The Amount of Nitrogen Used for Photosynthesis Modulates Molecular Evolution in Plants

Steven Kelly*,[1]

[1]Department of Plant Sciences, University of Oxford, Oxford, United Kingdom

***Corresponding author:** E-mail: steven.kelly@plants.ox.ac.uk.

Associate editor: Michael Purugganan

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 the strength of selection acting on transcript nitrogen content is influenced by the amount of nitrogen plants require to conduct photosynthesis. Specifically, plants that require more nitrogen to conduct photosynthesis experience stronger selection on transcript sequences to use synonymous codons that cost less nitrogen to biosynthesize. It is further shown that the strength of selection acting on transcript nitrogen cost constrains molecular sequence evolution such that genes experiencing stronger selection evolve at a slower rate. Together these findings reveal that the plant molecular clock is set by photosynthetic efficiency, and provide a mechanistic explanation for changes in plant speciation rates that occur concomitant with improvements in photosynthetic efficiency and changes in the environment such as light, temperature, and atmospheric CO_2 concentration.

Key words: **photosynthesis, efficiency, selection, evolution.**

Article

Introduction

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 biosynthesize 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 resources, nitrogen is scarcer and hence plant growth rate is generally nitrogen limited in both natural and agricultural environments (Ingestad and Lund 1979; Evans 1983; Ingestad and Ågren 1992; LeBauer and Treseder 2008). This limitation in growth is caused by the fact that synthesis of proteins required for photosynthetic carbon assimilation needs a substantial nitrogen investment (Evans 1989; Hohmann-Marriott and Blankenship 2011).

Photosynthetic nitrogen use efficiency (PNUE) 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 PNUE such that there is a large variation between different plant species (Rotundo and Cipriotti 2017). For example, plants that use the C_4 photosynthetic pathway exhibit higher nitrogen use efficiency when compared with plants that use C_3 photosynthesis. The cohort of changes that facilitated C_4 evolution enabled plants to reduce resource allocation to photosynthetic machinery without causing a corresponding reduction in photosynthetic rate (Oaks 1994). Thus, C_4 plants can achieve ~50% higher rates of photosynthesis than C_3 plants given the same amount of nitrogen (Evans and von Caemmerer 2000).

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 thymine/uracil only require 2. Although the sequence and abundance of proteins within a cell are functionally constrained, it is possible to encode the same polypeptide with multiple different nucleotide sequences by using different synonymous codons. This redundancy in the genetic code, coupled with the difference in nucleotide nitrogen content, means that it is possible to reduce the allocation of cellular resources to transcript sequences without altering protein sequence or presumed function (Seward and Kelly 2017). For example, a single A to T substitution in a highly expressed transcript such as RuBisCo small subunit (~5,000 transcripts per cell) saves an equivalent amount of nitrogen (~15,000 atoms) as is contained in three complete RuBisCo holoenzymes (~5,000 nitrogen atoms per hexadecamer). It was thus hypothesized that plants that require increased quantities of nitrogen to fix CO_2 would be more nitrogen limited and thus natural selection would favor codons in transcript sequences that required less nitrogen to biosynthesize.

Here, it is shown that plants that require more nitrogen to conduct photosynthesis experience stronger selection to minimize transcript nitrogen biosynthesis cost. Furthermore, it is demonstrated that the strength of selection acting on transcript sequence biosynthesis cost explains a

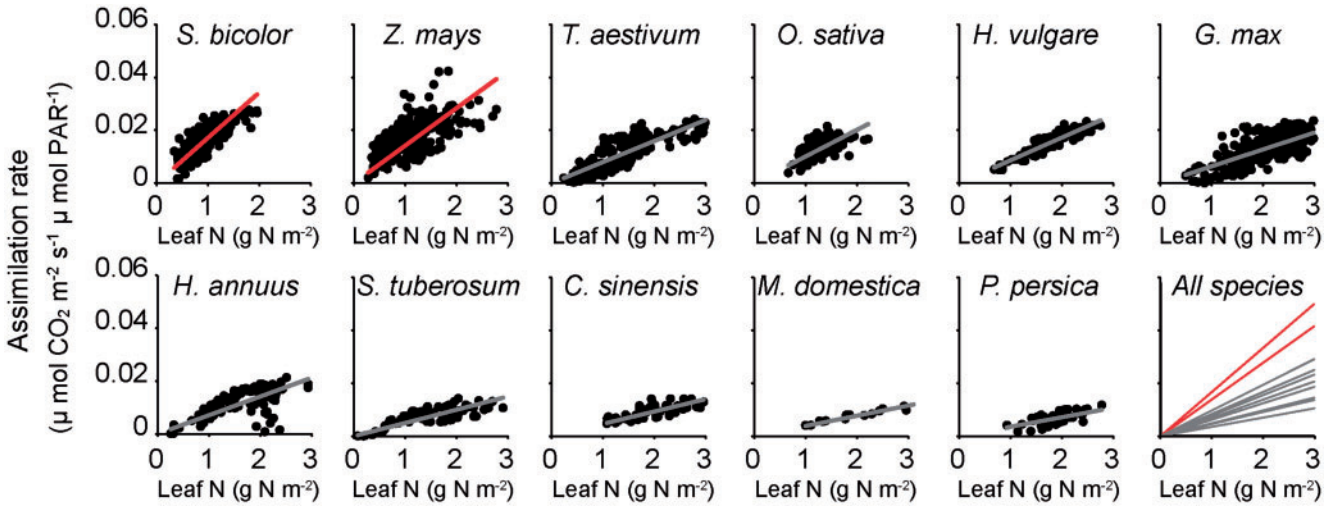

Fig. 1. Scatter plots depicting the relationship between light saturated photosynthetic rate and leaf nitrogen content for the 11 species used in this analysis. A higher resolution plot with R^2 values and slopes is provided as supplementary file S1, Supplementary Material online. Light saturated CO_2 assimilation rate (μ mol CO_2 $m^{-2}s^{-1}$ μ mol PAR^{-1}) and leaf nitrogen (g N m^{-2}). Each fitted line is also provided on the plot labeled "All Species" to make it easier to compare the range of relationships that are observed between light saturated CO_2 assimilation rate and leaf nitrogen. Fitted lines for C_4 species are shown in red and fitted lines for C_3 species shown in gray. The slope of each fitted line is the photosynthetic nitrogen use efficiency (PNUE) for each species. *Sorghum bicolor* (Muchow and Sinclair 1994; Anten et al. 1995; Tominaga et al. 2015), *Zea mays* (Muchow and Sinclair 1994; Lindquist and Mortensen 1999; Osaki et al. 2001; Paponov and Engels 2003; Drouet and Bonhomme 2004; Paponov et al. 2005; Vos et al. 2005), *Triticum aestivum* (Evans 1983; Fischer et al. 1998; Osaki et al. 2001; Müller et al. 2005), *Oryza sativa* (Anten et al. 1995; Osaki et al. 2001; Ohsumi et al. 2007; Hirasawa et al. 2010), *Hordeum vulgare* (Braune et al. 2009), *Glycine max* (Anten et al. 1995; Osaki et al. 2001; Maekawa and Kokubun 2005; Rotundo and Borras 2016), *Helianthus annuus* (Fredeen et al. 1991; Andrade et al. 1993; Gimenez et al. 1994; Trápani and Hall 1996; Bange et al. 1997; Rodríguez et al. 1998), *Solanum tuberosum* (Vos and van der Putten 1998, 2001), *Citrus sinensis* (Romero-Aranda and Syvertsen 1996), *Malus domestica* (Cheng and Fuchigami 2000), and *Prunus persica* (DeJong et al. 1989; Rosati et al. 1999; Malcolm et al. 2008).

significant proportion of variation in gene evolutionary rate, whereby genes that experience stronger selection to minimize cost are evolving slower than genes that experience weaker selection. Together these findings directly link photosynthetic efficiency of a plant to the rate at which its genes and genome evolve, and provide a mechanistic link between fluctuation in rates of plant diversification and changing environmental conditions.

