# Rubisco Adaptation Is More Limited by Phylogenetic Constraint Than by Catalytic Trade-off

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

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

[2]Doctoral Training Centre, University of Oxford, Oxford, United Kingdom

[3]Research School of Biology, Australian National University, Canberra, ACT, Australia

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

**Associate editor:** Julian Echave

## 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 highly constrained by catalytic trade-offs. However, these analyses did not consider the phylogenetic context of the enzymes that were analyzed. Thus, it is possible that the correlations observed were an artefact of the presence of phylogenetic signal in rubisco kinetics and the phylogenetic relationship between the species that were sampled. Here, we conducted a phylogenetically resolved analysis of rubisco kinetics and show that there is a significant phylogenetic signal in rubisco kinetic traits. We re-evaluated the extent of catalytic trade-offs accounting for this phylogenetic signal and found that all were attenuated. Following phylogenetic correction, the largest catalytic trade-offs were observed between the Michaelis constant for $CO_2$ and carboxylase turnover (∼21–37%), and between the Michaelis constants for $CO_2$ and $O_2$ (∼9–19%), respectively. All other catalytic trade-offs were substantially attenuated such that they were marginal (<9%) or non-significant. This phylogenetically resolved analysis of rubisco kinetic evolution also identified kinetic changes that occur concomitant with the evolution of $C_4$ photosynthesis. Finally, we show that phylogenetic constraints have played a larger role than catalytic trade-offs in limiting the evolution of rubisco kinetics. Thus, although there is strong evidence for some catalytic trade-offs, rubisco adaptation has been more limited by phylogenetic constraint than by the combined action of all catalytic trade-offs.**

***Key words:*** **evolution, rubisco, phylogenetic constraint, catalytic constraint, $C_4$ photosynthesis.**

Article

## Introduction

The vast majority of organic carbon on Earth entered the biosphere via the catalytic pocket of rubisco (ribulose-1,5-bisphosphate [RuBP] carboxylase/oxygenase) (Beer et al. 2010). Although 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 catalyzes the initial step of the Calvin–Benson–Bassham reductive pentose phosphate pathway, catalyzing the fixation of $CO_2$ onto the acceptor molecule RuBP to ultimately synthesize sugars. There is a diverse array of rubisco forms found across the tree of life. Plants and some bacteria contain Form I rubisco which is composed of both large (RbcL) and small (RbcS) subunits (Schneider et al. 1992; Tabita et al. 2008), whereas Form II, III, and IV rubisco found in other lineages are comprised of just the large subunit (Tabita et al. 2008; Banda et al. 2020). Although Form I rubisco contains two subunits, only the large subunit is essential for catalysis (Andrews 1988; Lee and Tabita 1990; Whitney et al. 2011), whereas the small subunit has an indirect effect on catalytic properties and activity (Andrews 1988; Lee and Tabita 1990; Lee et al. 1991; Read and Tabita 1992a, 1992b; Spreitzer et al. 2005; Ishikawa et al. 2011; Joshi et al. 2015; Fukayama et al. 2019; Martin-Avila et al. 2020).

Given that rubisco is the entry point for carbon into the global food chain it is perhaps unsurprising that it is the most abundant enzyme on Earth (Ellis 1979) with a global mass of approximately 0.7 gigatons (Bar-On and Milo 2019). However, this abundance is in part due to the inefficiency of rubisco as a catalyst. Specifically, rubisco has a low rate of $CO_2$ assimilation (Tcherkez et al. 2006; Savir et al. 2010) and is poorly able to discriminate $CO_2$ and $O_2$ (Ogren and Bowes 1971) causing it to catalyze both a carboxylation and an oxygenation reaction (Bowes et al. 1971; Chollet 1977; Sharkey 2020). Rubisco-mediated oxygenation of RuBP results in the production of 2-phosphoglycolate, which must then be metabolized to recover carbon and avoid depletion of metabolite pools (Eckardt 2005; Sharwood 2017). In plants, this carbon scavenging process is known as photorespiration, and consumes

ATP and reducing power and also liberates ammonia which must be re-assimilated (Peterhansel et al. 2010). Although the oxygenation reaction catalyzed by rubisco is not thought 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 serving a number of beneficial functions (Busch 2020), at its current rate rubisco oxygenation represents a substantial metabolic burden reducing the productivity of 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 (Flamholz and Shih 2020). 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; Flamholz and Shih 2020). These $CO_2$-concentrating mechanisms range in complexity from the carboxysome microcompartments in cyanobacteria (Kaplan et al. 1991; Espie and Kimber 2011), to the physical separation of primary $CO_2$ assimilation from rubisco-mediated photosynthetic $CO_2$ reduction in plants that conduct $C_4$ photosynthesis (Hatch 1987; Sage et al. 2012; Edwards 2019). The observation that evolution has resulted in 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 et al. 2011; Sharwood et al. 2016; Orr et al. 2017; Sharwood 2017; Sharkey 2020). This proposition that rubisco specificity 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 et al. 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 catalyze $CO_2$ fixation, and that contemporary rubisco are near perfectly adapted within this heavily constrained catalytic landscape (Tcherkez et al. 2006; Savir et al. 2010). However, new evidence has begun to question this paradigm of rubisco evolution. First, recent analyses 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; Iñiguez et al. 2020). Furthermore, engineering efforts to alter rubisco kinetics have produced enzyme variants that deviate from proposed catalytic trade-offs between $S_{C/O}$, $k_{catC}$, and $K_C$ (Wilson et al. 2018; Zhou and Whitney 2019). Finally, an updated examination of rubisco kinetics in the context of other enzymes has shown that it is not as inefficient a catalyst as often assumed (Bathellier et al. 2018). 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; Gunn et al. 2020).

Although the kinetic 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 have assumed that measurements of rubisco kinetic traits in different species are independent (Tcherkez, Farquhar and Andrews 2006; Savir et al. 2010; Flamholz et al. 2019; Iñiguez et al. 2020). However, this assumption has never formally been tested and is unlikely to be true because rubisco in all species are related to each other by descent from a single ancestral gene. This means that rubiscos in closely related species are more similar than rubiscos in species that are more distantly related, a feature 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). As sequence variation determines kinetic variation, closely related enzymes would be expected to 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, 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 a 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 inferred from the presence of correlations are, either wholly or in part, an artefact caused by this phylogenetic signal.

Here, we assess the presence of a phylogenetic signal in rubisco kinetic traits to evaluate the extent to which rubisco kinetic evolution is constrained by both phylogenetic effects and catalytic trade-offs. We demonstrate that there is a significant phylogenetic signal in all rubisco kinetic traits. This means that the similarity of kinetic measurements between species varies as a function of their evolutionary distance, and thus kinetic measurements in different species are non-independent. When this phylogenetic signal is correctly accounted for by using phylogenetic least squares regression, we reveal that inferred catalytic trade-offs are weak and that rubisco kinetic traits have been evolving largely independently of each other. Moreover, we find that phylogenetic constraints, most likely resulting from a slow rate of molecular evolution, have constrained rubisco kinetic evolution to a greater extent than catalytic trade-offs. This new insight offers encouragement to efforts which aim to increase yields in

food, fiber, and fuel crops by engineering rubisco variants with increased catalytic efficiency.

## Results

### Rubisco Kinetic Data

A data set 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 data set were normalized in order to overcome discrepancies between values determined using an oxygen electrode assay (Parry et al. 1989) and high precision gas-phase-controlled 3H-RuBP-fixation assays (Kane et al. 1994) (see Materials and Methods). To begin, the interrogation of this data was focused on the angiosperms because this was the group with the largest and most complete set of kinetic measurements, and to minimize any impact of long-branch effects (Su and Townsend 2015). It was also restricted to those species with measurements of $S_{C/O}$, maximum carboxylase turnover rate per active site ($k_{catC}$), and the Michaelis constant (i.e., the substrate concentration at half-saturated catalyzed 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 Materials and Methods). Of the 137 angiosperms 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.

### A Significant Phylogenetic Signal Exists in Rubisco Carboxylase-Related Kinetic Traits in Angiosperms

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. To assess whether rubisco in different angiosperms display similar kinetics as a consequence of their phylogenetic relationship, the kinetic traits were analyzed in the context of the phylogenetic tree by which the species are related (fig. 1). Here, all kinetic traits were subject to interrogation for a 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 detection of phylogenetic signal were implemented and the presence or absence of a phylogenetic signal in each trait was judged by the majority result (i.e., the result of $\geq$3 out of 5 methods tested). Out of the methods utilized, Pagel's lambda (Pagel 1999) and Blomberg's $K$ and $K^*$ (Blomberg et al. 2003) analyze 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 a 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 et al. 1985, 1986). For further discussion of the differences between these phylogenetic signal detection methods, see Münkemüller et al. (2012).

Irrespective of the methodological approach used for inference, a significant phylogenetic signal was observed in all carboxylase-related kinetic traits ($S_{C/O}$, $k_{catC}$, $K_C$, and $K_C^{air}$) in angiosperms (table 1; fig. 1). However, the strength of this signal varied across the different methods (table 1). In contrast, a phylogenetic signal was not detected for the oxygenase-related kinetic trait $K_O$ (table 1; fig. 1). 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, figs. S1 and S2 and table S1, Supplementary Material online). Overall, this means that the similarity in carboxylase-related (but not oxygenase-related) kinetic traits in different angiosperms is dependent on their phylogenetic relationship. Therefore, inferred correlations that assume independence between carboxylase-related kinetic trait values are incorrect, and correlation coefficients computed using such approaches have likely been overestimated (Felsenstein 1985; Grafen 1989; Pagel and Harvey 1989; Gittleman and Kot 1990; Abouheif 1999; Pagel 1999; Garland 2001; Blomberg et al. 2003).

### Significant Changes in Rubisco Kinetics Occur during the Evolution of $C_4$ Photosynthesis

Inspection of the data identified several dependencies in rubisco kinetic traits between $C_3$ and $C_4$ plants (fig. 2). 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 $S_{C/O} = 89.9\ mol.mol^{-1}$) (fig. 2; $P < 0.001$, $t$-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 $k_{catC} = 3.2\ s^{-1}$) (fig. 2; $P < 0.001$, $t$-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 $K_C = 19.0\ \mu M$, mean $K_C^{air} = 29.9\ \mu M$) than in $C_3$ species (mean $K_C = 15.4\ \mu M$, mean $K_C^{air} = 23.6\ \mu M$) (fig. 2; $P < 0.05$ and $P < 0.05$, $t$-test, respectively). In contrast, no significant difference was observed in $K_O$ between $C_3$ species (mean $K_O = 481.0\ \mu M$) and $C_4$ species (mean $K_O = 466.7\ \mu M$) (fig. 2; $P > 0.05$, $t$-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). 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 $K_O$ (fig. 2). Owing to a limited number of 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.

