Title

RuBisCO adaptation is more limited by phylogenetic constraint than by catalytic trade-off

Authors

Jacques W. Bouvier[1,2], David M. Emms[1], Timothy Rhodes[3], Jochem R. Nielsen[2], Jai S. Bolton[2], Alice Eddershaw[2], Amelia Brasnett[2], Ana Unitt[2], Spencer M. Whitney[3], Steven Kelly[1]

Affiliations

[1] Department of Plant Sciences, University of Oxford, South Parks Road, Oxford, OX1 3RB, United Kingdom

[2] Doctoral Training Centre, University of Oxford, Keble Road, Oxford, OX1 3NP, United Kingdom

[3] Research School of Biology, Australian National University, Canberra, Australian Capital Territory 2601, Australia

Corresponding Author

Name: Steven Kelly

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

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

Abstract

RuBisCO assimilates CO_2 to form the sugars that fuel life on earth. Correlations between RuBisCO kinetic traits across species have led to the proposition that RuBisCO adaptation is constrained by catalytic trade-offs. However, these correlations were founded on the assumption that kinetic measurements in different species are independent. Here, we show that this assumption is incorrect in angiosperms by evaluating the dependence of variation in RuBisCO kinetic traits on the phylogenetic tree that relates the enzymes. We show that there is significant phylogenetic signal in all carboxylase kinetic traits in angiosperms, and significant phylogenetic signal in the Michaelis constant for O_2 in species that conduct C_3 photosynthesis. When accounting for this non-independence, we show that the catalytic trade-off between carboxylase turnover and the Michaelis constant for CO_2 is weak (~30 % dependency) and that the correlations between all other RuBisCO kinetic traits are either not-significant or marginal (<9 % dependency). Finally, we demonstrate that phylogenetic constraints limit RuBisCO adaptation to a greater extent than catalytic trade-offs. Thus,

the biochemical landscape of RuBisCO adaptation in angiosperms is predominantly limited by phylogenetic constraint and a partial trade-off between carboxylase turnover and the Michaelis constant for CO_2.

Introduction

The vast majority of organic carbon on Earth entered the biosphere via the catalytic pocket of RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) (Beer *et al.*, 2010). Whilst there are several metabolic contexts in which this RuBisCO-mediated reaction can occur, the most important of these in terms of global net primary production is photosynthesis (Andersson and Backlund, 2008). Here, RuBisCO plays a central role in the initial step of the Calvin-Benson-Bassham reductive pentose phosphate pathway, catalysing the fixation of CO_2 onto the acceptor molecule ribulose 1,5-bisphosphate (RuBP) to ultimately synthesise carbohydrates. In plants, this reaction is carried out by RuBisCO Form IB which is structurally composed of both large (RbcL) and small (RbcS) subunits in a hexadecameric configuration (Schneider, Lindqvist and Brändén, 1992; Tabita *et al.*, 2008). Although this complex quaternary structure characteristic of all Form I (A-D) RuBisCO is the most common in nature (Andersson and Backlund, 2008), only the large subunit is essential for catalysis (Andrews, 1988; Lee and Tabita, 1990) as the active sites are located exclusively on this subunit (Whitney, Houtz and Alonso, 2011). In addition, disparate lineages of Form II, III and IV RuBisCO exist which catalyse CO_2 fixation and lack the small subunit entirely (Tabita *et al.*, 2008). However, despite not being involved in the catalytic chemistry directly, the small subunit does have an indirect effect on the catalytic properties (Read and Tabita, 1992b, 1992a; Spreitzer, Peddi and Satagopan, 2005; Ishikawa *et al.*, 2011; Joshi *et al.*, 2015; Fukayama *et al.*, 2019; Martin-Avila *et al.*, 2020) and activity (Andrews, 1988; Lee and Tabita, 1990; Lee, Berka and Tabita, 1991) of RuBisCO, and thus is important to the function of Form I RuBisCO.

Perhaps unsurprisingly given its role as the entry point for carbon into the global food chain, RuBisCO is the most abundant protein on Earth (Ellis, 1979) with a global mass of ~0.7 gigatons (Bar-On and Milo, 2019). This extraordinary abundance is in part due to the inefficiency of RuBisCO as a catalyst. Specifically, RuBisCO has a low rate of CO_2 assimilation (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010) and is poorly able to discriminate CO_2 and O_2 (Ogren and Bowes, 1971) causing it to catalyse both a carboxylation and an oxygenation reaction (Bowes, Ogren and Hageman, 1971;

Chollet, 1977; Sharkey, 2020). The RuBisCO-mediated oxygenation of RuBP results in the production of 2-phosphoglycolate which must then be metabolised to recover the fixed carbon and avoid depletion of metabolite pools (Eckardt, 2005; Sharwood, 2017). This carbon scavenging process, known as photorespiration, consumes ATP and reducing power and also liberates ammonia which must then be re-assimilated (Peterhansel *et al.*, 2010). Although the oxygenation reaction catalysed by RuBisCO is thought not to be deleterious in the anoxic environment prevalent when the enzyme first evolved (Nisbet *et al.*, 2007; Erb and Zarzycki, 2018), under current atmospheric conditions it can comprise a quarter of all RuBisCO reactions in terrestrial plants (Ehleringer *et al.*, 1991). Thus, despite oxygenation serving a number of beneficial functions (Busch, 2020), at its current rate it represents a substantial metabolic burden reducing the productivity of some plants by up to 50 % (Ogren, 1984; Bauwe *et al.*, 2012).

Given the high energic cost incurred by the RuBisCO oxygenation reaction, a number of photoautotrophic organisms have evolved mechanisms to reduce the frequency of its occurrence. Collectively referred to as CO_2-concentrating mechanisms, these function to increase the concentration of CO_2 relative to O_2 in the vicinity of RuBisCO and thus increase the relative frequency of carboxylation reactions (Meyer and Griffiths, 2013). The simplest of these CO_2-concentrating mechanisms work by compartmentalising RuBisCO in a protein shell or sub-cellular micro-compartment within which inorganic carbon is concentrated using active transport (Raven, 1997; Giordano, Beardall and Raven, 2005). More complex CO_2-concentrating mechanisms, such as those found in plants that conduct C_4 photosynthesis or crassulacean acid metabolism, require anatomical and biochemical adaptations of cells and tissue types (Bräutigam *et al.*, 2017; Edwards, 2019). These latter mechanisms utilise the oxygen insensitive enzyme phospho*enol*pyruvate carboxylase to separate primary CO_2 assimilation from RuBisCO-mediated photosynthetic CO_2 reduction either spatially (C_4) (Sage, Sage and Kocacinar, 2012) or temporally (crassulacean acid metabolism) (Dodd *et al.*, 2002). In doing so, both C_4 and crassulacean acid metabolism are able to largely circumvent the oxygenation reaction of RuBisCO resulting in highly successful adaptive phenotypes (Still *et al.*, 2003; Lüttge, 2004).

The observation that natural selection has devised an array of CO_2-concentrating mechanisms rather than improve the CO_2 specificity of RuBisCO has led many to question whether altering the kinetics of the enzyme is possible (Ogren, 1984; Parry *et al.*, 2007, 2013; Whitney, Houtz and Alonso, 2011; Sharwood, Ghannoum and Whitney, 2016; Araújo *et al.*, 2017; Sharwood, 2017; Sharkey, 2020). Here, the proposition that RuBisCO kinetics cannot be improved was supported by observations that the oxygenase and carboxylase activities of RuBisCO appear to be tightly linked (Badger and Lorimer, 1976; Chollet and Anderson, 1976). Subsequently, multiple studies have supported this suggestion by reporting strong antagonistic relationships between RuBisCO specificity ($S_{C/O}$), carboxylase turnover (k_{catC}) and the Michaelis constant (i.e., an inverse measure of substrate affinity for an enzyme) for CO_2 (K_C), as well as between K_C and the Michaelis constant for O_2 (K_O) (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Collectively, these studies have led to the hypothesis that severe kinetic trait trade-offs hamstring the inherent efficiency by which the enzyme can catalyse CO_2 fixation, and that contemporary RuBisCO are near perfectly adapted within this heavily constrained catalytic landscape (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). However, new evidence has started to overturn this paradigm of RuBisCO evolution. First, a recent analysis of the correlative nature of RuBisCO kinetics has demonstrated that associations between kinetic traits are weakened when a large number of species are considered (Flamholz *et al.*, 2019). Further, long-standing engineering efforts to improve RuBisCO have begun to reveal signs of success, where the enzyme variants evolved by directed evolution show deviations from the proposed catalytic trade-offs between $S_{C/O}$, k_{catC} and K_C (Wilson *et al.*, 2018; Zhou and Whitney, 2019). Thus together, these results indicate that RuBisCO kinetic traits are perhaps not as inextricably linked as originally thought, and suggest that there is scope for increasing the catalytic efficiency of the enzyme as has happened in nature for RuBisCO in some red algae (Andersson and Backlund, 2008).

Although the catalytic traits of RuBisCO appear to be correlated, there are flaws to inferring causality from this correlation. This is because previous analyses that have inferred correlations between kinetic traits have assumed that measurements of RuBisCO kinetic traits in different species are independent. However, this assumption has never formally been tested and is unlikely to be true because RuBisCO in all extant species are related to each other by descent from a single ancestral

gene. This means that the gene sequences encoding the enzyme in closely related species are more similar than the gene sequences from species drawn at random, a feature of RuBisCO which has long been exploited in systematics and evolutionary analyses to serve as an accurate proxy for the phylogenetic relationship between species (Gielly and Taberlet, 1994; APG, 1998, 2016). By extension of this fact, it should be expected that closely related enzymes will generally also have similar kinetics, with the extent of this similarity being dependent on the underlying tree describing the relationship between species. This phenomenon, which is known as phylogenetic signal (also known as phylogenetic co-variation), can cause spurious correlations in measured trait values between species unless the structure of the phylogenetic tree is taken into consideration (Felsenstein, 1985; Grafen, 1989; Pagel and Harvey, 1989; Garland, 2001). Thus, as previous analyses of RuBisCO kinetics have not assessed whether phylogenetic signal exists in RuBisCO kinetic traits, nor accounted for any phylogenetic signal which may exist, it is possible that the observed catalytic trade-offs are, either wholly or in part, an artefact caused by phylogenetic signal.

Here, we assess the phylogenetic signal and tree-like evolution of RuBisCO kinetic traits in a large dataset of angiosperms to re-evaluate the extent by which RuBisCO is constrained by phylogenetic effects and catalytic trade-offs. We demonstrate that there is a significant phylogenetic signal in RuBisCO specificity, carboxylase turnover and the Michaelis constant for CO_2 in angiosperms, and also a significant phylogenetic signal in the Michaelis constant for O_2 in species that conduct C_3 photosynthesis. This means that phylogenetic constraints cause the similarity of kinetic measurements between species to vary as a function of their evolutionary distance. When this phylogenetic dependency is correctly accounted for by using phylogenetic least squares regression, we reveal an attenuation in the interdependence between RuBisCO kinetic traits and show that previously inferred correlations have been over-estimated. Specifically, after phylogenetic correction we find evidence of a weak power-law correlation between carboxylase turnover and CO_2 affinity (~30 % dependency), and demonstrate that all other measured associations between specificity, carboxylase turnover and the Michaelis constant for both CO_2 and O_2 substrates are either non-significant or marginal (<9 % dependency). Moreover, we find that phylogenetic constraints, most likely resulting from low genetic variability and a slow rate of molecular evolution, constrain RuBisCO kinetic evolution to a greater extent than catalytic trade-offs. This new insight challenges the widely-

held belief that the adaptive landscape of RuBisCO is highly constrained by kinetic trait trade-offs, and offers encouragement to efforts which aim to increase yields in food, fibre and fuel crops by engineering RuBisCO variants with increased catalytic efficiency.

