

Surveying the C3-C4 landscape for positive selection in the Calvin Benson Cycle
enzyme SBPase

Teddy Gaskin

A thesis submitted for the degree of Master of Science (by Dissertation) in Biological
Sciences

School of Life sciences

University of Essex

Date of submission for examination: November 2023

Acknowledgments

I would first like to express my gratitude towards my supervisors Amanda Cavanagh, Christine Raines and Antonio Marco for their guidance throughout this project. I would also like to thank Kamel Chibani for his support and teaching during all my lab work and Chidi Afamefule for sharing vital data to this project. I also give thanks to the Essex plant group for giving me the opportunity to present my work.

Table of Contents

Abstract:	4
Introduction:	6
Methods:	22
Discussion:	35
References:	42
Appendix:	54

Abstract:

The Calvin-Benson cycle is the primary biochemical pathway for fixation of atmospheric carbon dioxide (CO₂) in over 85% of plants, including most major crop species. The cycle has three main phases: the initial carboxylation of RuBP by Rubisco, the reductive phase, and the regeneration of RuBP. As atmospheric concentrations of CO₂ increase, it is expected that control over carbon flux in the Calvin-Benson cycle will shift towards the rate of RuBP regeneration, making this a key target for future-proofing improvements to photosynthetic efficiency.

The enzyme SBPase plays a crucial role within the regeneration phase of the cycle, and previous studies have shown that overexpression and increased activity of SBPase translates to yield gains in many plants including major crop species. However, the natural variation of this enzyme has been relatively unexplored. If an increase in CO₂ concentrations were to impose selection pressure to improve the rate of RuBP regeneration, then variations in SBPase activity may be observed between C₄ species, with their carbon-concentrating mechanism, and their C₃ ancestors. The *Flaveria* (*Asteraceae*) genus includes C₃, C₃₋₄ intermediates and C₄ species which provides a convenient model to assess selection pressure along a C₃₋₄ gradient.

The purpose of this study is to determine whether there is positive selection pressure across a range of photosynthetic organisms but particularly the *Flaveria* genus to search for adaptive evolution in the transition from C₃ to C₄ photosynthesis. Using a bioinformatics approach, the ratio of synonymous/non-synonymous residues has been assessed in 116 SBPase coding sequences to

search for instances of positive selection. These residues may confer a higher performance of SBPase, therefore providing targets for precise transgenic manipulation of SBPase. The characteristics of SBPase enzymes in the *Flaveria* genus should also be assessed to determine if there is a link between variants under selection pressure and enzyme activity.

Introduction:

Through a combination of increasing world population, biofuel consumption and further stress under climate change, we face a rising demand for more food and fuel to the point that crop projections have predicted that global crop demands will have doubled from 2005 to 2050 (Tilman et al., 2011). Expanding the area of agricultural land is not a viable option to increase overall yield as much of the land that could be converted into crop fields, such as wetlands and forests, are ecosystems vital to maintaining biodiversity and providing environmental services (RSOL, 2009). The removal of these environments would also release stores of greenhouse gasses from the existing carbon sinks and oxidise carbon in the soil, in turn, further exacerbating climate change as one of the key stresses on our crop production (Godfray, 2014). Therefore, to meet these demands we must consider the factors that affect crop yields. The genetic yield potential of crops can be described as

$$Y = 0.487 \cdot St \cdot \epsilon_i \cdot \epsilon_c \cdot \epsilon_p$$

where Y is the yield potential and St is the total solar radiation incident on the earth's surface during a growing season (Zhu et al., 2010). Less than half of the spectrum of radiation is photosynthetically active (400-700nm), which is why St is multiplied by 0.487 to limit the yield potential to this part of the spectrum. This leaves three factors that limit the yield potential of crops: the light interception efficiency of crop plant leaves (ϵ_i), the conversion efficiency of intercepted radiative energy to biomass (ϵ_c), and the partitioning efficiency or harvest index (ϵ_p) which is the amount of biomass that can be harvested from the crop. Over the past century, great advances to the yield potential of crop plants have been realized through improvements to the ϵ_i and

ϵ_p . The ϵ_p of crops has benefited from selective breeding methods, a major factor of this being the introduction of high-yielding, semi-dwarf varieties of wheat and rice; the dwarfing of the stem has given a greater partitioning ratio to the grain of the plants and prevents lodging where other varieties would bend at the stem due to the increased weight of the grain, the shorter stem of dwarf varieties prevents this and inadvertently improves the ϵ_i throughout harsher weather conditions (Hedden, 2003). Otherwise, optimally spaced, larger-leafed crop variants that have been developed can intercept the majority of the available light during a growing season (Zhu et al., 2010).

However, these yield improvements have largely plateaued whilst demands have continued to increase. According to global yield trends from between 1961 and 2008, major crop yields are increasing yearly at a rate ($< 2\%$) far below what is required to meet predicted 2050 demands (Ray et al., 2013). To explore an untapped area for yield improvement, research efforts have turned to enhancing ϵ_c by improving the efficiency of photosynthesis via transgenic manipulation and modern breeding techniques to produce crops with greater yield. Unlike other components of yield potential, the photosynthetic conversion efficiency of sunlight to biomass is theoretically only at 4.3% and 6% for C3 and C4 plants respectively, (though actual field observations were only half of those values), indicating that there is significant room for improvement (Zhu et al., 2010).

The Calvin-Benson (CB) cycle is the primary biochemical pathway for the fixation of atmospheric CO_2 in over 85% of plants (Raghavendra, 2003), which includes our major crop species. The cycle makes use of 11 enzymes that catalyse 13 reactions that are typically described as belonging to three phases, starting with carboxylation. In this phase, Ribulose-1,5-bisphosphate carboxylase/oxygenase

(Rubisco) catalyses the carboxylation of ribulose biphosphate (RuBP), forming 3-phosphoglycerate which then continues into the reduction phase consuming ATP and NADPH in two reactions to form triose phosphates: glyceraldehyde phosphate and dihydroxyacetone phosphate. Some of the triose phosphates are used in the biosynthesis of sucrose and starch, whilst others are used in the final regeneration phase of the cycle which converts the triose phosphates into RuBP, ready to accept CO₂ and begin the cycle again (Sharkey, 2019; Fig. 1).

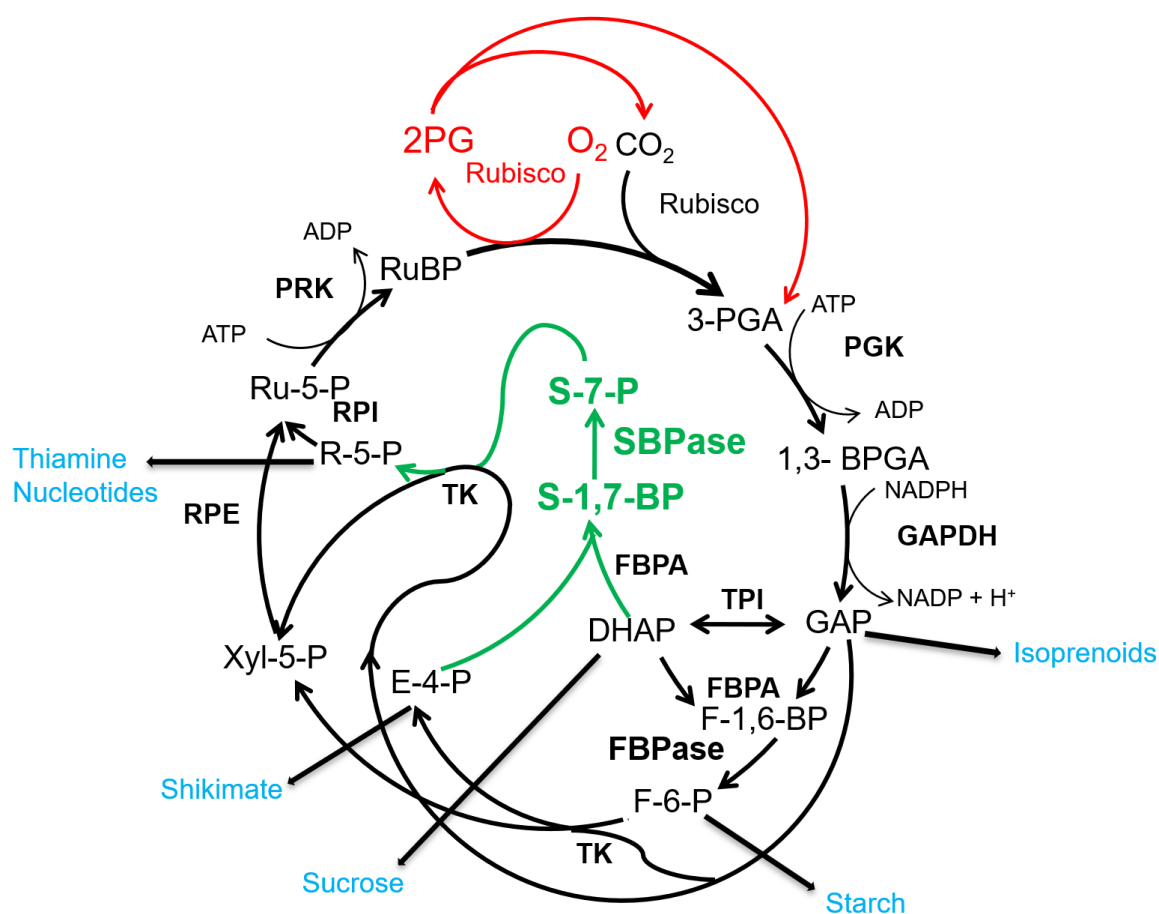


Figure 1. Schematic representation of the Calvin-Benson-Bassham cycle (Raines 2011). Rubisco and its reaction highlighted in red, SBPase and its reaction highlighted green.

Within this cycle, two biochemical processes are recognized to constrain net photosynthetic carbon assimilation (A): the fixation of carbon by Rubisco and the regeneration of RuBP (Farquhar 1979; Farquhar et al., 1980, Poolman et al., 2000). Rubisco is notorious for its inefficiency as a catalyst for carbon fixation due to its very slow catalytic rate and the competitive inhibition by O₂ (Parry et al., 2013). The oxygenation of RuBP produces one molecule of 3-PGA and one molecule of 2-phosphoglycolate (2-PG) which is recycled through the photorespiratory pathway. Due to the energy cost of photorespiratory reactions, the overall efficiency of A is lowered by ~40% (Walker et al., 2016). The degree of control exerted on A by Rubisco activity or RuBP regeneration varies with the concentration of CO₂ [CO₂] in the chloroplast; at ambient or lower [CO₂] carbon fixation by Rubisco is the primary factor limiting A, whereas at elevated [CO₂] it becomes RuBP regeneration where the pool of available RuBP molecules rapidly declines past 300 μbar of intercellular CO₂ partial pressure, requiring a faster rate of regeneration to further increase A (Fig. 2).

Following these limitations, past efforts to improve plant yield through photosynthesis have primarily focused on improving the carbon fixation rate and CO₂ specificity of Rubisco (Spreitzer, 2002). An interesting route for Rubisco improvement is to search for a “better” variation in nature by measuring the catalytic properties of Rubisco across different species, the structural analysis of these higher performing variants may reveal residues that translate to better enzyme performance with an amino acid substitution (Orr, et al. 2016). However, although Rubisco remains as an important target for photosynthetic improvement, it is necessary to consider other targets that may have more of an impact on crop yields in the context of a changing climate with rising atmospheric [CO₂].

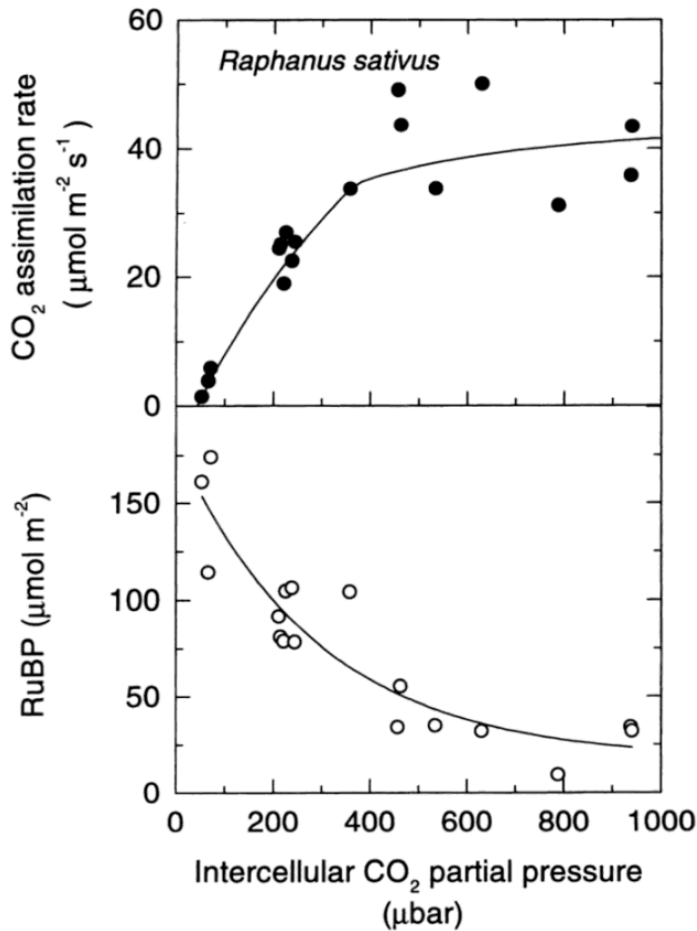


Figure 2. CO₂ assimilation rate and RuBP pool at varying intercellular CO₂ partial pressure measured in radish leaves with irradiance of 1400 μmol quanta m⁻²s⁻¹ and at a leaf temperature of 25 °C (von Caemmerer, 2000).

Through the use of antisense transgenic plants to lower the activity of certain enzymes and observe the effects on plant growth and photosynthetic activity, sedoheptulose-1,7-bisphosphatase (SBPase) and transketolase have been identified as enzymes that limit the rate of carbon flux in the CB cycle with only small decreases in activity (20-30%) (Harrison et al., 1997; Henkes et al., 2001). This is in contrast with other enzymes of the CB cycle such as fructose-1,6-bisphosphatase (FBPase) and aldolase that required a greater reduction in activity (>50%) before photosynthetic capacity was significantly affected (Kossmann et al., 1994; Haake et al., 1998). Therefore, alternative enzymes have been identified providing new targets

to explore for improving photosynthetic efficiency, particularly SBPase and its effect on the rate of RuBP regeneration.

SBPase

SBPase is the key enzyme of the CB cycle regulating RuBP regeneration. It functions in the regenerative phase where it catalyses the dephosphorylation of sedoheptulose-1,7-bisphosphate to sedoheptulose-7-phosphate alongside the dephosphorylation of fructose-1,6-bisphosphate by FBPase (Fig. 1). Originally, the two bisphosphatase reactions were thought to be catalysed by FBPase alone (Buchanan, et al. 1976). While this is true for some species of bacteria that have FBP enzymes that are known to catalyse both reactions (Gerbling et al., 1986), SBPase was eventually detected in spinach (Breazeale et al., 1978) then later isolated and purified in maize (Nishizawa and Buchanan, 1981), identifying it as an enzyme with a distinct function to its counterpart FBPase. The purified structure reveals SBPase as a homodimer with a molecular weight of 38 kDa that is immunologically distinct from FBPase, as shown by an immunodiffusion assay using polyclonal antibodies raised against FBPase. The poor recognition between SBPase and the antibodies signifies that the SBPase and FBPase structures are not related (Cadet and Meunier, 1988). Each dimer of SBPase likely folds the same way as FBPase, which is composed of a series of alpha helices and beta sheets connected by loops to fold into a hexahedron (Raines, et al. 1999; Villeret, et al. 1995). However, where SBPase forms a homodimer with the active site interfaced between the two subunits (Fig. 3), FBPase is a tetramer that requires an arginine residue from a neighbouring subunit to form the active sites (Zhang et al., 1994). Whilst there is extensive structural homology between the two enzymes, they likely have different evolutionary histories, as suggested by recent gene sequence analysis. FBPase is

derived from bacteria during an endosymbiotic event and SBPase from an archaeal gene already present in the host cell, which is an interesting scenario given their structural homology and identical regulatory systems they have evolved (Güttele et al., 2016).

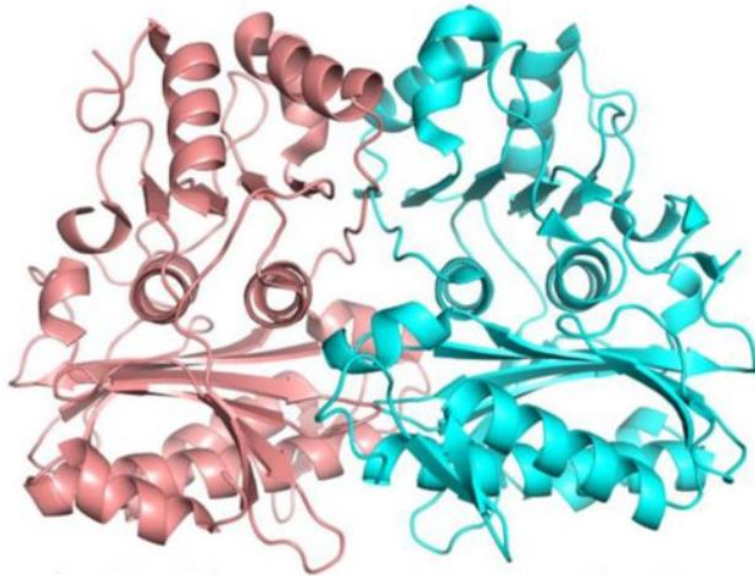


Figure 3. Crystal structure of SBPase homodimer from *Chlamydomonas reinhardtii* (Le Moigne, et al. 2022).

Both bisphosphatase enzymes are modulated by a ferredoxin/thioredoxin system, in which thioredoxin (TRX) is reduced by the enzyme ferredoxin-thioredoxin reductase with electrons donated by ferredoxin, TRX in turn reduces the disulfide bonds between cysteine residues 52 and 57 in SBPase to activate the enzyme (Güttele et al., 2016; Raines et al., 1999). The regulation of SBPase and its function as a branch point between the carboxylation and the regeneration of RuBP indicates its control of carbon flux in the CB cycle (Portis et al., 1977; Woodrow and Berry, 1988). This was further confirmed by the negative impacts in growth and CO₂ assimilation observed in antisense SBPase transgenic plants (Harrison et al., 1998).

SBPase as a target for improving photosynthesis

The deleterious effects of antisense repression of SBPase on plant growth prompted subsequent studies to explore the potential of manipulating SBPase activity to enhance plant productivity and yield. One of the earliest examples drove overexpression of a cyanobacterial FBPase/SBPase resulting in increased activity of these enzymes in transgenic tobacco plants (Miyagawa et al., 2001). With a 130% and 70% increase in overall SBP/FBPase activity, CO₂ fixation increased by 20% translating to a 50% increase in plant biomass. The RuBP content in the chloroplasts of transgenic plants was 180% larger than controls, suggesting a larger pool size of RuBP available for CO₂ fixation as a result of an increased rate of RuBP regeneration. This study demonstrates that non-plant enzymes have the potential to improve photosynthetic efficiency and the effectiveness of increased enzymatic activity in the CB cycle.

Further studies have focused on the overexpression of plant SBPase, typically using endogenous plant enzymes coupled with a constitutive promoter to drive overexpression. In particular, transgenic tobacco plants with up to a 150% increase in SBPase activity had an observed increase in A for young expanding leaves which translated to an accumulation of sucrose and starch products and an overall 30% increase in biomass (Lefebvre et al., 2005). Similar results have been observed across a number of species, including *Arabidopsis thaliana* (*Arabidopsis*) (Simkin et al., 2017), tomato (Ding et al., 2016) and more importantly, wheat (Driever et al., 2017). In transgenic wheat, the grain yield increased by 30-40% after cultivation under two types of growth regimens: one at a high plant density to reduce the chance of side shoot growth (tillering) and the other at low plant density to assess whether the effects of increased SBPase activity are consistent with potential field

conditions. Under both conditions higher A was observed, and alongside a biomass increase, there was a notable 30-40% increase in seed weight recorded. The extra weight came from a higher number of seeds formed per ear for high density plants or from a larger number of ears per plant where there are more instances of tillering at lower plant density, which displays the positive effect of SBPase activity regardless of density.

As $[\text{CO}_2]$ rises, theoretical models predict that the limitation of A will shift from Rubisco to RuBP regeneration (Fig. 4; Long et al., 2004). This is supported in work with transgenic plants overexpressing SBPase, where fully expanded tobacco leaves displaying a linear relationship between A and SBPase activity at saturating $[\text{CO}_2]$ where RuBP regeneration is typically limiting (Lefebvre et al., 2005). These results provided proof of concept that was later tested in replicated field conditions at elevated $[\text{CO}_2]$ (585ppm), achieving 14% increased light saturated A compared to WT controls resulting in a 12% increase in biomass (Rosenthal et al., 2011). Importantly, transgenic plants overexpressing SBPase displayed a greater growth benefit under high $[\text{CO}_2]$ conditions than WT controls (22% vs 13%).

These results show that there is a consistent positive effect from increased SBPase activity across a range of plant species and under future predicted atmospheric $[\text{CO}_2]$ where the role of SBPase becomes increasingly important as the control of carbon flux shifts towards RuBP regeneration. What is yet to be explored is the potential of natural variation in SBPase, using a similar approach taken with Rubisco (Orr et al., 2016).