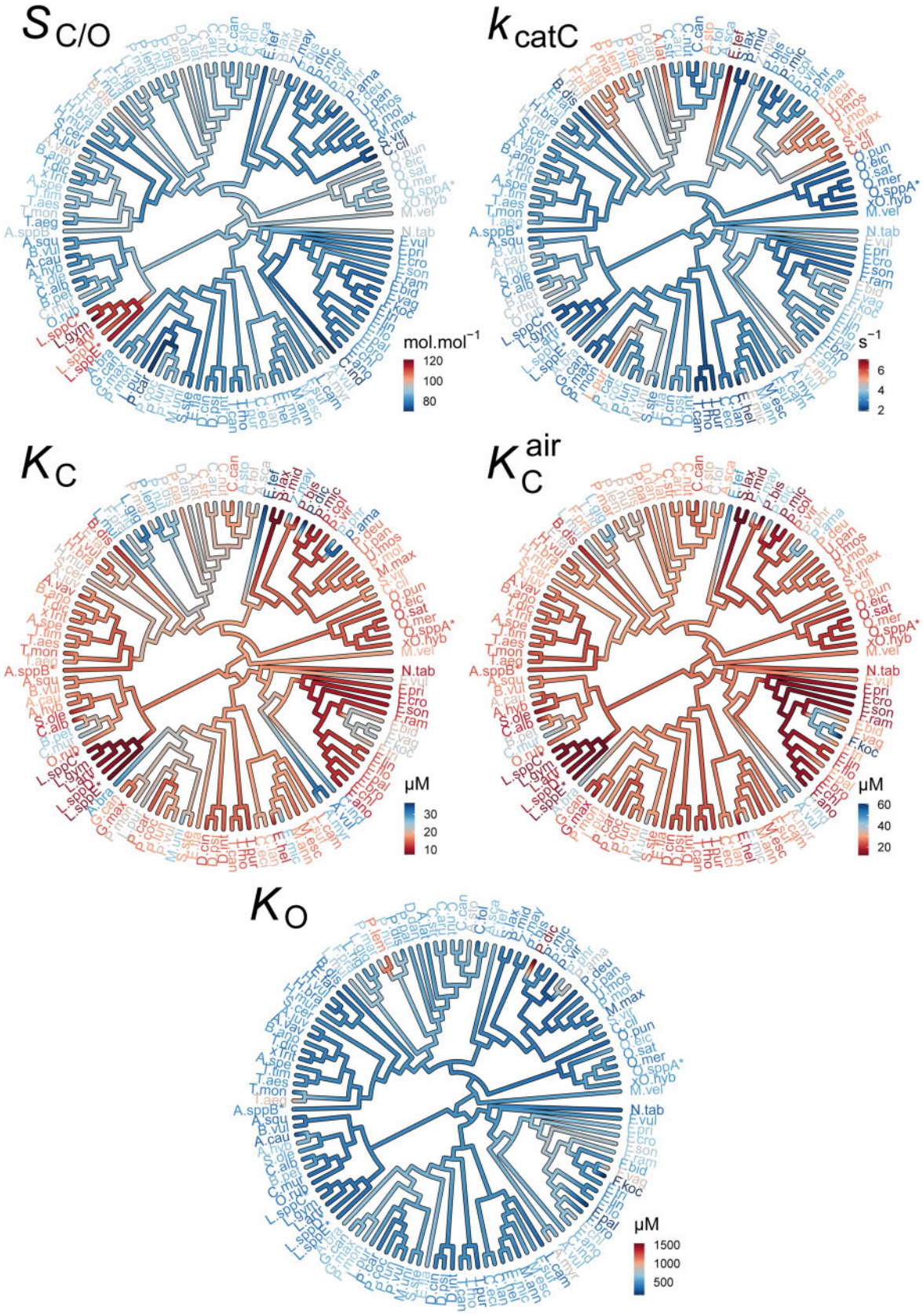


**Fig. 1.** The evolution of rubisco kinetic traits in angiosperms. Phylogenetic tree of angiosperms showing the kinetic trait values in the rubiscos used in this data set and the inferred ancestral kinetic traits for internal branches on the tree. Scale bars for color schemes are presented next to each tree. Species names have been abbreviated for legibility and are provided in full in supplementary file 1, table S4, Supplementary Material online. $S_{C/O}$: specificity. $k_{catC}$: carboxylase turnover per site. $K_C$: the Michaelis constant for $CO_2$. $K_C^{air}$ the inferred Michaelis constant for $CO_2$ in 20.95% $O_2$ air. $K_O$: the Michaelis constant for $O_2$.

**Table 1.** The Phylogenetic Signal Strength and Associated Significance Level in Rubisco Kinetic Traits in Angiosperms Using Five Different Signal Detection Methods.

| Kinetic Trait | C mean | | *I* | | *K* | | *K** | | Lambda | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *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 |

NOTE.—Statistics are rounded to three decimal places and 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 $\alpha =$ ns if $P > 0.05$.

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 (fig. 2), thus indicating that adaptation of this trait may occur early during the evolution of $C_4$ photosynthesis. All of the significant differences between $C_3$ and $C_4$ plants reported above were robust to correction for phylogenetic signal (supplementary file 1, table S2, Supplementary Material online). Collectively, these data demonstrate that there are adaptations 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.

## A Significant Phylogenetic Signal Exists in 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, fig. S1, Supplementary Material online), multiple independent transitions to $C_4$ photosynthesis are present in the data set. 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 an increased variability in $K_O$ (fig. 1; supplementary file 1, table S2, Supplementary Material online), it was hypothesized that a failure to account for kinetic differences associated with photosynthetic type may have confounded estimations of the 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 evolve similar kinetic trait values by convergence. To evaluate the extent to which these respective issues may have affected quantification of the phylogenetic signal, the above analyses were repeated using only the $C_3$ angiosperm species present (i.e., with $C_3$–$C_4$ intermediate, $C_4$-like, and $C_4$ species removed). In general, estimates of the phylogenetic signal in the carboxylase-related kinetic traits in $C_3$ species (table 2) agreed with those observed when all angiosperms were considered (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 (tables 1 and 2). In addition, the discrepancies in signal strength between the methods observed for all angiosperms (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 the phylogenetic signal measured by each test (Hardy and Pavoine 2012; Münkemüller et al. 2012). In summary, there is a statistically significant phylogenetic signal 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 angiosperms (table 1), a significant phylogenetic signal was observed in $K_O$ when only $C_3$ angiosperms were considered (table 2). Thus, both the oxygenase-related and carboxylase-related traits of rubisco have evolved in a tree-like manner in $C_3$ angiosperms. Furthermore, unlike the other carboxylase-related kinetic traits, the phylogenetic signal in $K_C^{air}$ was found to increase in strength when the analysis is restricted to $C_3$ angiosperms. This result is a corollary of the fact that $K_C^{air}$ is computed here from both $K_C$ and $K_O$. Thus, all kinetic traits of rubisco have a significant phylogenetic signal in $C_3$ angiosperms.

## Correlations between Kinetic Traits Are Weak in Angiosperms and Further Relaxed after Correcting for a Phylogenetic Signal

Given the finding that rubisco kinetic traits exhibit a significant phylogenetic signal (table 1; fig. 2), it is possible that previously reported correlations between rubisco kinetic traits (Tcherkez et al. 2006; Savir et al. 2010; Flamholz et al. 2019; Iñiguez et al. 2020) are an artefact of this signal. This is because prior analyses which did not account for the inherent phylogenetic structure (and non-independence) of this data (fig. 3A) may have overestimated correlation coefficients due to this underlying structure. Thus, in order to evaluate the extent to which phylogenetic signal may have influenced previous results (Tcherkez et al. 2006; Savir et al. 2010; Flamholz et al. 2019; Iñiguez et al. 2020), the correlations observed in the kinetic trait data using both phylogenetic and non-phylogenetic regression methods were compared (fig. 3B and C).

Using a standard non-phylogenetic approach, the relationships between kinetic traits of rubisco were consistent in both linear and least squares regression models (supplementary file

