INTRODUCTION

The functioning of leaves is driven by proteins. Each protein or protein-complex carries out specific transactions, ranging from bulk transfers of carbon between atmosphere and vegetation down to defences against specific pathogens or herbivores. The capacity of leaves to perform these functions is typically determined by the amount of protein doing the work (ref – Michaelis & Menten?). And yet, interpretation of what vegetation is doing across wide areas and many species continues to rely on measurements of leaf nitrogen content, both in comparative trait ecology (refs) and in models intended to capture contribution of vegetation to world carbon budgets (refs). Leaf nitrogen content is feasible to analyse across many hundreds of species and samples, although it obviously represents a pool of many different proteins with many different functions.

To date, methods for precisely quantifying protein amounts in leaves been too intensive to apply across large numbers of samples, requiring laborious benchtop assays to measure abundance of specific proteins of interest. Alternatively, protein amounts have been estimated indirectly using proxy measurements: chlorophyll can be used as a proxy for amount of light harvesting proteins, and abundance estimates of the carbon fixing enzyme Rubisco can be derived from leaf gas exchange measurements. As such, previous studies of how and why protein amounts vary in leaves have been limited in precision and in scope; see SuppMat for a more complete comment on this matter.

Here we apply new proteomics techniques that make possible one-pass quantification of almost all proteins in a leaf. We used quantitative mass spectrometry to analyse 320 leaves from 32 Australian eucalypt 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 a complete description of leaf protein allocation for all major protein functional categories, and describe the dominant biogeographical patterns in abundance of photosynthesis proteins.

RESULTS

*Protein composition of the average eucalypt leaf.*

Our mass spectrometric approach detected 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.

The majority (64%, SD X%) of leaf 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). [need to decide here whether we’re saying that in the large, this mixture is in line with previous intensive results from few species, versus it’s really different, vide previous discussions with steve about uws results]

Although protein allocation is known to vary between species, our numbers for the major categories are not dissimilar to those obtained using classical biochemical assays on domesticated species such as *Spinacea oleracea*, *Phaseolus vulgaris*, and *Cucumis sativus* (Evans & Seeman 1989).

*Linking leaf protein abundances with environment and functional traits*

Figure X illustrates a key advance made in our approach, whereby protein abundances are directly relatable to leaf functional and physiological traits as well as the environmental conditions in which sampled individuals were growing. We have included protein functional categories at a range of scales, from high-level categories such as ‘photosystems’, which include a number of subcategories, down to ‘isoprene synthase’, which represents a single enzyme.

*b.) first scatterplot panel*

Biogeographic patterns in abundance of photosynthesis proteins

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, Evan

s & Seeman 1989). Readers are invited to explore the dataset more deeply using the interactive data explorer at proteography.org.

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).

Analyses are presented with protein abundance represented both as a fraction of total leaf protein, and on a mass per leaf area basis. Protein fractional abundances render explicit the economics of allocating limited protein resources to the various functions required within leaves; that is, protein is invested in a given function at the expense of all other functions. Protein mass per unit leaf area describes how much of a protein is devoted to as given function per area presented to the sun. The following analyses allowed us to determine which of these two metrics/types of abundance were optimised under different environmental conditions.

**LIGHT**

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

In line with expectation, both photosystems protein per leaf area (PSarea) and fractional abundance of photosystems (PSfrac) showed a pronounced decline with increasing incident irradiance (Fig. 3b-v, X% per Y irradiance; Fig 3blah stat). No increase in Calvin cycle protein per leaf area (CCarea) was observed with greater light availability, but Calvin cycle fractional abundance (CCfrac) did increase marginally (Fig, %, stat); this fractional increase may be simple outcome of decreasing PSfrac, however.

**MAT**

Leaf protein content is known be higher in cool 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).

Both CCarea and PSarea declined notably with increasing MAT (stat, Fig. 3b-I, stat Fig XX % blah). Neither CCfrac nor PSfrac changed systematically over temperature gradients (Fig stat %).

**MAP**

We also expected that investment in Calvin cycle enzymes would 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).

CCarea declined strongly with increasing MAP (Fig x stat %), although no response of CCfrac over the precipitation gradient was observed (Fig. 3b-iii). This indicates that leaves alter their capacity to absorb CO2 from the mesophyll / extracellular spaces during stomatal closure by increasing the per leaf area amount of Calvin Cycle enzymes, rather than the amounts of these enzymes relative to other proteins.

PSarea showed no significant trend but PSfrac increased by x% over the precipitation gradient. We had no expectation of a direct effect of precipitation on investment in photosystem proteins, although cross-correlation between precipitation and vegetation canopy density may underlie this latter trend.

*c.) second scatterplot panel*

One obvious way CCarea can change is via changes in depth of mesophyll, 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 in the Calvin cycle – LMA relationship indicates that LMA responded to other requirements in addition to carboxylation capacity~~, reflecting the involvement of both nitrogen- and carbon-dominant components in determining LMA~~. Conversely, PSarea was not related to LMA (Fig. 3c-ii), and PSfrac declined declined notably as LMA increased.

The range of interspecific variation in PSfrac (0.09-0.23, 2.6-fold) was considerably higher than for CCfrac (0.30-0.39, 1.3-fold), and the correlation between PSarea and total leaf protein per area was somewhat weaker than that of CCarea (R2).

Together, these observations provide suggest that eucalypt leaves can adjust to different light conditions by optimising fractional protein allocation to the light harvesting apparatus (some stats and numbers), while adjustment of carboxylation capacity is largely achieved through bulk changes in per leaf area protein content.

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

Demands for extra Calvin cycle protein at low temperatures (due to temperature dependence of enzyme kinetics) and in water limited environments (to maximise CO2 drawdown at low stomatal conductance) were complementary: leaves sampled at cold dry sites required the most protein, while leaves from warm wet sites experienced neither constraint, having both CCarea and low LMA.

The role of LMA versus protein concentration (Calvin cycle protein as a fraction of leaf dry mass, CCconc) in determining CCarea depended interactively on MAP and MAT (Fig 3d-ii,iii). Low CCarea at warm, wet sites was more closely associated with low LMA than low protein concentration, while high CCarea cycle protein abundance at cold, dry sites was strongly associated with high CCconc. Thus plants construct cheaper leaves at warm wet sites, where photosynthetic reaction kinetics are increased and plants are not water limited.

SUMMARY

1. We believe this study is harbinger of widespread use of one-pass protein quantification to study ecological distribution of proteins, both those with well-understood function and those where function uncertain
   1. Potentially some kind of map?
   2. Possibly might choose to point out that strongly divergent ratios of quantity of certain protein pairs probably indicates strongly-divergent activity? – or could reserve that point to make elsewhere