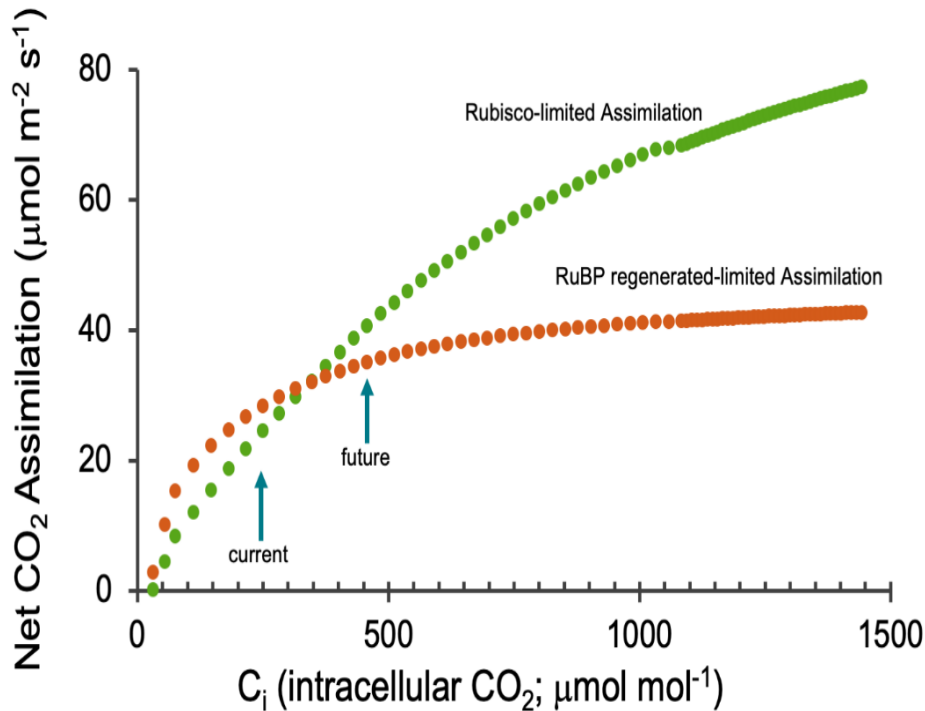


Figure 4. Modelled response of net CO₂ assimilation as a function of increasing [CO₂], illustrating regions limited by Rubisco (green dots) and RuBP regeneration (orange). Current and projected end of century C_i are labelled. Assimilation is modelled following von Caemmerer (2000), assuming a V_{cmax} of 137 µmol m⁻²s⁻¹, R_d of 2.1 µmol m⁻²s⁻¹ and J_{max} of 175 µmol m⁻²s⁻¹.

Neutral theory and adaptive evolution

Although the enzymes that catalyse the reactions of the CB pathway have been well characterised, with the exception of Rubisco, the degree of natural variation of each enzyme regarding their kinetic parameters and regulation is unknown as well as whether or not many of these enzymes undergo Darwinian selection (Raines, 2003). To demonstrate that any of these enzymes have evolved under selection pressure the null hypothesis must be rejected. In this case, it is the neutral theory of evolution, which suggests that the majority of evolutionary changes

at the molecular level are a result of neutral mutations that have become fixed at random rather than the process of positive selection acting upon advantageous mutations (Kimura, 1991). Neutral theory is based upon random chance or events that change the frequency of alleles in a population (genetic drift), whereas adaptive evolution is based upon directional changes in allele frequencies where mutations that generate new alleles that confer higher fitness levels for the organism increase in frequency within a population eventually reaching fixation (positive selection) or alternatively mutations that decrease the overall fitness of the organisms decrease in frequency in the population until the allele is 'purified' (negative selection) (Duret, 2008). The point of contention between these two theories is which of them accounts for the greater number of mutations that contribute to sequence divergence between species. Both theories recognise the existence of adaptive/non-mutations. The advantage neutral theory brings is its basis on simple assumptions that can be applied to mathematical theories that treat molecular variations quantitatively and allow for more complex theories such as positive selection to be tested against (Kimura, 1991).

Testing for positive selection

Evidence of adaptive evolution has already been shown in the enzyme Rubisco through phylogenetic analysis of sequences from a range of photosynthetic organisms (Kapralov & Filatov, 2007). This analysis is based upon the ratio of the rate of non-synonymous (i.e. base substitutions that result in a change to the amino acid produced in the sequence) to synonymous substitutions found throughout the sequence ($\omega = dN/dS$), this ratio is used as non-synonymous substitutions can be subject to positive selection whilst synonymous substitutions are mostly neutral; when ω is equal to 1 it indicates neutral mutations, $\omega < 1$ purifying/negative selection

and $\omega > 1$ positive selection which stems from the idea that advantageous mutations will be fixed at a higher rate than synonymous mutations.

Analysis of this ratio to test for positive selection can be achieved through likelihood ratio tests that assess how well the data fits to an evolutionary model compared to the null model of neutral evolution. Each evolutionary model will allow some variance in ω across the phylogeny, for example, a branch model could allow ω to vary on one specific branch or branches branch or branches of a phylogenetic tree (Yang, 1998) or a site model with ω variation across all codons (Suzuki and Gojobori, 1999). In the past, pairwise comparison was a common approach taken by averaging the rates of substitution across all codons in the gene between two sequences over a time period until sequence divergence, i.e. there is a single ω for all lineages and codons. However, this is not a very powerful method of testing as positive selection occurs typically over a few sites in a short time period so it is unlikely for the average dN to ever be higher than dS, resulting in very weak detection (Yang and Nielsen, 2002). More recently, a branch-site model has been developed that allows ω to vary across specific codons on a labelled branch (Yang and Nielsen, 2002). Allowing ω to differ over both lineages and specific sites improves the sensitivity for detection of positive selection; however, this particular method carries many assumptions that can generate false positives for positive selection when violated, so a large sample size of sequences or perfect sequence alignment would be required for effective testing (Zhang, 2004).

Model system for study

In the context of a future atmosphere in which CO₂ concentrations are increasing, it has been predicted through *in silico* evolutionary modelling that limitations to photosynthesis by SBPase activity increase during plant growth at higher CO₂ concentrations (Zhu et al., 2007). This is in line with the role of SBPase as under saturating CO₂ concentrations the control of CO₂ assimilation is likely to shift towards the rate of RuBP regeneration. C₄ and CAM photosynthetic organisms both have cellular mechanisms in place to separate initial carbon fixation and Rubisco reactions in order to maintain high CO₂ concentrations at the site of the CB cycle and limit photorespiration. CAM plants separate the carbon capture and CB cycle processes temporally: at night, the stomata open to take in and fix CO₂ with PEP carboxylase, which is then stored in the vacuole as malate or aspartate, then during the day the stomata close to prevent further gas exchange whilst decarboxylases release the CO₂ to be used in the CB cycle resulting in carbon concentration during the daytime (Bräutigam et al., 2017). Whereas the mechanism for carbon capture and concentration in C₄ plants is their leaf anatomy, at a glance, it consists of bundle sheath cells surrounded by mesophyll cells, CO₂ is initially fixed in the mesophyll cells by PEP carboxylase as oxaloacetic acid, which is converted to a 4 carbon acid (malate or aspartate) for storage, this intermediate then diffuses from the mesophyll cells to the bundle sheath cells to be broken down back into CO₂ by decarboxylases where it can then interact with Rubisco to enter the CB cycle (Fig. 5).

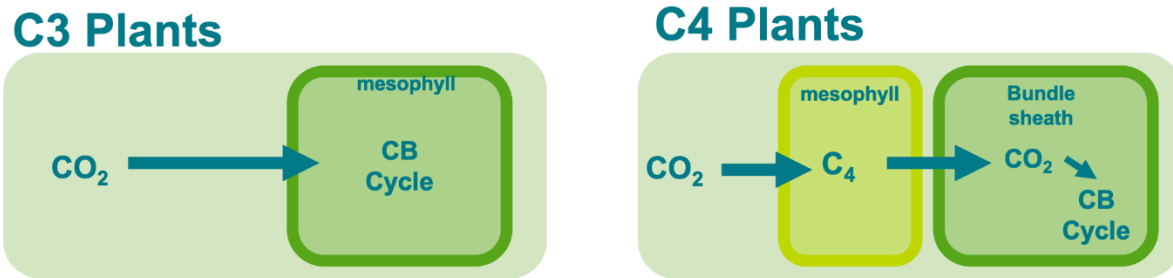


Figure 5. Schematic comparison between C3 and C4 photosynthesis.

It is thought that the evolution of the C4 pathway is a phenomenon of convergent evolution and is suggested to have evolved independently across 19 families of angiosperms over at least 70 times, likely to compensate for lower atmospheric $[\text{CO}_2]$ in the past and the accompanying higher rates of photorespiration (Sage, 2003; Sage, et al. 2012; Sage 2017). The evolution of the C4 pathway was in incremental steps, starting with anatomical adjustments to resemble what is known now as the Kranz anatomy, leading into integration of defining aspects of C4 photosynthesis such as enhanced activity of PEP carboxylase in the mesophyll cells. Each step of this evolution had to be stable on its own to not jeopardize the fitness of the plant, which is why there are C3 to C4 intermediates that resemble these steps in some plant families (Sage, 2003).

Flaveria (Asteraceae) species offer a unique gradient of C3 to C4 types including C3-4 intermediates. This gradient will provide a model system of the evolution from C3 to C4 photosynthesis for this study, which will identify the prevalence of positive selection amongst a phylogeny of C3/C4 and CAM organisms with a focus on *Flaveria*. Specifically, this work will identify where in the structure of SBPase this selection occurs, and determine whether there is any natural variation in SBPase activity along the C3 - C4 gradient. Due to the nature of the assumptions made by the statistical models used to detect positive selection, it is important to

confirm any findings with a further lab study. In this case, investigation of variations in SBPase activity would ideally confirm the presence of positive selection in species identified during phylogenetic analysis with corresponding higher levels of SBPase activity (Kapralov et al., 2010; Kubien et al., 2008). In C4 species, the CBC operates within the high [CO₂] environment of the bundle sheath, where carbon fixation is limited by RuBP regeneration (Fig 5). Enhancing RuBP regeneration rates in C3 crops (via transgenic overexpression of SBPase) enhances the CO₂ fertilisation effect on growth and yield (Rosenthal et al., 2011), supporting the idea that C3 photosynthetic biochemistry is not unbalanced for future climates. However, recent research showed no impact from SBPase overexpression in the C4 plant *Setaria viridis*, suggesting that the biochemistry of RuBP regeneration may already be optimised to the high [CO₂]-environment of the bundle sheath (Ermakova et al., 2022). One mechanism to support this would be inherent kinetic differences in C4 SBPase, conferring a higher maximal activity (V_{max}) than C3 counterparts. Therefore, this work will directly test the hypothesis that C4 *Flaveria* SBPase has a greater V_{max} compared to their C3 counterparts.

Aims and Objectives

Analysis of an instance of SBPase under positive selection may reveal structure - function relationships of the enzyme, in particular, there could be properties that lend themselves to performing better at higher CO₂ concentrations, this information could then be used to inform future transgenic engineering efforts to improve the operation of the CB cycle in crop plants for better yields in a future climate. In the absence of detectable selection pressure, where random neutral mutations are the overwhelming majority, this would indicate that there have been no optimized changes to SBPase as a result of directed evolution. Considering that

genetic manipulation can drive increases in plant growth and photosynthetic activity, there is potential for directed evolution experiments to identify the structure - function relationship of this enzyme to reveal targets for direct manipulation. This study will aim to first identify any residues under positive selection pressure in SBPase and characterize the natural diversity present amongst a variety of land plants and algae. Then, using the *Flaveria* genus, this work will explicitly test for positive selection in C4 vs C3 species. Finally, the activity of SBPase variants in the *Flaveria* genus will be characterized *in vitro* to find evidence that supports or contradicts the previous findings.

Methods:

Sequence alignment and phylogenetic analysis

All SBPase coding sequences, besides the *Flaveria* species, were obtained from phytozome or the NCBI GenBank (Goodstein, et al. 2011; Benson, et al. 2012). *Flaveria* sequences were obtained from past RNAseq experiments (Lyu, et al. 2015) following two approaches to generate coding sequences (C. Afamefule, personal communication). In the first approach, for the majority of *Flaveria* species with paired-end reads available, a *de novo* transcriptome assembly was carried out with Trinity (Grabherr, et al. 2011), using the Arabidopsis SBPase sequence as an alignment reference. The transcriptome for each species was run through BLAST to identify and select sequences that most resemble the Arabidopsis sequence (Altschul, et al. 1990). In the second approach, for the species *F. cronquistii*, *F. sonorensis* and *F. vaginata* for which only single end reads were available, RNA-seq reads from each species' transcriptome were selected based on whether they aligned to the Arabidopsis SBPase in BLAT (Kent, 2002). These selected reads were then aligned in bowtie2 (Langmead and Salzberg, 2012) to make consensus sequences for each species. A full list of the sequences and their respective source is available in the appendices (Appendix table 1).

These coding sequences were aligned with MUSCLE (Edgar, 2004) using default parameters and edited in MEGA11 software v 11.0.13 (Tamura, et al. 2021). The transit peptide encoding regions of all *Flaveria* species, *Z. mays* and *A. thaliana* sequences were identified using the cleavage site prediction tool TargetP 2.0 (Armenteros, et al. 2019). This region was then removed from all sequences, as well

as the last two codons to prepare for analysis, the final alignment is available in the appendices (Appendix figure 2). A phylogenetic tree was generated from this alignment (Appendix figure 1), also in MEGA11, using a maximum likelihood approach with the following parameters: the Tamura-Nei model, all sites from the edited alignment and all three codon positions used with uniform rates and bootstrapped 1000 times. This process was repeated using a smaller alignment of the 17 available *Flaveria* species and 6 other species forming an outgroup (Fig. 7a).

Test for positive selection

To detect positive selection, phylogenetic analysis was carried out using the codeml program included in the PAML v.4.10.6 package (Yang, 1997). This program enabled the use of likelihood ratio tests (LRTs) to compare different evolutionary models and assess their suitability to the data as well as determine if there are occurrences of positive selection. Four different sets of nested evolutionary models were selected that allow the ratio of synonymous/nonsynonymous substitutions ($\omega = dN/dS$) to vary across lineages and codon positions; ω acts as an indicator for selection pressure, $\omega = 1$ means neutral evolution, $\omega > 1$ positive selection and $\omega < 1$ purifying/negative selection. The significance of each LRT was calculated by taking twice the difference in log-likelihood values for each model to be applied to the chi-squared distribution with degrees of freedom equal to the number of parameters within each model (Yang, et al. 2000). The first LRT compared a null model of no variance in ω across all branches to a model that allows omega to vary across labelled lineages, in this case the branch containing all *Flaveria* species and further repeats by labelling the C3/4/intermediate lineages within the *Flaveria* genus. The second and third LRTs used site-specific models that allow ω to vary across codon positions, M1a (Nearly neutral) vs M2a (positive selection) and M7 (no codons with

$\omega > 1$) vs M8 (positive selection). The last LRT used the branch-site model to test for selection amongst the subset of codons within the labelled branch when compared to the null, with no variance of ω across codons and lineages and $\omega = 1$. Any amino acids that are potentially under selection pressure were identified by the Bayes Empirical Bayes approach (BEB) (Yang, et al. 2005), implemented by PAML using the parameters set from the site models and branch-site models to calculate the probability that a codon has an $\omega > 1$.

Generation of recombinant Flaveria SBPase

To test the functional impact of sequence variation among the *Flaveria* SBPase, a library of 17 recombinant proteins were expressed *in vitro*. To facilitate this, plasmid constructs were designed containing the SBPase from the coding sequence for each *Flaveria* species, each coding sequence was edited by removing the transit peptide region and adding an alanine (GCT) after the start codon to increase expression (Biovona, et al. 2010). These sequences were flanked by the *Escherichia virus T7* (T7) promoter + lac operator, and an N-terminal his-tag coding sequence + T7 terminator. The construct was cloned into the expression vector (pET-15b) for expression in *Escherichia coli* (Nico21) (GenScript, Oxford, United Kingdom).

Cell transformation

Plasmid constructs were introduced into *E. coli* cells via heat shock at 42 °C for 60 seconds and were suspended in 800 μ L Luria broth (LB) at 37°C for 1.5 hrs. Following this incubation, cells were centrifuged at 11,000 x g for 2 minutes and resuspended in 500 μ L fresh LB. 30 μ L of this bacterial suspension was plated on LB/Agar supplemented with 0.05 mg/ μ L for further incubation over 12 hours at 37°C.

Three colonies were selected for each construct and streaked onto fresh plates for further use.

Production test

A toothpick sample was taken from each streak plate colony from a random selection of *Flaveria* constructs (*F. palmeri*, *F. pringlei*, *F. cronquistii* and *F. ramosissima*) and grown in 5 mL of LB with 0.05 mg/μl Amp in a shaker (180 rpm) at 37°C for 2 hours. Protein expression was induced in the bacteria with 5 μl of 100mM Isopropyl β- d-1-thiogalactopyranoside (IPTG) and placed back into the shaker for 4 hours. 200 μl was taken from each clone and centrifuged at 11,000 x g for 2 minutes. The resulting pellet was resuspended in 10 μl of SDS-PAGE loading buffer (60 mM Tris-Cl pH 6.8, 5% βmercaptoethanol, 2% SDS, 10% glycerol and 2% Bromophenol Blue) Each sample was then denatured at 95 °C for 10 minutes and protein samples separated on a 15% SDS-PAGE gel alongside Nico21 cells for a control.

Bulk protein expression

Using clone two from each streak plate, a small culture was started in 5 mL of LB and 5 μl Amp and incubated in a shaker (180 rpm) at 37°C for 8 hrs. 200 μl of this culture was moved into 200 mL of LB and 0.05 mg/μl Amp for further incubation shaking at 37°C overnight. 45 mL of the overnight culture was added to 900 mL of LB and 900 μl of Amp and grown for 2 hours shaking at 37°C. After 2 hours SBPase expression was induced with 900 μl of 100mM IPTG, and the culture was then grown for a further 4 hours. The culture was centrifuged (1300 x g) for 15 minutes and the subsequent pellet was resuspended in 24 mL of 50 mM pH 8.0 phosphate buffer (0.2M sodium phosphate, dibasic dihydrate and 0.2M sodium phosphate, monobasic monohydrate) and stored at -20°C until purification.

Protein purification

Frozen cells were defrosted and lysed in a sonicator using 2 rounds of 50 kHz for 2 minutes, and the lysate was centrifuged at 25,000 x g for 30 minutes at 4°C, and again for 15 minutes with the supernatant. The supernatant was stored at 4°C to be loaded into a nickel-charged (HisPur Ni-NTA Resin, Thermo Fisher) IMAC column (Econo-column, Bio-Rad) after equilibration with 150 mL washing buffer (50 mM pH 8 phosphate buffer, 300 mM NaCl, 25 mM Imidazole). Upon loading the protein, the column was then washed with 300 mL washing buffer then eluted with 45 mL of elution buffer (50 mM pH 8 phosphate buffer, 300 mM NaCl, 250 mM Imidazole). The elution product was concentrated using a vivaspin 20 centrifugal concentrator MWCO 30kDa (Cole-Parmer Instrument Company LTD), centrifuging at 10°C, 1000 x g for between 5-10 minutes each spin until the protein was concentrated from 45 mL to 3 mL. Dialysis was performed on the remaining 3 mL overnight at 4 °C in 50 mM pH 8 phosphate buffer, then repeated with fresh buffer to remove excess salt and imidazole. Concentrated proteins were stored at 4°C until quantification.

Protein analysis

A preliminary check to assess the amount of protein was made by comparing samples against the SBPase of *A. thaliana* and *Z. mays* purified from a similar protocol. Proteins were first centrifuged at 16000 x g for 5 minutes at 4 °C then analysed using spectrophotometry with phosphate buffer (pH 8.0) as a blank, looking for peaks at 280nm for the presence of proteins.

Results:

Phylogenetic analysis

To test for positive selection in the SBPase gene, 116 SBPase coding sequences of different phototrophs were analyzed with likelihood ratio tests. A nested branch model has been used to assess the difference in omega values between *Flaveria* and other species reveals a significantly higher likelihood for the alternative model ($2\Delta\ln L = 29.98405$, $df = 1$, $p < 0.001$), which suggests there is a difference in selection pressure on the *Flaveria* branch compared to the background branches, however, the difference in the ω values calculated for the foreground (*Flaveria* clade) ($\omega = 0.0498$) and background branches (All other branches outside of the *Flaveria* clade) ($\omega = 0.0573$) was very small and not > 1 . When labelling the sub-branches to determine the selection pressure on C3/4 specific *Flaveria* lineages, there was no statistical support for the alternative model to suggest these branches are under positive selection. This prompted the use of site models and branch-site models to search for selection amongst codons and subsets of codons within the *Flaveria* lineages.

For further detection, a comparison was made between the site-specific models M1a vs. M2a and M7 vs. M8 to assess selection pressure focusing on the codons within the SBPase coding sequence. There was significant support for the fit of the M8 model over the M7 model ($2\Delta\ln L = 21.27987$, $df = 2$, $p < 0.001$), however, there was no statistical significance for the LRT between models M1a and M2a and no selection was detected at any codon positions with Bayes empirical Bayes

analysis (BEB), which left the branch-site model to further analyse subsets of codons within the *Flaveria* branch due to the significance of the branch model LRT.

Evidence of relaxed purifying selection in the Flaveria SBPase

This analysis with the branch-site model showed no significant support for the alternative model and there was no evidence of sites evolving with explicit positive selection at the *Flaveria* branch and its sub-branches, however, under BEB analysis, three potential sites were identified with relaxed purifying selection relative to the other codons, in reference to the edited *A. thaliana* SBPase coding sequence: 30 G (probability $\omega > 1 = 0.578$), 63 M ($p \omega > 1 = 0.543$) and 308 T ($p \omega > 1 = 0.699$). Based on the alignment (Fig. 6), there are no differences observed between the residues found at these sites in the *Flaveria* SBPase coding sequences.