Results

Variation in Photosynthetic Nitrogen Use Efficiency Modulates the Strength of Selection Acting on Transcript Nitrogen Cost

To test the hypothesis that variation in the amount of nitrogen used for photosynthesis influences the strength of selection acting on transcript biosynthesis cost, an analysis of molecular sequence evolution was conducted for 11 plant species for which both whole genome sequences (Goodstein et al. 2012) and accurate photosynthetic nitrogen use efficiencies (Rotundo and Cipriotti 2017) were available. This set of species included both C_3 and C_4 grasses, as well as C_3 herbs and trees encompassing a broad range of PNUE values (fig. 1 and supplementary table S1 sheet 1, Supplementary Material online). For each species, the strength of selection acting on transcript nitrogen cost [S_c] was inferred from the complete set of open reading frames in the respective genome using the SK model (Seward and Kelly 2016) implemented using CodonMuSe (Seward and Kelly 2017). Consistent with the hypothesis, those species that required more nitrogen to conduct photosynthesis exhibited stronger selection (more negative value for S_c) to minimize the nitrogen biosynthesis cost of transcript sequences ($R^2 = 0.62$, $P < 0.001$, fig. 2A). However, both PNUE and S_c exhibited significant phylogenetic signal (supplementary file S2, Supplementary Material online), and thus a phylogenetic least squares (PGLS) analysis was conducted to account for this effect. This PGLS approach makes the implicit assumption that traits evolve similarly across the phylogeny (Keck et al. 2016). This assumption is invalid for evolutionary transitions from C_3 to C_4 photosynthesis, as this change is concomitant with a rapid change in PNUE that is disproportionate to phylogenetic distance. Similarly, this assumption is also invalid for the evolution of root-nitrogen fixation as legumes export photosynthate for use in nitrogen fixation and thus the amount of carbon acquired per unit nitrogen in the leaf is an overestimate of the amount of carbon acquired by the plant. Thus, all PGLS model were built using data from the C_3 species (supplementary file S2, Supplementary Material online). Correction for phylogenetic signal did not remove the significant positive association between PNUE and S_c ($R^2 = 0.78$, $P = 0.004$, supplementary file S2, Supplementary Material online) and thus the amount of nitrogen used for photosynthesis in a plant modulates the strength of selection acting on its transcript nitrogen biosynthesis cost.

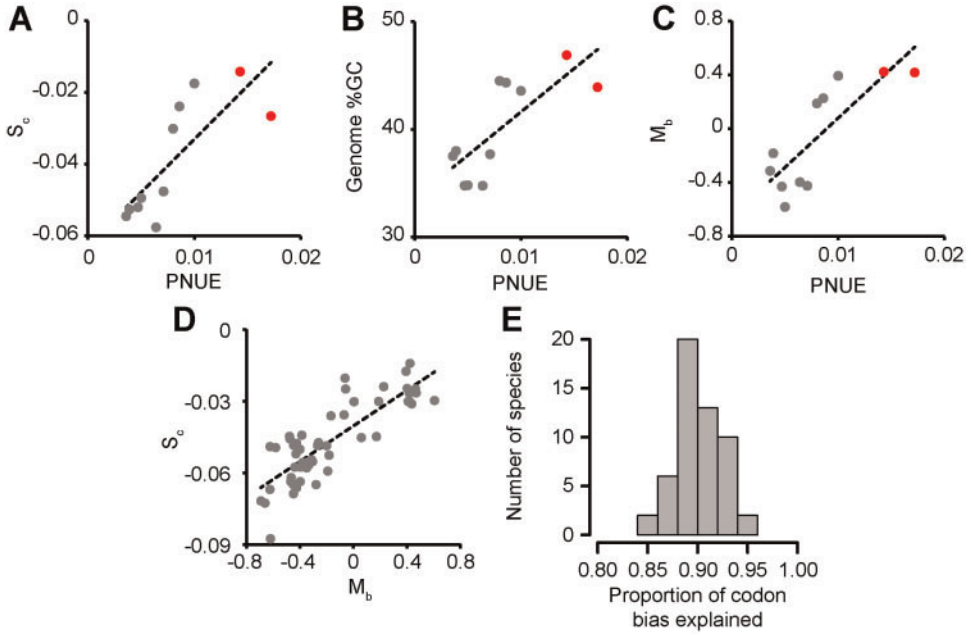

Fig. 2. Photosynthetic nitrogen use efficiency (PNUE) modulates selection acting on transcript biosynthesis cost and mutation bias. Plots in parts A to C depict the same species set. C_4 species are shown as red points and C_3 species are shown as gray points. Complete data sets are provided in supplementary file S1, Supplementary Material online. (*A*) The relationship between PNUE and the strength of selection acting on transcript biosynthesis cost (S_c, $R^2 = 0.62$) for these species. (*B*) The relationship between PNUE and the Genome wide GC content ($R^2 = 0.58$). (*C*) The relationship between PNUE and mutation bias acting on coding sequences here values >0 indicate mutation bias toward GC and values <0 indicate a mutation bias toward AT (M_b, $R^2 = 0.65$). (*D*) The relationship between S_c and M_b for all angiosperm species in Phytozome ($R^2 = 0.69$). (*E*) The proportion of codon bias that can be explained by the joint effects of mutation bias and selection acting on nitrogen biosynthesis cost.

It has previously been shown that the strength of selection acting on transcript biosynthesis cost and translational efficiency acts in proportion to the mRNA abundance of a gene (Seward and Kelly 2016, 2017). Comparison of gene-wise estimates for S_c and S_t with mRNA abundance estimates obtained from whole-plant RNA-Seq in *Arabidopsis thaliana* revealed that the same phenomenon also occurs in plants (supplementary file S3, Supplementary Material online). Thus, the magnitude of selection acting on an individual gene is modulated by both the abundance of the mRNA and the amount of nitrogen used for photosynthesis.

Variation in Photosynthetic Nitrogen Use Efficiency Influences Variation in Mutation Bias and Genome-Wide GC Content

As the nitrogen biosynthesis cost of DNA sequences varies (AT pairs require 7 and GC pairs require 8 nitrogen atoms), it was further hypothesized that those species that required more nitrogen to conduct photosynthesis would exhibit a stronger genome-wide mutation bias toward 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. 2*B*). This phenomenon was also apparent from the analysis of coding sequences, where codon mutation bias toward AT rich codons was stronger in species that had lower photosynthetic nitrogen use efficiencies ($R^2 = 0.65$, $P < 0.001$, fig. 2*C*). Like for S_c, both genome-wide GC content and mutation bias exhibited significant phylogenetic signal (supplementary file S2, Supplementary Material online). However, in contrast to the case for S_c correction for phylogenetic signal reduced the strength of the positive association with PNUE in C_3 species such that they failed to achieve statistical significance ($P \geq 0.05$, supplementary file S2, Supplementary Material online).

