will provide an easy-to-use tool for the exploration and analysis of stoichiometric characteristics of proteins and genomes that will be freely available to the research community.

### **Summary and outlook**

Here, we have reviewed recent findings that show how ecological perspectives can inform molecular evolution analyses and so reveal new patterns in the primary structure of biological molecules. The overall picture is one in which the signs of element limitations can be seen in the monomer usage patterns of the molecules involved in processing limiting elements [10]. This is evident in the proteomic and transcriptomic monomer-use patterns for major groups of biota (plants versus animals) that differ in the severity of direct element limitation [13], and in the monomer profiles associated with different cellular components [16].

It is probable that further patterns will emerge as additional studies proceed. Indeed this seems almost assured, as one of the most exciting aspects of biology in the post-genomic era is the wealth of genome sequence and gene expression data available. However, access to large volumes of information does not guarantee that key insights and discovery will emerge. In fact, such a 'data glut' might impede discovery because scientists well informed in biology are not always experts in handling

large-scale data sets and computational techniques. For this reason, we advocate the development of biologist-centered tools and databases [27]. In addition, insight is more likely to emerge by the synergistic collaboration of computationally skilled analysts with theoretical and empirical biologists working from clearly articulated conceptual frameworks. Biological stoichiometry is one such framework [7] and its emerging application in studies of macromolecular evolution and population genetics promises future insights for connecting the disparate realms of molecular evolution and ecosystem ecology.

### Acknowledgements

This work was supported by a National Science Foundation grant to J.J.E., S.K. and W. Fagan, and a National Institutes of Health grant to S.K.; W Fagan and three anonymous reviewers provided useful feedback on the article.

### References

- 1 Eyre-Walker, A. and Keightley, P.D. (2007) The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* 8, 610–618
- 2 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press
- 3 Subramanian, S. and Kumar, S. (2006) Higher intensity of natural selection on >90% of the human genes revealed by the intrinsic replacement mutation rates. *Mol. Biol. Evol.* 23, 2283–2287
- 4 Bohlin, J. and Skjerve, E. (2009) Examination of genome homogeneity in prokaryotes using genomic signatures. *PLoS ONE* 4, e8113
- 5 Wang, H-C. *et al.* (2006) On the correlation between genomic G1C content and optimal growth temperature in prokaryotes: data quality and confounding factors. *Biochem. Biophys. Res. Commun.* 342, 681–684
- 6 Bragg, J.G. *et al.* (2006) Variation among species in proteomic sulphur content is related to environmental conditions. *Proc. R. Soc. B.* 273, 1293–1300
- 7 Elser, J.J. and Hamilton, A.L. (2007) Stoichiometry and the new biology: the future is now. *PLoS Biol.* 5, e181 doi: 110.1371/journal.pbio.0050181
- 8 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
- 9 Scheibel, T. *et al.* (1997) *S. cerevisiae* and sulfur: a unique way to deal with the environment. *FASEB J.* 11, 917–921
- 10 Baudouin-Cornu, P. *et al.* (2001) Molecular evolution of protein atomic composition. *Science* 293, 297–300
- 11 Bragg, J.G. and Wagner, A. (2009) Protein material costs: single atoms can make an evolutionary difference. *Trends Genet.* 25, 5–8
- 12 Carlson, R.P. (2007) Metabolic systems cost-benefit analysis for interpreting network structure and regulation. *Bioinformatics* 23, 1258–1264
- 13 Elser, J.J. *et al.* (2006) Signatures of ecological resource availability in the animal and plant proteomes. *Mol. Biol. Evol.* 23, 10.1093
- 14 Li, N. *et al.* (2009) Low contents of carbon and nitrogen in highly abundant proteins: evidence of selection for the economy of atomic composition. *J. Mol. Evol.* 68, 248–255
- 15 Acquisti, C. *et al.* (2009) Ecological nitrogen limitation shapes the DNA composition of plant genomes. *Mol. Biol. Evol.* 26, 953–956
- 16 Acquisti, C. *et al.* (2009) From elements to biological processes: signatures of nitrogen limitation in the elemental composition of the catabolic apparatus. *Proc. R. Soc. B.* 276, 2605–2610
- 17 Bragg, J.G. and Hyder, C.L. (2004) Nitrogen versus carbon use in prokaryotic genomes and proteomes. *Proc. R. Soc. B.* 271, S374–S377
- 18 Hessen, D.O. *et al.* (2010) Genome streamlining and the elemental costs of growth. *Trends Ecol. Evol.* 25, 75–80
- 19 Sterner, R.W. and Elser, J.J. (2002) *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere*, Princeton University Press
- 20 DeLong, E.F. (2009) The microbial ocean from genomes to biomes. *Nature* 459, 200–206
- 21 Riesenfeld, C.S. *et al.* (2004) Metagenomics: genomic analysis of microbial communities. *Annu. Rev. Genet.* 38, 525–552
- 22 Jones, N. (2010) Undersea project delivers data flood. *Nature* 464
- 23 Van Mooy, B. *et al.* (2009) Phytoplankton in the ocean use non-phosphorus lipids in response to phosphorus scarcity. *Nature* 458, 69–72
- 24 Lv, J. *et al.* (2008) Association between the availability of environmental resources and the atomic composition of organismal proteomes: evidence from *Prochlorococcus* strains living at different depths. *Biochem. Biophys. Res. Commun.* 375, 241–246
- 25 Burke, J. *et al.* (2007) Crop evolution: from genetics to genomics. *Curr. Opin. Genet. Dev.* 17, 525–532
- 26 Stark, A. *et al.* (2007) Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* 450, 219–232
- 27 Kumar, S. and Dudley, J. (2007) Bioinformatics software for biologists in the genomics era. *Bioinformatics* 23, 1713–1717