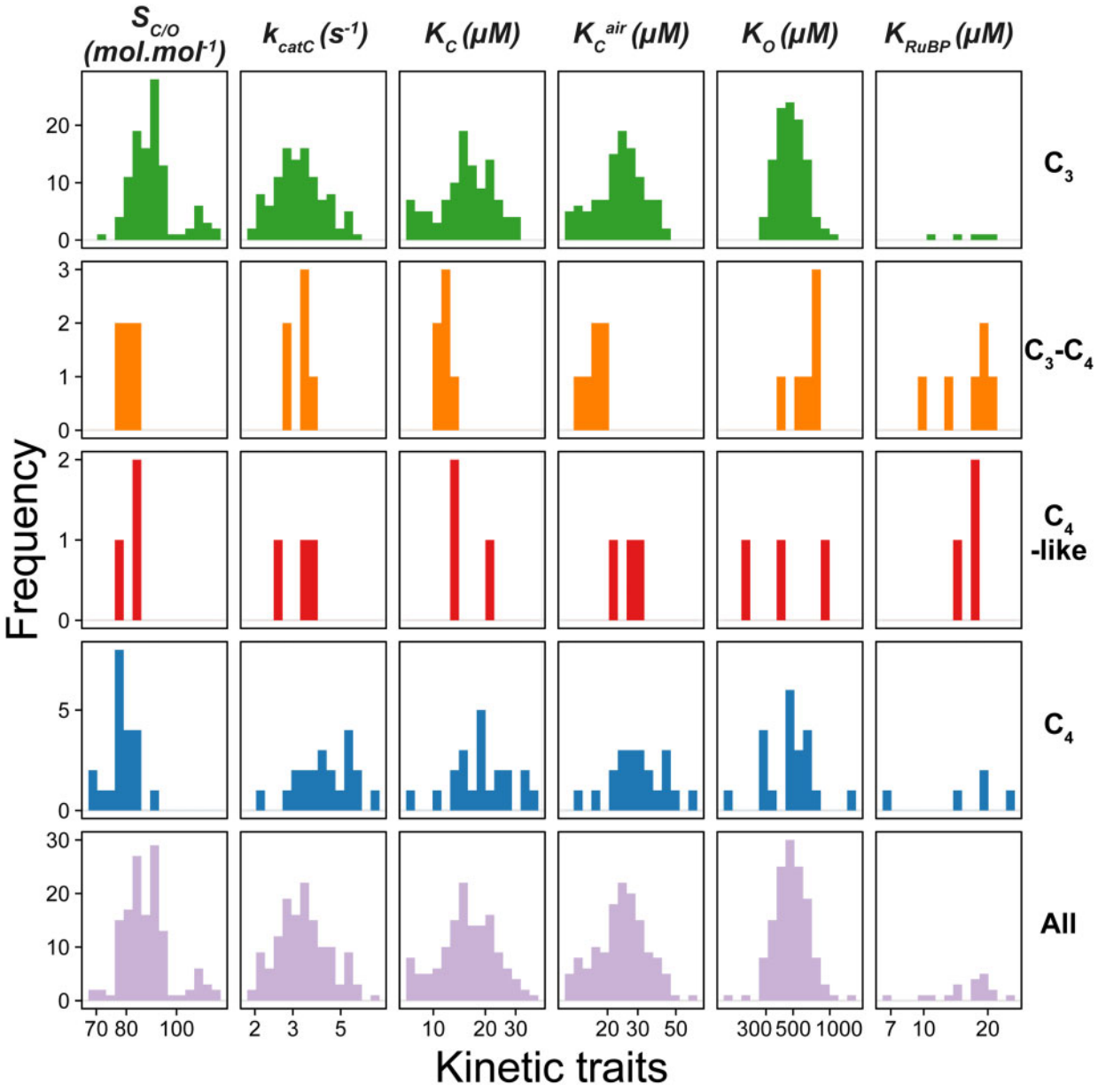


FIG. 2. The distributions of values for rubisco kinetic traits in angiosperms. Species are grouped by their photosynthetic type (rows). $S_{C/O}$: specificity. $k_{catC}$: carboxylase turnover per site. $K_C$: the Michaelis constant for $CO_2$. $K_C^{air}$ the inferred Michaelis constant for $CO_2$ in 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 ($C_3$; $n = 107$), $C_4$ photosynthesis ($C_4$; $n = 21$), $C_3$–$C_4$ intermediates ($C_3$–$C_4$; $n = 6$), $C_4$-like ($C_4$-like; $n = 3$). The *X* axis for all plots is on a log scale, where respective units are shown in column labels. The raw data set used can be found in supplementary file 2, Supplementary Material online.

1, fig. S3*A* and *B*, Supplementary Material online). The direction of the power–law relationships observed (fig. 3*B*) match those previously reported (Flamholz et al. 2019). Specifically, significant positive correlations were found between $k_{catC}$ and both $K_C$ and $K_C^{air}$ (fig. 3*B* and *C*). A significant positive correlation was also observed between the respective Michaelis constants for both $CO_2$ and $O_2$ substrates, $K_C$ and $K_O$ (fig. 3*B* and *C*). In addition, significant inverse power–law correlations were observed between $S_{C/O}$ and all other carboxylase-related kinetic traits, including $k_{catC}$, $K_C$, and $K_C^{air}$ (fig. 3*B* and *C*). In contrast, $K_O$ did not co-vary with either $S_{C/O}$ or $k_{catC}$ (fig. 3*B* and *C*), whereas $K_{RuBP}$ did not appear to be tightly linked to any kinetic trait from the limited number of observations that are available (supplementary file 1, fig. S3*A* and *B*, Supplementary Material online). Thus, across angiosperms, all pairwise relationships between the carboxylase-related kinetic traits $S_{C/O}$, $k_{catC}$, and either $K_C$ or $K_C^{air}$ were significant, whereas 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

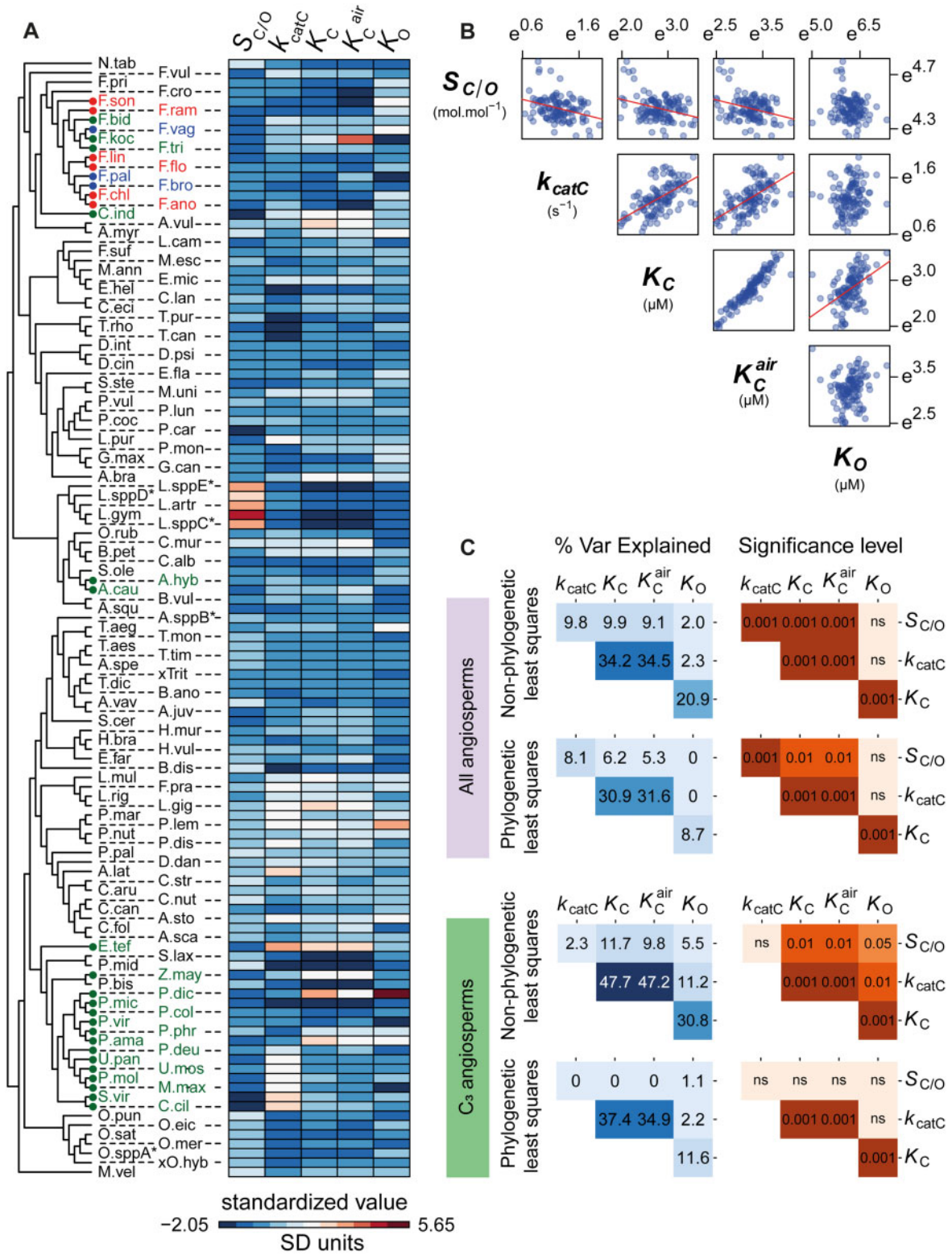


FIG. 3. The correlations between rubisco kinetic traits in angiosperms. (*A*) Heatmap depicting the variation in kinetic traits across the species used in this study (± SD away from each respective kinetic trait mean). Species labels on the tree are color coded by photosynthetic type ($C_3$: black, $C_3$–$C_4$ intermediates: red, $C_4$-like: blue, and $C_4$: green), and have been abbreviated for legibility (for full names refer to supplementary file 1, table S4, Supplementary Material online). (*B*) The relationships between all pairwise combinations of log transformed rubisco kinetic traits. (*C*) Pairwise correlation coefficients (percent variance explained) and associated *P*-values between rubisco kinetic traits assessed using non-phylogenetic least squares regression models or phylogenetic least squares regression models. Phylogenetic and non-phylogenetic least squares regressions were fit to both the complete set of angiosperms in the data set and the subset which perform $C_3$ photosynthesis. Significance values are represented as $\alpha$ 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 $\alpha =$ ns if $P > 0.05$.

**Table 2.** The Phylogenetic Signal Strength and Associated Significance Level in Rubisco Kinetic Traits in $C_3$ Species Using Five Different Signal Detection Methods.

| Kinetic Trait | *C* mean | | *I* | | *K* | | *K** | | Lambda | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *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 |

NOTE.—Statistics are rounded to three decimal places, and 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 $\alpha =$ ns if $P > 0.05$.

substantially reduced in magnitude when the analysis was focused solely on the angiosperms. This reduction in magnitude of correlation when taxonomic groups are removed is strongly indicative of phylogenetic signal in the data set and is analyzed in further detail in a subsequent results section. Within angiosperms, the strength of the correlation between $S_{C/O}$ and $K_C$ (9.9% variance explained; fig. 3C) is attenuated by 77% when compared with that previously reported using a larger range of rubisco variants (43.6% variance explained; Flamholz et al. 2019). Moreover, a 69% reduction was found in the dependency between $S_{C/O}$ and $k_{catC}$ in angiosperms (9.8% variance explained; fig. 3C) in comparison to that reported based on the larger range of species (31.4% variance explained; Flamholz et al. 2019), with the antagonistic correlation observed between $K_C$ and $K_O$ (20.9% variance explained; fig. 3C) also weakened by 33% relative to previous reports (31.4% variance explained; Flamholz et al. 2019). In contrast, the dependency between $K_C$ and $k_{catC}$ was 49% stronger when only angiosperms are assessed, increasing from 23.0% (Flamholz et al. 2019) to 34.2% in this study (fig. 3C). Therefore 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 (Flamholz et al. 2019; Iñiguez et al. 2020).