Results

RuBisCO kinetic data

A dataset comprising kinetic measurements for RuBisCO isolated from different photoautotrophs was obtained from (Flamholz *et al.*, 2019). Measurements of specificity ($S_{C/O}$) for CO_2 relative to O_2 (i.e., the overall carboxylation/oxygenation ratio of RuBisCO under defined concentrations of CO_2 and O_2 gases) in this dataset were normalised in order to overcome discrepancies between values determined using an oxygen electrode assay (Parry, Keys and Gutteridge, 1989) and high precision gas phase controlled 3H-RuBP-fixation assays (Kane et al., 1994) (see methods). Interrogation of this data was restricted to the angiosperms in order to minimize long-branch effects in subsequent analyses (Su and Townsend, 2015). It was also restricted to those species with measurements of $S_{C/O}$ (mol.mol^{-1}), maximum carboxylase turnover rate per active site (k_{catC}; s^{-1}), and the Michaelis constant (i.e., the substrate concentration in µM at half saturated catalysed rate) for both CO_2 (K_C) and O_2 (K_O). The Michaelis constant for CO_2 in 20.95 % O_2 air (K_C^{air}) was also inferred as a function of both K_C and K_O (see methods). Of the 137 species that satisfied these filtration criteria, 19 also had measurements of the Michaelis constant for RuBP (K_{RuBP}). From here on, these constants and rates are collectively termed kinetic traits, where $S_{C/O}$, k_{catC}, K_C and K_C^{air} are referred to as carboxylase-related kinetic traits, and K_O as the oxygenase-related kinetic trait.

Significant changes in RuBisCO kinetics traits are with the evolution of C_4 photosynthesis

Consistent with previous analyses (Flamholz *et al.*, 2019), all kinetic traits were log transformed to ensure they conformed to the distribution assumptions of the statistical analyses herein. Several dependencies in RuBisCO kinetic traits on photosynthetic type were found, with differences observed in carboxylase-related trait values between plant species that conducted C_3 and C_4 photosynthetic pathways (Figure 1). Specifically, the mean of the distribution of RuBisCO $S_{C/O}$ values in C_4 species (mean $S_{C/O}$ = 78.7 mol.mol^{-1}) was lower than that observed for RuBisCO in C_3 species (mean $S_{C/O}$ = 89.9 mol.mol^{-1}) (Figure 1; $p < 0.001$, t-test). Conversely, the mean of the distribution of RuBisCO k_{catC} values was higher in C_4 species (mean k_{catC} = 4.2 s^{-1}) than in C_3 species (mean k_{catC}

= 3.2 s^{-1}) (Figure 1; $p < 0.001$, t-test). The means of the distributions of both K_C and K_C^{air} were also found to be higher in C_4 species (mean K_C = 19.0 µM, mean K_C^{air} = 29.9 µM) than in C_3 plants (mean K_C = 15.4 µM, mean K_C^{air} = 23.6 µM) (Figure 1; $p < 0.05$ and $p < 0.05$, t-test, respectively). In contrast, no significant difference was observed in K_O between C_3 species (mean K_O = 481.0 µM) and C_4 species (mean K_O = 466.7 µM) (Figure 1; $p > 0.05$, t-test). However, variation in K_O was found to be considerably greater in C_4 species (95 % CI = [379.1, 574.6]) than in C_3 species (95 % CI = [457.1, 506.0]) ($p < 0.01$; Levene test). Moreover, although the restricted number of K_{RuBP} measurements did not allow statistical differences to be assessed between photosynthetic groups, the distribution of this trait appeared to show higher variability in C_4 species, similar to that observed for K_O. Taken together, this data supports a hypothesis that RuBisCO in C_4 plants have adapted to saturating CO_2 concentrations by increasing carboxylase turnover at the expense of lower specificity and CO_2 affinity, whilst RuBisCO affinity for O_2 (and possibly RuBP) appears to exhibit increased variability in C_4 species. These differences in the kinetic traits of RuBisCO between C_3 and C_4 plants were robust to phylogenetic correction (see below).

Owing to limited kinetic measurements for RuBisCO in C_3-C_4 intermediate and C_4-like species which respectively represent early and late transition states along the evolutionary continuum from C_3 to C_4 photosynthesis, it was not possible to assess changes in RuBisCO kinetics in these plants relative to the ancestral C_3 and derived C_4 photosynthetic types. Nevertheless, trait values of RuBisCO $S_{C/O}$ in both evolutionary intermediate C_3-C_4 and C_4-like states appear to closely resemble the distribution observed in C_4 species (Figure 1), thus indicating adaptation of this trait potentially occurred early during the evolution of C_4 photosynthesis.

Significant phylogenetic signal exists in RuBisCO carboxylase-related kinetic traits

To assess whether RuBisCO in different angiosperms display similar kinetics as a consequence of their phylogenetic relationship, the quantitative kinetic traits of the enzyme were analyzed in the context of the phylogenetic tree by which the species are related (Figure 2). Here, all kinetic traits were subject to interrogation for phylogenetic signal (Table 1) except for K_{RuBP}, which was omitted owing to the limited number of measurements available for this trait. For these analyses, several statistical tools varying in their approach to phylogenetic signal detection were implemented and the presence or absence of phylogenetic signal in each trait was judged by the majority result (i.e., the

result of ≥ 3 out of 5 methods tested). Out of the methods utilized, Pagel's lambda (Pagel, 1999) and Blomberg's K and K* (Blomberg, Garland and Ives, 2003) analyse the distribution of trait values in extant species using an explicit Brownian motion model of trait evolution in which the traits evolve stochastically on the underlying phylogenetic tree at a uniform rate and independently among branches. In contrast, Moran's I (Gittleman and Kot, 1990) and Abouheif's Cmean (Abouheif, 1999) do not invoke any specific aspect of evolutionary theory, but instead test for phylogenetic signal by assessing the correlation of trait values across evolutionary distance on the species tree using the concept of autocorrelation adopted from the field of spatial statistics (Cheverud, Dow and Leutenegger, 1985, 1986). For further discussion of the differences between these phylogenetic signal detection methods see (e.g., Münkemüller *et al.*, 2012).

Irrespective of the methodological approach used for inference, a significant phylogenetic signal was observed in all of the carboxylase-related kinetic traits, $S_{C/O}$, k_{catC}, K_C, and K_C^{air} (Table 1; Figure 2). However, the strength of this signal varied across the different methods (Table 1). In contrast, phylogenetic signal was not detected for the oxygenase-related kinetic trait K_O (Table 1; Figure 2). These measurements of phylogenetic signal were demonstrated to not suffer from overfitting due to the use of the *rbcL* gene to infer the phylogenetic tree (Supplementary File 1, Figure S1, Figure S2, and Table S1). Overall, this means that carboxylase-related (but not oxygenase-related) kinetic traits in different species are similar as a consequence of their phylogenetic relationship. Therefore, conventional approaches to measure correlations that assume independence between observations of carboxylase-related kinetic trait values are invalid, and correlation coefficients computed using such approaches have likely been over-estimated.

As the presence of phylogenetic signal in RuBisCO carboxylase-related kinetic traits violates the assumption that species measurements are independent, kinetic variations observed between C_3 and C_4 individuals (Figure 1) were re-analysed accounting for the evolutionary relationships among angiosperms. Here, differences in the distributions of $S_{C/O}$, k_{catC}, K_C and K_C^{air} between C_3 and C_4 species (Figure 1) were found to be robust after corrections were applied based on the structure of the phylogenetic tree (Supplementary File 1, Table S2). Thus, there is genuine adaptation in RuBisCO kinetics that are associated with the evolution of C_4 photosynthesis, such that the

emergence of the C_4 carbon concentrating mechanism is accompanied by a decreased specificity and CO_2 affinity, and an increased carboxylase turnover.

Significant phylogenetic signal exists in the RuBisCO K_O in C_3 plants

Based on the positions of C_3-C_4 intermediate, C_4-like, and C_4 species in the phylogenetic tree (Supplementary File 1, Figure S1), multiple independent evolutions toward C_4 photosynthesis are present in the dataset. Furthermore, given that transition to C_4 photosynthesis is found above to be associated with adaptive changes in RuBisCO kinetic traits including a reduction in $S_{C/O}$, an increase in k_{catC}, K_C and K_C^{air}, as well as increased variability in K_O (Figure 1; Supplementary File 1, Table S2), it was hypothesised that a failure to account for kinetic differences associated with photosynthetic type may have confounded estimations of phylogenetic signal. For example, kinetic modifications associated with the evolution of C_4 photosynthesis may cause larger differences in RuBisCO kinetics among closely related C_3 and C_4 species than expected based on evolutionary distance alone. Similarly, the independent evolution of C_4 photosynthesis in distantly related plant lineages could also cause evolutionarily distant species to have similar kinetic trait values by convergence. To evaluate the extent by which these respective issues might have affected quantification of phylogenetic dependency, the phylogenetic signal of RuBisCO kinetic traits was re-assessed using only the C_3 species present in Supplementary Figure S1 (i.e., with C_3-C_4 intermediate, C_4-like, and C_4 species removed). In general, estimates of phylogenetic signal in the carboxylase-related kinetic traits in C_3 species (Table 2) agreed with those observed when all species were considered on the phylogenetic tree (Table 1). Specifically, a phylogenetic signal of similar strength and significance was observed in $S_{C/O}$, k_{catC} and K_C for each of the detection methods across both sets of analyses (Table 1 and Table 2). In addition, the discrepancies in signal strength between the statistical methods previously observed (Table 1) were recapitulated in the analysis using only C_3 species (Table 2), thus indicating that these differences are not caused by a failure to control for photosynthetic type, but instead more likely represent distinctions in the assumptions and aspects of phylogenetic signal measured by each test (Hardy and Pavoine, 2012; Münkemüller *et al.*, 2012). In summary therefore, there is a statistical phylogenetic co-variation in RuBisCO specificity, carboxylase turnover and the Michaelis constant for CO_2 in angiosperms that is independent of photosynthetic type.

In contrast to the analysis of all species (Table 1) a significant signal was observed in K_O when the analysis was restricted to C_3 species (Table 2). Thus, both the oxygenase-related and carboxylase-related traits of RuBisCO evolve in a tree-like manner in C_3 species. Furthermore, unlike the other carboxylase-related kinetic traits, the phylogenetic signal in K_C^{air} is found to increase in strength when the analysis is restricted to C_3 species. This result is a corollary of the fact that K_C^{air} is computed here from both K_C and K_O. Taken together, these results reveal that all kinetic traits of RuBisCO have a significant phylogenetic dependency in C_3 species.

Correlations between kinetic traits are weak in angiosperms and further relaxed after correcting for phylogenetic signal

Given the finding that RuBisCO kinetic traits exhibit significant phylogenetic signal (Table 1; Figure 2), it is possible that previously reported associations between RuBisCO kinetic traits are an artefact of this signal. This is because prior analyses which did not account for the evolutionary relationships among (and non-independence of) species may suffer from over-estimated correlation coefficients as a consequence of the phylogenetic co-variation which is observed in the kinetic traits (Figure 3A). In order to evaluate the severity by which phylogenetic co-variation may have influenced previous results, the correlations observed in the kinetic trait data using both phylogenetic and non-phylogenetic regression methods were compared (Figure 3B and 3C).

Using a standard non-phylogenetic approach the relationships between kinetic traits of RuBisCO were consistent in both standard linear and least squares regression models (Supplementary File 1, Figure S3). The direction of correlations observed (Figure 3B) match those previously reported (Flamholz *et al.*, 2019). Specifically, significant positive power-law (log—log scale) relationships were found between k_{catC} and both K_C and K_C^{air} (Figure 3B and 3C). A significant positive correlation was also observed between the respective Michaelis constants for both CO_2 and O_2 substrates, K_C and K_O (Figure 3B and 3C). In addition, significant inverse power-law correlations were observed between $S_{C/O}$ and all other carboxylase-related kinetic traits investigated, including k_{catC}, K_C and K_C^{air} (Figure 3B and 3C). In contrast, K_O did not co-vary with either $S_{C/O}$ or k_{catC} (Figure 3B and 3C), whilst K_{RuBP} did not appear to be tightly linked to any kinetic trait from the limited number of observations that are available (Supplementary File 1, Figure S3). Thus, all pairwise relationships between the

carboxylase-related kinetic traits $S_{C/O}$, k_{catC} and either K_C or K_C^{air} were significant, whilst the oxygenase-related trait K_O was only correlated with K_C.

