**PROJECT TITLE Bioprospecting plant proteomes for more efficient photosynthesis enzymes**

**AIMS AND BACKGROUND**

The goal of this project is to develop an entirely new way to screen hundreds of plants for enzyme variants that have atypical, and therefore potentially useful, kinetic characteristics. The resulting methods will be able to quickly screen any species of land plant, and therefore will accelerate the development of new plant varieties across agricultural industries.

**The fundamental question of this project is: In conserved metabolic pathways, does abnormal isoform abundance indicate atypical enzyme characteristics?**

The two-step approach, reflected in the project aims below, will first quantify patterns of photosynthesis protein abundance and, second, characterize in greater detail the outliers from those patterns. Although the methods could be applied to many different plant proteins, this project will focus on discovering atypical variants of two key photosynthesis enzymes, Rubisco and Rubisco activase.

**Aim 1: Determine the genetic and environmental drivers of variation in photosynthesis enzyme abundance across native Australian plants.**

**Aim 2: Test the hypothesis that unusual isoform abundance relative to metabolically linked enzymes are predictors of atypical enzyme activities.**

The photosynthesis proteins in 700 leaf samples from different genotypes of plantation-grown and wild *Eucalytpus camaldulensis* (river red gum) will be measured by mass spectrometry and compared against each other and over 1000 samples of wild eucalypts and other Australian native plants. The Rubisco and Rubisco activase from outlier genotypes will then be quantified and kinetically characterized at the isoform level.

**Background**

Photosynthesis is important for climate modelling and crop productivity

Aim 1 relevant to climate modelling

Both aims relevant to improving crops

Rubisco and activase are high priority targets for crop improvement. Activity varies by spp and environment. Atwell’s work.

Australia a good place for it (Sc/o paper)

Quant proteomics not yet applied to large multi-species plant screens. Isoform quant across spp has been used in other fields, including industrially. Combining proteomics methods from medical research.

**PROJECT QUALITY AND INNOVATION**

This project will innovate in three ways:

1. *It will produce new knowledge about the influences of genetics and environment on the absolute amounts of key photosynthesis enzymes and their isoforms.* Recently developed, rapid, absolute protein quantification methods will be used across 120 genotypes of the important plantation timber species *Eucalyptus camaldulensis*, 31 additional eucalypt species, and 70 species of *Acacia* and Proteaceae.
2. *It will quantify the correlations between* in vivo *enzyme amounts and* in vitro *enzyme kinetics parameters.* A subset of Rubisco and Rubisco activase isoforms will be selected based on their abundance relative to other metabolically linked enzymes, and their catalytic parameters measured *in vitro* to establish links between *in vivo* enzymeamounts and enzyme kinetics.
3. *It will establish new proteomics methods, applicable across all vascular plants, for rapidly identifying enzyme isoforms potentially beneficial to crop plant improvement from among hundreds of genotypes.*

Crop scientists are limited in their ability to screen large populations for beneficial enzyme isoforms because existing approaches are time consuming or they imprecisely resolve consequential genetic differences from phylogenetic noise. This project will address those problems by using new proteomics methods I have developed to rapidly find phenotypically consequential isoform variants among large numbers of samples.

The ability to screen large numbers of genotypes by absolute enzyme quantification will complement and improve upon genomic and transcriptomic approaches because physiology is a product of enzyme kinetics and amounts, which are, at best, imprecisely reflected in sequencing data. My approach aims to identify important genetic variants much faster than reverse or forward genetics, or enzyme kinetics assays, based on the assumption that amounts and identities of photosynthesis enzyme isoforms reflect leaf physiology. It uses proteomics to bridge sequence data and physiology: enzyme sequence links to genotype, enzyme amount links to phenotype.

**Theoretical framework**

Two key assumptions underpin this project:

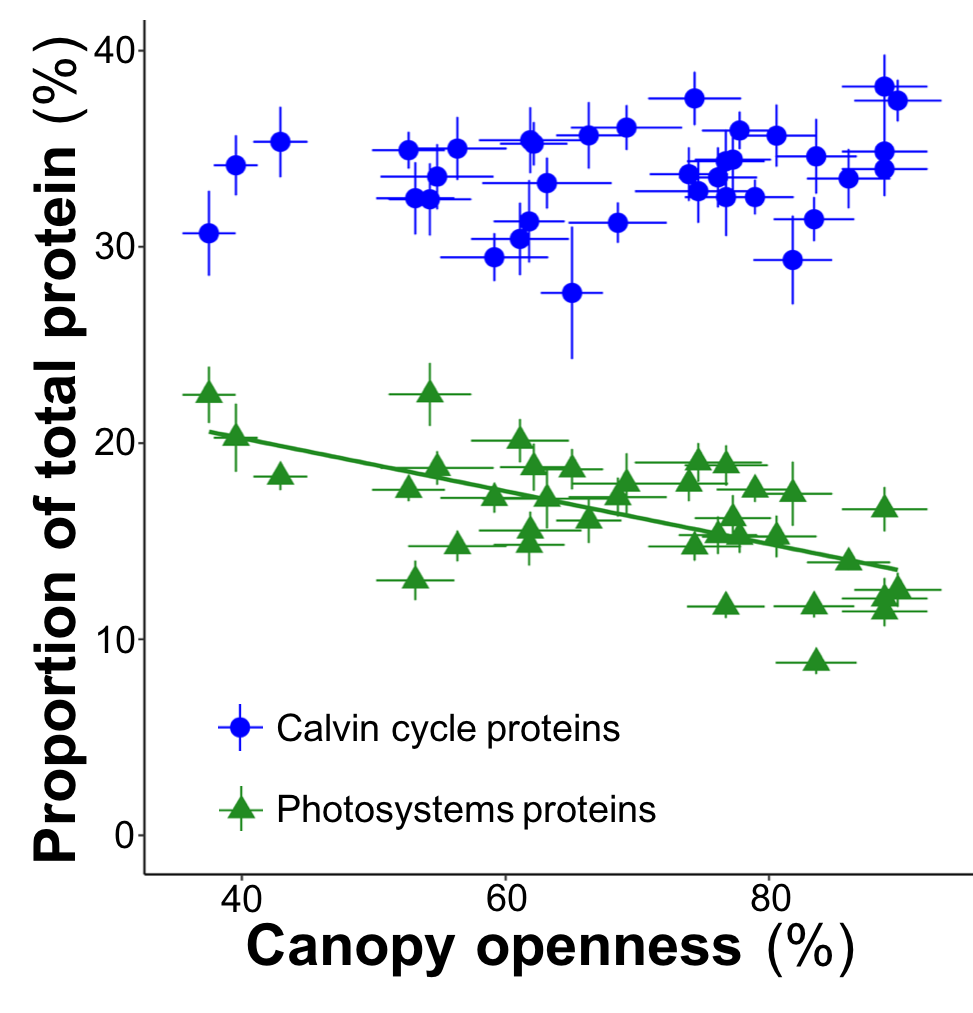
1. *Amounts of photosynthesis enzymes are coordinated so that there are no bottlenecks under average conditions of photosynthesis; no one enzyme amount limits photosynthesis.* This assumption is supported by large-scale physiology and modelling experiments that demonstrate a coordination between electron transport by the light-dependent reactions (*J*max) and carbon fixing by Rubisco (*V*cmax) so that neither is limiting (Maire et al 2012; Walker et al 2014; Quebbeman and Ramirez 2016). The fluxes of ATP and NADPH from the light-dependent reactions occur at rates coordinated with their consumption in the Calvin cycle, thus maximizing carbon assimilation per total investment in photosynthesis proteins (Quebbeman and Ramirez 2016). At times the fluxes might be out of balance (low light, temporary drought), but on average they are coordinated. Photosynthesis protein amounts are not adjusted to environmental conditions instantaneously, but instead reflect environmental conditions for approximately the preceding month (Walker et al 2014; Maire et al 2012).
2. *Amounts of Calvin cycle enzymes relative to each other are conserved, but vary predictably with environmental conditions, leaf traits, and whole plant traits.* Large-scale cross-species conservation of photosynthesis protein amounts relative to each other has not been previously studied, most likely because of technological limitations (Sulpice et al 2010). Enzyme activities differ across *Arabidopsis* genotypes, but those differences covary among related enzymes (Sulpice et al 2010). Amounts of photosynthesis metabolites are conserved across vascular plants and vary predictably with physiological measurements, with some interspecific differences (Gago et al 2016). Those observations suggest that enzyme amounts relative to each should be conserved across species, but vary around those conserved relationships for genetic and environmental reasons.

Combining the assumptions above provides the basis for the experimental approach through the following conclusions:

1. It should be possible to find trends in amounts of Calvin cycle enzymes across multiple species that vary sensibly with environmental conditions and plant traits.
2. Photosynthesis isoforms with atypical kinetic properties inherently contribute differently to the overall flux of photosynthesis metabolites than average isoforms.
3. In order to keep the flux of metabolites coordinated, amounts of kinetically abnormal isoforms, or their metabolic neighbors, must fall outside conserved trends.

*Approach summary*: Conserved trends in photosynthesis enzyme abundance across different levels of genetic diversity will be characterized by regression models of enzyme amounts over ranges of trait and environmental variables. Isoform outliers from the trends will be identified and characterized in kinetics experiments.

Figure 1 contains preliminary data from my current Discovery Project that demonstrates a sensible trend in photosynthesis proteins over a gradient of light availability in 32 eucalypt species sampled across Eastern Australia: protein investment in light capture decreases with increasing light. The lack of concomitant decline in Calvin cycle proteins supports the assumption that metabolic flux is coordinated among the enzymes of photosynthesis.



**Figure 1.** The amount of protein in photosystems, as a proportion of total protein, decreases with increasing light (R2 = 0.39; p < 10-12). Calvin cycle proteins were not correlated with light environment (p = 0.14). Points are species x site means, bars are S.E., 9 measurements per point (3 plants x 3 leaf ages).

Photosystem proteins are not universally expressed at maximal levels for two reasons:

1. A high flux from the photosystems without the capacity to fully utilize them in the Calvin cycle leads to photoinhibition and incurs a high cost of energy dissipation (Walker et al 2014).
2. Allocation is optimized to make the most of a scarce supply of nitrogen among nitrogen-rich proteins (Quebbeman and Ramirez 2016) and protein turnover is a major metabolic expense (Penning de Vries 1975). The second reason is applicable to all leaf proteins and supports the coordination assumption.