To exclude the possibility that low sample size caused the statistical test to fail, an additional analysis on a larger species set was conducted. If PNUE influences genome-wide GC content and mutation bias, then there should be a dependency between S_c and these traits in C_3 species from across the angiosperm phylogenetic tree. However, if there is no association between GC content, mutation bias and PNUE then S_c will also be independent of GC content and mutation bias. To investigate this, a larger set of C_3 angiosperm genomes on Phytozome were analyzed to determine whether there was a global, significant, positive association between S_c and GC content and mutation bias. As postulated, those C_3 species that exhibited stronger selection acting on transcript biosynthesis cost also exhibited lower genome-wide GC content ($R^2 = 0.69$, fig. 2*D* and supplementary table S1 sheet 2, Supplementary Material online). Correcting for phylogenetic signal did not remove the significant positive association ($R^2 = 0.21$, $P = 0.007$, supplementary file S2, Supplementary Material online). An analogous result was also obtained if

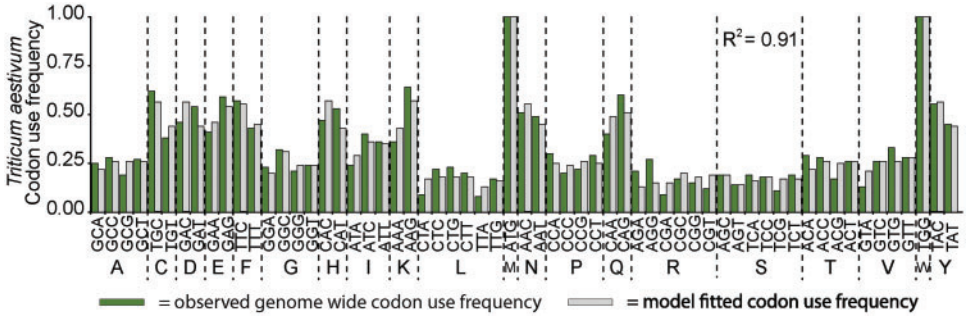

Fig. 3. A comparison of model-fitted and genome wide patterns of synonymous codon use. Example shown is from *Triticum aestivum*. The calculation of the coefficient of determination (R^2) did not include amino acids with only one codon (M or W). *T. aestivum* was chosen as it is close to the average R^2 value of the species analyzed. The complete set of plots can be found in supplementary file S3, Supplementary Material online.

estimates of mutation bias were obtained from genome sequence with no current gene annotation (supplementary file S2, Supplementary Material online). Therefore, the most parsimonious explanation is that variance in PNUE is a determinant of biased patterns of nucleotide use in both genome and transcriptome sequences. Furthermore, when mutation bias (as calculated by the SK model) and selection acting on transcript nitrogen cost are considered together, they are sufficient to explain ~90% of variance in genome-wide patterns of synonymous codon use in all plant species tested (fig. 2*E* and supplementary table S1 sheet 2, Supplementary Material online). To illustrate this explanatory power an example codon usage plot and model fit is provided for *Triticum aestivum* in figure 3. Analogous plots for each of the other eleven species are provided in supplementary file S4, Supplementary Material online. Other factors not included in this analysis account for the remaining unexplained variation in synonymous codons use.

Variation in Photosynthetic Nitrogen Use Efficiency Influences Variation in Nitrogen Content of Amino Acid Side Chains in Conserved Basic Sites

It has previously been shown that nitrogen limitation can cause a reduction in the nitrogen content of proteins in marine (Grzymski and Dussaq 2012) and parasitic microorganisms (Seward and Kelly 2016). Given the observed interaction between PNUE and the strength of selection acting on transcript biosynthesis cost, it was investigated whether a similar effect could be detected in the nitrogen content of amino acid side chains. Most amino acids used in construction of proteins contain a single nitrogen atom. However, six of the 20 also contain one or more nitrogen atoms in their side chains (R = 4 nitrogen atoms, H = 3, K = 2, N = 2, Q = 2, W = 2). Although, analogous redundancy to the codon code does not exist for amino acids, some amino acids exhibit similar functional properties. Of the 6 amino acids with nitrogen atoms in their side chains, three (R, H, and K) have basic side chains at neutral pH and thus could be considered to exhibit some biochemical redundancy with each other. Moreover, these three basic residues vary in their nitrogen content. Therefore, to determine whether variation in PNUE caused a concomitant variation in the nitrogen content of protein sequences, an analysis was conducted on 2,545 ungapped aligned basic sites in 124 ubiquitously conserved single copy genes in the 11 species. Here, ungapped sites that contained only basic residues were analyzed so it could be assumed that there is a functional constraint on the biochemical properties of the residue present, and that to some extent basic resides may be able to act redundantly at these positions. Consistent with the analysis of transcript sequences and genome GC content, those species that required more nitrogen to conduct photosynthesis contained fewer nitrogen atoms in amino acid side chains at conserved basic sites in ubiquitously conserved genes ($R^2 = 0.55$, $P = 0.008$, fig. 4*A*). However, correction for phylogenetic signal reduced the strength of the positive association with PNUE such that it failed to achieve statistical significance for the C_3 species within this group ($P \geq 0.05$).

As before, to exclude the possibility that low sample size caused the statistical test to fail, an additional analysis on a larger set of species was conducted. If PNUE influences nitrogen content in amino acid sequences, then there should be a dependency between S_c and amino acid nitrogen content at conserved basic sites across the angiosperm phylogenetic tree. As above, a larger set of C_3 species was analyzed and those species that exhibited stronger selection acting on transcript biosynthesis cost also exhibited lower nitrogen content at conserved basic sites ($R^2 = 0.23$, $P = 0.01$, fig. 4*B*). Correcting for phylogenetic signal did not remove the significant positive association ($R^2 = 0.14$, $P = 0.026$). Therefore, the most parsimonious explanation is that variance in PNUE also influences patterns of amino acid use in protein sequences.

Variation in the Strength of Selection Acting on Nitrogen Biosynthesis Cost Contributes to Variation in Gene Evolutionary Rate

Given that variance in PNUE is associated with variance in the strength of selection acting on gene sequences, it was postulated that this would cause a concomitant variance in molecular evolutionary rate of genes. Specifically, those genes that experience stronger selection to minimize transcript nitrogen cost would have lower rates of molecular evolution when compared with genes that experience weaker selection.