Although kinetic trade-offs inferred using non-phylogenetic methods were concordant in direction with those previously described (Flamholz *et al.*, 2019), they were substantially reduced in magnitude when the analysis was focussed solely on the angiosperms. For example, the strength of the correlation between $S_{C/O}$ and K_C in angiosperm RuBisCO (9.9 % variance explained; Figure 3C) is attenuated by a factor of 4.4-fold when compared to that previously reported using a larger range of photoautotrophs (43.6 % variance explained (Flamholz *et al.*, 2019)). Moreover, a 3.2-fold reduction was found in the dependency between $S_{C/O}$ and k_{catC} in angiosperms (9.8 % variance explained; Figure 3C) in comparison to that reported based on the larger range of photoautotrophs (31.4 % variance explained (Flamholz *et al.*, 2019)), whilst the antagonistic correlation observed between K_C and K_O (20.9 % variance explained; Figure 3C) was also weakened by a factor of 1.5-fold relative to previous reports (31.4 % variance explained (Flamholz *et al.*, 2019)). In contrast, the dependency between K_C and k_{catC} was ~1.5-fold stronger when only angiosperms are assessed, increasing from 23.0 % (Flamholz *et al.*, 2019) to 34.2 % in this study (Figure 3C). Thus overall, even in the absence of correctly accounting for the phylogenetic relationship between RuBisCO, the apparent catalytic trade-offs observed in angiosperms are weaker than previously thought. This supports the assertion that the biochemical landscape in angiosperms is relatively weakly constrained by catalytic trade-offs, and adds to a growing body of knowledge detailing differences in RuBisCO kinetics (Jordan and Ogren, 1981; Andersson and Backlund, 2008; Young *et al.*, 2016; Heureux *et al.*, 2017) and kinetic trade-offs (Iñiguez *et al.*, 2020) across the RuBisCO tree of life.

However, significant phylogenetic signal is present in RuBisCO kinetic traits (Table 1). Thus, phylogenetic generalized least squares regression analyses (Felsenstein, 1985) were conducted in order to correctly estimate the magnitude of the catalytic trade-offs whilst accounting for phylogenetic co-variation. In comparison to non-corrected correlations, the phylogenetic regression caused a reduction in the majority of kinetic trade-offs (Figure 3C). The largest reduction was observed was for the correlation between the Michaelis constants K_C and K_O. Here, the correlation was reduced by 2.4-fold (variance explained = 8.7 %) relative to methods which do not correctly account for the non-

independence of these measurements (variance explained = 20.9 %; Figure 3C). An analogous correlation was also observed when analyses were limited to C_3 species (variance explained = 11.6 %, Figure 3C). Thus, there has only been a marginal trade-off between K_C and K_O during the evolution of RuBisCO in angiosperms.

Phylogenetic correction also reduced the magnitude of the correlation between $S_{C/O}$ and each of the other carboxylase-related traits (Figure 3C). Here the dependency between $S_{C/O}$ and k_{catC} was reduced from 9.8 to 8.1 %, whilst that between $S_{C/O}$ and both K_C and K_C^{air} were reduced from 9.9 to 6.2 %, and from 9.1 to 5.3 %, respectively (Figure 3C). Furthermore, these correlations were not significant when considering only C_3 species (Figure 3C). Thus, there has been either a marginal or no trade-off between RuBisCO specificity and other carboxylase-related kinetic traits during the evolution of RuBisCO in angiosperms.

In contrast to the weak trade-offs above, the trade-off between k_{catC} and either K_C or K_C^{air} was robust to phylogenetic correction. Here, the strength of the correlation between these kinetic traits experienced a small proportional reduction when correctly accounting for phylogenetic covariation. Specifically, the correlation between k_{catC} and K_C decreased from 34.2 to 30.9 %, and the correlation between k_{catC} and K_C^{air} decreased from 34.5 to 31.6 % (Figure 3C). The phylogenetically-corrected correlations between these kinetic traits were of a similar magnitude when only C_3 species were considered (37.4 and 34.9 % respectively (Figure 3C)). Consequently, these trade-offs are at least 3-fold stronger than any other trade-off between other RuBisCO kinetic traits. Thus, the trade-off between CO_2 affinity and carboxylase turnover has been the largest catalytic constraint on the adaptation of RuBisCO during the evolution of angiosperms.

The evolution of RuBisCO kinetics is more limited by phylogenetic constraints than by catalytic trade-offs

As each of the RuBisCO kinetic traits contained significant phylogenetic signal (Table 1 and 2), the manner in which RuBisCO has adapted in angiosperms has been phylogenetically constrained. In order to assess the relative strength these phylogenetic constraints, the variance in kinetic traits partitioned between phylogenetic effects (i.e., the explanatory power of phylogenetic co-variation in the goodness-of-fit model and a measure of phylogenetic constraint) and non-phylogenetic effects

(i.e., the remaining explanatory power of the regression model, accounted for by the sum of all other constraints such as random effects and all kinetic trait trade-offs including those not assessed here) was quantified. This revealed that phylogenetic constraints explained a significant proportion of variation in all carboxylase related kinetic traits in all species (Figure 4A). Moreover, consistent with analysis of phylogenetic signal in C_3 species above, phylogenetic constraints also explained a significant proportion of variation in the oxygenase-related kinetic trait in C_3 species (Figure 4B). Furthermore, phylogenetic constraints explained a larger proportion of variation in kinetic traits than any catalytic trade-off (Figure 4C and 4D). Consequently, phylogenetic constraints have been a larger limitation on RuBisCO adaptation than constraints attributable to kinetic traits (Cumulative variance for phylogenetic constraints = 24.8%, cumulative variance for catalytic trade-offs = 9%), with this effect being more pronounced if limited to the C_3 species in the dataset (Cumulative variance for phylogenetic constraints = 43.4%, cumulative variance for catalytic trade-offs = 8.2%). Thus, phylogenetic constraints have limited RuBisCO kinetic evolution to a greater extent than catalytic trade-offs.

Discussion

The evolutionary landscape of RuBisCO has long been proposed to be constrained by catalytic trade-offs. In support of this hypothesis, antagonistic correlations between RuBisCO kinetic traits inferred from studies comparing limited numbers of species are commonly cited (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Specifically, strong dependencies are thought to occur between RuBisCO specificity ($S_{C/O}$), carboxylase turnover (k_{catC}) and the Michaelis constants for CO_2 (K_C) and O_2 (K_O), respectively (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Combined, these trade-offs are hypothesized to limit the capacity of RuBisCO to assimilate CO_2 at high rates by curtailing the inherent carboxylase activity of the enzyme, whilst also causing it to catalyse a reaction with O_2 which is energetically expensive and results in a loss of fixed carbon (Bowes, Ogren and Hageman, 1971; Chollet, 1977). However, all trade-offs have been inferred based on the assumption that RuBisCO in different species are independent. Here, we show that this assumption was incorrect and that significant phylogenetic constraint, and thus significant phylogenetic signal (also known as phylogenetic covariation), is found in all carboxylase-related

kinetic traits in all species. In addition, significant phylogenetic constraint is also found in the oxygenase-related kinetic trait in C_3 species. When accounting for the phylogenetic signal caused by this constraint, we demonstrate that previously reported kinetic trait correlations have been over-estimated. Specifically, there is weak trade-off between carboxylase turnover and the Michaelis constant for CO_2 (~30 % dependency), and either non-significant or marginal trade-offs between all other kinetic traits (<9 % dependency). Finally, we demonstrate that phylogenetic constraints have played a larger role than catalytic trade-offs in limiting the evolution of RuBisCO kinetics in angiosperms.

The presence of phylogenetic constraint, and thus phylogenetic signal, in RuBisCO kinetics traits simply means that the kinetic traits of RuBisCO are more similar among close relatives, with this similarity changing as a function of the phylogenetic tree which relates the species. This could be considered to be unsurprising given the fact that all extant enzymes are related by the process of descent with modification from a single ancestral enzyme (Nisbet *et al.*, 2007). However, such a tree-like pattern of evolution is not the rule for biological traits, and several mechanistic explanations have been put forward to account for the presence of such phylogenetic signal (Kamilar & Cooper, 2013; and references therein). In brief, it is often interpreted as suggestive of either neutral evolution under genetic drift (Felsenstein, 1985), or evolutionary stasis (Prinzing *et al.*, 2001; Ackerly, 2004; Cavender-Bares *et al.*, 2004; Moles *et al.*, 2005; Swenson and Enquist, 2007) under which adaptive change is mitigated by processes including stabilizing selection, pleiotropy and a lack of molecular variability or phenotypic plasticity (Maynard Smith *et al.*, 1985; Bradshaw, 1991; Edwards and Naeem, 1993; Wagner, 1995). Although it is possible that several of the above factors contribute to the phylogenetic constraint that gives rise to the phylogenetic signal, it is likely that low molecular variability is a key driver of this phenomenon in RuBisCO. This is because the catalytic subunit of RuBisCO is encoded by a single copy gene that is trapped in the haploid plastid genome and consequently is uniparentally inherited and does not recombine (Birky, 2001). In addition, the low rate of molecular evolution in RuBisCO is further entrenched by evolutionary constraints on the enzyme, including those imposed by the requirements for i) high levels of transcript and protein abundance (Kelly, 2018; Seward and Kelly, 2018), ii) maintaining complementarity to a wide array

of molecular chaperones which assist in protein folding and assembly (e.g. Raf1, Raf2, RbcX, BSD2, Cpn60/Cpn20) and metabolic regulation (e.g. RuBisCO activase) (Carmo-Silva *et al.*, 2015; Aigner *et al.*, 2017), and iii) the need to preserve overall protein stability within the molecular activity-stability trade-offs (Studer *et al.*, 2014; Duraõ *et al.*, 2015; Cummins, Kannappan and Gready, 2018). Combined, these features create a perfect storm of limitations which constrain the molecular evolution of RuBisCO resulting in significant phylogenetic constraint that is independent of any catalytic requirements.

It should be noted that RuBisCO in angiosperms is a holoenzyme composed structurally of both large and small subunits (Schneider, Lindqvist and Brändén, 1992; Tabita *et al.*, 2008). Thus, although the phylogenetic constraints identified in this study are related to the function of the complete holoenzyme, they apply predominantly to the *rbcL* gene encoding the large subunit. This is because the *rbcS* gene is nuclear encoded, multi-copy (in most plant species), undergoes recombination, and is less constrained by molecular chaperones. Moreover, as RbcS has an indirect influence on RuBisCO catalytic properties (Read and Tabita, 1992b, 1992a; Spreitzer, Peddi and Satagopan, 2005; Martin-Avila *et al.*, 2020) and activity (Andrews, 1988; Lee and Tabita, 1990; Lee, Berka and Tabita, 1991), it is possible that this subunit may have helped the RuBisCO holoenzyme to evolve more rapidly and escape some phylogenetic constraint. However, the exact contribution of the RbcS to variation in holoenzyme kinetics, or indeed whether the influence of this protein has already been optimised, remain still to be fully understood.

When accounting for phylogenetic covariation in RuBisCO kinetic traits, the association between the Michaelis constants for CO_2 and O_2 substrates across all angiosperms was reduced to below 9 %. Thus, changes in CO_2 affinity have occurred with only weak antagonism of O_2 affinity. Moreover, the phylogenetically corrected dependency between specificity and both carboxylase turnover and the Michaelis constant for CO_2 were similarly relaxed to less than a 9 % dependency. However, no antagonism of any kinetic trait on specificity was found for RuBisCO in C_3 plants. Thus, natural selection has altered RuBisCO specificity, and thus the overall carboxylation/oxygenation ratio of RuBisCO, either largely or completely independent of changes in CO_2 affinity or carboxylase turnover, and *vice versa*. Therefore, catalytic trade-offs between CO_2 and O_2 affinity, and between

specificity and other carboxylase-related kinetic traits, exert only very limited constraint on the evolution of RuBisCO kinetics in angiosperms.