Given that a significant phylogenetic signal is present in rubisco kinetic traits in angiosperms (tables 1 and 2), a phylogenetic generalized least squares regression analysis (Felsenstein 1985) was conducted to estimate the magnitude of the catalytic trade-offs when accounting for the inherent structure of the data. In comparison to phylogeny-unaware correlations, the phylogenetic regression resulted in a reduction in the majority of kinetic trade-offs (fig. 3C). The largest reduction observed was for the correlation between the Michaelis constants $K_C$ and $K_O$. Here, the correlation was reduced by 58% (variance explained = 8.7%) relative to methods which do not correctly account for the non-independence of these measurements (variance explained = 20.9%; fig. 3C). An analogous weak correlation was also observed when the phylogenetic analyses were limited to $C_3$ species (variance explained = 11.6%; fig. 3C). Thus, changes in rubisco $K_C$ have occurred largely independent of any change on $K_O$ during the diversification of the angiosperms.

Phylogenetic correction also resulted in less substantial reductions in the correlation between $S_{C/O}$ and each of the other carboxylase-related traits (fig. 3C). Here the dependency between $S_{C/O}$ and $k_{catC}$ was reduced by 18% from 9.8% to 8.1%, whereas the dependency between $S_{C/O}$ and both $K_C$ and $K_C^{air}$ was reduced by 37% and 42% from 9.9% to 6.2%, and from 9.1% to 5.3%, respectively (fig. 3C). Furthermore, these correlations were not significant when considering only $C_3$ species (fig. 3C). Thus, during the evolution of rubisco in angiosperms, changes to specificity have occurred with little or no effect on other carboxylase-related kinetic traits, and *vice versa*.

In contrast, the strength of the correlation between $k_{catC}$ and either $K_C$ or $K_C^{air}$ was robust to phylogenetic correction. Specifically, the dependency between $k_{catC}$ and $K_C$ only decreased by 10% from 34.2% to 30.9%, and the dependency between $k_{catC}$ and $K_C^{air}$ decreased by only 8% from 34.5% to 31.6% (fig. 3C). Furthermore, 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; fig. 3C). Thus, as rubisco kinetics have evolved in angiosperms, there has been a trade-off between $CO_2$ affinity and carboxylase turnover such that any change in one kinetic trait caused a partial change in the other, though with little impact on any further rubisco kinetic traits.

## The Evolution of Rubisco Kinetics Is More Limited by Phylogenetic Constraints Than by Catalytic Trade-Offs in Angiosperms

As rubisco kinetic traits contain a phylogenetic signal in angiosperms (tables 1 and 2), we sought to determine the extent to which the phylogenetic signal was caused by phylogenetic constraint. Here, phylogenetic constraint is considered to comprise all constraints which are embedded within the structure of the phylogenetic tree, that are independent of the kinetic constraints, and collectively act to impede the adaptive evolution of rubisco kinetics. For example, such phylogenetic constraints include processes which lead to neutral evolution (Felsenstein 1985) or evolutionary stasis (Prinzing et al. 2001; Ackerly 2004; Cavender-Bares et al. 2004; Moles et al. 2005; Swenson and Enquist 2007) of the trait in question (Ives 2019; Nevo et al. 2020). In order to assess the relative strength of such phylogenetic constraints on rubisco kinetics, the variance in kinetic traits partitioned between phylogenetic effects (i.e., the explanatory power of the phylogenetic tree in the goodness-of-fit model and a measure of phylogenetic constraint) and non-

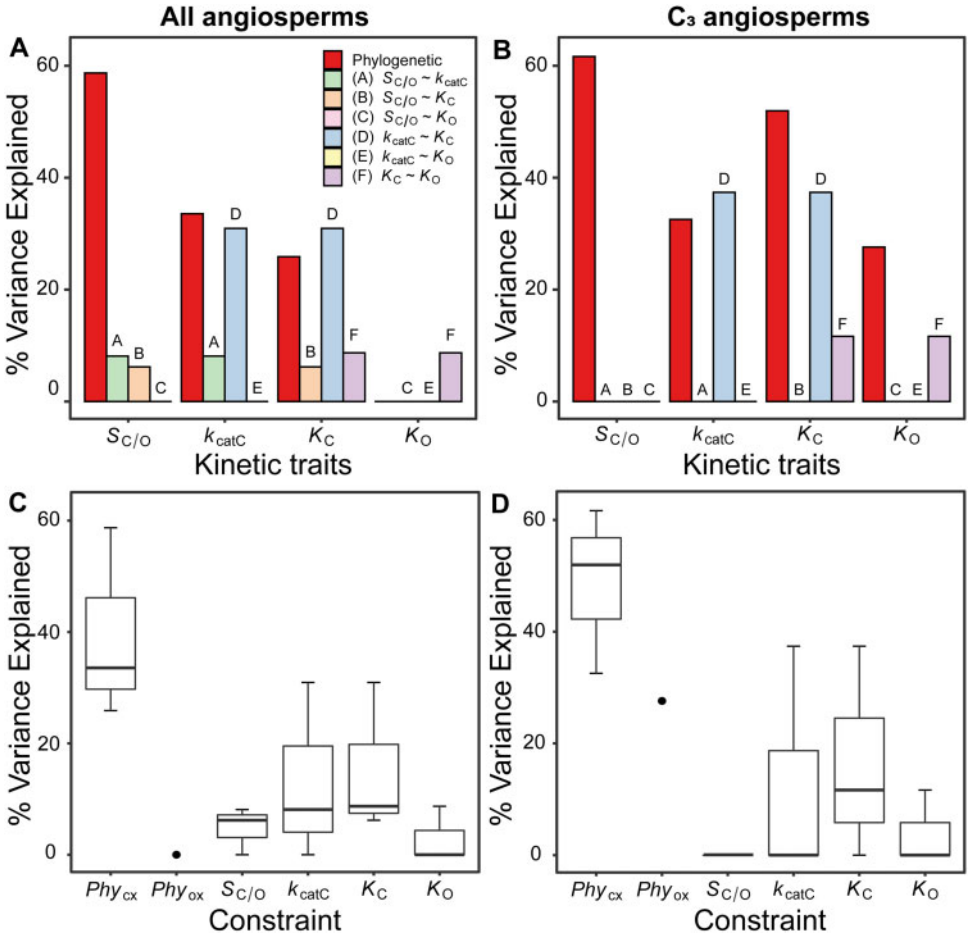


FIG. 4. The constraints on rubisco kinetic adaptation in angiosperms. (*A*) The variation (%) in rubisco kinetic traits across angiosperms that can be explained by phylogenetic constraint and each catalytic trade-off. (*B*) As in (A) but for $C_3$ angiosperms only. (*C*) Boxplot of all variation explained in each kinetic trait by kinetic trait correlations in comparison to variation explained by phylogeny in angiosperms. The phylogenetic constraints on the carboxylase-related traits $Phy_{CX}$ (includes $Phy_{Sc/o}$, $Phy_{Kcatc}$, and $Phy_{Kc}$) and phylogenetic constraints on the oxygenase-related trait $Phy_{ox}$ (includes $Phy_{Ko}$ only) are presented separately. (*D*) As in (C) but for $C_3$ angiosperms only.

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) was quantified. This analysis revealed that phylogenetic constraints explained a significant proportion of the variation in either the carboxylase-related kinetic traits across all angiosperms (fig. 4*A*), or in all kinetic traits across $C_3$ angiosperms (fig. 4*B*). With one exception (i.e., the phylogenetic constraint on $K_O$ in the larger species data set) the magnitude of variation explained by phylogenetic constraints was similar or larger to the variation explained by the strongest trade-off observed between $k_{catC}$ and $K_C$ (fig. 4*C* and *D*). Consequently, in angiosperms, the cumulative variance explained by phylogenetic constraints across all rubisco kinetic traits (29.5%) was larger than the cumulative variance for all catalytic trade-offs combined (9.0%). This effect was more pronounced for $C_3$ angiosperms (cumulative variance for phylogenetic constraints = 43.4%, cumulative variance for catalytic trade-offs = 8.2%). Thus, during the radiation of angiosperms phylogenetic constraints have restricted the evolution of rubisco kinetics to a greater extent than all catalytic trade-offs combined.

## Phylogenetic Signal, Weak Kinetic Trait Correlations, and Strong Phylogenetic Constraint Are Features of Rubisco Evolution in All Photosynthetic Organisms

Given the presence of phylogenetic signal and the impact of phylogenetic constraint on the evolution of rubisco kinetics in angiosperms, we sought to determine whether these findings were a unique feature of angiosperm rubisco or whether they were a more general phenomenon across the tree of life. To achieve this, the data set was expanded to include all species for which both kinetic measurements and an *rbcL* gene sequence were available. Analogous to the analysis of angiosperms, a strong and statistically significant phylogenetic signal was observed in $S_{C/O}$, $k_{catC}$, $K_C$, and $K_C^{air}$, but not in $K_O$ across all photosynthetic organisms (table 3). Similarly, a significant phylogenetic signal was also observed for $K_O$ when $C_3$–$C_4$, $C_4$-like, and $C_4$ angiosperms were omitted to control for the dependency in kinetic trait measurements on the tree associated with the convergent transition to $C_4$ photosynthesis in angiosperms (supplementary file 1, table S3, Supplementary Material online). Thus, there is a significant phylogenetic signal in rubisco kinetic traits in all photosynthetic organisms.

**Table 3.** The Phylogenetic Signal Strength and Associated Significance Level in Rubisco Kinetic Traits across All Studied Photosynthetic Organisms Using Five Different Signal Detection Methods.

| Kinetic Trait | *C* mean | | *I* | | *K* | | *K** | | Lambda | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *Stat* | α | *Stat* | α | *Stat* | α | *Stat* | α | *Stat* | α |
| $S_{C/O}$ | 0.712 | 0.001 | 0.229 | 0.001 | 0.005 | 0.001 | 0.002 | 0.001 | 1.005 | 0.001 |
| $k_{catC}$ | 0.538 | 0.001 | 0.29 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.975 | 0.001 |
| $K_C$ | 0.705 | 0.001 | 0.277 | 0.001 | 0.002 | 0.001 | 0.001 | 0.001 | 0.948 | 0.001 |
| $K_C^{air}$ | 0.609 | 0.001 | 0.299 | 0.001 | 0.001 | 0.001 | 0.001 | 0.01 | 0.924 | 0.001 |
| $K_O$ | 0.170 | 0.001 | 0.004 | ns | 0 | ns | 0 | ns | 0.603 | 0.001 |

NOTE.—Statistics are rounded to three decimal places and 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 $\alpha =$ ns if $P > 0.05$.

