1.) Introduction

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?). As such, quantifying protein amounts using proteomics methods provides direct information about how organisms are adapted to their environment (Diz & Calvete 2016). 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 mass spectrometric methods remain novel (refs from Steve).

We have developed proteomics methods which allow comprehensive extraction and absolute quantification of the top 2000-3000 most abundant proteins in leaves. 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) continental-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 sets of biochemical reactions within leaves as well as in the biosphere as a whole (Blankenship & Hartman 1998; Raven 2013, Evans & Seeman 1989).

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).

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 (Dong 2016, Ghimire 2016, 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~~.

Results:

*Study area and sampling:*

Wide coverage, over multiple biomes, varying almost independently in temp and precip (Fig 1a.)

*Protein composition of the average eucalypt leaf.*

In Fig 2a we show how protein resources were allocated to all major functions in an ‘average’ eucalypt leaf (based on 320 leaf samples). The majority (64%, SD X%) of protein was associated with photosynthesis: 36% was associated with the carbon fixing Calvin Cycle, 22% (SD X%) with the light reactions and 4% (SD X%) with photorespiration (Fig 2a). The most abundant individual protein complexes were Rubisco (30%, SD X%) of leaf protein, and photosystem II (X%, SD X%) (Fig 2b). Protein synthesis, folding and degradation was the second most abundant top-level category at X% (SD X%) (Fig 2a).

Our mass spectrometry approach allowed detection of X individual proteins per sample, on average. These proteins accounted for 99.9% of sample mass, among which the top 500 most abundant proteins represented 90% (Fig 2c). This is a higher degree of dominance by the top few proteins than observed in [comparison] (Fig 2d), reflecting the specialist nature of leaves as photosynthetic organs.

*Linking leaf protein abundances with environment and functional traits*

We were able to describe patterns of leaf protein abundance across environmental gradients, as well as in relation to key leaf functional traits ~~and physiological properties~~ (Fig 3a). Per leaf area abundances of all major protein functional categories were correlated with each other, as well as with leaf nitrogen per area (N\_area), leaf mass per area (LMA), and maximum photosynthetic rate (Amax). Patterns in proportional abundances of protein functional categories (indicating investment in a defined function relative to investment in all other functions) were less general.

*b.) first scatterplot panel*

Given the level of detail in our protein abundance dataset, it would have been possible to test a plethora of specific environment-protein abundance relationships. We decided to focus on photosynthesis here due to the dominance in leaves of proteins catalysing this set of processes. We selected several specific relationships of interest to the vegetation modelling community for deeper analysis; to date these relationships have only been investigated via proxies.

Calvin cycle proteins per leaf area reduced notably as sites became warmer (stat, Fig. 3b-i), and to a lesser extent with increasing precipitation (Fig. 3b-iii). The per leaf area abundance of Calvin cycle proteins was highly correlated with the total abundance of protein per area (Pearson’s r = 0.97), and environmental trends in Calvin cycle protein abundance were essentially identical to trends in leaf protein abundance.

Photosystem proteins per leaf area showed a pronounced decline with increasing incident irradiance (Fig. 3b-v, X% per Y irradiance). Per leaf area photosystem protein abundance declined substantially with increasing MAT (Fig. 3b-i) and was also strongly correlated with total leaf protein abundance (Pearson’s r = 0.82). No per leaf area response to MAP was observed (Fig. 3b-iii), however. Since MAP and incident irradiance were negatively correlated (i.e. denser canopies at wetter sites, Pearson’s r = -0.59) the lack of protein response to MAP could be explained by changing light conditions.

Proportional allocation of protein resources to Calvin cycle protein did not adjust over gradients of MAP or MAT (Fig. 3b-ii,iv) but increased marginally (stat) with increasing incident radiation (Fig. 3b-vi). Proportional photosystem protein abundance increased with increasing MAP (Fig. 3b-iv) and decreased with increasing incident irradiation to a similar extent as the per leaf area measure (Fig . BLAH). This latter response may explain the observed decline in Calvin cycle proteins as incident irradiance increased. The range of interspecific variation in photosystem protein proportional abundance (0.09-0.23, 2.6-fold) was considerably higher than for Calvin cycle proteins (0.30-0.39, 1.3-fold). These observations provide robust evidence that eucalypt leaves specifically optimise protein allocation to light capture in response to environmental conditions (some stats and numbers), while adjustment of carboxylation capacity is largely achieved through bulk changes in per leaf area protein content.

*c.) second scatterplot panel*

One obvious way Calvin Cycle protein per leaf area can change is via changes in depth of mesophyll and of leaf, and indeed adjustments in per leaf area Calvin cycle protein abundance occurred to some extent via changes in leaf mass per area (LMA) (Fig. 3c-i). The substantial scatter around the Calvin cycle – LMA relationship indicates that LMA responded to other requirements in addition to carboxylation capacity. Photosystem abundance did not increase per leaf area with increasing LMA (Fig. 3c-ii) and declined as a proportion of total leaf protein. Light harvesting capacity thus appears to be optimised for a given leaf area independently from leaf thickness.

Leaf nitrogen per area was a strong predictor of both Calvin cycle and photosystem protein abundance per leaf area, and no relative changes in these protein categories occurred with increasing nitrogen per area.

*d.) protein abundance/concentration/LMA multiple regressions*

We hypothesised that Calvin cycle protein abundance would be driven by temperature dependence of enzyme kinetics, and that maximisation of CO2 drawdown at low stomatal conductance in water-limited environments. Fig 3d-i shows that these demands were in fact complementary: leaves sampled at cold dry sites required the most protein, while leaves from warm wet sites experienced neither constraint, having both low protein content per area and low LMA.

The role of LMA versus protein concentration (Calvin cycle protein as a fraction of leaf dry mass) in determining per leaf area protein abundance depended interactively on MAP and MAT (Fig 3d-ii,iii). Low per leaf area Calvin cycle protein abundance at warm, wet sites was more closely associated with low LMA than low protein concentration, while high per leaf area Calvin cycle protein abundance at cold, dry sites was strongly associated with high Calvin cycle protein concentration.

Plants are able to build cheaper leaves at warm wet sites, where photosynthetic reaction kinetics are increased and plants are not water limited.

Hypotheses [include these above?]:

Optimal allocation theory gives rise to several expectations about how abundances of the different pools of photosynthetic leaf proteins should respond across gradients of temperature, light and water availability (Niinemets 1997, Maire 2012). 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 increase towards 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.