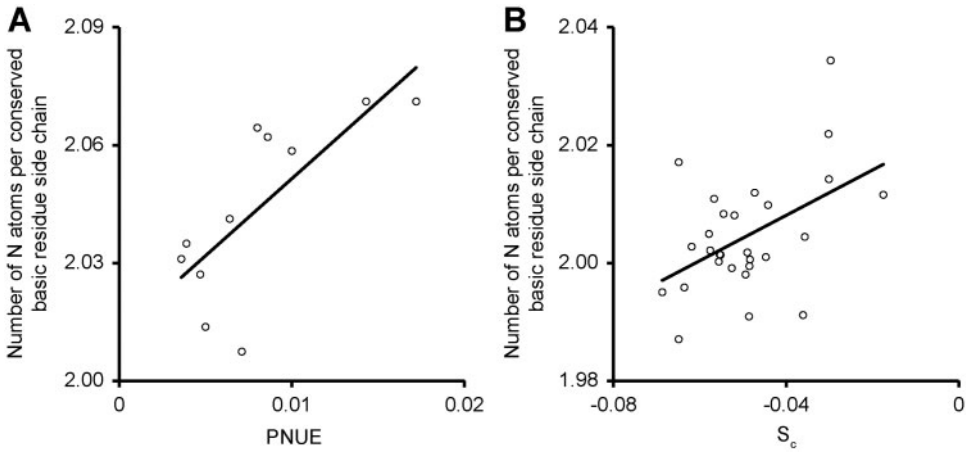

FIG. 4. PNUE modulates the amino acid use at conserved basic sites in protein sequences. (*A*) The relationship between PNUE and the number of nitrogen atoms in side chains of ungapped, conserved, basic sites in single copy orthologous genes ($R^2 = 0.56$, $P = 0.008$). (*B*) The relationship between S_c the number of nitrogen atoms in side chains of ungapped, conserved, basic sites in orthologous genes ($R^2 = 0.23$, $P = 0.01$).

This phenomenon occurs because spontaneous mutations that increase transcript biosynthesis cost will be more deleterious in genes that experience stronger selection to minimize cost irrespective of whether that mutation is synonymous or nonsynonymous (Seward and Kelly 2017). As mutations that are more deleterious will be lost more rapidly, this results in a lower molecular evolution rate for genes that experience stronger selection (Seward and Kelly 2017). This phenomenon has previously been observed for bacterial genes (Seward and Kelly 2017).

To investigate this, both the number of synonymous substitutions per synonymous site (K_s) and the number of nonsynonymous substitutions per nonsynonymous site (K_a) were estimated from pairwise alignments of single copy orthologous genes in a set of 38 plant species (fig. 5). The strength of selection acting on transcript nitrogen cost was also inferred for each individual gene using CodonMuSe (Seward and Kelly 2017). For each species pair, these data were subject to multiple regression analysis to estimate the proportion of variance in K_a or K_s that was explained by variance in S_c between that species pair (supplementary file S2 and table S1 sheet 3, Supplementary Material online). Consistent with the hypothesis, genes experienced stronger selection to minimize transcript nitrogen cost evolved more slowly than those that experience weaker selection (fig. 5 and supplementary file S2, Supplementary Material online). Moreover, variance in the strength of selection explained up to 10% of variance in synonymous site evolutionary rate (supplementary file S2, Supplementary Material online) and ~2% of variance in nonsynonymous site evolutionary rate across all species (fig. 5 and supplementary file S2, Supplementary Material online). Thus, as genome-wide values for S_c are linked to PNUE, it follows that the tempo of the plant molecular clock is modulated by changes in PNUE.

Discussion

There is substantial interspecies variation in the amount of nitrogen required to conduct photosynthesis in plants (Evans 1989; Rotundo and Cipriotti 2017). In this work, it is shown that this variation is a determinant of plant gene and genome composition, and modulates the rate at which plant gene sequences evolve. The findings presented here 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 with domesticated relatives that had been supplemented with nitrogen fertilizer for thousands for years (Acquisti et al. 2009).

Multiple factors have previously been proposed to bias the relative use of synonymous codons. These include but are not limited to; mutational biases during DNA replication and repair (Eyre-Walker 1991; Francino and Ochman 1999; Rao et al. 2011), selection due to difference in abundance of isoaccepting tRNAs (Plotkin et al. 2004), selection to modulate translational efficiency and accuracy (Sorensen et al. 1989; Akashi 1994; Shah and Gilchrist 2011), selection acting on altered gene splicing and protein folding (Shah and Gilchrist 2011), selection on RNA secondary structure (Vandivier et al. 2016), transcription-associated mutation bias (Comeron 2004), mRNA purine loading as a result of growth in high temperature (Lao and Forsdyke 2000; Paz et al. 2004), selection for certain dinucleotides and trinucleotides (Camiolo et al. 2015). The results presented here do not preclude these effects but rather build upon our understanding of factors affecting the relative use of synonymous codons.

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 (Lancaster 2010). Therefore, the mechanistic link between PNUE and molecular evolution presented here has significant implications for our understanding of the past, present, and future of plant evolution. For example, plants with higher PNUE, and thus with higher rates of molecular evolution, will have therefore higher rates of speciation and extinction. As a corollary, evolutionary adaptations that increase PNUE will also increase rates of speciation and extinction. For example, the suite of molecular and anatomical

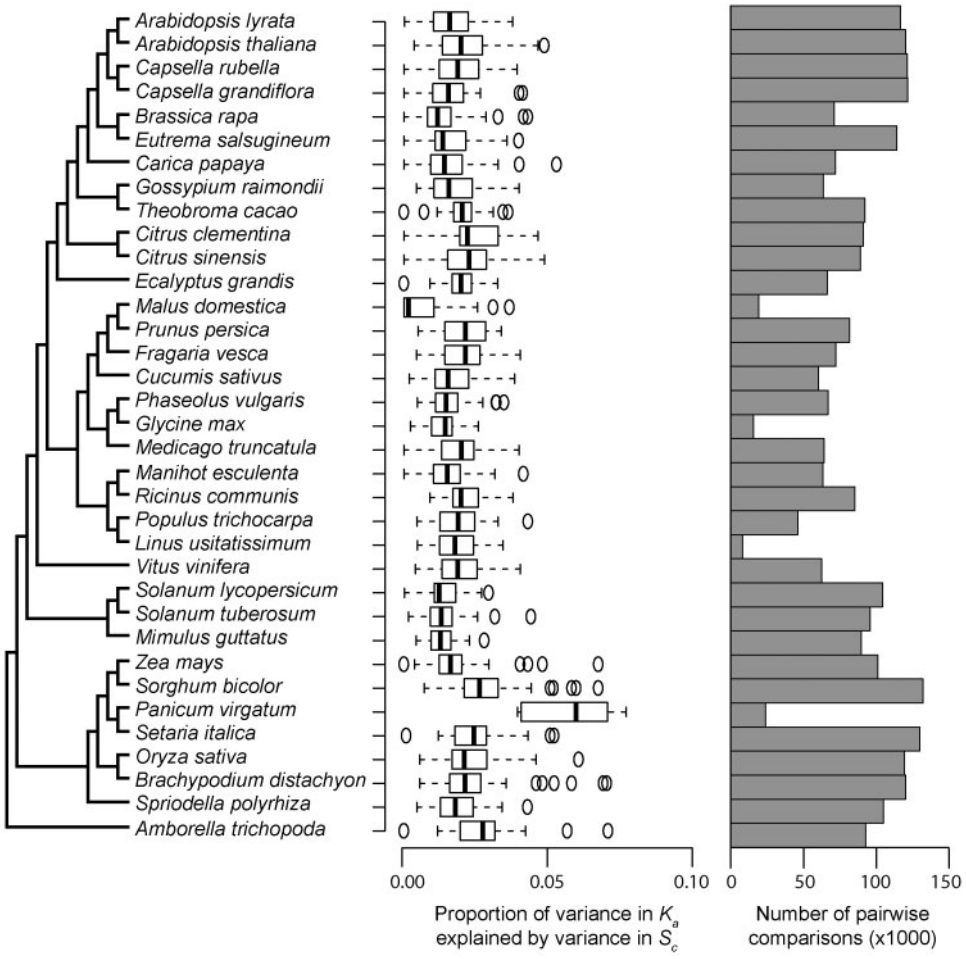

Fig. 5. The strength of selection acting on transcript biosynthesis cost constrains the rate of amino acid sequence evolution. Gray bars depict mean values estimated from all possible pairwise comparisons featuring the species under consideration. Phylogenetic tree adapted from Phytozome (Goodstein et al. 2012).