<i>A.thaliana</i>	V T K C E I G Q S L E E F L A Q A T P D K G L R T L L M C M G E A L R T I A F K	40
<i>Z.mays</i>	S A R C E I G D S L E E F L T K A T P D K N L I R L L I C M G E A M R T I A F K	
<i>F.palmeri</i>	V T R C E I G E S L E D F L S K A T P D K G L I R L L I S M S E A V R T I A F K	
<i>F.robusta</i>	V T R C E I G E S L E D F L S K A T P D K G L I R L L L S M S E A V R T I A F K	
<i>F.sonorensis</i>	V T R C E I G E S L E D F L S K A T P D K G L I R L L L S M S E A V R T I A F K	
<i>F.trinervia</i>	V T K C E I G E S L E D F L S K A T P D K G L I R L L I S M S E A V R T I A F K	
<i>A.thaliana</i>	V R T A S C G G T A C V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C	80
<i>Z.mays</i>	V R T A S C G G T A C V N S F G D E Q L A V D M L A N K L L F E A L E Y S H V C	
<i>F.palmeri</i>	V R T A S C G G T A C V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C	
<i>F.robusta</i>	V R T A S C G G T A C V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C	
<i>F.sonorensis</i>	V R T A S C G G T A C V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C	
<i>F.trinervia</i>	V R T A S C G G T A C V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C	
<i>A.thaliana</i>	I N L D D R T Q V A Y G S K N E I I R F E E T L Y G T S - R L K - N V P I G V T	326
<i>Z.mays</i>	N E L D E R T Q V A Y G S K N E I I R F E E T L Y G S S - R L A A S A T A T A R	
<i>F.palmeri</i>	V N L D D R T Q V A Y G S K N E I I R F E E M L Y G S S - R L K - N V P I G A S	
<i>F.robusta</i>	A N L D D R T Q V A Y G S K N E I V R F E E M L Y G S S - R L K - N V P I G V T	
<i>F.sonorensis</i>	V N L D D R T Q V A Y G S K N E I I R F E E M L Y G S S - R L K - G V P V G A S	
<i>F.trinervia</i>	V N L D D R T Q V A Y G S K N E I I R F E E M L Y G S S - R L K - N V P I G A S	

Figure 6. Multiple sequence alignment of SBPase translated protein sequences from selected species included in a larger alignment of 116 photosynthetic species. Residues highlighted in black were identified with relaxed purifying selection relative to other SBPase residues. Numbers indicate the last residue number of the aligned sequences.

Evidence of positive selection at residues of Flaveria SBPase

To further assess the selection pressure on the SBPase gene, focusing primarily within the *Flaveria* genus, a phylogenetic tree was made based on the *Flaveria* genus and a small outgroup of 6 species (Fig. 7a.) and phylogenetic analysis was repeated using this tree, using the C4 branch of the *Flaveria* for the branch and branch-site model foreground (Fig. 7a). Although there is no statistical support across the three models to suggest there is positive selection model would suit the data, BEB analysis during the site model tests detected three potential sites that may have been under positive selection pressure, in reference to the SBPase coding sequence of *A. thaliana*: 277 H ($p \omega > 1 = 0.998$; $\omega = 4.454 \pm 1.655$), 320 N ($p \omega > 1 = 0.991$; $\omega = 4.411 \pm 1.667$) and 326 T ($p \omega > 1 = 0.722$; $\omega = 3.279 \pm 2.144$). Furthermore, sites 277 H and 320 N have differences in residues between the *Flaveria* species as shown in the alignment (Fig. 7b.; Table 2).

Table 1. Summary of LRT results for both phylogenies, codon position numbers in reference to the aligned SBPase coding sequence of *A. thaliana* (NS = no significance)

Phylogeny	Model	Results
116 species	Branch (<i>Flaveria</i> clade)	$p < 0.001$, $\omega = 0.0498$
	Branch (C4 <i>Flaveria</i> branch)	NS
	Site (M1a vs M2a) + BEB	NS
	Site (M7 vs M8) + BEB	$p < 0.001$, no codons detected
	Branch-site + BEB	NS, 3 codons detected: 30G, 63M, 308T
<i>Flaveria</i> + outgroup	Branch (C4 <i>Flaveria</i> branch)	NS
	Site model + BEB	NS, 3 codons detected: 277H, 320N, 326T
	Branch-site + BEB	NS, no codons detected

Potential for biochemical changes due to differences between positively selected residues

At residue 277, 3 of the 17 *Flaveria* species have a Threonine whilst the others have a Histidine. A more even split is found at residue 320 where 10 species have an Asparagine and 7 have a Glycine (Table 2.). Residue differences at site 277 do not appear to be consistent with either C3 or C4 phenotype, however all C4-like and C4 species, excluding 1, adopt N at this position over G. The residue changes themselves may indicate some biochemical change but require further investigation.

Table 2. Residue differences in the SBPase gene of the *Flaveria* identified through BEB analysis.

Species	Photosynthesis	Residue	
		277	320
<i>F. cronquistii</i>	C3	T	G
<i>F. pringlei</i>	C3	H	G
<i>F. robusta</i>	C3	H	N
<i>F. angustifolia</i>	C3-4	H	G
<i>F. anomala</i>	C3-4	H	G
<i>F. chlorifolia</i>	C3-4	H	N
<i>F. floridana</i>	C3-4	H	N
<i>F. pubescens</i>	C3-4	H	N
<i>F. ramosissima</i>	C3-4	H	G
<i>F. sonorensis</i>	C3-4	T	G
<i>F. brownii</i>	C4 Like	H	N
<i>F. palmeri</i>	C4 Like	H	N
<i>F. vaginata</i>	C4 Like	H	N
<i>F. australasica</i>	C4	H	N
<i>F. bidentis</i>	C4	H	N
<i>F. kochiana</i>	C4	T	G
<i>F. trinervia</i>	C4	H	N

Residue number in reference to the edited SBPase coding sequence *A. thaliana* from the *Flaveria* focused alignment.

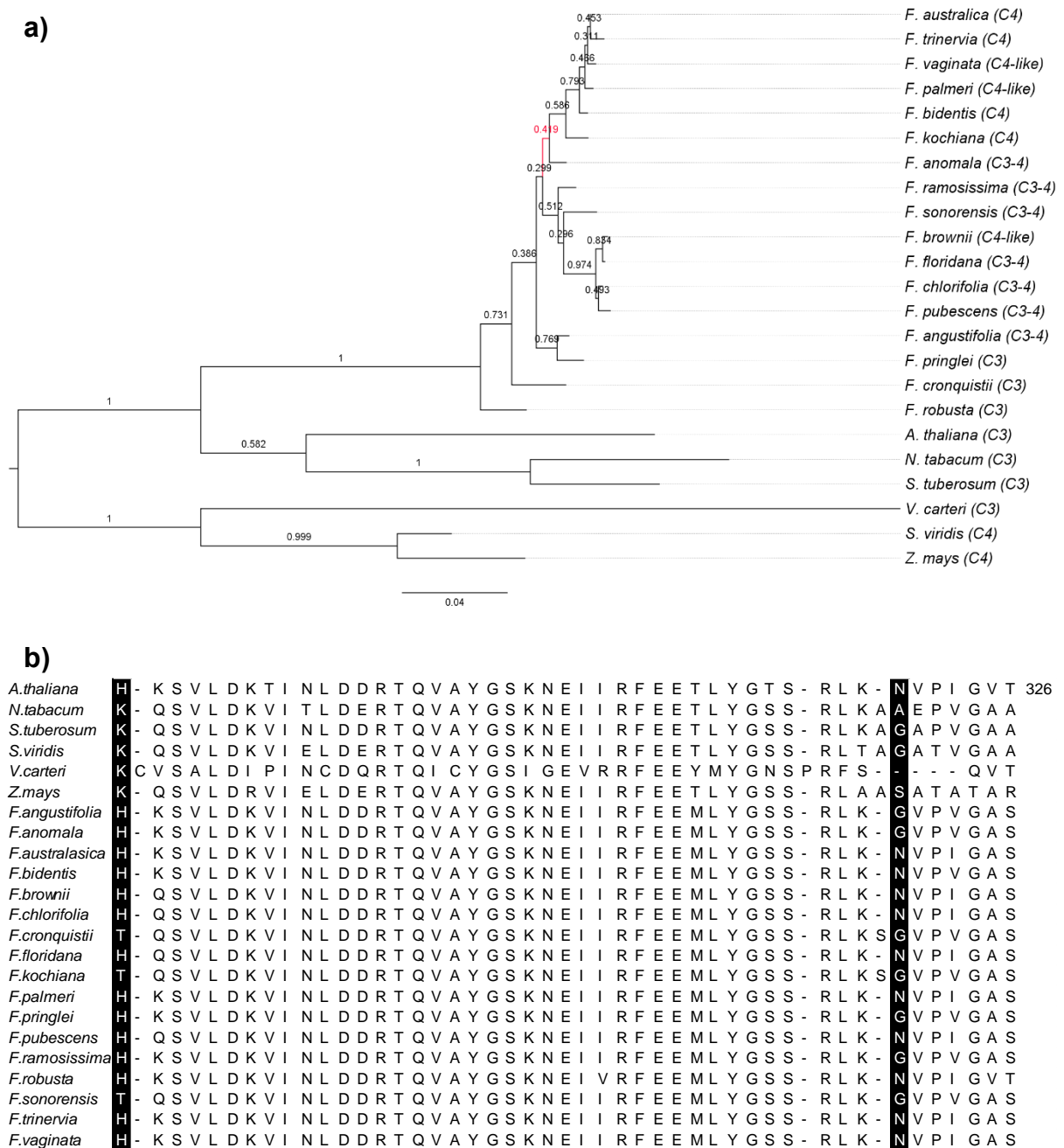


Figure 7. (a) Maximum likelihood phylogenetic tree based on SBPase coding sequences from the *Flaveria* genus and an outgroup of C3 and C4 photosynthetic species. Bootstrap values (/1000) labelled on branches and the foreground branch highlighted in red. **(b)** Multiple sequence alignment of SBPase translated protein sequences from the *Flaveria* genus, *A. thaliana*, *N. tabacum*, *S. tuberosum*, *S. viridis*, *V. carteri* and *Z. mays*. Residues highlighted in black were identified with BEB analysis during a site model test in PAML. Numbers indicate the last residue number of the aligned sequences.

Recombinant Flaveria SBPase expression

To characterise the activity of these SBPase variants, multiple constructs were made using pet15b as a base whilst inserting the SBPase coding sequence of each available *Flaveria* species with the transit peptide region removed and an alanine added after the start codon for greater expression. These constructs have been expressed in *E. coli* and the resulting proteins were purified, the presence of proteins was confirmed using spectroscopy by observing absorption at 280nm, in comparison to the absorption of SBPase proteins from *Z. mays* and *A. thaliana*, the absorption is significantly lower for 6 tested constructs: *F. trinervia*, *F. anomala*, *F. vaginata*, *F. bidentis*, *F. robusta* and *F. chlorifolia* (Fig. 8). As such, no conclusions can be drawn based on this concentration of protein.

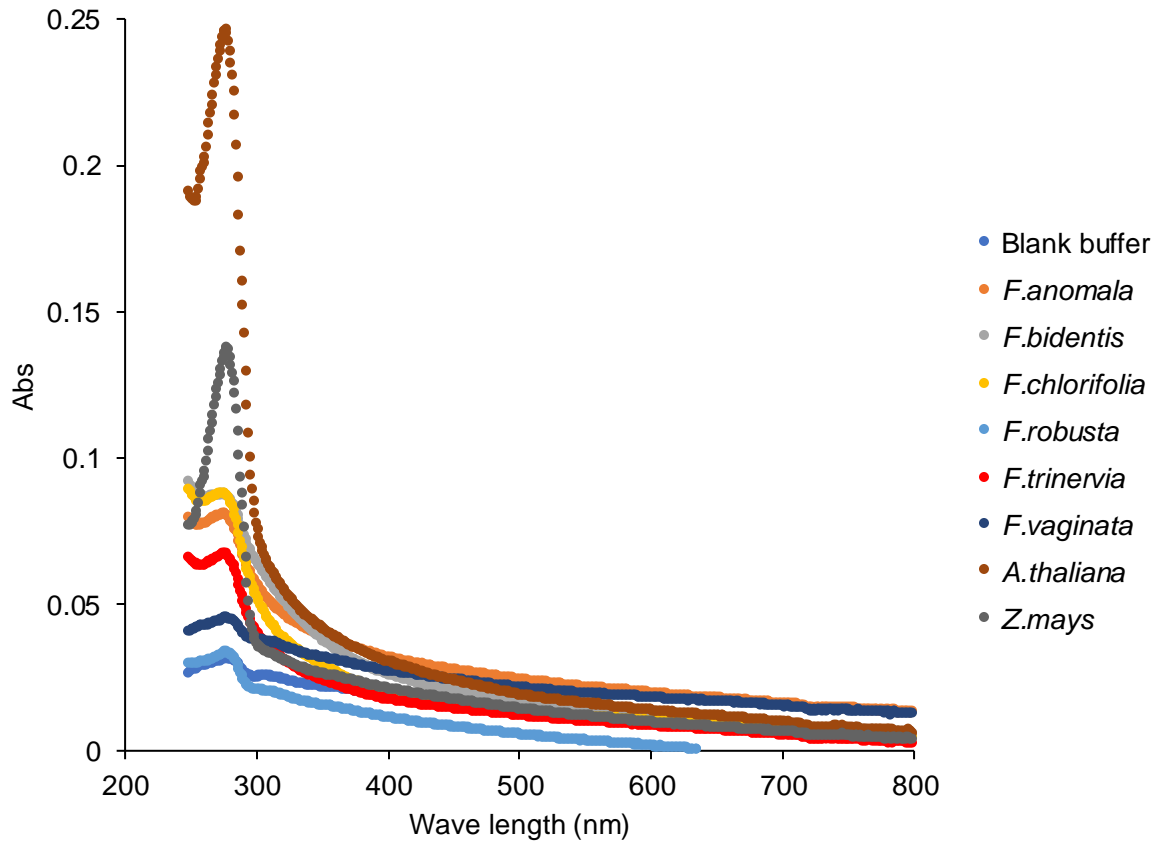


Figure 8. Light absorption between 250-800 nm for *Flaveria* SBPases diluted 10x with pH 8 phosphate buffer: *F. trinervia*, *F. anomala*, *F. vaginata*, *F. bidentis*, *F. robusta* and *F. chlorifolia*, compared to *Z. mays* and *A. thaliana* diluted 20x alongside a blank (pH 8 phosphate buffer).

Discussion:

The purpose of this study was to assess the degree of positive selection occurring within the CB cycle enzyme SBPase, identify any positively selected residues in the SBPase coding sequence and to determine whether there is any selection pressure on SBPase that is consistent with a C3 / C4 phenotype. This has been achieved through phylogenetic analysis of 116 photosynthetic species; generally, there is little evidence of positive selection throughout the SBPase gene. However, upon applying the site model and BEB analysis to a smaller tree two sites have been identified under positive selection pressure (Fig. 7), furthermore, there is some variation in the residues found at these sites between the *Flaveria* species. The residue differences at these sites may act as a filter for which species should be tested to determine a link between residue differences and SBPase activity.

The codon-based substitution (site) models used function under the basis that the non-synonymous / synonymous substitution rate ratio (ω) is an indicator of positive selection when ω is > 1 . Likelihood ratio tests are then used to compare the fit of a null model to the data which does not allow ω to be > 1 against a model that does. This particular analysis compared two sets of models, the first set: M1a (null) vs. M2a; M1a accounts for 2 site classes, $\omega = 1$ and $0 < \omega < 1$, whilst M2a includes an additional class of $\omega > 1$ (Nielsen and Yang, 1998). The second set compares M7 (null) vs. M8; the M7 model assumes a beta distribution across all sites whilst the alternative M8 model adds a category for positively selected sites (Yang, et al. 2000). Generally, if the null model can be rejected indicating positive selection across the gene, BEB analysis is then used to calculate the probability that any given

site belongs to a certain class; sites with a high probability ($p > 95\%$) of coming from a class with $\omega > 1$ are determined to be under positive selection (Yang, et al. 2005). In this case, although the null models could not be rejected based on the results of the LRTs, BEB analysis still identified some sites with a high probability of $\omega > 1$. Therefore, the advantage of this approach is that even if the average ω value across all sites is less than 1, as indicated by the LRTs, those few sites that may be under positive selection can be identified. Due to the smaller sample size in the phylogeny that was analysed to reveal these positively selected sites, there is a possibility of false positive detection. However, the introduction of the BEB over the naïve empirical bayes approach, which does not account for potential sampling errors, has reduced the rate of false positive detection, particularly in smaller data sets (Anisimova, et al. 2002; Yang, et al. 2005). For the purpose of identifying residues to be tested *in vitro* for their effect on SBPase activity, it is suitable.

The *Flaveria* genus was a key lineage included in the phylogeny used for analysis, it acted as a convenient model with a clear divide between C3 and C4 species as well as a branch of C3-4 intermediates, these three branches were labelled and tested under branch and branch-site models but showed no significant fit to a model that suggests that individual branches are under positive selection pressure. Similar research for the C3 cycle enzyme rubisco indicates that there are biochemical changes for the optimization of this enzyme under higher $[CO_2]$ brought about by the carbon concentrating mechanism of C4 photosynthesis. These changes are likely driven by positive selection unique to C4 *Flaveria* species, specifically on the large subunit of Rubisco (Kapralov, et al. 2010). In contrast, the residue changes suggested by this research on SBPase are found in all photosynthetic pathway phenotypes. Instead, these results would indicate that the SBPase of C4 *Flaveria*

species are not all optimized for a carbon concentrated environment, however, *in vitro* measurements of the performance for each enzyme variant would be required to confirm this. This is somewhat in line with the results of an SBPase sense study on *Setaria viridis* (Ermakova, et al. 2022), a C4 grass that showed no improvement to photosynthetic rates with greater SBPase expression. It was suggested that the spatial separation of starch and sucrose processing and the metabolites required to maintain a gradient to support the triose phosphate shuttle in C4 species (Furbank and Kelly, 2021) may lead to a difference in the distribution of control over RuBP regeneration between C4 cycle enzymes when compared to C3 species that hold all of these aspects in one cell type. Therefore, there may be less selection pressure on SBPase if the degree of control over RuBP regeneration is lessened in C4 plants. Alternatively, the lack of response to SBPase overexpression in *S. viridis* could be because this variant of SBPase is already optimized so further expression would have no effect on RuBP regeneration rate, but this is not supported by the lack of selection present in the SBPase gene. There is also the potential that the function of SBPase in C4 plants is optimized through the regulation of the enzyme rather than its individual performance. This could be achieved through greater expression of the enzyme or enhanced maintenance of the conditions required for enzyme operation such as a higher concentration of Mg^{2+} (Woodrow, et al. 1984). This is fitting in the context of an overall decrease in levels of SBP reported in C4 and C4-like species in the *Flaveria* genus (Borghi, et al. 2022) which could suggest that there is higher SBPase activity in the C4 species.

From the sites that were identified under positive selection, there are biochemical differences between these residues that could potentially have an impact on the activity of the enzyme. Particularly the swap between Histidine and

Threonine at residue 277; Histidine is known to be a key residue for use in the active site of enzymes. The imidazole group on the histidine side chain allows histidine to participate in multiple molecular interactions, for example, it can act as a coordinating ligand to a metallic cation, such as Mg^{2+} which is necessary for SBPase to become active, and the nitrogen and hydrogen atoms on the imidazole ring can fulfil proton acceptor and donator roles respectively to form hydrogen bonds (Liao, et al. 2013). Given that SBPase operates at pH 7-8 depending on light intensity, histidine will be partially protonated but mostly neutral, the swap to Threonine, which has a polar side chain and a net neutral charge, could have an effect on the enzyme's function due to the differences between these two amino acids and the loss of histidine's versatility. However, the involvement of this residue at the active site is unknown; it's possible that the SBPase active site is mostly homogenous with the FBPase active site (Ke, et al. 1990; Zhang, et al. 1993) but ultimately the purification and crystallization of SBPase is required to confirm this. At site 320 the swap between glycine and asparagine is found again in the majority of photosynthetic phenotypes, however all C4 and C4-like species apart from *F. kochiana* (C4) have an arginine residue at this site; the effect of glycine on the enzyme is likely limited, but asparagine is known to be capable of forming hydrogen bonds as an acceptor or donator with the amide group, particularly with protein backbones (Vasudev, et al. 2012), which may indicate some structural impact on the SBPase enzyme with or without asparagine. It is important to note that these inferences are purely speculative without having first tested the activity of these SBPase variants *in vitro* and also without knowing the position of these residues in the SBPase structure.

The range of interspecific variation in SBPase kinetic parameters remains under-investigated and poorly understood (Raines et al., 2022). The structure-functional relationships relating to enzyme regulation through the Trx system have driven the resolution of SBPase crystal structures, but the impact of other conserved amino acids on enzyme function remains unresolved. For example, a single amino acid change from Gly to Asp in rice SBPase at position 364 - far removed from the enzyme catalytic site or regulatory cysteines - severely impacts plant growth and yield, but the impact on SBPase kinetic performance remains uncharacterised (Li et al., 2020). Thus, it is crucial to test for associated kinetic changes in the residues identified to be under putative positive selection in the *Flaveria* genus (i.e. residues 277 or 320). However, there was a roadblock with the level of protein concentration obtained after purifying the SBPase proteins from recombinant expression. As a preliminary check, the protein concentrate from several *Flaveria* samples were compared to the *A. thaliana* and *Z. mays* SBPase purified using the same protocol. Spectroscopy was used to measure the absorption of light from these sample at 280 nm for a rudimentary comparison of the amount of protein present; the degree of absorption from all 6 *Flaveria* SBPase samples are significantly lower than *A. thaliana* and to a lesser extent *Z. mays* (Fig. 8). This difference is understated to a degree as the *A. thaliana* and *Z. mays* samples were diluted by a factor of 20 with pH 8 phosphate buffer whilst the *Flaveria* samples were only diluted by a factor of 10; for the purpose of activity testing, the amount of protein available in these samples as suggested by the spectral analysis is too small.

There are several explanations for why the degree of recombinant expression from these *Flaveria* constructs are smaller compared to other species: considering that this protocol has worked to purify SBPase proteins in the past, it is likely an

issue that stems from these particular variants of SBPase, perhaps the SBPase expressed from these species may be particularly insoluble which would have caused issues during the purification stage, the proteins may have precipitated and therefore not separated during dialysis from the contaminants introduced such as the high concentration of imidazole in the elution buffer. There are many other factors that could result in either lower expression or impeded purification; potential solutions could begin with ensuring the protocol is optimized to prevent protein precipitation at key steps, if this isn't the key issue there is the possibility of using an alternative expression vector such as a strain of yeast rather than *E. coli*, differences between the native environment and the expression vector can cause misfolding, however generally bacterial strains used for expression such as Nico21 have been designed to avoid these issues (Duong-ly and Gabelli, 2014).

