The University Of Sheffield.

Genetic and environmental enablers of the evolution of Crassulacean Acid Metabolism in submerged aquatic plants

by

Daniel Peter Wood

Department of Animal and Plant Sciences

University of Sheffield

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

January 2018

Contents

Acknowledgements

First and foremost I would like to thank my primary supervisor, Dr Pascal-Antoine Christin. He has been incredibly generous with his time, providing continuous discussion and guidance throughout the thesis. I have benefited greatly from his knowledge of evolution, phylogenetics, and bioinformatics. His boundless enthusiasm and encouragement has made the research process very enjoyable, often leaving me convinced that the seemingly insurmountable may be actually be surmountable after all. I am particularly grateful for his commitment towards the end of the thesis, putting in considerable time often at short notice.

Professor Colin Osborne has been a valuable source of ideas and advice, and I have appreciated the opportunity to participate in lab meetings with his group, which have provided a useful perspective at the physiological and ecological level. I am also grateful to Professor David Beerling for his valuable advice on long term carbon cycles and contribution to the design of the project. I am also grateful to Dr Guillaume Besnard for his sequencing assistance and advice.

I owe a particular debt to Dr Jill Oloffson and Dr Luke Dunning, who have contributed particularly to the fourth chapter of this thesis, and whose inexhaustible patience and advice in laboratory work and analytical technique (as well as careers, mince pies and jumping in lakes) has been invaluable, and I will miss visit my regular visits to their office. I am also grateful to Dr Marjorie Lundgren for advice, assistance and encouraging me to aim high.

It has been a privilege to have spent the last three years uncovering the mysteries of photosynthesis with my brothers-in-arms Jose Moreno-Villena and Matheus Bianconi, from whom I have learnt much and whose company I have thoroughly enjoyed. I have been glad of the valuable perspective of Nick Moody, look forward to seeing Lamia Munshi go on to success in her PhD, and been made to feel thoroughly at home by all members of the Christin lab. I am also grateful to members of the Osborne lab and attendees of the lab meetings at AWEC for their valuable perspectives. I would also like to thank the staff of the department of Animal and Plant Sciences, particularly Heather Walker, Rachel Tucker, Gavin Horsburgh, Irene Johnson and Maggi Killion for their

valuable support and assistance. I am grateful to Dr Scott McKenzie, Emma Baxter, Lidia Korba and Nick Stewart and Scottish Natural Heritage for providing samples for the fifth Chapter, and to Hannah Sewell for assistance in sample collection in Wales. I am also grateful to Dr James Hartwell and Professor Charles Wellman for agreeing to examine the thesis, and providing valuable comments and discussion. I have been glad of the support and comradeship of many of the PhD students in the department, and have enjoyed the company of my officemates. I am very lucky to have had the support, encouragement and friendship of Caitlin Higgott throughout the course of PhD.

I am grateful to the University of Sheffield for funding this research and hosting me throughout.

Finally, I thank my family, without whom none of this would be possible.

Summary

The evolution of complex traits is one of the great wonders of evolution. Traits such as camera eyes, flight and biochemical carbon concentrating mechanisms (CCMs) require the co-ordinated actions of multiple components for function, yet have evolved multiple times in diverse lineages of organisms and environments. In this thesis, I investigate the environmental and genetic enablers of the evolution of CCMs. I performed a literature review of these traits, leading me to focus on Crassulacean Acid Metabolism (CAM) as a study system. Whilst most commonly known in plants dwelling in hot and dry environments, this trait has also evolved in the submerged aquatic lycopod genus *Isoëtes,* offering a valuable comparison point to infer common evolutionary drivers. Using molecular dating to identify the temporal origins of extant aquatic CAM diversity, I found that aquatic CAM plants evolved and diversified at similar times to their terrestrial counterparts, implicating falling atmospheric CO_2 levels as a common environmental driver. To identify if this shared selection pressure invoked similar genetic modifications for CAM evolution in diverse lineages, I used RNA-sequencing to identify CAM genes in two aquatic CAM plants, the lycopod *Isoëtes lacustris* and the angiosperm *Littorella uniflora*. I found that the most highly expressed gene lineages in these species strongly overlap, but found that it stemmed from conservation of gene expression levels from the ancestral vascular plant. The similarity of *I. lacustris* and *L. uniflora* occurs despite the different mating systems of the angiosperms and lycopods, so I investigated the genetic structure of these species in the British Isles, finding higher levels of genetic structure in *Littorella*, likely a result of its reliance on emergent flowers for sexual reproduction. These results suggest that common environmental and genetic enablers contribute to trait evolution in divergent environmental and genomic backgrounds, leading to the repeated emergence of complex traits.

Introduction

1.1 Natural selection and adaptations

Living organisms resemble complex machines, with different components interacting to produce specific outputs (Paley, 1802). It was not until the nineteenth century that a naturalistic mechanism, evolution by natural selection, was proposed that could explain the origins of complexity in populations of organisms over time (Darwin, 1859). Natural selection is able to cause phenotypic changes when heritable variants of components exist, and when these variants result in differential probabilities of reproduction. While the mechanism of heritability was not known at the time the theory of natural selection was proposed, examination of DNA genomes identified them as heritable and encoding the vast majority of traits found in organisms. DNA replication errors lead to changes, such as base substitutions, insertions, deletions and rearrangements, some of which cause phenotypic alterations, by altering the expression patterns of genes, or the properties of the proteins they encode. Natural selection of variant genomes (genotypes) will result in those chance mutations that cause improved lifetime reproductive success (fitness) becoming more frequent in populations. Over time, this process can lead to new traits increasing the fitness of organisms in specific environments (adaptations).

In the production of adaptations involving changes to a few components, the mechanism can be understood intuitively. For example, a variant of the promoter in the human lactase gene prevents its downregulation in adulthood, which in certain populations (in which lactose formed a significant part of the adult diet) conferred an advantage, and was therefore selected (Swallow, 2003). In many cases, however, an adaptation requires large numbers of components for function. Over fifty gene products are involved in the bacterial chemotaxis/motility system (Macnab, 2004), and numerous traits result from the coordinated action of multiple anatomical and/or biochemical components, including camera eyes (Kozmik et al., 2008), powered flight (Brown, 1963) and most metabolic pathways (Weng, 2014). How traits with such large numbers of components, all of which are required for function, are assembled is less obvious. Indeed, those seeking to discredit the idea of evolution often claim that such events cannot occur, and

the traits in question are “irreducibly complex” (Behe, 1996). Identifying how, and under what circumstances, complex traits evolve is a key question for evolutionary biology, which remains only partially resolved.

1.2 Co-option and evolutionary potential of genes

The evolution of a complex trait requires the generation of suitable variation, and for these variants to spread throughout a population. New variants are generated by chance mutations, a process which generally acts to modify existing components rather than creating new ones *de nov*o. The components in complex adaptations are therefore co-opted from those present in ancestral organisms, in a process of exaptation (Gould and Vrba, 1982). The co-option of pre-existing components for a new function can be observed at multiple levels of biological organisation, from the recruitment of entire organisms into symbiotic functions (Sagan, 1967), to behavioural traits (Borgia and Coleman, 2000), and biochemical pathways (Huang et al., 2016). While the process was historically described at the phenotypic level (Gould and Vrba, 1982), it can easily be extend to the genotype. Whilst low-complexity features such as transcription factor binding sites may emerge *“de novo”* (Doniger and Fay, 2007), more complex features such as protein coding gene sequences are generally duplicated or fused from pre-existing copies (Feuk et al., 2006; Long et al., 2003; Zhou et al., 2008). Because new traits emerge from the diversification of existing genes, the accumulated genetic inventory of organisms is likely to determine the evolutionary trajectories available to their descendants. Indeed, the number of components an organism possesses that could be recruited to a particular new trait may be quite small, although the extent to which this constrains evolution is largely unknown.

At higher levels of biological organisation, the reality of a limited pool of components for recruitments is obvious – legs and arms are more suitable for recruitment for propulsion underwater than guts or ears, for instance. However, our understanding of what properties affect recruitment of components at lower levels of biological organisation, such as genes, to new functions, is limited. The molecular function or catalytic properties of a gene may affect its suitability for recruitment to a new function. The location of the expression of a gene or localisation of the gene product, and the level of expression in that location, may also be important factors

affecting the likelihood of a gene being recruited for a novel function (Christin et al., 2010). For example, genes recruited to venom function in poisonous snakes were ancestrally expressed in salivary glands (Hargreaves et al., 2014), and it has been suggested that the recruitment of a pancreatic trypsin for antifreeze functionality in arctic fish was facilitated by the presence of a signal peptide used for excretion into the digestive tract (Logsdon and Doolittle, 1997). In addition to these more “intrinsic” properties of particular genes, the genomic background of an organism as a whole may affect which components are recruited for new functions. The mutation rate varies across the genome (Wolfe et al., 1989; Lang and Murray, 2011), and genes located in regions undergoing more mutations could acquire adaptive mutations faster than genes located in other regions. In cases where the ancestral function of a gene is incompatible with co-option to new functions, gene duplication can result in relaxed selection and facilitate neofunctionalisation (Ohno, 1970; Zhang, 2003). This potentially makes frequently duplicated genes or genomic regions more likely sources of adaptive components. In addition, epistatic interactions between genes may result in homologs having different probabilities of co-option in different genetic backgrounds (Phillips, 2008; Griswold, 2015). Overall, the ensemble of genes present in genomes is likely to affect the potential of an organism to evolve new phenotypes. Understanding the factors that determine the ability to evolve certain traits is important. Across evolutionary time, different clades of organism have become ecologically dominant (e.g. dinosaurs, lycopods, Actinopterygian fish), resulting in different pools of components available for trait recruitment. The extent to which these pools of components varied in their suitability for recruitment to new functions therefore influences when and where a given trait evolved.

1.3 Environmental drivers of novel adaptations

In addition to the production of mutations suitable for recruitment into a new trait, these mutations must reach high frequencies in the population to trigger the evolution of novel traits. While drift and random processes can lead to the fixation of new mutations in some conditions, natural selection provides the most potent means by which multiple specific mutations can reach high frequencies in a population. Adaptive traits are often associated with particular environments that produce a new selective pressure. For example, the evolution of insectivory was favoured in plants inhabiting areas of low

nitrogen status (Givnish et al., 1984; Ellison and Gotelli, 2001). Understanding which particular aspects of an environment provide the selective advantage to a trait is therefore important to understanding when they evolve, and forms the basis of the definition of the "function" of a trait (Neander, 1991; Graur et al., 2015). This is especially important when a novel trait has multiple potential consequences. For example, aspects of bipedalism in humans proposed to provide the selective pressure that led to its evolution include increased viewing distance, tool carrying, thermal radiation avoidance and sexual selection (Parker, 1987; Hunt, 1994). The identification of the selective advantage of a trait is crucial for understanding the circumstances in which it will likely evolve, especially in different environmental contexts where not every consequence of the trait is likely to apply.

Disentangling the effects of genetic enablers and environmental drivers on the evolution of a given trait is complicated by the patchy distributions of organisms and environments in time and space. The lack of a trait in a certain environment may be due to a vicarious lack of organisms able to evolve this trait in that environment. For example, invasive species taking over niches indicate that a particular native genotype may not be the best suited for that environment, and the better genotype was simply unavailable in that location. Similarly, it is unclear whether the genotypes of modern cacti would be unsuitable for survival in the deserts of the Triassic; their absence is more likely to be a consequence of the relatively recent origins and global dominance of angiosperms, the group that includes cacti. The extent to which historical contingencies versus genotypic and environmental constraints affect the diversity of traits we see today, and potentially those that are likely to be seen in the future in the face of the significant depauperation of genotypes and changing environments in the Anthropocene (Ceballos et al., 2015), is difficult to discern. Experimental evolution studies have shed light on the factors promoting the evolution of certain traits (Blount et al., 2008; Kawecki et al., 2012), but such approaches are limited to fast evolving organisms. The origins of more complex traits, such as camera eyes, cannot be tracked using experimental systems. Instead, understanding the origins of complex traits in long-lived organisms require historical approaches. The tape of life (Gould, 1990) cannot be replayed, but traits that evolved multiple times independently represent natural replicates that can help with differentiating coincidence and causation. Such convergent

traits therefore represent outstanding systems to evaluate the factors that increase the accessibility to novel phenotypes.

1.4 Convergent traits as a study system

Convergent evolution, where distantly related species evolve similar phenotypes, represent valuable systems in which to disentangle the important causative factors underlying the emergence of novel adaptations. Well known examples of convergent trait evolution at large phylogenetic scales include the appearance of eyes, prehensile limbs, streamlining in fish, dolphins and ichthyosaurs, and the ability to fly in various extant and extinct groups (Conway-Morris, 2003). At smaller phylogenetic scales, phylogenetic reconstructions using genetic markers have been used to resolve relationships between taxa and allow distinguishing of convergent evolution from the common inheritance of a trait. These revealed more examples of convergent evolution at smaller evolutionary scales, such as the repeated origins of specific colour patterning in butterflies (Brower, 1994), hermaphroditism in nematodes (Kiontke et al., 2004), and industrial pollutant resistance in killifish (Reid et al. 2016). Identifying the factors associated with these repeated evolutionary events can identify the genetic and environmental enablers of adaptation.

Evolution is an inherently stochastic process, so the repeated involvement of a factor suggests it significantly increases the chances of the trait evolving. This applies equally to environmental and genetic factors. The frequency at which similar components are recruited, or similar environments are invaded, is an indicator of how important these factors are for the evolution of a complex trait. For example, distantly related proteins such as heat shock proteins, enolases and lipid binding proteins have been recruited for use as crystallins in vertebrate and cephallopod lenses, suggesting a wide potential pool of enzymes can be recruited for this function (Wistow, 1993; Tomarev and Piatigorsky, 1996). Conversely, the repeated origins of viviparity in reptiles are most frequently associated with transitions to colder climates (King and Lee, 2015; Li et al., 2017), which suggests that cold climates are an important driver of the evolution of this trait.

Many examples indicate that similar genes are repeatedly co-opted to produce convergent phenotypic changes (Martin and Orgogozo, 2013). For example, the *M1CR* coat locus has repeatedly been co-opted for changes in vertebrate colouration (Rosenblum et al., 2004; Gross et al., 2009; Hubbard et al., 2010), the same gene was repeatedly modified to confer tetrodotoxin resistance in garter snakes (Hague et al., 2017), and modifications of Prestin genes underlaid the origins of echolocation in a number of mammal lineages (Parker et al., 2013). These suggest that in some cases, the number of genes able to trigger a change in a given trait is fairly limited. Other cases show the opposite pattern, with divergent gene co-option or changes within a gene, such as in adaptation to high temperatures in *Escherichia coli* (Tenaillon et al., 2012), learning behaviour in *Drosophila* (Kawecki and Mery, 2006) or pelvic reduction in stickleback (Bell et al., 2007). Determining which factors result in similar versus divergent patterns of gene recruitment is important for understanding the evolutionary potential of different organisms and components. Genetic distance has been shown to be an important factor in determining the likelihood of repeated gene recruitment events, with more closely related species generally showing higher levels of convergence than more distantly related taxa (Christin et al., 2010; Conte et al., 2012; Storz, 2016). This is likely caused by the similarity of the ancestral state of close relatives and resultant similarities in the likelihood of recombination and mutation, the number of gene copies and propensity for duplication, the position of genes in regulatory networks, and the demography of a species, all of which may bias organisms to repeatedly co-opt the same genes (Stern, 2013). Many studies of convergent evolution focus on adaptations associated with relatively small numbers of genetic changes, and in relatively similar genetic backgrounds (Arendt and Reznick, 2008; Martin and Orgogozo, 2013). Studies of more complex adaptations that repeatedly evolved in a wide variety of genetic backgrounds are required to understand the factors that dictate the likelihood of evolving novel complex traits across the tree of life.

A complication in the study of convergent evolution is that the “function” of an adaptation might vary across environments. For example, the ecological drivers and “function” of bipedalism is likely to be different in birds and humans. This may be a trivial example, but determining whether similar phenotypes are functionally equivalent might be more complicated in other cases. Environment specific factors may influence the likelihood of a given trait evolving as well as the evolutionary trajectories likely to

be followed. For example, due to the refractive properties of air and water, transparency is a more effective means of camouflage underwater than above water, even though the same functions are performed (Speed and Arbuckle, 2017). Trait evolution may be correlated with particular environments, but the selection pressure within that environment may be unclear due to multiple environmental co-variates – for example, the exact aspect of cold habitats influencing viviparity is unclear. The situation is further complicated by the fact that closely related organisms are likely to inhabit similar environments (Gompel and Prud'homme, 2009), potentially leading to conflation of historical and environmental factors in the evolution of complex traits. Therefore, understanding the selective advantage of a trait in different environmental contexts is crucial in deciphering the likelihood of a complex trait evolving.

1.5 Carbon-concentrating mechanisms as convergent complex traits

The biochemical carbon concentrating mechanisms (CCMs) of higher plants represent highly convergent traits that occur in a wide variety of environments and genetic backgrounds. They therefore offer outstanding opportunities to disentangle the environmental and genetic determinants of complex trait evolution. These CCMs, the most common of which are known as C_4 and Crassulacean Acid Metabolsim (CAM), are a modification of the ancestral "C_3" form of photosynthesis. In C_3 plants, which represent the vast majority of species, atmospheric CO_2 is fixed directly by the enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO) to 3-phosophoglycerate. This reaction starts the Calvin cycle, which requires energy derived from the light-dependent reactions of photosynthesis to produce sugars. This system is highly conserved in land plants, but presents problems in certain ecological situations (Christin and Osborne, 2013). In situations where O_2 is abundant, the oxygenase activity of RuBisCO can result in the fixation of O_2 at the expense of CO_2, triggering the photorespiration pathway that consumes energy and releases CO_2 (Figure 1.1). Furthermore, the Calvin cycle can only take place in daylight hours, which limits the time available for CO_2 uptake, and in terrestrial environments necessitates the opening of stomata to facilitate gas exchange into the leaf. Open stomata also allow water to escape from the leaf, and as the daylight hours represent the hottest part of the day, this timing serves to exacerbate evaporative water loss (Figure 1.1). The C_3 system evolved in a high CO_2 environment 2.7 billion years ago where these side-effects were likely

minimal (Nisbet et al., 2007; Christin and Osborne, 2013). In certain situations, however, sophisticated CCMs evolved to mitigate these effects.

The CAM and C_4 CCMs involve separating the uptake of CO_2 from the environment and its subsequent photosynthetic fixation in the Calvin cycle. In CAM plants, this separation occurs temporally. CO2 is taken up at night, fixed nonphotosynthetically and stored in the photosynthetic mesophyll cells. During the following day, this stored CO_2 is released and photosynthetic fixation via the Calvin cycle takes place as in C_3 plants. This reduces the reliance of CO_2 uptake from the environment during the day, which can be advantageous in dry conditions where a reduced requirement to open stomata during the day can result in substantial increases in water use efficiency (Nobel, 1996), and rapid release of CO_2 can result in high concentrations around RuBisCO, increasing photosynthetic rate and reducing photorespiration (Cockburn, 1979; Lüttge, 2002; Pedersen et al., 2011). In C_4 plants, uptake of CO_2 from the environment and photosynthetic fixation both take place during the day but are separated spatially. CO_2 is taken up in mesophyll cells, but photosynthetic fixation does not occur in these cells. Instead, the CO_2 is fixed nonphotosynthetically and transported to the inner bundle sheath that is not in contact with the atmosphere and where RuBisCO is localised This increases the concentration of CO_2 around RuBisCO relative to O_2, which suppresses photorespiration. This is particularly advantageous in the low-CO_2 atmosphere that prevailed in the last 30 million years (Pagani et al. 2005)**,** especially at high temperatures in which CO_2 solubility decreases faster than O_2 solubility, and the CO_2:O_2 specificity of RuBisCO decreases, both of which promote the fixation of O_2 (Christin and Osborne, 2013).

The CAM and C_4 CCMs are complex, as defined previously, in that they involve a large number of components to act in concert to produce the phenotype (Figures 1.2 and 1.3). In CAM, the usual sequence of opening stomata during the day, vs. closed stomata at night, must be reversed to allow CO_2 entry during the night (Males and Griffiths, 2017). The nocturnal nonphotosynthetic fixation of CO_2 is catalysed by the enzyme phosphoenolpyruvate carboxylase (PPC), which fixes HCO_3^- to a three carbon substrate, phosphoenolpyruvate (PEP), resulting in the production of a four carbon (C_4) acid, oxaloacetate (OAA). HCO_3^- can form passively from hydration of CO_2, but this reaction can also be catalysed by carbonic anhydrase (CA; DiMario et al. 2017) . The large

quantities of PEP required are generated from the breakdown of transitory carbohydrates, such as starch and/or soluble sugars (Borland et al. 2016). The oxaloacetate produced is then converted to malate via the action of malate dehydrogenase (MDH). Malate can then be stored in the vacuole, which in CAM plants tend to be large (Silvera et al. 2010). Uptake occurs via transporters such as tonoplast dicarboxylate transporters (TDT) or aluminum-activated malate transporters (ALMT) in concert with vacuolar ATPases which provide the proton-motive force to facilitate uptake of malic acid (White and Smith, 1989; Brilhaus et al. 2016; Martinoia 2018). During the day, malic acid released from the vacuole and decarboxylated by one or several of; NADP- dependent malic enzyme (NADP-ME), NAD- dependent malic enzyme (NAD-ME) and MDH phosphoenolpryvuate carboxykinase (PCK), releasing CO_2 for photosynthetic fixation via the Calvin cycle. PCK-mediated decarboxylation yields PEP, whereas NAD-ME and NADP-ME-mediated decaboxylation yields pyruvate, which is converted to PEP via pyruvate *ortho*phosphate dikinase (PPDK). This PEP is then converted back into starch or soluble sugar synthesis for use the following night (Brilhaus et al. 2016). These enzymes are all present in C_3 plants but are often present at higher concentrations and/or possess amino acid differences increasing catalytic rates in CAM plants (Cushman and Bohnert, 1999; Aubry et al. 2011, Yang et al. 2017). Furthermore, the processes of carbon uptake, fixation and storage as malate must occur at higher rates during the night, and the subsequent processes of decarboxylation and PEP recycling must occur at higher rates during the day, to avoid futile cycling of metabolites. Diurnal fluctuations in the transcript, protein and post-translational modifications of CAM enzymes occurs to separate these reactions, in some cases involving multiple levels of regulation (Cushman and Bohnert 1999; Brilhaus et al. 2016). For example, PPC transcript levels increase at night in some species (Brilhaus et al. 2016; Yang et al. 2017), but protein activity can be enhanced at night via N-terminal phosphorylation by PPC kinase (PPCK), whose transcription and translation is higher during the dark period (Hartwell et al. 1999). Information from the circadian clock, as well as light levels and water, are integrated to produce these temporal rhythms (Wilkins, 1992, Hartwell et al. 1996, Ceusters et al. 2014, Males and Griffiths 2017). In addition, many CAM plants can switch from C_3 to CAM in response to drought or salinity, requiring the integration of environmental stress signalling and the CAM pathway (Dodd et al. 2002; Winter and Holtum 2014). Whilst the total number of genetic changes required for the efficient operation of a CAM cycle is not known, the

complexity of the pathway and requirement for sophisticated regulatory co-ordination suggest a large number of genetic changes are likely involved (Yang et al. 2015). C_4 photosynthesis involves many of the same biochemical reactions as CAM (Figure 1.3). The enzymes involved in carboxylation (CA, PPC) and decarboxylation (PCK, NAD-ME, NADP-ME) are shared between the two pathways (Edwards et al. 2012). Unlike in CAM photosynthesis, the oxaloacetate produced by PPC is not stored as malate, but transported either in the form of malate (produced from oxaloacetate via MDH) or aspartate (produced from oxaloaceatate via aspartate aminotransferase; Asp-AT) into the bundle sheath cells for decarboxylation and photosynthetic fixation. Similarly to CAM, complex regulatory changes are required to ensure the restriction of nonphotosynthetic fixation to the mesophyll cells, and decarboxylation and photosynthetic fixation to the bundle sheath cells (Schuler et al. 2016). Anatomical changes, such increases in the numbers, size and chloroplast content of bundle sheath cells compared to C_3 species, are also associated with the evolution of C_4 photosynthesis, as well as increased expression levels and altered catalytic properties of the C_4 enzymes (Schuler et al. 2016). In summary, CAM and C_4 photosynthesis represent complex traits, involving large numbers of components acting in a co-ordinated manner to produce a specific output. Our detailed understanding of the mechanistic basis of these complex traits makes them ideal candidates for studying the genetic and environmental enablers of their evolution.

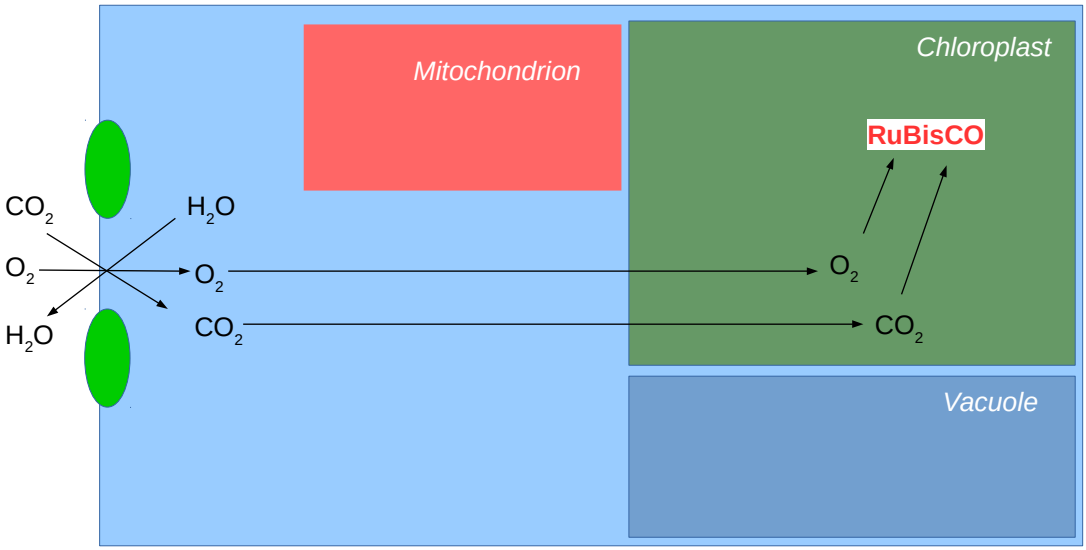


Figure 1.1: Schematic of the C_3 photosynthetic pathway

Schematic showing entry of CO_2 through open stomata (green ovals) into mesophyll cells during the day and direct fixation by RuBisCO. Consequentially, a) H_2O diffuses out of the leaf and b) O_2 enters, which can be fixed by RuBisCO, necessitating carbon recyling via photorespiration. Enzymes represented by white squares. See main text for acronyms.

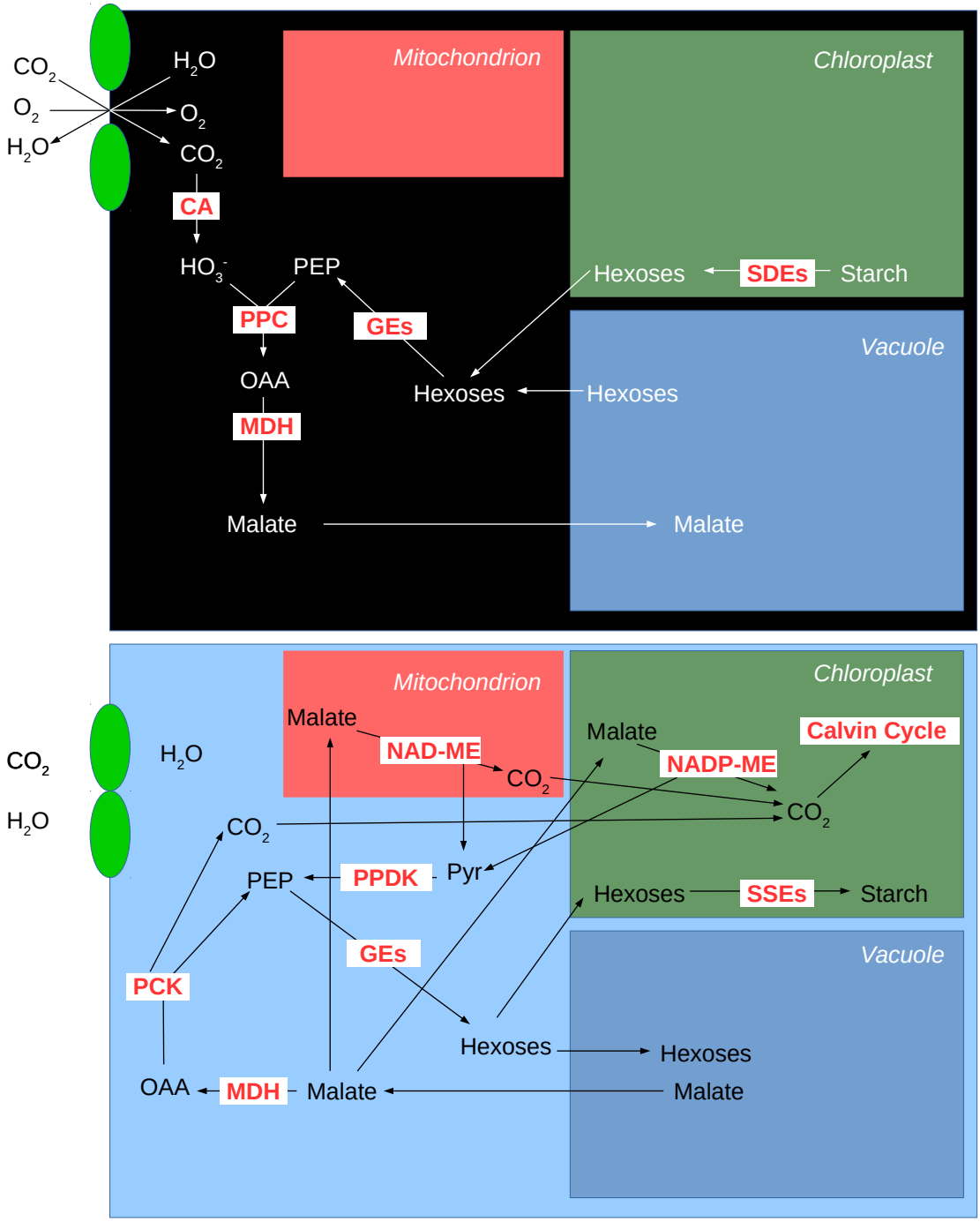


Figure 1.2: Schematic of the CAM photosynthetic pathway

At night (top panel), stomata (green ovals) open and CO_2 diffuses into leaf. H_2O diffuses out, but at a lesser rate than C_3 due to lower temperatures at night. CO_2 is hydrated by CA and fixed to PEP via PPC producing OAA. Hexoses are mobilised from the vacuole or from starch breakdown via starch breakdown enzymes (SBEs), and glycolytic/gluconeogenic enzymes (GEs) break hexoses down to produce PEP. OAA is converted to malate via MDH and stored in the vacuole. During the day (bottom panel), stomata are closed and water loss and gas exchanged are reduced. Malate is released from the vacuole and CO_2 is released via a) malate conversion to OAA via MDH, then decarboxylation via PCK in the cytosol, producing CO_2 and PEP, b) malate transported into the mitochondria/chloroplast and decarboxylated by NAD-ME/NADP-ME, producing CO_2 and pyruvate (Pyr), the latter of which is converted via PPDK to PEP. PEP is then converted to hexoses via GEs and either stored in the vacuole or converted to starch via starch synthesis enzymes (SSEs). Enzymes represented by white squares. See main text for acronyms.

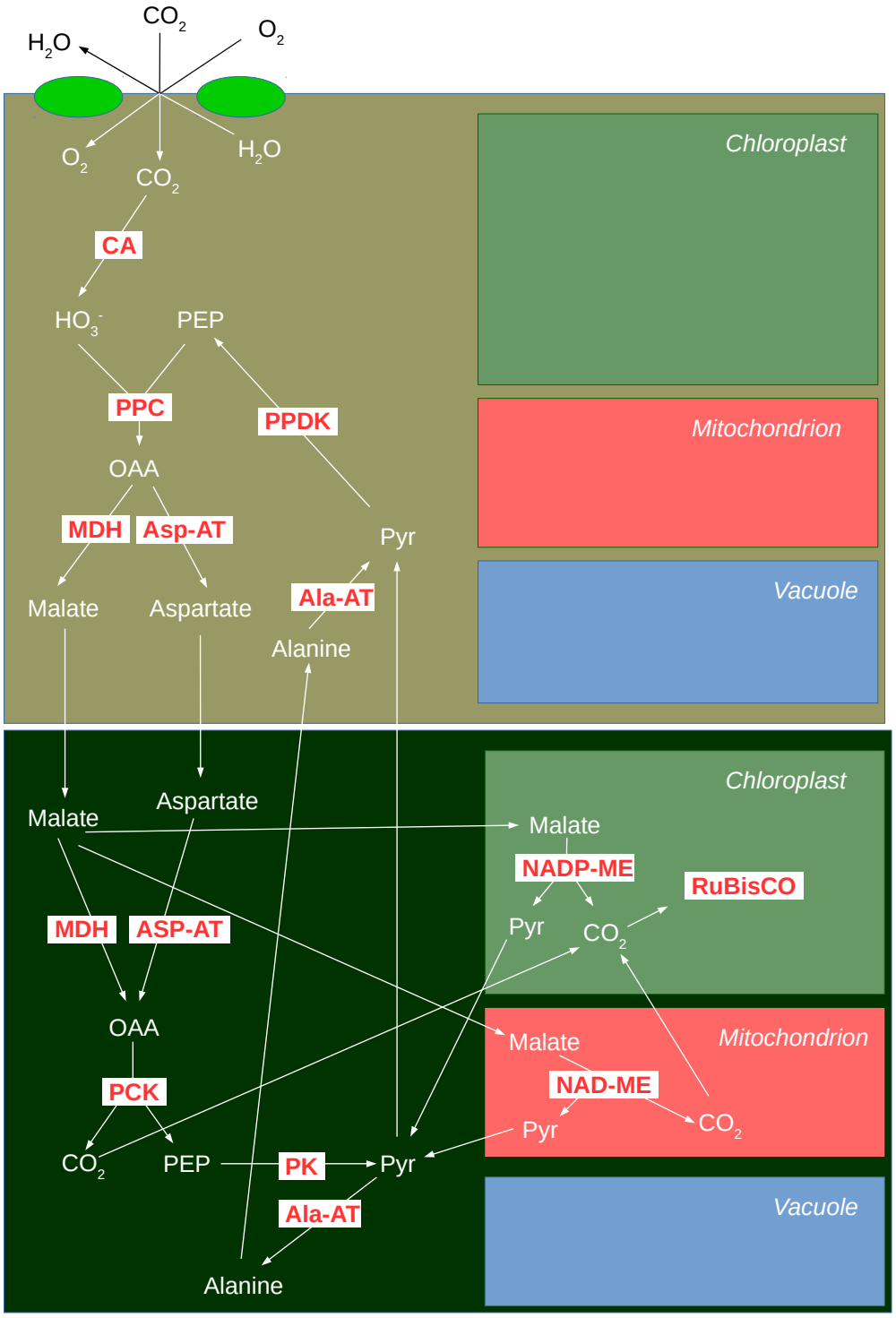


Figure 1.3: Schematic of the CAM photosynthetic pathway

CO_2 and O_2 enter the mesophyll cells during the day (top panel). CO_2 is hydrated by CA and fixed to PEP via PPC to OAA, converted to malate or aspartate via MDH or Asp-AT and transported into bundle sheath cells (bottom panel). Malate or aspartate can be decarboxylated via PCK to CO_2 and PEP, which is converted to pyruvate (Pyr) via pyruvate kinase (PK). Alternatively, malate can be transported into the mitochondria/chloroplasts and decarboxylated by NAD-ME/NADP-ME to CO_2 and pyruvate. CO_2 is carboxylated via RuBisCO, present exclusively in the bundle sheath cells. Pyruvate is transported back to the mesophyll cells, or converted to alanine via alanine-aminotransferase (Ala-AT) and transported to mesophyll cells before being converted back to pyruvate via Ala-AT. PEP is regenerated from pyruvate via PPDK. Enzymes represented by white squares. Acronyms are explained in the main text.

Despite the complexity of CCMs, they represent one of the most pervasive examples of convergent evolution (Conway-Morris, 2003). Molecular phylogenies have identified a large number of origins of CAM and C_4 photosynthesis (Sage et al., 2011, Edwards and Ogburn, 2012). C_4 photosynthesis is the most studied of the two CCMs, in part due to its importance in several important crop plants such as maize and sorghum (Sage and Zhu, 2011). Studies of convergent origins of C_4 have identified key environmental drivers, such as falling atmospheric CO_2 levels, that are associated with multiple C_4 origins in angiosperms (Christin et al., 2008; Christin et al., 2011a; Horn et al., 2014), with an additional effect of local factors, including temperature and aridity (Osborne and Freckleton, 2009; Edwards et al., 2010; Kadereit et al., 2012). From a genetic point of view, comparative analyses have demonstrated that some gene lineages were co-opted more often than expected by chance (Christin et al., 2007; Christin et al., 2013; 2015). The identity of genes co-opted was determined both by the genetic background (Christin et al., 2015), the ancestral expression levels (Christin et al. 2013; Emms et al., 2016; Moreno-Villena et al., 2018), and the cell and subcellular expression location (Christin et al., 2009; Clayton et al., 2017). Although there are fewer comparative studies focused on CAM origins, parallel radiation of CAM species have been identified in the past 30 MY in cacti and spurges (Arakaki et al., 2011; 10-15MYA) orchids (Silvera et al., 2009; Givnish et al. 2015; 20MYA) and bromeliads (Givnish et al., 2014; 5-15 MYA). This corresponds with a period of falling atmospheric CO_2 levels (Foster et al. 2017). Similarly, shared gene recruitment has been observed across CAM plants (Yang et al., 2017). It has also been suggested that the recruitment of genes for CAM photosynthesis was driven by gene expression levels within orchids (Silvera et al., 2014; Deng et al. 2016) and agaves (Gross et al. 2013; Abraham et al. 2016). These studies provide valuable insights into the potential biases of gene recruitment in these systems, but the recent global dominance of angiosperms (Lupia et al., 1999) means insights are limited to the most recent fraction of evolutionary diversification (Rensing, 2017). It is unclear, for example, whether CCMs would be as prevalent in today's relatively low CO_2 atmosphere if angiosperms had not risen to global dominance, or whether past atmospheres might have selected for CCMs in other groups of plants (Osborne and Beerling, 2006). The propensity of ancient clades of plants to evolve carbon concentrating mechanisms is unknown, but has been suggested (Decker and de Wit, 2006; Green, 2010; Cowling, 2013)**.** Similarly, the correlation between multiple independent climatic variables such as CO_2 levels, temperature and aridity makes the

relative importance of these factors unclear in many cases. It is therefore important to identify systems where genetic and environmental enablers of CCM evolution can be uncoupled.

1.6 A distant origin of CAM in lycopods

While most studies of CAM have focused on angiosperms, the genus *Isoëtes* offers a particularly powerful, yet largely unexplored system to investigate the environmental and genetic enablers of CAM evolution. This genus belongs to the lycopods, a basal lineage of vascular plants (Kenrick and Crane, 1997) that diverged from the rest of the vascular plants (including ferns, gymnosperms and angiosperms) over 393 million years ago (Granoff et al., 1976; Kenrick and Crane, 1997; Hoffman and Tomescu, 2013; Larsén and Rydin, 2015), and is the only surviving lineage of the once globally dominant Isoetalean lycopods (Pigg, 2001). It consists of approximately 200 species of aquatic and semi-aquatic plants, present in a wide range of environments from shallow pools to oligotrophic lakes (Taylor and Hickey, 1992). *Isoëtes howellii* was the first aquatic plant in which CAM was discovered (Keeley, 1981), with CAM activity subsequently found in the vast majority species (Keeley, 1982; Keeley, 1998). CAM activity in *Isoëtes* is evidenced by diurnal acidity changes, nighttime fixation of radiolabelled CO_2 which is stably converted to malic acid in the dark and subsequently decarboxylated and sufficient levels of carboxylase/decarboxylase activity for observed changes in malate (reviewed in Keeley, 1998). Whilst a number of angiosperms show diurnal acidity fluctuations, these are not necessarily exclusively associated with CAM activity, and only a small number of angiosperm genera (*Littorella, Vallisneria, Sagittaria* and *Crassula*) have been shown to perform CAM with high confidence (Keeley 1998). *Isoëtes* is significantly more speciose and widespread than these other genera (Keeley, 1998). When compared to the emblematic CAM angiosperms in arid systems, such as cacti and spurges, *Isoëtes* represents the most distant CAM relative, and occurs in one of the most divergent environmental contexts. It therefore offers an outstanding comparison point to explore the genetic and environmental factors affecting the evolution of CAM.