Figure 1 also demonstrates that, although there is a significant trend, there is still substantial scatter around that trend, which is the case for most of the protein-environment relationships in my Discovery Project data. Adding additional predictors (e.g. precipitation, temperature, leaf mass per area) in multiple regression models often explains protein variation from 50% to 70%. But, that leaves 30% to 50% unexplained variation, much of it interspecific, which is a major motivator for the proposed project.

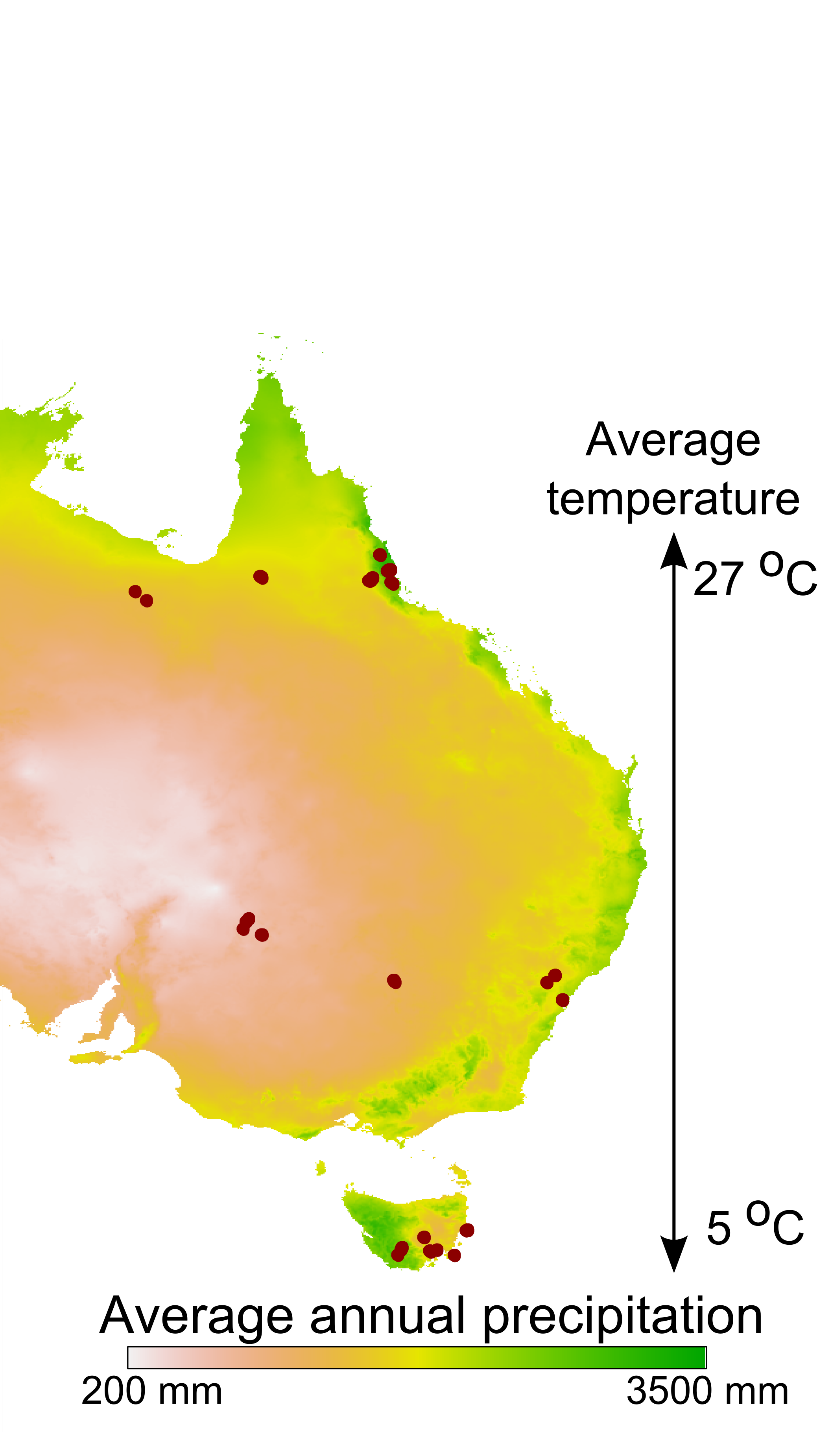
It is practically a circular argument to say that interspecific protein amount variation is caused by genetic differences between species, but that is the case. Those genetic differences sometimes encode species-specific isoforms with potentially atypical kinetics characteristics that could explain interspecific variation in protein abundance. Partitioning the variation in abundance of key photosynthesis enzymes among genetic and environmental predictors is Aim 1 of this proposal. Aim 2 tests the idea that outliers in abundance have atypical kinetic properties.

