bioRxiv preprint doi: https://doi.org/10.1101/162693; this version posted July 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Title (90 chars)

The amount of nitrogen used for photosynthesis governs molecular evolution in plants

Authors

Steven Kelly

Affiliations

Department of Plant Sciences, University of Oxford, Oxford, OX1 3RB, UK

Corresponding Author

Name: Steven Kelly

Email: steven.kelly@plants.ox.ac.uk

Telephone: +44 (0)1865 275123

Address: Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, UK

Keywords

Photosynthesis, efficiency, selection, evolution.

Abstract

Genome and transcript sequences are composed of long strings of nucleotide monomers (A, C, G and T/U) that require different quantities of nitrogen atoms for biosynthesis. Here it is shown that plants that require more nitrogen for photosynthesis use nucleotides with fewer nitrogen atoms in both genome and transcript sequences. These findings reveal that plant GC content, codon use, and gene sequence evolution are driven by photosynthetic nitrogen demand.

Main text

Cells are built from macromolecules (proteins, RNA, DNA, phospholipids and polysaccharides) that in turn are constructed from monomers (amino acids, nucleotides, fatty acids and sugars). The majority of plants can biosynthesise all of the monomers and macromolecules they require from inorganic carbon (CO_2) and nitrogen (NO_3/NH_4) obtained from their environment. Of these two

bioRxiv preprint doi: https://doi.org/10.1101/162693; this version posted July 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

resources, nitrogen is scarcer and as such plant growth rate is nitrogen limited in both natural and agricultural environments[1-3]. This limitation in growth is caused by the fact that synthesis of proteins required for photosynthetic carbon assimilation needs a substantial nitrogen investment[4]. Thus, the rate of macromolecule biosynthesis (and hence growth) in plants is predominantly limited by nitrogen availability.

Photosynthetic nitrogen use efficiency is the amount of carbon that can be fixed per unit of nitrogen invested by the plant. Multiple disparate anatomical, physiological and molecular factors contribute to variation in photosynthetic nitrogen use efficiency such that there is a large variation between different plant species[5]. For example, plants that use the C_4 photosynthetic pathway exhibit higher nitrogen use efficiency when compare to plants that use C_3 photosynthesis. The cohort of changes that facilitated C_4 evolution enabled plants to reduce resource allocation to protein synthesis without causing a corresponding reduction in photosynthetic rate[6]. Thus, C_4 plants can achieve ~50% higher rates of photosynthesis than C_3 plants given the same input resources[7].

While the sequence and abundance of proteins within a cell are functionally constrained, it is possible to encode the same polypeptide using multiple different nucleotide sequences. This redundancy in the genetic code means that the biosynthesis cost of transcripts can vary such that it is possible to reduce the allocation of cellular resources to transcript sequences without altering protein sequence or function[8]. This reduction in resource allocation is possible because nucleotide monomers (A, C, G and T/U) differ in their biosynthesis requirements, with different nucleotides requiring different quantities of nitrogen atoms for their construction. Adenine and guanine require 5 nitrogen atoms for their biosynthesis, cytosine requires 3, and thyamine/uracil only require 2. It was thus hypothesised that variation in photosynthetic nitrogen use efficiency between plant species would result in a selection-driven effect on the nitrogen content of plant gene sequences. Specifically, plants that require increased quantities of nitrogen to fix CO_2 would have less nitrogen available for other macromolecules in the cell and thus experience stronger selection to minimise nitrogen allocation to transcript sequences. To test this hypothesis, an analysis of molecular sequence evolution was conducted for 11 plant species for which both whole genome sequences and accurate photosynthetic nitrogen use efficiencies[5] were available. This set of species included

bioRxiv preprint doi: https://doi.org/10.1101/162693; this version posted July 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

both C_3 and C_4 grasses, as well as C_3 herbs and trees (Fig. 1a, Supplementary Table S1, Supplemental File S1).