Going forward, to carry out *in vitro* testing of SBPase activity, it may be simpler to isolate SBPase from leaves of grown *Flaveria* plants rather than troubleshoot the method of protein expression described in this study as there are multiple steps that could be the root cause of limited SBPase expression. The catalytic activity of these enzymes can be measured with spectrophotometry-based assays (Le Moigne, et al. 2022). Ideally measurements would be obtained for the maximal activity of SBPase (V_{max}), individual enzyme turnover (k_{cat}) and the midpoint redox potential to determine whether a change in residue at these sites affects the relationship between SBPase and activation of the TRX system (Güttele, et al. 2016). With *in vitro* measurements, the effect of these positively selected residues will be greater understood; depending on the findings, there may be some particular residues that confer greater performance of the SBPase enzyme which can be further explored in their application to other species' SBPase. With the limited

evidence of positive selection under site models, there is little incentive to explore positive selection in other species. It may be worthwhile to carry out phylogenetic analysis with other enzymes involved in RuBP regeneration to see if there is greater selection pressure on these enzymes particularly in C4 species.

References:

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, 215, 403-10.

Anisimova, M., Bielawski, J. P. and Yang, Z. (2002) Accuracy and Power of Bayes Prediction of Amino Acid Sites Under Positive Selection. *Molecular Biology and Evolution*, 19, 950-958.

Armenteros, J. J. A., Salvatore, M., Emanuelsson, O., Winther, O., Heijne, G. v., Elofsson, A. and Nielsen, H. (2019) Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, 2, e201900429.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W. (2013) GenBank. *Nucleic Acids Research*, 41, D36-42.

Bivona, L., Zou, Z., Stutzman, N., and Sun, P. D. (2010). Influence of the second amino acid on recombinant protein expression. *Protein expression and purification*, 74(2), 248–256.

Borghi, G. L., Arrivault, S., Günther, M., Barbosa Medeiros, D., Dell'Aversana, E., Fusco, G. M., Carillo, P., Ludwig, M., Fernie, A. R., Lunn, J. E. and Stitt, M. (2021) Metabolic profiles in C3, C3–C4 intermediate, C4-like, and C4 species in the genus *Flaveria*. *Journal of Experimental Botany*, 73, 1581-1601.

Buchanan, B. B., Schürmann, P. and Wolosiuk, R. A. (1976) Appearance of sedoheptulose 1,7-diphosphatase activity on conversion of chloroplast fructose 1,6-

diphosphatase from dimer form to monomer form. *Biochemical and Biophysical Research Communications*, 69, 970-978.

Bräutigam, A., Schlüter, U., Eisenhut, M. and Gowik, U. (2017) On the Evolutionary Origin of CAM Photosynthesis. *Plant Physiology*, 174, 473-477.

Breazeale, V. D., Buchanan, B. B. and Wolosiuk, R. A. (1978) Chloroplast Sedoheptulose 1,7-Bisphosphatase: Evidence for Regulation by the Ferredoxin/Thioredoxin System. *Zeitschrift für Naturforschung C*, 33, 521-528.

Cadet, F. and Meunier, J. C. (1988) Spinach (*Spinacia oleracea*) chloroplast sedoheptulose-1,7-bisphosphatase. Activation and deactivation, and immunological relationship to fructose-1,6-bisphosphatase. *Biochemical Journal*, 253, 243-8.

Ding, F., Wang, M., Zhang, S. and Ai, X. (2016) Changes in SBPase activity influence photosynthetic capacity, growth, and tolerance to chilling stress in transgenic tomato plants. *Scientific Reports*, 6, 32741.

Duret, L. (2008) Neutral theory: The null hypothesis of molecular evolution. *Nature Education* 1 218.

Driever, S. M., Simkin, A. J., Alotaibi, S., Fisk, S. J., Madgwick, P. J., Sparks, C. A., Jones, H. D., Lawson, T., Parry, M. A. J. and Raines, C. A. (2017) Increased SBPase activity improves photosynthesis and grain yield in wheat grown in greenhouse conditions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372.

Duong-Ly K. C. and Gabelli, S. B. (2014) Troubleshooting Recombinant Protein Expression: General. *Methods in Enzymology*, 541, 209-229.

Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792-7.

Ermakova, M., Lopez-Calcano, P. E., Furbank, R. T., Raines, C. A. and von Caemmerer, S. (2023) Increased sedoheptulose-1,7-bisphosphatase content in *Setaria viridis* does not affect C4 photosynthesis. *Plant Physiology*, 191, 885-893.

Farquhar, G. D. (1979) Models describing the kinetics of ribulose biphosphate carboxylase-oxygenase. *Archives of Biochemistry and Biophysics*, 193, 456-468.

Farquhar, G. D., von Caemmerer, S. and Berry, J. A. (1980) A biochemical model of photosynthetic CO₂ assimilation in leaves of C₃ species. *Planta*, 149, 78-90.

Furbank, R. T. and Kelly, S. (2021) Finding the C₄ sweet spot: cellular compartmentation of carbohydrate metabolism in C₄ photosynthesis. *Journal of Experimental Botany*, 72, 6018-6026.

Gerbling, K. P., Steup, M. and Latzko, E. (1986) Fructose 1,6-Bisphosphatase Form B from *Synechococcus leopoliensis* Hydrolyzes both Fructose and Sedoheptulose Bisphosphate. *Plant Physiology*, 80, 716-20.

Godfray, H. C. J. 2014. The challenge of feeding 9–10 billion people equitably and sustainably. *Journal of Agricultural Science* 152, S2–S8.

Güttele, D. D., Roret, T., Müller, S. J., Couturier, J., Lemaire, S. D., Hecker, A., Dhalleine, T., Buchanan, B. B., Reski, R., Einsle, O. and Jacquot, J. P. (2016) Chloroplast FBPase and SBPase are thioredoxin-linked enzymes with similar architecture but different evolutionary histories. *Proceedings of the National Academy of Sciences of the USA*, 113, 6779-84.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D. S. (2011) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40, D1178-D1186.

Haake, V., Zrenner, R., Sonnewald, U. and Stitt, M. (1998) A moderate decrease of plastid aldolase activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato plants. *The Plant Journal*, 14, 147-57.

Harrison, E. P., Willingham, N. M., Lloyd, J. C. and Raines, C. A. (1998) Reduced sedoheptulose-1,7-bisphosphatase levels in transgenic tobacco lead to decreased photosynthetic capacity and altered carbohydrate accumulation. *Planta*, 204, 27-36.

Hedden, P. (2003) The genes of the Green Revolution. *Trends in Genetics*, 19, 5-9.

Henkes, S., Sonnewald, U., Badur, R., Flachmann, R. and Stitt, M. (2001) A small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism. *Plant Cell*, 13, 535-51.

Kapralov, M. V. and Filatov, D. A. (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evolutionary Biology*, 7, 73.

Kapralov, M. V., Kubien, D. S., Andersson, I. and Filatov, D. A. (2010) Changes in Rubisco Kinetics during the Evolution of C4 Photosynthesis in *Flaveria* (Asteraceae) Are Associated with Positive Selection on Genes Encoding the Enzyme. *Molecular Biology and Evolution*, 28, 1491-1503.

Ke, H., Thorpe, C. M., Seaton, B. A., Lipscomb, W. N. and Marcus, F. (1990) Structure refinement of fructose-1,6-bisphosphatase and its fructose 2,6-

bisphosphate complex at 2.8 Å resolution. *Journal of Molecular Biology*, 212, 513-539.

Kent, W. J. (2002) BLAT—The BLAST-Like Alignment Tool. *Genome Research*, 12, 656-664.

Kimura, M. (1991) The neutral theory of molecular evolution: A review of recent evidence. *Japanese Journal of Genetics*, 66, 367–386.

Kossmann, J., Sonnewald, U. and Willmitzer, L. (1994) Reduction of the chloroplastic fructose-1,6-bisphosphatase in transgenic potato plants impairs photosynthesis and plant growth. *The Plant Journal*, 6, 637-650.

Kubien, D. S., Whitney, S. M., Moore, P. V. and Jesson, L. K. (2008) The biochemistry of Rubisco in *Flaveria*. *Journal of Experimental Botany*, 59, 1767-1777.

Langmead, B. and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357-359.

Lefebvre, S., Lawson, T., Zakhleniuk, O. V., Lloyd, J. C., Raines, C. A. and Fryer, M. (2005) Increased sedoheptulose-1,7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development. *Plant Physiology*, 138, 451-60.

Le Moigne, T., Santoni, M., Jomat, L., Lemaire, S. D., Zaffagnini, M., Chéron, N. and Henri, J. (2022) Structure of the photosynthetic Calvin-Benson-Bassham sedoheptulose-1,7-bisphosphatase SBPase from the model microalga *Chlamydomonas reinhardtii*. *bioRxiv*, 2022.10.28.514230.

Li, C., Li, N., Huang, R., Chen, C., Guo, J., Yang, X., Zhang, X., Sun, C., Deng, X., and Wang P. (2020). A single nucleotide substitution at the 3'-end of SBPase gene

involved in Calvin cycle severely affects plant growth and grain yield in rice. *BMC Plant Biol* 20, 345

Liao, S. M., Du, Q. S., Meng, J. Z., Pang, Z. W. and Huang, R. B. (2013) The multiple roles of histidine in protein interactions. *Chemistry Central Journal*, 7, 44.

Lyu, M. J., Gowik, U., Kelly, S., Covshoff, S., Mallmann, J., Westhoff, P., Hibberd, J. M., Stata, M., Sage, R. F., Lu, H., Wei, X., Wong, G. K. and Zhu, X. G. (2015) RNA-Seq based phylogeny recapitulates previous phylogeny of the genus *Flaveria* (Asteraceae) with some modifications. *BMC Evolutionary Biology*, 15, 116.

Miyagawa, Y., Tamoi, M. and Shigeoka, S. (2001) Overexpression of a cyanobacterial fructose-1,6-/sedoheptulose-1,7-bisphosphatase in tobacco enhances photosynthesis and growth. *Nature Biotechnology*, 19, 965-9.

Nielsen, R. and Yang, Z. (1998) Likelihood Models for Detecting Positively Selected Amino Acid Sites and Applications to the HIV-1 Envelope Gene. *Genetics*, 148, 929-936.

Nishizawa, A. N. and Buchanan, B. B. (1981) Enzyme regulation in C4 photosynthesis. Purification and properties of thioredoxin-linked fructose bisphosphatase and sedoheptulose bisphosphatase from corn leaves. *The Journal of Biological Chemistry*, 256, 6119-26.

Orr, D. J., Alcântara, A., Kapralov, M. V., Andralojc, P. J., Carmo-Silva, E. and Parry, M. A. J. (2016) Surveying Rubisco Diversity and Temperature Response to Improve Crop Photosynthetic Efficiency. *Plant Physiology*, 172, 707-717.

Parry, M. A. J., Andralojc, P. J., Scales, J. C., Salvucci, M. E., Carmo-Silva, A. E., Alonso, H. and Whitney, S. M. (2012) Rubisco activity and regulation as targets for crop improvement. *Journal of Experimental Botany*, 64, 717-730.

Paul, M., Pellny, T. and Goddijn, O. (2001) Enhancing photosynthesis with sugar signals. *Trends in Plant Science*, 6, 197-200.

Poolman, M. G., Fell, D. A. and Thomas, S. (2000) Modelling photosynthesis and its control. *Journal of Experimental Botany*, 51, 319-328.

Portis, A.R., Chon, C.J., Mosbach, A., Heldt, H.W., 1977. Fructose-and sedoheptulosebisphosphatase. The sites of a possible control of CO₂ fixation by light-dependent changes of the stromal Mg²⁺ concentration. *Biochimica et Biophysica Acta - Bioenergetics*. 461, 313–325.

Quick, W. P., Schurr, U., Scheibe, R., Schulze, E. D., Rodermeil, S. R., Bogorad, L. and Stitt, M. (1991) Decreased ribulose-1,5-bisphosphate carboxylase-oxygenase in transgenic tobacco transformed with "antisense" rbcS : I. Impact on photosynthesis in ambient growth conditions. *Planta*, 183, 542-54.

Raghavendra, A. S. (2003) PHOTOSYNTHESIS AND PARTITIONING | C₃ Plants. In: Thomas, B. (ed.) *Encyclopedia of Applied Plant Sciences*. Oxford: Elsevier.

Raines, C. A. (2003) The Calvin cycle revisited. *Photosynthesis Research*, 75, 1-10.

Raines, C. A. (2011) Increasing photosynthetic carbon assimilation in C₃ plants to improve crop yield: current and future strategies. *Plant Physiology*, 155, 36-42.

Raines, C. A., Lloyd, J. C. and Dyer, T. A. (1999) New insights into the structure and function of sedoheptulose-1,7-bisphosphatase; an important but neglected Calvin cycle enzyme. *Journal of Experimental Botany*, 50, 1-8.

Raines, C.A., Cavanagh, A.P., Afamefule, C., Chibani, K., Gherli, H., Lopez, P., Mengin, V., Moreno-García, B. and Wall, S., (2023) Understanding the biochemistry of C3 photosynthesis in crop plants. In *Understanding and improving crop photosynthesis* (pp. 3-30). Burleigh Dodds Science Publishing

Ray, D. K., Mueller, N. D., West, P. C. and Foley, J. A. (2013) Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLOS ONE*, 8, e66428.

Rosenthal, D. M., Locke, A. M., Khozaei, M., Raines, C. A., Long, S. P. and Ort, D. R. (2011) Over-expressing the C3 photosynthesis cycle enzyme Sedoheptulose-1-7 Bisphosphatase improves photosynthetic carbon gain and yield under fully open air CO₂ fumigation (FACE). *BMC Plant Biology*, 11, 123.

RSOL 2009. Reaping the benefits: science and the sustainable intensification of global agriculture. London: Royal Society.

Sage, R. F. (2004) The evolution of C4 photosynthesis. *New Phytologist*, 161, 341-370.

Sage, R.F. (2017) A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of Experimental Botany*, 68(2) e11-e28.

Sage, R. F., Sage, T. L. and Kocacinar, F. (2012) Photorespiration and the Evolution of C4 Photosynthesis. *Annual Review of Plant Biology*, 63, 19-47.

Sharkey T. D. (2019). Discovery of the canonical Calvin-Benson cycle.

Photosynthesis research, 140(2), 235–252.

Simkin, A. J., Lopez-Calcagno, P. E., Davey, P. A., Headland, L. R., Lawson, T., Timm, S., Bauwe, H. and Raines, C. A. (2017) Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphosphate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO₂ assimilation, vegetative biomass and seed yield in Arabidopsis. Plant Biotechnology Journal, 15, 805-816.

Solomon S. 2007 Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC. Cambridge, UK: Cambridge University Press.

South, P. F., Cavanagh, A. P., Liu, H. W. and Ort, D. R. (2019) Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. Science, 363, eaat9077.

Spreitzer, R. J. (2002) Rubisco: structure, regulatory interactions, and possibilities for a better enzyme. Annual Review of Plant Biology, v. 53, pp. 449-475-2002 v.53.

Suzuki, Y. and Gojobori, T. (1999) A method for detecting positive selection at single amino acid sites. Molecular Biology and Evolution, 16, 1315-28.

Tamura, K., Stecher, G. and Kumar, S. (2021) MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Molecular Biology and Evolution, 38, 3022-3027.

Tilman, D., Balzer, C., Hill, J. and Befort, B. L. (2011) Global food demand and the sustainable intensification of agriculture. *Proceedings of the National Academy of Sciences*, 108, 20260-20264.

Vasudev, P. G., Banerjee, M., Ramakrishnan, C. and Balaram, P. (2012) Asparagine and glutamine differ in their propensities to form specific side chain-backbone hydrogen bonded motifs in proteins. *Proteins: Structure, Function, and Bioinformatics*, 80, 991-1002.

Villeret, V., Huang, S., Zhang, Y., Xue, Y. and Lipscomb, W. N. (1995) Crystal structure of spinach chloroplast fructose-1,6-bisphosphatase at 2.8 Å resolution. *Biochemistry*, 34, 4299-4306.

Von Caemmerer, S., 2000. *Biochemical models of leaf photosynthesis*. Csiro publishing.

Walker, B. J., VanLoocke, A., Bernacchi, C. J. and Ort, D. R. (2016) The Costs of Photorespiration to Food Production Now and in the Future. *Annual Review of Plant Biology*, 67, 107-129.

Woodrow, I., Berry, J., 1988. Enzymatic regulation of photo-synthetic CO₂ fixation in C₃ plants. *Annual Review of Plant Biology*, 39, 533–594.

Woodrow, I. E., Murphy, D. J. and Latzko, E. (1984) Regulation of stromal sedoheptulose 1,7-bisphosphatase activity by pH and Mg²⁺ concentration. *Journal of Biological Chemistry*, 259, 3791-3795.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences*, 13(5), 555-556.

Yang, Z. (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution*, 15, 568-573.

Yang, Z. and Nielsen, R. (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, 19, 908-17.

Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A. M. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*, 155, 431-49.

Yang, Z., Wong, W. S. and Nielsen, R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, 22, 1107-18.

Zhang, J. (2004) Frequent False Detection of Positive Selection by the Likelihood Method with Branch-Site Models. *Molecular Biology and Evolution*, 21, 1332-1339.

Zhang, Y., Liang, J. Y. and Lipscomb, W. N. (1993) Structural similarities between fructose-1,6-bisphosphatase and inositol monophosphatase. *Biochemical and Biophysical Research Communications*, 190, 1080-3.

Zhang, Y., Liang, J. Y., Huang, S. and Lipscomb, W. N. (1994) Toward a mechanism for the allosteric transition of pig kidney fructose-1,6-bisphosphatase. *Journal of Molecular Biology*, 244, 609-24.

Zhu, X.-G., de Sturler, E. and Long, S. P. (2007) Optimizing the Distribution of Resources between Enzymes of Carbon Metabolism Can Dramatically Increase

Photosynthetic Rate: A Numerical Simulation Using an Evolutionary Algorithm. *Plant Physiology*, 145, 513-526.

Zhu, X. G., Long, S. P. and Ort, D. R. (2010) Improving photosynthetic efficiency for greater yield. *Annual Review of Plant Biology*, 61, 235-61.

Appendix:

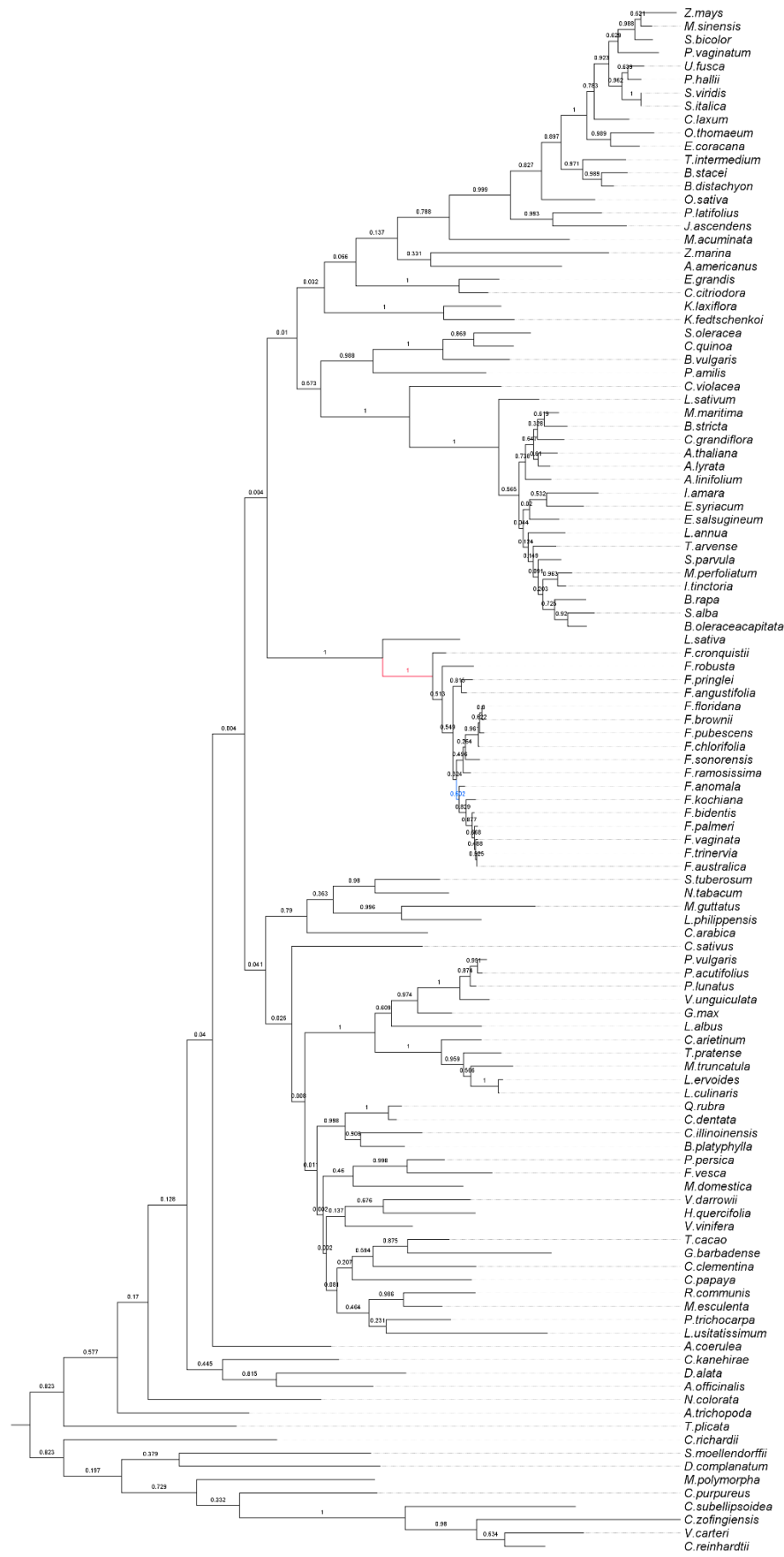


Figure 1. Maximum likelihood phylogenetic tree based on SBPase coding sequences from 116 photosynthetic species. Bootstrap values labelled on branches (/1000). Foreground branches are highlighted in red and blue. Since I did not test for C4 C3 differences in any clade other than Flaveria, Appendix 1 is not organised or delineated by photosynthetic type.

Table 1. List of photosynthetic species used as a source of SBPase coding sequences and their respective phytozome/NCBI page.