Analogous to the above analyses, accounting for the phylogenetic tree (supplementary file 1, fig. S4, Supplementary Material online) caused a substantial attenuation in the kinetic trait correlations in all species (fig. 5A; supplementary file 1 and figs. S3C, S5, and S6, Supplementary Material online). Specifically, when correcting for the phylogenetic signal in kinetic traits, a partial correlation between $k_{catC}$ and both $K_C$ and $K_C^{air}$ was observed (variance explained = 21.3% and 23.3%, respectively; fig. 5A). Furthermore, a partial correlation was also measured between $K_C$ and $K_O$ (variance explained = 18.6%; fig. 5A). However, correlations between all other pairwise combinations of kinetic traits were found to be either marginal, or not significant (fig. 5A). In addition, the dependency between $K_C$ and $K_O$ was attenuated to 13.4% when the $C_3$–$C_4$, $C_4$-like, and $C_4$ angiosperms were excluded from the analysis (supplementary file 1, fig. S7A, Supplementary Material online).

Evaluation of the phylogenetic constraints revealed that they explained a significant proportion of variation in the evolution of all rubisco kinetic traits (fig. 5B). Moreover, the phylogenetic constraints explained a larger proportion of kinetic trait variation than catalytic trade-offs (fig. 5C), such that the cumulate variation explained by phylogenetic constraints (56.1%) was larger than the combined effect of all catalytic trade-offs (8.0%). Analogous results were recovered when $C_3$–$C_4$, $C_4$-like, and $C_4$ species were removed from the analysis (cumulative variance for phylogenetic constraints = 61.4%, cumulative variance for catalytic trade-offs = 6.2%; supplementary file 1, fig. S7B and C, Supplementary Material online). Thus, phylogenetic constraints have been a critical limitation on rubisco adaptation in a diverse range of photoautotrophs and have presented a greater barrier to kinetic evolution than that imposed by all catalytic trade-offs combined.

## 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 et al. 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 et al. 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, while also causing it to catalyze a reaction with $O_2$ which is energetically expensive and results in a loss of fixed carbon (Bowes et al. 1971; Chollet 1977). However, all trade-offs have been inferred based on the assumption that rubisco in different species are independent (Tcherkez et al. 2006; Savir et al. 2010; Flamholz et al. 2019; Iñiguez et al. 2020). Here, we find that this assumption was incorrect and show that a significant phylogenetic signal is found in rubisco kinetic traits across the tree of life. We reevaluated the extent of rubisco catalytic trade-offs accounting for this phylogenetic signal and found that all catalytic trade-offs were attenuated. The largest trade-offs were observed between the Michaelis constant for $CO_2$ and carboxylase turnover (~21–37%), and between the Michaelis constants for $CO_2$ and $O_2$ (~9–19%), respectively. Furthermore, we demonstrated that all other catalytic trade-offs were either nonsignificant or substantially attenuated when the phylogenetic relationship of the species was correctly accounted for. Finally, we found that phylogenetic constraints have played a larger role than catalytic trade-offs in limiting the evolution of rubisco kinetics. Thus, rubisco kinetics have been evolving largely independently of each other in an adaptive landscape that is predominantly limited by phylogenetic constraint.

The presence of a phylogenetic signal in rubisco kinetic traits simply means that rubisco kinetics are more similar among close relatives, with this similarity changing as a function of the phylogenetic distance between species. This result is perhaps not surprising given that all extant rubisco are related by the process of descent with modification from a single ancestral enzyme (Nisbet et al. 2007). However, not all biological traits contain a phylogenetic signal (Blomberg et al. 2003; Kamilar and Cooper 2013). Furthermore, the functional consequences of changes to enzyme sequences are hard to predict (Minshull et al. 2005; Damborsky and Brezovsky 2014; Siegel et al. 2015; Carlin et al. 2016), with single amino acid substitutions often causing large effects in enzyme kinetics (Cleton-Jansen et al. 1991; Villar et al. 1997; Johnson et al. 2001). Thus, *a priori* it was unknown whether any or all of the rubisco kinetic traits would exhibit a phylogenetic signal. It will be interesting to see whether the presence of a phylogenetic signal in enzyme kinetic data is a phenomenon that is specific to rubisco, and if not it will likely be important to account for this non-independence when comparing the catalytic properties of enzymes across the tree of life.

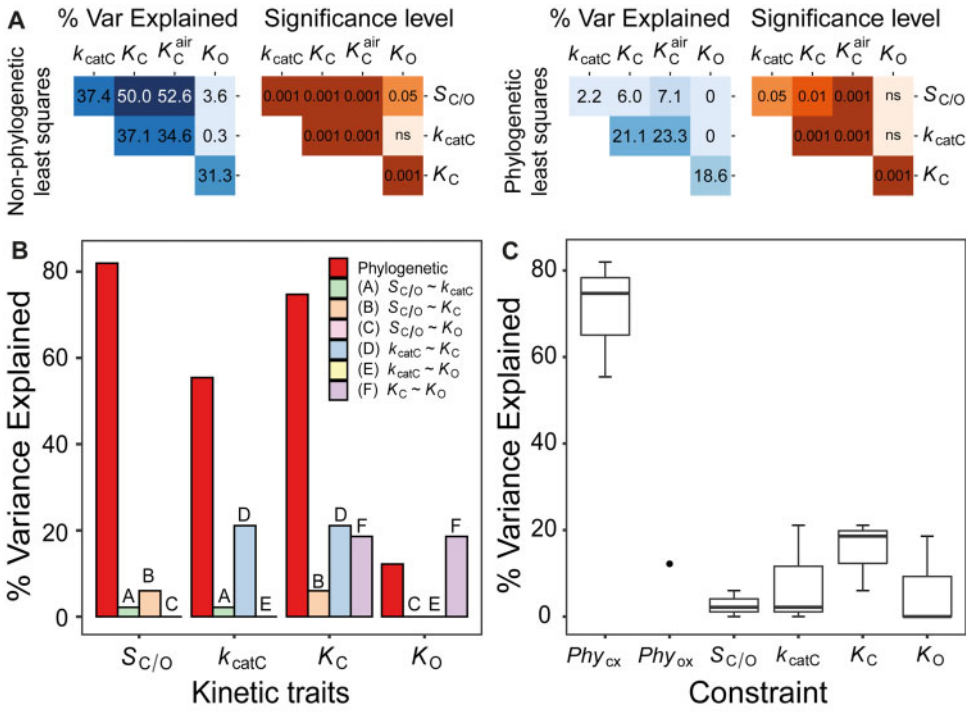


FIG. 5. Kinetic and phylogenetic constraints on rubisco adaptation across all photosynthetic organisms. (*A*) Pairwise correlation coefficients (percent variance explained) and associated *P*-values between different rubisco kinetic traits assessed using non-phylogenetic least squares regression models or phylogenetic least squares regression models. Significance values are represented as $\alpha$ 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 $\alpha =$ ns if $P > 0.05$. (*B*) The variation (%) in rubisco kinetic traits across all photosynthetic organisms that can be explained by phylogenetic constraint and each catalytic trade-off. (*C*) Boxplot of all variation explained in each kinetic trait by kinetic trait correlations in comparison to variation explained by phylogeny in all photosynthetic organisms. The phylogenetic constraints on the carboxylase-related traits $Phy_{CX}$ (includes $Phy_{Sc/o}$, $Phy_{Kcatc}$, and $Phy_{Kc}$) and phylogenetic constraints on the oxygenase-related trait $Phy_{ox}$ (includes $Phy_{Ko}$ only) are presented separately.

In this work we reveal that the phylogenetic signal in rubisco kinetics is caused by phylogenetic constraint on rubisco that is independent of the catalytic trade-offs. Phylogenetic constraint in this context includes all of the processes that collectively lead to slow rates of adaptation. These processes include 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 multiple factors contribute to the phylogenetic constraint detected in rubisco, it is likely that low molecular variability is a key driver of this phenomenon. For example, the rate of molecular evolution of rubisco is likely constrained by the requirements for 1) high levels of transcript and protein abundance (Kelly 2018; Seward and Kelly 2018), 2) 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 3) the need to preserve overall protein stability within the molecular activity-stability trade-offs (Studer et al. 2014; Duraõ et al. 2015; Cummins et al. 2018). In plants, these constraints would be further exacerbated due to the presence of the *rbcL* gene in the organellar genome that is uniparentally inherited and does not recombine (Birky 2001). For example, in angiosperms chloroplast-encoded genes evolve 10 times slower than nuclear-encoded genes (Wolfe et al. 1987; Smith 2015). Combined, these evolutionary constraints would hinder the kinetic adaptation of rubisco resulting in the phylogenetic constraint observed in this study.

The strongest catalytic trade-off detected in this study was the 21–37% dependency that was observed between $k_{catC}$ and both $K_C$ and $K_C^{air}$. This finding is compatible with the mechanistic models of rubisco (Farquhar 1979), and is supported by the recent discovery of rubisco variants which exhibit the highest $k_{catC}$ ever recorded at the expense of poor $CO_2$ affinities (i.e., $K_C$ values $>250\,\mu M$) (Davidi et al. 2020). Nevertheless, the dependency between $CO_2$ affinity and carboxylase turnover, despite being the strongest correlation that was observed, is substantially attenuated relative to the coefficients that are conventionally cited (Tcherkez et al. 2006; Savir et al. 2010). Therefore, 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 species such that variation in one kinetic trait only explains

approximately 21–37% of variation in the other. This fact 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 second strongest catalytic trade-off that was observed was the 9–19% dependency between $K_C$ and $K_O$. This trade-off is compatible with the fact that the singular active site of rubisco binds both $CO_2$ and $O_2$, and thus it is plausible that mutations that affect the active site will affect biding of both substrates, though not necessarily to equivalent extents. All other catalytic trade-offs were either marginal (<9%) or non-significant. Furthermore, the combined effect of all catalytic trade-offs can only account for 6–9% of total variation in rubisco kinetics between species, a substantially smaller component than can be explained by phylogenetic constraint (30–61%).