For each species, the strength of selection acting on transcript biosynthesis cost [S_c] was inferred for the complete set of open reading frames in the respective genome using the SK model[9] implemented using CodonMuSe[8]. Consistent with the hypothesis, those species that required more nitrogen to conduct photosynthesis exhibited stronger selection (more negative selection coefficient) to minimise the nitrogen biosynthesis cost of transcript sequences (R^2 = 0.62, $p < 0.001$, Fig. 1b). Because the biosynthesis cost of DNA sequences also varies (AT pairs require 7 and GC pairs require 8 nitrogen atoms), it was further hypothesised that those species that required more nitrogen to conduct photosynthesis would exhibit a stronger genome-wide mutation bias towards AT base pairs. Consistent with the hypothesis, those species that required more nitrogen to conduct photosynthesis had lower genome-wide GC content and thus invested less nitrogen in their genome sequences (R^2 = 0.58, $p < 0.001$, Fig. 1c). This mutation-driven phenomenon was also apparent from the analysis of coding sequences, where mutation bias towards AT was stronger in species that had lower photosynthetic nitrogen use efficiencies (R^2 = 0.65, $p < 0.001$, Fig. 1d). Moreover, when mutation bias and selection acting on transcript biosynthesis cost are considered together, they are sufficient to explain ~90% of variance in genome-wide patterns of synonymous codon use in plants (Supplemental Table 1, Supplemental File S2). Therefore, the photosynthetic nitrogen use efficiency of a plant is the major determinant of genome-wide GC content and of biased patterns of synonymous codon use.

Given the findings above it was postulated that variance in photosynthetic nitrogen use efficiency between species would result in a global positive correlation between mutation bias and selection acting on transcript biosynthesis cost across all plant species. To investigate this, the complete set of angiosperm genomes on Phytozome were subject to the same molecular evolution analysis as above. Consistent with the hypothesis, those species that exhibited stronger selection acting on transcript biosynthesis cost also exhibited lower genome-wide GC content (R^2 = 0.69, Fig. 1e, Supplemental Table 1). Thus, although detailed photosynthetic nitrogen use efficiency estimates are not available for these species, variance in selection acting on the biosynthesis cost of

bioRxiv preprint doi: https://doi.org/10.1101/162693; this version posted July 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

transcript sequences and the associated variance in the mutation bias of gene and genome sequences are consistent with molecular sequence evolution being driven by photosynthetic nitrogen use efficiency.

Given that the strength of selection acting on transcript biosynthesis cost has been shown to constrain the rates of molecular evolution of genes in bacteria[8], it was investigated whether the same effect was observed in plant genomes. Here the strength of selection acting on biosynthesis cost of transcript sequences was compared to the rate of synonymous substation (K_s) for the same genes. This revealed that plant genes that experienced stronger selection for minimisation of biosynthesis cost evolved more slowly than genes that experience weaker selection (Fig 1f). Moreover, variance in the strength of selection acting on the biosynthesis cost of transcript sequences explains ~30% of variation in molecular evolution rate between genes in plants (Fig. 1f). Thus, the tempo of the plant molecular clock is in part set by photosynthetic nitrogen use efficiency.

Collectively, these findings provide significant new insight into the relationship between metabolism, the environment and molecular evolution in plants. They are also compatible with previous reports that revealed that wild plants contained less nitrogen in their DNA when compared to domesticated relatives that had been supplemented with nitrogen fertiliser for thousands for years[10]. The new findings presented here have significant implications for our understanding of the past, present and future of plant evolution. First, the tempo of the plant molecular clock is set by photosynthetic nitrogen use efficiency such that plants with higher efficiency accumulate substitutions faster than those with lower efficiency. Thus, estimates of divergence time from molecular sequence data are dependent on past photosynthetic nitrogen use efficiency. Second, speciation and extinction rates in plants are a function of molecular substitution rate, such that lineages with higher rates of molecular substitution have higher rates of speciation and extinction[11]. Therefore, plants with enhanced photosynthetic nitrogen use efficiency, and thus with higher rates of molecular evolution, will have higher rates of speciation and extinction. As a corollary, evolutionary adaptations that increase photosynthetic nitrogen use efficiency will also increase rates of speciation and extinction. Therefore, this phenomenon provides a simple mechanistic

bioRxiv preprint doi: https://doi.org/10.1101/162693; this version posted July 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

explanation for the increase in rates of speciation that are concomitant with the evolution of C_4 photosynthesis[12] and the associated increase in photosynthetic nitrogen use efficiency. Finally, increases in atmospheric CO_2 concentration cause corresponding increases in photosynthetic nitrogen use efficiency in plants by reducing photorespiration and causing a corresponding reduction in the investment of cellular resources in photosynthesis protein production[13]. Thus, anticipated future increase in atmospheric CO_2 concentration will lead to an increase in the rate of plant mutation, speciation and extinction, independent of any effects due to habitat destruction or changes in land use.