In contrast, whilst the biochemical landscape of RuBisCO is only weakly constrained by the majority of catalytic trade-offs, a phylogenetically corrected ~30 % dependency was observed between k_{catC} and both K_C and K_C^{air} in angiosperms. Thus, there has been a more pronounced trade-off between CO_2 affinity and carboxylase turnover during the evolution of RuBisCO. This result is compatible with the mechanistic models of RuBisCO (Farquhar, 1979), and is supported by the recent discovery of enzyme variants which exhibit the highest k_{catC} ever recorded at the expense of poor CO_2 affinities (i.e. K_C values >250 µM) (Davidi *et al.*, 2020). Nevertheless, the dependency between CO_2 affinity and turnover, despite being the strongest pairwise correlation between kinetic traits observed, is still only partial (~30 % variance explained), and is substantially attenuated relative to the coefficients conventionally cited from studies investigating limited numbers of species (Tcherkez, Farquhar and Andrews, 2006; Savir *et al.*, 2010). Thus, although selecting for a greater RuBisCO carboxylase turnover is evolutionarily linked with a poorer affinity for CO_2 (higher K_C), and *vice versa*, significant plasticity exists in this relationship among angiosperms such that variation in one kinetic trait only explains 30 % of variation in the other. This explains why there is variability in the carboxylation efficiency among angiosperm RuBisCO (defined as k_{catC}/K_C), a core parameter which defines the initial slope of the response of CO_2 fixation rate to changes in CO_2 concentration within the aerobic environment of chloroplasts in C_3 species (Sharwood, 2017).

The phylogenetically resolved analysis of the evolution of RuBisCO kinetic traits also identified changes in kinetics traits associated with the evolution of C_4 photosynthesis. Specifically, $S_{C/O}$ was lower in C_4 species than in C_3 species while k_{catC}, K_C and K_C^{air} were higher in C_4 species than in C_3 species. Moreover, variation in K_O was found to be greater in C_4 species than in C_3 species. All of these changes are consistent with the lack of catalytic trade-off, and are neutral or adaptive in a C_4 context. For example, any change in K_O would effectively be neutral in the CO_2 saturated environment of the bundle sheath chloroplast as it would have only a marginal effect on in vivo carboxylation rate or carboxylation-to-oxygenation ratio, and thus would not cause a concomitant change to organism fitness. In contrast, an increase in k_{catC} in the same CO_2 saturated environment

would enable higher flux through RuBisCO and thus provide an energetic advantage to the cell. Accordingly, one would expect that elevated K_O in C_4 species would occur by neutral drift (Kimura, 1991), and that an increase K_C would occur given the weak antagonistic relationship between k_{catC} and K_C. Thus, the adaptations to RuBisCO kinetics that occur concomitant with the evolution of C_4 photosynthesis are consistent with the change in CO_2:O_2 ratio and weak catalytic trade-off that exists between k_{catC} and K_C. Moreover, the rate at which these changes occurred may have been facilitated by the increase rate of molecular evolution (Kelly, 2018) and diversification (Spriggs, Christinb and Edwards, 2014) that occurs concomitant with the evolution of C_4 photosynthesis.

Whilst every effort was taken to prevent systematic or methodological biases from influencing the results presented in this work, several factors outside of the control of the study may have led to the underestimation of phylogenetic dependency in traits. Specifically, experimental error in kinetic trait measurements may have constrained the detection of phylogenetic covariance, as has been shown in other studies (Rohlf, 2001). However, additional sources of measurement inaccuracy in the data analysed here may include i) biases introduced by a failure to account for the breadth of intraspecific variation among individuals, populations and ecotypes (Ives, Midford and Garland, 2007), ii) the fact that *in vitro* estimates of kinetics poorly correlate with the enzyme behaviour *in vivo* (e.g., Bar-Even *et al.*, 2011), or finally iii) inconsistencies associated with measurements being compiled from numerous sources in the literature (Flamholz et al., 2019). However, toward partially addressing the latter inaccuracy, the $S_{C/O}$ values used in our analyses have been normalised to avoid the known discrepancies between the rates measured using an oxygen electrode assay (Parry, Keys and Gutteridge, 1989) and those measured using high precision gas phase controlled ^{3}H-RuBP-fixation assays (Kane *et al.*, 1994) (see methods). Nevertheless, even though kinetic trade-offs measured here are reduced when compared to previous reports, it is possible that correlations have still been over-estimated due to the fact that the phylogenetic covariance inferred in this work represents a lower limit of the true signal due to these potential errors.

Given the importance of RuBisCO to life on Earth, the question as to why a “perfect” RuBisCO has not already materialized during evolution is legitimate. For example, although RuBisCO K_C is thought to be near optimal in C_3 plants (e.g., Figure 2) in light of the ~8 µM chloroplastic concentration of

CO_2 and the inherent limitations of CO_2 as a substrate, including its inertness, hydrophobicity and low molecular mass (Andrews and Whitney, 2003; Bar-Even *et al.*, 2011; Bathellier *et al.*, 2018), the observed k_{catC} (~3 s^{-1}) is generally low compared to the turnover rates of other enzymes in primary metabolism (Bar-Even *et al.*, 2011; Tcherkez, 2013; Davidi *et al.*, 2018). In addition, all known RuBisCO variants catalyse a promiscuous and energetically costly reaction with O_2. However, the revised kinetic trait trade-offs described in this study suggest biochemical constraints on the optimisation of carboxylase and oxygenase kinetic traits are either partial or weak, and thus are not predominantly responsible for slow adaptation of RuBisCO in angiosperms. Instead, the findings presented here unearth a novel and important role of phylogenetic constraints, which are found to restrict the adaptive evolution of RuBisCO kinetics to a larger extent than catalytic constraints. These phylogenetic constraints are most likely the consequence of the inherent low level of molecular evolution of the catalytic large subunit of RuBisCO, as a consequence of both the genetics and heredity of chloroplasts and non-catalytic constraints on the enzyme, as discussed above. By inducing a considerable lag in adaptive evolution, this phylogenetic constraint slows the adaptation of RuBisCO and helps explain why the enzyme is better suited to former environmental conditions.

The phylogenetically resolved analysis of angiosperm RuBisCO kinetics presented here reveals that the biochemical landscape of RuBisCO is weakly constrained by catalytic trade-offs in angiosperms. Instead, phylogenetic constraints have provided a more substantial limitation to the evolution of RuBisCO kinetics. Accordingly, it should be feasible in the current synthetic biology revolution to circumvent these evolutionary barriers to RuBisCO adaptation. Indeed, promising steps toward this goal have been already demonstrated using directed evolution of the enzyme to generate variants with improved catalytic traits in non-photosynthetic archaea (Wilson, Alonso and Whitney, 2016) and plant-like Form I hexadecameric RuBisCO from photosynthetic bacteria (Zhou and Whitney, 2019) and cyanobacteria (Wilson *et al.*, 2018). Thus, these findings presented here provide optimism for engineering RuBisCO in food, fibre and fuel crops to have improved catalytic efficiency.

Materials and Methods

Kinetic data

Kinetic measurements of RuBisCO (Form IB) from a large number of plant species were attained from (Flamholz *et al.*, 2019). $S_{C/O}$ values measured using the O_2 electrode method which calculate

[CO_2] using a pKa of 6.11 (Parry, Keys and Gutteridge, 1989) were normalized relative to $S_{C/O}$ values quantified using high precision gas phase controlled ^{3}H-RuBP-fixation assays (Kane *et al.*, 1994). Here, as the $S_{C/O}$ of wheat (*Triticum aestivum*) RuBisCO was measured using the method of (Kane *et al.*, 1994) at 88 mol.mol^{-1}, and *T. aestivum* is also used as an enzyme standard in the O_2 electrode $S_{C/O}$ studies (Orr *et al.*, 2016; Prins *et al.*, 2016), multipliers of either 0.88 (Prins *et al.*, 2016) or 0.85 (Orr *et al.*, 2016) were applied to normalise all $S_{C/O}$ measurements to 88 mol.mol^{-1} for *T. aestivum* RuBisCO and minimise methodological biases in the data.

All kinetic traits in the dataset were transformed on a log scale consistent with (Flamholz *et al.*, 2019), and the distributional assumptions of each were verified for analyses herein. Only angiosperm species with experimental measurements of all four principal experimentally measured kinetic traits of interest ($S_{C/O}$, k_{catC}, K_C and K_O) were taken forward for subsequent analysis. A small number of these species also had measurements available for K_{RuBP}. In addition, the measure of the Michaelis constant for CO_2 under 20.95 % ambient air, K_C^{air}, was inferred from K_C and K_O based on the formula $K_C + (K_C\ [O_2] / K_O)$, where 20.95 % [O_2] in water is 253µM. In cases where duplicate entries for a species were present in the kinetic dataset (including synonyms), the median value of their respective kinetic traits was used for inference. In this way, averages were also taken between *Triticum timonovum* and *Triticum timopheevii*, the former of which is a synthetic octoploid of the latter (Murashov and Morozova, 2008). The modified dataset containing corrected $S_{C/O}$ values and averages across duplicate entries for species is provided in Supplemental File 2.

Phylogenetic tree inference

As sequenced genomes or transcriptomes do not exist for many plants in the kinetic trait dataset, whole genome phylogenomic approaches could not be used to infer the species tree necessary in order to detect phylogenetic signal in measured kinetic traits of the RuBisCO holoenzyme. However, the *rbcL* gene that encodes the large subunit of RuBisCO has a long history of use in plants for phylogenetic inference of species relationships (Gielly and Taberlet, 1994; APG, 1998, 2016) and was available for all of the angiosperms that were used in the analysis. As such, *rbcL* was selected here for use in species tree inference. The complete coding sequences of *rbcL* for the 137 angiosperm species for which kinetic data was available can be found in Supplementary File 3. These sequences were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/) for all species except

Flaveria brownii which was acquired from the 1KP database (Leebens-Mack *et al.*, 2019). A multiple sequence alignment was performed using mafft L-INS-i (Katoh and Standley, 2013), and alignments were trimmed at the terminal ends to remove unaligned positions using AliView (Larsson, 2014). This trimmed nucleotide sequence alignment was used for subsequent phylogenetic analysis. A bootstrapped maximum likelihood phylogenetic tree was inferred by IQ-TREE (Nguyen *et al.*, 2015) using the ultrafast bootstrapping method with 1000 replicates and the Shimodaira-Hasegawa approximate likelihood ratio branch test. The resultant maximum likelihood phylogenetic tree was rooted manually using Dendroscope (Huson and Scornavacca, 2012). A number of nodes (n = 18) in the complete tree (Supplementary File 1, Figure S1) exhibited terminal zero-length branches due to 100 % sequence identify with other closely related species. These species were condensed into single data points (as their *rbcL* are 100 % identical) and the mean of their kinetic traits was used. This reduced the dataset to 119 unique enzymes. The phylogenetic tree inferred from these taxa (Supplementary File 1, Figure S1) closely matched the topology of the phylogenetic tree expected from the angiosperm phylogeny with only a few alterations (APG, 2016). Moreover, the topology of the *rbcL* tree most accurately reflects the sequence similarity of the RuBisCO genes and thus it was deemed a suitable tree for investigation of phylogenetic signal and its effects on correlations between RuBisCO kinetic traits.