The radical environmental differences between terrestrial and submerged aquatic CAM species makes it unclear whether the same selection pressure drove CAM in these

two groups of plants and therefore, by some definitions, whether the two types of CAM represent the same trait at all (Aulio, 1986a; Neander, 1991). Terrestrial CAM is primarily considered a water conserving mechanism rather than a carbon concentrating mechanism (Herrera, 2009; Edwards and Ogburn, 2012; Borland et al., 2014), and it is assumed that local aridity even in high CO_2 atmospheres may have driven CAM evolution, although atmospheric CO_2 concentrations are likely to play at least some role (Keeley and Rundel, 2003). In contrast, it is accepted that CAM in submerged aquatic plants functions to concentrate carbon (Keeley, 1981; Keeley, 1998; Silvera et al., 2010, Pedersen et al., 2011). CO_2 is often limiting in submerged aquatic environments because of the relatively slow diffusion of CO_2 in water compared to air (Raven, 1970). Many submerged aquatic phototrophs have evolved strategies to enhance carbon uptake, from biochemical methods such as bicarbonate uptake, pyrenoids (micro-compartments consisting of densely packed RuBisCO surrounded by carbon transporters, present in some algae and hornworts; Meyer et al. 2008), CAM and C_4 to morphological methods such as lacunae to enhance uptake from CO_2 rich sediments or emergent or floating leaves to access aerial CO_2 (Raven et al., 2008; Maberly and Gontero, 2017). However, the significance of carbon concentrations in submerged aquatic environments depend on a large number of factors, such as vegetation density, pH and temperature, which have been suggested as more significant factors than atmospheric CO_2 levels in CAM evolution (Cole et al., 1994; Keeley, 1998). These potential factors, combined with the presence of CAM in ancient terrestrial or submerged aquatic lineages such as cycads, *Weltwischia* or *Isoëtes,* have led some to suggest that CAM may have evolved relatively early during the history of vascular plants (Keeley and Rundel, 2003; Luttge, 2004; Raven et al., 2008; Silvera et al., 2010). Although the conserved morphology of these plants over evolutionary time (Taylor, 1981) is compatible with CAM in their ancient relatives, the lack of CAM in many succulent or isoetid plants indicates morphology is not always an indicator of CAM activity (Rundel et al., 1999; Heyduk et al., 2016), and these ancient lineages could have evolved CAM relatively recently (Edwards and Ogburn, 2012). Understanding the evolutionary history of *Isoëtes* and contrasting it to that of CAM angiosperms may therefore shed new light on the environmental drivers of CAM, and the extent to which traits in different environmental conditions can really have the same function.

In addition to understanding the environmental drivers of CAM evolution, a comparison between angiosperms and the *Isoëtes* genus can also provide valuable information on the genetic enablers underlying CAM photosynthesis. While some have suggested that early land plants might have used CCMs (Cowling, 2013), there is no strong evidence to support this. Based on the distribution of CAM in the phylogeny of plants (Edwards and Ogburn, 2012), *Isoëtes* and the different groups of angiosperms likely evolved CAM independently. The ancient divergence of extant lycopods from angiosperms occurred prior to several changes potentially affecting the evolvability of complex traits in angiosperms. These include multiple genome duplication events (Banks et al., 2011; Jiao et al., 2011), increased frequency of alternative splicing (Zhu et al., 2017), and the evolution of true leaves, seeds and flowers (Kenrick and Crane, 1997). Recruitment of divergent components for CAM in these groups of plants would indicate a widespread capacity to evolve carbon concentrating mechanisms from diverse sets of components, and that particular biases towards certain components in angiosperm carbon concentrating mechanisms are likely to have emerged relatively recently. By contrast, high levels of similarity in the genomic basis of CAM in *Isoëtes* and angiosperms would indicate biases in the suitability of genes for recruitment for CAM were present in the common ancestor of vascular plants. This would suggest that the gain or loss of components in the presumably C_3 ancestor of land plants would have consequences for the evolvability of traits hundreds of millions of years later.

1.7 *Littorella*, a distant yet similar relative of *Isoëtes*

Comparison between the CAM angiosperms from arid environments and the aquatic CAM lycopods offers an opportunity to study CAM evolution broadly. However, the evolutionary distance and ecological contrast represent two factors that may be conflated. Differences between *Isoëtes* and CAM angiosperms such as pineapple or cacti may be due to their divergent genomic backgrounds, or may be due to the particular requirements of CAM in an aquatic environment. These factors can be disentangled by considering *Littorella uniflora*, an aquatic CAM eudicot that is highly convergent in morphology and ecology with some members of *Isoëtes*, such as *Isoëtes lacustris* (Boston and Adams, 1985; Smolders et al., 2002). Both species occur in oligotrophic lakes of northern Europe, often growing side by side (Rørslett, 1991). Besides their CAM physiology, shared adaptations to these environments include

exploitation of CO_2 derived from the microbial activity in sediments (Wium-Andersek and Andersen, 1972) and associated internal lacunae allowing gas exchange between roots and leaves (Søndergaard and Sand-Jensen, 1979; Richardson et al., 1984; Boston et al., 1987; Madsen et al., 2002). Likely limited by the range of diffusion within these air spaces (Madsen et al., 2002), both exhibit the "isoetid" growth form, being relatively prostate and lacking stems. This may also facilitate CO_2 uptake in leaves from water near the sediment (Pedersen et al., 1995). Comparison of the genetic basis and environmental drivers of CAM in these two species allows determining the extent to which environmental versus genetic effects account for gene recruitment into novel functions. Higher similarity of *L. uniflora* to *I. lacustris* than terrestrial CAM angiosperms would indicate that the type of environment determines the changes linked to CAM evolution, whereas higher similarity of *Littorella* to terrestrial CAM plants would indicate a key role of the genomic background and therefore evolutionary history of the species in which CAM evolved.

Together, *I. lacustris* and *L. uniflora* form a powerful system to disentangle the enablers of CAM evolution. Besides the origins of CAM in terms of gene co-option and environmental drivers, the overlapping distribution of these species enables comparisons of the intraspecific, population-level dynamics. As with any trait, CCMs must be assembled via the repeated action of natural selection on novel mutations. The efficiency of the process will depend on the genetic structure of the populations, and the dynamics of migration and genetic exchanges (Garant et al., 2007; Habets et al., 2007). Submerged aquatic environments are highly fragmented, with small patches of suitable habitat surrounded by regions where growth is not possible (i.e. terrestrial habitats). Highly fragmented habitats potentially reduce gene flow, increase genetic drift and reduce the ability to evolve in response to environmental changes (Young et al., 1996; Jump and Peñuelas, 2005). This could lead to fundamental differences in evolutionary processes between submerged and terrestrial plant populations (Barrett et al., 1993), and consequent changes in responses to selection pressures and the fixation of mutations leading to the co-option of novel CAM components. In addition to environmental influences on the structure of metapopulations, intrinsic plant properties can affect demographic processes, one of the most important of which is the mating and reproductive system (Loveless and Hamrick, 1984). Mating systems are among the key differences between the major groups of land plants, with sexual reproduction in

lycopods occuring via homospory/heterospory and via flowering in angiosperms. In the lycopods, haploid gametophytic spores are released from the plant that are either bisexual (homospory) or unisexual (heterospory), with fusion of sperms and eggs occuring outside of the sporophyte, typically requiring damp environments (Petersen and Burd, 2017). In angiosperms, by contrast, fertilisation occurs within the flower, with pollen grains containing the microspores (which go on to produce sperms) typically delivered to flowers by animal vectors (Ackerman, 2000) – wetting of pollen often results in premature germination and inviability (Philbrick and Les, 1996). These differences characterize the pair of *I. lacustris* and *L. uniflora*, and may have important effects on their intraspecific genetic structure, and hence the efficiency of natural selection, in these otherwise highly convergent species. While population genetics studies in these two species are unlikely to bring direct insights into the origins of CAM photosynthesis, a trait likely to be progressively acquired over long evolutionary times, they would provide a first assessment of the efficiency of small-scale evolutionary processes in these plants, bridging the gap between the micro- and macro-evolutionary processes enabling the evolution of complex traits.

1.8 Thesis aims and structure

This thesis aims to identify the environmental and genetic factors facilitation the evolution of complex traits, with the example of carbon concentrating in land plants. This overarching goal was achieved through an innovative review of the literature followed by three original research projects, which represent different aspects of the same problem, the enablers of the evolution of CAM photosynthesis. This work capitalizes on the comparison of members of *Isoëtes* and angiosperms (especially *Littorella*) that evolved the same phenotype from highly divergent genomic backgrounds and in a variety of environmental settings. The adopted comparative framework allows differentiation of environmental and genetic influences, and the extent to which each of these enables the evolution of complex traits.

Previously in the literature, the existence of evolutionary enablers has been assessed mainly with C_4 photosynthesis as a model system, with CAM mainly been considered as a distinct trait. Because both traits possess a number of fundamental commonalities, we decided to evaluate them jointly in a review of the literature (Chapter I). This effort suggested that submerged aquatic CAM plants represent the best system in which to investigate the large scale factors underpinning CAM evolution, as these systems capture the extremes of genetic, environmental and demographic contexts in which these traits have evolved.

To understand the selective drivers of complex trait evolution, it was important to determine whether the environmental factors promoting CAM evolution in terrestrial and aquatic environments were similar. Atmospheric CO_2 levels were identified as the only plausible driver that both terrestrial and aquatic CAM could share (Chapter I), and as these only vary over large temporal timescales, a molecular dating approach was used to infer the evolutionary history of aquatic CAM species. *Isoëtes*, the most speciose genus of aquatic CAM plants (Keeley, 1998) and potentially the oldest, was the focus of this study (Chapter III). Chloroplast datasets are easily generated and commonly used for phylogenetics, but variability in the rate of chloroplast genome evolution led to the additional inclusion of transcriptome data for *I. lacustris*, which was used as a scaffold for genome-skimming data for other *Isoëtes* species. The results indicate a relatively

recent origin of extant *Isoëtes* in the Paleogene (43-66 MYA), coincident with falling concentrations of atmospheric CO_2, and the origins of other groups of submerged aquatic CAM plants, as well as terrestrial CAM plants (Chapter III). These result suggest that the selective pressures involved in the evolution of CAM are indeed shared across broad environmental contexts and between distant lineages of plants.

Having established the similar environmental drivers of CAM evolution in terrestrial and aquatic plants, we subsequently tested whether these traits were convergent at the genetic level (Chapter IV). Genes recruited to CAM photosynthesis are potentially recognisable by their increased expression levels and diurnal expression patterns, both of which can be established using RNA-seq (i.e. transcriptome sequencing). These approaches offer a quantitative estimate of transcript abundance of most genes, but also provide their sequences, which can be used to establish orthology among distantly-related, non-model species. A series of experiments were undertaken to attempt to switch the levels of CAM activity in *I. lacustris* and *L. uniflora*, and the transcriptomes of plants grown in different conditions were sequenced. These new datasets were combined with those generated in Chapter III to identify the genes putatively involved in the CAM cycle of each species. The core CAM cycles of *L. uniflora* and *I. lacustris* shared many similarities, but comparisons with terrestrial C_3 (*Arabidopsis thaliana*) and CAM (*Ananas comosus*) plants showed that similarities stemmed from conserved expression patterns across large evolutionary scales. My results suggest that the genes ancestrally most abundant were preferentially co-opted independently in *I. lacustris* and *L. uniflora*, and therefore that the characteristics of the last common ancestor of all vascular plants dictated evolutionary trajectories hundreds of million years later.

The shared environmental drivers and convergence at the genetic level of these very distantly related species shown in Chapters III and IV led me to investigate whether *I. lacustris* and *L. uniflora* also exhibit similar population structures, despite differences in their mating systems (Chapter V). Individuals from both species collected across the United Kingdom were genotyped using restriction-associated digest sequencing to assess the distribution of genetic diversity produced in the region during its post-glacial recolonisation. My results revealed more population structure in *L. uniflora* than *I. lacustris*, consistent with clonal reproduction underwater in the former.

The existence of ecotypes of *L. uniflora* associated with different types of lakes might further suggest that local adaptation is slow in this species, potentially representing a low evolutionary potential due to its reproductive system.

Overall, my work suggests that falling atmospheric CO_2 prompted the parallel evolution of carbon concentrating mechanisms in diverse environmental contexts and in various groups of plants. The resulting CAM phenotypes were achieved via the co-option of ancestrally abundant genes, so that the ancestral condition of all land plants led to convergent evolution at the genetic level millions of years later, despite environmental and demographic differences among plant lineages.

Chapter I: Evolution and diversification of C_4 and CAM photosynthesis in land plants

Pascal-Antoine Christin[1], Daniel Wood[1]

1. Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

This chapter is a modified version of the published book chapter:
Christin, P.-A., and Wood, D. (2016). C_4 and CAM Photosynthesis in Land Plants, Evolution and Diversification of. In Encyclopedia of Evolutionary Biology, (Academic Press), p.254-259.

Personal Contribution: I co-wrote the manuscript with Pascal-Antoine Christin.

2.1 Abstract

C_4 and CAM photosynthesis are complex assemblages of anatomical and biochemical novelties that increase photosynthetic efficiency in a variety of environments, including warm, arid, saline and CO_2-poor aquatic conditions. Despite their complexity, C_4 and CAM evolved numerous times independently in land plants. These origins were facilitated by the presence of enablers in some lineages of plants and the existence of evolutionary stable intermediates. Both C_4 and CAM lineages diversified long after their initial origins, when Miocene aridification and opening of biomes provided new opportunities. During this diversification, different integrations of these photosynthetic types within organisms led to a diversity of new ecological strategies.

2.2 Introduction

It is remarkable that almost all assimilation of inorganic carbon into food chains around the world is performed by a single enzyme, ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO; Raven, 2013). This enzyme is used by all photosynthetic organisms for the fixation of atmospheric CO_2 in the Calvin-Benson cycle, which constitutes the light-independent phase of photosynthesis. Despite this quasi-universality, the enzyme seems rather poorly suited for the current conditions of Earth (Tcherkez et al., 2006). RuBisCO is estimated to have evolved more than 2.7 billion years ago (Nisbet et al., 2007), on an earlier planet Earth that was very different to today. The atmosphere of this time was extremely rich in CO_2 and almost devoid of O_2 (Kasting, 1993). RuBisCO happened to evolve with a propensity to confuse the O_2 and CO_2 substrates, two featureless molecules (Tcherkez et al., 2006). This was not problematic in the O_2-free environment of RuBisCO early history and was consequently not counter-selected. However, O_2 became very abundant in the atmosphere following the expansion of photosynthetic organisms some 2.4 billion years and CO_2 levels continuously decreased during Earth's history (Bekker et al., 2004; Kaufman and Xiao, 2003). The emergence of an atmosphere where O_2 is more abundant than CO_2 revealed the flaws of RuBisCO to natural selection (Sage, 1999; Christin and Osborne, 2013). When O_2 is abundant, it will be incorporated in a significant proportion of RuBisCO reactions. The products of O_2 fixation by RuBisCO are toxic and need to be recycled by the photorespiratory cycle (Ogren, 1984). This cycle consumes energy and releases CO_2 whilst recycling ribulose-1,5-bisphosphate, and is therefore often considered a wasteful

process. In total, up to 29% of light-energy is dedicated to photorespiration in the current atmosphere (Skillman, 2008), which strongly decreases plant productivity in conditions where relative CO_2 concentration is low (Zelitch, 1973).

Despite its flaws, RuBisCO was never replaced by a better CO_2-fixing enzyme, even though some exist in some other pathways (reviewed in Rothschild, 2008). This is probably because it was too integrated in the photosynthetic metabolism, which happens to be the most successful autotrophic process. RuBisCO enzymes with higher specificity were however gradually selected, which came at the expense of catalytic efficiency (Tcherkez et al., 2006; Young et al., 2012). In conditions where CO_2 depletion is strongest, this evolutionary fix reached its limits and plants had to find additional tricks to prosper. One of these is represented by the CO_2-concentrating mechanisms (CCMs), which solve RuBisCO's deficiencies by concentrating CO_2 around the enzyme, reducing the relative concentration of oxygen and therefore the amount of photorespiration (Sage, 1999; Christin and Osborne, 2013). In land plants, the most frequent CCMs are C_4 and CAM (Crassulacean Acid Metabolism) photosynthesis, two adaptive novelties that represent exceptional evolutionary and ecological success stories. In this paper, we will review the history of C_4 and CAM plants, from their evolutionary origins to their recent diversification across the globe.

2.3 C_4 and CAM photosynthesis, two adaptations that reduce photorespiration

C_4 and CAM operate using the same biochemical mechanism, but diverge in their spatiotemporal organisation. They use an enzyme other than RuBisCO to fix atmospheric CO_2 into organic compounds, namely phosphoenolpyruvate carboxylase (PPC). This enzyme has no affinity for O_2, but its product cannot be directly integrated into the Calvin-Benson cycle. Instead, the resultant four-carbon product is transformed and transported via different carbon shuttles until CO_2 is finally released by one of three possible decarboxylating enzymes to feed RuBisCO and the Calvin-Benson cycle (Osmond, 1978; Hatch, 1987). Other enzymes are then involved to regenerate the intermediate compounds of the cycles (Figure. 2.1). These additional enzymatic reactions increase the energetic cost of carbon fixation (Kanai and Edwards, 1999). In both CCMs, PPC acts as an additional filter on the atmospheric gases that can reach RuBisCO. The main consequence of this filter is that mostly CO_2 is available for RuBisCO, and photorespiration is strongly decreased (Skillman, 2008). This effect

requires a segregation of PPC and RuBisCO activities, and an isolation of RuBisCO from atmospheric gases. This is achieved spatially in C_4 plants and temporally in CAM plants (Fig. 2.1).

In C_4 plants, PPC and RuBisCO activities are synchronized and happen during the light period, when the photosystems are active and provides ATP (Hatch and Osmond, 1976). PPC activity is localized in compartments within the leaf that are in direct contact with atmospheric gases, which reach them via diffusion through the stomata (Lundgren et al., 2014). RuBisCO activity is segregated in compartments that are nested deeper within the leaf, where contact with the atmosphere is limited. In most C_4 plants, PPC is localized in mesophyll cells while RuBisCO is segregated in bundle sheath cells, which surround the veins and are encircled by mesophyll cells (Figure 2.1; Lundgren et al., 2014). This segregation can however also be achieved within a single cell, in which case PPC and RuBisCO are segregated in different areas generated through a re-organization of subcellular components (Edwards et al., 2004).

The main effect of the C_4 trait is to decrease photorespiration, but this benefit is partially offset by the extra energetic cost of the C_4 reactions, so that it is advantageous only in conditions where photorespiration rates are high (Ehleringer and Björkman, 1977). While low atmospheric CO_2 concentration is a necessary precondition for photorespiration, temperature plays an important role too. The solubility of CO_2 decreases faster with temperature than that of O_2 and the specificity of RuBisCO decreases with temperature, so that the relative O_2 concentration at RuBisCO active sites increases with leaf temperature (Ku and Edwards, 1978; Jordan and Ogren, 1981). C_4 is consequently mainly advantageous in warm climates, where most C_4 plants are distributed (Ehleringer et al., 1997). Aquatic plants can similarly gain an advantage from the C_4 trait in high light, high temperature environments with limited dissolved CO_2 (Keeley and Rundel, 2003). Because C_4 plants can photosynthesise at low CO_2 concentrations, they can also maintain carbon assimilation despite limited exchange with the atmosphere. This allows for a closure of stomata, which limits water loss and provides an advantage in arid and saline conditions (Osborne and Sack, 2012). Finally, the C_4 trait increases the number of CO_2 molecules fixed per RuBisCO protein, which improves nitrogen-use efficiency and can confer an advantage in nutrient-poor environments (Brown, 1978). In summary, the C_4 CCM is advantageous in all conditions where the benefit of reducing photorespiration offsets the additional energetic requirements (Sage et al., 2012; Christin and Osborne, 2014).

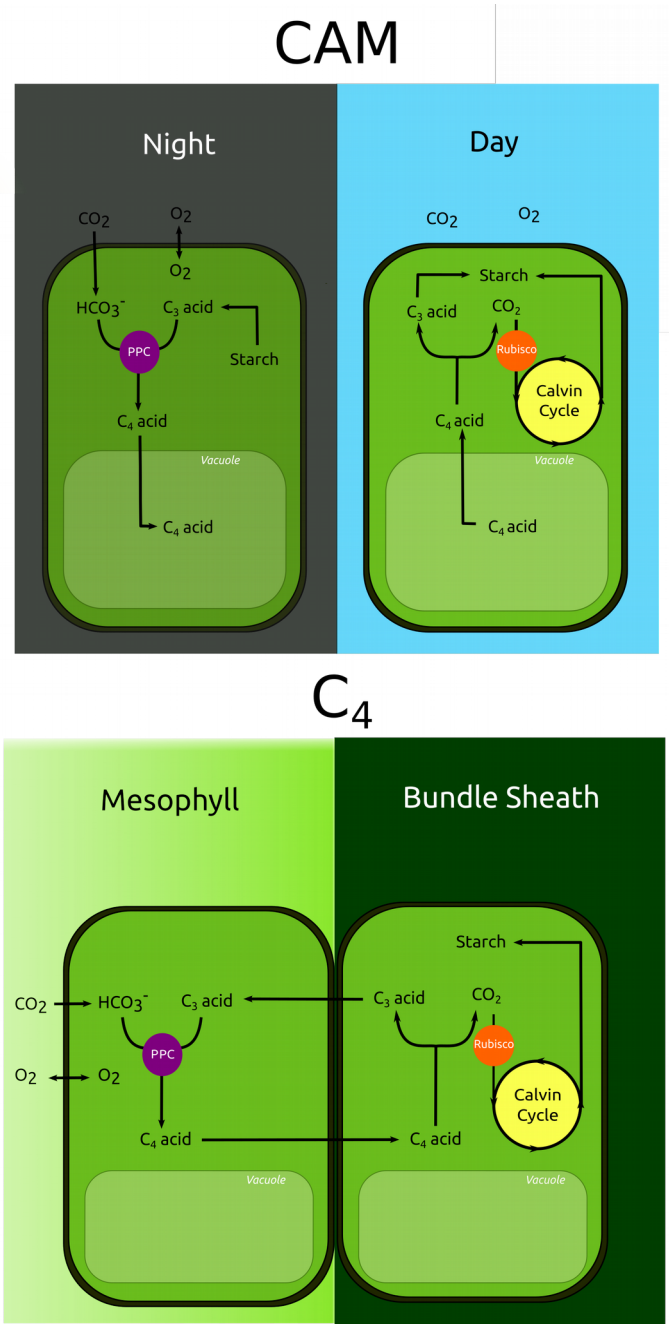


Figure 2.1: Schematics of the C_4 and CAM cycles.

In classical CAM, gas exchanges with the atmosphere take place at night. At this time, PPC is activated and catalyses the fixation of atmospheric CO_2 (in the form of HCO_3^-) to a C_3 acid to create a C_4 acid, which is stored in the vacuole (usually in the form of malate). In the day, PPC activity is downregulated and RuBisCO activity is upregulated. The C_4 acid leaves the vacuole and is decarboxylated, releasing CO_2 which can be fixed by RuBisCO and used to synthesise sucrose and starch via the Calvin cycle. In C_4, gas exchanges with the atmosphere take place primarily in the mesophyll cells. There, PPC is expressed and catalyses the fixation of atmospheric CO_2 (in the form of HCO_3^-) to a C_3 acid to create a C_4 acid, which is actively transported into the bundle sheath cells. Here it is decarboxylated, releasing CO_2 which can be fixed by RuBisCO and used to synthesise sucrose and starch by the Calvin cycle. Note in both cases the separation of RuBisCO activity and atmospheric oxygen.

In CAM plants, PPC and RuBisCO activities are not synchronized. Instead, the fixation of atmospheric CO_2 happens during the night, and the produced C_4 acids are stocked as malate in the vacuole. CAM is often associated with succulence, where large vacuoles allow the storage of high concentrations of malic acid (Nelson and Sage, 2008). The CO_2 is released from malate by decarboxylation during the light period to feed RuBisCO and the Calvin-Benson cycle, whose activity is synchronized with the photosystems (Osmond, 1978). The main consequence of CAM is to allow closure of stomata during the day, as atmospheric CO_2 is sequestrated in the plant during previous the night (Osmond, 1978). Stomatal opening leads to water losses through transpiration, which is exacerbated when stomata open during the higher temperatures of the light period. With water use efficiencies up to 16 fold higher than C_3 plants (Borland et al., 2009), CAM plants are consequently highly adapted to arid climates and saline environments, and several plants develop a CAM cycle only during periods of drought or high salt (Winter and Holtum, 2014). Some CAM plants can, in periods of extreme drought, keep their stomata closed throughout the diurnal cycle and recycle respiratory CO_2, which does not sustain growth but allows the plant to maintain functionality until more water is available (Lüttge, 2004). CAM species are also present in dry microenvironments in otherwise well-watered climates. For example, in tropical forests, epiphytism is a life strategy that is linked to limited water availability and there are consequently many CAM epiphytic orchids, bromeliads and ferns (Griffiths, 1989). CAM can even be advantageous when plants are fully immersed in water, which restricts CO_2 uptake during the day due to the low diffusivity of CO_2 in water (Keeley, 1998; Pedersen et al., 2011). When CO_2 fixation continues during the day, the night fixation of CO_2 by PPC extends the period of CO_2 uptake, potentially up to 24 hours. This process has also been shown to enhance carbon uptake in bromeliads in cloud forests where dew can inhibit gas exchange (Pierce et al. 2002). As in C_4, the additional reactions of CAM have an energetic cost and CAM plants can be associated with low growth rates (Lüttge, 2004). However, CAM Agaves and Opuntias rival the productivity of the most productive crops (Borland et al. 2009). Closing stomata comes at a cost, as it prevents oxygen produced during photosynthesis from leaving the from leaves, potentially increasing oxygenic stress towards the end of the day (Lüttge, 2002). In summary, CAM is associated with a variety of lifestyles, and similarly to C_4, can be associated with a number of ecological factors that all result in different ways from low CO_2 availability (Edwards and Ogburn, 2012).

2.4 Evolutionary origins of C_4 and CAM photosynthesis

The C_4 and CAM CCMs are traits of impressive complexity, which result from the co-ordinated action of multiple anatomical and biochemical components. Despite this complexity, each of them evolved multiple times independently. Over the last 15 years, phylogenetic efforts have elucidated the relationships between C_4 plants and those lacking this trait (e.g. Giussani et al., 2001; Kadereit et al., 2003; 2012; GPWGII 2012). These efforts have identified numerous monophyletic C_4 groups separated by other photosynthetic types in the phylogenetic trees. While some phylogenetic patterns might be interpreted in some cases as either multiple C_4 origins or fewer origins followed by losses of the C_4 trait (Duvall et al., 2003; Ibrahim et al., 2009), anatomical and biochemical differences among monophyletic C_4 groups as well as differences in the identity of genes co-opted to evolve the C_4 trait clearly point to a predominance of C_4 origins over losses (Christin et al., 2010), and it is estimated that C_4 originated more than 62 times independently in angiosperms (Sage et al., 2011). The number of CAM origins is not known with confidence, mainly because establishing whether specific plants are able to perform CAM can be challenging. For instance, some plants can switch to CAM depending on the environmental conditions or have a CAM cycle contributing to only part of their carbon assimilation (Winter and Holtum, 2014; Winter et al., 2015). In such cases, determining the photosynthetic type requires detailed physiological and biochemical analyses. Despite this uncertainty, CAM is present in distant phylogenetic groups, including lycophytes, gymnosperms, monocots and eudicots, and multiple origins are established in some groups (e.g. Crayn et al., 2004; Bone et al., 2015). Overall, it is estimated that the tally of CAM origins will most likely exceed the number of C_4 origins (Edwards and Ogburn, 2012). The apparent paradox between the complexity of CCMs and their recurrent origins is likely explained by the presence of CCM-like components in ancestors lacking these physiological adaptations. First, all currently known enzymes within the C_4 and CAM pathways exist in all plants, although they are ancestrally responsible for other functions (Aubry et al., 2011). They can in some cases be already abundant in photosynthetic organs, and it has been reported that C_4-like cellular localization of some enzymes existed before C_4 photosynthesis (Hibberd and Quick, 2002; Brown et al., 2010). The C_4-specific expression of some enzymes, moreover, co-opted pre-existing regulatory mechanisms (Brown et al., 2011). Low levels of CO_2 fixation in the dark, a possible precursor to

CAM, have similarly been detected in some plants (Ikeda and Yamada, 1981; Winter and Holtum, 2015), and the similarity of the reactions controlling stomatal opening to those of CAM has been suggested as a possible source of genetic pre-disposition for CAM evolution (Cockburn, 1981). C_4-like anatomical features also existed before the C_4 physiology (Muhaidat et al., 2011), and it has been shown that C_4 emerged from groups of plants that possessed C_4-like bundle sheaths (Christin et al., 2011b; Christin et al., 2013). Succulence, which is associated with enhanced water storage as well as providing a large vacuole for malic acid accumulation, has been suggested as an enabler of CAM evolution (Sage, 2002), and indeed osmotically active malate accumulation in the vacuole can potentially facilitate water uptake (Lüttge, 2004).

The evolution of CCMs consisted of the co-option of all the required anatomical and biochemical components. It is established that this happened in a stepwise manner for C_4, with the existence of evolutionarily stable intermediates (Sage, 2004; Christin et al., 2011b; Sage et al., 2012). These intermediates include plants with different degrees of C_4 physiology, such as several yellowtops (*Flaveria*), some heliotropes (*Heliotropium*), and the perennial wall-rocket (*Diplotaxis tenuifolia*; Sage et al., 2011). They are characterized by a weak CCM that relies on the segregation of photorespiratory reactions between the mesophyll and bundle sheath cells (Sage et al., 2012). This trait, referred to as the C_2 pathway, uses anatomical features that are close to the C_4 requirements (Sage et al., 2014), which may be advantageous in warm and dry conditions (Vogan and Sage, 2011). It therefore bridges the gap between the ancestral condition and C_4 plants, and models have suggested that the C_4 trait can be assembled through successively advantageous mutations from a C_2 ancestor (Heckmann et al., 2013; Mallmann et al., 2014).

Low levels of night-time CO_2 fixation, including recycling of respired CO_2, coupled with mostly RuBisCO-based daytime photosynthesis, allows plants to reduce stomatal conductance during the day and thus improve water-use efficiency. This physiological strategy can lead to the evolution of better integrated CAM systems, which further limit water losses, therefore potentially acting as evolutionary intermediates (e.g Sage, 2002; Edwards and Ogburn, 2012). The presence of CCM-like components in some lineages of plants eases the transition toward C_4 or CAM photosynthesis, and likely explains both the repeated origins of CCMs and their clustering in some parts of the phylogenetic tree (Sage, 2001; Sage et al., 2011; Edwards and Ogburn, 2012; Christin and Osborne, 2013).

2.5 Selective pressures and species diversification

While many environmental factors can be linked to the selective advantages of CCMs, changes in atmospheric gas concentrations are thought to be a necessary precondition for C_4 and CAM photosynthesis. Indeed, C_4 is predicted to gain an advantage at high temperatures only in extremely low CO_2 concentrations (Ehleringer and Björkman, 1977; Ehleringer et al., 1997) and the advantages of CAM in arid conditions are similarly tightly linked to CO_2 levels (Edwards and Ogburn, 2012). Molecular dating confirmed that all C_4 plants evolved in the relatively low-CO_2 atmosphere that persisted for the last 30 million years (Christin et al., 2008; Christin et al., 2011a), a time that might also have seen the emergence of some CAM groups, although earlier origins are possible (Edwards and Ogburn, 2012). While low CO_2 seems a necessary precondition for CCM evolution, it is not sufficient, and other factors that exacerbate photorespiration probably promoted the evolution of C_4 or CAM in the different lineages. In grasses, C_4 photosynthesis evolved in open habitats of the warm regions (Osborne and Freckleton, 2009; Edwards and Smith, 2010), while C_4 origins in Caryophyllales happened in dry and saline environments (Kadereit et al., 2012). In Bromeliaceae, CAM origins were more frequent in epiphytic taxa (Givnish et al., 2014), and aridity is generally seen as the main driver of CAM evolution (Edwards and Ogburn, 2012).

The evolutionary origins of CCMs are not linked in time to their ecological dominance. While C_4 origins are spread during the last 30 million years, the rise of C_4-dominated ecosystems is apparent in the fossil record in the last ten million years, and was driven mainly by C_4 grasses, which replaced either forested or open biomes depending on the geographical location (Edwards and Smith, 2010). This ecological dominance is also linked to increased numbers of species. The C_4 trait has indeed be shown to increase diversification rates, but again, this occurred long after the initial emergence of C_4, suggesting that C_4 diversification was influenced by other phenotypic traits and ecological changes in addition to the photosynthetic type (Bouchenak-Khelladi et al., 2014; Spriggs et al., 2014). The Miocene, which saw the rise to dominance and increased diversification of C_4 grasses, also witnessed the convergent radiation of the major CAM lineages, including cacti and other groups typical of arid climates (Arakaki et al., 2011; Horn et al., 2014). The selective pressures for the origins of the C_4 and CAM CCMs are therefore likely decoupled from those that increased their

ecological success. While Oligocene CO_2 decreases and local ecological conditions promoted the origins of CCMs, the aridification and expansion of open habitats during the Miocene likely triggered the expansion and species diversification of both CCMs.

2.6 Effect of CCMs on the ecological niche

Both CCMs confer a number of physiological characteristics that are potentially influenced by other attributes of the plant and can confer advantages in various conditions depending on how the new photosynthetic types are integrated within the organism. It is therefore not surprising that the ecology of plants with CCMs reflects that of the ancestors lacking such CCMs (Edwards and Ogburn, 2012; Christin and Osborne, 2014). The ecology of plants with CCMs is however also influenced by the changes that happened after the origin of the CCM. In grasses, inferences of the ancestral ecological niches concluded that C_4 lineages shifted to slightly more arid conditions compared to their ancestors (Edwards and Smith, 2010), and C_4 has been shown to increase statistically the rate of transition to both arid and saline habitats (Osborne and Freckleton, 2009; Bromham and Bennett, 2014). C_4 can therefore be considered as a niche opener. A recent study of C_4 ecological effects within a single species complex showed that C_4 initially broadened the niche without shifting it (Lundgren et al., 2015). This would allow young C_4 groups to explore new areas of the ecological space, with a possible subsequent specialization to more extreme conditions (Christin and Osborne, 2014). The evolution of CAM photosynthesis similarly facilitated the colonization of new niches. It has been suggested that CAM-like physiology might promote succulence, strengthening the ecological association to arid environments in some groups (Edwards and Ogburn, 2012). Phylogenies have confirmed that CAM facilitated the transition from humid to dry habitats in some terrestrial orchids (Bone et al., 2015). Similarly, CAM is likely lost in *Kalanchöe* expanding into wetter niches (Kluge et al., 2001). The ability of many CAM plants to plastically get rid of CAM photosynthesis may also open niches in stressful, changing environments such as near the shore in lakes (Aulio, 1985), and is thought to contribute to the wide range of niches occupied by the *Clusia* genus of trees (Lüttge, 2008).

2.7 Conclusion

C_4 and CAM photosynthesis are complex traits that were recurrently assembled by plants through the co-option of multiple anatomical and biochemical components. They represent an evolutionary strategy to address the affinity of RuBisCO for O_2, inherited from cyanobacteria, billions of years ago. Their main effects are to decrease photorespiration and increase water-use efficiency, which allow the colonization of a diversity of habitats. This includes warm, arid, saline, low-nutrient and aquatic habitats, all of which lead to a depletion of internal CO_2 concentrations. These ecological attributes favoured the spread and diversification of C_4 and CAM lineages during the Miocene, when open and arid environments expanded. In addition, the C_4 and CAM traits enable the exploration of new ecological niches when integrated with the other attributes of the organism. Overall, these properties contributed to the ecological success of C_4 and CAM plants, which nowadays cover most of the open habitats in the tropic and subtropical regions of the globe. C_4 photosynthesis alone is estimated to contribute up to 25% of terrestrial primary production (Still et al., 2003), and C_4 grasses have shaped multiple biomes, with an influence on major groups of herbivores, strong impacts on human evolution, and a key role in feeding the human population (Sage and Stata, 2015).

Chapter III: Palaeogene diversification of *Isoëtes*, a phylogenetically isolated lineage of CAM plants

Daniel Wood[1], Guillaume Besnard[2], David Beerling[1], Colin P Osborne[1], Pascal-Antoine Christin[1]

[1] Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, United Kingdom

[2] CNRS, Université de Toulouse, IRD, ENSFEA, UMR5174 EDB (Laboratoire Évolution & Diversité Biologique), 118 Route de Narbonne, 31062 Toulouse, France

Personal Contribution : I sourced the samples, performed the experiments and analyses and wrote the manuscript. Guillaume Besnard performed the next-generation sequencing. All authors provided comments on the manuscript.

3.1 Abstract

CAM is a complex trait most often associated with extremely arid, semi-arid and seasonally drought-prone environments, but the existence of aquatic CAM lineages demonstrates that different factors can select for this phenotype. Global factors, such as declining CO_2 levels in the atmosphere, could underlie all CAM origins, but the age of aquatic CAM lineages is not known with confidence. In this study, we combine transcriptomics and genomics with molecular dating to estimate the timing when the supposedly ancient aquatic CAM lineage *Isoëtes* diversified. Rate variation observed in genome-wide chloroplast markers hampers accurate dating, but nuclear markers place the origin of extant diversity within this group in the mid-Paleogene, 45-60 million years ago. This date coincides with the origins of terrestrial angiosperm CAM lineages such as cacti and spurges, with further diversification coincident with falling levels of CO_2. Using a land plant phylogeny, we further show that aquatic CAM in flowering plants also originated and diversified during the late Paleogene/Neogene. This suggests lowering atmospheric CO_2 levels likely played a significant role in facilitating the diversification of pre-existing CAM lineages in both aquatic and terrestrial environments. The “living fossil” *Isoëtes* benefited from the new ecological opportunities caused by global changes allowing it to rediversify and spread around the globe.

3.2 Introduction

Determining when and in which circumstances a given trait evolved is key to identifying its ecological drivers, but differentiating coincidence and causation is difficult when dealing with historical events. Convergent traits provide an opportunity to establish causation, since coincidences are unlikely to repeat themselves across multiple origins (Harvey and Pagel, 1998). Of particular interest is the effect of global events on the evolvability of different traits. Global conditions on Earth have varied tremendously during its history, including the average temperature, humidity and the composition of the atmosphere (Foster et al., 2017). While the effect of these changes on the diversification rates and success of some taxonomic groups has been widely discussed (Edwards et al., 2010; Nagalingum et al., 2011), these global changes may also have created the necessary preconditions to select for novel adaptations in combination with other, local factors. The importance of global factors is better tested with convergent traits for which local selective pressures are known to vary across the globe, effectively disconnecting global and local factors to evaluate their relative importance.