changes that facilitate the evolution C_4 photosynthesis result in a dramatic reduction in the amount of nitrogen required to conduct photosynthesis. The findings presented here predict that this increase in PNUE would cause a concomitant reduction in the strength of selection on gene sequences and therefore result in an increased rate of molecular evolution. Thus, PNUE-driven increase in molecular evolution rate provides a simple mechanistic explanation for the increase in rates of speciation that are observed concomitant with the evolution of C_4 photosynthesis (Spriggs et al. 2014).

Increases in atmospheric CO_2 concentration cause corresponding increases in PNUE in plants. In the short term, this increase in PNUE is caused by a reduction in the rate of photorespiration (Chollet and Ogren 1975). In the long term, plants also adapt to higher CO_2 concentration by reduction in the investment of cellular resources in photosynthesis protein production (Stitt and Krapp 1999). Thus, when atmospheric CO_2 increases, PNUE increases. The link between PNUE and molecular evolution presented here predicts that this increase in PNUE will cause a corresponding increase in molecular evolution rate, and thus an increase in the rate of plant diversification (fig. 6*A* and *B*). This therefore provides a mechanistic explanation for the observed relationship between plant diversification rates observed in the fossil record and changes in atmospheric CO_2 concentration (McElwain et al. 2011).

Similar to changes in CO_2 availability, changes in other environmental factors such as light availability (fig. 6*C* and *D*) and temperature (fig. 6*E* and *F*) also influence photosynthetic rate and thus PNUE. Unlike CO_2, these other environmental factors vary widely over the surface of the planet. For example, light intensity and temperature are not uniformly distributed on the surface of the earth, but instead decrease as a function of distance from the equator (fig. 6*D* and *F*). This variation is due to the curvature of the earth and the corresponding increase in the angle of the incident light. The findings presented here predict that plant diversification rates will be higher toward the equator where light and temperature are less limiting on photosynthesis and thus PNUE will be higher. These findings therefore provide additional insight into the plant species latitude diversity gradient (Mittelbach et al. 2007; Gillman and Wright 2014), where rates of plant diversification are higher in regions that are closer to the equator.

It has previously been shown that fossil plant genome size exhibits a strong positive correlation with atmospheric CO_2 concentration (Franks et al. 2012). Given that guard cell

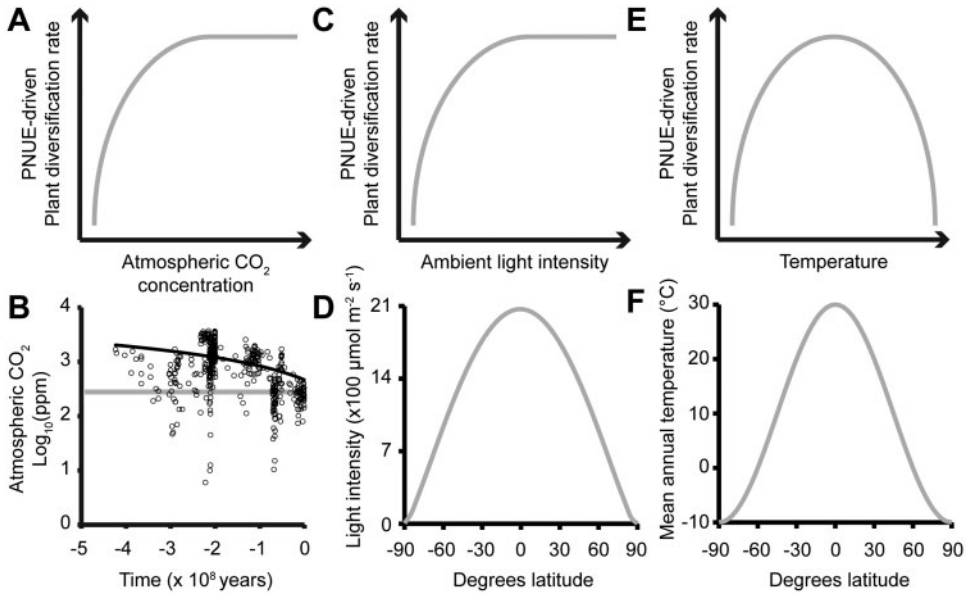

Fig. 6. Photosynthetic nitrogen use efficiency (PNUE) modulates plant diversification. (*A*) Cartoon depicting the proposed relationship between PNUE-driven plant diversification and atmospheric CO_2 concentration. (*B*) Changes in atmospheric CO_2 concentration (ppm = parts per million) over the last 500 My, least squares linear fit shown as solid black line. Gray horizontal line indicates preindustrial CO_2 levels (278 ppm) adapted from Foster et al. (2017). (*C*) Cartoon depicting the relationship between PNUE and light intensity. (*D*) The photon flux density on the surface of the earth at 0° longitude at noon on the vernal equinox. (*E*) Cartoon depicting the proposed relationship between PNUE-driven plant diversification and ambient temperature. (*F*) The average (contemporary) annual temperature at the surface of the earth (Randel et al. 2004).

volume is strongly linked to genome size, it was proposed that selection acting on guard cell volume adapted the aperture of stomata for different atmospheric CO_2 concentrations (Franks et al. 2012). The findings presented here may provide additional mechanistic insight into this phenomenon. Specifically, increases in atmospheric CO_2 concentration cause increases in PNUE. This increase in PNUE causes a concomitant reduction in the strength of selection to minimize resource allocation to transcript sequences, genome sequences (assessed in this work by changes in mutation bias and GC content) and protein sequences (assessed in this work by changes in nitrogen content of amino acid side chains). It follows that this reduction in selection likely also applies to genome size, such that increases in atmospheric CO_2 facilitate concomitant increases in genome size via reduction in selection to minimize resource allocation to DNA. Therefore, changes in PNUE provide an additional mechanistic explanation for the relationship between fossil plant genome size and atmospheric CO_2 concentration. It should be noted here that molecular sequence analysis was not conducted as the genome sequences for these fossil plant species no longer exist.

Conclusion

Plants build their genes and genomes from monomers assembled from inorganic carbon and nitrogen. Of these two, nitrogen is more limiting such that plants that require higher quantities of nitrogen to conduct photosynthesis have less nitrogen available for other uses and thus experience stronger selection to reduce nitrogen investment in gene sequences. A multitude of environmental factors can exacerbate or ameliorate PNUE. Therefore, both the environment and genetic factors can modulate the strength of selection acting to reduce nitrogen investment in gene sequences and hence modulate plant genome composition and molecular evolution. Hence, at multiple scales plant evolution is modulated by the amount of nitrogen required to conduct photosynthesis.