To confirm that the phylogenetic signal was not attributable to overfitting caused by the use of the *rbcL* gene to infer the phylogeny of RuBisCO, an analogous maximum likelihood phylogenetic tree was also inferred from a multiple sequence alignment in which codons containing non-synonymous nucleotide sequence changes were removed (Supplementary File 4). Due to the considerable loss of phylogenetic information accessible for tree building from this alignment, the species tree inferred using nucleotide sequences corresponding to ubiquitously conserved amino acid positions (Supplementary File 1, Figure S2) exhibited an additional number of species belonging to terminal zero-length branches (n = 13). As the sequences of these species are known to exhibit non-synonymous mutations which are not included in the tree, it is not appropriate to take means of their kinetic traits as above. As such, these data points were removed from analysis using only this tree. Use of this phylogenetic tree confirmed that the presence of phylogenetic signal in kinetic traits was

not due to overfitting, however as this tree was less accurate than the full-length alignment tree, it was not used for any subsequent analysis.

Phylogenetic signal analysis

The presence of phylogenetic signal in kinetic traits was assessed using five different phylogenetic signal detection methods (Gittleman and Kot, 1990; Abouheif, 1999; Pagel, 1999; Blomberg, Garland and Ives, 2003). Here, signal strength was estimated by assessing the distribution of trait values relative to the underlying species tree inferred from *rbcL* sequences in (Supplementary File 1, Figure S1) using methods which both depend on an explicit evolutionary model, such as Pagel's lambda (Pagel, 1999) and Blomberg's K and K* (Blomberg, Garland and Ives, 2003), as well as the spatial autocorrelation analyses of Moran's I (Gittleman and Kot, 1990) and Abouheif's Cmean (Abouheif, 1999). Implementation of these phylogenetic signal detection tools was performed using the phyloSignal function of the *phylosignal* package (Keck *et al.*, 2016) in the R environment.

Ancestral state estimation and mapping of kinetic traits to the phylogenetic tree

Ancestral state estimation was conducted to visualise the evolution of RuBisCO kinetic traits on the phylogenetic tree which relates the species. For this purpose, the kinetic dataset was mapped and scaled onto the species tree by employing the function *ggtree* in the ggtree package (Yu *et al.*, 2017). Here, terminal branches were coloured according to the measurement of the kinetic trait in the species which comprise the terminal branch, whereas internal branches were coloured based on values inferred in ancestral species using ancestral state estimation (Yu *et al.*, 2017).

Least squares and linear regression models

All regression models between pairwise combinations of kinetic traits were fit in the R environment. Phylogenetic generalized regression accounting for the phylogenetic non-independence between species was performed using the function pgls in the *caper* package (Comparative Analyses of Phylogenetics and Evolution in R) (Orme *et al.*, 2014). In each case, the structure of the phylogenetic signal was corrected for using branch length transformations of the phylogenetic tree based on the mean maximum likelihood estimates of lambda calculated for the traits being interrogated with kapa and delta held constant. Phylogenetic corrections to differences in kinetic trait values between C_3 and C_4 plants based on the phylogenetic non-independence of species were also applied using the pgls function (Orme *et al.*, 2014) with photosynthetic type incorporated as a factorial variable.

In order to partition the variance in RuBisCO kinetic traits explained by phylogenetic constraints as compared to non-phylogenetic constraints, the *rr2* package (Ives and Li, 2018) was employed in R. To assess the contribution of phylogeny in explaining the co-variation in kinetic trait values, the explanatory power of the phylogenetic component was measured by comparing full and reduced phylogenetic regression models using the partial R^2_{pred} inferential statistical, based on advice from (Ives, 2019) of this being the most intuitive and direct approach to understand how different model components explain variation. For this analysis, phylogenetic regression models were fit using the phylolm function in the *phylolm* package (Tung Ho and Ané, 2014) using Pagel's lambda model for the error term.

Acknowledgements

This work was funded by the Royal Society and the European Union's Horizon 2020 research and innovation program under grant agreement number 637765. JWB, JRN, JSB, AE, AB and AU were funded by the BBSRC through BB/J014427/1. TR and SMW were supported by the Australian Government through the Australian Research Council Centre of Excellence for Translational Photosynthesis CE140100015

Author Contributions

SK conceived the study. JWB conducted the analysis. JWB, TR, SMW, DME, and SK analyzed the data. SK and JWB wrote the manuscript. JRN, JSB, AE, AB and AU contributed to discussions and methodological development. All authors edited the manuscript.

References

Abouheif, E. (1999) *A method for testing the assumption of phylogenetic independence in comparative data*, *Evolutionary Ecology Research*. doi: 10.1007/s11295-009-0238-5.

Ackerly, D. D. (2004) 'Adaptation, niche conservatism, and convergence: Comparative studies of leaf evolution in the California chaparral', *American Naturalist*, 163(5), pp. 654–671. doi: 10.1086/383062.

Aigner, H. *et al.* (2017) 'Plant RuBisCo assembly in E. coli with five chloroplast chaperones including BSD2', *Science*. doi: 10.1126/science.aap9221.

Andersson, I. and Backlund, A. (2008) 'Structure and function of Rubisco', *Plant Physiology and Biochemistry*, 46(3), pp. 275–291. doi: 10.1016/j.plaphy.2008.01.001.

Andrews, T. J. (1988) 'Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits.', *The Journal of biological chemistry*.

Andrews, T. J. and Whitney, S. M. (2003) 'Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants', *Archives of Biochemistry and Biophysics*, 414(2), pp. 159–169. doi: 10.1016/S0003-9861(03)00100-0.

APG (1998) *An Ordinal Classification for the Families of Flowering Plants*, *Annals of the Missouri Botanical Garden*. doi: 10.2307/2992015.

APG (2016) 'An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV', *Botanical Journal of the Linnean Society*. Oxford University Press, 181(1), pp. 1–20. doi: 10.1111/boj.12385.

Araújo, W. L. *et al.* (2017) 'Engineering photosynthesis: Progress and perspectives',

F1000Research. doi: 10.12688/f1000research.12181.1.

Badger, M. R. and Lorimer, G. H. (1976) 'Activation of ribulose-1,5-bisphosphate oxygenase. The role of Mg2+, CO2, and pH', *Archives of Biochemistry and Biophysics*, pp. 723–729. doi: 10.1016/0003-9861(76)90565-8.

Bar-Even, A. *et al.* (2011) 'The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters', *Biochemistry*, 50(21), pp. 4402–4410. doi: 10.1021/bi2002289.

Bar-On, Y. M. and Milo, R. (2019) 'The global mass and average rate of rubisco', *Proceedings of the National Academy of Sciences of the United States of America*. PNAS, 116(10), pp. 4738–4743. doi: 10.1073/pnas.1816654116.

Bathellier, C. *et al.* (2018) 'Rubisco is not really so bad', *Plant Cell and Environment*, pp. 705–716. doi: 10.1111/pce.13149.

Bauwe, H. *et al.* (2012) 'Photorespiration has a dual origin and manifold links to central metabolism', *Current Opinion in Plant Biology*, pp. 269–275. doi: 10.1016/j.pbi.2012.01.008.

Beer, C. *et al.* (2010) 'Terrestrial gross carbon dioxide uptake: Global distribution and covariation with climate', *Science*, 329(5993), pp. 834–838. doi: 10.1126/science.1184984.

Birky, C. W. (2001) 'The Inheritance of Genes in Mitochondria and Chloroplasts: Laws, Mechanisms, and Models', *Annual Review of Genetics*. doi: 10.1146/annurev.genet.35.102401.090231.

Blomberg, S. P., Garland, T. and Ives, A. R. (2003) 'Testing for phylogenetic signal in comparative data: Behavioral traits are more labile', *Evolution*. Oxford, UK: Blackwell Publishing Ltd, pp. 717–745. doi: 10.1111/j.0014-3820.2003.tb00285.x.

Bowes, G., Ogren, W. L. and Hageman, R. H. (1971) *Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase*, *Biochemical and Biophysical Research Communications*. doi: 10.1016/0006-291X(71)90475-X.

Bradshaw, A. (1991) 'Genostasis and the limits to evolution', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 333(1267), pp. 289–305.

Bräutigam, A. *et al.* (2017) 'On the evolutionary origin of CAM photosynthesis', *Plant Physiology*. American Society of Plant Biologists, pp. 473–477. doi: 10.1104/pp.17.00195.

Busch, F. A. (2020) 'Photorespiration in the context of Rubisco biochemistry, CO2 diffusion and metabolism', *Plant Journal*, 101(4), pp. 919–939. doi: 10.1111/tpj.14674.

Carmo-Silva, E. *et al.* (2015) 'Optimizing Rubisco and its regulation for greater resource use efficiency', *Plant, Cell and Environment*, pp. 1817–1832. doi: 10.1111/pce.12425.

Cavender-Bares, J. *et al.* (2004) 'Phylogenetic overdispersion in Floridian oak communities', *American Naturalist*, 163(6), pp. 823–843. doi: 10.1086/386375.

Cheverud, J. M., Dow, M. M. and Leutenegger, W. (1985) 'the Quantitative Assessment of Phylogenetic Constraints in Comparative Analyses: Sexual Dimorphism in Body Weight Among Primates', *Evolution*. Wiley, 39(6), pp. 1335–1351. doi: 10.1111/j.1558-5646.1985.tb05699.x.

Cheverud, J. M., Dow, M. M. and Leutenegger, W. (1986) 'A Phylogenetic Autocorrelation Analysis of Sexual Dimorphism in Primates', *American Anthropologist*. Wiley-Blackwell, 88(4), pp. 916–922. doi: 10.1525/aa.1986.88.4.02a00090.

Chollet, R. (1977) *The biochemistry of photorespiration*, *Trends in Biochemical Sciences*. doi: 10.1016/0968-0004(77)90364-4.

Chollet, R. and Anderson, L. L. (1976) 'Regulation of ribulose 1,5-bisphosphate carboxylase-oxygenase activities by temperature pretreatment and chloroplast metabolites', *Archives of Biochemistry and Biophysics*, pp. 344–351. doi: 10.1016/0003-9861(76)90173-9.

Cummins, P. L., Kannappan, B. and Gready, J. E. (2018) 'Directions for optimization of photosynthetic carbon fixation: Rubisco's efficiency may not be so constrained after all', *Frontiers in Plant Science*, 9. doi: 10.3389/fpls.2018.00183.

Davidi, D. *et al.* (2018) 'A Bird's-Eye View of Enzyme Evolution: Chemical, Physicochemical, and Physiological Considerations', *Chemical Reviews*, 118(18), pp. 8786–8797. doi: 10.1021/acs.chemrev.8b00039.

Davidi, D. *et al.* (2020) 'Highly active rubiscos discovered by systematic interrogation of natural sequence diversity', *The EMBO Journal*. John Wiley & Sons, Ltd. doi: 10.15252/embj.2019104081.

Dodd, A. N. *et al.* (2002) 'Crassulacean acid metabolism: Plastic, fantastic', *Journal of Experimental Botany*, pp. 569–580. doi: 10.1093/jexbot/53.369.569.

Duraõ, P. *et al.* (2015) 'Opposing effects of folding and assembly chaperones on evolvability of Rubisco', *Nature Chemical Biology*, 11(2), pp. 148–155. doi: 10.1038/nchembio.1715.

Eckardt, N. A. (2005) 'Photorespiration revisited', *Plant Cell*. American Society of Plant Biologists, pp. 2139–2141. doi: 10.1105/tpc.105.035873.

Edwards, E. J. (2019) 'Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis', *New Phytologist*, pp. 1742–1755. doi: 10.1111/nph.15851.

Edwards, S. V. and Naeem, S. (1993) 'The phylogenetic component of cooperative breeding in perching birds', *American Naturalist*. Am Nat, 141(5), pp. 754–789. doi: 10.1086/285504.

Ehleringer, J. R. *et al.* (1991) *Climate change and the evolution of C4 photosynthesis*, *Trends in Ecology and Evolution*. doi: 10.1016/0169-5347(91)90183-X.

Ellis, R. J. (1979) *The most abundant protein in the world*, *Trends in Biochemical Sciences*. doi: 10.1016/0968-0004(79)90212-3.

Erb, T. J. and Zarzycki, J. (2018) *A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme*, *Current Opinion in Biotechnology*. doi: 10.1016/j.copbio.2017.07.017.