**Experimental approach**

The proposed project will determine the genetic versus environmental contributions to variation in enzyme amounts of key photosynthesis proteins, down to the isoform level for Rubisco and Rubisco activase. To do so it will make comparisons across:

1. 120 genotypes of *E. camaldulensis* subsp *camaldulensis* in a common garden of 422 genotypes at the Hawkesbury Institute for the Environment (HIE, Richmond, NSW). The collection is part of a Science and Industry Endowment Fund (SIEF) collaboration among HIE, ANU, and CSIRO. The genotypes were sourced from across the natural range of *E. camaldulensis* (Figure 2) at 26 provenances. There are 4 families (mothers) per provenance, 5 seedlings per mother, and several clones per seedling. Physiology data and RNAseq data from the 120 genotypes already exist and will be provided by Prof David Tissue and Dr Paul Rymer (HIE).
2. *E. camaldulensis* mother trees at 18 of the 26 provenances for the HIE garden.
3. 32 species of eucalypts (*Eucalyptus*, *Corymbia*, *Angophora*) from my current Discovery Project sampled from Eastern Australia (Figure 2). All the samples have already been collected and analyzed by the same proteomics methods that will be used in this proposed project.
4. 27 species of *Acacia* and 43 species of Proteaceae from my current Discovery Project sampled from the same sites as the 32 eucalypts, plus additional rainforest sites. The *Acacia* samples have already been analyzed; the Proteaceae will have been analyzed by May 2017.

The two Aims of this proposal both contain two hypotheses. The hypotheses for Aim 1 are:



**Figure 2.** Dots mark sites sampled from my current Discovery Project from which proteomics data will be provided for leaves of eucalypts, Acacia, and Proteaceae. The previously collected samples span two major gradients, temperature and rainfall, that are large enough to cover most of the vegetated climate space on Earth. The range of E. camaldulensis, covered by the 26 provenances of the HIE common garden, is shaded in blue.

*Hypothesis 1A:* *Rubisco and Rubisco activase isoform sequence diversity does not vary substantially across provenances.* This hypothesis is based on previous genomic analysis of the same population, which found that most genetic diversity existed within individuals; i.e. they are highly heterozygous (Dillon et al 2015). There is high gene flow across the population and little genetic structure across provenances. Hypothesis 1B is based on a contrasting result from a physiology experiment with genotypes from the same population.

*Hypothesis 1B: Rubisco and Rubisco activase isoforms vary substantially in abundance across genotypes and environments.* Glasshouse grown *E. camaldulensis* from the HIE collection demonstrated significant physiological differences across 14 genotypes from 6 provenances (Blackman et al 2016). The differences in physiology, in spite of nearly insignificant genetic trends across the population, suggest that proteins and their isoforms are differentially expressed across the genotypes.

Understanding isoform abundance variation across those scales will create fundamental knowledge of how phenotypic variation arises across species and communities. It will also inform future experimental designs by answering the question: *How much does protein variation across genotypes in controlled environments reflect variation in the wild?* In combination with enzyme kinetic measurements it will answer the question: *Can genetically driven isoform variation among wild plants be separated from confounding environmental effects?* This second question relates to the hypotheses of Aim 2.

*Hypothesis 2A: Amounts of Rubisco and Rubisco activase isoforms relative to neighboring enzymes are highly conserved, but vary predictably across environmental gradients.*

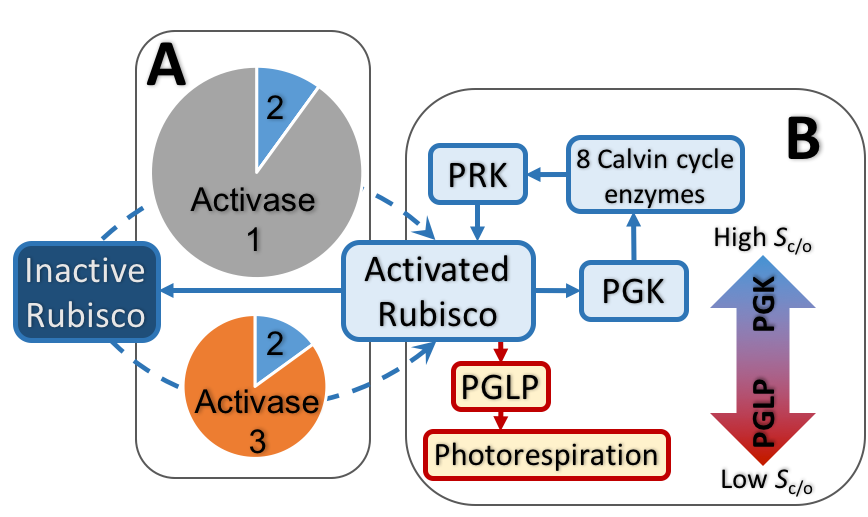
*Hypothesis 2B: Outliers of trends from Hypothesis 2A predict atypical isoform enzyme characteristics.*

Figure 3 demonstrates how ratios of photosynthesis enzymes can be conserved and vary predictably across environmental gradients and how outliers can be detected using absolute protein quantification data. Rubisco activase and Calvin cycle proteins are strongly correlated. However, there is more Rubisco activase relative to Calvin cycle proteins at lower amounts and higher temperatures (1:4) than at higher amounts and lower temperatures (1:5). That trend toward more activase per Calvin cycle at higher temperatures fits with observations that activase is highly temperature sensitive.