Higher plants had to develop key innovations to adapt to global changes, in particular biochemical carbon concentrating mechanisms such as C_4 and Crassulacean Acid Metabolism (CAM) photosynthesis. These traits separate the initial uptake of atmospheric CO_2 and its subsequent photosynthetic fixation (Hatch, 1987; Osmond, 1978). This separation occurs spatially in C_4 plants and temporally in CAM plants, allowing efficient photosynthesis in conditions that restrict CO_2 availability (Edwards and Ogburn, 2012). CAM and C_4 require numerous changes in gene expression patterns as well as leaf anatomy, yet both represent excellent examples of convergent evolution, each having evolved more than 60 times independently in various groups of plants (Sage et al., 2011; Edwards and Ogburn, 2012). The recurrent origins of C_4 photosynthesis have been associated with low atmospheric CO_2 concentrations that prevailed over the last 30 million years (Christin and Osborne, 2014), but the selective pressures underlying CAM photosynthesis are generally assumed to vary among groups (Keeley and Rundel, 2003; Edwards and Ogburn, 2012).

Terrestrial CAM plants open their stomata fix CO_2 at night, allowing stomata to close during the day when temperatures are higher and the air is drier, which reduces water losses (Martin et al., 1988). The CAM trait is therefore usually associated with

arid environments (Edwards and Ogburn, 2012; Keeley, 2014; Winter and Holtum, 2014). However, the discovery of CAM in submerged aquatic plants (Keeley, 1981) showed clearly that CAM is not solely an adaptation to aridity. Instead, it has been suggested that CAM and other carbon concentrating mechanisms provide a benefit underwater when CO_2 levels are low (Giordano et al., 2005), with the night-time uptake of CAM being advantageous in avoiding daytime competition, or when CO_2 levels are constantly very low (Keeley, 1981; Keeley and Rundel, 2003; Pedersen et al., 2011). In terrestrial plants, low CO_2 levels force increased stomal conductance (Farquhar et al., 1978), thereby increasing CAM advantages in arid environments, and a low CO_2 world has been suggested as a potential global precondition for CAM evolution (Keeley and Rundel, 2003; Edwards and Ogburn, 2012). Testing this hypothesis is however complicated by uncertainty in the taxonomic distribution of CAM and paucity of fossil CAM plants (Ehleringer and Monson, 1993), so that the timing of CAM evolution cannot be directly evaluated.

Molecular dating represents a powerful alternative, and has shown that terrestrial CAM lineages concomitantly increased in diversification in a period of low CO_2 levels coupled with global aridification (Silvera et al., 2009; Arakaki et al., 2011; Givnish et al., 2014). The timing of aquatic CAM evolution remains however poorly studied. The prevalence of CAM in one of the earliest diverging groups of land plants (*Isoëtes*, in the Lycopods) has led to the widespread assumption that aquatic CAM in *Isoëtes* is very ancient (Griffiths, 1992; Ehleringer and Monson, 1993). This was corroborated by the existence of fossils closely resembling *Isoëtes* dating from the Jurassic (Ash and Pigg, 1991), and similar fossils from the Permian and the Triassic (Retallack, 1997; Cantrill and Webb, 1998; Kustatscher et al., 2010; McLoughlin et al., 2015). While the relationship between these fossil species and extant lineages of *Isoëtes* is unclear, recent molecular dating studies seem to confirm the old origin of extant *Isoëtes* (Larsén and Rydin, 2015; Kim and Choi, 2016; Pereira et al., 2017). These previous studies were however based on a limited number of markers, mainly from the chloroplast genome, where high rate variation can make dating estimates strongly dependent on model assumptions (Christin et al., 2014). Given the large evolutionary distance between *Isoëtes* and its sister group *Selaginella*, we conclude that the divergence times need to be reevaluated using a combination of phylogenomic methods in order to accurately resolve the timing of CAM diversification across aquatic plants.

In this study, we generate genomic datasets for multiple *Isoëtes* species and other angiosperms, and apply different molecular dating approaches to (i) estimate the time to the most recent common ancestor of extant *Isoëtes* based on nuclear and plastid genomes. We then analyse representatives of the other aquatic CAM lineages and their non-CAM relatives to (ii) evaluate the coincidence of CAM origins and diversification with falling levels of atmospheric CO_2 and the origins of CAM in terrestrial plants. Our results shed new light on the global factors driving convergent innovations across distantly related land plants.

3.3 Materials and Methods

General approach

In this study, we generated genome-wide DNA datasets for six *Isoëtes* species selected to capture the deeper divergence events within this group (Larsén and Rydin, 2015) and analyzed different genome partitions in isolation to get accurate estimates of divergence times within the group. Herbarium specimens represent a useful source of DNA (Besnard et al., 2014), particularly for globally distributed, hard to access groups such as *Isoëtes*. Low-coverage whole-genome scans can be applied to these samples, and will yield high coverage for genomic fractions present as multiple copies, such as the organellar genomes (Bakker et al., 2016). However, high evolutionary rate variation in chloroplast markers is known from seed plants (Bousquet et al., 1992; Ruhfel et al., 2014), which can affect the results of dating methods that differ in their assumptions of rate heterogeneity (Christin et al., 2014). Previous studies of the chloroplast marker *rbcL* in *Isoëtes* indicate much higher rates of sequence evolution in *Selaginella* than in *Isoëtes* (Karol et al., 2010; Ruhfel et al., 2014; Larsén and Rydin, 2015), urging for a consideration of nuclear markers, which can be more useful for molecular dating if they show less variation in rates among branches (Christin et al., 2014). Genome skimming can provide nuclear sequences, but low coverage makes *de novo* assembly difficult. However, the sequencing reads can be mapped to a reference dataset, providing phylogenetically informative characters (Olofsson et al., 2016). A reference genome is available for *Selaginella*, but it is too distant from *Isoëtes* to allow accurate read mapping. Transcriptomes provide high coverage of expressed protein-encoding genes, which represent regions of the genome allowing read mapping across distinct species (Olofsson et al., 2016). We consequently decided to generate and assemble a transcriptome for a single *Isoëtes* species, which was used as a reference to map reads from low-coverage whole-genome sequencing datasets obtained from several *Isoëtes* species sampled from herbarium collections and selected to capture the deeper divergence events within *Isoëtes*. The sequencing data were used to obtain chloroplast and nuclear alignments for five *Isoëtes* species as well as a number of other land plants sequenced in other studies. The resulting datasets included the main lineages of land plants, allowing the incorporation of fossil evidence providing calibration points spread across the tree. The widely available chloroplast marker *rbcL* was then retrieved for a

large number of species representing the other, species poor lineages of aquatic CAM plants (Keeley, 1998) and molecular dating was performed, enabling comparison to the age of *Isoëtes*.

Sequence acquisition

Live *Isoëtes lacustris* were sampled from Cwm Idwal, Wales and maintained at the University of Sheffield in 40 x 30 x 25 cm transparent plastic containers, with a substrate of sand to a depth of 5cm, and the containers filled to the top with deionised water. These were placed in a Conviron growth chamber with a 12h light/dark cycle, 495 µmol m^2s^{-1} light, light and dark temperatures of 18 and 20 °C and CO_2 at 400ppm for six days. To maximise the number of transcripts retrieved, leaves from three individuals were sampled 3 hours after the onset of the dark period and 3 hours after the onset of the light period light and flash frozen immediately in liquid nitrogen. Individuals of *Littorella uniflora* also sampled from Cwm Idwal were grown under a variety of conditions and leaves were sampled as described above.

RNA was extracted from the sampled leaves using the Qiagen RNeasy® Plant Mini Kit, following the manufacturer protocol, with the addition of on-column DNase I digestion (Qiagen RNase-Free DNase Set). 2.5µl SUPERase-In RNase inhibitor was added to 50µl of extracted RNA to stabilise it, and RNA was quantified using a gel electrophoresis, RNA 6000 Nano chips in an Agilent 2100 Bioanalyser, and a Nanodrop 8000. Samples were then prepared for Illumina sequencing using the TruSeq® RNASample Prep Kit v2. Paired-end sequencing was performed on an Illumina HiSeq 2500 platform available at the Core Genomic Facility of the University of Sheffield in rapid mode for 100 cycles, with 24 libraries pooled per flow cell (other samples were from the same or different projects).

DNA from herbarium specimens of five *Isoëtes* species were acquired from the DNA Bank from the Royal Botanical Gardens, Kew. This was supplemented with one silica gel dried leaf of *I. lacustris* and *L. uniflora* collected from the field (Cwm Idwal, Wales). Whole genome sequencing of these seven samples was performed at the Genotoul from the University of Toulouse, using previously described protocols (Lundgren et al., 2015; Olofsson et al., 2016). Each sample was sequenced on a 24^{th} of a flow cell, with other samples from distinct projects. Raw sequencing reads were cleaned using NGS QC toolkit v2.3.3 (Patel and Jain, 2012) by removing adapter sequences,

reads with ambiguous bases and reads with less than 80% of positions with a quality score above 20. Low quality bases (q<20) were removed from the 3' end of remaining reads. Species identity was confirmed by assembling the nuclear ribosomal internal transcribed spacer (nrITS) using NOVOPlasty (Dierckxsens et al., 2017) and comparing to *Isoëtes* nrITS sequences from the NCBI database using blastn (Altschul et al., 1990; see Supplementary Table 3.1).

Chloroplast data matrix

Cleaned reads from the *Isoëtes* and *Littorella* genome skimming datasets corresponding to the chloroplast genomes were assembled using NOVOPlasty, with a 25 bp kmer and a seed sequence of a conserved portion of *rbcL* for *Isoëtes* and *Littorella*, respectively. Chloroplast genome assemblies from 24 additional species representing the major Embryophyte taxa, including two *Selaginella* species, were downloaded from NCBI database (see Supplementary Table 3.2*).* Chloroplast protein-coding genes were identified using DOGMA (Wyman et al., 2004) and coding sequences were extracted using TransDecoder v2.1.0 (Haas et al., 2013). A total of 65 genes were identified and aligned by predicted amino acids using t-coffee (Notredame et al., 2000) and MAFFT v7.164b (Katoh et al., 2002). Gene alignments were manually inspected using AliView (Larsson, 2014) and eight of them were discarded due to poor homology and alignment difficulties among these distantly related species. The remaining 57 chloroplast genes were concatenated, producing a 55,742 bp matrix, with 33,496 polymorphic and 25,502 parsimoniously informative sites. A maximum likelihood phylogeny was generated using RaxML v8.2.11 (Stamatakis, 2014), with a GTR+G+I model of sequence evolution, determined to be the best fit model using likelihood ratio tests. The same matrix was later used for molecular dating (see below).

Nuclear data matrices

Cleaned RNAseq reads of *Isoëtes lacustris* were assembled using Trinity v2.3.2 (Haas et al., 2013). The longest open reading frames (ORFs) were extracted using TransDecoder and for each unigene the contig with the longest ORF was used to assemble a reference dataset. Cleaned reads from the whole-genome sequencing (genome skimming) datasets were then separately mapped to this reference dataset

using bowtie2 v2.3.2 (Langmead and Salzberg, 2012) in local mode to avoid excluding reads from exon/intron overlaps. Alignments with MAPQ quality below 20 were excluded using SAMtools v1.5 (Li et al 2009). The SAMtools mpileup utility was then used to generate separate consensus sequences for each species of the reads mapping to each *I. lacustris* transcript. Transcriptome and coding sequence data from seven additional species representing different Embryophyte groups were downloaded (see Supplementary Table 3.3), and ORFs extracted.

Gene duplication and losses are common in nuclear genomes, so a combined reciprocal best blast and phylogenetic approach was taken to identify groups of co-orthologs covering *I. lacustris* and the other land plants. Families of homologous ORFs generated by the method of Vilella et al., (2009) were downloaded from EnsemblPlants. In total, 4,516 highly conserved homolog families among land plants (containing at least one sequence from *Physcomitrella patens, Selaginella moellendorffii, Amborella trichopoda, Oryza sativa, Arabidopsis thaliana* and *Theobroma cacao)* were used for subsequent ortholog identification. Reciprocal best protein BLAST searches assigned ORFs of *I. lacustris* and the additional Embryophyte species to homolog families, with a minimum match length of 150 and e-value of 10^{-7}. The expanded homolog families were then aligned according to their protein sequences using MAFFT, and phylogenies were constructed using RaxML and the GTR+G+I model, which fits most genes well and is therefore appropriate for constructing large numbers of gene trees (Fisher et al., 2016; Dunning et al., 2017). The longest sequence of each monospecific clade of sequences was kept to remove transcripts representing the same gene or genes that duplicated after the divergence from all other species. The datasets were then realigned and a new phylogeny was inferred. Sets of 1:1 orthologs were then identified as clades containing exactly one gene per species, resulting in 30,258 groups of co-orthologs. Of these, 2,165 contained more than nine species, including *I. lacustris, S. moellendorffii* and either *Physcomitrella patens* or *Ceratodon purpurea,* which were needed to use some of the fossil calibration points (see below). Orthogroups were realigned, and consensus sequences of the genome skimming data were added to the alignments. Only the 782 orthogroups containing sequences for *I. coromandelina*, a representative of the earliest diverging *Isoëtes* lineage based on a previous study (Larsén and Rydin, 2015), were considered further. New phylogenetic trees were inferred from these datasets, and genes failing to recover the monophyly of the vascular plants, *Isoetopsids (Isoëtes + Selaginella*) or *Isoëtes* were considered phylogenetically uninformative and excluded.

The remaining 181 datasets were deemed suitable for the phylogenetic problem addressed here, and were used for molecular dating. A phylogenetic tree was inferred separately for each of these markers, and a maximum likelihood phylogeny was also inferred using the 841,154 bp concatenated alignment, which was 33.96% complete with 443,864 polymorphic and 316,350 phylogenetically informative sites.

Calibration points and molecular dating

Time-calibrated trees were inferred from the different markers using the same calibration points. For the chloroplast dataset, trees were rooted by constraining liverworts and the rest of the land plants to be monophyletic (Kenrick and Crane, 1997). For the nuclear dataset, which only contained bryophytes and vascular plants, the tree was rooted by enforcing the monophyly of each of these two groups.

A maximum age constraint of the crown node of all land plants was designated based on the appearance of cryptospores in the fossil record. These abundant spores are considered a likely synapomorphy of early land plants (Wellman, 2010). Their appearance in the fossil record is therefore likely to occur soon after the origins of land plants, making them appropriate for setting a maximum age for land plants (Larsén and Rydin, 2015, Morris et al., 2018). The earliest unequivocal cryptospores are found in the early Middle Ordovician (473-471 MYA; Rubinstein et al., 2010). However, pre-Middle Ordovician terrestrial sediments are rare (Wellman and Strother, 2015), and as no unequivocal cryptospores are found in pre-Ordovician rocks (Kenrick, 2003; Morris et al., 2018) the beginning of the Ordovician (485 MYA) was used as a conservative upper limit for the age of land plants. This maximum age was used to constrain the crown node of the liverworts + rest of vascular plants in the chloroplast dataset, and the crown node of the bryophytes + vascular split in the nuclear dataset. The minimum age of the same node in both cases was constrained by the earliest vascular plant macrofossil, *Baragwanathia longifolia* from the Ludlow epoch in the Silurian at 421 MYA (Garratt et al., 1984; Kenrick and Crane, 1997; Magallón et al., 2013).

A third calibration point was assigned to the crown node of the *Isoetopsida* (*Isoëtes* plus *Selaginella*). Isoetalean lycopsid trees are considered to form a clade within the *Isoetaceae*, based on synapomorphies including bipolar growth from a shootlike “rhizomorph” structure and secondary woody tissue (Pigg, 2001). Arborescent lycopsids are known from the Frasnian (382.7-372.2MYA; (Stein et al., 2012; Berry and

Marshall, 2015), although the rhizomorph root structure could not be identified in these early fossils. However, discovery of a putatively homosporous arborescent lycopsid (the *Isoetales* are heterosporous) suggests that arborescence could be a convergent phenotype within the lycopods (Xu et al., 2012). As multiple examples of isoetalean arborescent lycopsids, including rhizomorphs, are known from Fammenian strata (358.9 to 372.2MYA; Wang and Berry, 2003; Cressler and Pfefferkorn, 2005), a minimal age of 358 MYA was implemented using a uniform distribution between 358 and 485MYA.

Molecular dating was performed using r8s (Sanderson, 2003) and BEAST (Drummond and Rambaut, 2007), two commonly used relaxed-clock methods that differ in their general approach and the strategy used to assign rates to internal branches of the phylogeny. r8s is a semiparametric method that uses a penalised likelihood approach to assign rates among branches (Sanderson, 2002). The smoothing parameter, which determines the extent to which rates vary among branches, is determined for each dataset using an empirical approach (Sanderson, 2003). The method takes a phylogram as an input, assumes no uncertainty in topology, and uses a simplified model of nucleotide substitution. BEAST is a highly parametrised Bayesian method that samples trees generated from nucleotide data using an explicit model of sequence evolution (Drummond et al., 2006). Rates are uncorrelated across the tree, but an overall distribution of rates is assumed.

For BEAST, version 1.5.4 was used. A lognormal clock was used with a GTR + G + I model of nucleotide substitution with four rate categories and a birth-death speciation prior. For the concatenated chloroplast markers, four independent analyses were run for at least 20,000,000 generations and appropriate burn-in periods (at least 10%) were assigned by inspection using Tracer v1.6 (Drummond and Rambaut, 2007). For individual nuclear genes, BEAST was run for 3,000,000 generations (based on observing convergence times with a subset of genes) with a burn-in of 50%.

For r8s, version 1.81 was used, with the “TN” algorithm manual and additive penalty function. Cross validation was performed for a range of smoothing parameters from 10^{-2} to 10^{6}, increasing by a power of $10^{0.5}$ each time, and the identified smoothing parameter was used for molecular dating. Confidence intervals were obtained by generating 100 bootstrap pseudoreplicates using seqboot (Felsenstein, 2002) and obtaining branch lengths for each of these using RaxML, constraining the trees to the topology generated by the full dataset. These trees were then individually dated using r8s, providing a distribution of ages across the pseudoreplicates.

Dating the origin of other aquatic CAM lineages

In order to compare the crown dates obtained for *Isoëtes* with other aquatic CAM groups, molecular dating was performed for a tree including other high-confidence CAM lineages (*Isoëtes*, *Vallisneria*, *Sagittaria*, *Crassula* and *Littorella*; Keeley, 1998), and non-CAM taxa separating them in the phylogeny. The chloroplast marker *rbcL* was used for this analysis, since it was the only one available that was sufficiently conserved to allow accurate alignment across all of these taxa. A total of 118 *rbcL* sequences representing the main embryophyte lineages and families/orders containing aquatic CAM taxa were downloaded from NCBI nucleotide database (Supplementary Table 3.4). Sequences were aligned using MAFFT. Molecular dating using BEAST v1.5.4 was performed as described previously, but with an additional fossil constraint on the eudicots. The presence of tricolpate pollen in the fossil record from 125-135MYA has been used to set a maximal age for the emergence of eudicots (Anderson et al., 2005, Christin et al., 2014, Magallón et al., 2015) since abundance of pollen means that its appearance in the fossil record is unlikely to occur significantly after its evolution. A minimum age for the crown node of the eudicots was assigned as 112MYA, using a fossil flower from the Early Cretaceous with affinities to the Ranunculales (von Balthazar et al., 2005; Magallón et al., 2015). The crown node of eudicots (*Ranunculus macranthus* + the rest of the eudicots) was constrained with a uniform distribution between 112 and 135MYA, and the stem node (eudicots + *Ceratophyllum demersum*) constrained to between 125 and 485MYA. The monophyly of the following taxa was constrained to get time estimates independent of the resolving power of *rbcL* for these deep nodes: land plants excluding *Marchantia*, vascular plants, lycopods, Isoetaceae, seed plants, angiosperms, monocots, eudicots, *Saxifragales, Plantaginaceae, Alismataceae, Hydrocharitcae*, *Plantago* + *Littorella* and *Plantago.* r8s was run as previously, with the same constraints as described for BEAST. The topology of the input tree was constrained to that of the maximum clade credibility tree from the BEAST analysis, with a GTR+G+I model in RAxML used to generate branch lengths.

3.4 Results

Phylogenetic reconstruction and dating based on the chloroplast genome

The maximum likelihood phylogeny based on chloroplast markers recapitulated major land plant relationships and expected relationships within the *Isoëtes*, with *I. coromandelina* representing the earliest divergence within the group (Figure 3.1). The tree was well resolved, with only the *Ceratophyllum*/eudicot split receiving less than 95% bootstrap support. Branch length variation was high, particularly between *Isoëtes* and *Selaginella*, with the latter having accumulated approximately 4.5 times more substitutions than *Isoëtes* since their most recent common ancestor (Figure 3.1).

r8s estimated the age of the crown group of *Isoëtes* at 24.15 MYA with an optimum smoothing parameter of 1000 identified by cross validation, and a 95% bootstrap confidence interval of 23.38-27.40 MYA (Table 3.1). Decreasing the value of the smoothing parameter resulted in an increased age of the *Isoëtes* crown group, with a log smoothing value of 0.01 giving a crown age of *Isoëtes* of 218 MYA (Figure 3.2). Whilst low smoothing values result in overfitted models that perform poorly in cross validation, high levels of smoothing may produce rates that are nevertheless poor predictors of branch lengths in particular parts of the tree. For high smoothing values, the ratio of the “effective” rate (the branch length divided by the estimated time elapsed) to the rate assigned by the model was 0.33 for the stem branch of *Isoëtes* (Figure 3.3), showing that the branch is significantly shorter than would be expected for the assigned rate and divergence time. On the other hand, the average ratio for the crown branch lengths was 1.33, indicating that the crown branches are longer than would be expected for the assigned rates and divergence times (Figure 3.3).

BEAST estimated the crown of *Isoëtes* at 23.7 MYA with a 95% highest posterior density (HPD) interval of 6.2-48.9 (Table 3.1), similar to the value obtained with the optimum level of smoothing in r8s. Unlike r8s, BEAST rates can vary independently throughout the tree, but a prior distribution of rates across the tree is assigned – in this case a lognormal distribution. Rates in the maximum clade credibility tree follow a lognormal distribution with the *Isoëtes* stem branch being assigned the lowest rate in the tree, with crown branches assigned rates closer to the average rates in the rest of the tree (Figure 3.4). For both r8s and BEAST, a date of 23-24MYA is obtained via the implicit or explicit inference of a decrease in the rate of evolution along

the stem branch, with rates in the crown branches being more similar to those in the rest of the tree. This assumption results from the model, and is not necessarily correct, urging for independent evidence.

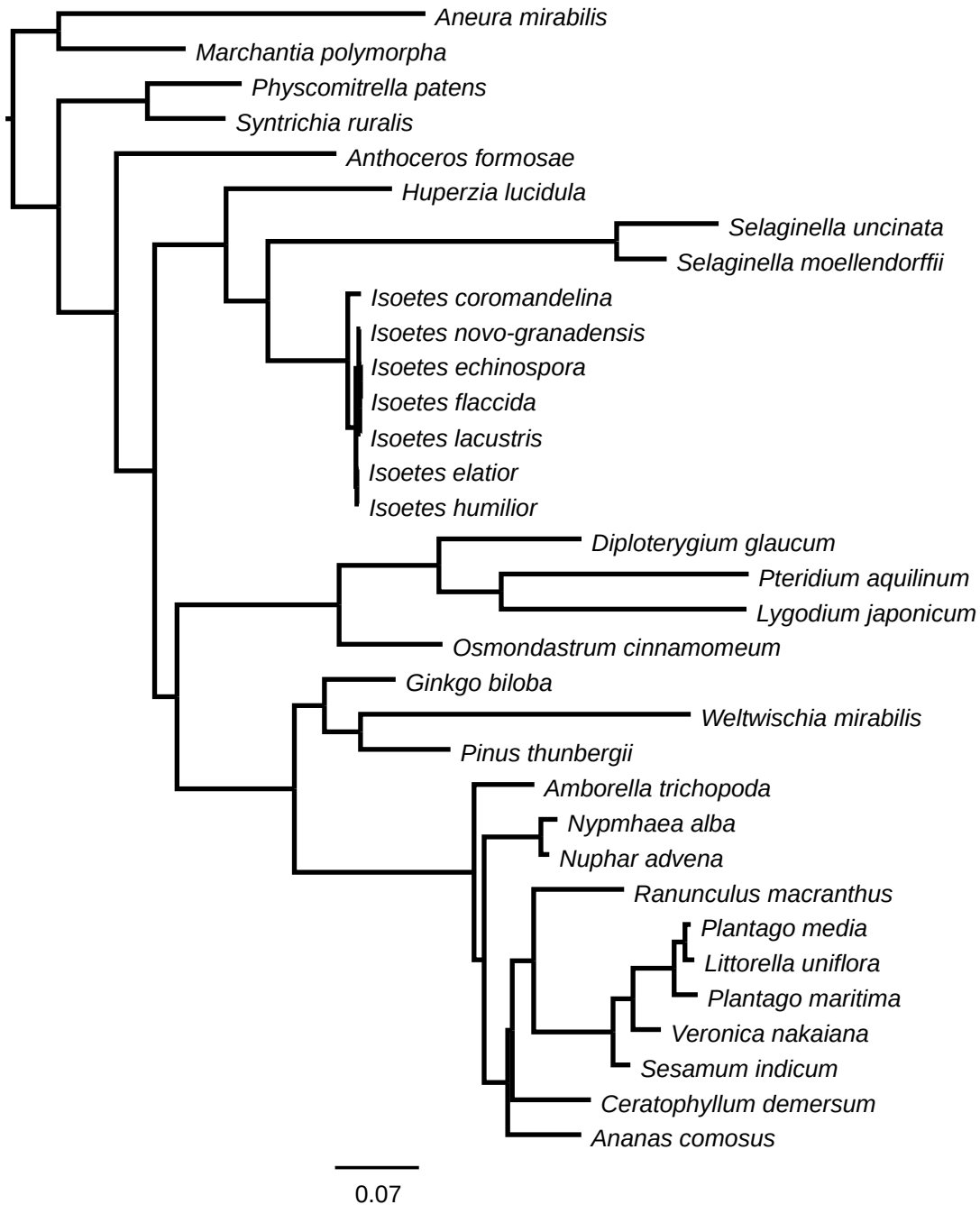


Figure 3.1 Maximum likelihood phylogram of concatenated chloroplast markers

Branch lengths are proportional to the number of substitutions per site, with scale bar representing 0.07 substitutions per site. All bootstraps support values are 100 with the exception of the branch separating *A. comosus* from the clade containing *C. demersum*, which has a support value of 76.

Table 3.1: Estimates of *Isoëtes* crown date

Analysis	***Isoëtes* crown age - BEAST (95% CI)**	***Isoëtes* crown age - r8s (95% CI)**
Chloroplast concatenated markers	23.3 (5.5-51.7)	24.18 (22.3-25.8)
Nuclear concatenated markers		58.94 (57.3-60.1)
Nuclear individual markers	47.6 (24.1, 90.1)	46.43 (16.1, 85.8)

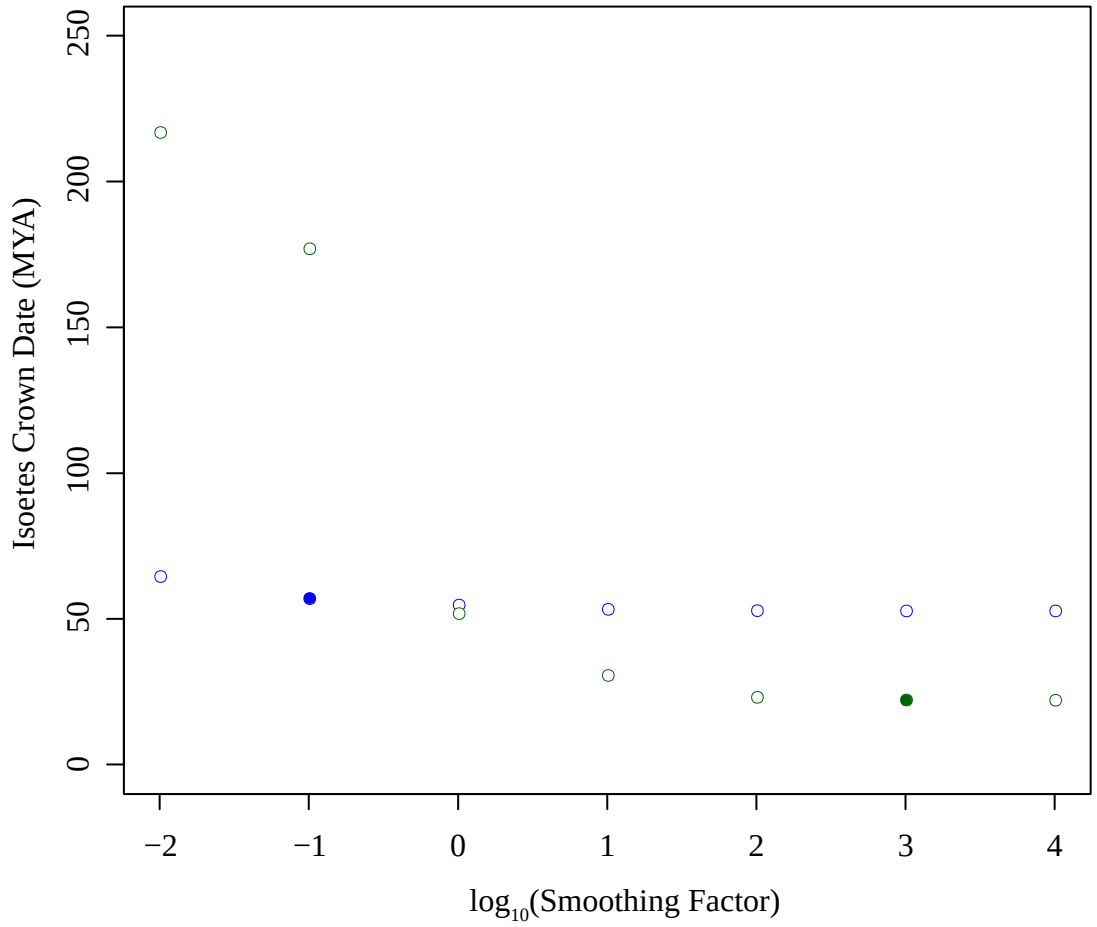


Figure 3.2: **Effect of different smoothing factors on *Isoëtes* crown date estimation in r8s**
Estimated crown dates for *Isoëtes* produced by r8s for concatenated chloroplast (green) and nuclear (blue) datasets for a range of smoothing factors. The best fitting smoothing factor, as identified by cross validation, is highlighted for each dataset by a filled circle.

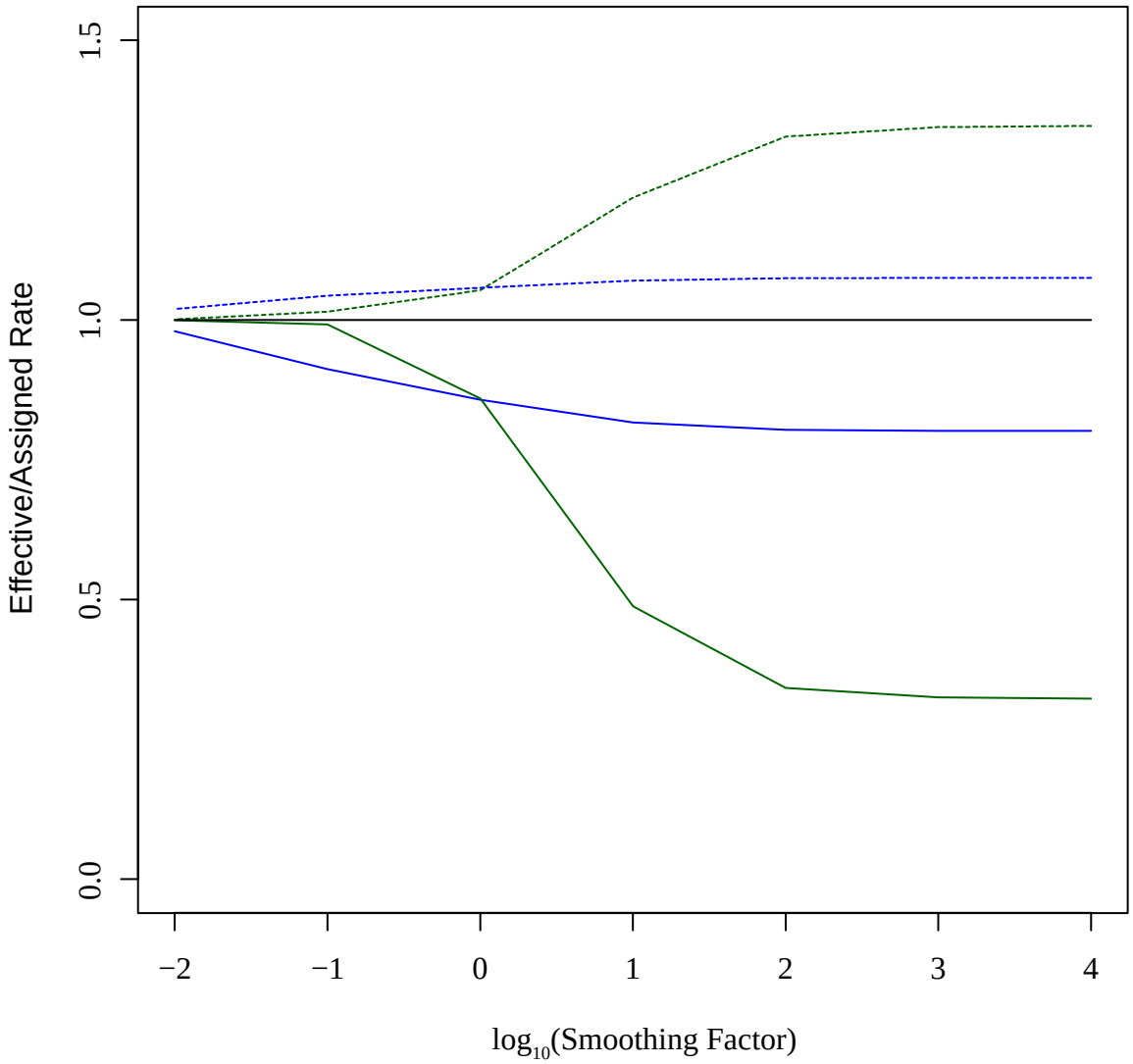


Figure 3.3: Rate assignment on crown and stem branches of *Isoëtes* in r8s

The ratio of effective vs. assigned rates for different smoothing factors in r8s for the stem branch of *Isoëtes* (dashed lines) and the average ratio for the *Isoëtes* crown branches (solid lines) for the concatenated chloroplast (green) and nuclear (blue) datasets. Solid grey line represents Effective/Assigned rate of 1.

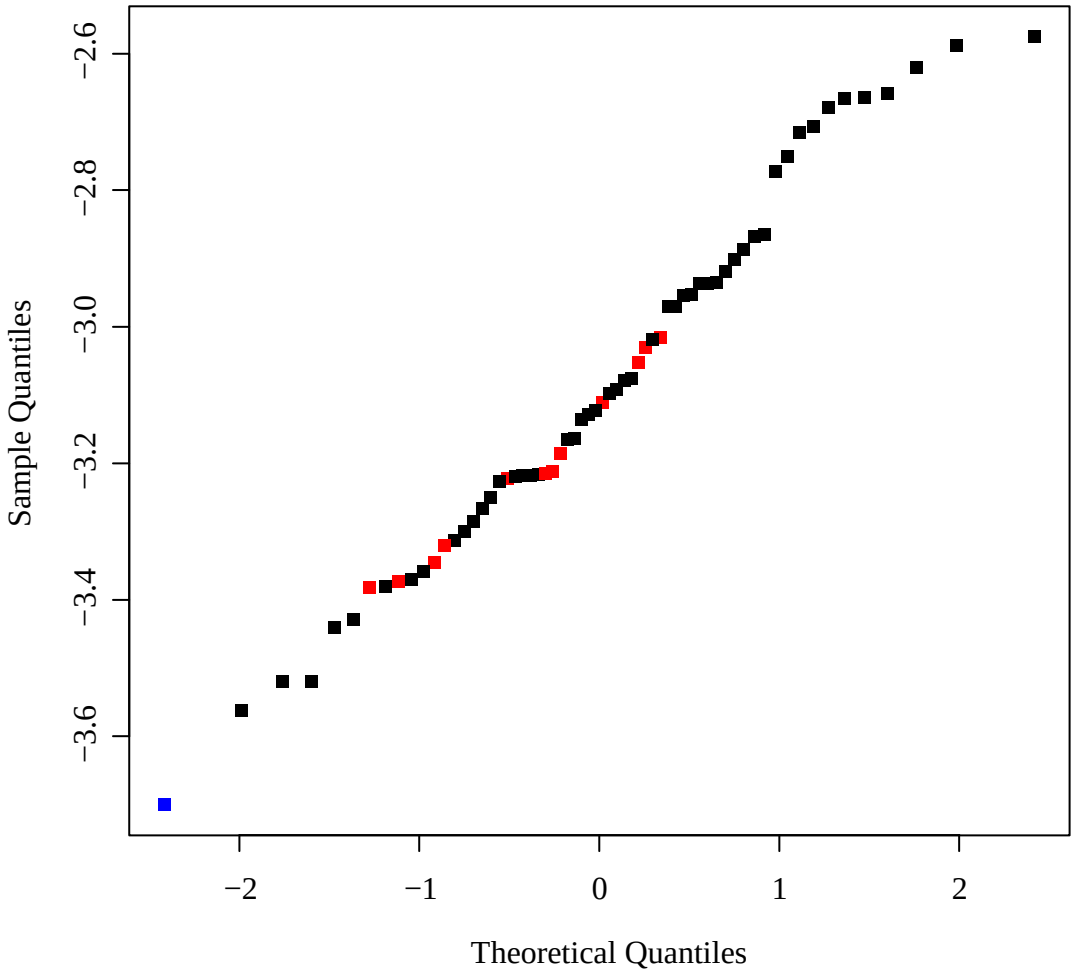


Figure 3.4: Quantile-Quantile plot of BEAST rates for concatenated chloroplast markers
Quantile-quantile plot of log10 transformed branch rates for the concatenated chloroplast dataset in BEAST. The *Isoëtes* stem branch (blue) and crown branches (red) are highlighted.

Phylogenetic reconstruction and dating based on nuclear markers

The *I. lacustris* transcriptome was assembled into 88,340 transcripts with an average length of 753bp. The concatenated nuclear phylogram also recapitulated major land plant relationships, and the topology of the *Isoëtes* clade was consistent with that of the chloroplast phlyogeny, with *I. coromandelina* again being sister to all other species (Figure 3.5). Despite overall longer branch lengths in the concatenated nuclear phylogeny, variation among groups was reduced. Particularly, the total branch lengths from the common ancestor of *Isoëtes* and *Selaginella* were much more similar than in the chloroplast phylogeny, with *Selaginella* having accumulated approximately 1.25

times more mutations than *Isoëtes* since their common ancestor. However, the ratio of the average crown branch length to stem length in the *Isoëtes* lineage was very similar between the nuclear and chloroplast markers; approximately 5.8 for the chloroplast dataset and 5.6 for the nuclear dataset (Figure 3.1, Figure 3.5).