Farquhar, G. D. (1979) 'Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase', *Archives of Biochemistry and Biophysics*, 193(2), pp. 456–468. doi: 10.1016/0003-9861(79)90052-3.

Felsenstein, J. (1985) *Phylogenies and the comparative method.*, *American Naturalist*. doi: 10.1086/284325.

Flamholz, A. I. *et al.* (2019) 'Revisiting Trade-offs between Rubisco Kinetic Parameters', *Biochemistry*, pp. 3365–3376. doi: 10.1021/acs.biochem.9b00237.

Fukayama, H. *et al.* (2019) 'Rubisco small subunits of C4 plants, Napier grass and guinea grass confer C4-like catalytic properties on Rubisco in rice', *Plant Production Science*. doi: 10.1080/1343943X.2018.1540279.

Garland, T. (2001) 'Phylogenetic comparison and artificial selection', in. Springer, Boston, MA, pp. 107–132. doi: 10.1007/978-1-4757-3401-0_9.

Gielly, L. and Taberlet, P. (1994) *The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences.*, *Molecular Biology and Evolution*. doi: 10.1093/oxfordjournals.molbev.a040157.

Giordano, M., Beardall, J. and Raven, J. (2005) 'CO2 Concentrating mechanisms in algae: Mechanisms, environmental modulation, and evolution', *Annual Review of Plant Biology*. Annual Reviews Inc, pp. 99–131. doi: 10.1146/annurev.arplants.56.032604.144052.

Gittleman, J. L. and Kot, M. (1990) 'Adaptation: Statistics and a Null Model for Estimating Phylogenetic Effects', *Systematic Zoology*. Society of Systematic Zoology, pp. 227–241. doi: 10.2307/2992183.

Grafen, A. (1989) 'The phylogenetic regression.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*. doi: 10.1098/rstb.1989.0106.

Hardy, O. J. and Pavoine, S. (2012) 'Assessing phylogenetic signal with measurement error: A comparison of mantel tests, blomberg et al.'s K, and phylogenetic distograms', *Evolution*. John Wiley & Sons, Ltd, 66(8), pp. 2614–2621. doi: 10.1111/j.1558-5646.2012.01623.x.

Heureux, A. M. C. *et al.* (2017) 'The role of Rubisco kinetics and pyrenoid morphology in shaping the CCM of haptophyte microalgae', *Journal of experimental botany*. doi: 10.1093/jxb/erx179.

Huson, D. H. and Scornavacca, C. (2012) 'Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks', *Systematic Biology*. Oxford University Press, pp. 1061–1067. doi: 10.1093/sysbio/sys062.

Iñiguez, C. *et al.* (2020) 'Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating mechanisms', *Plant Journal*, 101(4), pp. 897–918. doi: 10.1111/tpj.14643.

Ishikawa, C. *et al.* (2011) 'Functional incorporation of sorghum small subunit increases the catalytic turnover rate of rubisco in transgenic rice', *Plant Physiology*. doi: 10.1104/pp.111.177030.

Ives, A. and Li, D. (2018) ‘rr2: An R package to calculate R^2s for regression models’, *Journal of Open Source Software*, 3(30), p. 1028. doi: 10.21105/joss.01028.

Ives, A. R. (2019) ‘R 2 s for Correlated Data: Phylogenetic Models, LMMs, and GLMMs’, *Systematic Biology*, 68(2), pp. 234–251. doi: 10.1093/sysbio/syy060.

Ives, A. R., Midford, P. E. and Garland, T. (2007) ‘Within-species variation and measurement error in phylogenetic comparative methods’, *Systematic Biology*, 56(2), pp. 252–270. doi: 10.1080/10635150701313830.

Jordan, D. B. and Ogren, W. L. (1981) ‘Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase’, *Nature*. doi: 10.1038/291513a0.

Joshi, J. *et al.* (2015) ‘Role of small subunit in mediating assembly of red-type Form I Rubisco’, *Journal of Biological Chemistry*. doi: 10.1074/jbc.M114.613091.

Kamilar, J. M. and Cooper, N. (2013) ‘Phylogenetic signal in primate behaviour, ecology and life history’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1618). doi: 10.1098/rstb.2012.0341.

Kane, H. J. *et al.* (1994) ‘An improved method for measuring the CO2/O2 specificity of ribulosebisphosphatte carboxylase-oxygenase’, *Australian Journal of Plant Physiology*. doi: 10.1071/PP9940449.

Katoh, K. and Standley, D. M. (2013) ‘MAFFT multiple sequence alignment software version 7: Improvements in performance and usability’, *Molecular Biology and Evolution*. Oxford University Press, pp. 772–780. doi: 10.1093/molbev/mst010.

Keck, F. *et al.* (2016) ‘Phylosignal: An R package to measure, test, and explore the phylogenetic signal’, *Ecology and Evolution*, pp. 2774–2780. doi: 10.1002/ece3.2051.

Kelly, S. (2018) ‘The amount of nitrogen used for photosynthesis modulates molecular evolution in plants’, *Molecular Biology and Evolution*, 35(7), pp. 1616–1625. doi: 10.1093/molbev/msy043.

Kimura, M. (1991) ‘The neutral theory of molecular evolution: A review of recent evidence’, *The Japanese Journal of Genetics*. doi: 10.1266/jjg.66.367.

Larsson, A. (2014) ‘AliView: A fast and lightweight alignment viewer and editor for large datasets’, *Bioinformatics*. Oxford University Press, pp. 3276–3278. doi: 10.1093/bioinformatics/btu531.

Lee, B., Berka, R. M. and Tabita, F. R. (1991) ‘Mutations in the small subunit of cyanobacterial ribulose-bisphosphate carboxylase/oxygenase that modulate interactions with large subunits’, *Journal of Biological Chemistry*.

Lee, B. and Tabita, F. R. (1990) ‘Purification of Recombinant Ribulose-l,5-bisphosphate Carboxylase/Oxygenase Large Subunits Suitable for Reconstitution and Assembly of Active L8S8 Enzyme’, *Biochemistry*. doi: 10.1021/bi00492a007.

Leebens-Mack, J. H. *et al.* (2019) ‘One thousand plant transcriptomes and the phylogenomics of green plants’, *Nature*. doi: 10.1038/s41586-019-1693-2.

Lüttge, U. (2004) ‘Ecophysiology of Crassulacean Acid Metabolism (CAM)’, *Annals of Botany*. Oxford University Press, pp. 629–652. doi: 10.1093/aob/mch087.

Martin-Avila, E. *et al.* (2020) ‘Modifying Plant Photosynthesis and Growth via Simultaneous Chloroplast Transformation of Rubisco Large and Small Subunits.’, *The Plant cell*. doi: 10.1105/tpc.20.00288.

Maynard Smith, J. *et al.* (1985) ‘Developmental constraints on evolution’, *The Quarterly Review of Biology*.

Meyer, M. and Griffiths, H. (2013) ‘Origins and diversity of eukaryotic CO2-concentrating mechanisms: Lessons for the future’, *Journal of Experimental Botany*, pp. 769–786. doi: 10.1093/jxb/ers390.

Moles, A. T. *et al.* (2005) ‘A brief history of seed size’, *Science*, 307(5709), pp. 576–580. doi: 10.1126/science.1104863.

Münkemüller, T. *et al.* (2012) ‘How to measure and test phylogenetic signal’, *Methods in Ecology and Evolution*, 3(4), pp. 743–756. doi: 10.1111/j.2041-210X.2012.00196.x.

Murashov, V. V. and Morozova, Z. A. (2008) ‘Comparative morphogenesis of Triticum timopheevii (Zhuk.) and synthetic octoploid species T. timonovum Heslot et Ferrary’, *Moscow University*

Biological Sciences Bulletin. Heidelberg: Allerton Press, Inc., pp. 127–133. doi: 10.3103/s0096392508030073.

Nguyen, L. T. *et al.* (2015) 'IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies', *Molecular Biology and Evolution*, pp. 268–274. doi: 10.1093/molbev/msu300.

Nisbet, E. G. *et al.* (2007) 'The age of Rubisco: The evolution of oxygenic photosynthesis', *Geobiology*, 5(4), pp. 311–335. doi: 10.1111/j.1472-4669.2007.00127.x.

Ogren, W. L. (1984) 'Photorespiration: Pathways, Regulation, and Modification', *Annual Review of Plant Physiology*. Annual Reviews, 35(1), pp. 415–442. doi: 10.1146/annurev.pp.35.060184.002215.

Ogren, W. L. and Bowes, G. (1971) 'Ribulose diphosphate carboxylase regulates soybean photorespiration', *Nature New Biology*. Nat New Biol, 230(13), pp. 159–160. doi: 10.1038/newbio230159a0.

Orme, D. *et al.* (2014) *Caper: Comparative analyses of phylogenetics and evolution in R*, *R package version 0.5.2/ r121*. Available at: http://r-forge.r-project.org/projects/ caper/ (Accessed: 31 May 2020).

Orr, D. J. *et al.* (2016) 'Surveying rubisco diversity and temperature response to improve crop photosynthetic efficiency', *Plant Physiology*. doi: 10.1104/pp.16.00750.

Pagel, M. (1999) 'Inferring the historical patterns of biological evolution', *Nature*. Nature Publishing Group, pp. 877–884. doi: 10.1038/44766.

Pagel, M. D. and Harvey, P. H. (1989) 'Comparative methods for examining adaptation depend on evolutionary models.', *Folia primatologica; international journal of primatology*. doi: 10.1159/000156417.

Parry, M. A. *et al.* (2013) 'Rubisco activity and regulation as targets for crop improvement | Journal of Experimental Botany | Oxford Academic', *Journal of Experimental Bitany*.

Parry, M. A. J. *et al.* (2007) 'Prospects for increasing photosynthesis by overcoming the limitations of Rubisco', in *Journal of Agricultural Science*. doi: 10.1017/S0021859606006666.

Parry, M. A. J., Keys, A. J. and Gutteridge, S. (1989) 'Variation in the specificity factor of C3 higher plant rubiscos determined by the total consumption of ribulose-P2', *Journal of Experimental Botany*. doi: 10.1093/jxb/40.3.317.

Peterhansel, C. *et al.* (2010) 'Photorespiration', *The Arabidopsis Book*. BioOne, 8(8), p. e0130. doi: 10.1199/tab.0130.

Prins, A. *et al.* (2016) 'Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis', *Journal of Experimental Botany*. doi: 10.1093/jxb/erv574.

Prinzing, A. *et al.* (2001) 'The niche of higher plants: Evidence for phylogenetic conservatism', *Proceedings of the Royal Society B: Biological Sciences*, 268(1483), pp. 2383–2389. doi: 10.1098/rspb.2001.1801.

Raven, J. A. (1997) 'CO2-concentrating mechanisms: A direct role for thylakoid lumen acidification?', *Plant, Cell and Environment*. Blackwell Publishing Ltd, 20(2), pp. 147–154. doi: 10.1046/j.1365-3040.1997.d01-67.x.

Read, B. A. and Tabita, F. R. (1992a) 'A Hybrid Ribulosebisphosphate Carboxylase/Oxygenase Enzyme Exhibiting a Substantial Increase in Substrate Specificity Factor', *Biochemistry*. doi: 10.1021/bi00139a018.

Read, B. A. and Tabita, F. R. (1992b) 'Amino Acid Substitutions in the Small Subunit of Ribulose-1, 5-bisphosphate Carboxylase/Oxygenase That Influence Catalytic Activity of the Holoenzyme', *Biochemistry*. doi: 10.1021/bi00117a031.

Rohlf, F. J. (2001) *Comparative methods for the analysis of continuous variables: Geometric interpretations*, *Evolution*. doi: 10.1111/j.0014-3820.2001.tb00731.x.

Sage, R. F., Sage, T. L. and Kocacinar, F. (2012) ' Photorespiration and the Evolution of C 4 Photosynthesis ', *Annual Review of Plant Biology*, pp. 19–47. doi: 10.1146/annurev-arplant-042811-105511.