Methods

Data sources

The genome sequence and corresponding set of representative gene models for each species were downloaded from Phytozome[14]. The *Helianthus annuus* genome was obtained from[15]. Photosynthetic measurements and leaf nitrogen measurements were obtained from[5].

Inference of selection acting on codon usage bias

To obtain the number of tRNA genes in each genome, tRNAscan[9] was run on each of the plant genomes. For each species the tRNAscan output file and the complete set of representative coding sequences was analysed using CodonMuSe[8] and values for the selection coefficients of mutation bias (M_b), selection acting on biosynthesis cost (S_c), and selection acting on translational efficiency (S_c) were obtained (Supplemental Table S1).

Inference of molecular evolution rates

To identify orthologous gene pairs between each species pair the complete set of representative protein sequences for each species were subject to ortholog inference using OrthoFinder[16]. Orthologues for each species pair were subject to multiple sequence alignment using MAFFT using the L-LNS-i algorithm[17]. The resulting protein sequence alignments were rethreaded with their corresponding coding sequences and subject to molecular rate analysis using KaKs calculator[18] using the default settings.

bioRxiv preprint doi: https://doi.org/10.1101/162693; this version posted July 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

References (20 references)

1 Ingestad, T. & Lund, A.-B. Nitrogen Stress in Birch Seedlings. *Physiologia Plantarum* **45**, 137-148, doi:10.1111/j.1399-3054.1979.tb01678.x (1979).

2 Ingestad, T. & I. Ågren, G. Theories and methods on plant nutrition and growth. *Physiologia Plantarum* **84**, 177-184, doi:10.1111/j.1399-3054.1992.tb08781.x (1992).

3 LeBauer, D. S. & Treseder, K. K. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology* **89**, 371-379 (2008).

4 Hohmann-Marriott, M. F. & Blankenship, R. E. Evolution of photosynthesis. *Annual review of plant biology* **62**, 515-548, doi:10.1146/annurev-arplant-042110-103811 (2011).

5 Rotundo, J. L. & Cipriotti, P. A. Biological limits on nitrogen use for plant photosynthesis: a quantitative revision comparing cultivated and wild species. *New Phytologist* **214**, 120-131, doi:10.1111/nph.14363 (2017).

6 Oaks, A. Efficiency of Nitrogen Utilization in C3 and C4 Cereals. *Plant physiology* **106**, 407-414 (1994).

7 Evans, J. R. & von Caemmerer, S. in *Studies in Plant Science* Vol. Volume 7 (eds P. L. Mitchell J.E. Sheehy & B. Hardy) 53-71 (Elsevier, 2000).

8 Seward, E. A. & Kelly, S. Selection-Driven Cost-Efficiency Optimisation Of Transcript Sequences Determines The Rate Of Gene Sequence Evolution In Bacteria. *bioRxiv*, doi:10.1101/136861 (2017).

9 Seward, E. A. & Kelly, S. Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms. *Genome biology* **17**, 226, doi:10.1186/s13059-016-1087-9 (2016).

10 Acquisti, C., Elser, J. J. & Kumar, S. Ecological Nitrogen Limitation Shapes the DNA Composition of Plant Genomes. *Molecular biology and evolution* **26**, 953-956, doi:10.1093/molbev/msp038 (2009).

11 Lancaster, L. T. Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. *BMC evolutionary biology* **10**, 162, doi:10.1186/1471-2148-10-162 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/162693; this version posted July 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

12 Spriggs, E. L., Christin, P. A. & Edwards, E. J. C4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PloS one* **9**, e97722, doi:10.1371/journal.pone.0097722 (2014).

13 Stitt, M. & Krapp, A. The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. *Plant, Cell & Environment* **22**, 583-621, doi:10.1046/j.1365-3040.1999.00386.x (1999).

14 Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic acids research* **40**, D1178-1186, doi:10.1093/nar/gkr944 (2012).

15 Badouin, H. *et al.* The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **546**, 148-152, doi:10.1038/nature22380 (2017).

16 Emms, D. M. & Kelly, S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology* **16**, 157, doi:10.1186/s13059-015-0721-2 (2015).

17 Katoh, K. & Standley, D. M. MAFFT: iterative refinement and additional methods. *Methods in molecular biology* **1079**, 131-146, doi:10.1007/978-1-62703-646-7_8 (2014).