Because temperature significantly explains more variation in Rubisco activase than Calvin cycle protein amounts alone, a multiple regression model is more appropriate for detecting outliers than the model in Figure 3, but Figure 3 is shown with one predictor so that it can be visualized more easily. The proposed project will use multiple regression models and mixed models to identify trends and outliers.

Figure 4 outlines the strategy for identifying isoforms with atypical activity. The strategy differs for Rubisco and Rubisco activase—activase kinetics will be inferred from its amount directly; Rubisco specificity will be inferred from amounts of neighboring enzymes.

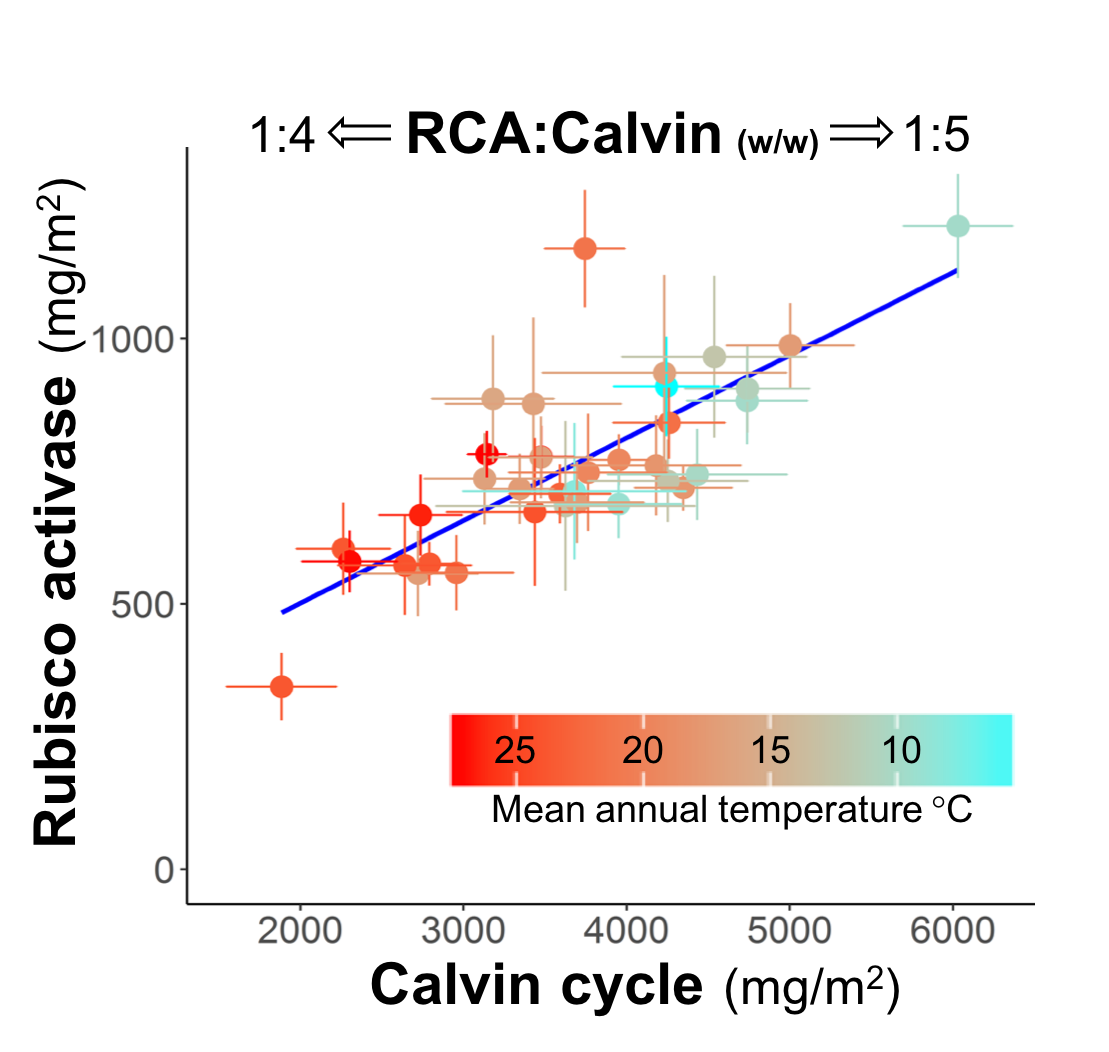
In preliminary results, PGLP and PGK are strongly correlated (R2 = 0.72, not shown) and both protein amounts decrease with increasing temperature. The ratio of PGLP:PGK ranges from 1:10 at high temperatures to 1:15 at low temperatures, which is consistent with decreasing Rubisco specificity and increasing photorespiration with increasing temperature.



**Figure 4.** Enzyme isoform activity will be inferred from protein abundance in two ways: A) by directly measuring the amount of the enzyme of interest relative to its metabolic neighbors (Rubisco activase); B) by indirectly inferring activity from neighboring enzymes in contrasting metabolic pathways (Rubisco oxygenation versus carboxylation activity).