The phylogenetically resolved analysis of rubisco kinetic evolution also identified changes in kinetic traits associated with the evolution of $C_4$ photosynthesis. Specifically, $S_{C/O}$ was lower in $C_4$ species than in $C_3$ species, whereas $k_{catC}$, $K_C$, 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. These differences in rubisco kinetics would likely be either neutral or adaptive in a $C_4$ context. For example, any change in $K_O$ would effectively be neutral under the elevated $CO_2$ environment of the bundle sheath chloroplast, as it would have only a marginal effect on the *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 elevated $CO_2$ environment would enable higher flux through rubisco and thus provide an energetic advantage. Accordingly, one would expect that an increased variation in $K_O$ in $C_4$ species would occur by neutral drift (Kimura 1991; Savir et al. 2010), and that an increased $k_{catC}$ would confer a selective advantage even if it came at the expense of a partial reduction in $K_C$. 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 the weak catalytic trade-off that exists between $k_{catC}$ and $K_C$. Here, despite the phylogenetic constraints limiting rubisco adaptation, the increased rate at which these kinetic changes occurred in $C_4$ species may have been facilitated by the higher rates of molecular evolution (Kelly 2018) and diversification (Spriggs et al. 2014) that occur concomitant with the evolution of $C_4$ photosynthesis.

Although every effort was taken to prevent systematic or methodological biases from influencing the results, several factors may have led to the underestimation of phylogenetic signal in the data. For example, experimental error in kinetic measurements, and/or inconsistencies associated with measurements being compiled from numerous sources, may have hindered the detection of phylogenetic signal, as has been shown in other studies (Rohlf 2001). However, to help mitigate this problem, the $S_{C/O}$ values used in this analysis were normalized to avoid the discrepancy between the rates measured using an oxygen electrode assay (Parry et al. 1989) and those measured using high precision gas-phase-controlled $^3$H-RuBP-fixation assays (Kane et al. 1994) (see Materials and Methods). Thus, improvements in both the accuracy and breadth of rubisco kinetic measurements across species will lead to a concomitant improvement in our understanding of how rubisco kinetics have evolved.

Given the importance of rubisco to life on Earth, the question as to why a "perfect" rubisco has not already evolved is legitimate. For example, although rubisco $K_C$ is thought to be near optimal in $C_3$ plants 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 $s^{-1}$ per site) has often been considered low (Bar-Even et al. 2011; Tcherkez 2013; Davidi et al. 2018). In addition, all known rubisco variants catalyze a promiscuous and energetically costly reaction with $O_2$. However, a recent review of rubisco kinetics relative to those of other enzymes has argued that rubisco is actually not such a bad catalyst (Bathellier et al. 2018). Indeed, the phylogenetically informed analysis of rubisco presented here demonstrates that the kinetic traits have been able to evolve largely independently of each other, with kinetic evolution primarily limited by phylogenetic constraint. These constraints induce a lag in adaptive evolution that help to explain why the enzyme is better suited to former environmental conditions.

The study presented here highlights the importance of considering phylogenetic relationships when conducting comparative analyses of enzyme kinetics across species. In doing so, it reveals that rubisco evolution has been only weakly constrained by catalytic trade-offs. Instead, phylogenetic constraints, caused by factors that limit the pace of molecular evolution, have provided a more substantial hindrance to rubisco kinetic evolution. Accordingly, it should be feasible in the current synthetic biology revolution to circumvent this evolutionary barrier on rubisco optimization. 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 archea (Wilson et al. 2016), photosynthetic bacteria (Zhou and Whitney 2019), and cyanobacteria (Wilson et al. 2018). Thus, our findings provide optimism for engineering rubisco in food, fiber, and fuel crops to have improved catalytic efficiency.

## Materials and Methods

### Kinetic Data

Kinetic measurements of rubisco 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 et al. 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) in order to minimize methodological biases in the data. Specifically, as rubisco from wheat (*Triticum aestivum*) was measured in both the $O_2$ electrode studies (Orr et al. 2016; Prins et al. 2016) as well

as in the high precision method by Kane et al. (1994), multipliers were applied to all $S_{C/O}$ measurements derived from $O_2$ electrode assays using wheat as an enzyme standard. The distribution of $S_{C/O}$ values in angiosperms before and after normalization can be seen in supplementary file 1, fig. S8, Supplementary Material online).

All kinetic traits in the data set were log transformed consistent with (Flamholz et al. 2019), and the distributional assumptions of each were verified for analyses herein. For the angiosperm focused analysis, only angiosperms with experimental measurements of all four principal kinetic traits of interest ($S_{C/O}$, $k_{catC}$, $K_C$, and $K_O$) were taken forward for subsequent analysis. However, all species in the data set with more than one kinetic trait measurement were considered for subsequent analyses of all photosynthetic organisms. In addition, an estimate 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 253 μM.

In cases where duplicate entries for a species were present in the kinetic data set (including synonyms), the median value of their kinetic traits was used for inference. In this way, medians were also taken for *Triticum timonovum* and *Triticum timopheevii*, the former of which is a synthetic octoploid of the latter (Murashov and Morozova 2008). The modified data set containing corrected $S_{C/O}$ values and no duplicate entries is provided in supplementary file 2, Supplementary Material online. Estimates of $K_{RuBP}$ are provided where available. It should also be noted that values of $k_{catC}$ are presented as units per active site.

## Phylogenetic Tree Inference

As sequenced genomes or transcriptomes do not exist for many species in the kinetic trait data set, whole genome phylogenomic approaches could not be used to infer the species tree necessary in order to detect a phylogenetic signal in the kinetic traits of rubisco . However, the *rbcL* gene that encodes the large subunit of rubisco has a long history of use for phylogenetic inference of species relationships (Gielly and Taberlet 1994; APG 1998, 2016) and was available for all of the angiosperms, and the majority of photosynthetic organisms, that were considered in the analyses. Accordingly, *rbcL* was selected here for use in species tree inference. The coding sequences of *rbcL* for the 137 angiosperm species for which kinetic data was available can be found in supplementary file 3, Supplementary Material online. The coding sequences for *rbcL* for the complete set of 181 photosynthetic organisms for which both kinetic data and sequencing data were available can be found in supplementary file 4, Supplementary Material online. Gene 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). Multiple sequence alignments were 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). These trimmed nucleotide sequence alignments were used for subsequent phylogenetic analysis. Bootstrapped maximum-likelihood phylogenetic trees were 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 best fitting model of sequence evolution was inferred from the data automatically by IQ-TREE. The resultant maximum-likelihood phylogenetic trees were rooted manually using Dendroscope (Huson and Scornavacca 2012). A number of nodes in the angiosperm tree (supplementary file 1, fig. S1, Supplementary Material online) and a number of nodes in the tree of all photosynthetic organisms exhibited terminal zero-length branches due to 100% sequence identify with other closely related species ($n = 18$ and $n = 23$, respectively). 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 data set to 119 angiosperms and 158 photosynthetic organisms. The phylogenetic tree inferred from the angiosperm taxa (supplementary file 1, fig. S1, Supplementary Material online) 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 phylogenetic trees inferred from the *rbcL* gene most accurately reflects the sequence similarity of rubisco, and thus were deemed as suitable 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 inferred using IQ-TREE (Nguyen et al. 2015) following the methods described above but based on a multiple sequence alignment in which columns containing non-synonymous nucleotide sequence changes were removed (supplementary file 5, Supplementary Material online). Due to the considerable loss of phylogenetic information accessible for tree building from this alignment, the species tree inferred using the nucleotide sequences corresponding to these ubiquitously conserved amino acid positions (supplementary file 1, fig. S2, Supplementary Material online) exhibited an additional number ($n = 13$) of angiosperm zero-length terminal branches. 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 the analysis using only this tree. Use of this phylogenetic tree confirmed that the presence of the 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 a phylogenetic signal in kinetic traits was assessed using five different phylogenetic signal detection methods (Gittleman and Kot 1990; Abouheif 1999; Pagel 1999; Blomberg et al. 2003). Here, signal strength was estimated by assessing the distribution of trait values relative to the respective underlying species tree inferred from the *rbcL*

sequences using methods which both depend on an explicit evolutionary model, such as Pagel's lambda (Pagel 1999) and Blomberg's *K* and *K** (Blomberg et al. 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. For further discussion of the differences between phylogenetic signal detection methods, see Münkemüller et al. (2012).

### Ancestral State Estimation and Mapping of Kinetic Traits to the Phylogenetic Tree

Ancestral state estimation was conducted to visualize the evolution of rubisco kinetic traits on the phylogenetic tree which relates the angiosperms. For this purpose, the kinetic data set was mapped and scaled onto the angiosperm species tree by employing the function *ggtree* in the ggtree package (Yu et al. 2017). Here, terminal branches were colored according to the measurement of the kinetic trait in the species which comprise the terminal branch, whereas internal branches were colored 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 phylogenetic signal was corrected for by using branch length transformations of the phylogenetic tree based on the mean maximum-likelihood estimates of lambda calculated for each trait, with kapa and delta held constant. In cases where the mean maximum-likelihood estimates of lambda exceeded the upper limit of the model, this value was set to 1. 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 in the caper package (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 with non-phylogenetic constraints, the rr2 package (Ives and Li 2018) was employed in R. Here, to assess the extent to which phylogeny can explain the 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). 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.

## Supplementary Material

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

## Acknowledgments

The authors thank Avi Flamholz for his comments on the manuscript. This work was funded by the Royal Society and the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 637765). J.W.B., J.R.N., J.S.B., A.E., A.B., and A.U. were funded by the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/M011224/1 and BB/P003117/1). T.R. and S.M.W. were supported by the Australian Government through the Australian Research Council Centre of Excellence for Translational Photosynthesis (CE140100015).

## Data Availability

All data used in this study are provided in the Supplementary Material online.

## Author Contributions

S.K. conceived the study. J.W.B. conducted the analysis. J.W.B., T.R., S.M.W., D.M.E., and S.K. analyzed the data. S.K. and J.W.B. wrote the manuscript. J.R.N., J.S.B., A.E., A.B., and A.U. 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. *Evol Ecol Res*. 1(8):895–909.

Ackerly DD. 2004. Adaptation, niche conservatism, and convergence: comparative studies of leaf evolution in the California chaparral. *Am Nat*. 163(5):654–671.

Aigner H, Wilson RH, Bracher A, Calisse L, Bhat JY, Hartl FU, Hayer-Hartl M. 2017. Plant RuBisCo assembly in *E. coli* with five chloroplast chaperones including BSD2'. *Science* 358(6368):1272–1278.

Andersson I, Backlund A. 2008. Structure and function of Rubisco. *Plant Physiol Biochem*. 46(3):275–291.

Andrews TJ. 1988. Catalysis by cyanobacterial ribulose-bisphosphate carboxylase large subunits in the complete absence of small subunits. *J Biol Chem*. 263(25):12213–12219.