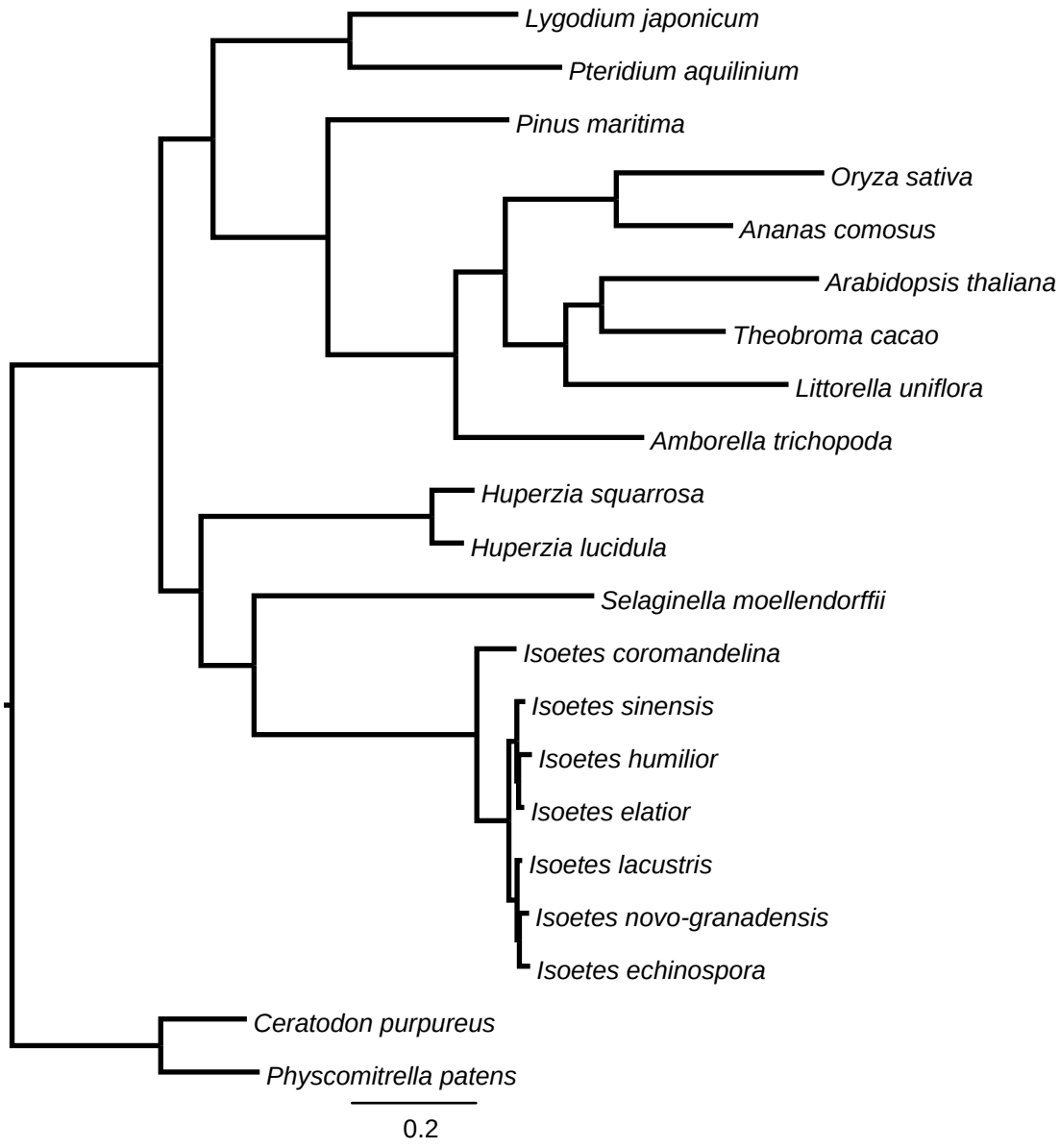


Figure 3.5 Maximum likelihood phylogram of concatenated nuclear markers
Branch lengths are proportional to the number of substitutions per site, with scale bar representing 0.2 substitutions per site. All bootstraps support values are 100.

Dating of the concatenated matrix of nuclear markers in r8s gave an estimated crown node age of *Isoëtes* as 58.9 MYA (Table 3.1), with an estimated stem node age of 358 MYA, at an optimum smoothing value of 0.1. Unlike with the chloroplast markers, the date of the *Isoëtes* crown node was similar across all smoothing values tested (Figure 3.2). Increased smoothing values led to increases in the disparity between effective and assigned rates (Figure 3.3), although this was reduced in comparison with the concatenated chloroplast alignment (0.82 vs 0.33 for the stem branch and 1.25 vs. 1.33 for the crown branch for a smoothing value of 10^6). The conservation of the effective rates in the stem and crown branch of *Isoëtes* across a range of smoothing parameters indicates that the average rates predicted across the entire nuclear tree are a relatively good fit to the stem and branch rates of *Isoëtes* (Figure 3.3), implying similar rates across the *Isoëtes* stem and crown branches. This is consistent with the highly similar branch length ratios between the stem and crown branches of *Isoëtes* in the chloroplast and nuclear trees (Figure 3.1 and Figure 3.5). If a large rate change had taken place, it would have had to have affected the chloroplast and nuclear rates in a very similar way to produce such similar branch lengths; as can be seen in taxa such as *Selaginella*, which has relatively high chloroplast rates that are not coupled with those of the nuclear genome (Figure 3.1 and Figure 3.5).

Dating individual nuclear genes in r8s resulted in a wide range of optimum smoothing values (Supplementary Figure 3.1). Low smoothing values frequently resulted in gradient check failures, indicating a single optimum solution was not reached (Supplementary Figure 3.1). For genes reaching a single optimum, the median estimated crown date for *Isoëtes* was 46.4MYA with 95% of estimates between 16.1 and 85.8MYA and 50% of results between 31.9 and 58.3MYA (Figure 3.6), with estimated dates forming a unimodal distribution. Low values of the smoothing parameter resulted in the majority of genes failing gradient checks and a wide range of age estimates for those that passed – a smoothing parameter of 0.01 resulted in a median estimate of 112.5MYA with 95% of data between 61.1 and 225.8MYA and 50% between 86.6 and 150.2MYA. Smoothing values above 10 gave similar results to when optimum values from cross validation were assigned to each gene individually (Figure 3.6). As with the chloroplast datasets, increasing smoothing values resulted in a decreased effective/assigned stem rate and increased effective/assigned crown rate (Figure 3.7), although the disparities for the optimum smoothing values were reduced compared to the chloroplast data (0.82 vs 0.33 for the stem branch and 1.25 vs. 1.33 for the crown

branch), indicating the globally optimum smoothing values for the individual nuclear markers fit the stem and crown branches of the *Isoëtes* better than in the chloroplast dataset. The median optimum smoothing value of 316 was significantly higher than the value of 0.1 obtained for the concatenated alignment. Differences in rates between individual genes across species, and the presence of relatively incomplete alignments, may have reduced the ability to detect autocorrelation of rates in individual gene lineages, leading to higher optimal smoothing values, although these still predicted a similar date to the concatenated markers.

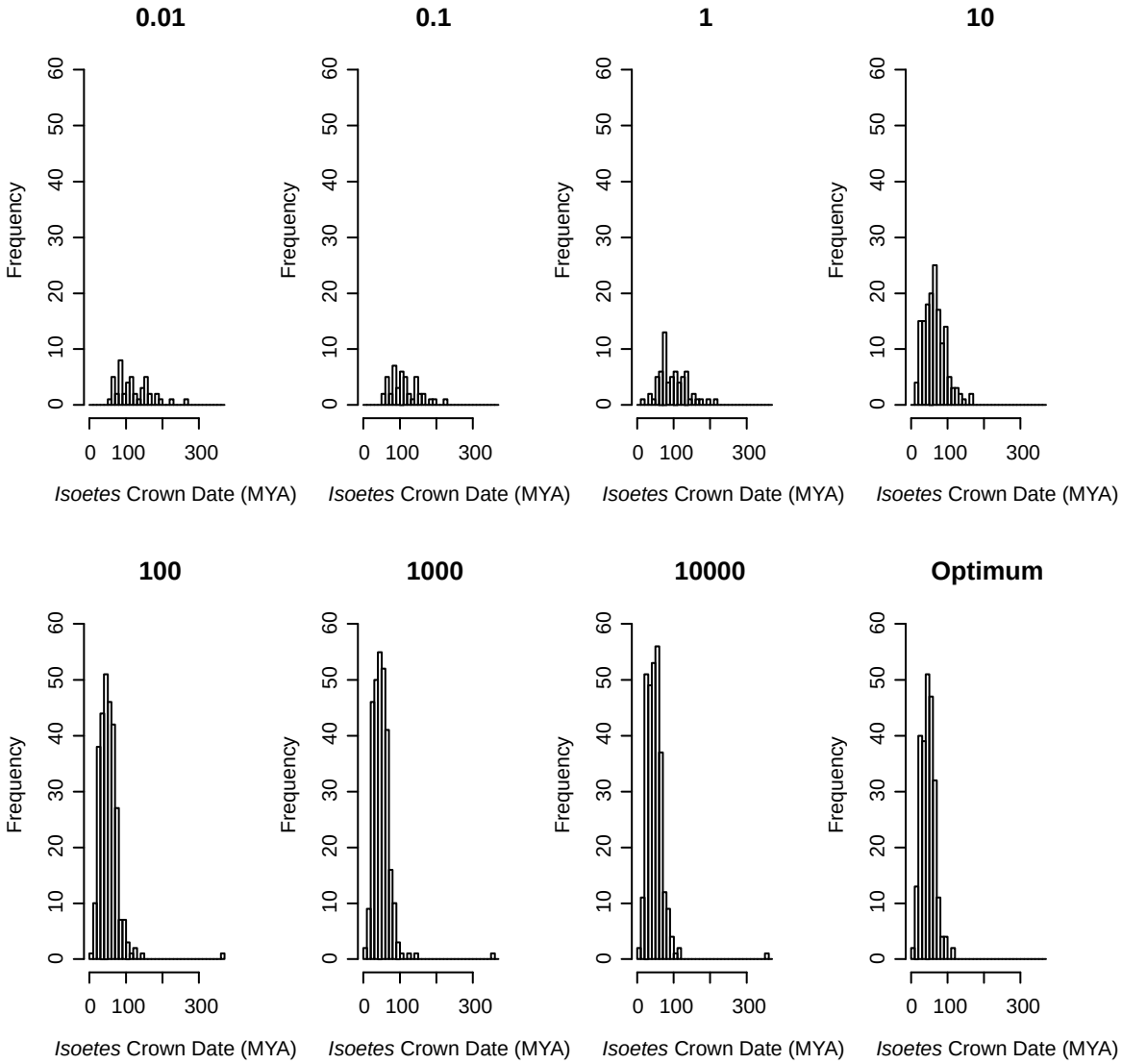


Figure 3.6: ***Isoëtes* crown dates for individual nuclear genes for different smoothing values in r8s**

Histograms showing estimated *Isoëtes* crown group dates for individual nuclear genes in r8s that pass gradient checks for a range of assigned smoothing values, and the histogram of estimates where each gene is assigned its optimum smoothing value based on cross validation (final panel).

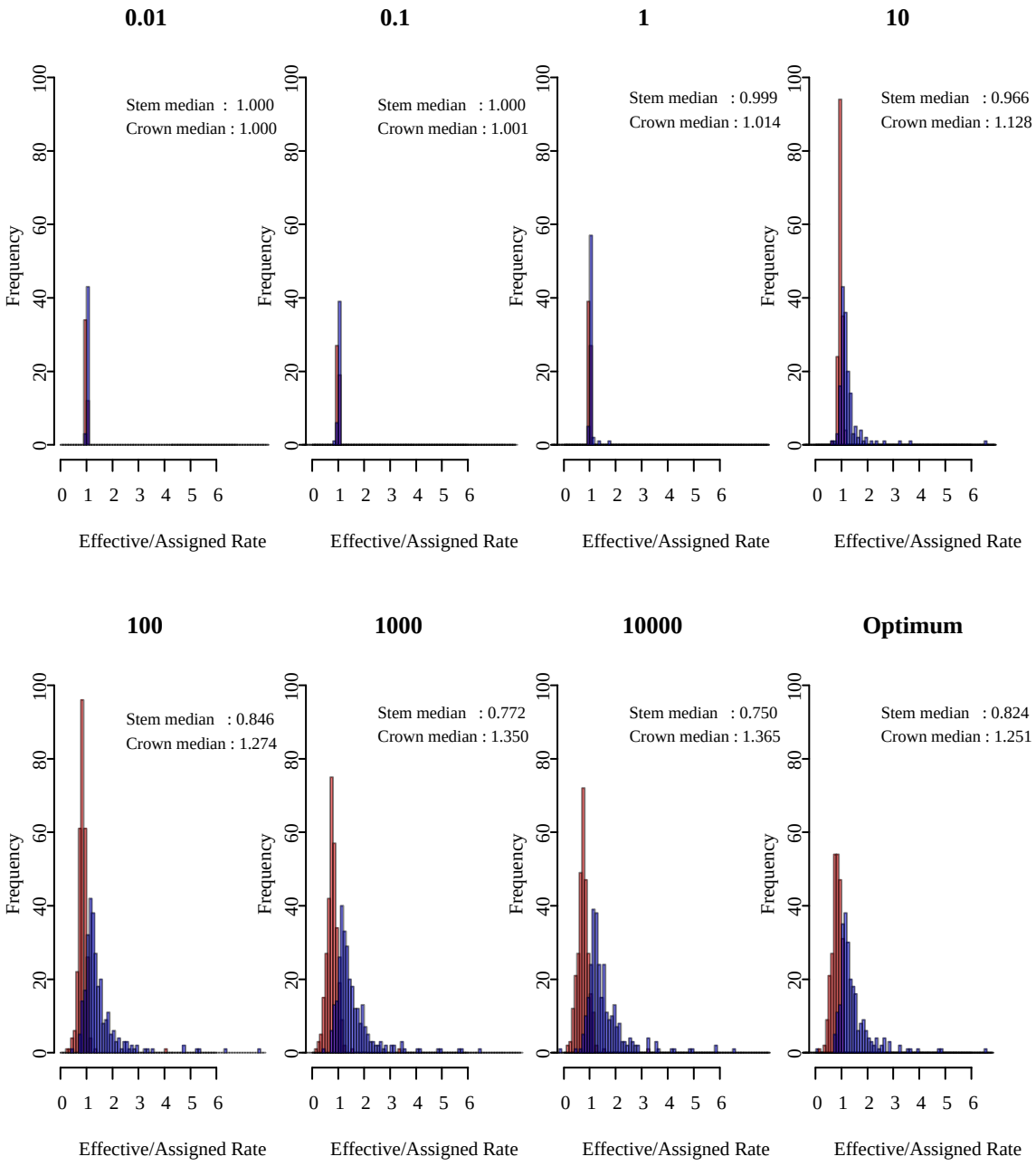


Figure 3.7: **Effective/Assigned Rate ratios for individual nuclear genes in r8s**

Histograms of the ratio of effective vs. assigned branch rates for the stem (red) and average value for crown (blue) branches of *Isoëtes* for individual nuclear genes in r8s that passed gradient checks for a range of assigned smoothing values, and the histogram of estimates where each gene is assigned its optimum smoothing value based on cross validation (final panel). Median values are displayed in the top righthand corner of each panel.

Dating individual genes using BEAST gave a median result of 54.7 MYA for the crown of *Isoëtes*, with 95% of estimates between 45.6 and 120.0 MYA, and 50% between 45.6 MYA and 68.5 MYA. The ages obtained for individual genes were highly correlated between r8s and BEAST (linear model, slope = 0.94, p-value < 0.001; $R^2 = 0.64$; Figure 3.8). Linear modelling suggested a significant but small effect of the percent completeness of the alignments on the estimate for the crown age of *Isoëtes*, with a larger effect from the average completeness of *Isoëtes* sequences (Supplementary Table 3.5). However, the adjusted R^2 for this effect was 0.059 in r8s and 0.042 in BEAST, indicating that the completeness of the alignment has relatively little impact on the estimated dates.

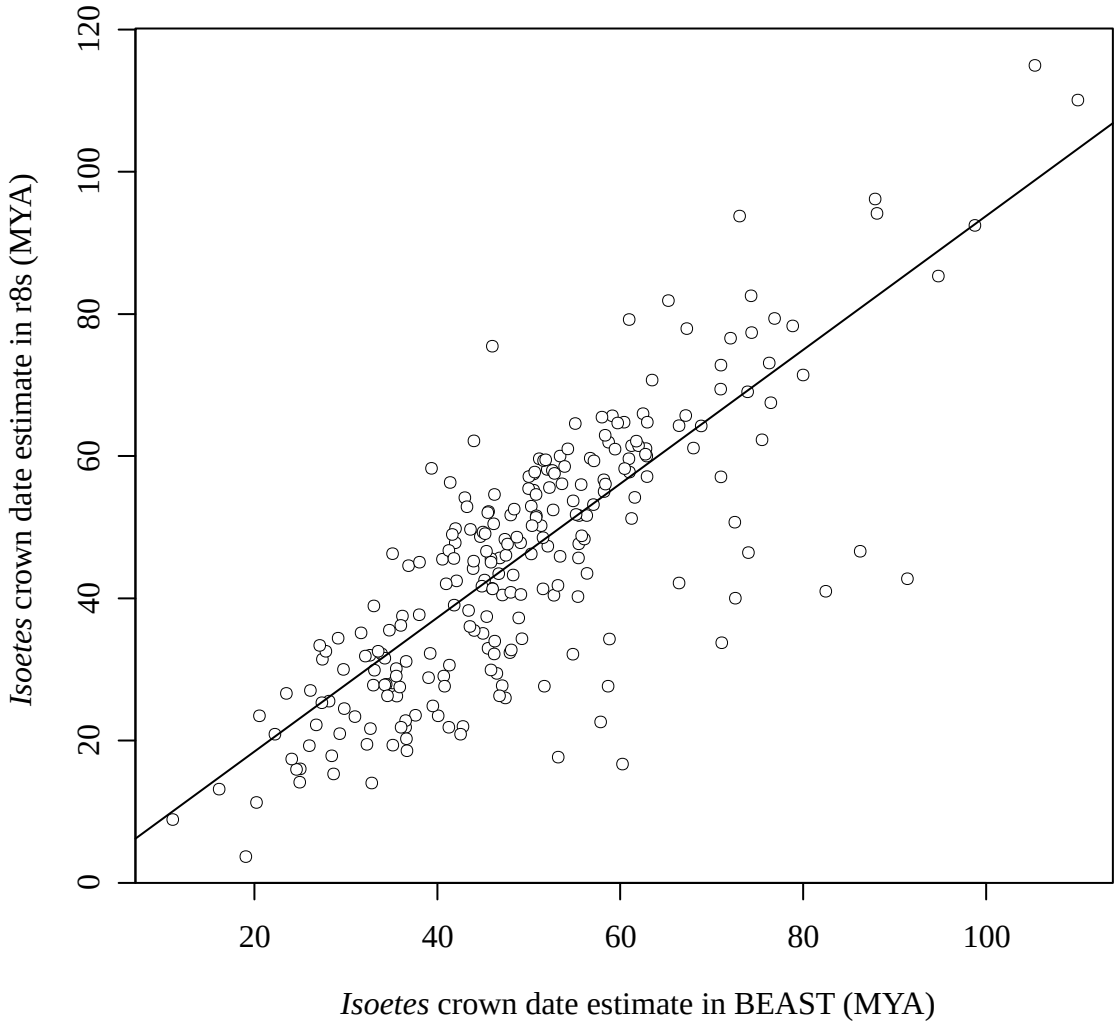


Figure 3.8: r8s versus BEAST *Isoëtes* crown estimates for individual nuclear genes
Scatterplot of estimates of the *Isoëtes* crown date in r8s and BEAST for each individual nuclear gene. Line represents output of linear model.

Molecular dating of other aquatic CAM species

Molecular dating of the *rbcL* alignment using BEAST gave an estimate for the *Isoëtes* crown node of 47.8 MYA, with a 95% HPD of 16.8-90.6 MYA, with estimates of the crown ages of *Vallisneria, Sagittaria, Littorella* and *Crassula* also occuring in the last 35.8 MYA (Table 3.2). r8s gave an optimum smoothing factor of 1000 and produced an older crown date for *Isoëtes* at 65.2 MYA (33.0-82.9 95% CI), with crown ages predicted to be 5-15 MY younger than the BEAST estimates (Table 3.2).

Table 3.2: Molecular dating other other aquatic CAM genera

Genus	BEAST crown node age (95% CI)	r8s crown node age (95% CI)
Isoëtes	47.8 (16.8-90.6)	65.2 (33.0-82.9)
Sagittaria	30.7 (15.2-48.2)	15.5 (12.4-20.1)
Vallisneria	21.6 (8.5-37.5)	11.9 (7.9-16.1)
Crassula	35.5 (16.2-57.0)	29.5 (23.7-35.9)
Littorella	30.7 (16.7-46.3)	26.2 (20.8-29.8)

3.5 Discussion

Differences between methods and partitions, and the age of Isoëtes

Very different dates for the crown of *Isoëtes* were found on the chloroplast and nuclear datasets, at 23-24 MYA and 45-60 MYA, respectively (Table 3.1). These differences are unlikely to be caused by the methods, since BEAST and r8s produced almost identical dates (Table 3.1, Figure 3.8), despite the very distinct ways in which these two programs deal with rate variation among branches. Similarly, the incongruence is probably not due to missing data since the completeness of datasets for *Isoëtes* did not correlate with the age estimates. Instead, the incompatibilities between nuclear and chloroplast results probably arise from differences in how branch rates are assigned between the different datasets. High levels of branch length variation were observed between *Selaginella* and *Isoëtes* chloroplast markers (Figure 3.1), a pattern consistent with previous seed plant studies on this organellar genome (Bousquet et al., 1992; Ruhfel et al., 2014). The particularly large increase in *Selaginella* branch lengths inferred from the chloroplast dataset relative to shorter *Isoëtes* branches has been found previously (Karol et al., 2010; Ruhfel et al., 2014; Larsén and Rydin, 2015). The high levels of variability in

these datasets make low levels of smoothing in r8s relatively poor fits to the data, effectively forcing a single rate on the tree that is determined by the average branch length. In turn, this results in a rate fitted to *Isoëtes* that is a poor match to its relatively short branch lengths (Figure 3.3). As a result, the fitted models predict more changes along the stem branch and fewer changes along the crown branches than occur in the data. Similarly in BEAST, the lognormal prior distribution results in a relatively low rate assignment on the stem branch compared to the crown branches, which leads to a better fit to the lognormal distribution across branches than if all crown branches were low (Figure 3.4). The high rate variability and low branch lengths in the chloroplast datasets, and the relatively short *Isoëtes* branch lengths result in disparity between branch lengths and estimated rates and divergence times, making the chloroplast dating unreliable (Figure 3.3). By contrast, the individual and concatenated nuclear datasets have a reduced disparity between estimated and effective rates, and are consistent across genes (Figure 3.3, Figure 3.7), in the concatenated versus individual datasets, and between BEAST and r8s (Figure 3.8). This consistency indicates the nuclear dataset is a more appropriate dataset for estimating divergence times using the models of rate distribution in r8s and BEAST.

We therefore conclude, based on our genome-wide analyses, that the extant *Isoëtes* diversity originated during the Paleogene, likely between 60 and 45 MYA (Table 3.1). This conclusion is in sharp contradiction with previous estimates of the crown group *Isoëtes* at 147 and 251 MYA (Larsén and Rydin, 2015, Kim and Choi, 2016, Pereira et al., 2017). The study of Larsén and Rydin (2015) was based on a relatively small number of markers with only *rbcL* aligning with sequences outside of the genus, and the noncoding markers available solely for *Isoëtes* species caused an increase in crown branch dates without altering the stem branch (Supplementary Table 3.6). On the other hand, Kim and Choi (2016) and Pereira et al. (2017) constrained the crown node of *Isoëtes* with fossils that do not contain features that reliably confirm that the most basal split in extant *Isoëtes* had taken place (Ash and Pigg, 1991; Retallack, 1997), and should therefore be considered stem node calibrations. The present study is based on a dramatically increased number of markers from both chloroplast and nuclear genes that provide informative sites within the *Isoëtes* but can also be aligned to species outside the genus. Our nucleotide data are consequently homogeneously distributed among taxonomic groups, and the fossil evidence is used in a very conservative way, relying solely on external fossils to estimate the age of the crown *Isoëtes*, our group of interest.

Extant Isoëtes diversified more recently than previously thought

The origins of extant *Isoëtes* crown group diversity in the Palaeogene considerably postdates the earliest appearance of modern *Isoëtes*-like forms in the Jurassic and Cretaceous (Pigg, 2001). Whilst it is possible that a single lineage of *Isoëtes* arose in the Jurassic and only diversified in the Paleogene, this seems unlikely. The decline of aquatic lycopsid megaspores in the Late Cretaceous (Kovach and Batten, 1993) is consistent with extant *Isoëtes* being the remnants of a previously more diverse lineage. The decline of submerged aquatic lycopods was hypothesised to coincide with the radiation of angiosperms into submerged aquatic habitats (Kovach and Batten, 1993; Greb and DiMichele, 2006), and is consistent with the decline of other ancient lineages such as gymnosperms and ferns during this period (Wolfe, 1997; Lupia et al., 1999). This would explain a low diversity of *Isoëtes* in the late Cretaceous, but does not explain why the lineage would then diversify during the Palaeogene and the subsequent Neogene despite competition from submerged aquatic angiosperms. Recent estimates indicate CO_2 levels remained relatively high in the Paleogene until approximately 30 MYA (Foster et al., 2017). However, most of the extant diversity of Isoëtes emerged relatively long after the crown of the group (Larsén and Rydin, 2015), and our revised dates therefore indicate that a significant proportion of *Isoëtes* speciation events took place in a low-CO_2 atmosphere during the past 30 million years (Foster et al., 2017). Declining levels of atmospheric CO_2 may have allowed expansion from palustrine environments (wetlands) uncoupled from atmospheric CO_2 levels to lacustrine (lakes) environments that would be more susceptible to global changes in atmospheric CO_2 (Keeley and Rundel, 2003).

The date of origin for CAM in *Isoëtes* is unclear. The prevalence of CAM in extant *Isoëtes* suggests the most recent common ancestor of extant *Isoëtes* was CAM, but CAM may have evolved at any point along the stem branch of the *Isoëtes*, which spans more than 300 million years (Edwards and Ogburn, 2012). The relative morphological similarity of extant and fossil *Isoëtes* does not suggest recent evolutionary innovations, but the fact that some extant *Isoëtes* lack CAM whilst possessing similar morphology (Keeley, 1998) shows that the transition to CAM would be hard to infer from the fossil record. A relatively recent origin of CAM in *Isoëtes* could explain the reversal in decline of *Isoëtes* species diversity in the Paleogene,

followed by diversification to present species number as falling CO_2 levels opened up new niches to allow expansion.

Common selective pressures for aquatic and terrestrial CAM?

The diversification of *Isoëtes* during the Paleogene and subsequent Neogene suggests that the environments of these periods was favourable for aquatic CAM photosynthesis. This is consistent with the contemporaneous origins of other lineages of aquatic CAM plants during the last 60-20 million years (Table 3.2). As with *Isoëtes*, neither the crown nor the stem dates necessarily coincide with the origins of CAM in these groups. Only six of 30 *Sagittaria* species and two of 14 *Vallisneria* species have been tested for CAM activity, with significant support for CAM activity only found in two species in *Sagittaria* and one *Vallisneria* species, with weak support in the other *Vallisneria* species (Keeley, 1998). CAM is therefore likely to have originated after the origin of the genus in both of these cases, which would place the transitions during the late Paleogene/Neogene, when atmospheric CO_2 levels were low. Most terrestrial members of *Crassula* are CAM, and an aquatic lifestyle likely evolved from terrestrial CAM within the genus. The closest relatives of *Littorella* are the terrestrial, temperate *Plantago*, and CAM certainly evolved after the split of *Littorella* from *Plantago*. Most origins of aquatic CAM therefore occurred shortly before or after the drastic atmospheric CO_2 decline during the Oligocene, and the diversification of the groups most likely took place in a low-CO_2 world. These results suggest that low atmospheric CO_2 levels might not be necessary for the evolution of submerged aquatic CAM, but are associated with increased diversification of groups with this phenotype.

The hypothesised influence of atmospheric CO_2 levels on CAM evolution in submerged aquatic plants sheds light on the origins of CAM in terrestrial plants. Recent aridification events are likely to have had an important impact on the expansion of CAM species in particular environments (Arakaki et al., 2011; Givnish et al., 2014), and similarly it is likely that the expansion and contraction of CAM-favourable microhabitats in aquatic systems also plays an important role in diversification. A diverse range of local factors that limit daytime CO_2 uptake have resulted in the tremendous diversity of environments supporting CAM plants, from the arid deserts, to the tropical forests supporting epiphytic CAM, and the small ponds and lakes across the globe. Local factors appear to ultimately drive the origins of CAM in these individual

environments, but the global effect of declining atmospheric CO_2 levels has presented an opportunity for convergent diversification of CAM plants in a wide variety of environments. Importantly, lineages belonging to groups as distant as the lycopods and angiosperms responded convergently to the global challenge created by atmospheric changes.

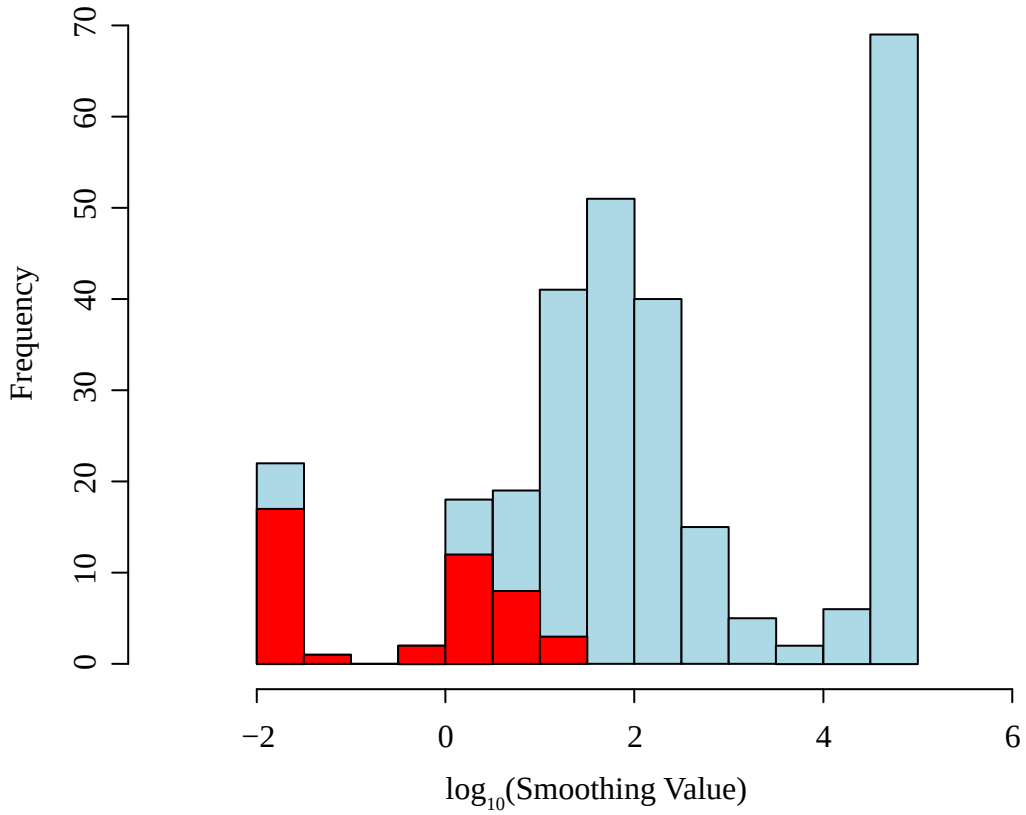
3.6 Conclusions

Using genomic datasets covering the land plants, this study has dated the origins of extant species diversity in *Isoëtes*, showing that this group of aquatic CAM plants diversified in the last 45-60 million years. This finding strongly contrasts with the rich record of *Isoëtes*-like fossils dated back to the Triassic, showing that extant representatives of ancient lineages represent relatively recent radiations, as shown for other groups (Nagalingum et al., 2011). The revised timing of *Isoëtes* diversification places these lycopods on the same time scale as CAM lineages within the flowering plants, including other aquatic groups, but also the emblematic terrestrial CAM lineages, such as cacti (Arakaki et al., 2011). The convergent and parallel expansions of CAM groups are consistent with falling levels of atmospheric CO_2 acting as an enabler of CAM plant diversification in highly diverse environments. While global atmospheric changes led to extreme species diversity in cacti and spurges, we now have demonstrated that “living fossils” such as *Isoëtes* were also able to take advantage of new ecological opportunities caused by environmental change.

Acknowledgements

We thank Luke T. Dunning, Jill K. Oloffsson, Jose J. Moreno-Villano and Matheus E. Bianconi for advice in DNA sequencing and computational analyses, Charles Wellman for advice on phylogenetic assignment of cryptospores, and Hanna Sewell for assistance in live plant collection. Catarina Rydin kindly provided alignments from Larsén and Rydin (2015).

3.7 Supplementary Information



Supplementary Figure 3.1: Histogram of optimum smoothing values identified by cross validation for individual nuclear genes. Proportion of genes for each smoothing value that fail gradient checks are highlighted in red.

Supplementary Table 3.1: *Isoëtes nrITS* species assignment

Recorded species name	Collection location	Top blast hit	Assigned species
Isoëtes nuttallii	Alaska, USA	*Isoëtes echinospora* (KT288378.1)	*Isoëtes echinospora*[1]
Isoëtes andicola	Casapalca, Peru	*Isoëtes melanopoda* (DQ284996.1)	*Isoëtes novo-granadensis*[2]
Isoëtes humilior	Unknown	*Isoëtes humilior* (KT288381.1)	*Isoëtes humilior*
Isoëtes elatior	Tasmania, Australia	*Isoëtes drummondii* (DQ2844993.1)	*Isoëtes elatior*[3]
Isoëtes coromandelina	Arajasthan, India	*Isoëtes coromandelina* (DQ284992.1)	*Isoëtes coromandelina*

1. *I. echinospora* is known to co-occur with *I. nuttallii* in Alaska (Britton et al., 1999).

2. *I. novo-granadensis* is native to Peru (Leon and Young, 1996) and gave a similar match score to this sample compared to *I. melanopoda,* which is native to North Amercia (Brunton and Britton, 2006).

3. Assigned to *I. elatior* as no other sequence data for this species is available.

Supplementary Table 3.2: Chloroplast Data Sources

Species	Source
Amborella trichopoda	NCBI: NC_005086
Ananas comosus	NCBI: NC_026220
Aneura mirabilis	NCBI: NC_0105359
Ceratophyllum demersum	NCBI: EF614270
Diploterygium glaucum	NCBI: NC_024158
Ginkgo biloba	NCBI: NC_016986
Huperzia lucidula	NCBI: NC_006861
Isoëtes flaccida	NCBI: NC_014675
Lygodium japonicum	NCBI: KF225593
Marchantia polymorpha	NCBI: NC001319
Nuphar advena	NCBI: DQ354691
Nymphaea alba	NCBI: AJ627251
Osmundastrum cinnamomeum	NCBI: BC_024157
Physcomitrella patens	NCBI: AP005672
Pinus thunbergii	NCBI: D17510
Plantago maritima	NCBI: NC_028519
Plantago media	NCBI: NC_028520
Pteridium aquilinum	NCBI: NC_014348
Ranunculus macranthus	NCBI: DQ359689
Selaginella mollendorffii	NCBI: HM173080
Selaginella uncinata	NCBI: AB197035
Sesamum indicum	NCBI: JN637766
Syntrichia ruralis	NCBI: FJ46412
Veronica nakaiana	NCBI: NC_031153

Supplementary Table 3.3: Transcriptome Data Sources

Species	Data type	Source
Ananas comosus	Assembly (v3.0)	Phytozome
Ceratodon purpurea	Assembly	Szövényi et al., 2015
Huperzia lucidula	Reads	http://www.medplantrnaseq.org/
Huperzia squarrosa	Reads	http://www.medplantrnaseq.org/
Isoëtes sinensis	Reads	SRA: SRR1648119
Lygodium japonicum	Assembly (v1.0)	http://bioinf.mind.meiji.ac.jp/kanikusa
Pinus pinaster	Assembly (v3.0)	http://www.scbi.uma.es/sustainpinedb/
Pteridium aquilinum	Assembly	Der et al., 2011

Supplementary Table 3.4: *rbcL* Data Sources

Species	Source
Acorus americanus	NCBI: DQ069499
Alisma lanceolatum	NCBI: HM849753
Altingia gracilipes	NCBI: DQ352379
Amborella trichopoda	NCBI: NC_005086
Angelonia pubescens	NCBI: AF123672
Anthoceros formosae	NCBI: NC_004543
Antirrhinum majus	NCBI: GQ997015
Arabidopsis thaliana	NCBI: NC_000932
Aragoa cleefii	NCBI: AJ459247
Astonia australiensis	NCBI: HQ456499
Bacopa caroliniana	NCBI: AF123670
Baldellia ranunculoides	NCBI: DQ859163
Burnatia enneandra	NCBI: JN547809
Callitriche nana	NCBI: AY289597
Callitriche truncata	NCBI: AF248025
Ceratophyllum demersum	NCBI: NC_009962
Chaenorhinum minus	NCBI: KM360709
Chamaedorea seifrizii	NCBI: HQ182421
Crassula helmsii	NCBI: KM360736
Crassula marnierana	NCBI: L01899
Crassula perforata	NCBI: AF274594
Crassula vaillantii	NCBI: HM849922
Digitalis lanata	NCBI: KY085895
Digitalis purpurea	NCBI: X83720
Dillenia indica	NCBI: GQ997181
Diphasium jussiaei	NCBI: AJ133256
Echinodorus berteroi	NCBI: KU499839
Echinodorus horizontalis	NCBI: KU499840
Elodea nuttallii	NCBI: KM360763
Franklinia alatamaha	NCBI: HM100377
Gnetum parvifolium	NCBI: NC_011942
Gonocarpus depressus	NCBI: JQ933346
Gossypium raimondii	NCBI: NC_016668
Halophila engelmannii	NCBI: HEU80699
Halophila ovalis	NCBI: KX527483
Helanthium parvulum	NCBI: HQ456504
Hemiphragma heterophyllum	NCBI: AF123667
Hosta ventricosa	NCBI: HQ182431

Huperzia selago	NCBI: KX761187
Huperzia serrata	NCBI: KX426071
Hydrilla verticillata	NCBI: AB004891
Hydrocleys martii	NCBI: HQ901564
Isoëtes bradei	NCBI: AF404493
Isoëtes coromandelina	NCBI: DQ294242
Isoëtes kirkii	NCBI: AF404499
Isoëtes lacustris	NCBI: AJ010855
Isoëtes malinverniana	NCBI: DQ294245
Isoëtes orcuttii	NCBI: DQ294247
Lagarosiphon major	NCBI: LMU80703
Lemna minor	NCBI: NC_010109
Limnobium spongia	NCBI: LSU80704
Limnophyton angolense	NCBI: JF781049
Limnophyton sp.	NCBI: JF781062
Linaria simplex	NCBI: KX282855
Littorella uniflora	NCBI: HM850128
Lycopodiella inundata	NCBI: Y07938
Lycopodium clavatum	NCBI: Y07936
Magnolia kwangsiensis	NCBI: NC_015892
Marchantia polymorpha	NCBI: NC_001319
Mollugo verticillata	NCBI: HQ621337
Najas browniana	NCBI: HM240486
Najas flexilis	NCBI: JX978472
Najas marina	NCBI: LC128123
Nechamandra alternifolia	NCBI: AB506768
Nechamandra sp.	NCBI: KJ994537
Nicotiana tabacum	NCBI: NC_001879
Nuphar advena	NCBI: DQ069501
Nymphaea alba	NCBI: NC_006050
Oryza sativa	NCBI: NC_001320
Paeonia suffruticosa	NCBI: KT944727
Penstemon serrulatus	NCBI: KX678660
Peridiscus lucidus	NCBI: AY380356
Phlegmariurus wilsonii	NCBI: Y07933
Phoenix dactylifera	NCBI: NC_013991
Phylloglossum drummondii	NCBI: Y07939
Physcomitrella patens	NCBI: NC_005087
Pinus nelsonii	NCBI: NC_011159
Plantago cylindrica	NCBI: KY293571
Plantago lanceolata	NCBI: L36454
Plantago maritima	NCBI: NC_028519
Psilotum nudum	NCBI: NC_003386
Ranunculus macranthus	NCBI: DQ069502
Rhodiola sexifolia	NCBI: KP115076
Rhododendron simsii	NCBI: GQ997829
Ribes aureum	NCBI: L11204
Sagittaria filiformis	NCBI: KX397946
Sagittaria guayanensis	NCBI: JF781054
Sagittaria isoetiformis	NCBI: JF781061
Sagittaria latifolia	NCBI: L08767
Sagittaria lichuanensis	NCBI: KT899952
Sagittaria natans	NCBI: JF781055
Sagittaria subulata	NCBI: HM850330

Saxifraga umbellulata	NCBI: MF197627
Scoparia dulcis	NCBI: KJ773875
Sedum rubrotinctum	NCBI: L01956
Selaginella ciliaris	NCBI: EU126658
Selaginella moellendorffii	NCBI: HM173080
Selaginella pedata	NCBI: KY023113
Selaginella sanguinolenta	NCBI: KY023142
Selaginella selaginoides	NCBI: KY023148
Selaginella sinensis	NCBI: AJ295868
Selaginella uncinata	NCBI: KR028155
Sesamum indicum	NCBI: NC_016433
Spirodela polyrhiza	NCBI: NC_015891
Stratiotes aloides	NCBI: KM360999
Thalassia hemprichii	NCBI: KX527484
Theobroma cacao	NCBI: NC_014676
Umbilicus horizontalis	NCBI: HM850434
Umbilicus schmidtii	NCBI: KP279363
Vallisneria americana	NCBI: EF143015
Vallisneria caulescens	NCBI: EF143009
Vallisneria rubra	NCBI: AY870370
Vallisneria spinulosa	NCBI: EF143017
Vallisneria spiralis	NCBI: VSU80712
Veronica americana	NCBI: KT178133
Veronica cupressoides	NCBI: AY034014
Welwitschia mirabilis	NCBI: NC_010654
Yucca schidigera	NCBI: DQ069504

Supplementary Table 3.5: Effects of gene properties on estimated dates

Relationship	BEAST gradient	BEAST p-value[1]	BEAST R^2 value	r8s gradient	r8s p-value[1]	r8s R^2 value
Average *Isoëtes* alignment completeness	0.291	**0.0007**	0.042	0.4	**7.4E-05**	0.059
Average genome skimming sample completeness	0.329	**0.0018**	0.036	0.35	**0.0045**	0.029
I. coromandelina alignment completeness	0.355	**0.00047**	0.047	0.414	**0.0005**	0.049
Average alignment completeness	0.16	**0.038**	0.014	0.25	**0.0051**	0.027
Percentage of polymorphic sites	-0.087	0.22	0.0022	-0.056	0.495	-0.0022
Percentage of phylogenetically informative sites	-0.045	0.616	-0.0031	0.06	0.53	-0.025
Percentage completeness of Selaginella	0.11	**0.029**	0.016	0.16	**0.0071**	0.0254
Average completeness of *I. sinensis* and *I. lacustris*	0.09	0.092	0.0076	0.152	**0.016**	0.02

1. Significant values ($p < 0.05$) highlighted in bold.

Supplementary Table 3.6: Reanalysis of dataset of Larsén and Rydin, 2015

Analysis[1]	*Isoëtes crown date* (95% CI)	*Isoëtes stem date* (95% CI)
All *rbcL* + *Isoëtes nrITS* + Isoëtes *atpB-rbcL* intergenic spacer[i]	145.8 (88.2-208.7)	382.1 (353.9-410.4)
All *rbcL* + Isoëtes *nrITS*[ii]	169.2 (99.6-244.7)	381.1 (353.5-410.4)
All *rbcL* + Isoëtes *atpB-rbcL* intergenic spacer[iii]	64.3 (38.2-94.8)	383.6 (354.7-413.2)
All *rbcL*[iv]	41.1 (23.0-63.2)	380.6 (351.1-412.4)

1. The alignment from Larsén and Rydin, 2015 was re-analysed using the same constraints and BEAST settings as the previous paper. The dataset contains *rbcL* sequences for *Isoëtes* species and other Embryophyte groups, and additional highly variable sequences for *nrITS* and the atpB-rbcL intergenic spacer for *Isoëtes* only. *Isoëtes* species lacking an rbcL sequence were excluded from the analysis. The entire dataset (i) gave similar estimates of the *Isoëtes* crown age to Larsén and Rydin, 2015, but removal of the *atpB-rbcL* intergenic spacer (ii) reduced ages for the Isoëtes crown, and removal of either *nrITS* (iii) or both *Isoëtes*-specific markers (iv) resulted in ages consistent with the present study.