18 Wang, D., Zhang, Y., Zhang, Z., Zhu, J. & Yu, J. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics, proteomics & bioinformatics* **8**, 77-80, doi:10.1016/S1672-0229(10)60008-3 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/162693; this version posted July 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figures

Figure 1

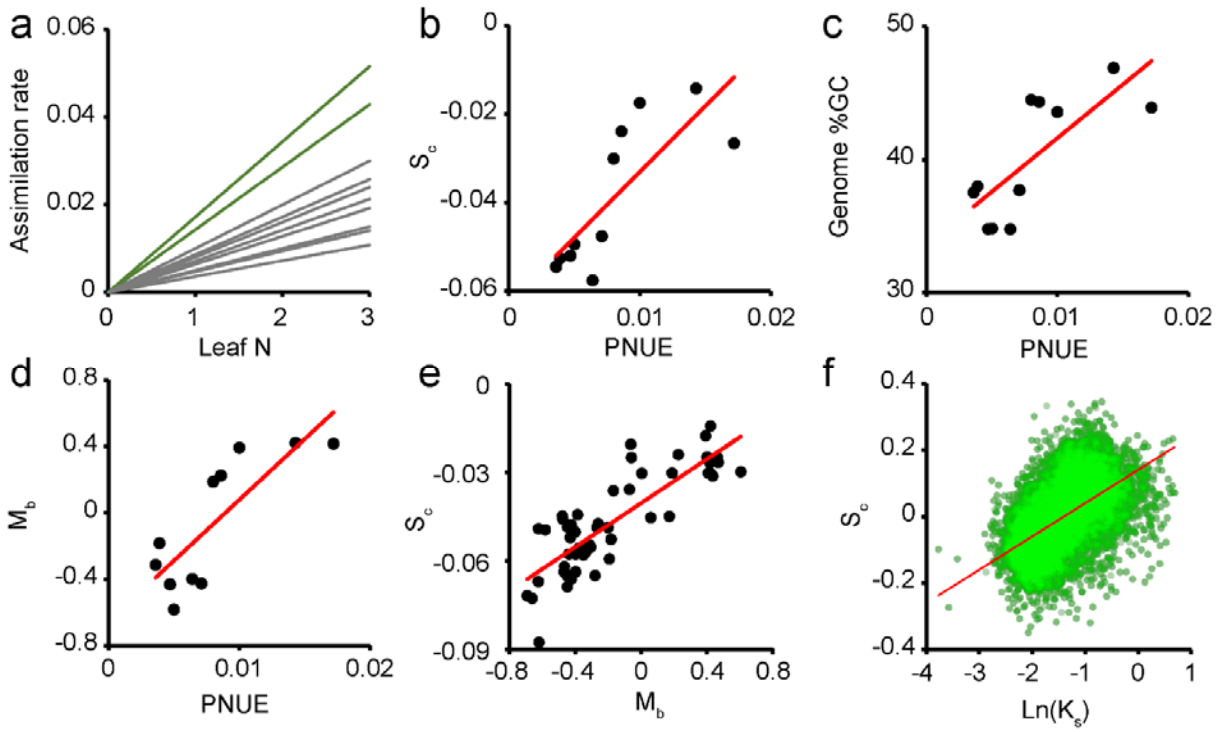

Figure 1. Photosynthetic nitrogen use efficiency (PNUE) drives selection acting on codon biosynthesis cost and mutation bias. Plots in parts a to e depict the same species set. a) The range of relationships between light saturated CO_2 assimilation rate (µ mol CO_2 m^{-2} s^{-1} µ mol PAR^{-1}) and leaf nitrogen (g N m^{-2}). C_4 species shown in green and C_3 species shown in grey. Complete datasets provided in Supplemental file 1. The slope of the line is the PNUE for that species. b) The relationship between PNUE and the strength of selection acting on codon biosynthesis cost (S_c, R^2 = 0.62) for these species. c) The relationship between PNUE and the Genome wide GC content (R^2 = 0.58). d) The relationship between PNUE and the strength of mutation bias acting on coding sequences (M_b, R^2 = 0.65). e) The relationship between S_c and M_b for all angiosperm species in Phytozome (R^2 = 0.69). f) The relationship between S_c and K_s for *Z. mays* and *S. bicolor* (R^2 = 0.32).