Andrews TJ, Whitney SM. 2003. Manipulating ribulose bisphosphate carboxylase/oxygenase in the chloroplasts of higher plants. *Arch Biochem Biophys*. 414(2):159–169.

Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW. 1991. Climate change and the evolution of C4 photosynthesis. *Trends Ecol Evol*. 6(3):95–99.

APG. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Bot J Linnean Soc*. 181(1): 1–20.

Badger MR, Lorimer GH. 1976. Activation of ribulose-1,5-bisphosphate oxygenase. The role of Mg2+, CO2, and pH'. *Arch Biochem Biophys*. 175(2):723–729.

Banda DM, Pereira JH, Liu AK, Orr DJ, Hammel M, He C, Parry MAJ, Carmo-Silva E, Adams PD, Banfield JF, et al. 2020. Novel bacterial clade reveals origin of form I Rubisco'. *Nat Plants*. 6(9):1158–1166.

Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik DS, Milo R. 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters'. *Biochemistry* 50(21):4402–4410.

Bar-On YM, Milo R. 2019. The global mass and average rate of rubisco'. *Proc Natl Acad Sci U S A*. 116(10):4738–4743.

Bathellier C, Tcherkez G, Lorimer GH, Farquhar GD. 2018. Rubisco is not really so bad'. *Plant Cell Environ*. 41(4):705–716.

Bauwe H, Hagemann M, Kern R, Timm S. 2012. Photorespiration has a dual origin and manifold links to central metabolism'. *Curr Opin Plant Biol*. 15(3):269–275.

Beer C, Reichstein M, Tomelleri E, Ciais P, Jung M, Carvalhais N, Rödenbeck C, Arain MA, Baldocchi D, Bonan GB, et al. 2010. Terrestrial gross carbon dioxide uptake: global distribution and covariation with climate'. *Science* 329(5993):834–838.

Birky CW. 2001. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. *Annu Rev Genet*. 35(1):125–148.

Blomberg SP, Garland T, Ives AR. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57(4):717–745.

Bowes G, Ogren WL, Hageman RH. 1971. Phosphoglycolate production catalyzed by ribulose diphosphate carboxylase. *Biochem Biophys Res Commun*. 45(3):716–722.

Bradshaw A. 1991. Genostasis and the limits to evolution. *Philos Trans R Soc Lond B Biol Sci*. 333(1267):289–305.

Busch FA. 2020. Photorespiration in the context of rubisco biochemistry, $CO_2$ diffusion and metabolism. *Plant J*. 101(4):919–939.

Carlin DA, Caster RW, Wang X, Betzenderfer SA, Chen CX, Duong VM, Ryklansky CV, Alpekin A, Beaumont N, Kapoor H, et al. 2016. Kinetic characterization of 100 glycoside hydrolase mutants enables the discovery of structural features correlated with kinetic constants'. *PLoS One* 11(1):e0147596.

Carmo-Silva E, Scales JC, Madgwick PJ, Parry MAJ. 2015. Optimizing rubisco and its regulation for greater resource use efficiency. *Plant Cell Environ*. 1817–1832.

Cavender-Bares J, Ackerly DD, Baum DA, Bazzaz FA. 2004. Phylogenetic overdispersion in Floridian oak communities. *Am Nat*. 163(6):823–843.

Cheverud JM, Dow MM, Leutenegger W. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution* 39(6):1335–1351.

Cheverud JM, Dow MM, Leutenegger W. 1986. A phylogenetic autocorrelation analysis of sexual dimorphism in primates. *Am Anthropol*. 88(4):916–922.

Chollet R. 1977. The biochemistry of photorespiration. *Trends Biochem Sci*. 2(7):155–159.

Chollet R, Anderson LL. 1976. Regulation of ribulose 1,5-bisphosphate carboxylase-oxygenase activities by temperature pretreatment and chloroplast metabolites'. *Arch Biochem Biophys*. 176(1):344–351.

Cleton-Jansen AM, Dekker S, van de Putte P, Goosen N. 1991. A single amino acid substitution changes the substrate specificity of quinoprotein glucose dehydrogenase in *Gluconobacter oxydans*. *Mol General Genet*. 229(2):206–212.

Cleton-Jansen AM, et al. 1991. A single amino acid substitution changes the substrate specificity of quinoprotein glucose dehydrogenase in *Gluconobacter oxydans*. *Mol General Genet*. 229(2):206–212.

Cummins PL, Kannappan B, Gready JE. 2018. Directions for optimization of photosynthetic carbon fixation: rubisco's efficiency may not be so constrained after all. *Front Plant Sci*. 9:183.

Damborsky J, Brezovsky J. 2014. Computational tools for designing and engineering enzymes. *Curr Opin Chem Biol*. 19:8–16.

Davidi D, Longo LM, Jabłońska J, Milo R, Tawfik DS. 2018. A bird's-eye view of enzyme evolution: chemical, physicochemical, and physiological considerations. *Chem Rev*. 118(18):8786–8797.

Davidi D, Shamshoum M, Guo Z, Bar-On YM, Prywes N, Oz A, Jablonska J, Flamholz A, Wernick DG, Antonovsky N, et al. 2020. Highly active rubiscos discovered by systematic interrogation of natural sequence diversity. *EMBO J*. 39(18):e104081.

Duraõ P, et al. 2015. Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat Chem Biol*. 11(2):148–155.

Eckardt NA. 2005. Photorespiration revisited. *Plant Cell*. 17(8):2139–2141.

Edwards EJ. 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytol*. 223(4):1742–1755.

Edwards SV, Naeem S. 1993. The phylogenetic component of cooperative breeding in perching birds. *Am Nat*. 141(5):754–789.

Ehleringer JR, Sage RF, Flanagan LB, Pearcy RW. 1991. Climate change and the evolution of C4 photosynthesis. Trends Ecol Evol. 6(3):95–99.

Ellis RJ. 1979. The most abundant protein in the world. *Trends Biochem Sci*. 4(11):241–244.

Erb TJ, Zarzycki J. 2018. A short history of RubisCO: the rise and fall (?) of Nature's predominant CO2 fixing enzyme. *Curr Opin Biotechnol*. 49:100–107.

Espie GS, Kimber MS. 2011. Carboxysomes: cyanobacterial RubisCO comes in small packages. *Photosynth Res*. 109(1–3):7–20.

Farquhar GD. 1979. Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase. *Arch Biochem Biophys*. 193(2):456–468.

Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat*. 125(1):1–15.

Flamholz AI, Prywes N, Moran U, Davidi D, Bar-On YM, Oltrogge LM, Alves R, Savage D, Milo R. 2019. Revisiting trade-offs between rubisco kinetic parameters. *Biochemistry* 58(31):3365–3376.

Flamholz A, Shih PM. 2020. Cell biology of photosynthesis over geologic time. *Curr Biol*. 30(10):R490–R494.

Fukayama H, Kobara T, Shiomi K, Morita R, Sasayama D, Hatanaka T, Azuma T. 2019. Rubisco small subunits of C4 plants, Napier grass and guinea grass confer C4-like catalytic properties on Rubisco in rice. *Plant Prod Sci*. 22(2):296–300.

Garland T. 2001. Phylogenetic Comparison and Artificial Selection. Boston: Springer. p. 107–132.

Gielly L, Taberlet P. 1994. The use of chloroplast DNA to resolve plant phylogenies: noncoding versus rbcL sequences. *Mol Biol Evol*. 11(5):769–777.

Gittleman JL, Kot M. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst Zool*. 39(3):227–241.

Grafen A. 1989. The phylogenetic regression. *Philos Trans R Soc Lond B Biol Sci*. 326(1233):119–157.

Gunn LH, Martin Avila E, Birch R, Whitney SM. 2020. The dependency of red Rubisco on its cognate activase for enhancing plant photosynthesis and growth. *Proc Natl Acad Sci U S A*. 117(41):25890–25896.

Hardy OJ, Pavoine S. 2012. Assessing phylogenetic signal with measurement error: a comparison of mantel tests, Blomberg et al.'s K, and phylogenetic distograms. *Evolution* 66(8):2614–2621.

Hatch MD. 1987. C4 photosynthesis: a unique elend of modified biochemistry, anatomy and ultrastructure. *BBA Rev Bioenerget*. 895(2):81–106.

Huson DH, Scornavacca C. 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol*. 61(6):1061–1067.

Iñiguez C, Capó-Bauçà S, Niinemets Ü, Stoll H, Aguiló-Nicolau P, Galmés J. 2020. Evolutionary trends in RuBisCO kinetics and their co-evolution with CO2 concentrating mechanisms. *Plant J*. 101(4):897–918.

Ishikawa C, Hatanaka T, Misoo S, Miyake C, Fukayama H. 2011. Functional incorporation of sorghum small subunit increases the catalytic turnover rate of rubisco in transgenic rice. *Plant Physiol*. 156(3):1603–1611.

Ives A, Li D. 2018. rr2: an R package to calculate $R^2$s for regression models. *JOSS* 3(30):1028.

Ives AR. 2019. R 2 s for correlated data: phylogenetic models, LMMs, and GLMMs. *Syst Biol*. 68(2):234–251.

Johnson ET, Ryu S, Yi H, Shin B, Cheong H, Choi G. 2001. Alteration of a single amino acid changes the substrate specificity of dihydroflavonol 4-reductase. *Plant J*. 25(3):325–333.

Joshi J, Mueller-Cajar O, Tsai Y-CC, Hartl FU, Hayer-Hartl M. 2015. Role of small subunit in mediating assembly of red-type Form I Rubisco'. *J Biol Chem*. 290(2):1066–1074.

Kamilar JM, Cooper N. 2013. Phylogenetic signal in primate behaviour, ecology and life history. *Philos Trans R Soc B*. 368(1618):20120341.

Kane HJ, et al. 1994. An improved method for measuring the CO2/O2 specificity of ribulosebisphosphatte carboxylase-oxygenase. *Aust J Plant Physiol*. doi: 10.1071/PP9940449.