Materials and Methods

Data Sources

The genome sequences and corresponding set of representative gene models for each species were downloaded from Phytozome V12 (Goodstein et al. 2012). The *Helianthus annuus* genome was obtained from (Badouin et al. 2017). Photosynthetic measurements and leaf nitrogen measurements were obtained from Rotundo and Cipriotti (2017).

Inference of Selection Acting on Codon Usage Bias

To obtain the number of tRNA genes in each genome, tRNAscan (Lowe and Eddy 1997) was run on each of the plant genomes. For each species the tRNAscan output file and the complete set of representative coding sequences was analyzed using CodonMuSe (Seward and Kelly 2017). This provided the values for mutation bias (M_b) as well as the composite parameters of selection acting on transcript biosynthesis cost (S_c), and selection acting on translational efficiency (S_t) for each species in this analysis. CodonMuSe by default estimates the proportion of variance in codon use that can be explained by the mutation bias and these selective forces.

Phylogenetic Tree Inference for the 11 Species with PNUE Data

The complete set of proteomes for the 11 species used in this analysis was subject to orthogroup inference using OrthoFinder (Emms and Kelly 2015). In the case of hexaploid wheat genome, only proteins derived from genes present in the wheat A genome were used for orthogroup inference. Orthogroups containing proteins derived from single copy genes in each of the 11 species were selected and aligned using the MAFFT (Katoh and Standley 2016) L-INS-i algorithm. These alignments were trimmed to remove any columns containing gap characters and then concatenated to form a multiple sequence alignment containing 4,949 aligned amino acid positions in each species. This alignment and was subject to bootstrapped maximum likelihood phylogenetic tree inference using IQ-TREE (Nguyen et al. 2015) while estimating the best fitting model of sequence evolution from the data. The best fitting model was inferred to be JTTDCMut + F+G4 by Bayesian information criterion. This tree was used for the phylogenetic least squares analysis and is provided in supplementary file S2, Supplementary Material online.

K_a and K_s Estimation and Comparison with S_c

The predicted proteins from 38 species were downloaded from Phytozome. These species were subject to orthogroup and ortholog inference using OrthoFinder (Emms and Kelly 2015). All 1,406 pairwise comparisons between species were subsequently conducted. Each pairwise comparison comprised the following steps. 1) The full set of single copy orthologs for the species pair under consideration were isolated. 2) The protein sequences for each orthologous pair were aligned using MAFFT (Katoh et al. 2005) L-INS-i and the coding sequences rethreaded back through the protein sequence alignment. 3) The resulting coding sequence alignments were parsed to remove any gap-containing columns. 4) Ungapped alignments containing >100 aligned codons were subject to K_a and K_s inference using KaKsCalculator v2.0 (Wang et al. 2010) using the default settings. Additional data filtering and quality control were carried out as described in supplementary file S2, Supplementary Material online. Individual estimates for S_c and S_t were obtained for each gene in the 38 species using CodonMuSe. Here, the value for M_b in each inference was set to the genome-wide value estimated from an analysis of all genes. Pairwise species comparisons that had > 100 genes satisfying all filtration criteria were selected for further analysis.

The value for K_a and K_s are dependent on several factors:

$$K_x = f\left(T_{d1}, T_{d2,}\, S_1, S_2,\, M_1, M_2,\, N_{e_1}, N_{e_2}\right), \qquad (1)$$

where K_x is either K_a or K_s, T_{d1} is the divergence time in number of generations between species 1 and the most recent common ancestor of the species pair being analyzed, S_1 is the strength of selection acting on the sequence of the gene in species 1, M_1 is the mutation rate species 1, and N_{e1} is the effective population size of species 1. S_c is a composite parameter (Seward and Kelly 2016) that is a product of a component of the selection coefficient S_1 and the effective population size N_{e1}. Thus, each pairwise species comparison was subject to multiple regression analysis using the lm function in R using the following model:

$$\ln K_x = \beta_1 S_{c_1} + \beta_2 S_{c_2},$$

where β_1 thus incorporates both T_{d1} and M_1. Thus, the multiple regression evaluates the component of variance in K_a or K_s that is attributable to both S_{c1} and S_{c2}. The natural log of the K_a *and* K_s estimates were taken as both K_a and K_s are log-normally distributed whereas S_c is normally distributed. All data were confirmed to be normally distributed by the Shapiro–Wilks test for normality prior to use in regression analysis. The mean of the adjusted R^2 for all pairwise comparisons featuring a given species was taken as an estimate the proportion of variance that is explained by variation in S_c for that species.

Quantification of mRNA Abundance

To provide whole-organism mRNA abundance estimates the NCBI SRA database was searched for RNA-Seq samples from whole plants. A single experiment containing three biological replicates of whole-plant RNA-Seq from *Arabidopsis thaliana* 8 day old seedlings was obtained from BioProject PRJNA384979 (Major et al. 2017). The raw reads were downloaded, subject to quality filtering using trimmomatic (Bolger et al. 2014). This was done to remove low quality bases and read-pairs as well as contaminating adaptor sequences prior to quantification. Sequences were searched for all common Illumina adaptors (the default option) and the settings used for read processing by trimmomatic were LEADING: 10 TRAILING: 10 SLIDINGWINDOW: 5: 15 MINLEN: 25. Following trimming, the processed reads were subject to quantification estimation using the complete set of transcript sequences for protein coding genes *A. thaliana* from Phytozome v12 using Salmon v0.9.1 (Patro et al. 2017) with the –seqBias option enabled. TPM values for multiple transcript variants were summed so that a single TPM estimate was provided for each gene for each biological replicate. The mean TPM value of the three biological replicates was taken as the abundance estimate for that gene. Transcripts with mean TPM values ≥ 1 were selected for analysis.

Analysis of Amino Acid Side Chains

Two data sets were constructed to analyze the effect of variation in PNUE on amino acid nitrogen content. The first focused on single copy orthologous genes. Here, the 11 species with PNUE data were subject to orthogroup inference using OrthoFinder (Emms and Kelly 2015). Orthogroups containing only single copy genes ($N = 130$) were identified. The amino acid sequences were aligned using MAFFT L-INS-I (Katoh and Standley 2014) and ungapped aligned positions containing only basic residues (R, H, and K) were selected for further analysis. Of the 130,124 alignments contained ungapped basic positions comprising a total of 2,545 aligned positions. The mean number of N atoms per ungapped aligned position was calculated as the mean of these 2,545 aligned positions.

An equivalent set of single copy orthologous genes was not available for the larger species analysis. This was because of gene duplication and loss which meant that there were no orthogroups present as a single copy gene in all species under consideration. Thus an analogous analysis was performed. Here, the amino acid sequences from orthogroups containing all species were aligned ($N = 4{,}996$). Of these alignments, 1,098 contained ungapped basic positions found in at least one representative sequence from all species with a mean number of 11,835 sites per species. The mean number of N atoms per ungapped aligned position was calculated as the mean of these sites.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

S.K. is a Royal Society University Research Fellow. This work was supported by the Royal Society, the European Union's Horizon 2020 research and innovation programme under grant agreement number 637765, and the Biotechnology and Biological Sciences Research Council (BBSRC) through BB/P003117/1.