**A:** Pie charts represent different hypothetical activase compositions in different species. The top species has two isoforms, 1 & 2, the bottom species 2 & 3. A comparison of total activase amounts, relative to neighboring enzymes, would predict that the bottom species has a more efficient isoform. Isoform quantification would identify Activase 3 as the most likely explanation for the difference in total activase amount and, therefore, likely a more efficient isoform than isoform 1.

**B:** The efficiency of Rubisco is a product of several kinetic characteristics and activation states, which confound direct inferences based on Rubisco amount. Therefore, this study will make indirect inferences about Rubisco specificity (Sc/o) for carboxylation reactions over oxygenation reactions. Ratios of neighboring enzymes from the subsequent Calvin cycle versus photorespiration pathways will used as indicators of Rubisco specificity. **Shaded arrow:** Genotypes with higher than average ratios of PGLP (phosphoglycolate phosphatase) to PGK (phosphoglycerate kinase) would be predicted to have lower Rubisco specificities and vice versa.



**Figure 3.** Per leaf area amounts of Rubisco activase (RCA) vary predictably with amounts of total Calvin cycle enzymes (R2 = 0.62) in wild eucalypts from broad temperature and precipitation gradients. Ratios vary from 1:4 at low protein amounts to 1:5 at high protein amounts. There is also a trend in temperature with protein amounts, which corresponds with a strong trend of decreasing total protein with increasing temperature (R2 = 0.50, not shown).

There are two clear outliers from the RCA:Calvin cycle trend, which suggests those genotypes potentially have RCA isoforms that differ kinetically from average eucalypt RCA. Points are means (bars = S.E.) for 36 species x site observations, 32 eucalypt species (4 species sampled at 2 sites), 9 samples each, 3 plants x 3 leaf ages.

Once outliers for Rubisco and activase are identified from isoform quantification, then they will be used in kinetics experiments to test Hypothesis 2B and link enzyme kinetics with isoform amounts.

**Methods**

*Protein extraction and mass spectrometry sample preparation.* Leaf protein will be extracted using a modification of Chapman et al (2013). Modifications include: the addition of ovalbumin as an internal standard; leaf powder solvent washes before phenol extraction; phenol extraction at 80 °C, followed by two re-extractions. Precipitated protein will be dissolved in urea-SDS buffer, alkylated, assayed (FluoroProfile kit, Sigma) and 50 μg re-extracted by a modification of Wessel and Flugge (1984), then Lys-C and trypsin digested in a Rapigest (Waters) buffer.

*Mass spectrometry*. Samples will be analyzed on a Sciex TripleTOF mass spectrometer using a SWATH method (Gillet et al 2012) with a 60 min acetonitrile gradient. A reference ion library will be created by running the same method, but in data dependent acquisition (DDA) mode for representative samples of each genotype. The DDA results will be matched, allowing for amino acid mutations (ProteinPilot, Sciex), to predicted protein sequences from *E. grandis* (Myburg et al 2014), with chloroplast and mitochondrial sequences from Bayly et al (2013) and Uniprot.

*RNAseq*. Predicted protein sequences for selected enzymes of interest (Rubisco, Rubisco activase, Calvin cycle and photorespiration enzymes) will be added from RNAseq data provided by Dr Paul Rymer and from RNAseq data for the 12 additional eucalypt species I have studied previously. I have on hand 50 mg leaf samples from the additional eucalypt species that will be used for RNA extraction (Chang et al 1993); cDNA library creation and Illumina NextSeq sequencing will be done by the ANU Biomolecular Resource Facility.

*Absolute protein quantification*. A synthetic protein (QconCAT) I have on hand contains isotope labelled standards for peptides conserved across all vascular plants (57 peptides from 34 proteins, including ovalbumin). It will be added during protein digestions to produce internal peptide standards (Pratt et al 2006; Pertl-Obermeyer et al 2016).

SWATH data will be exported to R and proteins represented in the QconCAT will be quantified by comparing sample peptide areas to the QconCAT peptide areas; corrections for protein losses during sample handing will be calculated based on QconCAT measured ovalbumin amounts versus expected amounts.

Absolute abundance of proteins not represented in the QconCAT will be estimated from SWATH data by comparison to ovalbumin using the top 2 peptide areas and top 2 ions per peptide to minimize the effects of different ionization efficiencies across proteins (Ludwig et al 2012).

*Isoform quantification*. A second QconCAT, manufactured by PolyQuant, will contain ~60 peptides that differentiate Rubisco and Rubisco activase isoforms among the genotypes in this project. The ~700 *E. camaldulensis* samples will be analyzed by SWATH before the creation of the isoform QconCAT, but the SWATH data will be re-interrogated (Liu et al 2013) to calculate isoform abundance based on QconCAT calibration curves created from a subset of representative samples, standardized to ovalbumin. In that way isoform abundance will be calculated retrospectively without having to re-run all 700 *E. camaldulensis* samples and the other 330 eucalypt samples for which I already have SWATH data.