Chapter IV: Conserved expression patterns across vascular plants drove the convergent evolution of CAM in aquatic lycopods and angiosperms

Daniel Wood[1], Pascal-Antoine Christin[1]

1. Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN

Personal Contribution: I collected the plants, performed the experiments, analysed the data and wrote the manuscript. Pascal-Antoine Christin provided assistance throughout and comments on the manuscript.

4.1 Abstract

The evolution of numerous adaptive traits involves the co-option of existing genes into new functions, and the genomic constituents of an organism are thus likely to constrain future evolutionary trajectories. Some properties of genes have been shown to increase their suitability for a given new function, thereby boosting the frequency at which they are co-opted. However, whether this increased suitability can be maintained over large evolutionary times remains unknown.

In this study, we investigate gene co-option linked to the recurrent origins of Crassulacean acid metabolism (CAM). Comparing independent origins of CAM in plants that diverged more than 400 million years ago (lycopods versus angiosperms), we show convergent gene co-option among these distant relatives. Surprisingly, we also demonstrate that transcriptome-wide expression patterns are partially conserved across the phylogeny of land plants, independently of the photosynthetic type. The ancestrally most highly expressed genes have been recurrently co-opted for CAM, explaining the commonalities between lycopods and angiosperms, but also between terrestrial and aquatic CAM. These results indicate that changes in expression patterns have been maintained over hundreds of millions of years, influencing the way novel adaptations are realized by very distant descendants.

4.2 Introduction

During evolution, organisms acquire novel adaptations that allow them to survive and thrive in a diversity of environments. This process involves changes in the inherited DNA, where the regulatory or coding sequences of ancestral genes are modified to trigger novel expression patterns and/or different catalytic properties (Christin et al., 2010; Martin and Orgogozo, 2013). While some adaptations require small quantitative adjustments in existing traits, others involve the emergence of novel biochemical pathways, in which case genes are co-opted from other functions to integrate into the new pathway (Christin et al., 2015; Huang et al., 2016). The likelihood of evolving new traits via gene co-option is therefore likely to depend on the existence of suitable genes in the ancestor of a group (Christin et al., 2010). While possessing genes encoding enzymes with the appropriate catalytic reactions is a very plausible prerequisite, other factors increasing gene co-optability likely include genetic redundancy and possessing properties close to those required in the new trait, such as particular gene product localisation, expression levels and allosteric interactions (Christin et al., 2010; Martin and Orgogozo, 2013; Rosenblum et al., 2014). However, whether genes maintain the properties facilitating co-option over large evolutionary scales remains largely unknown.

Adaptive traits that evolved repeatedly in different groups constitute outstanding systems to assess the factors that facilitate innovation. In plants, CO_2-concentrating mechanisms represent highly convergent novelties that are based on a large number of modifications (Sage et al., 2011; Edwards and Ogburn, 2012; Borland et al., 2014). These include C_4 and CAM photosynthesis, which both concentrate atmospheric CO_2 before its fixation by RuBisCO, the enzyme responsible for carbon assimilation in all oxygenic photosynthetic organisms. This concentration relies on the coordinated action of multiple enzymes, which are segregated spatially in C_4 plants and temporally in CAM plants (Hatch, 1987; Osmond, 1978). In CAM plants, atmospheric CO_2 is taken up at night, hydrated via carbonic anhydrase (CA) and fixed to phosphoenolpyruvate (PEP) by phosphoenolpyruvate carboxylase (PPC), resulting in the production of oxaloacetate. PEP is supplied by the breakdown of starch and/or soluble sugars during the night (Borland et al. 2016). This is converted via malate-dehydrogenase (MDH) to malate, which is actively transported into the vacuole using the proton-motive force provided by vacuolar ATPases. During the day, malate is exported from the vacuole and

decarboxylated by one or both of the following pathways; a) conversion via MDH to oxaloacetated, then decarboxylated to produce CO_2 and PEP via phosphoenolpyruvate carboxykinase (PCK), or b) conversion via NADP- or NAD-dependent malic enzyme (NADP-ME/NAD-ME) to oxaloacetate, which is then converted to PEP via pyruvate, phosphate dikinase (PPDK). This PEP is converted back into storage carbohydrates via the enzymes of gluconeogenesis (Borland et al. 2016). These processes must be temporally separated to avoid futile cycling, with some CAM genes showing diurnal expression patterns (Brilhaus et al. 2015; Ming et al. 2015; Yang et al. 2017). Information from the circadian clock, as well as light levels and water, are integrated to produce the temporal rhythms that characterise CAM photosynthesis (Wilkins, 1992; Hartwell et al. 1996; Ceusters et al. 2014; Males and Griffiths 2017). C_4 photosynthesis involves the same core biochemistry as CAM, but separation of nonphotosynthetic and photosynthetic carbon fixation occurs spatially, in the mesophyll and bundle sheath cells, respectively.

All enzymes of the CAM/C_4 cycles are believed to have existed in the C_3 ancestors, but were responsible for different, mainly non-photosynthetic functions (Aubry et al., 2011; Maier et al., 2011; Huang et al., 2015; Ting et al., 2017). From a biochemical point of view, the evolution of CO_2-concentrating mechanisms therefore involved the co-option of multiple genes, with their subsequent upregulation and in some cases adaptation of the kinetic properties for the new catalytic context (Tausta et al., 2002; Svensson et al., 2003; Cushman et al., 2008; Yang et al., 2017; Moreno-Villena et al., 2018).

Genes of the C_4/CAM pathways are encoded by gene families consisting of multiple gene copies, with duplicates created during recurrent gene or genome duplications (Christin et al., 2013; Christin et al., 2015; Zhang et al., 2016; Yang et al., 2017). After millions of years of independent evolution, the duplicates differ in their expression levels and catalytic properties (Tausta et al., 2002; Moreno-Villena et al., 2018). Comparative analyses have shown that independent C_4 origins within some groups tended to co-opt the same duplicate more often than expected by chance (Christin et al., 2013; 2015).

In grasses, the ancestrally highly expressed genes were more likely to be co-opted, so that transcript abundance in leaves seems to have facilitated the evolution of the CO_2-concentrating mechanism (Moreno-Villena et al. 2018). However, the identity of genes co-opted for C_4 however differed among C_4 origins occurring in distant clades

of flowering plants (Christin et al., 2015), suggesting that gene suitability might not be maintained over long evolutionary periods. Convergent gene recruitment for CAM has been reported both within (Deng et al., 2016; Zhang et al., 2016) and among groups of angiosperms (Yang et al., 2017), although systematic tests are still missing. In addition, angiosperms represent only a small part of the diversity of land plants, and while C_4 photosynthesis is restricted to this group (Sage et al., 2011), CAM is also found in distant lineages of land plants (Edwards and Ogburn, 2012), providing an opportunity to compare gene co-option among very distantly related groups.

CAM photosynthesis is classically associated with arid systems, where it allows cacti and other succulents to thrive (Osmond, 1978; Han and Felker, 1997). It is also widespread in various groups of epiphytes, and evolved in many clades spread across the phylogeny of flowering plants (Edwards and Ogburn, 2012). In addition, several groups of aquatic plants use CAM in low-CO_2 environments (Keeley, 1998). While several of these occur in angiosperms, aquatic CAM is also present in *Isoëtes*, a genus of lycopods (Keeley, 1981). These basal vascular plants diverged from angiosperms more than 400 million years ago (Kenrick and Crane, 1997), and it is therefore possible to compare CAM origins among some of the most distantly related land plants. The environmental conditions in which aquatic CAM occurs are very different to those of terrestrial CAM, with increased water availability but reduced light levels, rates of gas exchange and temperature. Gene recruitment patterns in *Isoëtes* compared to terrestrial CAM plants could therefore be confounded by these effects. A submerged aquatic CAM eudicot, *Littorella uniflora*, is frequently found in sympatry with *I. lacustris*, and shares many morphological adaptations for living in cold, oligotrophic lakes in northern Europe (Wium-Andersek and Andersen, 1972; Boston and Adams, 1985; Boston et al., 1987; Madsen et al., 2002). Comparison of these two species with C_3 and CAM terrestrial plants can therefore be used to determine whether factors determining gene co-option are retained over long evolutionary periods, and whether environmental conditions coupled with phylogenetic identity affect the realized biochemical phenotype.

In this study, we compared the transcriptomes of distantly related lineages of aquatic CAM plants in the lycopods and flowering plants, and those of terrestrial CAM and C_3 flowering plants, to test for convergent gene co-option across large evolutionary scales. Newly generated transcriptome data are used to (i) identify the CAM-specific genes in an aquatic lycopod (*Isoëtes lacustris*) and an aquatic CAM flowering plant

(*Littorella uniflora*) and (ii) determine whether the same genes were co-opted in these two distant lineages. The data are then compared to expression data from a C_3 flowering plant (*Arabidopsis thaliana*) to (iii) determine whether gene expression patterns are conserved over large evolutionary scales, independently of the photosynthetic type. A terrestrial CAM flowering plants (*Ananas comosus*) is then integrated to (iv) determine whether gene co-option is convergent, or differs, between CAM in terrestrial and aquatic environments. Together, these investigations shed new light on the factors facilitating adaptive transitions in land plants.

4.3 Materials and Methods

Sampling to assess diurnal expression patterns

Genes involved in the CAM biochemical pathway are expressed in photosynthetic tissues at high levels, and some are diurnally regulated (Cushman et al., 2008). These predicted patterns were used to identify genes co-opted for the CAM pathway in *Isoëtes lacustris*, using previously generated RNA-seq data consisting of three samples collected three hours after the end of the light period and three samples three hours after the end of the dark period (Chapter III). CAM activity was verified by performing acid titrations on leaf samples collected one hour before the end of the light period and one hour after the end of the dark period on the same days as those used for RNA-seq. The accumulation of C_4 acids during the night and their use during the day in CAM plants results in higher levels of malate in pre-dawn samples than pre-dusk samples (Osmond, 1978).

Leaf samples were weighed and ground in liquid nitrogen. The resulting powder was resuspended in 45 ml of boiling 20% ethanol and the mixture was boiled for a further 10 minutes. A pH indicator (100 µl 0.2% o-cresolpthalein) was added to the cooled mixture, which was titrated using 0.01N NaOH. Titratable acidity was calculated using the following formula:

$$A = Y * (N – B) / M$$

where A is the titratable acidity (umol H^+ g^{-1}), Y is the molarity of NaOH (mM), N is the volume of NaOH added until the sample changed colour (ml), B is the volume of NaOH

added to a blank sample of equivalent volume before a colour change (ml) and M is the mass of the sample (g).

The genes co-opted for CAM by *Littorella uniflora* were identified using individuals collected from Llyn Idwal, Wales. These were placed in a Sanyo Versatile Environmental Test Chamber in individual transparent plastic glasses 9 cm high with a diameter of 5 cm. A 3.7 cm layer of a mixture of sand, gravel, and sediment from Llyn Idwal was used as the substrate. Because this species has been reported to vary in its CAM activity as a function of the environment (Aulio, 1985; Madsen, 1987a) different treatments were used to attempt to capture this variation. A layer of clingfilm was stretched over each glass and perforated with four 0.5cm holes to produce a uniform limitation of gas exchange between treatments. Temperature was set at 12-15°C and humidity at 50%, in accordance with meteorological data for Llyn Idwal in summer, in a 15/9h light/dark cycle with lights at 150-200 μmol m^{-2} s^{-1}. The individuals were randomly assigned to three different treatments. First, a high-light submerged environment was achieved by filling the glasses to the top with deionised water, topping up when necessary. Second, low-light submerged conditions were achieved by covering some of the filled up glasses with four layers of muslin cloth, on top of the clingfilm and around the sides of the glass, which reduced light levels to ~60 μmol m^{-2} s^{-1}. Third, for the high-light terrestrial conditions, the sediment was saturated with deionised water, but the plant leaves were exposed to the atmosphere. The position of glasses within the chamber was randomized and changed once a week. After two weeks in these conditions, a minimum of three individuals were sampled for RNA extraction and acid titrations as described above.

Sampling in high- and low-CO_2 conditions

CAM is hypothesised to enable enhanced photosynthesis at low levels of CO_2 (Keeley, 1981; Keeley, 1998; Pedersen et al., 2011), and the strength of the CAM cycle has been reported to vary between environments and seasons, potentially due to variations in CO_2 availability (Keeley et al., 1983a; Keeley et al., 1983b; Keeley and Busch, 1984). To investigate the effects of reduced CO_2 availability on CAM activity and gene expression patterns, RNA-seq data were collected from plants grown in ambient and low CO_2 conditions. *Littorella uniflora* and *I. lacustris* individuals collected in Cwm Idwal and stored in large ponds outside for six months were transferred to 40 x 30 x 25cm

transparent plastic containers, with a 3.5 cm layer of soil consisting in a mixture of silica sand and Humax Sterilised Loam (East Riding Horticulture) soil in a ratio of 10:1. Two replicated containers were placed in each of two Conviron growth chambers, with CO_2 levels set to 400ppm and 180ppm, respectively. The day/night cycle was set to 14/10h, with a temperature of 25/20 C. Light levels were of 500 µmol $m^{-2} s^{-1}$, and the relative humidity was set to 60%. After eight days in these conditions, a minimum of three individuals of each species were sampled as described previously. To gain precision, malate levels in these samples were quantified using an ultra-high performance mass spectrometer. Soluble metabolites were extracted from freeze-dried leaves using a methanol/chloroform extraction, and run on a mass spectrometer coupled to a high performance liquid chromatograph in negative ion mode at the Sheffield BiOMICS Mass Spectrometry Facility.

Transcriptome sequencing and assembly, and transcript abundance estimates

RNA was extracted from frozen leaves using the Qiagen RNeasy® Plant Mini Kit, with a DNase I on-column digestion step (Qiagen Rnase-Free DNase Set). RNA quality and quantity were assessed using gel electrophoresis and a Nanodrop 8000, and library preparation was performed using the TruSeq RNA Sample Prep Kit v2 (Illumina®). Paired-end reads were generated on an Illumina HiSeq 2500 platform in rapid mode with 100 cycles. A total of 24 libraries were pooled per flow cell lane. Note that because the photosynthetic type of the *L. uniflora* individuals grown in low-light submerged conditions was not clear (see Results), these were not used for transcriptome analyses.

Reads were cleaned using NGS QC toolkit to remove adapter sequences, reads with ambiguous bases and reads with less than 80% high quality bases (q>20). Low quality bases (q<20) were further removed from the 3' end of reads. All the reads from each species were combined to assemble species-level reference transcriptomes using trinity v2.3.2. The longest isoform from each trinity contig was selected, and its longest coding sequence as predicted by TransDecoder was used in subsequent analyses. This resulted in 215,303 contigs for *L. uniflora* and 190,305 for *I. lacustris*.

Because aquatic plants are prone to contaminations by algae, contigs of potential algal origin were identified and removed. Using a stringent filtering, only contigs with a protein percentage similarity at least 5% higher when compared to a land plant than to the alga *Volvox carteri* (Supplementary Table 4.1) were retained. Reads were mapped to

the remaining contigs using bowtie2 v2.3.2. The majority of *L. uniflora* samples from the ambient vs. low CO_2 experiment had very low read mapping to the filtered contigs as a result of large levels of rRNA contamination, and were therefore excluded from the rest of the analysis. Contigs receiving at least five reads per million of mapped reads (rpm) in each sample were used in subsequent analyses. These successive filtering steps reduced the number of contigs from 215,303 to 8,751 in *L. uniflora* and 190,305 to 8,960 in *I. lacustris*.

To avoid problems created by genes assembled in separate, non-overlapping contigs, or redundancy among contigs, contigs were grouped by land plant co-orthologs, based on their similarity to co-orthologs of the model lycopod *Selaginella moellendorffii* and the model flowering plant *Arabidopsis thaliana.* Groups of co-orthologs represent genes descended by a combination of speciation and/or gene duplication from a single gene in the last common ancestor of *S. moellendroffii* and *A. thaliana*. The 5,081 gene families from Ensembl database containing at least one gene sequence from each of *S. moellendorffii* and *A. thaliana* were selected, and a BLAST search identified reciprocal best matches between *A. thaliana* and *S. moellendorfii*. Pairs of reciprocal best matches were considered as co-orthologs, which can be compared among land plants. To provide thresholds for assigning *I. lacustris* and *L. uniflora* contigs to groups of co-orthologs, the dataset consisting of *A. thaliana* sequences included in groups of co-orthologs was compared to itself using a BLAST search, and the percentage similarity of the best BLAST match was used as a the upper threshold of false positives. *Isoëtes lacustris* and *L. uniflora* contigs were subsequently compared to *A. thaliana* sequences placed in groups of co-orthologs and assigned to the group of the best matching sequence if the similarity was above that observed among *A. thaliana* sequences. This resulted in multiple contigs assigned to the same group of co-orthologs, which represent either duplicates that arose after the split of the studied species or redundant contigs from the same gene. The read abundance was computed per group of co-orthologs in number of reads per million of mapped reads per kilobase (RPKM), using the length of the alignable sequences.

Equivalent estimates of transcript abundances were obtained for *Ananas comosus*, a terrestrial CAM angiosperm, and *Arabidopsis thaliana*, a terrestrial C_3 angiosperm, using existing datasets (Supplementary Table 4.1) and methods described previously, although the step of filtering by comparison to *V. carteri* sequences was not performed for these taxa where contamination by algae is unlikely.

Differential gene expression analyses

As both *I. lacustris* and *L. uniflora* performed CAM in all the experimental conditions (see Results), genes differentially expressed between the light and dark time points were detected using simultaneously all treatments. Samples with more than one million reads mapping to the filtered species-level contigs were used in the analyses, resulting in 18 samples out of 18 for *I. lacustris* (nine for the light and nine for the dark conditions). The *I. lacustris* samples included the six generated in Chapter III, plus the six from the low-CO_2 treatment and the six from the ambient-CO_2 treatment. Out of the 12 *L. uniflora* samples from the high-light submerged and high-light terrestrial treatments, nine had more than one million reads mapping to the species-level contigs and were retained for differential expression analyses (five for the light and four for the dark conditions). Differential gene expression analyses were then performed using the edgeR R package (Robinson et al., 2010), with library sizes normalised by default and RNA composition normalised by the trimmed mean of M-values (TMM) method. Generalized linear models were used, with the experimental condition included as a co-factor, in addition to the light/dark condition. Common, trended and tagwise dispersions were estimated using the *estimateDisp* function and likelihood ratio tests were performed to identify differentially expressed genes using a 5% false discovery rate (FDR). The same approach was used to identify *I. lacustris* genes differentially expressed between the high- and low-CO_2 treatments, both diurnally expressed and differentially expressed between day and night independent of time. This was performed on the set of plants generated in this study only, including the time points and CO_2 levels as factors in the generalized linear model.

Gene co-option and biased recruitment

To estimate whether gene co-option for CAM is convergent among distant relatives, we identified the most abundant groups of co-orthologs from each gene family. The average value for the orthologs in each gene family during the day were calculated, and the most highly expressed ortholog was identified. Gene families present in every species were used to estimate how often the same ortholog was the most abundant between species. A null model was calculated by performing 10,000 simulations of randomly assigning

each species an ortholog within each gene family, for all the gene families considered. These analyses were conducted first over the whole transcriptome, and then only on CAM-related genes – those containing orthologs to genes identified as CAM-related in Brilhaus et al., 2016 (Supplementary Table 4.2). In the case of convergent gene co-option, similarity would be expected only for pairs of species sharing the same physiological character state. However, conserved expression patterns across evolutionary scales would result in overall commonality, independently of the photosynthetic states.

4.4 Results

CAM activity detected in most conditions

Clear excess of the titrable acidity in the predawn versus predusk samples was observed for *L. uniflora* in both high-light environments, albeit with high levels of variability between samples (Figure 4.1). Differences were less marked in the low-light submerged treatment, but the high level of titratable acidity means that CAM activity cannot be ruled out (Figure 4.1). Clear decreases from the predawn to the predusk titratable acidity were also observed in *I. lacustris* (Figure 4.1). The levels of malate similarly decreased between predawn and predusk samples of *I. lacustris* and *L.* uniflora grown in both the ambient and low CO_2 treatments (Figure 4.2), and we concluded that reduced levels of CO_2 did not significantly affect CAM activity in *I. lacustris* or *L. uniflora* under the growth conditions used.

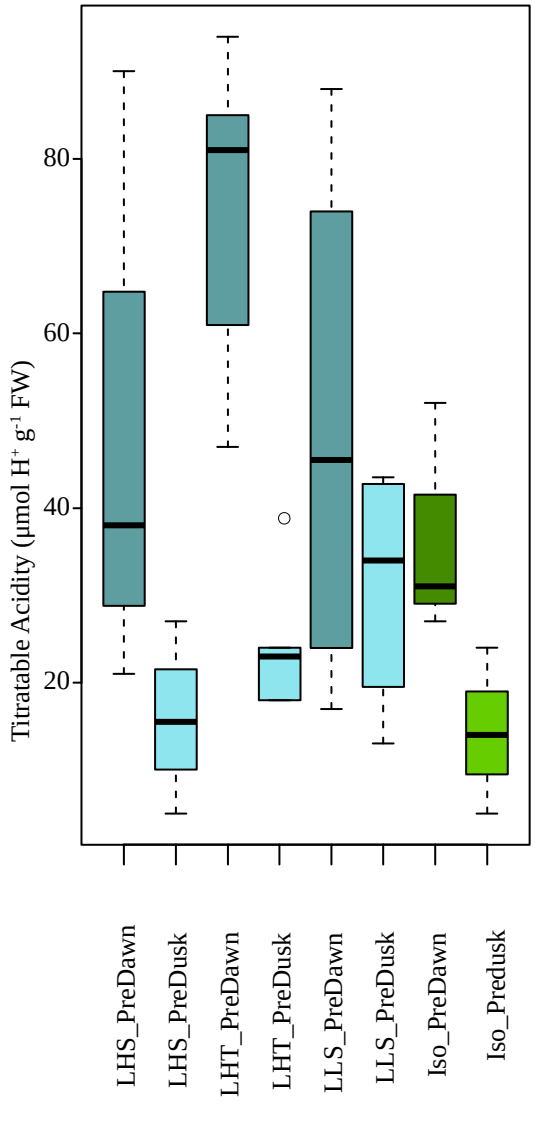


Figure 4.1: Diurnal acid fluctuations measured by acid titrations
Titratable acidity measured in µmol of H^+ per gram of fresh tissue, one hour before dawn (PreDawn) and one hour before dusk (PreDusk) of *Littorella uniflora* in high-light submerged (LHS), high light terrestrial (LHT) and low-light submerged (LLS) conditions, and *Isoetes lacustris* (Iso). n=3-5.

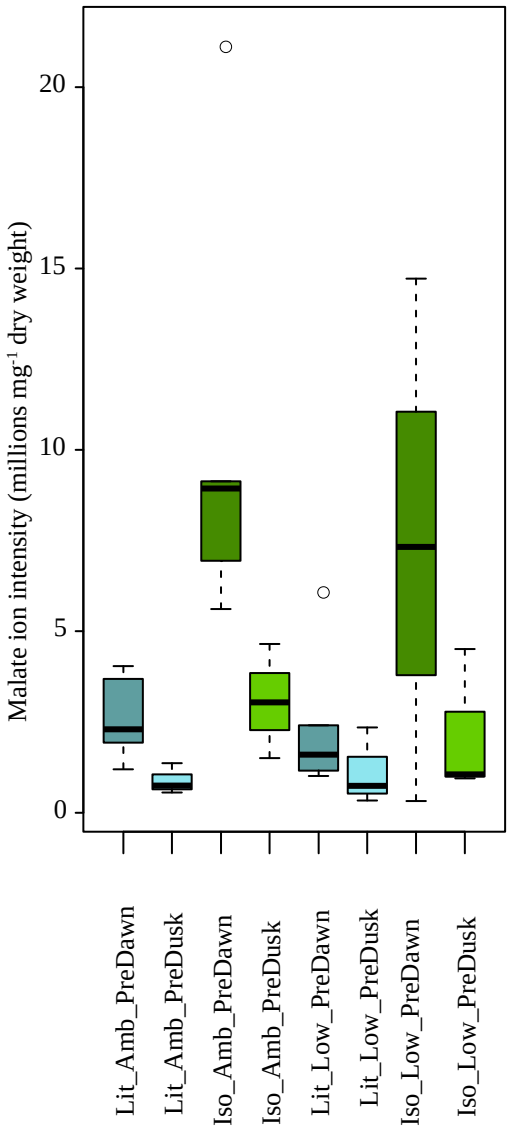


Figure 4.2 Diurnal fluctuation of malate levels measured by mass spectrometry
Malate level measurements expressed in millions of malate ions detected per mg of dry tissue, one hour before dawn (PreDawn) and one hour before dusk (PreDusk) of *Littorella uniflora* (Lit) and *Isoetes lacustris* (Iso) in ambient (Amb) and low (Low) CO_2 conditions. n=3-7.

Diurnally expressed, CAM-related genes and effects of CO_2 levels in I. lacustris

A total of 445 groups of orthologs from *I. lacustris* were differentially expressed between day and night, of which 279 were most abundant during the day and 166 during the night (Supplementary Table 4.3). Differentially expressed orthologs included those orthologous to *A. thaliana* circadian clock and photoperiod sensing genes such as Gigantea, Early Flowering 4-Like (ELF4), and Constitutive photomorphogenesis protein 1 (COP1) (Supplementary Table 4.3).

A total of 97 orthologs encoding CAM-related enzymes were identified, of which 31 were significantly differentially expressed between the day and the night (Table 4.1). The ortholog for β-CA was significantly upregulated during the day (14,000 RPKM vs. 6,800 RPKM for the night; Table 4.1). No ortholog for PPC was differentially expressed between the day and night, but the ortholog to the so-called "bacterial-type" PPC ortholog (PPC4) was very highly expressed at both time points (36,000 RPKM in the day vs. 28,000 RPKM in the night) compared to the "plant-type" PPC ortholog (PPC2), which was expressed at 980 rpkm during the day vs. 630 RPKM at night (Table 4.1). No significant expression of PPC kinase (PPCK1), a key activator of night-time PPC activity in other CAM species (Hartwell et al. 1999), was detected. The decarboxylase PCK was highly expressed, significantly more during the day (23,000 RPKM during the day vs. 11,000 RPKM during the night; Table 4.1). One of three MDH orthologs (MDH2) was highly expressed in both day and night samples (7,500 RPKM in the day vs. 5,200 RPKM at night). Other CAM-related genes were found at moderate (NADP-ME, NAD-ME1, NAD-ME2, PPDK) or low levels (Table 4.1).

Table 4.1: Expression levels of *Isoëtes lacustris* CAM co-orthologs

***Arabidopsis thaliana* ortholog**	**Day RPKM**	**Night RPKM**	**log2(Fold change rpm)[1]**	**P-value**	**Description**
AT1G68750	36000	28000	0.22	0.18	Phosphoenolpyruvate carboxylase 4 (PPC4)(
AT3G04120	39000	22000	0.76	0.00028	Glyceraldehyde-3-phosphate dehydrogenase GAPC1, cytosolic
AT5G65690	23000	11000	0.84	3.70E-07	Phosphoenolpyruvate carboxykinase 2
AT2G01140	18000	12000	0.47	0.011	Fructose-bisphosphate aldolase 3, chloroplastic
AT5G64860	14000	10000	0.36	0.027	4-alpha-glucanotransferase DPE1, chloroplastic/amyloplastic
AT3G01390	15000	7400	0.82	0.00016	Vacuolar membrange ATPase 10 (VMA10)
AT1G70410	14000	6800	0.81	5.80E-05	Beta carbonic anhydrase 4
AT4G11150	12000	7100	0.55	0.00065	V-type proton ATPase subunit E1
AT3G46970	11000	5600	0.7	0.0028	Alpha-glucan phosphorylase 2, cytosolic
AT4G38510	9200	4900	0.75	2.50E-06	V-type proton ATPase subunit B2
AT2G36530	8800	4800	0.65	0.0024	Low expression of osmotically responsive genes 2 (LOS2)
AT5G43330	7500	5200	0.41	0.12	Malate dehydrogenase 2, cytoplasmic
AT3G58730	6500	4900	0.17	0.35	V-type proton ATPase subunit D
AT1G78900	7500	3900	0.74	3.80E-06	Vacuolar ATP synthase subunit A (VHA-A)
AT1G42970	6100	4900	0.14	0.42	Glyceraldehyde-3-phosphate dehydrogenase GAPB, chloroplastic
AT3G29320	7400	3500	0.77	2.20E-07	Alpha-glucan phosphorylase 1
AT3G23920	5400	4000	0.24	0.38	Beta-amylase
AT5G04360	5100	4200	0.16	0.46	Pullulanase 1, chloroplastic
AT3G55440	4900	3900	0.34	0.13	Triosephosphate isomerase, cytosolic
AT1G09780	5000	2700	0.66	0.00046	2,3-bisphoshoglycerate-independent phosphoglycerate mutase 1 (IPGAM1)
AT5G48300	4200	3300	0.32	0.17	Glucose-1-phosphate adenylyltransferase small subunit, chloroplastic
AT3G12780	4600	2800	0.59	0.00062	Phosphoglycerate kinase
AT3G55800	3800	3300	0.01	0.94	Sedoheptulose-1,7-bisphosphatase, chloroplastic
AT5G24300	4200	2600	0.42	0.045	Starch synthase, chloroplastic/amyloplastic
AT3G42050	4200	2600	0.51	0.0085	V-type proton ATPase subunit H
AT1G19450	3300	2900	-0.04	0.83	Sugar transporter ERD6-like 4
AT1G12840	4000	1800	0.91	3.50E-07	V-type proton ATPase subunit C
AT1G27680	4000	1800	0.98	1.40E-06	Glucose-1-phosphate adenylyltransferase large subunit 2, chloroplastic
AT4G26530	3600	1800	1.1	1.50E-06	Fructose-bisphosphate aldolase
AT2G22480	3000	2300	0.14	0.47	ATP-dependent 6-phosphofructokinase 5, chloroplastic
AT1G16300	3100	1300	1.1	6.20E-11	Glyceraldehyde-3-phosphate dehydrogenase GAPCP2, chloroplastic
AT3G54050	1800	1800	-0.15	0.26	Fructose-1,6-bisphosphatase 1, chloroplastic
AT1G43670	2200	1400	0.47	0.0021	Fructose-1,6-bisphosphatase, cytosolic
AT4G02620	2100	1100	0.8	0.0018	V-type proton ATPase subunit F
AT5G03650	1500	1400	0.01	0.95	1,4-alpha-glucan-branching enzyme 2-2, chloroplastic/amyloplastic
AT1G11720	1700	1100	0.53	0.0051	Starch synthase 3
AT1G12000	2100	670	1.3	7.40E-11	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta 1

AT5G16150	1300	1400	-0.18	0.24	Plastidic glucose transporter 4
AT5G46800	1300	1400	-0.16	0.4	Mitochondrial carnitine/acylcarnitine carrier-like protein
AT5G22620	1500	1100	0.26	0.042	Probable 2-carboxy-D-arabinitol-1-phosphatase
AT5G08680	1200	1200	0.03	0.91	ATP synthase subunit beta-3, mitochondrial
AT5G19760	1300	1000	0.15	0.43	Mitochondrial dicarboxylate/tricarboxylate transporter DTC
AT1G32900	1700	580	1.4	1.40E-08	Granule-bound starch synthase 1, chloroplastic/amyloplastic
AT5G42740	1500	670	0.94	6.70E-08	Glucose-6-phosphate isomerase
AT2G40840	750	1200	-0.45	0.19	4-alpha-glucanotransferase DPE2
AT1G10760	1100	830	0.18	0.39	Alpha-glucan water dikinase 1, chloroplastic
AT4G29130	1000	820	0.18	0.27	Phosphotransferase
AT1G70730	960	830	0.01	0.91	Phosphoglucomutase
AT1G69830	1100	600	0.67	0.00015	Alpha-amylase 3, chloroplastic
AT5G54800	1100	540	0.78	0.0024	Glucose 6-phosphate/phosphate trans;locator 1 (GPT1)
AT1G20950	740	880	-0.35	0.054	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha
AT2G42600	980	630	0.45	0.0059	Phosphoenolpyruvate carboxyklase 2 (PPC2)
AT3G48680	820	760	0	0.99	Gamma carbonic anhydrase-like 2, mitochondrial
AT4G25000	930	640	0.31	0.19	Alpha-amylase 1
AT3G47520	1000	560	0.58	0.0027	Malate dehydrogenase, chloroplastic
AT1G15690	660	860	-0.61	0.0075	Vacuolar proton pyrophosphatase 1 (VHP1)
AT4G15530	870	570	0.48	0.022	Pyruvate, phosphate dikinase 1, chloroplastic
AT5G58330	780	650	0.06	0.7	Malate dehydrogenase
AT3G52180	800	530	0.52	0.00091	Starch-excess 4 (SEX4)
AT3G01510	590	670	-0.3	0.02	Like SEX4 1 (LSF1)
AT5G26570	660	580	0.1	0.61	Phosphoglucan, water dikinase, chloroplastic
AT2G26900	700	500	0.22	0.2	Bile acid: sodium symporter family protein 2 (BASS2)
AT4G26270	660	520	0.14	0.32	ATP-dependent 6-phosphofructokinase 3
AT2G21170	500	620	-0.39	0.022	Triosphosphate isomerase
AT5G01340	580	530	-0.14	0.38	Mitochondrial succinate-fumarate transporter 1
AT4G24620	480	440	0	1	Glucose-6-phosphate isomerase
AT2G39930	530	300	0.64	4.10E-06	Isoamylase 1 (ISA1)
AT3G19490	420	350	-0.1	0.45	Sodium/proton antiporter 1
AT5G17520	530	240	1	1.40E-05	Maltose excess protein 1, chloroplastic
AT3G01180	400	330	0.28	0.18	Starch synthase, chloroplastic/amyloplastic
AT4G00490	390	230	0.49	3.00E-04	Beta-amylase 2, chloroplastic
AT1G16780	330	280	0	0.97	Pyrophosphate-energized membrane proton pump 3
AT3G25410	400	180	0.78	2.50E-05	Probable sodium/metabolite cotransporter BASS3, chloroplastic
AT1G03310	300	250	0.04	0.77	Isoamylase 2 (ISA2)
AT2G20780	250	220	-0.28	0.22	Probable polyol transporter 4
AT5G12860	260	190	0.25	0.21	Dicarboxylate transporter 1, chloroplastic
AT5G59250	250	200	-0.04	0.8	D-xylose-proton symporter-like 3, chloroplastic
AT5G64380	200	220	-0.31	0.046	Inositol monophosphate family protein
AT1G30220	190	210	-0.27	0.1	Probable inositol transporter 2
AT5G33320	190	210	-0.36	0.012	Phosphoenolpyruvate/phosphate translocator 1, chloroplastic
AT5G17530	200	170	0.09	0.45	phosphoglucosamine mutase family protein

AT2G29560	170	150	0.03	0.88	Cytosolic enolase 3
AT1G79900	190	130	0.4	0.033	Mitochondrial arginine transporter BAC2
AT3G10940	190	120	0.35	0.26	Phosphoglucan phosphatase LSF2, chloroplastic
AT5G47810	170	100	0.45	0.0036	ATP-dependent 6-phosphofructokinase 2
AT2G33820	100	150	-0.52	0.00087	Mitochondrial substrate carrier family protein
AT3G01200	120	120	-0.25	0.07	Pyruvate, phosphate dikinase regulatory protein 2
AT1G05030	130	96	0.23	0.33	Probable plastidic glucose transporter 1
AT3G54110	110	110	-0.14	0.44	Uncoupling protein 1 (UCP1)
AT1G74030	130	97	0.16	0.55	Enolase 1, chloroplastic
AT2G13560	97	83	0.16	0.44	NAD-dependent malic enzyme 1, mitochondrial
AT1G78560	87	83	-0.19	0.32	Probable sodium/metabolite cotransporter BASS1, chloroplastic
AT5G64290	91	74	0.05	0.77	Dicarboxylate transporter 2, chloroplastic
AT1G67300	85	78	-0.08	0.64	Putative plastidic glucose transporter 2
AT3G20440	67	76	-0.34	0.03	1,4-alpha-glucan-branching enzyme 3, chloroplastic/amyloplastic
AT4G09020	66	35	0.8	0.013	Isoamylase 3, chloroplastic
AT5G47560	34	35	-0.21	0.12	Tonoplast dicarboxylate transporter

1. Positive values represent genes more highly expressed during the day.

Out of 5,015 orthologs which were included in the low CO_2 versus ambient CO_2 comparison of *I. lacustris*, only 33 were significantly differentially expressed between the day and the night in both conditions, likely reflecting a reduction in the statistical power with a smaller number of replicates (3 vs. 9). Another 99 were significantly differentially expressed between the day and the night, but only in ambient CO_2 conditions. These include Gigantea orthologs, as well as MDH, β-CA and PCK (Supplementary Table 4.4). Conversely, 23 genes showed a light/dark expression pattern only in low CO_2, including COP1 (Supplementary Table 4.4).