Species	Source
<i>Acorus americanus</i>	https://phytozome-next.jgi.doe.gov/report/gene/aamericanus_v1_1/acora.09g005500
<i>Alyssum linifolium</i>	https://phytozome-next.jgi.doe.gov/report/gene/alinifolium_v1_1/alyli.0008s0308
<i>Amborella trichopoda</i>	https://phytozome-next.jgi.doe.gov/report/gene/atrichopoda_v1_0/evm_27.tu.amtr_v1.0_scaffold00010.221
<i>Aquilegia coerulea</i>	https://phytozome-next.jgi.doe.gov/report/gene/acoerulea_v3_1/aqcoe1g084600
<i>Arabidopsis lyrata</i>	https://phytozome-next.jgi.doe.gov/report/gene/alyrata_v2_1/al5g37090
<i>Arabidopsis thaliana</i>	https://phytozome-next.jgi.doe.gov/report/gene/athaliana_tair10/at3g55800
<i>Asparagus officinalis</i>	https://phytozome-next.jgi.doe.gov/report/gene/aofficinalis_v1_1/evm.tu.asparagusv1_07.40
<i>Beta vulgaris</i>	https://phytozome-next.jgi.doe.gov/report/gene/bvulgaris_el10_1_0/el10ac2g04782
<i>Betula platyphylla</i>	https://phytozome-next.jgi.doe.gov/report/gene/bplatyphylla_v1_1/bpchr06g31320
<i>Boechera stricta</i>	https://phytozome-next.jgi.doe.gov/report/gene/bstricta_v1_2/bostr.20505s0170
<i>Brachypodium distachyon</i>	https://phytozome-next.jgi.doe.gov/report/gene/bdistachyonabr2_v1/brdisv1abr21021953m.g
<i>Brachypodium stacei</i>	https://phytozome-next.jgi.doe.gov/report/gene/bstacei_v1_1/brast01g087000
<i>Brassica oleracea capitata</i>	https://phytozome-next.jgi.doe.gov/report/gene/boleraceacapitata_v1_0/bol045435
<i>Brassica rapa</i>	https://phytozome-next.jgi.doe.gov/report/gene/brapafpsc_v1_3/brara.g01673
<i>Capsella grandiflora</i>	https://phytozome-next.jgi.doe.gov/report/gene/cgrandiflora_v1_1/cagra.0003s0005

<i>Carica papaya</i>	https://phytozome-next.jgi.doe.gov/report/gene/cpapaya_asgpbv0_4/evm.tu.supercontig_19.221
<i>Carya illinoensis</i>	https://phytozome-next.jgi.doe.gov/report/gene/cillinoensis_v1_1/caril.01g292400
<i>Castanea dentata</i>	https://phytozome-next.jgi.doe.gov/report/gene/cdentata_v1_1/caden.06g192400
<i>Ceratodon purpureus</i>	https://phytozome-next.jgi.doe.gov/report/gene/cpurpureusgg1_v1_1/cepurg1.ug201900
<i>Ceratopteris richardii</i>	https://phytozome-next.jgi.doe.gov/report/gene/crichardii_v2_1/ceric.06g067900
<i>Chasmanthium laxum</i>	https://phytozome-next.jgi.doe.gov/report/gene/claxum_v1_1/chala.01g284300
<i>Chenopodium quinoa</i>	https://phytozome-next.jgi.doe.gov/report/gene/cquinoa_v1_0/aur62022835
<i>Chlamydomonas reinhardtii</i>	https://phytozome-next.jgi.doe.gov/report/gene/creinhardtii_v5_6/cre03.g185550
<i>Chromochloris zofingiensis</i>	https://phytozome-next.jgi.doe.gov/report/gene/czofingiensis_v5_2_3_2/cz01g42020
<i>Cicer arietinum</i>	https://phytozome-next.jgi.doe.gov/report/gene/carietinum_v1_0/ca_10310
<i>Cinnamomum kanehirae</i>	https://phytozome-next.jgi.doe.gov/report/gene/ckanehirae_v3/ckan_01023500
<i>Citrus clementina</i>	https://phytozome-next.jgi.doe.gov/report/gene/cclementina_v1_0/ciclev10031765m.g
<i>Cleome violacea</i>	https://phytozome-next.jgi.doe.gov/report/gene/cviolacea_v1_1/clevi.0027s0319
<i>Coccomyxa subellipsoidea</i>	https://phytozome-next.jgi.doe.gov/report/gene/csubellipsoideac_169_v2_0/fgenesh1_pm_25_num_2
<i>Coffea arabica</i>	https://phytozome-next.jgi.doe.gov/report/gene/carabica_v0_5/evm.tu.scaffold_2596.420
<i>Corymbia citriodora</i>	https://phytozome-next.jgi.doe.gov/report/gene/ccitriodora_v2_1/cocit.k0555
<i>Cucumis sativus</i>	https://phytozome-next.jgi.doe.gov/report/gene/csativus_v1_0/cucsa.164600
<i>Dioscorea alata</i>	https://phytozome-next.jgi.doe.gov/report/gene/dalata_v1_1/dioal.2372s0026

<i>Diphasiastrum complanatum</i>	https://phytozome-next.jgi.doe.gov/report/gene/dcomplanatum_v3_1/dicom.20g031100
<i>Eleusine coracana</i>	https://phytozome-next.jgi.doe.gov/report/gene/ecoracana_v1_1/eleco.r07.1bg0090250
<i>Eucalyptus grandis</i>	https://phytozome-next.jgi.doe.gov/report/gene/egrandis_v2_0/eucgr.j00242
<i>Euclidium syriacum</i>	https://phytozome-next.jgi.doe.gov/report/gene/esyriacum_v1_1/eusyr.0028s0064
<i>Eutrema salsugineum</i>	https://phytozome-next.jgi.doe.gov/report/gene/esalsugineum_v1_0/thhalv10010434m.g
<i>Fragaria vesca</i>	https://phytozome-next.jgi.doe.gov/report/gene/fvesca_v2_0_a2/gene29967-v1.0-hybrid
<i>Glycine max</i>	https://phytozome-next.jgi.doe.gov/report/gene/gmaxwm82isu_01_v2_1/gmisu01.18g028300
<i>Gossypium barbadense</i>	https://phytozome-next.jgi.doe.gov/report/gene/gbarbadense_v1_1/gobar.a10g247700
<i>Hydrangea quercifolia</i>	https://phytozome-next.jgi.doe.gov/report/gene/hquercifolia_v1_1/hyque.04g072900
<i>Iberis amara</i>	https://phytozome-next.jgi.doe.gov/report/gene/iamara_v1_1/ibeam.3930s0005
<i>Isatis tinctoria</i>	https://phytozome-next.jgi.doe.gov/report/gene/itinctoria_v1_1/isati.3757s0013
<i>Joinvillea ascendens</i>	https://phytozome-next.jgi.doe.gov/report/gene/jascendens_v1_1/joasc.12g118900
<i>Kalanchoe fedtschenkoi</i>	https://phytozome-next.jgi.doe.gov/report/gene/kfedtschenkoi_v1_1/kaladp0045s0221
<i>Kalanchoe laxiflora</i>	https://phytozome-next.jgi.doe.gov/report/gene/klaxiflora_v1_1/kalax.0009s0212
<i>Lactuca sativa</i>	https://phytozome-next.jgi.doe.gov/report/gene/lsativa_v8/lsat_1_v5_gn_2_122641
<i>Lens culinaris</i>	https://phytozome-next.jgi.doe.gov/report/gene/lculinaris_v1/lcu.2rby.3g038090
<i>Lens ervoides</i>	https://phytozome-next.jgi.doe.gov/report/gene/lervoides_v1/ler.1drt.3g040260

<i>Lepidium sativum</i>	https://phytozome-next.jgi.doe.gov/report/gene/lsativum_v1_1/lesat.0151s0044
<i>Lindenbergia philippensis</i>	https://phytozome-next.jgi.doe.gov/report/gene/lphilippensis_v1_1/liphi.01g239900
<i>Linum usitatissimum</i>	https://phytozome-next.jgi.doe.gov/report/gene/lusitatissimum_v1_0/lus10028261.g
<i>Lunaria annua</i>	https://phytozome-next.jgi.doe.gov/report/gene/lannua_v1_1/luann.0034s0210
<i>Lupinus albus</i>	https://phytozome-next.jgi.doe.gov/report/gene/lalbus_v1/lalb_chr15g0084341
<i>Malcolmia maritima</i>	https://phytozome-next.jgi.doe.gov/report/gene/mmaritima_v1_1/mamar.0023s0261
<i>Malus domestica</i>	https://phytozome-next.jgi.doe.gov/report/gene/mdomestica_v1_1/md01g1146300
<i>Manihot esculenta</i>	https://phytozome-next.jgi.doe.gov/report/gene/mesculenta_v8_1/manes.07g124800
<i>Marchantia polymorpha</i>	https://phytozome-next.jgi.doe.gov/report/gene/mpolymorpha_v3_1/mapoly0024s0048
<i>Medicago truncatula</i>	https://phytozome-next.jgi.doe.gov/report/gene/mtruncatula_mt4_0v1/medtr3g070100
<i>Mimulus guttatus</i>	https://phytozome-next.jgi.doe.gov/report/gene/mguttatus_v2_0/migut.f01436
<i>Miscanthus sinensis</i>	https://phytozome-next.jgi.doe.gov/report/gene/msinensis_v7_1/misin06g333100
<i>Musa acuminata</i>	https://phytozome-next.jgi.doe.gov/report/gene/macuminata_v1/gsmua_achr1g17270_001
<i>Myagrum perfoliatum</i>	https://phytozome-next.jgi.doe.gov/report/gene/mperfoliatum_v2_1/myper.0005s0848
<i>Nicotiana tabacum</i>	https://www.ncbi.nlm.nih.gov/nuccore/kj808757.1?report=fasta
<i>Nymphaea colorata</i>	https://phytozome-next.jgi.doe.gov/report/gene/ncolorata_v1_2/nycol.f00895
<i>Oropetium thomaeum</i>	https://phytozome-next.jgi.doe.gov/report/gene/othomaeum_v1_0/oropetium_20150105_12813

<i>Oryza sativa</i>	https://phytozome-next.jgi.doe.gov/report/gene/osativakitaake_v3_1/oskitaake04g051600
<i>Panicum hallii</i>	https://phytozome-next.jgi.doe.gov/report/gene/phallii_v3_2/pahal.5g087500
<i>Paspalum vaginatum</i>	https://phytozome-next.jgi.doe.gov/report/gene/pvaginatum_v3_1/pavag03g339100
<i>Pharus latifolius</i>	https://phytozome-next.jgi.doe.gov/report/gene/platifolius_v1_1/phala.03g292200
<i>Phaseolus acutifolius</i>	https://phytozome-next.jgi.doe.gov/report/gene/pacutifolius_v1_0/phacu.cvr.001g288400
<i>Phaseolus lunatus</i>	https://phytozome-next.jgi.doe.gov/report/gene/plunatus_v1/pl01g0000393700.v1
<i>Phaseolus vulgaris</i>	https://phytozome-next.jgi.doe.gov/report/gene/pvulgaris_v2_1/phvul.001g240200
<i>Populus trichocarpa</i>	https://phytozome-next.jgi.doe.gov/report/gene/ptrichocarpa_v4_1/potri.008g063800
<i>Portulaca amilis</i>	https://phytozome-next.jgi.doe.gov/report/gene/pamilis_v1_0/fun_030206
<i>Prunus persica</i>	https://phytozome-next.jgi.doe.gov/report/gene/ppersica_v2_1/prupe.4g136700
<i>Quercus rubra</i>	https://phytozome-next.jgi.doe.gov/report/gene/qrubra_v2_1/qurub.m184400
<i>Ricinus communis</i>	https://phytozome-next.jgi.doe.gov/report/gene/rcommunis_v0_1/29610.t000022
<i>Schrenkiella parvula</i>	https://phytozome-next.jgi.doe.gov/report/gene/sparvula_v2_2/sp5g06680
<i>Selaginella moellendorffii</i>	https://phytozome-next.jgi.doe.gov/report/gene/smoellendorffii_v1_0/99338
<i>Setaria italica</i>	https://phytozome-next.jgi.doe.gov/report/gene/sitalica_v2_2/seita.5g384700
<i>Setaria viridis</i>	https://phytozome-next.jgi.doe.gov/report/gene/sviridis_v2_1/sevir.5g389900
<i>Sinapis alba</i>	https://phytozome-next.jgi.doe.gov/report/gene/salba_v3_1/sialb.0006s0743
<i>Solanum tuberosum</i>	https://phytozome-next.jgi.doe.gov/report/gene/stuberosum_v6_1/soltu.dm.05g022620
<i>Sorghum bicolor</i>	https://phytozome-next.jgi.doe.gov/report/gene/sbicolor_v3_1_1/sobic.003g359100

<i>Spinacia oleracea</i>	https://phytozome-next.jgi.doe.gov/report/gene/soleracea_spov3/spov3_chr4.03206
<i>Theobroma cacao</i>	https://phytozome-next.jgi.doe.gov/report/gene/tcacao_v2_1/thecc.10g027000
<i>Thinopyrum intermedium</i>	https://phytozome-next.jgi.doe.gov/report/gene/tintermedium_v3_1/thint.s03g384000
<i>Thlaspi arvense</i>	https://phytozome-next.jgi.doe.gov/report/gene/tarvense_v1_1/thlar.0021s0771
<i>Thuja plicata</i>	https://phytozome-next.jgi.doe.gov/report/gene/tplicata_v3_1/thupl.29382069s0008
<i>Trifolium pratense</i>	https://phytozome-next.jgi.doe.gov/report/gene/tpratense_v2/tp57577_tgac_v2_gene29802
<i>Urochloa fusca</i>	https://phytozome-next.jgi.doe.gov/report/gene/ufusca_v1_1/uofu.5g089500
<i>Vaccinium darrowii</i>	https://phytozome-next.jgi.doe.gov/report/gene/vdarrowii_v1_2/vadar_g35736
<i>Vigna unguiculata</i>	https://phytozome-next.jgi.doe.gov/report/gene/vunguiculata_v1_2/vigun01g224100
<i>Vitis vinifera</i>	https://phytozome-next.jgi.doe.gov/report/gene/vvinifera_v2_1/vit_213s0019g03350
<i>Volvox carteri</i>	https://phytozome-next.jgi.doe.gov/report/gene/vcarteri_v2_1/vocar.0015s0215
<i>Zea mays</i>	https://phytozome-next.jgi.doe.gov/report/gene/zmays_refgen_v4/zm00001d042840
<i>Zostera marina</i>	https://phytozome-next.jgi.doe.gov/report/gene/zmarina_v3_1/zosma01g40910

	1	50
<i>A.americanus</i>	T	K
<i>A.coerulea</i>	T	K
<i>A.linifolium</i>	T	K
<i>A.lyrata</i>	T	K
<i>A.officinalis</i>	T	K
<i>A.thaliana</i>	T	K
<i>A.trichopoda</i>	T	R
<i>B.distachyon</i>	T	R
<i>B.oleraceacapitata</i>	T	K
<i>B.platyphylla</i>	T	R
<i>B.rapa</i>	T	K
<i>B.stacei</i>	T	R
<i>B.stricta</i>	T	K
<i>B.vulgaris</i>	T	K
<i>C.arabica</i>	T	K
<i>C.arietinum</i>	T	K
<i>C.citriodora</i>	T	R
<i>C.clementina</i>	T	K
<i>C.dentata</i>	T	R
<i>C.grandiflora</i>	T	K
<i>C.illinoensis</i>	T	R
<i>C.kanehirae</i>	T	R
<i>C.laxum</i>	T	R
<i>C.papaya</i>	T	K
<i>C.purpureus</i>	T	R
<i>C.quinoa</i>	T	K
<i>C.reinhardtii</i>	T	Q
<i>C.richardii</i>	V	R
<i>C.sativus</i>	T	R
<i>C.subellipsoidea</i>	T	Q
<i>C.violacea</i>	T	R
<i>C.zofingiensis</i>	T	K
<i>D.alata</i>	T	K
<i>D.complanatum</i>	T	R
<i>E.coracana</i>	A	R
<i>E.grandis</i>	T	R
<i>E.salsugineum</i>	T	K
<i>E.syracum</i>	T	K
<i>F.angustifolia</i>	T	R
<i>F.anomala</i>	T	R
<i>F.australica</i>	T	R
<i>F.bidentis</i>	T	R
<i>F.brownii</i>	T	R
<i>F.chlorifolia</i>	T	R
<i>F.cronquistii</i>	T	R
<i>F.floridana</i>	T	R
<i>F.kochiana</i>	T	R
<i>F.palmeri</i>	T	R
<i>F.pringlei</i>	T	R
<i>F.pubescens</i>	T	K
<i>F.ramosissima</i>	T	R
<i>F.robusta</i>	T	R
<i>F.sonorensis</i>	T	R
<i>F.trinervia</i>	T	K
<i>F.vaginata</i>	T	R
<i>F.vesca</i>	P	K
<i>G.barbadense</i>	T	K
<i>G.max</i>	T	R
<i>H.quercifolia</i>	T	K