Savir, Y. *et al.* (2010) 'Cross-species analysis traces adaptation of Rubisco toward optimality in a

low-dimensional landscape', *Proceedings of the National Academy of Sciences of the United States of America*, 107(8), pp. 3475–3480. doi: 10.1073/pnas.0911663107.

Schneider, G., Lindqvist, Y. and Brändén, C. I. (1992) 'RUBISCO: Structure and mechanism', *Annual Review of Biophysics and Biomolecular Structure*. doi: 10.1146/annurev.bb.21.060192.001003.

Seward, E. A. and Kelly, S. (2018) 'Selection-driven cost-efficiency optimization of transcripts modulates gene evolutionary rate in bacteria', *Genome Biology*, 19(1). doi: 10.1186/s13059-018-1480-7.

Sharkey, T. D. (2020) 'Emerging research in plant photosynthesis', *Emerging Topics in Life Sciences*. Portland Press Ltd. doi: 10.1042/etls20200035.

Sharwood, R. E. (2017) 'Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops', *New Phytologist*, pp. 494–510. doi: 10.1111/nph.14351.

Sharwood, R. E., Ghannoum, O. and Whitney, S. M. (2016) 'Prospects for improving CO2 fixation in C3-crops through understanding C4-Rubisco biogenesis and catalytic diversity', *Current Opinion in Plant Biology*. doi: 10.1016/j.pbi.2016.04.002.

Spreitzer, R. J., Peddi, S. R. and Satagopan, S. (2005) 'Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco', *Proceedings of the National Academy of Sciences of the United States of America*. doi: 10.1073/pnas.0508042102.

Spriggs, E. L., Christinb, P. A. and Edwards, E. J. (2014) 'C4 photosynthesis promoted species diversification during the miocene grassland expansion', *PLoS ONE*. doi: 10.1371/journal.pone.0097722.

Still, C. J. *et al.* (2003) ' Global distribution of C 3 and C 4 vegetation: Carbon cycle implications ', *Global Biogeochemical Cycles*, pp. 6-1-6–14. doi: 10.1029/2001gb001807.

Studer, R. A. *et al.* (2014) 'Stability-activity tradeoffs constrain the adaptive evolution of RubisCO', *Proceedings of the National Academy of Sciences of the United States of America*, 111(6), pp. 2223–2228. doi: 10.1073/pnas.1310811111.

Su, Z. and Townsend, J. P. (2015) 'Utility of characters evolving at diverse rates of evolution to resolve quartet trees with unequal branch lengths: Analytical predictions of long-branch effects', *BMC Evolutionary Biology*, p. 86. doi: 10.1186/s12862-015-0364-7.

Swenson, N. G. and Enquist, B. J. (2007) 'Ecological and evolutionary determinants of a key plant functional trait: Wood density and its community-wide variation across latitude and elevation', *American Journal of Botany*, 94(3), pp. 451–459. doi: 10.3732/ajb.94.3.451.

Tabita, F. R. *et al.* (2008) 'Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships', *Journal of Experimental Botany*, 59(7), pp. 1515–1524. doi: 10.1093/jxb/erm361.

Tcherkez, G. (2013) 'Modelling the reaction mechanism of ribulose-1,5-bisphosphate carboxylase/oxygenase and consequences for kinetic parameters', *Plant, Cell and Environment*, 36(9), pp. 1586–1596. doi: 10.1111/pce.12066.

Tcherkez, G. G. B., Farquhar, G. D. and Andrews, T. J. (2006) 'Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized', *Proceedings of the National Academy of Sciences of the United States of America*, 103(19), pp. 7246–7251. doi: 10.1073/pnas.0600605103.

Tung Ho, L. S. and Ané, C. (2014) 'A linear-time algorithm for gaussian and non-gaussian trait evolution models', *Systematic Biology*, 63(3), pp. 397–408. doi: 10.1093/sysbio/syu005.

Wagner, P. J. (1995) 'Testing Evolutionary Constraint Hypotheses with Early Paleozoic Gastropods', *Paleobiology*. doi: 10.1017/S0094837300013294.

Whitney, S. M., Houtz, R. L. and Alonso, H. (2011) 'Advancing our understanding and capacity to engineer nature's CO2-sequestering enzyme, Rubisco', *Plant Physiology*, 155(1), pp. 27–35. doi: 10.1104/pp.110.164814.

Wilson, R. H. *et al.* (2018) 'An improved Escherichia coli screen for Rubisco identifies a protein-protein interface that can enhance CO2-fixation kinetics', *Journal of Biological Chemistry*. doi: 10.1074/jbc.M117.810861.

Wilson, R. H., Alonso, H. and Whitney, S. M. (2016) ‘Evolving Methanococcoides burtonii archaeal Rubisco for improved photosynthesis and plant growth’, *Scientific Reports*. doi: 10.1038/srep22284.

Young, J. N. *et al.* (2016) ‘Large variation in the Rubisco kinetics of diatoms reveals diversity among their carbon-concentrating mechanisms’, *Journal of Experimental Botany*. doi: 10.1093/jxb/erw163.

Yu, G. *et al.* (2017) ‘ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data’, *Methods in Ecology and Evolution*. doi: 10.1111/2041-210X.12628.

Zhou, Y. and Whitney, S. (2019) ‘Directed evolution of an improved Rubisco; in vitro analyses to decipher fact from fiction’, *International Journal of Molecular Sciences*, 20(20). doi: 10.3390/ijms20205019.

Figures

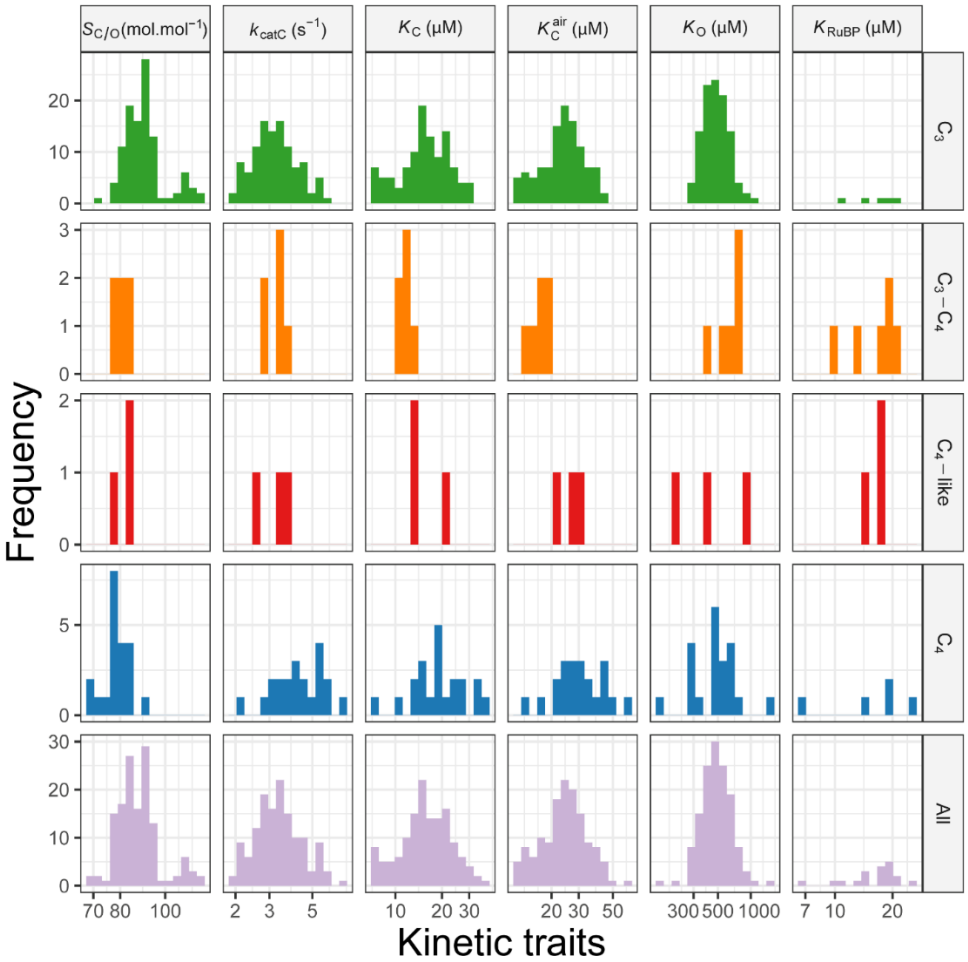

Figure 1. The distributions of values for RuBisCO kinetic traits in angiosperms. Species are grouped by their photosynthetic types (rows). $S_{C/O}$: specificity. k_{catC}: carboxylase turnover. K_C: the Michaelis constant for the CO_2. K_C^{air} the inferred Michaelis constant for CO_2 in 20.95 % O_2 air. K_O: the Michaelis constant for O_2. K_{RuBP}: the Michaelis constant for ribulose 1,5-bisphosphate. Plants have been classified as those which perform C_3 photosynthesis (C_3), C_4 photosynthesis (C_4), C_3-C_4 intermediate (C_3-C_4), C_4-like (C_4-like). The X axis for all plots is on a log scale, where respective units are shown in column labels.

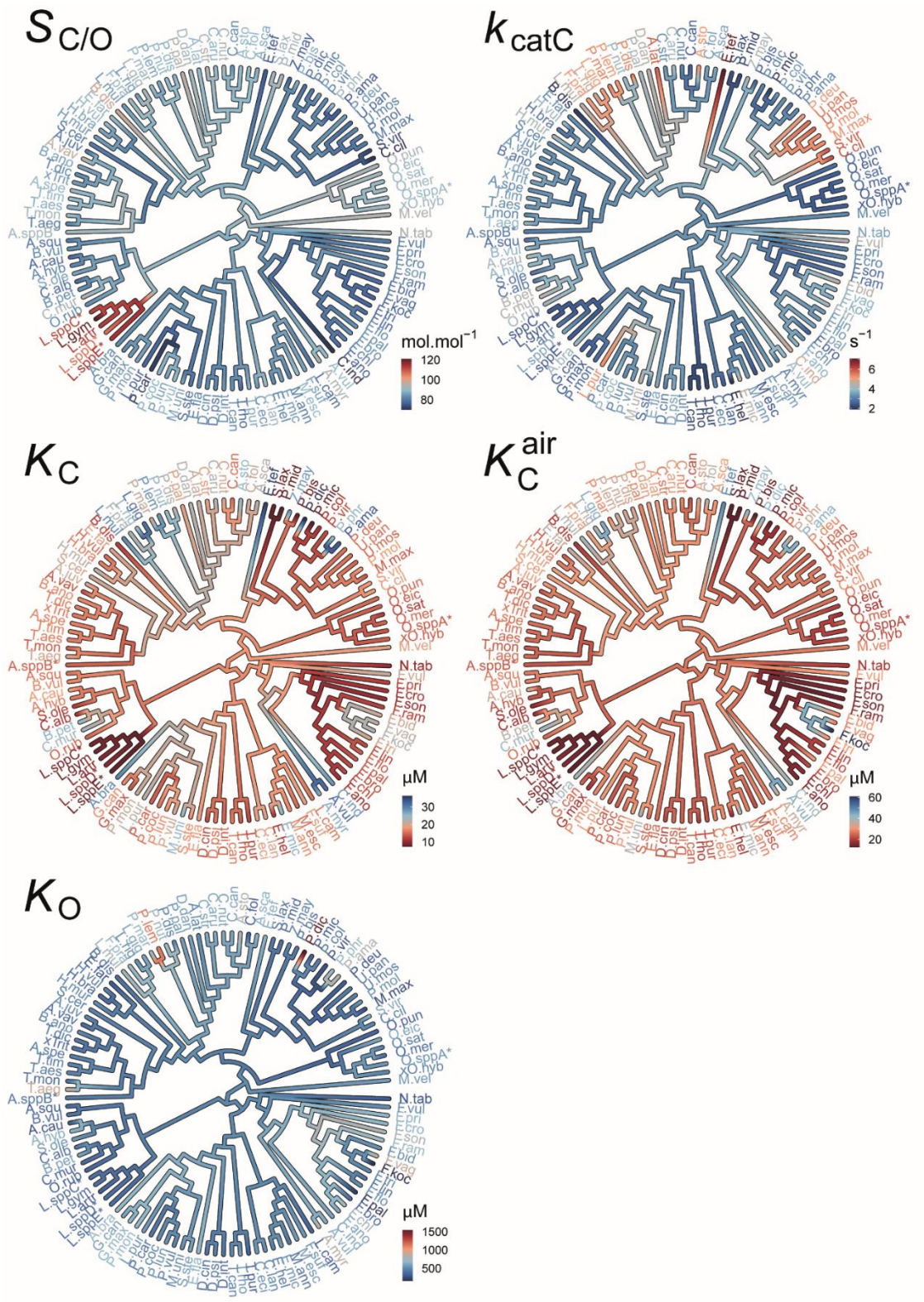

Figure 2. The evolution of RuBisCO kinetic traits in angiosperms. Phylogenetic tree of angiosperm species depicting the kinetic trait values in of the species used in this dataset and the maximum likelihood inferred ancestral kinetic traits for internal branches on the tree. Scale bars for colour schemes are presented next to each tree. Species names have been abbreviated for legibility and are provided in full in Supplementary File 1, Table S3.