Independent of time, a total of 65 orthologs were differentially expressed between low and high CO_2, of which 52 were more abundant in the ambient-CO_2 treatment. Orthologs encoding enzymes involved in photosynthesis and downstream processes such as chlorophyll-binding proteins, photosystem II proteins, accumulation and replication of chloroplasts and cellulose synthase were downregulated in the low CO_2 conditions, Similarly, early-light induced protein 1 (ELIP1), which binds to free chlorophyll and prevents the production of singlet oxygen (Hutin et al. 2003), was upregulated in the low CO_2 conditions, consistent with excess light relative to available CO_2 (Table 4.2). Orthologs with roles in CAM that were downregulated in low CO_2 include the glycolytic/gluconeogenic enzymes enolase and glucose-6-phosphate isomerase (G6PI), and the H+ transporter vacuolar proton pyrophosphatase 1 (VHP1).

Table 4.2: Orthologs differentially expressed between high and low CO_2 conditions independent of time

Arabidopsis thaliana ortholog	log2(Fold Change)[1]	P-value	Description
AT5G05340	4.7	0.00037	Peroxidase 52
AT3G24480	4.5	5.50E-07	Leucine-rich repeat extensin-like protein 4
AT1G15820	4	1.40E-06	Chlorophyll a-b binding protein, chloroplastic
AT5G54270	3.7	4.40E-07	Chlorophyll a-b binding protein 3, chloroplastic
AT4G32410	3.4	1.00E-04	Cellulose synthase
AT3G14110	3.3	5.50E-13	Tetratricopeptide repeat (TPR)-like superfamily protein
AT1G03630	3.3	1.40E-06	Protochlorophyllide reductase C, chloroplastic
AT3G46780	3.1	1.20E-10	Protein plastid transcriptionally active 16, chloroplastic
AT1G58290	3.1	2.70E-10	Glutamyl-tRNA reductase 1, chloroplastic
AT1G44446	2.9	5.40E-06	Chlorophyll oxygenase 1 (CH1)
AT3G18890	2.8	9.00E-06	Translocon at the inner envelope membrane of chloroplasts 62
AT5G23060	2.7	9.60E-08	Calcium sensing receptor, chloroplastic
AT3G04260	2.7	2.00E-05	Plastid transcriptionally active 3
AT1G15690	2.6	8.40E-08	Vacuolar proton pyrophosphatase 1 (VHP1)
AT1G68890	2.6	2.00E-04	Protein PHYLLO, chloroplastic
AT5G61250	2.6	0.00028	Glucoronidase 1 (GUS1)
AT2G05100	2.6	0.00019	photosystem II light harvesting complex gene 2
AT1G54520	2.6	2.60E-10	Putative uncharacterized protein
AT1G20950	2.4	1.40E-06	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha
AT5G05740	2.3	2.70E-06	Probable zinc metalloprotease EGY2, chloroplastic
AT3G50820	2.3	0.00016	Oxygen-evolving enhancer protein 1-2, chloroplastic
AT3G08940	2.3	0.00056	Chlorophyll a-b binding protein CP29.2, chloroplastic
AT5G10150	2.2	5.80E-05	Protein UPSTREAM OF FLOWERING LOCUS C
AT4G01690	2.1	2.50E-09	Protoporphyrinogen (PPOX)
AT2G44060	2.1	0.00012	
AT4G15560	2.1	1.50E-05	1-deoxy-d-xylulose 5-phosphate synthase 1 (DXS1)
AT5G56510	2.1	4.60E-06	Pumilio homolog 12
AT1G74030	2.1	8.80E-08	Enolase 1, chloroplastic
AT5G42480	2	7.00E-06	Protein ACCUMULATION AND REPLICATION OF CHLOROPLASTS 6, chloroplastic
AT5G13630	2	0.00024	Magnesium-chelatase subunit ChlH, chloroplastic
AT3G15850	2	7.70E-05	Fatty acid desaturase 5
AT4G24750	2	7.20E-06	Rhodanese-like domain-containing protein 11, chloroplastic
AT1G07180	2	8.00E-05	Internal alternative NAD(P)H-ubiquinone oxidoreductase A1, mitochondrial
AT4G25450	1.9	2.50E-06	ABC transporter B family member 28
AT5G10690	1.9	1.90E-05	Pentatricopeptide repeat-containing protein At5g10690
AT3G17040	1.9	4.10E-06	Protein high chlorophyll fluorescent 107
AT1G30950	1.9	0.00017	Putative uncharacterized protein At1g30950
AT2G36250	1.8	0.00018	Cell division protein FtsZ homolog 2-1, chloroplastic
AT5G42740	1.8	0.00014	Glucose-6-phosphate isomerase
AT2G36530	1.8	0.00045	Low expression of osmotically responsive genes 2 (LOS2)
AT3G56940	1.8	0.00012	Magnesium-protoporphyrin IX monomethyl ester
AT4G34350	1.8	0.00019	4-hydroxy-3-methylbut-2-enyl diphosphate reductase, chloroplastic
AT3G14900	1.8	0.00065	Putative uncharacterized protein At3g14900
AT1G74470	1.7	0.00035	Geranylgeranyl diphosphate reductase, chloroplastic
AT5G24460	1.7	3.80E-06	Putative uncharacterized protein

AT1G15980	1.6	0.00035	Photosynthetic NDH subunit of subcomplex B 1, chloroplastic
AT4G24620	1.5	0.00022	Glucose-6-phosphate isomerase
AT5G37360	1.4	0.00034	Acclimation of photosynthesis to environment 1
AT3G24550	1.4	0.00018	Proline-rich receptor-like protein kinase PERK1
AT1G23360	1.4	0.00017	2-phytyl-1,4-beta-naphthoquinone methyltransferase, chloroplastic
AT1G73600	1.3	0.00011	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein
AT2G40490	1.2	0.00049	Uroporphyrinogen decarboxylase 2, chloroplastic
AT2G19570	-1.3	0.00034	Cytidine deaminase 1
AT1G07080	-1.4	0.00062	GLT domain-containing protein
AT2G41250	-1.4	0.00048	Haloacid dehalogenase-like hydrolase domain-containing protein
AT2G21960	-1.4	0.00051	Expressed protein
AT5G66380	-1.5	0.00052	Folate transporter 1, chloroplastic
AT5G24850	-1.5	0.00054	Cryptochrome DASH, chloroplastic/mitochondrial
AT3G58800	-1.6	0.00026	Putative uncharacterized protein At3g58800/T20N10 150
AT2G26560	-1.7	0.00031	Patatin-like protein 2
AT4G28740	-1.7	2.80E-05	
AT5G40670	-1.9	6.70E-05	Cystinosin homolog
AT3G56290	-2	0.00044	Potassium transporter
AT3G22840	-3.4	2.40E-07	Early light inducible protein 1 (ELIP1)

1. Positive values represent genes more highly expressed in ambient CO_2.

Diurnally expressed and CAM-related genes in L. uniflora

A total of 202 *L. uniflora* orthologs were differentially expressed between the day and night samples, of which 81 were more abundant during the day and 121 during the night (Supplementary Table 4.5). These include proteins involved in chlorophyll binding, photosystem proteins, as well as circadian clock genes such Reiveille and a cryptochrome ortholog.

A total of 99 orthologs to CAM-related genes were identified in *L. uniflora,* of which 15 were differentially expressed between the day and the night (Table 4.3). Similarly to *Isoëtes*, no gene for PPC was significantly differentially expressed between the day and night. The ortholog to the “plant-type” PPC (PPC2) was more highly expressed during the night than during the day (2,600 RPKM in the day vs. 7,900 RPKM at night), although the ortholog to the “bacterial-type” PPC4 was more highly expressed (11,000 RPKM in the day vs. 11,000 RPKM at night; Table 4.3). Similarly to *Isoëtes,* no PPCK1 homolog was highly expressed in the samples of *L. uniflora*. The decarboxylase PCK was again highly expressed, and significantly more abundant during the day (29,000 RPKM in the day vs 6,100 RPKM at night). Orthologs of PPDK, as well as the PPDK regulatory protein (PPDK-RP) that catalyses the reversible

phosphorylation of PPDK (Chastain et al. 2002), were significantly upregulated during the day (1,100 RPKM in the day vs. 800 RPKM at night and 200 RPKM in the day vs. 140rpkm at night, respectively), although as with *I. lacustris* malic enzyme transcript levels levels were relatively low (Table 4.3). Orthologs associated with starch breakdown for PEP provision such as Glucose 6-phosphate/phosphate translocator 1 (GPT1) and Starch-excess 4 (SEX4) were upregulated during the night, and starch synthase orthologs upregulated during the day (Table 4.3).

Table 4.3: Expression levels of *Littorella uniflora* CAM co-orthologs

Locus	Day RPKM	Night RPKM	log2(Fold Change)[1]	P-value	Description
AT5G65690	29000	6100	2.3	1.10E-08	phosphoenolpyruvate carboxykinase 2
AT2G01140	23000	11000	1.4	8.00E-06	Fructose-bisphosphate aldolase 3, chloroplastic
AT4G26530	12000	15000	-0.046	0.92	Fructose-bisphosphate aldolase
AT3G26650	12000	12000	0.064	0.85	Glyceraldehyde-3-phosphate dehydrogenase GAPA1, chloroplastic
AT1G68750	11000	11000	0.34	0.54	PPC4
AT3G04120	9200	11000	-0.034	0.91	Glyceraldehyde-3-phosphate dehydrogenase GAPC1, cytosolic
AT5G43330	8700	9000	0.23	0.58	Malate dehydrogenase 2, cytoplasmic
AT3G12780	5700	8600	-0.42	0.32	Phosphoglycerate kinase
AT2G36530	5400	8600	-0.53	0.11	Low expression of osmotically responsive genes 2 (LOS2)
AT1G42970	8600	4100	1.1	0.0015	Glyceraldehyde-3-phosphate dehydrogenase (GAPB), chloroplastic
AT1G16300	5900	6600	0.017	0.98	Glyceraldehyde-3-phosphate dehydrogenase (GAPCP2), chloroplastic
AT2G42600	2600	7900	-1.4	0.0042	Phosphoenolpyruvate carboxylase 2 (PPC2)
AT2G22780	5300	4400	0.6	0.091	Malate dehydrogenase
AT1G12000	4400	4900	-0.047	0.92	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta 1
AT3G55440	3900	5400	-0.34	0.3	Triosephosphate isomerase, cytosolic
AT3G29320	4900	4100	0.3	0.48	Alpha-glucan phosphorylase 1
AT1G70410	3800	4000	0.49	0.07	Beta carbonic anhydrase 4
AT1G78900	3800	4000	0.24	0.47	Vacuolar catalytic subunut A
AT1G09780	3100	4600	-0.12	0.74	Phosphoglycerate mutase, 2,3-bisphosphate glycerate-independent (IPGAM1)
AT3G55800	4200	3500	0.47	0.19	Sedoheptulose-1,7-bisphosphatase, chloroplastic
AT4G11150	2900	2900	0.33	0.29	V-type proton ATPase subunit E1
AT1G15690	2800	2500	0.31	0.46	Vacuolar proton pyrophosphatase 1 (VHP1)
AT4G38510	2300	2800	0.14	0.58	V-type proton ATPase subunit B2
AT1G43670	1300	2400	-0.97	0.0083	Fructose-1,6-bisphosphatase, cytosolic
AT1G32900	2800	740	2	2.00E-05	Granule-bound starch synthase 1, chloroplastic/amyloplastic
AT3G54050	1400	1800	-0.17	0.62	Fructose-1,6-bisphosphatase 1, chloroplastic
AT5G48300	1800	1400	0.74	0.045	Glucose-1-phosphate adenylyltransferase

					small subunit, chloroplastic
AT1G27680	1600	1500	0.36	0.28	Glucose-1-phosphate adenylyltransferase large subunit 2, chloroplastic
AT2G20780	1100	2000	-0.44	0.3	Probable polyol transporter 4
AT5G51820	1600	1500	0.44	0.28	Phosphoglucomutase, chloroplastic
AT5G58330	2100	990	1.6	1.60E-05	Malate dehydrogenase
AT3G58730	1300	1400	0.18	0.55	V-type proton ATPase subunit D
AT1G12840	1200	1400	0.14	0.66	V-type proton ATPase subunit C
AT5G54800	700	1900	-1.4	0.00064	Glucose 6-phosphate/phosphate trans;locator 1 (GPT1)
AT4G02620	1200	1300	0.26	0.34	V-type proton ATPase subunit F
AT5G03650	1200	1300	0.41	0.13	1,4-alpha-glucan-branching enzyme 2-2, chloroplastic/amyloplastic
AT3G52180	650	1800	-1.2	0.00032	Starch-excess 4 (SEX4)
AT5G08680	1000	1400	-0.05	0.83	ATP synthase subunit beta-3, mitochondrial
AT3G23920	720	1500	-0.057	0.9	Beta-amylase
AT3G42050	1100	1100	0.075	0.78	V-type proton ATPase subunit H
AT4G26270	970	1200	0.0064	0.98	ATP-dependent 6-phosphofructokinase 3
AT4G15530	1100	810	0.88	4.00E-04	Pyruvate, phosphate dikinase 1, chloroplastic
AT1G79750	730	1100	-0.49	0.053	Malic enzyme
AT1G10760	700	1100	-0.63	0.04	Alpha-glucan water dikinase 1, chloroplastic
AT4G24570	77	1700	-3.1	0.007	Dicarboxylate carrier 2 (DIC2)
AT5G24300	1100	580	1.2	0.001	Starch synthase, chloroplastic/amyloplastic
AT5G12860	1000	630	1.3	5.70E-08	Dicarboxylate transporter 1, chloroplastic
AT5G42740	790	690	0.66	0.027	Glucose-6-phosphate isomerase
AT1G53240	650	820	0.15	0.52	Malate dehydrogenase 1, mitochondrial
AT1G20950	440	1000	-0.91	0.068	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha
AT5G64860	680	760	-0.051	0.9	4-alpha-glucanotransferase DPE1, chloroplastic/amyloplastic
AT1G47260	620	810	0.026	0.91	Gamma carbonic anhydrase 2
AT1G69830	470	940	-0.36	0.18	Alpha-amylase 3, chloroplastic
AT5G19760	680	730	0.29	0.26	Mitochondrial dicarboxylate/tricarboxylate transporter DTC
AT1G11720	540	500	0.42	0.1	Starch synthase 3
AT2G26900	460	580	0.0059	0.98	Bile acid: sodium symporter 2 (BASS2)
AT2G21170	470	520	-0.039	0.89	Triosephosphate isomerase
AT4G24620	470	490	0.15	0.66	Glucose-6-phosphate isomerase
AT5G59250	410	520	0.2	0.37	D-xylose-proton symporter-like 3, chloroplastic
AT3G46970	180	710	-1.2	0.0081	Alpha-glucan phosphorylase 2, cytosolic
AT1G70730	370	510	-0.23	0.3	Phosphoglucomutase
AT2G13560	270	590	-0.45	0.28	NAD-dependent malic enzyme 1, mitochondrial
AT5G26570	390	470	-0.22	0.41	Phosphoglucan, water dikinase, chloroplastic
AT3G01180	700	90	3.1	9.50E-07	Starch synthase, chloroplastic/amyloplastic
AT5G46800	580	210	2	4.80E-08	Mitochondrial carnitine/acylcarnitine carrier-like protein
AT4G29130	300	330	0.34	0.16	Phosphotransferase
AT3G48680	250	370	-0.15	0.48	Gamma carbonic anhydrase-like 2, mitochondrial
AT5G17530	290	330	-0.066	0.82	phosphoglucosamine mutase family protein
AT1G30220	330	230	0.81	0.0089	Probable inositol transporter 2
AT3G47520	230	310	0.27	0.44	Malate dehydrogenase, chloroplastic
AT1G19450	150	340	-0.45	0.12	Sugar transporter ERD6-like 4
AT5G33320	180	300	-0.074	0.79	Phosphoenolpyruvate/phosphate translocator

					1, chloroplastic
AT2G39930	180	270	-0.17	0.63	Isoamylase 1 (ISA1)
AT3G18440	210	230	0.24	0.49	Aluminum-activated malate transporter 9
AT2G40840	160	270	-0.35	0.14	4-alpha-glucanotransferase DPE2
AT1G12580	190	210	0.12	0.69	Phosphoenolpyruvate carboxylase-related kinase 1
AT3G01510	140	230	-0.37	0.34	Like SEX4-1 (LSF1)
AT1G67300	220	130	1.6	2.90E-05	Putative plastidic glucose transporter 2
AT3G01200	200	140	0.89	0.0014	Pyruvate, phosphate dikinase regulatory protein 2
AT5G16150	83	240	-0.88	0.0076	Plastidic glucose transporter 4
AT1G16780	110	200	-0.47	0.056	Pyrophosphate-energized membrane proton pump 3
AT3G25410	160	140	0.58	0.063	Probable sodium/metabolite cotransporter BASS3, chloroplastic
AT5G64290	130	170	0.071	0.79	Dicarboxylate transporter 2, chloroplastic
AT4G09020	120	170	-0.085	0.74	Isoamylase 3, chloroplastic
AT3G54110	100	160	-0.33	0.22	Uncoupling protein 1 (UCP1)
AT5G22620	110	140	-0.29	0.35	Probable 2-carboxy-D-arabinitol-1-phosphatase
AT4G17260	64	180	-1	0.00095	L-lactate dehydrogenase
AT3G10940	110	130	-0.14	0.69	Phosphoglucan phosphatase LSF2, chloroplastic
AT5G01340	68	140	-0.35	0.2	Mitochondrial succinate-fumarate transporter 1
AT5G04360	98	94	0.31	0.23	Pullulanase 1, chloroplastic
AT4G00490	78	110	-0.31	0.21	Beta-amylase 2, chloroplastic
AT5G64380	79	100	-0.15	0.6	AT5g64380/MSJ1 22
AT5G17520	74	100	-0.31	0.37	Maltose excess protein 1, chloroplastic
AT1G14140	68	83	-0.04	0.88	Mitochondrial uncoupling protein 3
AT3G19490	51	99	-0.44	0.09	Sodium/proton antiporter 1
AT2G29560	59	89	-0.17	0.5	Cytosolic enolase 3
AT1G74030	39	100	-0.71	0.032	Enolase 1, chloroplastic
AT1G03310	74	56	0.72	0.0069	Isoamylase 2 (ISA2
AT3G20440	43	71	-0.45	0.054	1,4-alpha-glucan-branching enzyme 3, chloroplastic/amyloplastic

1. Positive values represent genes more highly expressed during the day.

Patterns of gene co-option

Of 829 gene families with multiple lineages expressed in leaves during the day, *I. lacustris* and *L. uniflora* shared the most highly abundant ortholog in 497 cases, which is significantly higher than expected by chance ($p < 0.001$, Figure 4.3). However, similar numbers of most highly abundant orthologs are shared between other pairs of species (Figure 4.4). A large proportion of most highly abundant orthologs are shared between all four species, significantly more than expected by chance ($p < 0.001$). Fifty orthologs were the most abundant exclusively in *I. lacustris* and *L. uniflora*, but this was fewer than for any of the other pairwise comparisons (Figure 4.4).

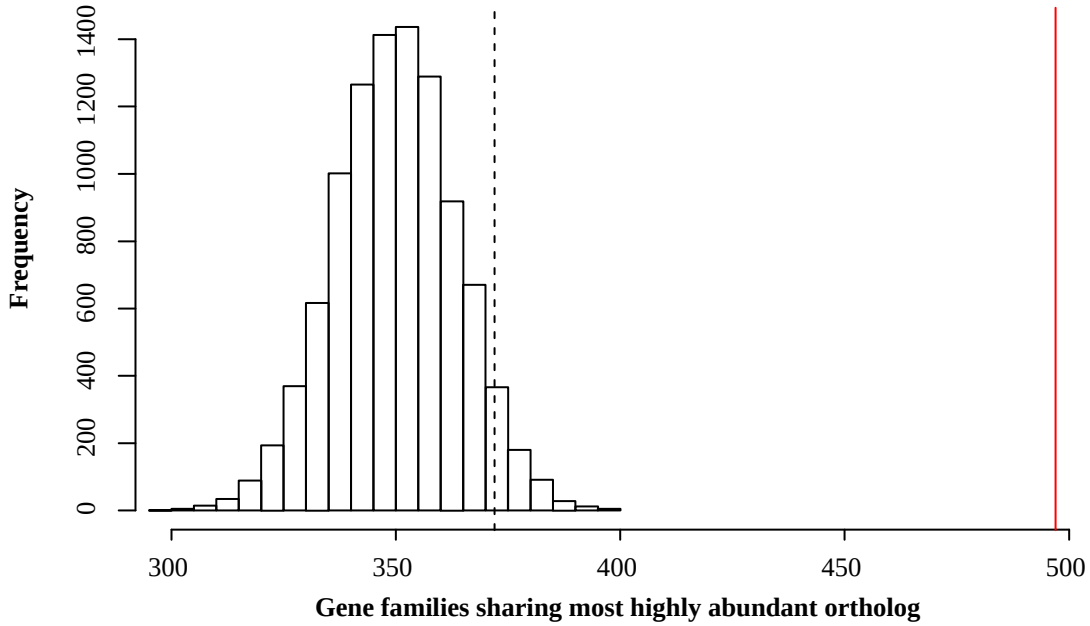


Figure 4.3 Gene families sharing most highly abundant ortholog in *Littorella uniflora* and *Isoëtes lacustris*.
Histogram represents number of shared orthologs observed from 10,000 simulations randomly assigning each of 829 gene families a most highly expressed ortholog and summing the number of times these were the same in *L. uniflora* and *I. lacustris*. The dashed line represents the threshold of significance ($p < 0.05$). The red line represents the observed number of gene families sharing the most highly abundant ortholog from RNA-seq data.

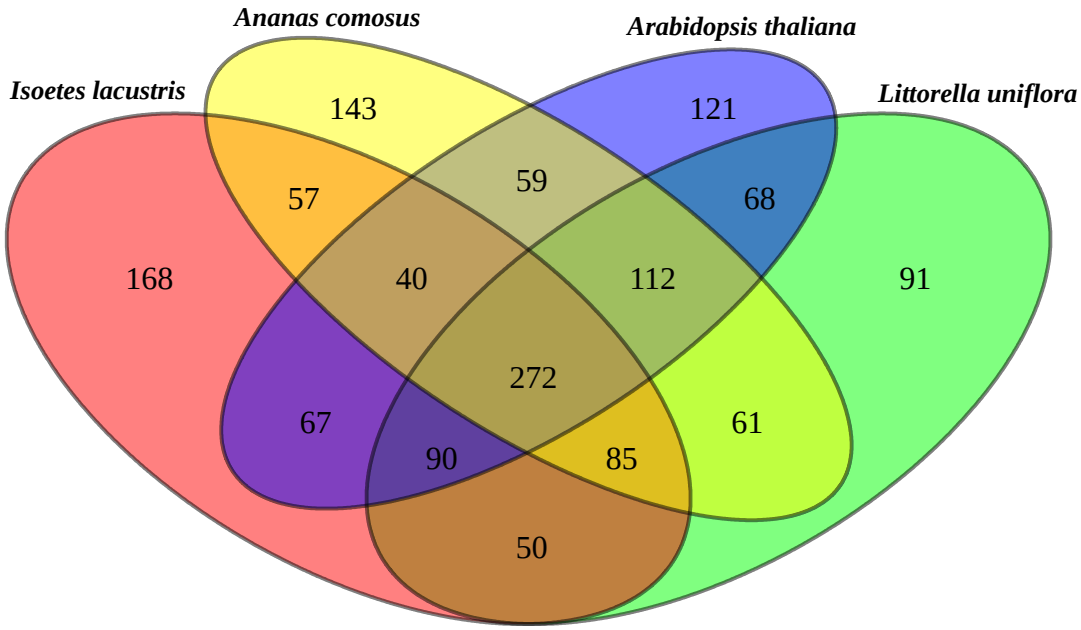


Figure 4.4 Overlap of shared most abundant orthologs within gene families across land plants
Venn diagram representing the numbers of times 829 gene families share the same most abundant ortholog between *Isoëtes lacustris* (red oval), *Ananas comosus* (yellow oval), *Arabidopsis thaliana* (blue oval) and *Littorella uniflora* (green oval).

Of the gene families included in the analyses, 31 encode enzymes and transporters that can be posited to play a role in CAM photosynthesis (Brilhaus et al., 2016). A similar pattern was observed in these families – whilst the number of shared highest expressed orthologs in *I. lacustris* and *L. uniflora* was 18, significantly higher than expected by chance ($p = 0.014$, Figure 4.5), only a single ortholog was exclusively the most highly expressed in *I. lacustris* and *L. uniflora* (Figure 4.6) - this was the "bacterial-type" PPC, as opposed to the "plant-type" PPC in *A. comosus* and *A. thaliana*. Fewer orthologs were the most highly expressed in CAM species exclusively (four) than in all species (ten), in angiosperms (five) or between *I. lacustris*, *A. comosus* and *A. thaliana* (five, Figure 4.6).

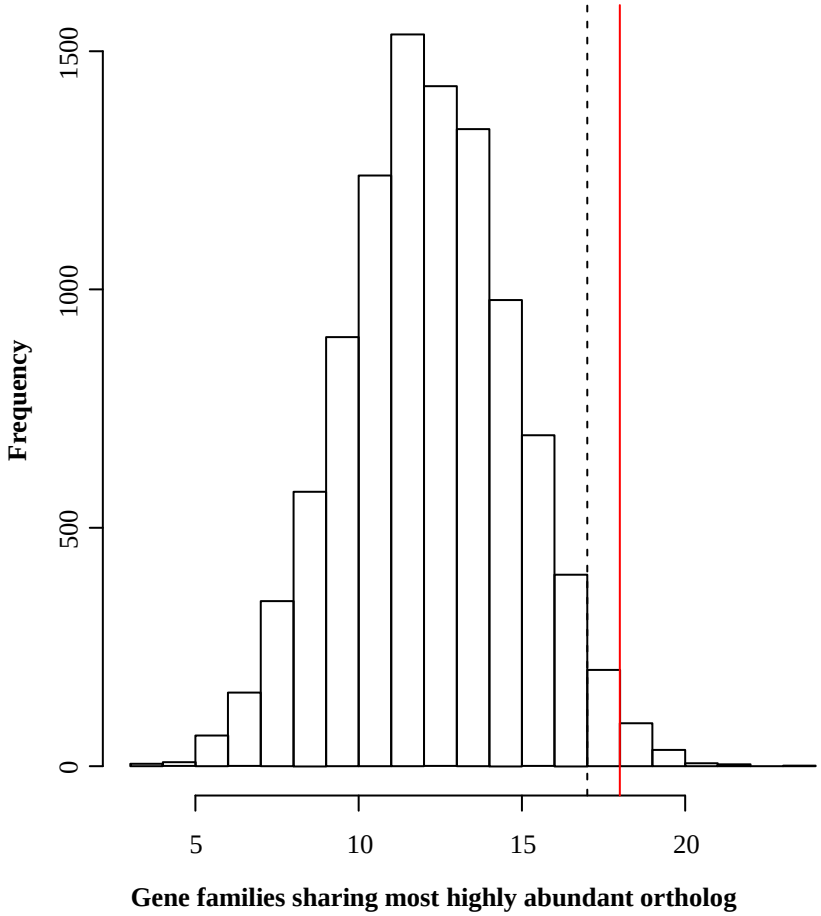


Figure 4.5: CAM gene families sharing most highly abundant ortholog in *Littorella uniflora* and *Isoetes lacustris*.

Histogram represents number of shared orthologs observed from 10,000 simulations randomly assigning each gene family a most highly expressed ortholog and summing the number of times these were the same in *L. uniflora* and *I. lacustris*. The dashed line represents the threshold of significance ($p < 0.05$). The red line represents the observed number of gene families sharing the most highly abundant ortholog from RNA-seq data.

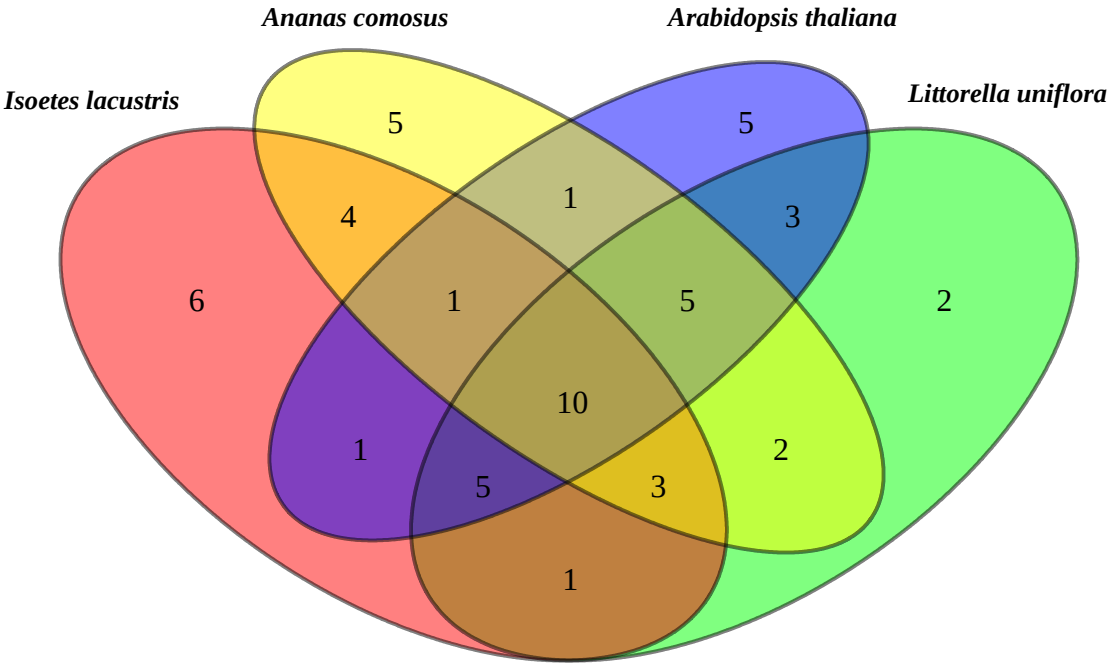


Figure 4.6 Overlap of shared most abundant orthologs within CAM gene families across land plants Venn diagram representing the numbers times 31 gene families with previously proposed roles in CAM share the same most abundant ortholog between *Isoëtes lacustris* (red oval), *Ananas comosus* (yellow oval), *Arabidopsis thaliana* (blue oval) and *Littorella uniflora* (green oval).

4.5 Discussion

CAM cycles in Isoëtes *and* Litorella

Published evidence suggests that the CAM physiology in the aquatic genera *Litorella* and *Isoëtes* is plastic, with the levels of CAM activity modulated by environmental factors (Aulio, 1985, Madsen, 1987a, Baattrup-Pedersen and Madsen, 1999, Keeley, 1998). In this study, plants were grown in a variety of conditions, some of which were expected to reduce the level of CAM activity. With the exception of the low-light submerged conditions in *Littorella*, all results indicated an accumulation of malate during the night and its use during the day (Figure 4.1, 4.2). Even in the low-light treatment, malate levels were high, so that a weak CAM cycle cannot be excluded. CAM plasticity in these accessions of *Littorella* and *Isoëtes* therefore appears to be limited, but our sampling is too small to exclude intraspecific variation in the level of plasticity.

Aquatic CAM activity has long been hypothesised to be a response to relatively low levels of CO_2 in vernal pools and oligotrophic environments (Keeley, 1981; Keeley, 1998), and it is therefore notable that low levels of atmospheric CO_2 do not appear to increase CAM activity (Figure 4.2). Despite a lack of effect on the CAM activity, low CO_2 appears to dampen the circadian rhythm, as evidenced by a reduction in the number of diurnally expressed genes (Supplementary Table 4.4). Low carbon concentrations are likely to reduce the rates of photosynthesis, which explains the reduced abundances of transcripts associated with photosynthesis in the low-CO_2 plants (Table 4.2). Previous studies have identified fluxes in photosynthetic products as drivers of circadian rhythms in *Arabidopsis* (Haydon et al., 2013), and the reduced diurnal patterns in *Isoëtes* might result from lower photosynthetic rates. It is however surprising, given the status of CAM as a circadian trait evolving in response to low CO_2 levels, to observe this pattern in *Isoëtes*.

High levels of genes encoding enzymes of the CAM pathway were observed in both species. Genes for PPC and PCK were particularly highly expressed in both *I. lacustris* and *L. uniflora* (Table 4.1, Table 4.3). Both species express high levels of PPC4, a distant paralog of the PPC genes used by all C_4 plants and terrestrial CAM species screened so far (Christin et al., 2015; Deng et al., 2016; Moreno-Villena et al., 2018). Previous studies of PPC4 have suggested it acts in a hetero-octameric complex with the “plant-type” PPCs (Gennidakis et al., 2007; O’Leary et al., 2011; Ting et al., 2017) but in *I. lacustris* the bacterial-type PPC is expressed at 20-30x the level of the “plant-type” PPC. We therefore hypothesize that the two distant isoforms do not interact closely in *Isoëtes*. *Littorella uniflora* also expresses the “bacterial-type” PPC at high levels, but the “plant-type” PPC is expressed at comparable levels, particularly during the night (Table 4.3), and the role each plays in CAM is unclear. Diurnal variability in the size of the PPC complex in *Littorella* has been previously reported (Groenhof et al., 1988), consistent with roles for both lineages of PPC. High levels of bacterial PPCs have been associated with recycling of respiratory CO_2 in angiosperm fruits that have limited gas exchange (Park et al., 2012; Ting et al., 2017), which is consistent with the important role of respiratory CO_2 recycling in *I. lacustris* and *L. uniflora* (Madsen, 1987b; Madsen et al., 2002).

PCK is the most plausible candidate for the role of CAM-specific decarboxylase in both species, despite previous reports of high NADP-ME activity in *L. uniflora* (Groenhof et al., 1988). There may be a degree of flexibility in the identity of the

decarboxylase, either by plasticity within individuals or local adaptation within populations. By contrast, the high transcript level for PCK in *I. lacustris* is consistent with high PCK activity in *I. howellii (*Keeley and Busch, 1984). Overall, the core CAM pathways of *I. lacustris* and *L. uniflora* seem to be achieved using similar enzymes, and therefore represent a case of convergent gene co-option.

Gene co-option driven by conservatism of expression levels

For the multiple gene families with paralogs shared across land plants, the same gene lineage is usually the most expressed in both *Litorella* and *Isoëtes*, which is true for CAM-related genes (Figure 4.5), but also across the whole transcriptomes (Figure 4.3). Given the large divergence time between the two species, the shared identity of the most abundant orthologs could be interpreted as convergent evolution linked to their CAM pathways and other convergent aspects of their phenotypes. However, many of the same orthologs are also the most abundant in the C_3 flowering plant *A. thaliana* and the terrestrial CAM plant *A. comosus* (Figure 4.4, 4.6). Therefore, instead of convergent evolution, the shared identity of the most expressed genes reflects conservation of abundance in leaves within gene families from at least the common ancestor of vascular plants. This occurs despite subsequent duplications and losses of gene copies within each lineage (Jiao et al., 2011). For example, the expansion of plant-type PPCs in angiosperms (Christin et al., 2015) suggests that the overall functions of lineages of genes within a family is retained independently of duplication (Panchy et al., 2016). Broad functional innovations such as the independent evolution of leaves in lycopods and euphyllophytes (Tomescu, 2009) or CAM in *I. lacustris*, *L. uniflora* and *A. comosus* do not appear to be associated with large rearrangements of relative transcription within homolog families (Figure 4.4, 4.6). Even within genes with previously identified roles in CAM (Brilhaus et al., 2016), many of the most highly abundant lineages are also the most abundant in *A. thaliana*, and few lineages have a different most highly abundant lineage exclusively in CAM plants (Figure 4.6). This pattern suggests that despite the complex evolutionary changes associated with CAM photosynthesis, ancestral expression levels still largely determine which lineages are recruited for functions in CAM within each homolog group. However, the conservatism concerns only the identity of the most highly expressed gene within each gene family. Indeed, the evolution of CAM, whether in aquatic or in terrestrial settings, is accompanied by large

expression increases of the ancestrally most highly expressed gene (Figure 4.6, Table 4.1, Table 4.3), a pattern mirroring that observed in grasses (Moreno-Villena et al., 2018).

The conclusion that expression patterns and the ensuing co-option bias is maintained across land plants at first sight contradicts the findings of Christin et al., 2015. This previous study concluded that different gene lineages were co-opted in C_4 eudicots and C_4 monocots. This apparent paradox likely stems from the differences in the number of co-orthologs defined at the angiosperm and land plant levels. Indeed, each group of land plant co-orthologs potentially encompasses multiple groups of angiosperm co-orthologs, because of more recent gene or genome duplications. The conclusion that eudicots and monocots co-opted different genes for C_4 might indicate that, while expression differences among ancient duplicates are maintained, more recent duplicates diverge in their expression phenotype. This phenomenon of neofunctionalization of duplicates is predicted since redundant genes are less likely to be retained, and the early origins of angiosperms have been accompanied by multiple rounds of genome duplications (Jiao et al., 2011), likely increased the divergence of recent paralogs among its subgroups.

Overall, our results suggest that ancestral gene expression levels are a crucial determinant of suitability for recruitment into CAM and C_4, across multiple evolutionary scales. Expression levels are broadly conserved between lineages, but subsequent gene duplications within lineages can lead to changes in relative expression levels resulting in different within-lineage biases between phylogenetic groups.

4.6 Conclusions

In this study, we investigated the gene expression patterns of two ecologically convergent but phylogenetically divergent submerged aquatic CAM plants. The CAM phenotypes of the studied species that diverged 400 million years ago are generated using similar enzymes, and in many cases orthologous genes. While this pattern is suggestive of strong convergent evolution, the expression patterns of genes in the aquatic CAM are broadly shared with taxa of other photosynthetic types, suggesting that they mainly stem from conserved expression patterns across land plants. CAM-related genes do however reach exceptional levels in CAM plants, and we suggest that the gene lineages that remained elevated during the diversification of land plants were recurrently co-opted for the CAM trait, which involved further increases in transcript abundances. The convergent evolution of some complex traits is therefore influenced by changes in the ancestral expression patterns of gene lineages that occurred hundreds millions of years in the past.

Acknowledgements
We thank Professor Colin Osborne and Professor David Beerling, who made extremely important contributions to the design of this study.

