1.) Introducing *ecological proteography*

Over the last decade we have seen a step change in our ability to investigate the molecular bases of biological adaptation and evolution. Rapidly progressing ‘omics’ technologies now allow researchers to characterise complete sets of biological molecules in non-model organisms (Wang, Gerstein & Snyder 2009; Ekblom & Galindo 2010; Armengaud et al. 2014). Extent of protein expression determines an organism’s capacity to perform biochemical functions where the rate at which work is done is a function of the amount of protein doing the work (ref – Michaelis & Menten?). In photosynthesis, for example, light energy captured by light harvesting proteins is used to catalyse carbon uptake from the atmosphere by enzymatic carboxylation. The rate at which plants are able to perform this process is a function of the abundance of photosynthetic proteins in their leaves (Evans 1989?). As such, quantifying protein amounts using proteomics methods provides direct information about how organisms are adapted to their environment (Diz & Calvete 2016).

Proteomics is becoming a useful tool in ecological and evolutionary research because genomic resources are increasingly available for non-model organisms (Armengaud et al. 2014; Baer & Millar 2016; Diz & Calvete 2016). ~~Comparative data on the abundances of proteins or groups of proteins which perform key biochemical functions measured under different environmental conditions could provide mechanistically explicit insight into organism-environment relationships~~. Comparative ecological proteomics requires rapid, scalable methods for absolute quantification of protein abundances in wild organisms. To date however, absolute quantification of plant proteins has proved to be challenging: proteins are difficult to extract from plant tissues and the required data-dependent mass spectrometric methods remain novel (refs from Steve).

We have developed proteomics methods which allow comprehensive extraction of leaf proteins and absolute quantification of the top 2000-3000 most abundant proteins. This allows us to compare protein abundances between samples, which has been demonstrated in model organisms under controlled environments (need refs from Steve) but not in a large-scale study of wild plants.

Using this new technology, we have conducted (a/the first) continent-scale ecological proteomics experiment to characterise the influence of ~~biogeographic and~~ environmental controls on leaf protein expression. We analysed 320 eucalypt leaves across 32 species sampled from Tasmania, New South Wales, and Queensland, spanning large gradients of mean annual precipitation and temperature (200-3200 mm, 5-27 °C, respectively) (Fig. 1).

The resulting dataset describes protein abundances at all levels of functional organisation, from broad groupings down to individual protein subunits. We provide the most complete description to date of leaf protein allocation for all major protein functional categories.

We have concentrated the initial analysis of this dataset on photosynthesis, as it represents one of the most important and abundant set of biochemical reactions in the biosphere (Blankenship & Hartman 1998; Raven 2013). ~~The 500 most abundant proteins account for >90% total leaf protein by weight and the majority of these are involved in photosynthesis and photorespiration (see Fig. X in Methods).~~

Much of what is known about variation in photosynthetic capacity in wild plants is derived from measurements of leaf nitrogen content (Wright et al 2004, Hikosaka 2010), on the basis that photosynthetic proteins comprise the largest pool of nitrogen in leaves (Evans & Seeman 1989, something else). The actual relationship between leaf nitrogen content and photosynthetic carbon assimilation varies substantially, largely in relation to how nitrogen resources are allocated to different functions within photosynthesis (Evans & Seemann 1989, Wright et al 2004). ~~Quantifying these sources of variation has been the focus of substantial research effort since the 1980’s (Niinemets & Tenhunen 1997, Niinemets 2007, lots of others).~~

Mathematical models of photosynthesis describe two important processes in photosynthetic carbon assimilation: carboxylation of ribulose-1,6-bisphosphate (RuBP) by the enzyme Rubisco, and regeneration of RuBP using energetic products derived from the light reactions of photosynthesis (Farquhar, von Caemmerer & Berry 1980; Farquhar, von Caemmerer S & Berry 2001). Theoretically, either of these processes can limit the rate of photosynthesis, depending on whether leaves are light or CO2 limited (REF). Proportional allocation of protein resources to the light-capturing photosystem complexes and carbon fixing Calvin cycle enzymes is thought to be optimised such that carboxylation and regeneration of RuBP are co-limiting in leaves under average daytime conditions (Haxeltine & Prentice 1996; Chen et al. 2009; Maire et al. 2012, Niinemets & Tenhuenen 1997).