1

50

I.amara TKCEI GQSLEEFLLTQATPDKGLRLLTCMGEALRTI AFKVRTASC GGTTAC

I.tinctoria TKCEI GQSLEEFLLTQATPDKGLRLLTCMGEALRTI AFKVRTASC GGTTAC

J.ascendens TRCEI GDSLEEFLLKATPDKNLI RLLI CMGEAMRTI AFKVRTASC GGTTAC

K.fedtschenkoi TKAELGDSLEEFLLTKSTKDKGLI RLLMCMGEALRTI AFKVRTASC GGTTAC

K.laxiflora TKAELGDSLEEFLLTKSTKDKGLI K - - - - - ALRTI AFKVRTASC GGTTAC

L.albus TRCEI GDSLEEFLLTQATPDKGLI SLLI SMGEALRTI SFKVKTASC GGTTQC

L.annua TKCEI GQSLEEFLLTQATPDKGLRLLTCMGEALRTI AFKVRTASC GGTTAC

L.cullinaris TKCEI GDSLEEFLLSKATPDKGLI RLMVCMGEALRTI SFKVKTASC GGTTQC

L.ervoides TKCEI GDSLEEFLLSKATPDKGLI RLMVCMGEALRTI SFKVKTASC GGTTQC

L.philippensis TKCEI GDSLEEFLLTKSTSDKGLI RLLMCMGEALRTI AFKVRTASC GGTTAC

L.satīva TKCEI GESLEDFLLSKATPDKGLI RLLTCMGEALRTI AFKVRTASC GGTTAC

L.satīvum TKCEI GQSLEEFLLTQATPDKGLRLLTCMGEALRTI AFKVRTASC GGTTAC

L.usitatissimum IRSEI GDSLEEFLLTQATPDKGLI RVMTCMGEALRTI GFKVRTASC GGTTQC

M.acuminata TKCEI GDSLEEFLLTQATPDKNLI RLLMAMGEALRTI SFKVRTASC GGTTAC

M.domestica IKCEI GDSLEEFLLAKATPDKGLI RLLLSMGEALRTI SFKVRTASC GGTTAC

M.esculenta TRCEI GDSLEEFLLTQATPDKGLARLLMCMGEALRTI AFKVRTASC GGTTAC

M.guttatus TKCEI GDSLEEFLLSKSTKDKGLI RLMTCMGEALRTI AFKVRTASC GGTTAC

M.maritima TKCEI GQSLEEFLLAQATPDKGLRLLTCMGEALRTI AFKVRTASC GGTTAC

M.perfoliatum TKCEI GQSLEEFLLTQATPDKGLRLLTCMGEALRTI AFKVRTASC GGTTAC

M.polymorpha TRAEI GDSLEEFLLFKNTKNI NLRQLMLSMGLAI KKI SFKVRTASC GATAC

M.sinensis TRCEI GDSLEEFLLTQATPDKNLI RLLI CMGEAMRTI AFKVRTASC GGTTAC

M.truncatula TKCEI GDSLEEFLLSKATPDKGLI RLMVCMGEALRTI SFKVKTASC GGTTQC

N.colorata TRCEI GDSLEEFLLNKSTEDKNLVRLVCMGEALRTI AFKVRTASC GGTTAC

N.tabacum TKCEI GDSLEEFLLTKSTSDKGLI RLLMCMGEALRTI AFKVRTASC GGTTAC

O.satīva TRCEI GDSLEEFLLTQATPDKNLI RLLI CMGEAMRTI SFKVRTASC GGTTAC

O.thomaeum TRCEI GDSLEEFLLTQATPDKNLI RLLI CMGEAMRTI AFKVRTASC GGTTAC

P.acutifolius TRCEI GDGLEEFLLTQATPDKGLI RLLVSMGEALRTI SFKVKTASC GGTTQC

P.amilis TKCEI GDSLEEFLLTQATPDKGLI RLLMCMGEALRTI AFKVRTASC GGTTQC

P.hallii ARCEI GDSLEEFLLTQATPDKNLI RLLTCMGEAMRTI AFKVRTASC GGTTAC

P.latifolius TRCEI GDSLEEFLLTQATPDKNLI RLLMMSMGEAMRTI SFKVRTASC GGTTAC

P.lunatus TKCEI GDGLEEFLLTQATPDKGLI RLLVSMGEALRTI SFKVKTASC GGTTQC

P.persica PKCAI GDSLEEFLLTQATPDKKLI TLLI SMGEALRTI GFKVRTASC GGTTAC

P.trichocarpa TRCEI GDSLEEFLLAKATPDKGLI RVMCMGEALRTI AFKVRTASC GGTTAC

P.vaginatūm ARCEI GDSLEEFLLTQATPDKNLI RLLI CMGEALRTI SFKVRTASC GGTTAC

P.vulgaris TRCEI GDGLEEFLLTQATPDKGLI RLLVSMGEALRTI SFKVKTASC GGTTQC

Q.rubra TRCEI GDSLEEFLLKATPDKGLI RLMTCMGEALRTI SFKVKTASC GGTTAC

R.communis TRCEI GDSLEEFLLAKATPDKGLARLLMCMGEALRTI AFKVRTASC GGTTAC

S.alba TKCEI GQSLEEFLLTQATQDKGLRLLTCMGEALRTI AFKVRTASC GGTTAC

S.bicolor TRCEI GDSLEEFLLTQATPDKNLI RLLI CMGEAMRTI AFKVRTASC GGTTAC

S.italica TRCEI GDSLEEFLLTQATPDKNLI RLLTCMGEAMRTI AFKVRTASC GGTTAC

S.moellendorffii - - AELGDSLEEFLLATSTKDKSLARLLTCMGEALRTI SFKVKTASC GATAC

S.oleracea TKCEI GDSLEEFLLAKATTDKGLI RLLMCMGEALRTI GFKVRTASC GGTTQC

S.parvula TKCEI GQSLEEFLLAQATPDKGLRSLLMCMGEALRTI AFKVRTASC GGTTAC

S.tuberosum TKCEI GDSLEEFLLSKSTSDKGLI RLLMCMGEALRTI AFKVRTASC GGTTAC

S.viridis TRCEI GDSLEEFLLTQATPDKNLI RLLTCMGEAMRTI AFKVRTASC GGTTAC

T.arvense TKCEI GQSLEEFLLTQATPDKGLRLLTCMGEALRTI AFKVRTASC GGTTAC

T.cacao TRCEI GDSLEEFLLTQATPDKGLI RLLMCMGEALRTI AFKVRTASC GGTTAC

T.intermedium TRCAI GDSLEEFLLTQATPDKNLI RLLI CMGEAMRTI AFKVRTASC GGTTAC

T.plicata IKCEI GDSLEEFLLNKSTPNKNLI RLMVCMGEAMRTI AFKVRTASC GATAC

T.pratense TKCEI GDSLEEFLLSKATPDKGLVRLVCMGEALRTI AFKVKTASC GGTTQC

U.fusca ARCEI GDSLEEFLLTQATPDKNLI RLLTCMGEAMRTI AFKVRTASC GGTTAC

V.carteri TQAKI GDSLAEFLVEATPDKPLRQLMMSMAEAVRTI GHKVRTASC AGTTAC

V.darrowii ARCALGDSLEEFLLKATPDKGLI RLMTCMGEALRTI SFKVRTASC GGTTAC

V.unguiculata TRCEI GDSLEEFLLTQATPDKGLI RLLVSMGEALRTI SFKVKTASC GGTTQC

V.vinifera TRCEI GDSLEEFLLTQATPDKGLI RLLMCMGEALRTI AFKVRTASC GGTTAC

Z.marina TKCEI GDSLEEFLLTQATPDKGLI RLLVCMGEAMRTI AFKVRTASC GGTTAC

Z.mays ARCEI GDSLEEFLLTQATPDKNLI RLLI CMGEAMRTI AFKVRTASC GGTTAC

A.americanus V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H F C K Y A C S E E V P E L Q D M G G P V E G G
A.coerulea I N T F G D E Q L A V D M L A D K L L F E A L T Y S H F C K Y A C S E E V P E L Q D M G G P V E G G
A.linifolium V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
A.lyrata V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
A.officinalis V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H F C K Y A C S E E V P E L Q D M G G P V E G G
A.thaliana V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
A.trichopoda V N S F G D E Q L A V D I L A N Q L L F E A L R Y S H F C K Y A C S E E V P E L Q D M G G P V E G G
B.distachyon V N S F G D E Q L A V D M L A D K L L F E A L E Y S H V C K Y A C S E E V P E L Q D M G G P V D G G
B.oleraceacapitata V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
B.platyphylla V N S F G D E Q L A V D M L A D K L L F E A L T Y S H F C K Y A C S E E V P E L Q D M G G P V E G G
B.rapa V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
B.stacei V N S F G D E Q L A V D M L A D K L L F E A L E Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
B.stricta V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
B.vulgaris I N T F G D E Q L A I D V L A D K L L F E A L N Y S H V C K Y A C S E E V P E L Q E M G G P V E G G
C.arabica V N S F G D E Q L A V D L L A N N V L F E A L K Y S H I V K Y G C S E E V P E L Q D M E G P A E G G
C.arietinum V N S F G D E Q L A V D M L A N N L L F E A L K H S H F C K Y A C S E E I P E L Q D M G G P T E G G
C.citriodora I N S F G D E Q L A V D V L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P I E G G
C.clementina V N S F G D E Q L A V D M L A D K L L F E A L T Y S H F C K Y A C S E E V P E L Q D M G G P A E G G
C.dentata V N S F G D E Q L A V D M L A D K L L F E A L N Y S H F C K Y A C S E E V P E L Q D M G G P A E G G
C.grandiflora V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P A E G G
C.illinoensis V N S F G D E Q L A V D M L A D K L L F E A L T Y S H F C K Y A C S E E V P E L Q D M G G P V K G G
C.kanehirae V N S F G D E Q L A V D M L A D K L L F E A L T Y S H V C K Y A C S E E L P E L Q D M G G P V E G G
C.laxum V N S F G D E Q L A V D M L A N K L L F D A L E Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
C.papaya V N S F G D E Q L A V D M L A N K L L F E A L Q Y S H F C K Y A C S E E V P E L Q D M G G P V E G G
C.purpureus I N T F G D E Q L A V D M L A D K L L F E A L K H S H V C K Y A C S E E E P V L Q D M E G E - - - G
C.quinoa V N T F G D E Q L A I D V L A D K L L F E A L N Y S H F C K Y A C S E E I P E L Q D M G G P V D G G
C.reinhardtii V N S F G D E Q L A V D M V A D K L L F E A L K Y S H V C K L A C S E E V P E P V D M G G E - - - G
C.richardii L N T F G D E Q L A V D L L A N K L L F E A L R Y S H V C K W A C S E E D P V P L D M E G E - - - G
C.sativus V N S F G D E Q L A V D M L A D K L L F E A L R Y S H F C K Y A C S E E V P E L Q D M G G P V E G G
C.subellipsoidea V N T F G D E Q L A V D L L A D K L L F E A L K F S G V V K Y A C S E E V P E P V D M Q G E - - - G
C.violacea V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
C.zofingiensis V N S F G D E Q L A V D M L A D K L L F E A L R F S H V C K Y A C S E E V P E P V D V G G E - - - G
D.alata V N S F G D E Q L A V D M L A N N L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
D.complanatum I N T F G D E Q L A V D L L A N K L L F E A L K F S H F C K Y A C S E E D P E L L D M G G P V N G G
E.coracana V N S F G D E Q L A V D M L A N K L L F D A L E Y S H V C K Y A C S E E V P T L Q D M G G P V E G G
E.grandis V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H F C K Y A C S E E V P E L Q D M G G P V Q G G
E.salsugineum V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
E.syracum V N S F G D E Q L A V D M L A D K L L F E A L Q Y S H V C K Y A C S E E V P E L Q D M G G P A E G G
F.angustifolia V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.anomala V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.australica V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.bidentis V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.brownii V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.chlorifolia V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.cronquistii V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.floridana V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.kochiana V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.palmeri V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.pringlei V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.pubescens V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.ramosissima V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.robusta V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.sonorensis V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.trinervia V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.vaginata V N S F G D E Q L A V D V L A D K L L F D A L S H S H F C K Y A C S E E V P E L Q D M G G P V E G G
F.vesca V N S F G D E Q L A V D M L A D K L L F E A L T Y S H V C K Y A C S E E V P E L Q D M G G P V E G G
G.barbadense V N S F G D E Q L A V D L L A N K L L F E A L T Y S H F C K Y A C S E E I P E L Q D M G G P A E G G
G.max V N T F G D E Q L A V D L L A N Q L L F E A L N Y S H F C K Y A C S E E N P E L L D M G G P V E G G
H.quercifolia V N S F G D E Q L A V D M L A D K L L F E A L T Y S H Y C K Y A C S E E V P E L Q D M G G P V E G G

51

100

I.amara VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
I.tinctoria VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
J.ascendens VNSFGDEQLAVDLLANKLLFEALEYSHVCKYACSEEVPELQEMGGPV QGG
K.fedtschenkoi VNSFGDEQLAVDMLANNLLFEALTYSHFCKYACSEEVPELQDMGGPA EGG
K.laxiflora VNSFGDEQLAVDMLANKLLFEALTYSHFCKYACSEEVPELQDMGGPV EGG
L.albus VNSFGDEQLAVDMLANKLLFDALKYSHVCKYACSEEVPELQDMGGPI EGG
L.annua VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
L.cullinaris VNSFGDEQLAVDMLANNLLFEALKHSHFCKYACSEEIPELQDMGGPV EGG
L.ervoides VNSFGDEQLAVDMLANNLLFEALKHSHFCKYACSEEIPELQDMGGPV EGG
L.philippensis VNSFGDEQLAVDMLANKLLFEALQYSHFCKYACSEEVPELQDMGGPV EGG
L.satīva VNSFGDEQLAVDMLADKLLFDALTYSHFCKYACSEEVPELQDMGGPV EGG
L.satīvum VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPA EGG
L.usitatissimum VNSFGDEQLAVDMLADKLLFEALEHSHFCKYACSEEVPELQDMGGPV EGG
M.acuminata VNSFGDEQLAVDLLANNLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
M.domestica VNSFGDEQLAVDMLADKLLFEALSYSHVCKYACSEEVPELQDMGGPV EGG
M.esculenta VNSFGDEQLAVDMLADKLLFEALAYSHFCKYACSEEVPELQDMGGPT EGG
M.guttatus VNSFGDEQLAVDMLADKLLFEALEYSHYCKYACSEEVPELQDMGGPA EGG
M.maritima VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
M.perfoliatum VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
M.polymorpha VNTFGDEQLAVDMLANKVLF EALLHSHVCKYACSEEEPNLVDMGGPT EGG
M.sinensis VNSFGDEQLAVDMLANKLLFEALEYSHVCKYACSEEVPELQDMGGPV EGG
M.truncatula VNSFGDEQLAVDMLANNLLFEALKHSHFCKYACSEEIPELQDMGGPV DGG
N.colorata VNTFGDEQLAVDLLADKLLFEALRYSHVCKYACSEEEPELQDMEGPA EGG
N.tabacum VNSFGDEQLAVDMLANKLLFDALTYSHVCKYACSEEVPELQDMGGPAI GG
O.satīva VNSFGDEQLAVDMLADKLLFEALEYSHVCKYACSEEVPELQDMGGPV DGG
O.thomaeum VNSFGDEQLAVDMLANKLLFDAL EYSHVCKYACSEEVPELQDMGGPV EGG
P.acutifolius VNSFGDEQLAVDMLANQLLFDALNYSHFCKYACSEEEPELQDMGGPV EGG
P.amilis VNSFGDEQLAVDLLANQLLFEALTYSHVCKYACSEEVPELQDMGGPT EGG
P.hallii VNSFGDEQLAVDMLANKLLFDAL EYSHVCKYACSEEVPELQDMGGPV DGG
P.latifolius VNSFGDEQLAVDMLADKLLFEALEYSHVCKYACSEEVPELQDMGGPV EGG
P.lunatus VNSFGDEQLAVDMLANQLLFDALNYSHFCKYACSEEEPELQDMGGPV EGG
P.persica VNSFGDEQLAVDMLADKLLFEALTYSHVCKYACSEEVPELQDMGGPA EGG
P.trichocarpa VNSFGDEQLAVDMLANNLLFEALTYSHFCKYACSEEVPELQDMGGPV EGG
P.vaginatūm VNSFGDEQLAVDMLANKLLFDAL EYSHVCKYACSEEIPELQDMGGPV EGG
P.vulgaris VNSFGDEQLAVDMLANQLLFDALNYSHFCKYACSEEEPELQDMGGPV EGG
Q.rubra VNSFGDEQLAVDMVADKLLFEALNYSHFCKYACSEEVPELQDMGGPV EGG
R.communis VNSFGDEQLAVDMLANQLLFEALNYSHYCKYACSEEVPELQDMGGPA EGG
S.alba VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
S.bicolor VNSFGDEQLAVDMLANKLLFEALEYSHVCKYACSEEVPELQDMGGPV EGG
S.italica VNSFGDEQLAVDMLANKLLFDAL EYSHVCKYACSEEVPELQDMGGPV EGG
S.moellendorffii VNTFGDEQLAVDLLANKLLFEALRF SHVCKYACSEEDPEL LDMGGPV EGG
S.oleracea VNTFGDEQLAIDV LADKLLFEALNYSHFCKYACSEELPELQDMGGPV DGG
S.parvula VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
S.tuberosum VNSFGDEQLAVDMLADKLLFEALTYSHFCKYACSEEVPELQDMGGPA EGG
S.viridis VNSFGDEQLAVDMLANKLLFDAL EYSHVCKYACSEEVPELQDMGGPV EGG
T.arvense VNSFGDEQLAVDMLADKLLFEALQYSHVCKYACSEEVPELQDMGGPV EGG
T.cacao VNSFGDEQLAVDLLANKLLFEALTYSHFCKYACSEEVPELQDMGGPA EGG
T.intermedium VNSFGDEQLAVDMLADKLLFEALEYSHVCKYACSEEVPELQDMGGPV EGG
T.plicata INTFGDEQLAVDLIANKLLFESLRYSHFCKYACSEENPVPEDMGGPV EGG
T.pratense VNSFGDEQLAVDMLANNLLFEALKHSHFCKYACSEEIPELQDMGGPV EGG
U.fusca VNSFGDEQLAVDMLANKLLFDAL EYSHVCKYACSEEVPELQDMGGPV EGG
V.carteri VNSFGDEQLAVDMVADKLLFEALKYSHVCKLACSEEVPEPVDMGGE - - G
V.darrowii VNSFGDEQLAVDMLADKLLFEALNYSHFCKYACSEEVPELQDMGGPV EGG
V.unguiculata VNSFGDEQLAVDMLANQLLFEALNYSHFCKYACSEEEPELQDMGGPV EGG
V.vinifera VNSFGDEQLAVDMVANQLLFEALNYSHFCKHACSEEVPELQDMGGPV EGG
Z.marina VNTFGDEQLAVDMLADKLLFEALTYSHFCKYACSEEVPELQDMGGPV EGG
Z.mays VNSFGDEQLAVDMLANKLLFEALEYSHVCKYACSEEVPELQDMGGPV EGG

101	<p><i>A.americanus</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>A.coerulea</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>A.linifolium</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>A.lyrata</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>A.officinalis</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>A.thaliana</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G I T G G D Q V A A A M G I Y G P</p> <p><i>A.trichopoda</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>B.distachyon</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>B.oleraceacapitata</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>B.platyphylla</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A S A M G I Y G P</p> <p><i>B.rapa</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G A D Q V A A A M G I Y G P</p> <p><i>B.stacei</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>B.stricta</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>B.vulgaris</i> F S V A F D P L D G S S I V D T N F S V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>C.arabica</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I F G P</p> <p><i>C.arietinum</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L I G V T G K D Q V A A A M G V L G P</p> <p><i>C.citriodora</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>C.clementina</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>C.dentata</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>C.grandiflora</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>C.illinoensis</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A S A M G I Y G P</p> <p><i>C.kanehirae</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>C.laxum</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G A D Q V A A A M G I Y G P</p> <p><i>C.papaya</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R E Q V A A A M G I Y G P</p> <p><i>C.purpureus</i> F C V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T E I T G R D Q A A S A M G I Y G P</p> <p><i>C.quinoa</i> F S V A F D P L D G S S I V D T N F S V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>C.reinhardtii</i> F C V A F D P L D G S S I V D T N F A V G T I F G V W P G D K L T N I T G R E Q V A A G M G I Y G P</p> <p><i>C.richardii</i> Y S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T D I T G R D L A A S A M G I Y G P</p> <p><i>C.sativus</i> F S V A F D P L D G S S I V D T N F S V G T I F G V W P G D K L T G V K G S D Q V A A A M G V Y G P</p> <p><i>C.subellipsoidea</i> F S V A F D P L D G S S I V D T N F T V G T I F G I W P G N K L T G I T G R D Q A A A G M G I Y G P</p> <p><i>C.violacea</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>C.zofingiensis</i> Y C V A F D P L D G S S I V D T N F A V G T I F G V W P G D K L K D I N G R Q Q A A A G M G I Y G P</p> <p><i>D.alata</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T S V T G R D Q V A A A M G I Y G P</p> <p><i>D.complanatum</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A G M G I Y G P</p> <p><i>E.coracana</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P</p> <p><i>E.grandis</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>E.salsugineum</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G A D Q V A A A M G I Y G P</p> <p><i>E.syracum</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G A D Q V A A A M G I Y G P</p> <p><i>F.angustifolia</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.anomala</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.australica</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.bidentis</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.brownii</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.chlorifolia</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.cronquistii</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.floridana</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.kochiana</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L N G V T G R D Q V A A A M G I Y G P</p> <p><i>F.palmeri</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.pringlei</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.pubescens</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.ramosissima</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.robusta</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.sonorensis</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.trinervia</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.vaginata</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P</p> <p><i>F.vesca</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G A D Q V A A A M G I Y G P</p> <p><i>G.barbadense</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I F G P</p> <p><i>G.max</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G I T G R D Q V A A A M G V L G P</p> <p><i>H.quercifolia</i> F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R E Q V A A A M G I Y G P</p>	150
-----	--	-----

101

150

I.amara F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
I.tinctoria F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
J.ascendens F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
K.fedtschenkoi F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I F G P
K.laxiflora F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I F G P
L.albus F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G V L G P
L.annua F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
L.cullinaris F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L I G V T G R E Q V A A A M G V L G P
L.ervoides F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L I G V T G R E Q V A A A M G V L G P
L.philippensis F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G A Q Q V A A A M G I F G P
L.satīva F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
L.satīvum F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G A D Q V A A A M G I Y G P
L.usitatissimum F S V A F D P L D G S S I V D T N F S V G T I F G V W P G D K L T G V T G R D Q V A A A M G V Y G P
M.acuminata F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P
M.domestica F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P
M.esculenta F S V A F D P L D G S S I V D T N F S V G T I F G V W P G D K L T G V T G R D Q V A A A M G V Y G P
M.guttatus F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G S Q Q V A A A M G I F G P
M.maritima F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
M.perfoliatum F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
M.polymorpha F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R E Q V A A A M G I Y G P
M.sinensis F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
M.truncatula F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L I G V T G R D Q V A A A M G V L G P
N.colorata F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G I T G R D Q A A A A M G I Y G P
N.tabacum F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G I T G R D Q V A A A M G I F G P
O.satīva F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
O.thomaeum F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
P.acutifolius F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A G M G V M G P
P.amilis F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V S G R D Q V A A A M G I Y G P
P.hallii F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
P.latifolius F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
P.lunatus F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G V M G P
P.persica F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G S D Q V A A A M G I Y G P
P.trichocarpa F S V A F D P L D G S S I V D T N F S V G T I F G V W P G E K L I G V T G R D Q V A A A M G V Y G P
P.vaginatūm F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
P.vulgaris F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G V M G P
Q.rubra F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P
R.communis F S V A F D P L D G S S I V D T N F S V G T I F G V W P G D K L T G V T G R D Q V A A A M G V Y G P
S.alba F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
S.bicolor F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
S.italica F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
S.moellendorffii F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G N D Q V A A A M G I Y G P
S.oleracea F S V A F D P L D G S S I V D T N F S V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P
S.parvula F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
S.tuberosum F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G I T G R E Q V A A A M G I F G P
S.viridis F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
T.arvense F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
T.cacao F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I F G P
T.intermedium F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
T.plicata F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A S A M G I Y G P
T.pratense F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L I G V T G A D Q V A A A M G V L G P
U.fusca F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P
V.carteri F C V A F D P L D G S S I V D T N F S V G T I F G V W P G D K L V N I T G R E Q A A A G M G I Y G P
V.darrowii F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P
V.unguiculata F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G V M G P
V.vinifera F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P
Z.marina F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G R D Q V A A A M G I Y G P
Z.mays F S V A F D P L D G S S I V D T N F T V G T I F G V W P G D K L T G V T G G D Q V A A A M G I Y G P