4.7 Supplementary Information

Supplementary Table 4.1: Data sources

Species	Datatype	Source
Ananas comosus	Protein coding sequences (v3)	Phytozome
Ananas comosus	Reads[1]	Ming et al., 2015
Arabidopsis thaliana	Protein coding sequences (v10)	Phyotzome
Arabidopsis thaliana	Reads[2]	Stroud et al., 2014
Volvox carteri	Protein coding sequences (v2.1)	Phytozome

1. 8am samples from green leaves used – see (Ming et al., 2015)
2. Control samples used – see (Stroud et al., 2014)

Supplementary Table 4.2: CAM-related Homologs

Arabidopsis Homolog	Homolog Group	Description
AT1G79750	1	Malic enzyme
AT2G13560	1	NAD-dependent malic enzyme 1 mitochondrial
AT2G19900	1	NADP-dependent malic enzyme 1
AT4G00570	1	NAD-dependent malic enzyme 2 mitochondrial
AT5G11670	1	NADP-dependent malic enzyme 2
AT5G25880	1	Malic enzyme
AT1G12000	2	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta 1
AT1G20950	2	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha
AT1G76550	2	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit alpha 2
AT4G04040	2	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta 2
AT1G05030	3	Probable plastidic glucose transporter 1
AT1G08890	3	Sugar transporter ERD6-like 1
AT1G08900	3	Sugar transporter ERD6-like 2
AT1G08920	3	ERD (early response to dehydration) six-like 1
AT1G08930	3	ERD6
AT1G19450	3	Sugar transporter ERD6-like 4
AT1G30220	3	Probable inositol transporter 2
AT1G54730	3	Sugar transporter ERD6-like 5
AT1G67300	3	Putative plastidic glucose transporter 2
AT1G75220	3	Sugar transporter ERD6-like 6
AT1G79820	3	Probable plastidic glucose transporter 3
AT2G16120	3	Putative polyol transporter 1
AT2G16130	3	Putative polyol transporter 2
AT2G18480	3	Probable polyol transporter 3
AT2G20780	3	Probable polyol transporter 4
AT2G35740	3	Probable inositol transporter 3
AT2G43330	3	Inositol transporter 1

AT2G48020	3 Sugar transporter ERD6-like 7
AT3G03090	3 D-xylose-proton symporter-like 1
AT3G05150	3 Major facilitator superfamily protein
AT3G05155	3 Major facilitator superfamily protein
AT3G05160	3 Sugar transporter ERD6-like 10
AT3G05165	3 Major facilitator superfamily protein
AT3G05400	3 Sugar transporter ERD6-like 12
AT3G18830	3 PMT5
AT3G20460	3 Putative sugar transporter ERD6-like 13
AT4G04750	3 Major facilitator superfamily protein
AT4G04760	3 Sugar transporter ERD6-like 15
AT4G16480	3 Inositol transporter 4
AT4G36670	3 At4g36670
AT5G16150	3 Plastidic glucose transporter 4
AT5G17010	3 D-xylose-proton symporter-like 2
AT5G18840	3 Sugar transporter ERD6-like 16
AT5G27350	3 Sugar transporter ERD6-like 17
AT5G27360	3 Major facilitator superfamily protein
AT5G59250	3 D-xylose-proton symporter-like 3 chloroplastic
AT1G03310	4 ISA2
AT2G39930	4 ISA1
AT4G09020	4 Isoamylase 3 chloroplastic
AT2G22480	5 ATP-dependent 6-phosphofructokinase 5 chloroplastic
AT4G26270	5 ATP-dependent 6-phosphofructokinase 3
AT4G29220	5 ATP-dependent 6-phosphofructokinase 1
AT4G32840	5 ATP-dependent 6-phosphofructokinase
AT5G47810	5 ATP-dependent 6-phosphofructokinase 2
AT5G56630	5 ATP-dependent 6-phosphofructokinase 7
AT5G61580	5 ATP-dependent 6-phosphofructokinase
AT1G12900	6 glyceraldehyde 3-phosphate dehydrogenase A subunit 2
AT1G13440	6 Glyceraldehyde-3-phosphate dehydrogenase GAPC2 cytosolic
AT1G16300	6 Glyceraldehyde-3-phosphate dehydrogenase GAPCP2 chloroplastic
AT1G42970	6 Glyceraldehyde-3-phosphate dehydrogenase GAPB chloroplastic
AT1G79530	6 Glyceraldehyde-3-phosphate dehydrogenase GAPCP1 chloroplastic
AT3G04120	6 Glyceraldehyde-3-phosphate dehydrogenase GAPC1 cytosolic
AT3G26650	6 Glyceraldehyde-3-phosphate dehydrogenase GAPA1 chloroplastic
AT1G10760	7 Alpha-glucan water dikinase 1 chloroplastic
AT4G24450	7 Alpha-glucan water dikinase 2
AT5G26570	7 Phosphoglucan water dikinase chloroplastic
AT1G74030	8 Enolase 1 chloroplastic
AT2G29560	8 Cytosolic enolase 3
AT2G36530	8 LOS2
AT1G78560	9 Probable sodium/metabolite cotransporter BASS1 chloroplastic
AT2G26900	9 BASS2
AT3G25410	9 Probable sodium/metabolite cotransporter BASS3 chloroplastic

AT4G12030	9	Probable sodium/metabolite cotransporter BASS5 chloroplastic
AT4G22840	9	Probable sodium/metabolite cotransporter BASS6 chloroplastic
AT1G61800	10	glucose-6-phosphate/phosphate translocator 2
AT3G01550	10	Phosphoenolpyruvate/phosphate translocator 2 chloroplastic
AT4G03950	10	Glucose-6-phosphate/phosphate-translocator-like protein 1
AT5G17630	10	Xylulose 5-phosphate/phosphate translocator chloroplastic
AT5G33320	10	Phosphoenolpyruvate/phosphate translocator 1 chloroplastic
AT5G54800	10	GPT1
AT2G32290	11	Beta-amylase 6
AT2G45880	11	Beta-amylase 7
AT3G23920	11	Beta-amylase
AT4G00490	11	Beta-amylase 2 chloroplastic
AT4G15210	11	Beta-amylase 5
AT4G17090	11	Beta-amylase 3 chloroplastic
AT5G18670	11	Inactive beta-amylase 9
AT5G45300	11	beta-amylase 2
AT5G55700	11	Inactive beta-amylase 4 chloroplastic
AT3G29320	12	Alpha-glucan phosphorylase 1
AT3G46970	12	Alpha-glucan phosphorylase 2 cytosolic
AT1G23190	13	Probable phosphoglucomutase cytoplasmic 1
AT1G70730	13	Phosphoglucomutase
AT5G51820	13	Phosphoglucomutase chloroplastic
AT2G40840	14	4-alpha-glucanotransferase DPE2
AT5G64860	14	4-alpha-glucanotransferase DPE1 chloroplastic/amyloplastic
AT1G05610	15	ADP-glucose pyrophosphorylase small subunit 2
AT1G27680	15	Glucose-1-phosphate adenylyltransferase large subunit 2 chloroplastic
AT2G21590	15	Probable glucose-1-phosphate adenylyltransferase large subunit chloroplastic
AT4G39210	15	Glucose-1-phosphate adenylyltransferase
AT5G19220	15	Glucose-1-phosphate adenylyltransferase large subunit 1 chloroplastic
AT5G48300	15	Glucose-1-phosphate adenylyltransferase small subunit chloroplastic
AT5G12860	16	Dicarboxylate transporter 1 chloroplastic
AT5G64280	16	Dicarboxylate transporter 2.2 chloroplastic
AT5G64290	16	Dicarboxylate transporter 2 chloroplastic
AT1G43670	17	Fructose-1 6-bisphosphatase cytosolic
AT3G54050	17	Fructose-1 6-bisphosphatase 1 chloroplastic
AT3G55800	17	Sedoheptulose-1 7-bisphosphatase chloroplastic
AT5G64380	17	AT5g64380/MSJ1 22
AT1G53310	18	Phosphoenolpyruvate carboxylase 1
AT1G68750	18	PPC4
AT2G42600	18	PPC2
AT3G14940	18	Phosphoenolpyruvate carboxylase 3
AT3G42628	18	Phosphoenolpyruvate carboxylase-related / PEP carboxylase-like protein
AT1G04410	19	Malate dehydrogenase
AT5G43330	19	Malate dehydrogenase 2 cytoplasmic
AT5G56720	19	Malate dehydrogenase

AT5G58330	19 Malate dehydrogenase
AT1G69830	20 Alpha-amylase 3 chloroplastic
AT1G76130	20 Probable alpha-amylase 2
AT4G25000	20 Alpha-amylase 1
AT1G11720	21 SS3
AT1G32900	21 Granule-bound starch synthase 1 chloroplastic/amyloplastic
AT3G01180	21 Starch synthase chloroplastic/amyloplastic
AT4G18240	21 Probable starch synthase 4 chloroplastic/amyloplastic
AT5G24300	21 Starch synthase chloroplastic/amyloplastic
AT5G65685	21 UDP-Glycosyltransferase superfamily protein
AT1G14140	22 Mitochondrial uncoupling protein 3
AT1G79900	22 Mitochondrial arginine transporter BAC2
AT2G22500	22 Mitochondrial uncoupling protein 5
AT2G33820	22 Mitochondrial substrate carrier family protein
AT3G54110	22 UCP1
AT4G03115	22 Mitochondrial substrate carrier family protein
AT4G24570	22 DIC2
AT5G01340	22 Mitochondrial succinate-fumarate transporter 1
AT5G09470	22 DIC3
AT5G19760	22 Mitochondrial dicarboxylate/tricarboxylate transporter DTC
AT5G46800	22 Mitochondrial carnitine/acylcarnitine carrier-like protein
AT5G58970	22 Mitochondrial uncoupling protein 2
AT1G20260	23 VAB3
AT1G76030	23 V-type proton ATPase subunit B1
AT1G78900	23 VHA-A
AT4G38510	23 V-type proton ATPase subunit B2
AT5G08670	23 ATP synthase alpha/beta family protein
AT5G08680	23 ATP synthase subunit beta-3 mitochondrial
AT5G08690	23 ATP synthase subunit beta-2 mitochondrial
ATCG00480	23 ATP synthase subunit beta chloroplastic
AT2G21170	23 TIM
AT3G55440	24 Triosephosphate isomerase cytosolic
AT1G53240	24 Malate dehydrogenase 1 mitochondrial
AT2G22780	25 Malate dehydrogenase
AT3G15020	25 Malate dehydrogenase 2 mitochondrial
AT3G47520	25 Malate dehydrogenase chloroplastic
AT3G53910	25 Malate dehydrogenase-like protein
AT4G17260	25 L-lactate dehydrogenase
AT5G09660	25 Malate dehydrogenase
AT2G36390	26 SBE2
AT3G20440	26 1 4-alpha-glucan-branching enzyme 3 chloroplastic/amyloplastic
AT5G03650	26 1 4-alpha-glucan-branching enzyme 2-2 chloroplastic/amyloplastic
AT3G01510	27 LSF1
AT3G10940	27 Phosphoglucan phosphatase LSF2 chloroplastic
AT3G52180	27 SEX4

AT1G15690	28	VHP1
AT1G16780	28	Pyrophosphate-energized membrane proton pump 3
AT1G78920	28	Pyrophosphate-energized membrane proton pump 2
AT2G01140	29	Fructose-bisphosphate aldolase 3 chloroplastic
AT2G21330	29	Fructose-bisphosphate aldolase 1 chloroplastic
AT2G36460	29	Fructose-bisphosphate aldolase 6 cytosolic
AT3G52930	29	Fructose-bisphosphate aldolase
AT4G26520	29	Aldolase superfamily protein
AT4G26530	29	Fructose-bisphosphate aldolase
AT4G38970	29	Fructose-bisphosphate aldolase
AT5G03690	29	Fructose-bisphosphate aldolase 4 cytosolic
AT1G19580	30	GAMMA CA1
AT1G47260	30	GAMMA CA2
AT3G48680	30	Gamma carbonic anhydrase-like 2 mitochondrial
AT5G63510	30	Gamma carbonic anhydrase like 1
AT5G66510	30	gamma carbonic anhydrase 3
AT4G24620	31	Glucose-6-phosphate isomerase
AT5G42740	31	Glucose-6-phosphate isomerase

Supplementary Table 4.3: Diurnally expressed genes in *Isoëtes lacustris*

Available at ORDA - The University of Sheffield Research Data Catalogue and Repository.

Supplementary Table 4.4: Diurnally expressed genes in *Isoëtes lacustris* in high and low CO_2 conditions

Arabidopsis thaliana co-ortholog	Ambient log2(Fold Change)[1]	Ambient p-value	Low log2(Fold Change)[1]	Low p-value	Description
AT3G55580	4	6.20E-28	3.3	3.90E-20	Regulator of chromosome condensation (RCC1) family protein
AT2G42670	2.7	4.70E-19	1.4	3.30E-06	Protein of unknown function (DUF1637)
AT2G25620	2.5	5.60E-18	2.7	4.60E-19	Probable protein phosphatase 2C 22
AT4G12710	2.4	7.40E-15	1.5	1.40E-06	ARM repeat protein
AT3G25640	2	1.20E-13	2.3	1.60E-15	Protein of unknown function, DUF617
AT1G13640	-1.9	8.70E-11	-1.3	1.30E-05	Phosphatidylinositol 4-kinase gamma 6
AT4G27970	1.7	1.20E-10	1.1	3.30E-05	S-type anion channel SLAH2
AT1G54520	1.9	2.00E-10	2.1	4.50E-12	Putative uncharacterized protein
AT2G31840	1.9	7.10E-10	0.44	0.17	Thioredoxin-like fold domain-containing protein MRL7L, chloroplastic
AT1G12000	2	1.10E-09	0.91	0.0049	Pyrophosphate--fructose 6-phosphate 1-phosphotransferase subunit beta 1
AT2G24100	1.5	1.50E-09	1.1	1.50E-05	At2g24100
AT1G07180	2.2	1.90E-09	0.87	0.018	Internal alternative NAD(P)H-ubiquinone oxidoreductase A1, mitochondrial
AT3G17800	2.8	6.00E-09	3.4	1.10E-11	Protein of unknown function (DUF760)
AT2G32120	1.6	6.30E-09	0.97	0.00061	Heat shock 70 kDa protein 8

AT4G20070	1.8	7.60E-09	0.83	0.0057	Allantoate deiminase
AT5G02810	1.8	1.20E-08	1.7	7.80E-08	Two-component response regulator-like APRR7
AT1G07280	-1.8	1.50E-08	-1.4	6.90E-06	At1g07280/F22G5 32
AT2G39130	1.3	7.60E-08	0.81	0.00086	Amino acid transporter AVT1C
AT3G10690	-1.6	9.20E-08	-1.3	2.10E-05	DNA gyrase subunit A, chloroplastic/mitochondrial
AT1G60600	-1.4	1.00E-07	-1.1	8.50E-05	2-carboxy-1,4-naphthoquinone phytyltransferase, chloroplastic
AT3G17630	1.4	1.20E-07	0.58	0.03	Cation/H(+) antiporter 19
AT2G44130	1.4	2.80E-07	0.88	0.00079	F-box/kelch-repeat protein At2g44130
AT3G47520	1.4	3.70E-07	0.29	0.29	Malate dehydrogenase, chloroplastic
AT5G18640	1.4	4.40E-07	0.71	0.01	alpha/beta-Hydrolases superfamily protein
AT2G41250	-1.4	5.30E-07	-0.33	0.24	Haloacid dehalogenase-like hydrolase domain-containing protein
AT3G05170	1.3	7.20E-07	0.8	0.0025	Phosphoglycerate mutase-like protein AT74
AT4G37790	1.5	1.40E-06	0.077	0.8	Homeobox-leucine zipper protein HAT22
AT1G71090	1.3	1.50E-06	0.52	0.053	Protein PIN-LIKES 2
AT3G12120	-1.5	2.50E-06	-1.6	2.00E-07	FAD2
AT4G08330	1.9	2.60E-06	1.1	0.0045	Uncharacterized protein At4g08330, chloroplastic
AT2G27050	1.4	4.00E-06	1.2	1.00E-04	ETHYLENE INSENSITIVE 3-like 1 protein
AT4G38690	1.6	4.70E-06	1.1	0.0013	1-phosphatidylinositol phosphodiesterase-related protein
AT3G55640	1.2	5.00E-06	0.92	0.00031	Ca-dependent solute carrier-like protein
AT3G22840	2.1	5.80E-06	3.1	1.60E-10	ELIP1
AT2G44740	-2.2	6.50E-06	-0.35	0.46	Cyclin-U4-1
AT1G34540	1.6	6.60E-06	1.6	4.90E-06	CYP94D1
AT4G24670	1.6	8.50E-06	0.67	0.063	Tryptophan aminotransferase-related protein 2
AT1G16300	1.3	8.60E-06	1	0.00081	Glyceraldehyde-3-phosphate dehydrogenase GAPCP2, chloroplastic
AT5G19530	2.2	8.70E-06	0.55	0.26	Thermospermine synthase ACAULIS5
AT1G05850	-3.5	1.00E-05	-0.5	0.49	Chitinase-like protein 1
AT4G37680	1.1	1.30E-05	0.45	0.11	heptahelical protein 4
AT1G22770	1.3	1.30E-05	0.56	0.061	Protein GIGANTEA
AT5G50180	1.4	1.40E-05	1.2	0.00051	At5g50180
AT3G02580	1.2	1.50E-05	0.36	0.19	Delta(7)-sterol-C5(6)-desaturase 1
AT1G68660	1.4	1.70E-05	0.87	0.0054	ATP-dependent Clp protease adapter protein CLPS1, chloroplastic
AT1G18335	1.5	2.00E-05	0.11	0.78	Acyl-CoA N-acyltransferases (NAT) superfamily protein
AT3G24160	1.4	2.00E-05	0.75	0.02	PMP
AT2G39740	-1.3	2.30E-05	-1.1	0.00052	Protein HESO1
AT2G31400	-1.5	2.50E-05	-0.76	0.035	Pentatricopeptide repeat-containing protein At2g31400, chloroplastic
AT5G43830	1.1	2.80E-05	0.89	0.00065	Aluminium induced protein with YGL and LRDR motifs
AT3G49050	1.3	3.30E-05	0.73	0.018	Alpha/beta-Hydrolases superfamily protein
AT5G54470	1.6	3.50E-05	2.2	1.80E-08	BBX29
AT1G11530	1.6	3.50E-05	0.63	0.1	Thioredoxin-like protein CXXS1
AT2G32500	-1.5	3.80E-05	-1.3	0.00035	At2g32500
AT3G57680	1.2	4.00E-05	1.4	4.70E-06	Carboxyl-terminal-processing peptidase

					3, chloroplastic
AT3G07090	1	7.00E-05	0.52	0.048	PPPDE putative thiol peptidase family protein
AT3G61320	1.2	7.10E-05	0.62	0.044	Bestrophin-like protein
AT5G22510	-2	8.30E-05	-0.91	0.065	Alkaline/neutral invertase E, chloroplastic
AT1G57680	-1.4	8.40E-05	-0.44	0.21	Putative uncharacterized protein At1g57680
AT5G42740	1.3	9.40E-05	0.8	0.016	Glucose-6-phosphate isomerase
AT1G44542	1.1	1.00E-04	0.31	0.27	At1g44542
AT2G39210	1	0.00012	0.83	0.0025	At2g39210/T16B245
AT4G16600	0.93	0.00012	0.57	0.023	Glycosyltransferase
AT1G78230	1.5	0.00012	1.2	0.0038	Outer arm dynein light chain 1 protein
AT2G13650	1.1	0.00014	0.47	0.11	GDP-mannose transporter
AT5G67480	-1.7	0.00015	-0.97	0.026	BTB and TAZ domain protein 4
AT1G55370	-1.1	0.00016	-0.74	0.015	NDF5
AT2G40980	0.86	0.00018	0.93	0.00012	Protein kinase family protein
AT5G07200	1.4	2.00E-04	0.61	0.12	Gibberellin 20 oxidase 3
AT1G18900	-0.99	0.00021	-0.34	0.23	Pentatricopeptide repeat-containing protein
AT3G62830	0.98	0.00022	0.12	0.65	UDP-glucuronic acid decarboxylase 2
AT5G36890	0.95	0.00024	0.42	0.11	Beta-glucosidase 42
AT5G17640	1.1	0.00026	1.1	0.00048	At5g17640
AT1G42540	-0.96	0.00026	-0.52	0.059	Glutamate receptor 3.3
AT5G61250	-1.8	0.00027	-0.68	0.18	GUS1
AT1G55020	-2.8	0.00027	-1.6	0.034	Linoleate 9S-lipoxygenase 1
AT1G22040	-0.91	0.00028	-1.1	2.60E-05	F-box/kelch-repeat protein At1g22040
AT5G63190	1.1	0.00029	0.97	0.0012	MA3 domain-containing protein
AT5G65690	1.1	0.00029	0.47	0.12	phosphoenolpyruvate carboxykinase 2
AT5G01410	1.1	0.00029	0.71	0.022	RSR4
AT3G47160	0.94	3.00E-04	0.84	0.0014	RING/U-box superfamily protein
AT3G15140	1.1	0.00032	0.39	0.19	Uncharacterized exonuclease domain-containing protein At3g15140
AT5G28910	-1.7	0.00033	-1.4	0.0046	At5g28910
AT2G25610	1.4	0.00033	0.92	0.018	ATPase, F0/V0 complex, subunit C protein
AT5G20950	-1.1	0.00033	-0.1	0.72	Glycosyl hydrolase family protein
AT1G21410	1.1	0.00036	0.72	0.018	F-box protein SKP2A
AT3G19970	-0.99	4.00E-04	-0.87	0.0037	AT3g19970/MZE19 2
AT4G11010	-3	4.00E-04	-1.5	0.063	Nucleoside diphosphate kinase
AT4G35750	1.2	0.00043	0.4	0.25	At4g35750
AT4G34710	1	0.00044	0.93	0.002	Arginine decarboxylase 2
AT4G35080	-0.85	0.00044	-1.1	9.80E-06	High-affinity nickel-transport family protein
AT4G09620	0.96	0.00045	0.35	0.21	Mitochondrial transcription termination factor family protein
AT5G61840	0.89	0.00046	0.5	0.058	GUT1
AT1G32730	0.79	0.00051	0.17	0.48	F6N181
AT5G06720	-3.5	0.00055	-1.2	0.18	Peroxidase 53
AT3G24480	-2.2	6.00E-04	0.66	0.29	Leucine-rich repeat extensin-like protein 4
AT5G49800	1.2	0.00061	0.49	0.2	At5g49800
AT3G52950	1	0.00061	1.1	0.00036	CBS domain-containing protein CBSCBSPB3
AT2G26690	1.9	0.00061	0.19	0.74	Protein NRT1/ PTR FAMILY 6.2
AT2G15890	1.2	0.00062	0.88	0.0095	CCG-binding protein 1

AT5G63850	2.1	0.00063	0.1	0.86	AAP4
AT1G18350	-1.1	0.00065	-0.71	0.038	MKK7
AT3G27460	0.91	0.00065	0.36	0.18	SGF29 tudor-like domain-containing protein
AT1G70410	1.1	0.00066	0.4	0.23	Beta carbonic anhydrase 4
AT3G15790	0.86	0.00067	0.48	0.06	Methyl-CpG-binding domain-containing protein 11
AT5G42650	-2.6	0.00071	-0.5	0.49	Allene oxide synthase, chloroplastic
AT5G67030	1.2	0.00074	1.3	0.00015	Zeaxanthin epoxidase, chloroplastic
AT2G26770	0.86	0.00077	0.14	0.6	Stomatal closure-related actin-binding protein 1
AT2G01830	1.2	0.00079	1.4	9.90E-05	Histidine kinase 4
AT5G01990	0.9	8.00E-04	0.17	0.53	Protein PIN-LIKES 6
AT2G25870	0.82	0.00082	0.6	0.015	At2g25870
AT3G08600	-1.2	0.00083	-0.38	0.34	AT3g08600/F17O14 7
AT3G07400	1.1	0.00083	0.6	0.06	F21O31 protein
AT4G37470	1.2	0.00083	1	0.0065	KAI2
AT3G51670	1	0.00084	0.58	0.056	Patellin-6
AT5G60580	1.3	0.00088	0.17	0.67	RING/U-box domain-containing protein
AT5G49720	-1.9	0.00092	-0.72	0.2	Endoglucanase 25
AT1G10510	-1.1	0.00093	-0.21	0.55	Emb2004
AT4G08900	0.88	0.00096	0.73	0.0072	Arginase 1, mitochondrial
AT4G17940	1	0.00098	0.63	0.039	Putative uncharacterized protein
AT4G17650	0.89	0.001	0.45	0.12	Polyketide cyclase / dehydrase and lipid transport protein
AT1G13990	0.85	0.001	0.4	0.15	
AT4G32410	-2	0.0011	-0.23	0.71	Cellulose synthase
AT1G78660	0.87	0.0011	0.37	0.17	Gamma-glutamyl hydrolase 1
AT3G63240	0.92	0.0011	0.86	0.0028	Type I inositol polyphosphate 5-phosphatase 4
AT5G20220	0.89	0.0011	1.3	3.80E-06	zinc knuckle (CCHC-type) family protein
AT5G58110	1.1	0.0012	0.63	0.058	AT5g58110/k21l19 90
AT5G64090	-0.78	0.0012	-0.35	0.15	At5g64090
AT1G52890	-1	0.0012	-0.96	0.0024	NAC domain-containing protein 19
AT5G67070	-1.4	0.0012	-0.77	0.081	Protein RALF-like 34
AT1G15310	0.71	0.0013	0.28	0.24	Putative signal recognition particle 54 kDa subunit
AT5G13650	1.1	0.0013	1.2	0.00034	Putative TypA-like translation elongation factor SVR3
AT3G18060	0.89	0.0016	1.1	0.00017	Actin-interacting protein 1-2
AT2G03390	0.84	0.0026	1.1	0.00011	Clp protease adapter protein ClpF, chloroplastic
AT1G27461	0.87	0.0029	1.2	9.50E-05	Putative uncharacterized protein
AT1G70780	0.87	0.003	1.4	2.90E-06	At1g70780
AT1G01500	0.76	0.0043	1.2	2.00E-05	Uncharacterized protein At1g01500
AT5G65010	0.85	0.0053	1.3	2.90E-05	asparagine synthetase 2
AT3G54500	0.72	0.0061	1.8	5.70E-11	
AT5G10150	0.92	0.0098	1.6	9.50E-05	Protein UPSTREAM OF FLC
AT5G15880	-0.81	0.013	-1.2	0.00048	At5g15880
AT1G49630	0.73	0.03	1.3	0.00021	Presequence protease 2, chloroplastic/mitochondrial
AT5G64670	-0.56	0.035	-1	0.00018	Putative uncharacterized protein
AT2G23290	-0.6	0.043	-1.4	6.70E-06	At2g23280
AT2G24270	0.47	0.063	0.97	0.00014	aldehyde dehydrogenase 11A3
AT4G26850	0.4	0.18	1.1	0.00019	GDP-L-galactose phosphorylase 1

AT1G62750	0.41	0.26	1.3	0.00055	Elongation factor G, chloroplastic
AT5G37055	-0.31	0.26	-1.2	1.00E-04	SWR1 complex subunit 6
AT1G54115	0.26	0.29	1.4	3.60E-07	Cation/calcium exchanger 4
AT1G73820	-0.26	0.32	-1	0.00018	Ssu72-like family protein
AT1G03770	-0.25	0.34	-1.1	0.00011	Putative E3 ubiquitin-protein ligase RING1b
AT4G27390	0.11	0.66	0.95	0.00026	AT4g27390/M4I22 200
AT2G32950	0.12	0.68	1.2	2.50E-05	COP1
AT2G28840	-0.066	0.86	-1.6	5.90E-05	Putative E3 ubiquitin-protein ligase XBAT31
AT5G35170	0.035	0.91	1.2	0.00022	Adenylate kinase 5, chloroplastic

1. Positive values represent genes more highly expressed during the day.

Supplementary Table 4.5: Diurnally expressed genes in *Littorella uniflora*

Available at ORDA - The University of Sheffield Research Data Catalogue and Repository.

Chapter V: Contrasting phylogeographic structures between freshwater lycopods and angiosperms in the British Isles

Daniel P. Wood[*1], Jill K. Olofsson[*,1], Scott W. McKenzie[2], Luke T. Dunning[1]

[1]Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

[2] Ecus Ltd., Brook Holt, 3 Blackburn Road, Sheffield, S61 2DW.

[*]Authors contributed equally to this work

Personal Contribution: I co-designed the study (with the help of Luke Dunning), collected some samples, performed part of the lab work (with the help of Jill Olofsson), and interpreted the results and wrote the paper. The analyses were performed by Jill Olofsson and the other samples were collected by Scott McKenzie and Luke Dunning. All the authors helped with the writing.

5.1 Abstract

Aquatic plants face many novel challenges compared to their terrestrial counterparts. The habitat they occupy is typically highly fragmented, with isolated water bodies surrounded by swathes of "dry desert". This can result in reduced gene flow, inbreeding, and potentially local extinction. The level of gene flow and degree of genetic structure in these species is also likely to be influenced by the mating system they adopt. To test this hypothesis we compare the phylogeographic structure of two freshwater plants in the British Isles, the largely clonal angiosperm *Littorella uniflora*, and the heterosporous lycopod *Isoëtes lacustris*. We sampled both plants from geographically spread lakes where they co-occur, and used restriction site-associated DNA sequencing (RAD-Seq) to infer their relationships. Genetic structure among lakes is higher in the angiosperm, which we associate with reduced sexual reproduction, and hence lower levels of gene flow between lakes. Further, we found evidence of lineage-specific association to certain lake nutrient type in *L. uniflora*, which might result from environmental filtering of specific ecotypes. Overall, we conclude that the reproductive system of lycopods, which is less specialized to terrestrial conditions, provides an advantage following the secondary colonisation of aquatic habitats by enabling frequent genetic exchanges between populations and potentially allow faster adaptation.

5.2 Introduction

The transition from aquatic to terrestrial environments has happened multiple times in both animals and plants (Vermeij and Dudley, 2000). This is typically accompanied by multiple challenges related to survival and reproduction (Li, 2014). In plants, the ancestral mode of reproduction is inherently linked to the presence of water (Renzaglia et al., 2000), and the adaptation to dry conditions once plants became terrestrial required increasing degrees of specialization of the reproductive system (Banks, 2009; Linkies et al., 2010; Niklas and Kutschera, 2010; Qiu et al., 2012). Several lineages subsequently made the transition back to aquatic environments, which is likely to disproportionately affect their dispersal abilities depending on their reproductive strategy.

In basal groups of land plants such as mosses, ferns, and lycophytes, male gametes are flagellated and dessication intolerant, with sexual reproduction often requiring damp habitats even in terrestrial environments (Banks, 2009). Secondarily aquatic species of these groups are therefore able to reproduce sexually underwater (Rury, 1978; Nagalingum et al., 2006; Hutsemékers et al., 2013). By contrast, submerged flowering plants (angiosperms) share the mating systems of their terrestrial ancestors, and generally only sexually reproduce above the water using flowers (Cox, 1988; Laushman, 1993), although sexual reproduction underwater has evolved in some taxa (Philbrick, 1988). The type of dispersal propagules will further affect dispersal in aquatic environments. Water-borne propagules will be efficient for dispersal within the aquatic environments, but the production of dry-resistant dispersal units, such as fruits and seeds, may facilitate the dispersal across the "dry desert" between isolated aquatic habitats (Li, 2014).

Gene flow between populations is determined by the dispersal ability and mating systems affect the genetic structure of populations, which will in turn impact their adaptive potential and resilience to environmental change (Loveless and Hamrick, 1984). While population size and their spatial distribution will also influence the intraspecific genetic structure, in plants, the reproductive system is arguably the most important factor (Loveless and Hamrick 1984; Holsinger, 2000). This has important evolutionary consequences (Morjan and Rieseberg, 2004; Eckert et al., 2010; Schiffers et al., 2014; Barrett and Harder, 2017), particularly in highly fragmented habitats (Young et al., 1996; Aguilar et al., 2006). Habitat fragmentation is especially likely for plants from freshwater habitats, such as rivers and lakes. These environments are

ephemeral in evolutionary time, and not necessarily directly connected to other suitable habitats, leading to high risks of local extinction, small effective population sizes, and inbreeding depression (Barrett et al., 1993). Despite these limitations, plants are ubiquitous in freshwater environments and indeed have very large species ranges compared to terrestrial plants, a paradox that has long fascinated biologists (Darwin, 1859; Barrett et al., 1993). Solving this paradox requires estimating effective dispersal rates and gene flow using population genetics approaches. A number of studies have inferred the genetic structure of angiosperms and more basal groups of plants (e.g. (Lokker et al., 1994; Dong et al., 2007; Hutsemekers et al., 2010; Korpelainen et al., 2013; Zhu et al., 2015; Hofstra and de Winton, 2016; Martínez-Garrido et al., 2017). However, genetic structure has never been directly compared between angiosperms and basal vascular plants colonizing the same freshwater environments.

Basal land plants and angiosperms co-occur within lakes with nutrient regimes that range from oligotrophic to mesotrophic. These environments are typically highly fragmented. In particular, the lycopod *Isoëtes lacustris* and the angiosperm *Littorella uniflora* co-occur in lakes across Northern Europe (Murphy, 2002). Despite 400 million years of independent evolution (Kenrick and Crane, 1997), these two species exhibit convergent ecological and phenotypic traits. Both species have independently adapted to the carbon-depleted aquatic environments via a relatively slow growth rate, evergreen leaves, isoetid growth form, internal lacunae allowing access to sediment CO_2, and Crassulacean Acid Metabolism (CAM; Keeley, 1981; Richardson et al., 1984; Boston, 1986; Keeley, 1998; Madsen et al., 2002). While their distribution, ecology and vegetative types are convergent, these two species retain divergent reproductive systems corresponding to their taxonomic groups. Submerged *L. uniflora* propagates asexually, by producing short stolons (Robe and Griffiths, 1998), although the buoyancy and longevity of floating whole plants (Spierenburg et al., 2013) may also allow asexual dispersal over short distances within lakes. Flowering, and therefore sexual reproduction, can only occur when water levels decrease during the summer, exposing plants near the shores to the air (Robe and Griffiths, 1998). Rates of outcrossing are unknown in *L. uniflora*, although Tessene, (1968) found possible evidence of self-incompatibility in the closely related *L. americana*. Because emersion might be limited to some populations and some years (Hoggard et al., 2003), genetic exchanges might be limited in *L. uniflora*. Seed dispersal might however occur over long distances, with long distance dispersal by birds considered the most likely mechanism (Thorne, 1972;

Hoggard et al., 2003). Little is known about how these traits influence the population genetic structure of *L. uniflora* (Hoggard et al., 2003).

The reproduction of *I. lacustris* occurs via the fusion of micro- and mega-spores. Because spores disperse in the water (Vöge, 2006), genetic exchanges are possible between submerged plants, although rates of outcrossing versus selfing in these populations are unknown. In contrast to flowering plants, little is known about the between-lake dispersal mechanism of heterosporous lycopopds (Larsén and Rydin, 2015; Troia et al., 2016), with water fowl- and wind-mediated dispersal being the most prominent suggestions (Brunton, 2001; Hoot et al., 2006; Troia, 2016). However, long distance dispersal in this species may still be challenging, as drying spores of the two closely related species *I. lacustris* and *I. echinospora* resulted in failure to germinate (Kott and Britton, 1982). Whilst a number of studies of *Isoëtes* species suggest some geographic structure, many of these are based on endangered species that have suffered population decline, and the age of the populations are unknown (Jin-ming et al., 2005; Kim et al., 2009; Hofstra and de Winton, 2016).

In this study, we contrast the intraspecific structure of *L. uniflora* and *I. lacustris* in Britain. Ice sheets covered most of Northern Europe, including Britain, until about 12,000 years ago, after which point these geographic areas were subsequently recolonised from refugia (Cottrell et al., 2002; Hoarau et al., 2007). Both studied species were present in refugia in Ireland prior to recolonisation, and are recorded arriving at similar times in paleolakes throughout Europe (Godwin, 1984; Birks, 2000). As a result, populations of *I. lacustris* and *L. uniflora* in Britain are highly similar in ecology and demographic history, and therefore represent an excellent system in which to understand the effects of their contrasting reproductive systems on population genetic structure, and its implications for adaptive evolution in these species. Using restriction site-associated DNA sequencing (RAD-Seq) of population samples spread from Snowdonia in Wales, to Aberdeenshire in Scotland and the Outer and Inner Hebrides of the Scottish Isles, we (i) infer the intraspecific genetic structure for each species, (ii) test for elevated differentiation in *L. uniflora* resulting from limited opportunities for sexual reproduction and (iii) test for genetic differentiation among nutrient types of lakes. Overall, this first parallel phylogeographic investigation of a freshwater lycopod and an angiosperm sheds new light onto the effect of sexual reproductive strategies on the build-up of the genetic structure, coupled to habitat specialization.

5.3 Material and Methods

Plant material and sequencing

Samples of *Isoëtes* sp. and *Littorella uniflora* were collected from the Scottish mainland and the Outer and Inner Hebrides in August-September 2016, dried and stored in silica gel. In addition, individual samples of *I. lacustris* and *L. uniflora* were collected in 2016 from Cwm Idwal in Snowdonia, Wales (Figs 5.1 and 5.5 and Supplementary Table 5.1). Lake type was classified according to the Scottish Natural Heritage standing water database and the scheme of Duigan et al., (2007).

DNA was extracted from silica-dried leaf material using the Qiagen PlantMini Extraction kit following the manufacturer's protocol, with the exception of the elution step, which was performed once with 50 µl AE buffer. Double digested restriction associated DNA (ddRAD) libraries were built following the protocol of Soria-Carrasco et al., 2014, using a modified common indexed adaptor to allow for paired-end sequencing (Peterson et al., 2012). In short, DNA extract (approximately 200-700 ng) was double-digested with *Eco*RI and *Mse*I. Barcoded adaptors were ligated to the *Eco*RI side and a common adaptor was ligated to the *Mse*I side. Following ligation, libraries were PCR amplified using standard Illumina sequencing primers. A total of 96 samples from the same and different projects were pooled based on relative estimates of library concentrations. The library pool was size selected by gel extraction, with a target size of 300-600 bp, and purified using the Qiagen QIAQuick Gel Extraction Kit (Qiagen). Paired-end sequencing (125 bp) was performed on one HiSeq2500 lane at the Edinburgh Genome Centre following standard protocols.

Raw sequencing data were cleaned using the trimmomatic tool kit (Bolger et al., 2014), removing adaptor and primer sequences with the ILLUMINACLIP option in palindrome mode. The expected primer and adaptor sequences where supplied to the program and a maximum of two mismatches were allowed. The cleaned reads were further trimmed by removing low quality bases, removing bases with $q < 3$ from both the 5' and 3' ends. Furthermore, bases with a quality score below 15 in a four base sliding window were also removed. Only reads longer than 36 bp after trimming were kept for downstream analyses. The cleaned reads were de-multiplexed and barcodes were removed using the processRADtag.pl script from the program STACKS (Catchen et al., 2013).

Assembly and analyses of chloroplast genomes

Cleaned and trimmed reads were mapped onto previously assembed plastomes of *I. lacustris* and *L. uniflora* collected from Llyn Idwal, Wales (Chapter III), using bowtie2 v.2.2.3 (Langmead and Salzberg, 2012) with default settings for paired end reads. Base calls for each plastid genomic position were extracted using in-house developed shell-scripts (Olofsson et al., 2016) and maximum likelihood phylogenies were inferred in RAxML v8.2.11 (Stamatakis, 2014) under a GTR+G substitution model. Node support was evaluated with 100 bootstrap replicates.

Identification and analyses of nuclear polymorphisms

RAD loci were *de novo* assembled using the program ipyrad v.0.7.2 (Eaton, 2014), with default parameters for clustering and assembly. To avoid incorporation of plastid and mitochondrial loci in the final assembly, only clusters with coverage below 100x were processed. The maximum number of alleles per single nucleotide polymorphism (SNP) was set to two and only loci present in at least 40% of samples were incorporated in the final assembly. All samples from each genus were used for two separate clusterings. For *Isoëtes* a second assembly was performed using only the samples of the species *I. lacustris* (see Results).

A random single nucleotide polymorphism (SNP) with less than 60% missing data was extracted from each assembled RAD locus using vcftools v. 0.1.15 (Danecek et al., 2011). The resulting SNP dataset was used for phylogenetic and genetic structure analyses. A maximum likelihood phylogeny was inferred for each genus in RAxML under a GTR+G substitution model and node support was evaluated with 100 bootstrap pseudo-replicates. Principal component analyses (PCA) were performed in the R package adegenet (Jombart, 2008) using the dudi.pca function. Pairwise F_{ST} between different geographic regions and lake types and homozygosity were calculated in vcftools v. 0.1.15.