This expectation gives rise to several predictions of how abundances of the different pools of photosynthetic leaf proteins should respond across gradients of temperature, light and water availability. Abundance of both Calvin cycle enzymes and photosystems should increase towards colder environments, to compensate for lower enzyme activity at lower temperatures (Raven & Geider 1988). This effect has been observed for Rubisco in a number of studies (summarised by Hikosaka et al 2006). Rates of primary photochemistry performed by the light harvesting apparatus may be less temperature sensitive, however (Raven & Geider 1988).

Allocation to photosystem complex proteins should be greatest where photosynthesis is light-limited (Niinemets 2007), and investment in Calvin cycle enzymes should increase with light availability, since capacity for carboxylation of RuBP determines the rate of light-saturated photosynthesis (Farquhar et al. 1980).

Investment in Calvin cycle enzymes should be greater at drier sites. By effecting greater internal CO2 drawdown, rate of CO2 uptake can be maintained at lower stomatal conductance, reducing the water cost of photosynthesis for dryland plants (Wright et al. 2001a,b, Scalon & Wright 2017). No direct effect of precipitation on investment in photosystem proteins is expected, although cross-correlation between precipitation and vegetation canopy density could influence this relationship.

More recently, an increasing number of biosphere models have incorporated a leaf nitrogen allocation component in an attempt to improve estimates of photosynthesis (Ghimire 2016 refs, Dong Ning refs).

While substantial progress has been made in characterising allocation of leaf protein to different functional pools, there some important limitations of approaches taken so far. Most crucially, protein quantification depends on complete extraction from leaves. The difficulty of extracting protein varies according to protein solubility (e.g. soluble, membrane-associated or cell wall-associated fractions) and between species (REF), and different extraction methods may also preferentially extract different protein fractions (Makino & Osmond 1990). Extraction of protein from leaves with high phenol content (e.g. eucalypts) is especially challenging due to the chemistry involved (see Warren 2000 for ref). Furthermore, most studies do not quantify what fraction of total protein was actually extracted from leaves (Warren refs?). Thus meaningful comparisons of protein abundances can be difficult to make even within a sample, still more difficult between species, and may not be reliable across studies.

Secondly, protein amounts are often estimated by measuring a proxy (e.g. chlorophyll for ‘pigment protein complexes’), and then calculating the protein amount using a ‘stock’ ratio of proxy amount to protein amount (Niinemets & Tenheunen 1997, Ghimire 2016). For example, Evans & Poorter (2001) estimated amounts of ‘pigment protein complex’ by measuring chlorophyll and using values of 38·5 and 41 moles of pigment protein nitrogen per mole of chlorophyll, for low- and high-light-grown plants, respectively. These values were sourced from measurements made on model organisms in the 1970’s and 80’s. Amounts of rubisco and electron transport proteins can similarly be estimated using equations that relate gas exchange parameters to protein amounts (Niinemets & Tenheunen 1997, Ghimire 2016, Evans & Poorter 2001). This approach is problematic for rubisco, since it only quantifies rubisco in its active conformation (REF). In a study of Australian species, the concentration of rubisco measured by radioimmune assay varied between 40% and 600% of that estimated from enzyme kinetics and gas exchange measurements (Warren et al 2000).

Finally, work to date has either made accurate measurements of a small number of species grown in controlled conditions, used plant traits, gas exchange measurements or environmental variables to estimate nitrogen allocation to the different functions of photosynthesis (Ning Dong, Ghimire, Dechant 2017). These limitations prevent us from accurately parameterising regional scale models of vegetation function.

Because we are able to comprehensively extract leaf protein and rapidly quantify over 2000 individual leaf proteins, our approach represents a substantial increase in accuracy and specificity with which we can investigate allocation of protein to functions of interest, and greatly expands the scope of what is possible ~~in plant proteomics and related disciplines~~.

Finally, accurate quantification

using standard ratios of protein amount per amount of a measured proxy

and most studies do not quantify how much they actually extracted.

* Incomplete extraction of leaf protein
* Lab-based estimations of protein pools which use ‘stock’ ratios of measured to estimated quantities
* Estimation of protein pools using gas exchange measurements (especially problematic for rubisco, as it only estimates active rubisco)
* Limited species coverage / glasshouse grown plants (although we fall squarely into the category of ‘limited taxonomic coverage’)

The problem with all of these things is that where people have talked about protein amounts, we’ve either done it in the lab for a limited set of species, estimation using equations or inference that Rubisco should be the protein of interest (CO2 drawdown stuff).

Lab measurements of rubisco rely on complete extraction from leaves. The difficulty of extracting rubisco varies across species, and most studies don’t quantify how much they actually extracted. Extraction is notably difficult from leaves with high phenol content (such as eucs).