151	<p><i>A.americanus</i> R T T Y V V A L K D L P G T H E F L L L D E G K W Q H V K D T T Q I G E G K M F S P G N L R A T F D</p> <p><i>A.coerulea</i> R T T Y V I A L K D I P G T H E F L L L D E G K W Q H V K D T T E I G E G K M F S P G N L R A T F D</p> <p><i>A.linifolium</i> R T T Y V L A V K G F P G T H E F L L L D E G K W Q H V K E T T E I G E G K M F S P G N L R A T F D</p> <p><i>A.lyrata</i> R T T Y V L A V K G F P G T H E F L L L D E G K W Q H V K E T T E I A E G K M F S P G N L R A T F D</p> <p><i>A.officinalis</i> R T T Y V L A L K D I P G T H E F L L L D E G K W Q H V K D T T S I G E G K M F S P G N L R A T F D</p> <p><i>A.thaliana</i> R T T Y V L A V K G F P G T H E F L L L D E G K W Q H V K E T T E I A E G K M F S P G N L R A T F D</p> <p><i>A.trichopoda</i> R T T Y V L A L K D L P G T H E F L L L D E G K W Q H V K D T T E I G E G K M F S P G N L R A T S D</p> <p><i>B.distachyon</i> R T T F V V A L K D C P G T H E F L L L D E G K W Q H V K D T T T I G E G K M F S P G N L R A T F D</p> <p><i>B.oleraceacapitata</i> R T T Y V L A V K G F P G T H E F L L L D E G K W Q H V K E T T E I N E G K M F S P G N L R A T F D</p> <p><i>B.platyphylla</i> R T T Y V L A L K D V P G T H E F L L L D E G K W Q H V K D T T E I G E G K L F S P G N L R A T F D</p> <p><i>B.rapa</i> R T T Y V L A V K G F P G T H E F L L L D E G K W Q H V K E T T E I N E G K M F S P G N L R A T F D</p> <p><i>B.stacei</i> R T T F V V A L K D C P G T H E F L L L D E G K W Q H V K D T T T I G E G K M F S P G N L R A T F D</p> <p><i>B.stricta</i> R T T Y V L A V K G F P G T H E F L L L D E G K W Q H V K E T T E I A E G K M F S P G N L R A T F D</p> <p><i>B.vulgaris</i> R T T Y V L A L K D F P G T H E F L L L D E G K W Q H V K E T T E I N E G K L F C P G N L R A T S D</p> <p><i>C.arabica</i> R T T Y V L A L K D L P G T H E F L L L D E G K W Q H V K D T T E I G E G K M F S P G N L R A T S D</p> <p><i>C.arietinum</i> R T T Y V L A L K D F P G T H E F L L L D E G K W Q H V K E T N E I G E G K L F S P G N L R A T F D</p> <p><i>C.citriodora</i> R T T Y V L A V K D V P G T H E F L L L D E G K W L H V K D T T E I G E G K M F S P G N L R A T F D</p> <p><i>C.clementina</i> R T T Y V I A I K D F P G T H E F L L L D E G K W Q H V K E T T E I G E G K M F S P G N L R A T F D</p> <p><i>C.dentata</i> R T T Y V L A L K D I P G T H E F L L L D E G K W Q H V K D T T E I N E G K L F S P G N L R A T F D</p> <p><i>C.grandiflora</i> R T T Y V L A I K G F P G T H E F L L L D E G K W Q H V K E T T E I A E G K M F S P G N L R A T F D</p> <p><i>C.illinoensis</i> R T T Y V L A I K D F P G T H E F L L L D E G K W Q H V K E T T E I G E G K L F S P G N L R A T F D</p> <p><i>C.kanehirae</i> R T T Y I I A L K D A P G T H E F L L L D E G K W Q H V K D T Y E I G E G K L F S P G N L R A T F D</p> <p><i>C.laxum</i> R T T Y V I A L K D C P G T H E F L L L D E G K W Q H V K D T T S I G E G K M F S P G N L R A T F D</p> <p><i>C.papaya</i> R T T Y V L A I K D I P G T H E F L L L D E G K W Q H V K E T T E I G E G K L F S P G N L R A T Y D</p> <p><i>C.purpureus</i> R T T Y V I A I N G F P G T H E F L L L D D G K W Q H V K E T T E I R E G K L F S P G N L R A T F D</p> <p><i>C.quinoa</i> R T T Y V L A L K D Y P G T H E F L L L D E G K W Q H V K E T T E I N E G K L F C P G N L R A T S D</p> <p><i>C.reinhardtii</i> R T V F C I A L K D A P G C H E F L L M D D G K W M H V K E T T H I G E G K M F A P G N L R A T F D</p> <p><i>C.richardii</i> R T T Y V L A L A D V P G T H E F L L M D D G K W Q H V K E T T E I G E G K L F S P G N L R A T F D</p> <p><i>C.sativus</i> R T T Y V L A L K D F P G T H E F L L L D E G K W Q H V K E T T E I G E G K L F S P G N L R A T F D</p> <p><i>C.subellipsoidea</i> R T I Y C I A L K D A P G C H E F V L Q D D G K W L H I K E T T E I G E G K M F S P G N L R A T F D</p> <p><i>C.violacea</i> R T T Y V V A L K G Y P G T H E F L L L D E G K W Q H V K E T T E I G E G K M F S P G N L R A T F D</p> <p><i>C.zofingiensis</i> R T V F C L A I T G C P G T H E F L L M D D G K W V H V K E T T E I G E G K L F A P G N L R A T F D</p> <p><i>D.alata</i> R T T Y I I A L K D C P G T H E F L L L D E G K W Q H V K D T T S I G E G K I F S P G N L R A T A D</p> <p><i>D.complanatum</i> R T T Y V L A L K D F P G T H E F L L L D D G K W Q H V K E T T E I G E G K L F S P G N L R A T F D</p> <p><i>E.coracana</i> R T T Y V I A L K D C P G T H E F L L L D E G K W Q H V K D T T S I G E G K M F S P G N L R A T F D</p> <p><i>E.grandis</i> R T T Y V L A L K D C P G T H E F L L L D E G K W Q H V K D T T E I G E G K M F S P G N L R A T F D</p> <p><i>E.salsugineum</i> R T T Y V L A V K G F P G T H E F L L L D E G K W Q H V K E T T E I A E G K M F S P G N L R A T F D</p> <p><i>E.syracum</i> R T T Y V L A V K G F P G T H E F L L L D E G K W Q H V K E T T E I N E G K M F S P G N L R A T F D</p> <p><i>F.angustifolia</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.anomala</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.australica</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.bidentis</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.brownii</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.chlorifolia</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.cronquistii</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.floridana</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.kochiana</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.palmeri</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.pringlei</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.pubescens</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.ramosissima</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.robusta</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.sonorensis</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.trinervia</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.vaginata</i> R T T Y V I A I K G F P G T H E F L L L D E G K W Q H V K E T T E I S E G K M F S P G N L R A T F D</p> <p><i>F.vesca</i> R T T Y V L A I K G F P G T H E F L L L D E G K W Q H V K E T T E I G E G K M F S P G N L R A T F D</p> <p><i>G.barbadense</i> R T T Y V L A L K D I P G T H E F L L L D E G K W Q H V K D T Y E I G E G K M F S P G N L R A T F D</p> <p><i>G.max</i> R T T Y V L A L K D F P G T H E F L L L D E G K W Q H V K E T T E I G E G K L F S P G N L R A T S D</p> <p><i>H.quercifolia</i> R T T Y V L A L K D V P G T H E F L L L D E G K W Q H V K D T T E I G E G K M F S P G N L R A T F D</p>	200
-----	--	-----

151

200

I.amara RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI NEGKMFSPGNLRATFD
I.tinctoria RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI NEGKMFSPGNLRATFD
J.ascendens RTTTYIVALKDSPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
K.fedtschenkoi RTTTYVLALKDIPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATYD
K.laxiflora RTTTYVLALDDIPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATYD
L.albus RTTTYVLSLKDYPGTHEFLLLDEGKWQHVKETTEI NEGKIFSPGNLRATVD
L.annua RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI GEGKMFSPGNLRATFD
L.cullinaris RTTTYVLALKDFPGTHEFLLLDEGKWQHVKETHEI GEGKLFSPGNLRATFD
L.ervoides RTTTYVLALKDFPGTHEFLLLDEGKWQHVKETHEI GEGKLFSPGNLRATFD
L.philippensis RTTTYVLALKDIPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATSD
L.satīva RTTTYVIAIKGFPGTHEFLLLDEGKWQHVKETTEI SEGKMFSPGNLRATFD
L.satīvum RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI AEGKMFSPGNLRATFD
L.usitatissimum RTTTYVLALKDYPGTYEFLLLDEGKWQLVKETNEI GEGKLFSPGNLRATFD
M.acuminata RTTTYIIALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKIFSPGNLRATFD
M.domestica RTTTYVLAIKGFPGTHEFLLLDEGKWQHVKETTEI GEGKMFSPGNLRATFD
M.esculenta RTTTYVLALKDYPGTHEFLLLDEGKWQHVKETTEV GEGKLFSPGNLRATFD
M.guttatus RTTTYVLALKDVPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
M.maritima RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI AEGKMFSPGNLRATFD
M.perfoliatum RTTTYVAVKGFPGTHEFLLLDEGKWQHVKETTEI NEGKMFSPGNLRATFD
M.polymorpha RTTTYVCLAEVPGTHEFLLMDDGTWQHVKETTTI GEGKLFSPGNLRATYD
M.sinensis RTTTYIVALKDPCPGTHEFLLLDEGKWQHVKDTTTI GEGKMFSPGNLRATFD
M.truncatula RTTTYVLALKDFPGTHEFLLLDEGKWQHVKETHEI GEGKLFSPGNLRATFD
N.colorata RTTTYVLALKDIPGTHEFLLLDEGKWQHVKDTTQI GEGKMFSPGNLRATFD
N.tabacum RTTTYVVALKDVPGTHEFLLLDEGKWQHVKDTTEI EEGKMFSPGNLRATFD
O.satīva RTTTYIIALKDCPGTHEFLLLDEGKWQHVKDTTTI GEGKMFSPGNLRATFD
O.thomaeum RTTTYVIALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
P.acutifolius RTTTYVLAIKGFPGTHEFLLLDEGKWQHVKETTEI GEGKLFSPGNLRATSD
P.amilis RTTTYVLAIKGFPGTHEFLLLDEGKWQHVKETTEI GEGKMFSPGNLRATVD
P.hallii RTTTYVIALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
P.latifolius RTTTYVIALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
P.lunatus RTTTYVLAIKGFPGTHEFLLLDEGKWQHVKETTEI GEGKLFSPGNLRATSD
P.persica RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI GEGKLFSPGNLRATFD
P.trichocarpa RTTTYVLALKDYPGTHEFLLLDEGKWQHVKETTEV GEGKLFSPGNLRATFD
P.vaginatūm RTTTYIIALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
P.vulgaris RTTTYVLAIKGFPGTHEFLLLDEGKWQHVKETTEI GEGKLFSPGNLRATSD
Q.rubra RTTTYVLALKDIPGTHEFLLLDEGKWQHVKDTTEI NEGKLFSPGNLRATFD
R.communis RTTTYVLALKDYPGTHEFLLLDEGKWQHVKETTEV GEGKLFSPGNLRATFD
S.alba RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI NEGKMFSPGNLRATFD
S.bicolor RTTTYIVALKDPCPGTHEFLLLDEGKWQHVKDTTTI GEGKMFSPGNLRATFD
S.italica RTTTYVIALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
S.moellendorffii RTTTYVMAIKGHPGTHEFLLLDDGKWQHVKETTSI GEGKLFSPGNLRATFD
S.oleracea RTTTYVLALKDYPGTHEFLLLDEGKWQHVKETTEI NEGKLFSPGNLRATSD
S.parvula RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI AEGKMFSPGNLRATFD
S.tuberosum RTTTYVLALKDVPGTHEFLLLDEGKWQHVKDTTEI GEGKMFSPGNLRATFD
S.viridis RTTTYVIALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
T.arvense RTTTYVLAVKGFPGTHEFLLLDEGKWQHVKETTEI AEGKMFSPGNLRATFD
T.cacao RTTTYVLALKDSPGTHEFLLLDEGKWQHVKDTTEI GEGKMFSPGNLRATSD
T.intermedium RTTFVVALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
T.plicata RTTTYVLALKDAPGTHEFLLMDDGKWHHVKDTTEI GEGKLFSPGNLRATFD
T.pratense RTTTYVLAIKGFPGTHEFLLLDEGKWQHVKETHEI SEGKLFSPGNLRATFD
U.fusca RTTTYVIALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
V.carteri RTVFCIALKDCPGTHEFLLMDDGKWMHVKETTEI GEGKMFAPGNLRATFD
V.darrowii RTTTYVLALKDIPGTHEFLLLDEGKWLVHVKDTTEI GEGKMFSPGNLRATFD
V.unguiculata RTTTYVLAIKGFPGTHEFLLLDEGKWQHVKETTEI GEGKLFSPGNLRATSD
V.vinifera RTTTYVLALKDIPGTHEFLLLDEGKWQHVKDTTEI GEGKLFSPGNLRATFD
Z.marina RTTTYVVALKDCPGTHEFLLLDEGKWQHVKDTTSSI GEGKMFSPGNLRATFD
Z.mays RTTTYIVALKDCPGTHEFLLLDEGKWQHVKDTTTI GEGKMFSPGNLRATFD

201		250
<i>A.americanus</i>	NP DY QK LI NYYVSEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>A.coerulea</i>	NP AY AK LI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGVFTNVI SPTTKAK	
<i>A.linifolium</i>	NSEYSKLV DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>A.lyrata</i>	NSEYSKLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>A.officinalis</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSTKAK	
<i>A.thaliana</i>	NSEYSKLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>A.trichopoda</i>	NP DY AK LI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSAKAK	
<i>B.distachyon</i>	NP DY DK LV NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>B.oleraceacapitata</i>	NSEYSKLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>B.platyphylla</i>	NPEYNKLI NYYVSEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTSKAK	
<i>B.rapa</i>	NSEYSKLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>B.stacei</i>	NP DY DK LV NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>B.stricta</i>	NSEYNKLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>B.vulgaris</i>	NAEYDKLI QYYVREKYTL RY TGGMVP DV NQL I VKEKGI FTNVASPSAKAK	
<i>C.arabica</i>	NP DY AK LI DYYVREKYTL RY TGGMVP DV NQ I I VKEKGI FTNVASPSAKAK	
<i>C.arietinum</i>	NANYAKLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>C.citriodora</i>	NP DY GK LI EYYVKEKYTL RY TGGMVP DV NQ I I VKEKGVFTNVVSPSAKAK	
<i>C.clementina</i>	NP DY DK LI NYYVKQKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSSKAK	
<i>C.dentata</i>	NP DY NK LI NYYVNEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTSKAK	
<i>C.grandiflora</i>	NSEYSTLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>C.illinoensis</i>	NP DY NK LI NYYVNEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSTKAK	
<i>C.kanehirae</i>	NP AY EK LI NYYVREKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>C.laxum</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSAKAK	
<i>C.papaya</i>	NPEYDKLI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGVFTNVI SPSSKAK	
<i>C.purpureus</i>	NPEYEKLI NYYVSEKYTL RY TGGMVP DV NQ I I VKERGI FTNVI SPTTKAK	
<i>C.quinoa</i>	NP DY DK LI QYYVREKYTL RY TGGMVP DV NQ I I VKEKGI FTNVASPSAKAK	
<i>C.reinhardtii</i>	NP AY ER LI NFYLGEKYTL RY TGGMVP DV FQ I I VKEKGVFTNVT SPTTKAK	
<i>C.richardii</i>	NPEYEKLI NYYVGEKYTL RY TGGMVP DV NQ I I VKEKGVFTNVSSPSTKAK	
<i>C.sativus</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSTKAK	
<i>C.subellipsoidea</i>	NP NY ER LI SHYI GEKYTL RY TGGMVP DV FQ I I VKEKGVFTNVT SPSTKAK	
<i>C.violacea</i>	NPEYSKLV DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>C.zofingiensis</i>	NP AY ER LMSYYLGEKYTL RY TGGMVP DV FQ I I VKEKGVFTNVT SPSTKAK	
<i>D.alata</i>	NP AY SN LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSSKAK	
<i>D.complanatum</i>	NP DY EK LI NYYVSEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTTKAK	
<i>E.coracana</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSAKAK	
<i>E.grandis</i>	NP DY GK LI EYYVKEKYTL RY TGGMVP DV NQ I I VKEKGVFTNVVSPSAKAK	
<i>E.salsugineum</i>	NSEYSKLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTAKAK	
<i>E.syracum</i>	NSEYSKLI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPTARAK	
<i>F.angustifolia</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.anomala</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.australica</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.bidentis</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.brownii</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.chlorifolia</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.cronquistii</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.floridana</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.kochiana</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.palmeri</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.pringlei</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.pubescens</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.ramosissima</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.robusta</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.sonorensis</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.trinervia</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.vaginata</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	
<i>F.vesca</i>	NP DY GK LI DYYVKEKYTL RY TGGMVP DV NQ I I VKEKGVFTNVI SPTTKAK	
<i>G.barbadense</i>	NPEYDKLI NYYVREKYTL RY TGGMVP DV NQ I I VKEKGI FTNVASPSAKAK	
<i>G.max</i>	NP DY AK LI DYYVNEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVT SPSAKAK	
<i>H.quercifolia</i>	NP DY DK LI NYYVKEKYTL RY TGGMVP DV NQ I I VKEKGI FTNVI SPTTKAK	

201	<i>I.amara</i> <i>I.tinctoria</i> <i>J.ascendens</i> <i>K.fedtschenkoi</i> <i>K.laxiflora</i> <i>L.albus</i> <i>L.annua</i> <i>L.cullinaris</i> <i>L.ervoides</i> <i>L.philippensis</i> <i>L.satīva</i> <i>L.satīvum</i> <i>L.usitatissimum</i> <i>M.acuminata</i> <i>M.domestica</i> <i>M.esculenta</i> <i>M.guttatus</i> <i>M.maritima</i> <i>M.perfoliatum</i> <i>M.polymorpha</i> <i>M.sinensis</i> <i>M.truncatula</i> <i>N.colorata</i> <i>N.tabacum</i> <i>O.satīva</i> <i>O.thomaeum</i> <i>P.acutifolius</i> <i>P.amilis</i> <i>P.hallii</i> <i>P.latifolius</i> <i>P.lunatus</i> <i>P.persica</i> <i>P.trichocarpa</i> <i>P.vaginatūm</i> <i>P.vulgaris</i> <i>Q.rubra</i> <i>R.communis</i> <i>S.alba</i> <i>S.bicolor</i> <i>S.italica</i> <i>S.moellendorffii</i> <i>S.oleracea</i> <i>S.parvula</i> <i>S.tuberosum</i> <i>S.viridis</i> <i>T.arvense</i> <i>T.cacao</i> <i>T.intermedium</i> <i>T.plicata</i> <i>T.pratense</i> <i>U.fusca</i> <i>V.carteri</i> <i>V.darrowii</i> <i>V.unguiculata</i> <i>V.vinifera</i> <i>Z.marina</i> <i>Z.mays</i>	NSEYNKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NSEYSKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPEYAKLI DFYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYAKLI EFYI KEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPEYEKLVNYYVRQKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NSEYSKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPTAKAK NTNYAKLV DYYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NTNYAKLV DYYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYAKLI DYYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPSAKAK NPDYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGV FTNVI SPTTKAK NSEYSKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYEKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPTTKAK NPDYDKLI NYVYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSSKAK NPDYDKLI SYVYVKEKYTL RYTGGMVPDVNQI I VKEKGV FTNVI SPTTKAK NPDYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPSKAK NPDYAKLI DYYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPTAKAK NSEYSKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPEYEKLI NYVYVSEKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPTTKAK NTEYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NKDYAKLV DYYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPEYQKLI NYVYVNEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPTSKAK NADYAKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPEYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPEYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSAKAK NVNYAKLI DYYVNEKYTL RYTGGMVPDVNQI I VKEKGI FTNVASPSAKAK NSDYDKLI QYYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSAKAK NPDYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSSKAK NLDYAKLI DYYVNEKYTL RYTGGMVPDVNQI I VKEKGI FTNVASPSAKAK NPDYSKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPTTKAK NPDYEKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGV FTNVI SPTSKAK NPEYEKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NVNYAKLI DYYVNEKYTL RYTGGMVPDVNQI I VKEKGI FTNVASPSAKAK NPDYKLVNYYVNEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPTSRAK NPDYDKLI NYVYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPSKAK NSEYSKLV DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPEYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSAKAK NPQYEKLI NYVYVTERYTL RYTGGMVPDVNQI I VKEKGV FTNVTSPSTKAK NADYAKLI QYYI KEKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPTAKAK NSEYSKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYAKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVI SPTAKAK NPDYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSAKAK NSEYSKLI DYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPEYDKLI NYVYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVASPSAKAK NPDYDKLVNYYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYEKLI NYVYVSEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSAKAK NANYAKLV DYYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NPDYDKLI NYVYVKEKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSAKAK NPAYERLI NFYLG EKYTL RYTGGMVPDV FQI I VKEKGV FTNVTSPPTTKAK NPDYNKLI NYVYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPATAKAK NLDYAKLI DYYVNEKYTL RYTGGMVPDVNQI I VKEKGI FTNVASPSAKAK NPDYDKLI NYVYVREKYTL RYTGGMVPDVNQI I VKEKGI FTNVTSPSSKAK NSNYAKLV DYYVREKYTL RYTGGMVPDVNQI I VKEKGV FTNVI SPSAKAK NPEYDKLI NYVYVKEKYTL RYTGGM- - - - - I VKEKGI FTNVTSPATAKAK	250
-----	--	---	-----