Author Contributions

S.K. conducted the study and wrote the manuscript.

References

Acquisti C, Elser JJ, Kumar S. 2009. Ecological nitrogen limitation shapes the DNA composition of plant genomes. *Mol Biol Evol.* 26(5):953–956.

Akashi H. 1994. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics* 136(3):927–935.

Andrade A, Wolfe DW, Fereres E. 1993. Leaf expansion, photosynthesis, and water relations of sunflower plants grown on compacted soil. *Plant Soil* 149(2):175–184.

Anten NPR, Schieving F, Werger MJA. 1995. Patterns of light and nitrogen distribution in relation to whole canopy carbon gain in C3 and C4 mono- and dicotyledonous species. *Oecologia* 101(4):504–513.

Badouin H, Gouzy J, Grassa CJ, Murat F, Staton SE, Cottret L, Lelandais-Briere C, Owens GL, Carrere S, Mayjonade B, et al. 2017. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* 546(7656):148–152.

Bange MP, Hammer GL, Rickert KG. 1997. Effect of specific leaf nitrogen on radiation use efficiency and growth of sunflower. *Crop Sci.* 37(4):1201–1208.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Braune H, Müller J, Diepenbrock W. 2009. Integrating effects of leaf nitrogen, age, rank, and growth temperature into the photosynthesis-stomatal conductance model LEAFC3-N parameterised for barley (*Hordeum vulgare* L.). *Ecol Model.* 220(13–14):1599–1612.

Camiolo S, Melito S, Porceddu A. 2015. New insights into the interplay between codon bias determinants in plants. *DNA Res.* 22(6):461–470.

Cheng L, Fuchigami LH. 2000. CO2 assimilation in relation to nitrogen in apple leaves. *J Horticult Sci Biotechnol.* 75(4):383–387.

Chollet R, Ogren WL. 1975. Regulation of photorespiration in C3 and C4 species. *Bot Rev.* 41(2):137–179.

Comeron JM. 2004. Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. *Genetics* 167(3):1293–1304.

DeJong TM, Day KR, Johnson RS. 1989. Partitioning of leaf nitrogen with respect to within canopy light exposure and nitrogen availability in peach (*Prunus persica*). *Trees* 3(2):89–95.

Drouet JL, Bonhomme R. 2004. Effect of 3D nitrogen, dry mass per area and local irradiance on canopy photosynthesis within leaves of contrasted heterogeneous maize crops. *Ann Bot.* 93(6):699–710.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* 16:157.

Evans JR. 1983. Nitrogen and photosynthesis in the flag leaf of wheat (*Triticum aestivum* L.). *Plant Physiol.* 72(2):297–302.

Evans JR. 1989. Photosynthesis and nitrogen relationships in leaves of C3 plants. *Oecologia* 78(1):9–19.

Evans JR, von Caemmerer S. 2000. Would C4 rice produce more biomass than C3 rice?*. In: Sheehy JE, PL Mitchell, Hardy B, editors. Studies in plant science. Amsterdam, The Netherlands: Elsevier. p. 53–71.

Eyre-Walker AC. 1991. An analysis of codon usage in mammals: selection or mutation bias? *J Mol Evol.* 33(5):442–449.

Fischer RA, Rees D, Sayre KD, Lu Z-M, Condon AG, Saavedra AL. 1998. Wheat yield progress associated with higher stomatal conductance and photosynthetic rate, and cooler canopies. *Crop Sci.* 38(6):1467–1475.

Foster GL, Royer DL, Lunt DJ. 2017. Future climate forcing potentially without precedent in the last 420 million years. *Nat Commun.* 8:14845.

Francino MP, Ochman H. 1999. Isochores result from mutation not selection. *Nature* 400(6739):30–31.

Franks PJ, Freckleton RP, Beaulieu JM, Leitch IJ, Beerling DJ. 2012. Megacycles of atmospheric carbon dioxide concentration correlate with fossil plant genome size. *Philos Trans R Soc Lond B Biol Sci.* 367(1588):556–564.

Fredeen AL, Gamon JA, Field CB. 1991. Responses of photosynthesis and carbohydrate-partitioning to limitations in nitrogen and water availability in field-grown sunflower*. *Plant Cell Environ.* 14(9):963–970.

Gillman LN, Wright SD. 2014. Species richness and evolutionary speed: the influence of temperature, water and area. *J Biogeogr.* 41(1):39–51.

Gimenez C, Connor DJ, Rueda F. 1994. Canopy development, photosynthesis and radiation-use efficiency in sunflower in response to nitrogen. *Field Crops Res.* 38(1):15–27.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS. 2012. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40(Database issue):D1178–D1186.

Grzymski JJ, Dussaq AM. 2012. The significance of nitrogen cost minimization in proteomes of marine microorganisms. *ISME J.* 6(1):71–80.

Hirasawa T, Ozawa S, Taylaran R, Ookawa T. 2010. Varietal differences in photosynthetic rates in rice plants, with special reference to the nitrogen content of leaves. *Plant Prod Sci.* 13(1):53–57.

Hohmann-Marriott MF, Blankenship RE. 2011. Evolution of photosynthesis. *Annu Rev Plant Biol.* 62:515–548.

Ingestad T, Ågren GI. 1992. Theories and methods on plant nutrition and growth. *Physiol Plantarum* 84(1):177–184.

Ingestad T, Lund A-B. 1979. Nitrogen stress in birch seedlings. *Physiol Plantarum* 45(1):137–148.

Katoh K, Kuma K, Miyata T, Toh H. 2005. Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Inform.* 16(1):22–33.

Katoh K, Standley DM. 2014. MAFFT: iterative refinement and additional methods. *Methods Mol Biol.* 1079:131–146.

Katoh K, Standley DM. 2016. A simple method to control over-alignment in the MAFFT multiple sequence alignment program. *Bioinformatics* 32(13):1933–1942.

Keck F, Rimet F, Bouchez A, Franc A. 2016. phylosignal: an R package to measure, test, and explore the phylogenetic signal. *Ecol Evol* 6(9):2774–2780.

Lancaster LT. 2010. Molecular evolutionary rates predict both extinction and speciation in temperate angiosperm lineages. *BMC Evol Biol.* 10:162.

Lao PJ, Forsdyke DR. 2000. Thermophilic bacteria strictly obey Szybalski's transcription direction rule and politely purine-load RNAs with both adenine and guanine. *Genome Res.* 10(2):228–236.

LeBauer DS, Treseder KK. 2008. Nitrogen limitation of net primary productivity in terrestrial ecosystems is globally distributed. *Ecology* 89(2):371–379.

Lindquist JL, Mortensen DA. 1999. Ecophysiological characteristics of four maize hybrids and *Abutilon theophrasti*. *Weed Res.* 39(4):271–285.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25(5):955–964.

Maekawa T, Kokubun M. 2005. Correlation of leaf nitrogen, chlorophyll and Rubisco contents with photosynthesis in a supernodulating soybean genotype Sakukei 4. *Plant Prod Sci.* 8(4):419–426.