Photosystems / ‘pigment associated proteins’ have never really been directly quantified anyway (using radioimmune assays or similar as for rubisco.)

Estimation from equations is dodgy as well. A study of protein composition in phenol-rich leaves Australian native plants showed Rubisco content to be present at between 60 – 600 % of the abundance predicted by models (Warren et al 2000).

- we know there are some fundamental relationships between major environmental variables and the size of protein pools associated with carboxylation and light capturing - temperature, water, light

- Responses of major protein groups to these env vars have been investigated extensively as a means to understanding the fundamentals of how the photosynthetic apparatus is optimised to its environment – ‘coordination’ and ‘optimality theories’

  - light response (Evans & Poorter 2001 would do, Niinemets too)

  - temperature response (see below, Berry and Bjorkman 1980)

  - precip response (Wright paper, forget which one – 2005?)

  - Evans & Poorter 2001, ref Niinemets and a few others, have directly determined Rubisco content and electron transport proteins in wild plants, but the species coverage remains limited and generally to glasshouse grown plants.

  - 'Pigment associated proteins' aren't usually quantified directly - measurements are made by measuring chlorophyll and then multiplying by a stock 'N per chlorophyll' number derived from Evans work (1989, paper with pie chart and tables?)

-Make the point here that allocation / abundance in relation to env vars has been looked at in detail for limited sets of species, or estimated for a wide range of species using a combination of gas exchange, functional trait and environmental data. Rubisco is often measured directly in the lab but other protein pools are estimated using proxies. ‘Pigment associated protein’ in particular (focusing on it here because it’s such a big pool) is almost always estimated from chlorophyll measurements using a stock ‘N per chlorophyll’ number derived from Evans work (Evans & Seeman 1989?).

Previously, leaf chlorophyll content expressed in nitrogen equivalents has been used as a proxy for investment in light capturing machinery (Niinemets & Tenhunen 1997), while Rubisco abundance has typically been estimated using gas exchange methods to estimate rates of carboxylation (ref). Our leaf protein abundance dataset provides the opportunity to directly test hypotheses about molecular adaptation of the photosynthetic apparatus to environmental conditions.

- It is possible to test a vast range of environment-function relationships using this dataset. In this initial analysis, we have opted to address variation in the abundance of photosynthesis proteins across fundamental environmental gradients: MAT, MAP and canopy irradiance. These relationships are of longstanding interest across multiple disciplines in the plant sciences.

More recently, an increasing number of biosphere models have incorporated N allocation in an attempt to improve estimates of photosynthesis (see Steve’s refs in presentation, Ghimire 2016 refs, Dong Ning refs).

Following co-ordination theory, we derived a set of predictions about differential investment in light capture and carbon assimilation along gradients of temperature, precipitation and light availability (see Fig. 1b):

a.) Investment in both Calvin cycle enzymes and photosystems should increase towards colder environments, to make up for the associated thermodynamic reduction of biochemical reaction rates (Hikosaka et al. 2006).

b.) Investment in Calvin cycle enzymes should be greater at drier sites. By effecting greater internal CO2 drawdown, rate of CO2 uptake can be maintained at lower stomatal conductance, reducing the water cost of photosynthesis for dryland plants (Wright et al. 2005) (other more ecophys oriented refs?). No direct effect of precipitation on investment in photosystem proteins is expected, although cross-correlation between precipitation and vegetation canopy density could influence this relationship.

c.) Investment in photosystem complex proteins should be greatest where photosynthesis is light-limited (Niinemets 2007), and investment in Calvin cycle enzymes should increase with light availability, since capacity for carboxylation of RuBP determines the rate of light-saturated photosynthesis (Farquhar et al. 1980).

Figure 1.) (left) Location of sampling sites across eastern Australia. Sites are marked by red triangles; 2.) (right) Hypotheses about differential investment in light capture and carbon assimilation proteins along gradients of temperature, precipitation and light availability (represented here as canopy density). Red up arrows indicate a predicted increase, blue down arrows indicate a predicted decrease, black ‘X’ indicates no predicted trend. The environmental gradients described here can be more or less overlaid across the map in Fig. 1. It is worth noting that although distinct mechanisms underlie hypotheses regarding canopy density and precipitation, the two variables are strongly related. [this caption needs refining, also need to standardise display of units – ‘mean annual precip (mm)’ vs ‘mean annual temperature’]