251	300
<i>A.americanus</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G T Q - S V L D R V I H S L D D R T Q V A Y G S K N
<i>A.coerulea</i>	L R L L F E V A P L G F L V E K A G G Y S S - D G K Q - S V L D K V I N V L D E R T E V A Y G S K N
<i>A.linifolium</i>	L R L L F E V A P L G L L I E N A G G F S S - D G H K - S V L D K T I I N L D D R T Q V A Y G S K N
<i>A.lyrata</i>	L R L L F E V A P L G L L I E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>A.officinalis</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G K Q - S V L D K V I S G L D E R T Q V A Y G S K N
<i>A.thaliana</i>	L R L L F E V A P L G L L I E N A G G F S S - D G H K - S V L D K T I I N L D D R T Q V A Y G S K N
<i>A.trichopoda</i>	L R L L F E V A P L G L L V E K A G G Y S S - D G K Q - S V L D K V I S T L D E R T Q V A Y G S K N
<i>B.distachyon</i>	L R L L F E V A P L G F L I E K A G G H S S - D G K Q - S V L D K V I T V L D E R T Q V A Y G S K N
<i>B.oleraceacapitata</i>	L R L L F E V A P L G L L V E N A G G F S S - D G Y N - S V L D K T I I N L D D R T Q V A Y G S K N
<i>B.platyphylla</i>	L R L L F E V A P L G L L V E Q A G G Y S S - D G H Q - S V L D K V I H N L D D R T Q V A Y G S K N
<i>B.rapa</i>	L R L L F E V A P L G L L V E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>B.stacei</i>	L R L L F E V A P L G F L I E K A G G H S S - D G K Q - S V L D K V I T V L D E R T Q V A Y G S K N
<i>B.stricta</i>	L R L L F E V A P L G L L I E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>B.vulgaris</i>	L R L L F E V A P L G L L I E K A G G Y S S - D G T Q - S I L D K V I E N L D D R T Q V A Y G S K N
<i>C.arabica</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G N K - S V L D K V I N S L D D R T Q V A Y G S K N
<i>C.arietinum</i>	L R L L F E V A P L G L L I E K A G G Y S S - D G H K - S V L E K V I E N I D D R T Q V A Y G S K N
<i>C.citriodora</i>	L R L L F E V A P L G F L I E K A G G F S S - D G K Q - S V L D K V I N N L D D R T Q V A Y G S K N
<i>C.clementina</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G K I - S V L D K V I N D L D D R T Q V A Y G S K N
<i>C.dentata</i>	L R L L F E V A P L G L L I E K A G G Y S S - D G K Q - S I L D K V I N N L D D R T Q V A Y G S K N
<i>C.grandiflora</i>	L R L L F E V A P L G L L I E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>C.illinoensis</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G R Q - S V L D K V I N N L D D R T Q V A Y G S K N
<i>C.kanehirae</i>	L R L L F E V A P L G F L V E K A G G F S S - D G T Q - S V L E K V I V S L D D R T Q V A Y G S K N
<i>C.laxum</i>	L R L L F E V A P L G F L I E K A G G H S S - D G K Q - S V L D K V I N E L D E R T Q V A Y G S K N
<i>C.papaya</i>	L R L L F E V A P L G Y L I E K A G G Y S S - D G Y Q - S V L D K T V N N L D D R T Q V A Y G S K N
<i>C.purpureus</i>	L R L L F E V A P L A L L I E N A G G Y S S - D G T Q - S V L D R E V I N T D D R T Q V A Y G S R D
<i>C.quinoa</i>	L R L L F E V A P L G F L I E K A G G H S S - D G T K - S V L D I V V Q N L D D R T Q V A Y G S K N
<i>C.reinhardtii</i>	L R I L F E V A P L A L L I E K A G G A S S C D G K A V S A L D I P I L V C D Q R T Q I C Y G S I G
<i>C.richardii</i>	L R L L F E V A P L G F L I E K A G G F S S - D G T Q - S V L D K V V K S T D D R T Q V A Y G S K Q
<i>C.sativus</i>	L R L L F E V A P L G F L V E K A G G Y S S - D G R Q - S V L D K V I E N L D E R T Q V A Y G S K E
<i>C.subellipsoidea</i>	L R I L F E V A P L A L L I E K A G G H S S C D G L L V S G L D V E I T A H D Q R T Q I C Y G S K A
<i>C.violacea</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G Y Q - S V L D R T I N N L D D R T Q V A Y G S K N
<i>C.zofingiensis</i>	L R I L F E V A P L A L M V E N A G G A S S C D G K S V S A L D I P I L G Y D Q R T E I C F G S I G
<i>D.alata</i>	L R L L F E V A P L G F L I E K A G G H S S - D G E K - S V L D K V I T S L D E R T Q V A Y G S K N
<i>D.complanatum</i>	L R L L F E V A P L G L L I E K A G G Y S S - D G K Q - S V L D K V I I N T D D R T Q V A Y G S K N
<i>E.coracana</i>	L R L L F E V A P L G L L I E K A G G H S S - D G K Q - S V L D K V I N E L D E R T Q V A Y G S K N
<i>E.grandis</i>	L R L L F E V A P L G F L I E K A G G F S S - D G K Q - S V L D K V I I N L D D R T Q V A Y G S K N
<i>E.salsugineum</i>	L R L L F E V A P L G L L V E N A G G F S S - D G Y K - S V L D K T I I N L D D R T Q V A Y G S K N
<i>E.syracum</i>	L R L L F E V A P L G L L V E N A G G F S S - D G Y N - S V L D K I I V N L D D R T Q V A Y G S K N
<i>F.angustifolia</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.anomala</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.australica</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.bidentis</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.brownii</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.chlorifolia</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.cronquistii</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G T Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.floridana</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.kochiana</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G T Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.palmeri</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.pringlei</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.pubescens</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.ramosissima</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.robusta</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I A N L D D R T Q V A Y G S K N
<i>F.sonorensis</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G T Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.trinervia</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.vaginata</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>F.vesca</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I V N L D D R T Q V A Y G S K N
<i>G.barbadense</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G Q K - S V L D K V I E N L D D R T Q V A Y G S K N
<i>G.max</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G H Q - S V L D K V I T N I D E R T Q V A Y G S K N
<i>H.quercifolia</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G K Q - S V L D K V V H N L D D R T Q V A Y G S K N

<i>I.amara</i>	L R L L F E V A P L G L L V E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>I.tinctoria</i>	L R L L F E V A P L G L L I E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>J.ascendens</i>	L R L L F E V A P L G F L V E K A G G Y S S - D G K Q - S V L D K V I N N L D E R T Q V A Y G S K N
<i>K.fedtschenkoi</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G T K - S V L D R V V N V L D E R T Q V A F G S K N
<i>K.laxiflora</i>	L R L L F E V A P L G Y L I E K A G G Y S S - D G T Q - S V L D K V I N F L D E R S Q V A Y G S K N
<i>L.albus</i>	L R L L F E V A P L G L L I E K A G G Y S S - D G H Q - S V L D K V I S N I D E R T Q V A Y G S K N
<i>L.annua</i>	L R L L F E V A P L G L L V E N A G G F S S - D G Y K - S V L D K T I I N L D D R T Q V A Y G S K N
<i>L.cullinaris</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G Y K - S V L D K V V E N I D D R T Q V A Y G S K N
<i>L.ervoides</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G Y K - S V L D K V V E N I D D R T Q V A Y G S K N
<i>L.philippensis</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G K Q - S V L H K V I T N L D E R T Q V A Y G S K N
<i>L.satīva</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G T Q - S V L D K V I V N L D D R T Q V A Y G S L N
<i>L.satīvum</i>	L R L L F E V A P L G L L I E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>L.usitatissimum</i>	L R L L F E V A P L G L L V E N A G G Y S S - D G H Q - S V L D K V I K E L D D R T Q V A Y G S K N
<i>M.acuminata</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G T R - S V L D K V I E N L D D R T Q V A Y G S K N
<i>M.domestica</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G Q Q - S V L D K V V V N L D D R T Q V A Y G S K N
<i>M.esculenta</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G H Q - S V L D K E I K N L D D R T Q V A Y G S K N
<i>M.guttatus</i>	L R L L F E V A P L G F L I E K A G G F S S - D G K Q - S V L D K V I I N L D D R T Q V A Y G S K N
<i>M.maritima</i>	L R L L F E V A P L G L L I E N A G G F S S - D G H K - S V L D K T I I N L D D R T Q V A Y G S K N
<i>M.perfoliatum</i>	L R L L F E V A P L G L L I E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>M.polymorpha</i>	L R L L F E V A P L G L L V E K A G G Y S S - D G K I - S V L D K V V V N T D D R T Q V A Y G S R N
<i>M.sinensis</i>	L R L L F E V A P L G F L M E K A G G H S S - D G K Q - S V L D K V I N E L D E R T Q V A Y G S K N
<i>M.truncatula</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H K - S V L D K V I E N I D D R T Q V A Y G S K N
<i>N.colorata</i>	L R L L F E V A P L G F L V E K A G G Y S S - D G K Q - S V L D R V I E N L D D R T Q V A Y G S K N
<i>N.tabacum</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G K Q - S V L D K V I G T L D E R T Q V A Y G S K N
<i>O.satīva</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G K Q - S V L D K V I N N L D E R T Q V A Y G S K N
<i>O.thomaeum</i>	L R L L F E V A P L G L L I E K A G G H S S - D G K Q - S V L D K V I N E L D E R T Q V A Y G S K N
<i>P.acutifolius</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G Y Q - S V L D K V I N N I D E R T Q V A Y G S K N
<i>P.amilis</i>	L R L L F E V A P L G L L V E N A G G Y S S - D G A Q - S V L D K V I N N L D E R T Q V A Y G S K N
<i>P.hallii</i>	L R L L F E V A P L G F L I E K A G G H S S - D G K Q - S V L D R V I N E L D E R T Q V A Y G S K N
<i>P.latifolius</i>	L R L L F E V A P L G L L I E K A G G Y S S - D G K Q - S V L N K V I N N L D E R T Q V A Y G S K N
<i>P.lunatus</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G Y Q - S V L D K V I N N I D E R T Q V A Y G S K N
<i>P.persica</i>	L R L L F E V A P L G L L V E N A G G F S S - D G H Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>P.trichocarpa</i>	L R L L F E V A P L G L L V E K A G G Y S S - D G H K - S V L D K E I I N L D D R T Q V A Y G S K N
<i>P.vaginatūm</i>	L R L L F E V A P L G L L V E K A G G H S S - D G K Q - S V L D K V I N E L D E R T Q V A Y G S K N
<i>P.vulgaris</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G Y Q - S V L D K V I N N I D E R T Q V A Y G S K N
<i>Q.rubra</i>	L R L L F E V A P L G L L I E K A G G Y S S - D G K Q - S I L D K V I N N L D E R T Q V A Y G S K N
<i>R.communis</i>	L R L L F E V A P L G L L I E K A G G Y S S - D G F Q - S V L D K E I K G L D D R T Q V A Y G S K D
<i>S.alba</i>	L R L L F E V A P L G L L I E N A G G F S S - D G H K - S V L D K T I I N L D D R T Q V A Y G S K N
<i>S.bicolor</i>	L R L L F E V A P L G L L I E K A G G H S S - D G K Q - S V L D K V I N E L D E R T Q V A Y G S K N
<i>S.italica</i>	L R L L F E V A P L G F L M E K A G G H S S - D G K Q - S V L D K V I N E L D E R T Q V A Y G S K N
<i>S.moellendorffii</i>	L R L L F E V A P L G L L I E N A G G F S S - D G H Q - S V L D K L V V N T D D R T Q V A Y G S K A
<i>S.oleracea</i>	L R L L F E V A P L G F L I E K A G G H S S - E G T K - S V L D I E V K N L D D R T Q V A Y G S L N
<i>S.parvula</i>	L R L L F E V A P L G L L I E N A G G F S S - D G H K - S V L E K T I I N L D D R T Q V A Y G S K N
<i>S.tuberosum</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G K Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>S.viridis</i>	L R L L F E V A P L G F L M E K A G G H S S - D G K Q - S V L D K V I N E L D E R T Q V A Y G S K N
<i>T.arvense</i>	L R L L F E V A P L G L L I E N A G G F S S - D G Y K - S V L D K T I V N L D D R T Q V A Y G S K N
<i>T.cacao</i>	L R L L F E V A P L G F L I E K A G G Y S S - D G H R - S V L D K L I N N L D D R T Q V A Y G S K N
<i>T.intermedium</i>	L R L L F E V A P L G F L I E K A G G H S S - D G K Q - S V L D K V I S V L D E R T Q V A Y G S K N
<i>T.plicata</i>	L R L L F E V A P L G L L I E K A G G H S S - D G K Q - S V L D R V I I N V D D R T Q V A Y G S K N
<i>T.pratense</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G Y K - S V L D K V I E N I D D R T Q V A Y G S K N
<i>U.fusca</i>	L R L L F E V A P L G F L I E K A G G H S S - D G K Q - S V L D R V I N E L D E R T Q V A Y G S K N
<i>V.carteri</i>	L R I L F E V A P L A L L V E K A G G A S S C D G K C V S A L D I P I L N C D Q R T Q I C Y G S I G
<i>V.darrowii</i>	L R L L F E V A P L G F L I E Q A G G H S S - D G K Q - S V L D K V I V N L D D R T Q V A Y G S K N
<i>V.unguiculata</i>	L R L L F E V A P L G L L I E N A G G Y S S - D G H Q - S V L D K V I N N I D E R T Q V A Y G S K N
<i>V.vinifera</i>	L R L L F E V A P L G F L I E K A G G F S S - D G T Q - S V L D K V I N N L D E R T Q V A Y G S K N
<i>Z.marina</i>	L R L L F E V A P L G F L I E K A G G H S S - D G T Q - S V L D K V I T A L D E R T Q V A Y G S K N
<i>Z.mays</i>	L R L L F E V A P L G F L M E K A G G Y S S - D G K Q - S V L D R V I N E L D E R T Q V A Y G S K N

301	326
<i>A.americanus</i>	E I I R F E E T L Y G S S R L K - - S G A P V G A T
<i>A.coerulea</i>	E I I R F E E T L Y G S S R L K - - A G E P V G A A
<i>A.linifolium</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>A.lyrata</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>A.officinalis</i>	E I I R F E E T L Y G S S R L K - - A G A P V G A A
<i>A.thaliana</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>A.trichopoda</i>	E I I R F E E T L Y G S S R L K - - A G T P L A A A
<i>B.distachyon</i>	E I I R F E E T L Y G S S R L A - - A G A T V G A T
<i>B.oleraceacapitata</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>B.platyphylla</i>	E I I R F E K T L Y G S S R L E - - A G V P V G A A
<i>B.rapa</i>	E I I R F E E T L Y G T S R L K - - - N V P I G A T
<i>B.stacei</i>	E I I R F E E T L Y G T S R L A - - A G A T V G A T
<i>B.stricta</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>B.vulgaris</i>	E I I R F E K T L Y G S S R L E - - E P V P V G A A
<i>C.arabica</i>	E I I R F E E T L Y G S S R L K - - A P E P V G A A
<i>C.arietinum</i>	E I I R F E E T L Y G S S R L Q - - G G V P V G A T
<i>C.citriodora</i>	E I I R F E E M L Y G S S R L K - - A G V P V G A A
<i>C.clementina</i>	E I I R F E E T L Y G S S R L K - - G G V P V G A A
<i>C.dentata</i>	E I I R F E K T L Y G S S R L E - - A G V P V G A A
<i>C.grandiflora</i>	E I I R F E E T L Y G S S R L K - - - T A P I G V T
<i>C.illinoensis</i>	E I V R F E K T L Y G S S R L E - - A G V P V G A A
<i>C.kanehirae</i>	E I I R F E E T L Y G S S R L K - - A G A P V G A A
<i>C.laxum</i>	E I I R F E D T L Y G S S R L T - - A G A T V G A A
<i>C.papaya</i>	E I I R F E E T L Y G S S R L K - - S G I P V G A A
<i>C.purpureus</i>	E I I R F E E T L F G A S R L K - - - - A E L G V A
<i>C.quinoa</i>	E I I R F E K T L Y G S S R L E - - E P V P V G A A
<i>C.reinhardtii</i>	E V R R F E E Y M Y G T S P R F - - - - - S E K V A
<i>C.richardii</i>	E I I R F E E T L Y G S S R L S - - - S V L S A A A
<i>C.sativus</i>	E I I R F E E T L Y G S S R F K - - S G V P V G A A
<i>C.subellipsoidea</i>	E V Q R F E E Y V Y G Q S P R F - - - - A E V E A N
<i>C.violacea</i>	E V I R F E E T L Y G S S R I N - A D G V P V G A A
<i>C.zofingiensis</i>	E V R R F E E Y M Y G T S P R F - - S E K K A S V A
<i>D.alata</i>	E I I R F E E T L Y G S S R L K - - A V E P V G A A
<i>D.complanatum</i>	E I I R F E E T L Y G A S R L A - - - K V P A Q A V
<i>E.coracana</i>	E I I R F E E T L Y G S S R L A - - A G A T V G A A
<i>E.grandis</i>	E I I R F E E M L Y G S S R L K - - A G V P V G A A
<i>E.salsugineum</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>E.syracum</i>	E I I R F E E T L Y G K S R L K - - - N V P I G V T
<i>F.angustifolia</i>	E I I R F E E M L Y G S S R L K - - - G V P V G A S
<i>F.anomala</i>	E I I R F E E M L Y G S S R L K - - - G V P V G A S
<i>F.australiana</i>	E I I R F E E M L Y G S S R L K - - - N V P I G A S
<i>F.bidentis</i>	E I I R F E E M L Y G S S R L K - - - N V P V G A S
<i>F.brownii</i>	E I I R F E E M L Y G S S R L K - - - N V P I G A S
<i>F.chlorifolia</i>	E I I R F E E M L Y G S S R L K - - - N V P I G A S
<i>F.cronquistii</i>	E I I R F E E M L Y G S S R L K - - S G V P V G A S
<i>F.floridana</i>	E I I R F E E M L Y G S S R L K - - - N V P I G A S
<i>F.kochiana</i>	E I I R F E E M L Y G S S R L K - - S G V P V G A S
<i>F.palmeri</i>	E I I R F E E M L Y G S S R L K - - - N V P I G A S
<i>F.pringlei</i>	E I I R F E E M L Y G S S R L K - - - G V P V G A S
<i>F.pubescens</i>	E I I R F E E M L Y G S S R L K - - - N V P I G A S
<i>F.ramosissima</i>	E I I R F E E M L Y G S S R L K - - - G V P V G A S
<i>F.robusta</i>	E I V R F E E M L Y G S S R L K - - - N V P I G V T
<i>F.sonorensis</i>	E I I R F E E M L Y G S S R L K - - - G V P V G A S
<i>F.trinervia</i>	E I I R F E E M L Y G S S R L K - - - N V P I G A S
<i>F.vaginata</i>	E I I R F E E M L Y G S S R L K - - - N V P I G A S
<i>F.vesca</i>	E I I R F E E T L Y G S S R L K - - E A V P V G A A
<i>G.barbadense</i>	E I I R F E E T L Y G S S R L N - - A G V P V G A T
<i>G.max</i>	E I I R F E E T L Y G K S R L K - - D G V A V G A A
<i>H.quercifolia</i>	E I I R F E E T L Y G S S R L K - - S G V P V G A A

301	326
<i>I.amara</i>	E I I R F E E T L Y G T S R L K - - - N V P V G A T
<i>I.tinctoria</i>	E I I R F E E T L Y G T S R L K - - - N V P V G V T
<i>J.ascendens</i>	E I I R F E E T L Y G S S R L K - - A G T P V G A A
<i>K.fedtschenkoi</i>	E I I R F E E T L Y G S S R L R - E A G L P V G A A
<i>K.laxiflora</i>	E I I R F E E T L Y G S S R L K - E A G V P V G A A
<i>L.albus</i>	E I I R F E E T L Y G K S R L K - - E G V P V G A S
<i>L.annua</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>L.cullinaris</i>	E I I R F E E T L Y G S S R L K - - G G V P V G A S
<i>L.ervoides</i>	E I I R F E E T L Y G S S R L K - - G G V P V G A S
<i>L.philippensis</i>	E I I R F E E T L Y G S S R L K - - A G Q P V G A A
<i>L.satīva</i>	E I I R F E E T L Y G S S R L K - - S G V P V G A S
<i>L.satīvum</i>	E V I R F E E T L Y G T S R L N - - - N V A V G V T
<i>L.usitatissimum</i>	E I I R F E E T L Y G K S R L K E A G G V P V G A A
<i>M.acuminata</i>	E I I R F E E T L Y G S S R L N - K A G A P V G A A
<i>M.domestica</i>	E I I R F E E T L Y G S S R L K - - - G V P V G A A
<i>M.esculenta</i>	E I I R F E E T L Y G K S R L K - A G G V P V G A A
<i>M.guttatus</i>	E I I R F E E T L Y G S S R L K - - T A A P V G A S
<i>M.maritima</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>M.perfoliatum</i>	E I I R F E E T L Y G T S R L K - - - N V P V G V T
<i>M.polymorpha</i>	E I I R F E E A L H G T S R V A - - - - A L V G A S
<i>M.sinensis</i>	E I I R F E E T L Y G S S R L A - - A G A T V G A A
<i>M.truncatula</i>	E V I R F E E T L Y G S S R L K - - G G V P V G A S
<i>N.colorata</i>	E I I R F E E T L Y G S S R L K - - V G T P V G A A
<i>N.tabacum</i>	E I I R F E E T L Y G S S R L K - - A A E P V G A A
<i>O.satīva</i>	E I I R F E E T L Y G S S R L T - - A G A T V G A A
<i>O.thomaeum</i>	E I I R F E E T L Y G S S R L A - - A G A T V G A A
<i>P.acutifolius</i>	E I I R F E E T L Y G K S R L K - - G G V A V G A A
<i>P.amilis</i>	E V L R F E K T L Y G S S R L E - - E P V P V G S S
<i>P.hallii</i>	E I I R F E E T L Y G S S R L T - - A G A T V G A A
<i>P.latifolius</i>	E I I R F E E T L Y G S S R L K - - A G A P V G A A
<i>P.lunatus</i>	E I I R F E E T L Y G K S R L K - - G G V A V G A A
<i>P.persica</i>	E I I R F E E T L Y G S S R L K - - E A V P V G A A
<i>P.trichocarpa</i>	E I I R F E E T L Y G K S R L K - A E G V P V G A A
<i>P.vaginatū</i>	E I I R F E E T L Y G S S R L A - - A G A T V G A A
<i>P.vulgaris</i>	E I I R F E E T L Y G K S R L K - - G G V A V G A A
<i>Q.rubra</i>	E I I R F E K T L Y G S S R L E - - A G V P V G A A
<i>R.communis</i>	E I I R F E E T L Y G K S R L K - S G G V P V G A T
<i>S.alba</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>S.bicolor</i>	E I I R F E E T L Y G S S R L A - - A G A T V G A A
<i>S.italica</i>	E I I R F E E T L Y G S S R L T - - A G A T V G A A
<i>S.moellendorffii</i>	E I I R F E E T L Y G S S R L A - - - K V A A A K L
<i>S.oleracea</i>	E I I R F E K T L Y G S S R L E - - E P V P V G A A
<i>S.parvula</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>S.tuberosum</i>	E I I R F E E T L Y G S S R L K - - A G A P V G A A
<i>S.viridis</i>	E I I R F E E T L Y G S S R L T - - A G A T V G A A
<i>T.arvense</i>	E I I R F E E T L Y G T S R L K - - - N V P I G V T
<i>T.cacao</i>	E I I R F E E T L N G S S R L K - - A G V P V G A A
<i>T.intermedium</i>	E I I R F E E T L Y G S S R L A - - A S A T V G A T
<i>T.plicata</i>	E I I R F E E T L Y G S S R L S - L K G A P V G A K
<i>T.pratense</i>	E I I R F E E T L Y G S S R L K - - G G V P V G A S
<i>U.fusca</i>	E I I R F E E T L Y G S S R L T - - A G A T V G A A
<i>V.carteri</i>	E V R R F E E Y M Y G N S P R F - - - - - S Q V T
<i>V.darrowii</i>	E I I R F E E T L Y G S S R L K - - G G V P V G A A
<i>V.unguiculata</i>	E I I R F E E T L Y G K S R L K - - G G V A V G A A
<i>V.vinifera</i>	E I I R F E E T L Y G S S R L K - - A G V P V G A A
<i>Z.marina</i>	E I I R F E E T M Y G S S R L N - K T S S S V G A V
<i>Z.mays</i>	E I I R F E E T L Y G S S R L A A S A T A T A R A L

Figure 2. Multiple protein sequence alignment of SBPase coding sequences from 116 photosynthetic species with the transit peptide region and last two residues removed. Residue numbers at the top of the alignments are in reference to the edited *A. thaliana* SBPase sequence. DNA alignment available at this link:

<https://essexuniversity.box.com/s/0nzju6pmxxx813iqquanhbyr31ckh2x5>