Major IT, Yoshida Y, Campos ML, Kapali G, Xin XF, Sugimoto K, de Oliveira Ferreira D, He SY, Howe GA. 2017. Regulation of growth-defense balance by the JASMONATE ZIM-DOMAIN (JAZ)-MYC transcriptional module. *New Phytol.* 215(4):1533–1547.

Malcolm P, Holford P, McGlasson B, Barchia I. 2008. Leaf development, net assimilation and leaf nitrogen concentrations of five Prunus rootstocks in response to root temperature. *Sci Horticult.* 115(3):285–291.

McElwain JC, Willis KJ, Niklas KJ. 2011. Long-term fluctuations in atmospheric CO2 concentration influence plant speciation rates. In: Parnell JAN, Jones MB, Waldren S, Hodkinson TR, editors. Climate change, ecology and systematics. Cambridge: Cambridge University Press. p. 122–140.

Mittelbach GG, Schemske DW, Cornell HV, Allen AP, Brown JM, Bush MB, Harrison SP, Hurlbert AH, Knowlton N, Lessios HA, et al. 2007. Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography. *Ecol Lett.* 10(4):315–331.

Muchow RC, Sinclair TR. 1994. Nitrogen response of leaf photosynthesis and canopy radiation use efficiency in field-grown maize and sorghum. *Crop Sci.* 34(3):721–727.

Müller J, Wernecke P, Diepenbrock W. 2005. LEAFC3-N: a nitrogen-sensitive extension of the CO2 and H2O gas exchange model LEAFC3 parameterised and tested for winter wheat (*Triticum aestivum* L.). *Ecol Model.* 183(2–3):183–210.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274.

Oaks A. 1994. Efficiency of nitrogen utilization in C3 and C4 cereals. *Plant Physiol.* 106(2):407–414.

Ohsumi A, Hamasaki A, Nakagawa H, Yoshida H, Shiraiwa T, Horie T. 2007. A model explaining genotypic and ontogenetic variation of leaf photosynthetic rate in rice (*Oryza sativa*) based on leaf nitrogen content and stomatal conductance. *Ann Bot.* 99(2):265–273.

Osaki M, Shinano T, Kaneda T, Yamada S, Nakamura T. 2001. Ontogenetic changes of photosynthetic and dark respiration rates in relation to nitrogen content in individual leaves of field crops. *Photosynthetica* 39:205–213.

Paponov IA, Engels C. 2003. Effect of nitrogen supply on leaf traits related to photosynthesis during grain filling in two maize genotypes with different N efficiency. *J Plant Nutr Soil Sci.* 166(6):756–763.

Paponov IA, Sambo P, Erley G. S a'm, Presterl T, Geiger HH, Engels C. 2005. Grain yield and kernel weight of two maize genotypes differing in nitrogen use efficiency at various levels of nitrogen and carbohydrate availability during flowering and grain filling. *Plant Soil* 272(1–2):111–123.

Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. 2017. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* 14(4):417–419.

Paz A, Mester D, Baca I, Nevo E, Korol A. 2004. Adaptive role of increased frequency of polypurine tracts in mRNA sequences of thermophilic prokaryotes. *Proc Natl Acad Sci U S A.* 101(9):2951–2956.

Plotkin JB, Robins H, Levine AJ. 2004. Tissue-specific codon usage and the expression of human genes. *Proc Natl Acad Sci U S A.* 101(34):12588–12591.

Randel W, Udelhofen P, Fleming E, Geller M, Gelman M, Hamilton K, Karoly D, Ortland D, Pawson S, Swinbank R, et al. 2004. The SPARC intercomparison of middle-atmosphere climatologies. *J Climate* 17(5):986–1003.

Rao Y, Wu G, Wang Z, Chai X, Nie Q, Zhang X. 2011. Mutation bias is the driving force of codon usage in the *Gallus gallus* genome. *DNA Res.* 18(6):499–512.

Rodríguez D, Zubillaga MM, Ploschuk EL, Keltjens WG, Goudriaan J, Lavado RS. 1998. Leaf area expansion and assimilate production in sunflower (*Helianthus annuus* L.) growing under low phosphorus conditions. *Plant Soil* 202:133–147.

Romero-Aranda R, Syvertsen JP. 1996. The influence of foliar-applied urea nitrogen and saline solutions on net gas exchange of citrus leaves. *J Am Soc Horticult Sci.* 121:501–506.

Rosati A, Esparza G, DeJong TM, Pearcy RW. 1999. Influence of canopy light environment and nitrogen availability on leaf photosynthetic characteristics and photosynthetic nitrogen-use efficiency of field-grown nectarine trees. *Tree Physiol.* 19(3):173–180.

Rotundo JL, Borras L. 2016. Reduced soybean photosynthetic nitrogen use efficiency associated with evolutionary genetic bottlenecks. *Funct Plant Biol*. 43:862–869.

Rotundo JL, Cipriotti PA. 2017. Biological limits on nitrogen use for plant photosynthesis: a quantitative revision comparing cultivated and wild species. *New Phytol.* 214(1):120–131.

Seward EA, Kelly S. 2016. Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms. *Genome Biol.* 17(1):226.

Seward EA, Kelly S. 2017. Selection-driven cost-efficiency optimisation of transcript sequences determines the rate of gene sequence evolution in bacteria. *bioRxiv*. 136861. doi: https://doi.org/10.1101/136861.

Shah P, Gilchrist MA. 2011. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc Natl Acad Sci U S A.* 108(25):10231–10236.

Sorensen MA, Kurland CG, Pedersen S. 1989. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol.* 207(2):365–377.

Spriggs EL, Christin PA, Edwards EJ. 2014. C4 photosynthesis promoted species diversification during the Miocene grassland expansion. *PLoS One* 9(5):e97722.

Stitt M, Krapp A. 1999. The interaction between elevated carbon dioxide and nitrogen nutrition: the physiological and molecular background. *Plant Cell Environ.* 22:583–621.

Tominaga J, Yabuta S, Fukuzawa Y, Kawasaki S-I, Jaiphong T, Suwa R, Kawamitsu Y. 2015. Effects of vertical gradient of leaf nitrogen content on canopy photosynthesis in tall and dwarf cultivars of sorghum. *Plant Prod Sci.* 18(3):336–343.

Trápani N, Hall AJ. 1996. Effects of leaf position and nitrogen supply on the expansion of leaves of field grown sunflower (*Helianthus annuus* L.). *Plant Soil* 184(2):331–340.

Vandivier LE, Anderson SJ, Foley SW, Gregory BD. 2016. The conservation and function of RNA secondary structure in plants. *Annu Rev Plant Biol.* 67:463–488.

Vos J, Putten PEL. v d, Birch CJ. 2005. Effect of nitrogen supply on leaf appearance, leaf growth, leaf nitrogen economy and photosynthetic capacity in maize (*Zea mays* L.). *Field Crops Res.* 93(1):64–73.

Vos J, van der Putten PEL. 1998. Effect of nitrogen supply on leaf growth, leaf nitrogen economy and photosynthetic capacity in potato. *Field Crops Res.* 59(1):63–72.

Vos J, van der Putten PEL. 2001. Effects of partial shading of the potato plant on photosynthesis of treated leaves, leaf area expansion and allocation of nitrogen and dry matter in component plant parts. *Eur J Agronomy* 14(3):209–220.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* 8(1):77–80.