*In vitro measurements of Rubisco catalytic parameters* (*V*cmax, *V*omax, *K*c, *K*o, *k*catc, *k*cato, *S*c/o). Rubisco will be extracted from leaves of 18 species/genotypes using the *Eucalyptus*-specific method of Sharwood et al (2017). Tobacco Rubisco will be used as a reference. The following will be measured at 25 °C, 35 °C, and 45 °C by 14CO2 fixation assays by varying [CO2] and [O2] (Sharwood et al 2017; Sharwood et al 2016 Nature Plants): *V*cmax and *V*omax, the substrate saturated turnover rates for CO2 and O­2, respectively; *K*c and *K*o, the Michaelis constants (*K*m) for CO2 and O­2.

The *in vitro* concentrations of Rubisco active sites will be measured by [14C]carboxyarabinitol-P2 binding (Sharwood et al 2008 Plant Phys). The substrate turnover numbers, *k*catc and *k*cato will be calculated as either *V*cmax or *V*omax divided by the number of Rubisco active sites, respectively.

The CO2/O2 specificity (*S*c/o) of Rubisco purified by anion exchange chromatography (Sharwood et al 2017; Sharwood et al 2008) will be determined at 25 °C, 35 °C, and 45 °C by measuring the amounts of radio-labelled carboxylation and oxygenation products of [1-14C]RuBP by HPLC and scintillation counting (Kane et al 1994).

Rubisco integrity in the enzyme extracts will be analyzed by SDS-PAGE. Rubisco isoform compositions in the enzyme extracts will be quantified by SWATH including isotope labelled internal peptide standards.

*In vitro measurement of recombinant Rubisco activase (RCA) isoform activities.* Twenty Rubisco activase isoforms will be expressed in *E. coli* using commercially synthesized genes (GeneArt). The rates at which they activate Rubisco will be quantified at four temperatures. His-tagged RCA-ubiquitin fusion proteins will be purified by nickel affinity chromatography and the ubiquitin removed enzymatically (Baker et al 2005; Hendrickson et al 2008; Wachter et al 2013). Purified RCA isoforms will be quantified by protein assays and by SWATH.

RCA activity and temperature response will be determined by *in vitro* Rubisco activation across several temperatures (Carmo-Silva and Salvucci 2011; Barta et al 2010). Purified Rubisco from the eucalypt Rubisco *S*c/o measurements and from *Arabidopsis* will be inactivated by pre-incubation with RuBP (forming Rubisco ER). The rate at which RCA activates Rubisco ER will be determined radiometrically by measuring 14CO2 incorporation into acid stable products over time (Barta et al 2010).

*Statistical analysis.* Relationships among isoform abundance, enzyme kinetics, environmental conditions, and genotypes will be evaluated in R using linear regression, multiple linear regression, and linear mixed models.

**DECRA CANDIDATE**

I will be based at the ARC Centre of Excellence for Translational Photosynthesis, working full time. I will work closely with Prof John Evans, A/Prof Spencer Whitney, DECRA fellow Dr Rob Sharwood, Dr Victoria Clarke, and Dr Adam Carroll (head of the ANU Research School of Biology mass spectrometry facility). All provided input into this proposal, Sharwood and Carroll have helped with my current Discovery Project, and Clarke worked for me as a postdoc on my Discovery Project before moving to Evans’ group at ANU.

I have allocated 0.8 FTE to this project because I expect to receive additional funding to extend my proteomics methods to crop research at ANU. I have experience raising funds in both applied and basic research, >$1m since my PhD (2012) in total.

I have experience in all aspects of the project except RNA extraction, which is a well-established method and expertise with it is readily available at ANU.

I have relevant field work experience and personally collected ~4500 leaf samples across Eastern Australia for my current Discovery Project. The fieldwork for the proposed project is much more straightforward than that because most are growing in a plantation and I already have the GPS coordinates for the wild plants..

I developed the quantitative protein extraction method that will be used for the proposed project and have performed it, or have supervised technicians performing it, over 1300 times across over 100 species.

I developed the mass spectrometry methods, including data analysis, and have used it with over 800 samples. I have experience with the required statistical methods through my Discovery Project and also my current grapevine yield modelling project.

My main area of expertise is protein chemistry and I am experienced in protein purification and enzyme kinetics, although assaying Rubisco or Rubisco activase activity will be new to me. As an undergraduate, and three years afterwards, I worked as a volunteer and casual technician in the enzymology lab of Prof Christopher Halkides, Univ North Carolina—Wilmington. Tasks included recombinant protein expression in *E. coli*, purification, and enzyme assays. My honours project studied grape berry enzyme kinetics across species and my PhD involved characterizing grape proteases and recombinant protease by enzymatic assays.

I manage all aspects of my current Discovery Project and grapevine yield project. The Discovery Project is achieving exactly what was outlined in the proposal, perfectly on budget, and will fulfil the project aims by its end in 2017. The grapevine yield project is ahead of schedule and I am already working toward commercializing its outcomes, which was originally planned as a follow up to the project.

My main motivation for this proposal is developing applied research applications for the proteomics methods I developed for the basic research questions of my Discovery Project. I have a track record of delivering real-world outcomes in wine research, I am currently working toward making the conserved peptide standards (QconCAT) commercially available, and I hope to develop my proposed approach to proteome screening at the CoE for Translational Photosynthesis because it is the perfect place to translate my methods into real-world outcomes in crop improvement.