Kaplan A, Schwarz R, Lieman-Hurwitz J, Reinhold L. 1991. Physiological and molecular aspects of the inorganic carbon-concentrating mechanism in cyanobacteria. *Plant Physiol*. 97(3):851–855.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
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.
Kelly S. 2018. The amount of nitrogen used for photosynthesis modulates molecular evolution in plants. *Mol Biol Evol*. 35(7):1616–1625.
Kimura M. 1991. The neutral theory of molecular evolution: a review of recent evidence. *Jpn J Genet*. 66(4):367–386.
Larsson A. 2014. AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30(22):3276–3278.
Lee B, Berka RM, Tabita FR. 1991. Mutations in the small subunit of cyanobacterial ribulose-bisphosphate carboxylase/oxygenase that modulate interactions with large subunits. *J Biol Chem*. 266(12):7417–7422.
Lee B, Tabita FR. 1990. Purification of recombinant ribulose-l,5-bisphosphate carboxylase/oxygenase large subunits suitable for reconstitution and assembly of active L8S8 enzyme. *Biochemistry* 29(40):9352–9357.
Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, Gitzendanner MA, Graham SW, Grosse I, Li Z, Melkonian M, Mirarab S and Porsch M. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*. 574(7780):679–685.
Martin-Avila E, Lim Y-L, Birch R, Dirk LMA, Buck S, Rhodes T, Sharwood RE, Kapralov MV, Whitney SM. 2020. Modifying plant photosynthesis and growth via simultaneous chloroplast transformation of rubisco large and small subunits. *Plant Cell* 32(9):2898–2916.
Maynard Smith J, Burian R, Kauffman S, Alberch P, Campbell J, Goodwin B, Lande R, Raup D, Wolpert L. 1985. Developmental Constraints and Evolution: A Perspective from the Mountain Lake Conference on Development and Evolution. *Q Rev Biol*. 60(3):265–287.
Meyer M, Griffiths H. 2013. Origins and diversity of eukaryotic CO2-concentrating mechanisms: lessons for the future. *J Exp Bot*. 64(3):769–786.
Minshull J, Ness JE, Gustafsson C, Govindarajan S. 2005. Predicting enzyme function from protein sequence. *Curr Opin Chem Biol*. 9(2):202–209.
Moles AT, Ackerly DD, Webb CO, Tweddle JC, Dickie JB, Westoby M. 2005. A brief history of seed size. *Science* 307(5709):576–580.
Münkemüller T, Lavergne S, Bzeznik B, Dray S, Jombart T, Schiffers K, Thuiller W. 2012. How to measure and test phylogenetic signal. *Methods Ecol Evol*. 3(4):743–756.
Murashov VV, Morozova ZA. 2008. Comparative morphogenesis of *Triticum timopheevii* (Zhuk.) and synthetic octoploid species *T. timonovum* Heslot et Ferrary. Moscow Univ Biol Sci Bull. 63(3):127–133.
Nevo O, Valenta K, Kleiner A, Razafimandimby D, Jeffrey JAJ, Chapman CA, Ayasse M. 2020. The evolution of fruit scent: phylogenetic and developmental constraints. *BMC Evol Biol*. 20(1):138.
Nguyen L-T, 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.
Nisbet EG, Grassineau NV, Howe CJ, Abell PI, Regelous M, Nisbet RER. 2007. The age of rubisco: the evolution of oxygenic photosynthesis. *Geobiology* 5(4):311–335.
Ogren WL. 1984. Photorespiration: pathways, regulation, and modification. *Annu Rev Plant Physiol*. 35(1):415–442.
Ogren WL, Bowes G. 1971. Ribulose diphosphate carboxylase regulates soybean photorespiration. *Nat New Biol*. 230(13):159–160.
Orme D, et al. 2014. Caper: comparative analyses of phylogenetics and evolution in R, R package version 0.5.2/r121. Available from: http://r-forge.r-project.org/projects/ caper/ (accessed: May 31, 2020).
Orr DJ, Alcântara A, Kapralov MV, Andralojc PJ, Carmo-Silva E, Parry MAJ. 2016. Surveying rubisco diversity and temperature response to improve crop photosynthetic efficiency. *Plant Physiol*. 172(2):707–717.
Orr DJ, Pereira AM, da Fonseca Pereira P, Pereira-Lima ÍA, Zsögön A, Araújo WL. 2017. Engineering photosynthesis: progress and perspectives. *F1000Res*. 6:1891.
Pagel M. 1999. Inferring the historical patterns of biological evolution'. *Nature* 401(6756):877–884.
Pagel MD, Harvey PH. 1989. Comparative methods for examining adaptation depend on evolutionary models. *Folia Primatol*. 53(1-4):203–220.
Parry MAJ, Andralojc PJ, Scales JC, Salvucci ME, Carmo-Silva AE, Alonso H, Whitney SM. 2013. Rubisco activity and regulation as targets for crop improvement. *J Exp Bot*. 64(3):717–730.
Parry MAJ, Madgwick PJ, Carvalho JFC, Andralojc PJ.. 2007. Prospects for increasing photosynthesis by overcoming the limitations of Rubisco. *J Agric Sci*. 145:31–43.
Parry MAJ, Keys AJ, Gutteridge S. 1989. Variation in the specificity factor of C3 higher plant rubiscos determined by the total consumption of ribulose-P2. *J Exp Bot*. 40(3):317–320.
Peterhansel C, Horst I, Niessen M, Blume C, Kebeish R, Kürkcüoglu S, Kreuzaler F. 2010. 'Photorespiration'. *Arabidopsis Book*. *BioOne* 8(8):e0130.
Prins A, Orr DJ, Andralojc PJ, Reynolds MP, Carmo-Silva E, Parry MAJ. 2016. Rubisco catalytic properties of wild and domesticated relatives provide scope for improving wheat photosynthesis. *J Exp Bot*. 67(6):1827–1838.
Prinzing A, Durka W, Klotz S, Brandl R. 2001. The niche of higher plants: evidence for phylogenetic conservatism. *Proc Biol Sci*. 268(1483):2383–2389.
Read BA, Tabita FR. 1992a. A hybrid ribulosebisphosphate carboxylase/oxygenase enzyme exhibiting a substantial increase in substrate specificity factor. *Biochemistry* 31(24):5553–5560.
Read BA, Tabita FR. 1992b. Amino acid substitutions in the small subunit of ribulose-1, 5-bisphosphate carboxylase/oxygenase that influence catalytic activity of the holoenzyme. *Biochemistry* 31(2):519–525.
Rohlf FJ. 2001. Comparative methods for the analysis of continuous variables: geometric interpretations. *Evolution* 55(11):2143–2160.
Sage RF, Sage TL, Kocacinar F. 2012. Photorespiration and the evolution of c 4 photosynthesis. *Annu Rev Plant Biol*. 63(1):19–47.
Savir Y, Noor E, Milo R, Tlusty T. 2010. Cross-species analysis traces adaptation of Rubisco toward optimality in a low-dimensional landscape. *Proc Natl Acad Sci U S A*. 107(8):3475–3480.
Schneider G, Lindqvist Y, Brändén CI. 1992. RUBISCO: structure and mechanism. *Annu Rev Biophys Biomol Struct*. 21(1):119–143.
Seward EA, Kelly S. 2018. Selection-driven cost-efficiency optimization of transcripts modulates gene evolutionary rate in bacteria. *Genome Biol*. 19(1):102.
Sharkey TD. 2020. Emerging research in plant photosynthesis. *Emerg Top Life Sci*. 4(2):137–150.
Sharwood RE. 2017. Engineering chloroplasts to improve Rubisco catalysis: prospects for translating improvements into food and fiber crops. *New Phytol*. 213(2):494–510.
Sharwood RE, Ghannoum O, Whitney SM. 2016. Prospects for improving CO2 fixation in C3-crops through understanding C4-Rubisco biogenesis and catalytic diversity. *Curr Opin Plant Biol*. 31:135–142.
Siegel JB, Smith AL, Poust S, Wargacki AJ, Bar-Even A, Louw C, Shen BW, Eiben CB, Tran HM, Noor E, et al. 2015. Computational protein design enables a novel one-carbon assimilation pathway. *Proc Natl Acad Sci U S A*. 112(12):3704–3709.
Smith DR. 2015. Mutation rates in plastid genomes: they are lower than you might think. *Genome Biol Evol*. 7(5):1227–1234.
Spreitzer RJ, Peddi SR, Satagopan S. 2005. Phylogenetic engineering at an interface between large and small subunits imparts land-plant kinetic properties to algal Rubisco. *Proc Natl Acad Sci U S A*. 102(47):17225–17230.
Spriggs EL, Christin PA, Edwards EJ. 2014. C4 photosynthesis promoted species diversification during the iocene grassland expansion. *PLoS One* 9(5):e97722.

Studer RA, Christin P-A, Williams MA, Orengo CA. 2014. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. *Proc Natl Acad Sci U S A*. 111(6):2223–2228.

Su Z, Townsend JP. 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 Evol Biol*. 15:86.

Swenson NG, Enquist BJ. 2007. Ecological and evolutionary determinants of a key plant functional trait: wood density and its community-wide variation across latitude and elevation. *Am J Bot*. 94(3):451–459.

Tabita FR, Satagopan S, Hanson TE, Kreel NE, Scott SS. 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. *J Exp Bot*. 59(7):1515–1524.

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

Tcherkez GGB, Farquhar GD, Andrews TJ. 2006. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci U S A*. 103(19):7246–7251.

Tung Ho LS, Ané C. 2014. A linear-time algorithm for Gaussian and non-Gaussian trait evolution models. *Syst Biol*. 63(3):397–408.

Villar KD, Mitsuzawa H, Yang W, Sattler I, Tamanoi F. 1997. Amino acid substitutions that convert the protein substrate specificity of farnesyltransferase to that of geranylgeranyltransferase type I. *J Biol Chem*. 272(1):680–687.

Wagner PJ. 1995. Testing evolutionary constraint hypotheses with early paleozoic gastropods. *Paleobiology* 21(3):248–272.

Whitney SM, Houtz RL, Alonso H. 2011. Advancing our understanding and capacity to engineer nature's CO2-sequestering enzyme, Rubisco. *Plant Physiol*. 155(1):27–35.

Wilson RH, Martin-Avila E, Conlan C, Whitney SM. 2018. An improved *Escherichia coli* screen for Rubisco identifies a protein-protein interface that can enhance CO2-fixation kinetics. *J Biol Chem*. 293(1):18–27.

Wilson RH, Alonso H, Whitney SM. 2016. Evolving *Methanococcoides burtonii* archaeal Rubisco for improved photosynthesis and plant growth. *Sci Rep*. 6(1):22284.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*. 84(24):9054–9058.

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol*. 8(1):28–36.

Zhou Y, Whitney S. 2019. Directed evolution of an improved Rubisco; in vitro analyses to decipher fact from fiction. *Int J Mol Sci*. 20(20):5019.