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 relationship between leaf nitrogen and photosynthetic carbon assimilation varies substantially across species, and in relation to how nitrogen resources are allocated to different functions within photosynthesis (Evans & Seemann 1989, Wright et al 2004).

The relationship is particularly weak for many Australian native plants (Warren 2000).

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).

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).

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). According to the co-ordination hypothesis of C3 photosynthetic acclimation, proportional allocation of protein resources to the light capturing photosystem complexes and carbon fixing Calvin cycle enzymes should 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).

- we know there are some fundamental relationships between N allocation and major environmental variables - 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)

  - 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.

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’]