**FEASIBILITY**

I am confident that the four hypotheses of this project will be answered conclusively because the experimental design, comparison across 1700 genotypes, is extremely statistically powerful. The proteomics methods have already detected trends and outliers across eucalypt species as illustrated in Figure 1, specifically with Rubisco activase as illustrated in Figure 3, and with photorespiration versus Calvin cycle enzymes (not shown).

The only uncertainty is whether or not isoform abundance indicates atypical enzyme activity—the fundamental question of this proposal. Using enzyme activities as proxies for protein amounts is commonplace in plant research (e.g. Supice et al 2010) and estimates of protein amounts based on total leaf nitrogen and photosynthesis measurements underpin models from the scale of the cell to the Earth’s surface (Farquhar et al 2001; modelling papers ). At its simplest, this proposal is about making those assumptions in the other direction because the technology now exists to accurately measure many proteins simultaneously, which is much faster than enzymatic assays and adds the important capability of isoform discrimination.

Meeting project goals within the budget and timeline (below) is feasible because I base them on costs and time requirements of my current project, which has already used the same methods to analyze more samples than will be analyzed in the proposed project. There is uncertainty in the time requirement for the enzyme assays; therefore I have allocated nine months to them and proposed a modest number of samples for it. If time and budget allow, then I will use more samples for the kinetics experiments. Critically, the assays will be done in collaboration with Whitney, a world leading expert in the area.



**BENEFIT**

*Output 1: Quantitative knowledge of how genetic versus environmental variables affect photosynthesis enzyme amounts and variation, including isoforms for Rubisco and Rubisco activase.* Quantitative compositions of photosynthesis enzymes, and measures of how their variation is affected by genetics and environment, across large numbers of species is currently unknown. Therefore, this output, as a high-impact publication, will be original in its methodology, its accuracy, and its scope.

*Benefit of Output 1:* Photosynthesis models from the scale of single cells, to leaves, to the entire photosynthetic surface of Earth rely on estimates of nitrogen allocation among different functions of photosynthesis based knowingly on oversimplified and overgeneralized treatment of the individual enzymes (Farquhar et al 2001; modelling papers; UQ systems biol paper; Plant Phys metabolite paper). This output will make possible more accurate, systems biology-based models on all those scales by filling in an entire missing category of key information—actual measured protein amounts.

*Output 2: Two demonstrations, using Rubisco and Rubisco activase, that enzyme and isoform amounts can be used as indicators for enzyme activity in cross-genotype comparisons.*

*Benefit of Output 2:* The findings specific to Rubisco and Rubisco activase will be beneficial for crop development because they increase knowledge of isoform variation in abundance and activity for two of the most important enzyme targets for crop improvement. The output is theoretically significant because it could be applied to screens at different scales and with different methods than this project. The output will benefit metabolic engineering approaches for improving crop plants and it could also be applied to industrial microbes.

*Output 3: The analytical and computational methods for this project will be published in a high-impact methods journal.*

*Benefit of Output 3:* Absolute protein quantification is becoming increasingly common in plant research, but not across multiple species or on hundreds of samples at a time because of technological limitations that my methods overcome. My existing methods are already unique because the internal standard (QconCAT) can be used with any species of vascular plant, the mass spectrometry method is very fast, and the cost per sample is low (1/3 the cost of whole transcriptome sequencing). Extending the methods to identifying potentially beneficial isoforms will dramatically increase the adoption of my methods because they will provide a fast and informative new capability to crop researchers. Crop development will benefit through faster discovery of new beneficial isoforms.

Cost-effective & value for money

**COMMUNICATION OF RESULTS**

Aim 1 will produce basic knowledge about plants across >100 species and most of Earth’s vegetated climate space, which are unprecedented scales for quantitative plant protein data. It will be broadly of interest to plant physiologists, ecologists, and climate scientists. For those reasons I will target high-impact generalist journals such as Nature and PNAS.

Aim 2 will produce outcomes more specific to plant scientists and will be targeted at leading plant journals such as Nature Plants, Plant Physiology, and Plant Cell and Environment.

The theoretical framework behind this project will be applicable to other fields and I will promote it through a commentary article, such as the minireview format of Molecular and Cellular Proteomics.

I will promote my work through seminars in Canberra and Sydney. Since 2014 I have organized the Sydney Plant Ecophysiology group meetings, a seminar series across Sydney universities, and I will promote my research at organisations connected with that group.

In year 3 of the project I will present its results at the meeting of the American Society for Mass Spectrometry and I will arrange to present it at at least one US research institution during the same trip.

**MANAGEMENT OF DATA**

All data and documents will be synced with local drives using purchased Google Drive storage, which is what I use currently. All lab and field notebooks will be scanned regularly and stored in Google Drive.

Mass spectrometry data will be made publicly available by uploading to ProteomeXchange (Vizcaino et al 2014; Deutsch et al 2017), which will make both the mass spectrometry and peptide/protein identification data publicly available through several online resources such as PRIDE, Peptide Atlas, and GPMDB.

Currently, ProteomeXchange does not support geographical or environmental metadata. Therefore, I will make them available as electronic supplementary material associated with articles. I will also include all calculated protein amounts for all samples as supplementary material.

**REFERENCES**