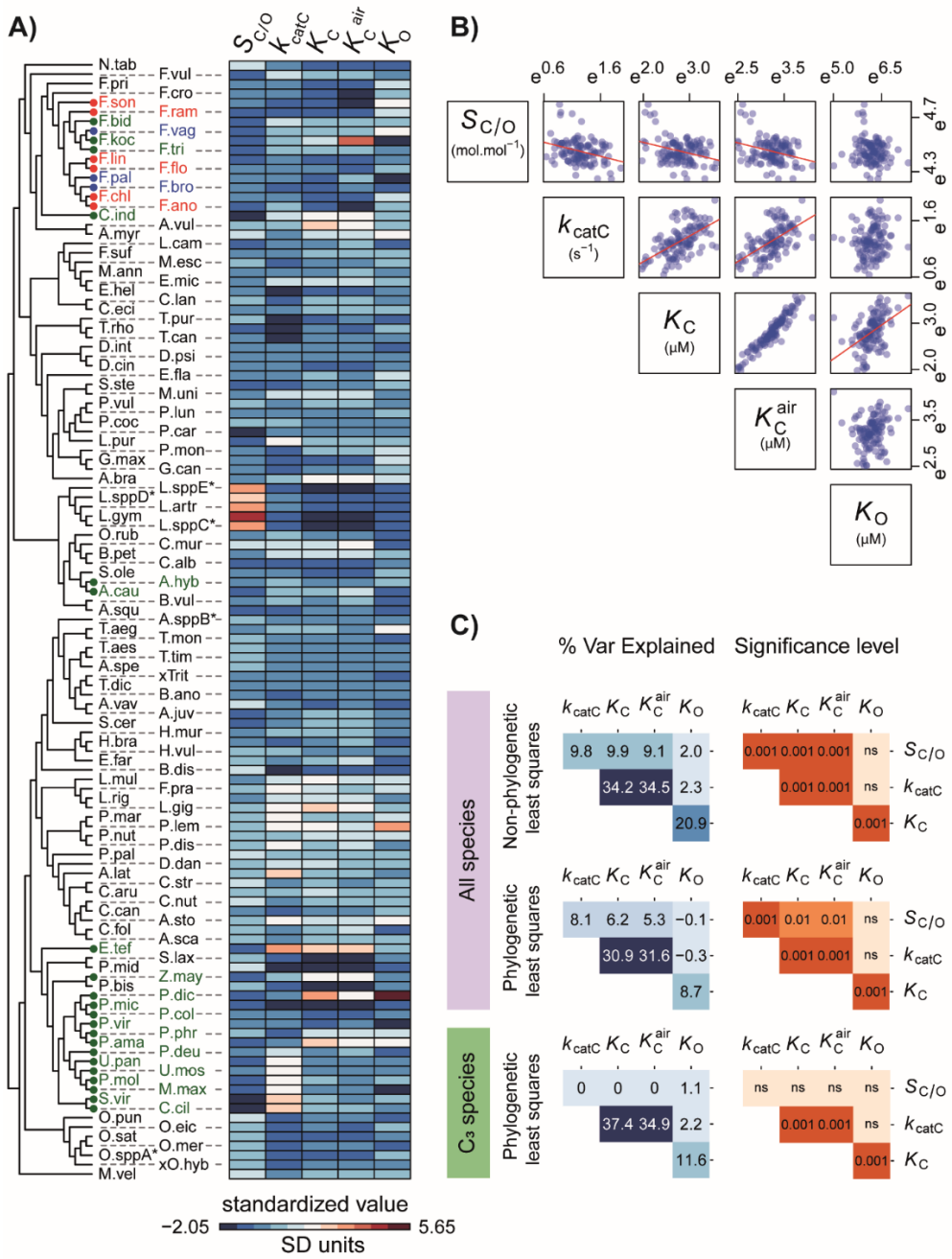

Figure 3. The correlations between RuBisCO kinetic traits in angiosperms. **A**) Heatmap depicting the normalised variation in kinetic traits across the species set used in this study (± S.D. away from each respective kinetic trait mean). Species labels on the tree are colour coded by photosynthetic type (C_3: black, C_3-C_4 intermediates: red, C_4-like: blue, and C_4: green), and have been abbreviated for legibility (refer to Supplemental File 1 Table S3). **B)** Trends in relationships between all pairwise combinations of log transformed RuBisCO kinetic traits shown in A. **C**) Pairwise correlation coefficients (percent variance explained) and associated *p*-values between different RuBisCO kinetic traits assessed using non-phylogenetic least squares regression models or phylogenetic least

squares regression models. Phylogenetic least squares regression was fit to both the complete set of angiosperms in the dataset and the subset which perform C_3 photosynthesis. Significance values are represented as α levels, where; $\alpha = 0.001$ if $p < 0.001$, $\alpha = 0.01$ if $0.001 < p < 0.01$, $\alpha = 0.05$ if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

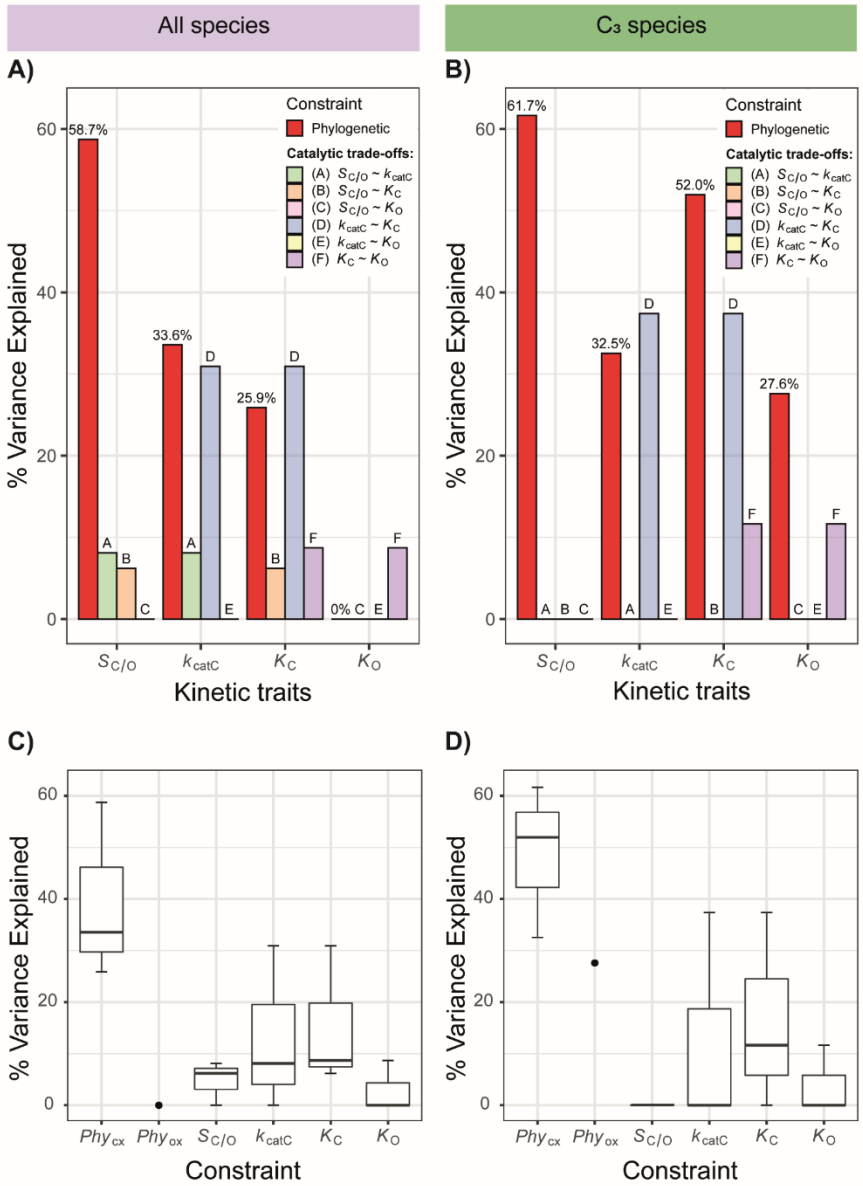

Figure 4. The evolutionary constraints on RuBisCO kinetic adaptation. **A)** The variation (%) in RuBisCO kinetic traits across all species which can be explained by phylogenetic constraint or individual catalytic trade-off. **B)** as in A but for just the C_3 species. **C)** Boxplot of all variation explained by each kinetic trait in comparison to variation explained by phylogeny across all species. The phylogenetic constraints on the carboxylase-related traits (Phy_{CX}) and phylogenetic constraints on the oxygenase-related trait (Phy_{ox}) are presented separately. **D)** As in C but for just the C_3 species.

Tables

Table 1. Phylogenetic signal strength and associated significance level in RuBisCO kinetic traits in angiosperms using five different signal detection methods. Statistics are rounded to 3 decimal places and significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

Kinetic	**C mean**		**I**		**K**		**K ***		**Lambda**	
Trait	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***
$S_{C/O}$	0.516	0.001	0.425	0.001	0.001	ns	0.001	ns	0.879	0.001
k_{catC}	0.350	0.001	0.224	0.001	0.003	0.001	0.003	0.001	0.968	0.001
K_C	0.282	0.001	0.248	0.001	0.001	0.01	0.002	0.01	0.902	0.001
K_C^{air}	0.234	0.001	0.242	0.001	0.001	0.05	0.001	0.05	0.392	ns
K_O	0.032	ns	-0.070	ns	0	ns	0	ns	0	ns

Table 2. Phylogenetic signal strength and associated significance level in RuBisCO kinetic traits in C_3 species using five different signal detection methods. Statistics are rounded to 3 decimal places, and significance values are represented as α levels, where; α = 0.001 if $p < 0.001$, α = 0.01 if $0.001 < p < 0.01$, α = 0.05 if $0.01 < p < 0.05$, and α = ns if $p > 0.05$.

Kinetic	**C mean**		**I**		**K**		**K ***		**Lambda**	
Trait	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***	***Stat***	***α***
$S_{C/O}$	0.533	0.001	0.453	0.001	0	ns	0.001	ns	0.814	0.001
k_{catC}	0.387	0.001	0.234	0.001	0.002	0.01	0.002	0.01	0.913	0.001
K_C	0.449	0.001	0.341	0.001	0.001	0.05	0.001	0.05	0.948	0.001
K_C^{air}	0.398	0.001	0.317	0.001	0.001	0.05	0.002	0.01	0.947	0.001
K_O	0.279	0.001	0.167	0.01	0	ns	0	ns	0.743	0.001