5.4 Results

Genetic structure within L. uniflora

A mean of 83% of the chloroplast genome of *L. uniflora* was covered by the filtered reads (Supplementary Table 5.2). The inferred plastid phylogeny was overall poorly resolved, with low support values (Figure 5.2). Interestingly, some geographically distant populations were grouped together (e.g. samples 20 and 45 or 19 and 30), while geographically close populations were placed in different parts of the tree (e.g. sample 41 and 45 or 36 and 37). Overall, a high diversity was observed, including within the single lake from Wales (sample A, B and C).

The number of cleaned reads per sample varied from 810,000 to 2.6 million, probably reflecting variation in the quality and quantity of input DNA and libraries (Supplementary Table 5.3). A total of 128,359 RAD loci were assembled for *L. uniflora*. After filtering, 14,669 of these with 1.7% polymorphic sites were retained for analyses. The level of homozygosity was moderately high (average F = 0.55, Supplementary Table 5.3).

The first two principal components (PC) in *L. uniflora* explained 16.7% and 12.4% of the variation in the data, respectively (Figure 5.3A). The first PC separated a distinct pair of two samples (30 and 19) from all others, mirroring the chloroplast phylogeny (Figure 5.2). The remaining samples formed three groups on the second PC, one of which corresponded to the Welsh samples (A, B and C), while the two others represent different types of lake independently of geography (41 and 45 from mesotrophic lakes, and 20, 37, 38 and 39 from oligotrophic lakes – Figure 5.1, Figure 5.3A). This pattern was broadly recapitulated in the maximum likelihood nuclear phylogeny, which placed the two distinct samples (30 and 19) as identified in the PCA as monophyletic and sister to all other samples (Figure 5.4). Among the remaining samples, the monophyly of the mesotrophic and oligotrophic groups was strongly supported (97 and 82; Figure 5.4). However, some important incongruences are observed between the chloroplast and nuclear phylogenies, such as a lack of clustering by lake type in the chloroplast phylogenies (Figure 5.2, Figure 5.4). Pairwise F_{ST} values (Table 5.1) show a moderate differentiation based on geographic origin with values ranging from 0.14-0.22 between populations from different region. However, pairwise F_{ST} among phylogenetic groups mostly confirms the genetic structure we observe.

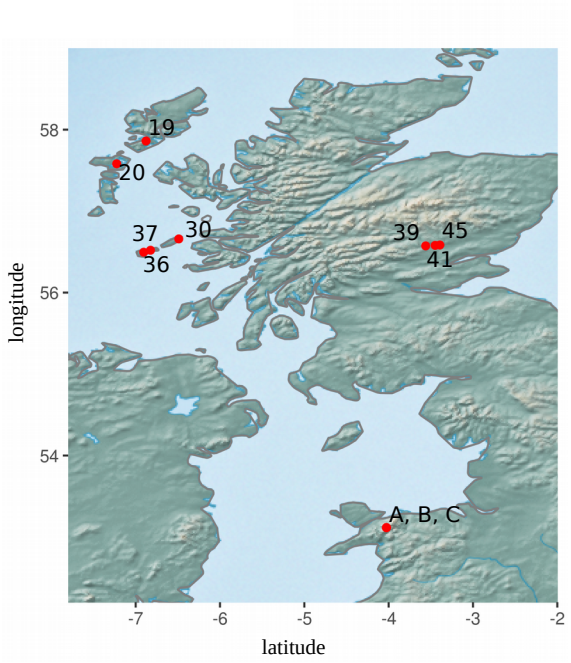


Figure 5.1: *L. uniflora* sampling locations
Locations of *Littorella uniflora* samples collected from Cwm Idwal in Wales (W – A,B, C), Aberdeenshire in the Scottish Mainland (SM – 39, 41, 35), Coll and Tirree (South Hebrides – SH; 30,36,37) and Uist, Harris and Lewis (North Hebrides – NH; 20, 19).

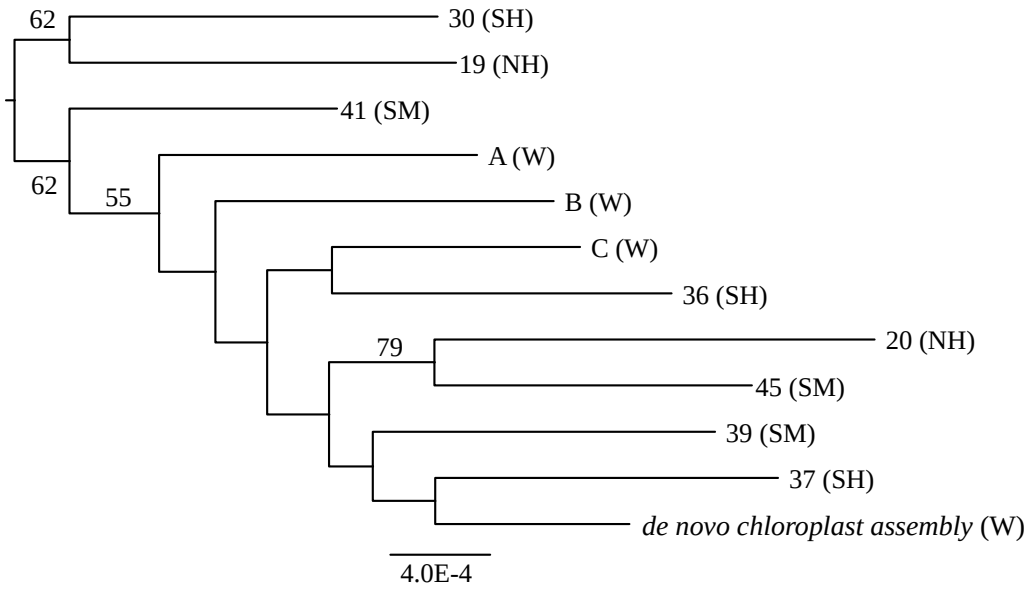


Figure 5.2: Maximum likelihood phylogeny of *Littorella uniflora* chloroplasts

Maxmimum likelihoods based on whole chloroplast alignments. Branch labels represent bootstrap supports out of 100. Scale bar represents substitutions per site. Bootstrap supports less than 50 are not shown. Tip labels represent samples (with location in brackets, see Figure 5.1).

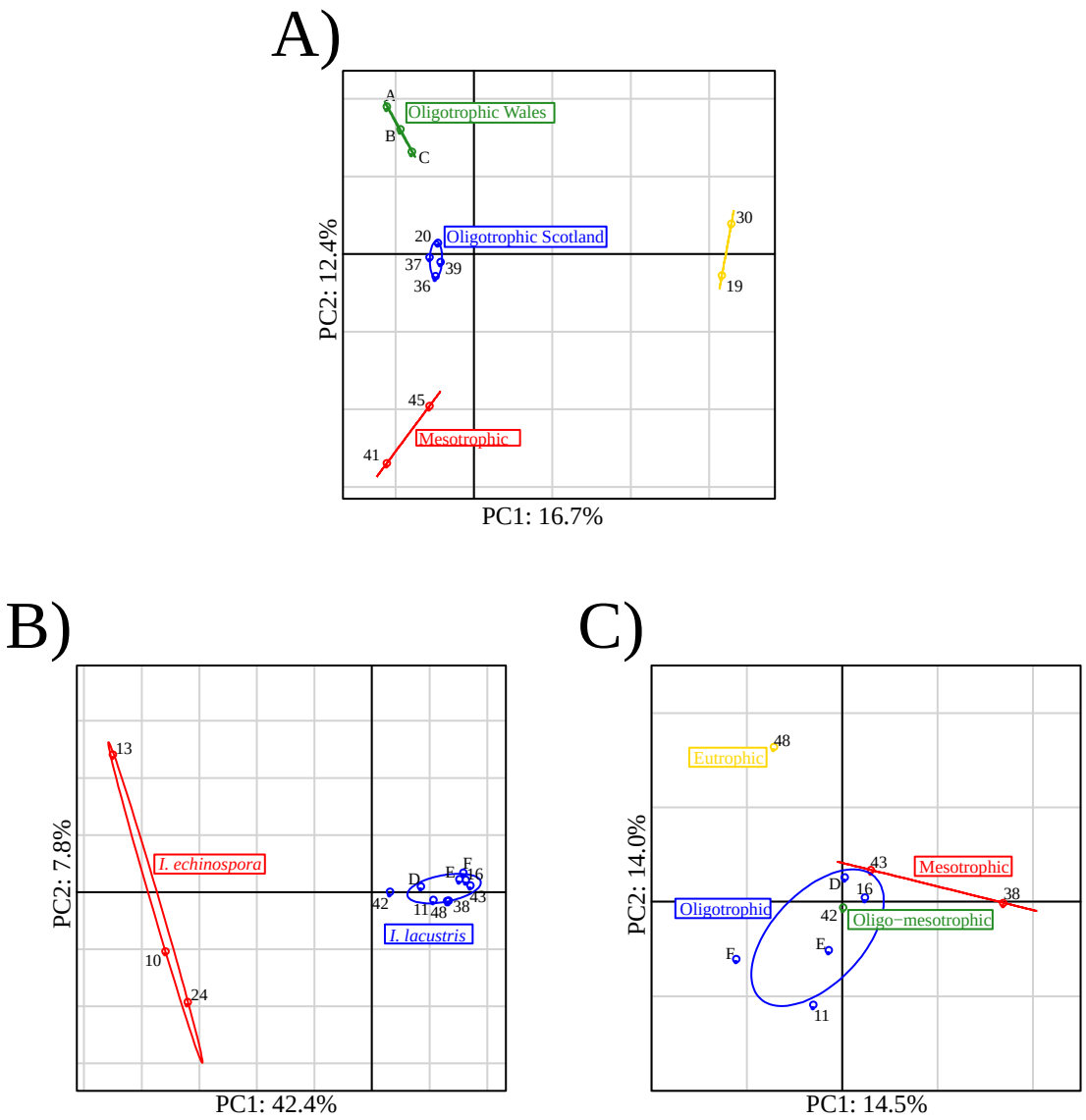


Figure 5.3: Principal Component Analyses of nuclear SNPs

Plots showing samples plotted against the first two principal components for the nuclear single nucleotide polymorphisms (SNPs) for A) *Littorella uniflora*, B) All *Isoëtes* samples and C) *Isoëtes lacustris*, with individual samples labelled. Axes correspond to the percentage of variation in the data explained by each principal component. Coloured ovals correspond to lake type in A) and C), and species of *Isoëtes* in B).

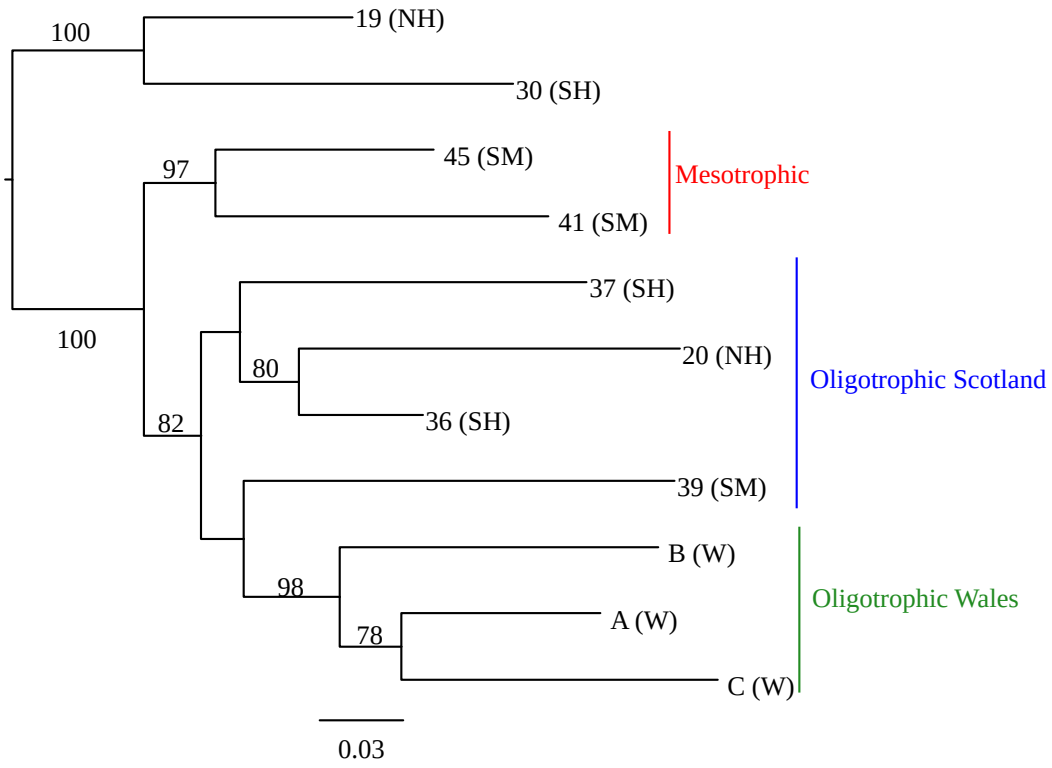


Figure 5.4: Maximum likelihood phylogeny of *Littorella uniflora* nuclear SNPs

Maximum likelihood phylogeny for *Littorella uniflora* nuclear single nucleotide polymorphisms (SNPs). Branch labels represent bootstrap supports out of 100. Scale bar represents substitutions per site. Bootstrap supports less than 50 are not shown. Labels refer to lake type/geographic regions.

Table 5.1: Pairwise F_{st} of *Littorella uniflora*

Geographic group[1]	**W (3)**	**SM (3)**	**SH (3)**	**NH (2)**
W (3)	*			
SM (3)	0.193	*		
SH (3)	0.170	0.139	*	
NH (2)	0.219	0.196	0.160	*
Lake type group[2]	**W O (3)**	**S O (4)**	**S M (2)**	**19 and 30 (2)**
W O (3)	*			
S O (4)	0.159	*		
S M (2)	0.223	0.160	*	
19 and 30 (2)	0.253	0.203	0.250	*

1. Pairwise F_{st} values for samples separated by geographic groups. W = Wales, SM = Scottish Mainland, SH = South Hebrides, NH = Northern Hebrides, see Figure 5.1. Number in brackets indicates number of samples in each group.
2. Pairwise F_{st} values for samples separated by lake type as inferred from genetic clustering. W O = Wales oligotrophic samples, S O = Scottish Oligotrophic samples, SM = Scottish Mesotrophic samples, 19 and 30 = samples 19 and 30 (See Figure 5.3A, Figure 5.1). Numbers in brackets indicate number of samples in each group.

Genetic structure within I. lacustris

An average of 51% of the plastome of *I. lacustris* was covered by sequencing reads (Supplementary Table 5.2). The phylogeny inferred from plastomes revealed two divergent groups within *Isoëtes*, with a bootstrap support of 100 (Figure 5.6). Comparison of previously published *I. lacustris* and *I. echinospora* sequences identified a diagnostic SNP in the *trnL* gene, which suggested the members of the smaller clade were *I. echinospora* and those of the larger clade were *I. lacustris* (Figure 5.6). Bootstrap support within the *I. lacustris* group was generally low.

In total 134,378 RAD loci were assembled for *Isoëtes*, of which 16,451 were retained after filtering (Supplementary Table 5.4). These loci contained 4.4% polymorphic sites. A second assembly was performed using only *I. lacustris* samples, which resulted in a total of 99,672 RAD loci, of which 19,855 were retained after filtering, with 3.5% showing polymorphisms (Supplementary Table 5.5). On average the samples of *I. lacustris* have a lower level of homozygosity than *L. uniflora* (F = 0.32 vs. F = 0.55; Supplementary Table 5.3, Supplementary Table 5.5).

The principal component analysis performed on all *Isoëtes* samples clearly separated the two *Isoëtes* species identified in the chloroplast phylogeny (Figure 5.3B). Similarly, the nuclear phylogeny of *Isoëtes* clearly separated the two species into two highly supported monophyletic clades (Figure 5.7). Within *I. lacustris*, evidence of clustering is less clear than in *L. uniflora*, with samples broadly distributed over the first PC (explaining 14.5% of the variation) with little clustering between the regions or lake types (Figure 5.3C). The second PC explains 14.0% of the variation and broadly separates one sample (48), from a eutrophic loch on North Uist (Outer Hebrides), from the rest of the samples. Branch support values within the nuclear phylogeny of *I. lacustris* are low, and no clustering by geography or lake type is evident (Figure 5.8). F_{ST} values between geographic regions were generally lower in *I. lacustris* than *L. uniflora* (0.09-0.11 vs 0.14-0.22), with similar levels of differentiation between the Wales, Scottish Mainland and Northern Hebrides samples (0.09-0.11, Table 5.2). Oligotrophic and mesotrophic samples showed limited genetic differentiation (F_{ST} = 0.10, Table 5.2).

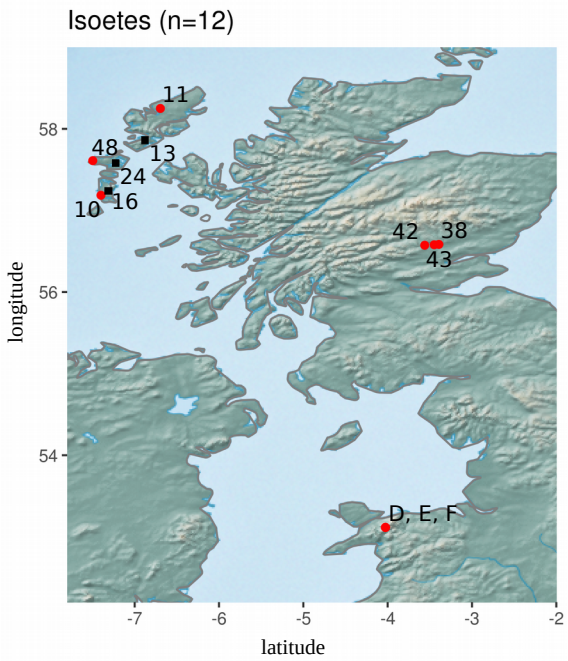


Figure 5.5: Map
Isoetes samples were collected from Cwm Idwal in Wales (W – A, B, C), Aberdeenshire in the Scottish Mainland (SM – 38, 42, 43) and Uist, Harris and Lewis (North Hebrides – NH; 10, 16, 24, 48, 13, 11). Red markers represent *Isoetes lacustris*, black markers represent *Isoetes echinospora* (see main text).

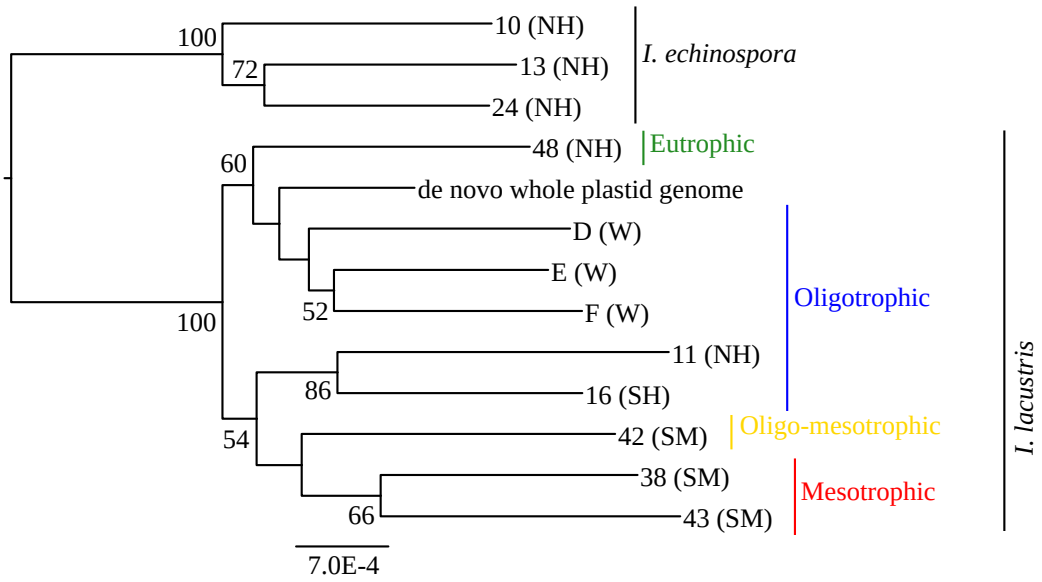


Figure 5.6: Maximum likelihood phylogeny of *Isoëtes uniflora* chloroplasts

Maxmimum likelihoods based on whole chloroplast alignments. Branch labels represent bootstrap supports out of 100. Scale bar represents substitutions per site. Bootstrap supports less than 50 are not shown. Tip labels represent samples (with location in brackets, see Figure 5.1). Letters in brackets refer to sampling locations, see Figure 5.5.

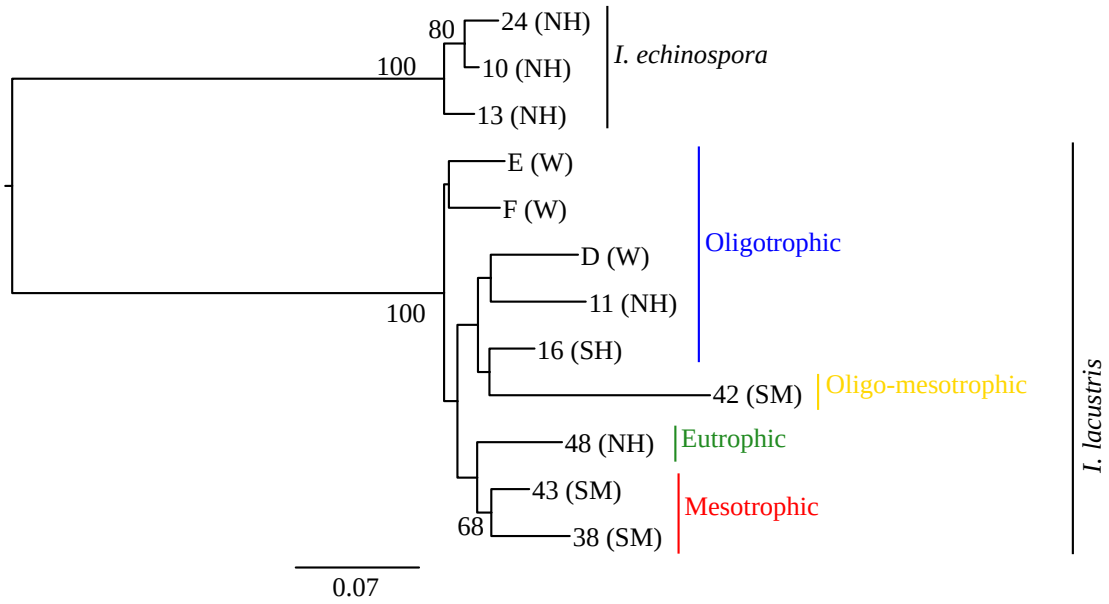


Figure 5.7: Maximum likelihood phylogeny of *Isoëtes* species using nuclear SNPs
Maximum likelihood phylogeny for *Isoëtes* nuclear single nucleotide polymorphisms (SNPs). Branch labels represent bootstrap supports out of 100. Scale bar represents substitutions per site. Bootstrap supports less than 50 are not shown. Labels refer to lake types and species. Letters in brackets refer to sampling locations, see Figure 5.5.

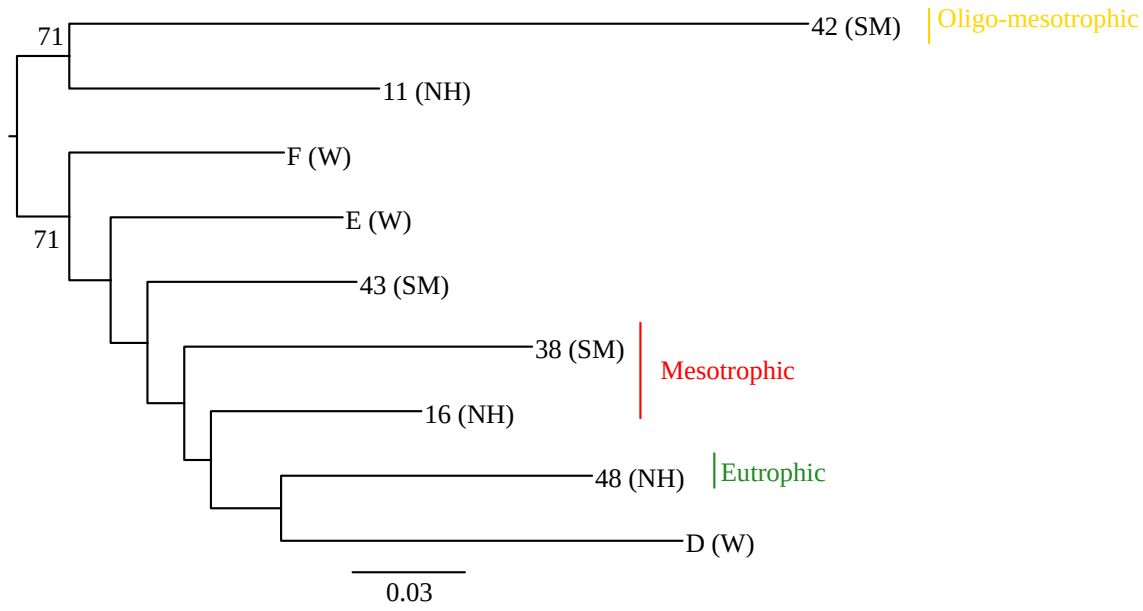


Figure 5.8: Maximum likelihood phylogeny of *Isoëtes lacustris* using nuclear SNPs
Maximum likelihood phylogeny for *Isoëtes lacustris* nuclear single nucleotide polymorphisms (SNPs). Branch labels represent bootstrap supports out of 100. Scale bar represents substitutions per site. Bootstrap supports less than 50 are not shown. Labels refer to lake types. Letters in brackets refer to sampling locations, see Figure 5.5.

Table 5.2 Genetic differentiation between *Isoëtes lacustris* groups

Geographic group[1]	**W (3)**	**SM (3)**	**NH (3)**
W (3)	*		
SM (3)	0.1093602	*	
NH (3)	0.08713628	0.1133132	*
Lake type group[2]	**O (6)**	**M (2)**	**E (1)**
O (6)	*		
M (2)	0.09916852	*	
E (1)	0.1371196	0.2337294	*

1. Pairwise F_{st} values for samples separated by geographic groups. W = Wales, SM = Scottish Mainland, NH = Northern Herbides, see Figure 5.5. Number in brackets indicates number of samples in each group.
2. Pairwise F_{st} values for samples separated by lake type as inferred from genetic clustering. O = oligotrophic and oligotrophic-mesotrophic samples, M = Mesotrophic samples, E = Eutrophic sample. (See Figure 5.3C, Figure 5.5). Numbers in brackets indicate number of samples in each group.

5.5 Discussion

Different waves of colonization of the British Isles

As the ice sheets retreated in post-glacial Britain, *L. uniflora* and *I. lacustris* were both early colonisers of the exposed aquatic habitats (e.g. Birks, 2000). However, this pattern does not seem to have involved a single wave of colonization from a limited number of sources. We identified for both species divergent genetic lineages in geographically close lakes. The cohabitation of distinct genetic groups is consistent with multiple, independent colonizations (e.g. Prentice et al., 2008; Rosenthal et al., 2008; Hedrén, 2009; Schenekar et al., 2014). The distinct group of individuals of *Littorella* identified in some of the Hebridean lakes (samples 30 and 19) might represent glacial survivors (Westergaard et al., 2011) or post-glacial colonisation from a distinct glacial refugia (Jiménez-Mejías et al., 2012). While these two scenarios cannot be distinguished without additional sampling beyond the British Isles, the coexistence of different genetic groups indicates that the freshwater plant populations are not homogenized. This view is moreover supported by the overall high chloroplast diversity coupled with a lack of a clear isolation by distance.

Higher population structure in Littorella

A higher level of genetic structure is observed in *L. uniflora* compared to *I. lacustris*, in terms of phylogenetic resolution, clustering in the principal component analyses, and pairwise genetic distances (Figures 5.2-5.4, Figures 5.6-5.8; Table 5.1- 5.2). These results are consistent with the hypothesis of more frequent sexual reproduction in *I. lacustris* than in *L. uniflora*. A low rate of sexual reproduction in *L. uniflora* could potentially also explain higher levels of nucleotide diversity in *I. lacustris*. One alternative explanation is that migration between lakes is higher in *I. lacustris* than in *L. uniflora*. However, we find the latter scenario unlikely as desiccation boosts the germination of *L. uniflora* seeds (Arts and van der Heijden, 1990), while it reduces that of *I. lacustris* spores (Kott and Britton, 1982), and the similar colonisation times of these species observed in paleolakes (Godwin, 1984; Birks, 2000) suggests similar rates of dispersal. Establishing the causal mechanism for the higher genetic structure in *Littorella* would require additional studies, but our results suggest than gene exchanges in freshwater plants are more effective in lycopods capable of sexual reproducing underwater than in flowering plants with sexual reproduction only in emergent flowers. Rather than being linked to the effectiveness of dispersal among lakes, we suggest that the observed pattern stems from the rate of intrapopulation genetic exchanges and the resulting impact of rare migrants on the different genetic pools.

Some genetic lineages are associated with different types of lakes

Because of our limited sampling, the type of lake is highly correlated with geographic distance in our study. However, an effect of habitat type is suggested for *L. uniflora*, where the population from the oligotrophic Loch of Lowe (39) in Scotland clusters more closely to those of the oligrophic Hebridean and Welsh lakes more than 250km away, as opposed to the populations from the other Aberdeenshire lakes only 6-9km away that are mesotrophic (Figures 4.1 and 4.3A). This pattern suggests that selection was acting on migrants of *L. uniflora* colonising the lake, effectively filtering genotypes by the nutrient conditions. *Littorella uniflora* shows increased growth rates in response to nutrient levels elevated from the low levels typically found in oligotrophic lakes (Christiansen et al., 1985), but declines in growth in more high-nutrient habitats (e.g. Farmer and Spence, 1986; Robe and Griffiths, 1992), potentially due to competition or

nitrogen toxicity (Robe and Griffiths, 1994; Smolders et al., 2002), indicating that nutrient status is likely to exert strong selective pressures. Transplant experiments between eutrophic and oligotrophic lakes in Cumbria found some evidence of adaptation to increased nutrient levels (Robe and Griffiths, 1992), suggesting the existence of ecotypes specialising in lakes of different nutrient status. This ecological filtering would also be consistent with higher levels of homozygosity observed in *Littorella*, due to reduced hybrid fitness. We suggest that the capacity to thrive in mesotrophic lakes evolved in some *L. uniflora* populations before or at the early stages of the colonization of the British Isles, limiting the subsequent migration to different lake types.

Within *I. lacustris*, there was less evidence of genetic associations between samples due to nutrient type than geography, although a single sample in a eutrophic lake that was relatively highly differentiated from the other populations (Figure 5.3C). Growth of *Isoëtes* is also likely highly influenced by nutrient levels (Gacia and Ballesteros, 1994; Arts, 2002), so that local adaptation might be expected. Our results do not test for local adaptation, but indicate that genetic lineages within *I. lacustris* are not restricted to specific lake types. More genetic exchanges as a result of frequent sexual reproduction would increase the pool of adaptive alleles available to the populations, potentially facilitating adaptation to complex, heterogenous aquatic environments (Santamaría, 2002; Becks and Agrawal, 2010; Luijckx et al., 2017). The extent to which these exchanges could contribute to adaptation to particular lake types would be dependent on multiple factors, such as rates of migration, the strength of selection and the genetic architecture of the trait (Rundle and Nosil, 2005; Leimu and Fischer, 2008). Testing the extent to which different *Isoëtes* populations are adapted to varying nutrient conditions would require dedicated experiments (Blanquart et al., 2013), but our results suggest that the ability to reproduce sexually underwater could facilitate the spread of adaptive alleles between populations in *I. lacustris*.

5.6 Conclusions

In this study, we compared the genetic structure within the British Isles of two freshwater plants belonging to very divergent groups; the lycopod *I. lacustris* and the angiosperm *L. uniflora*. Our investigations revealed higher levels of population structure in *L. uniflora* than in *I. lacustris* and we suggest this stems from increased opportunity for underwater sexual reproduction in the lycopod *I. lacustris*. *Littorella uniflora,* inheriting the angiosperm mating system from its terrestrial ancestors, relies on above water structures for the production of flowers and seeds, making sexual reproduction dependent on fluctuating water levels. The ancestral reproductive system of lycopods that was less specialized for terrestrial conditions, facilitates genetic exchanges in secondarily aquatic habitats. Furthermore, certain lineages of *Littorella* appeared to be restricted to lakes of particular nutrient status. We suggest that this results from early adaptation of some populations to new habitats following by strong ecological filtering. This pattern is not observed in *I. lacustris*, which could be explained by frequent genetic exchanges in this species allowing the potentially more rapid spread of adaptive alleles among lineages.

Acknowledgements

We would like to thank Scottish Natural Heritage for funding the fieldwork in Scotland, and Emma Baxter, Lidia Korba and Hannah Sewell for assistance in sampling, Nick Stewart for performing sample identification and Pascal-Antoine Christin for providing very useful comments on the manuscript. Jill Olofsson and Luke Dunning are funded by a NERC grant NE/M00208X/1 and an ERC grant ERC-2014-STG-638333.

5.7 Supplementary Information

Supplementary Table 5.1: Sample Locations

Sample	Genus	Species	Lake	Location	National Grid Reference	Lake type
10	*Isoëtes*	*echinospora*	Loch Na Cuithe	South Uist	NF 73910 23510	Machair
13	*Isoëtes*	*echinospora*	Loch Laxdale	Harris	NG 10840 96122	Oligotrophic
24	*Isoëtes*	*echinospora*	Loch Scadabhaigh	North Uist	NF 87830 66612	Oligotrophic
11	*Isoëtes*	*lacustris*	Loch Laxavat Ard	Lewis	NB 24639 38628	Oligotrophic
16	*Isoëtes*	*lacustris*	Loch Faoilean	South Uist	NF 79704 29017	Oligotrophic
38	*Isoëtes*	*lacustris*	Loch Marlee	Perthshire	NO 14549 44449	Mesotrophic
42	*Isoëtes*	*lacustris*	Loch of Lowes	Perthshire	NO 04202 43517	Oligo-mesotrophic
43	*Isoëtes*	*lacustris*	Loch Clunie	Perthshire	NO 11149 44000	Mesotrophic
48	*Isoëtes*	*lacustris*	Loch Grogary	North Uist	NF 71868 71132	Eutrophic
D	*Isoëtes*	*lacustris*	Llyn Idwal	Snowdonia	SH 64503 59677	Oligotrophic
E	*Isoëtes*	*lacustris*	Llyn Idwal	Snowdonia	SH 64503 59677	Oligotrophic
F	*Isoëtes*	*lacustris*	Llyn Idwal	Snowdonia	SH 64503 59677	Oligotrophic
19	*Littorella*	*uniflora*	Loch Laxdale	Harris	NG 10840 96122	Oligotrophic
20	*Littorella*	*uniflora*	Loch Scadabhaigh	North Uist	NF 87830 66612	Oligotrophic
30	*Littorella*	*uniflora*	Loch An T-Saigart	Coll	NM 25027 60911	Oligotrophic
36	*Littorella*	*uniflora*	Loch Riaghain	Tiree	NM 03427 46886	Oligotrophic
37	*Littorella*	*uniflora*	Loch A'Chlair	Tiree	NL 98275 44521	Oligotrophic
39	*Littorella*	*uniflora*	Loch of Lowes	Perthshire	NO 04202 43517	Oligo-mesotrophic
41	*Littorella*	*uniflora*	Loch Clunie	Perthshire	NO 11149 44000	Mesotrophic
45	*Littorella*	*uniflora*	Loch Marlee	Perthshire	NO 14549 44449	Mesotrophic
A	*Littorella*	*uniflora*	Llyn Idwal	Snowdonia	SH 64503 59677	Oligotrophic
B	*Littorella*	*uniflora*	Llyn Idwal	Snowdonia	SH 64503 59677	Oligotrophic
C	*Littorella*	*uniflora*	Llyn Idwal	Snowdonia	SH 64503 59677	Oligotrophic

Supplementary Table 5.2: Coverage of the chloroplast genome from RAD-seq data

Sample	Genus	Species	% chloroplast coverage
10	*Isoëtes*	*echinospora*	38.5
13	*Isoëtes*	*echinospora*	38.7
24	*Isoëtes*	*echinospora*	*34.9*
11	*Isoëtes*	*lacustris*	*32.9*
16	*Isoëtes*	*lacustris*	*61.1*
38	*Isoëtes*	*lacustris*	*49.1*
42	*Isoëtes*	*lacustris*	*51.2*
43	*Isoëtes*	*lacustris*	*58.9*
48	*Isoëtes*	*lacustris*	*45.2*
D	*Isoëtes*	*lacustris*	*61.8*
E	*Isoëtes*	*lacustris*	*67.7*
F	*Isoëtes*	*lacustris*	*69.9*
19	*Littorella*	*uniflora*	*79.2*
20	*Littorella*	*uniflora*	*79.2*
30	*Littorella*	*uniflora*	*74.7*
36	*Littorella*	*uniflora*	*81.8*
37	*Littorella*	*uniflora*	*83.9*
39	*Littorella*	*uniflora*	*86.1*
41	*Littorella*	*uniflora*	*85.4*
45	*Littorella*	*uniflora*	*83.3*

A	*Littorella*	*uniflora*	*86.5*
B	*Littorella*	*uniflora*	*83.6*
C	*Littorella*	*uniflora*	*88.1*

Supplementary Table 5.3: *Littorella* RAD loci assembly statistics

Sample	Number of raw reads	Number of filtered reads	Loci in assembly (total 14,669)	Percentage coverage	F
19	1776273	1714419	11646	58.7	0.39207
20	928303	872503	8839	44.5	0.64897
30	1357972	1311865	9927	50	0.57042
36	2605150	2442373	13425	67.6	0.23717
37	1052085	1031690	8727	44	0.67591
39	813179	798731	6435	32.4	0.76217
41	1540588	1500158	11202	56.4	0.53883
45	2199039	2077674	12664	63.8	0.35576
A	2101413	1979030	12197	61.4	0.45061
B	1408609	1362492	8691	43.8	0.64581
C	1335051	1294779	7837	39.5	0.74772

Supplementary Table 5.4: Combined *Isoëtes* RAD-loci

Species	Sample	Number of raw reads	Number of filtered reads	Number of loci in assembly (total 16,451)	Percentage coverage
Isoëtes echinospora	10	1449515	1418605	6167	37.5
Isoëtes echinospora	13	2317696	2230866	8825	53.6
Isoëtes echinospora	24	1138801	1107718	5547	33.7
Isoëtes lacustris	11	782066	756049	6786	41.2
Isoëtes lacustris	16	1626396	1555625	12038	73.2
Isoëtes lacustris	38	1139299	1103562	9084	55.2
Isoëtes lacustris	42	329893	321378	2105	12.8
Isoëtes lacustris	43	2014018	1923331	13005	79.1
Isoëtes lacustris	48	1085263	1051138	8851	53.8
Isoëtes lacustris	D	942854	928519	5716	34.7
Isoëtes lacustris	E	1982271	1910507	12495	76
Isoëtes lacustris	F	2396914	2303946	13353	81.2

Supplementary Table 5.5: *Isoëtes lacustris* RAD-loci assembly statistics

Sample	Number of raw reads	Number of filtered reads	Loci in assembly (total 19,855)	Percentage coverage	F
11	782066	756049	8093	40.8	0.5142
16	1626396	1555625	14734	74.2	0.2321
38	1139299	1103562	10921	55	0.35323
42	329893	321378	2552	12.9	0.54424
43	2014018	1923331	15998	80.6	0.08458
48	1085263	1051138	10610	53.4	0.40972
D	942854	928519	6935	34.9	0.49812
E	1982271	1910507	15165	76.4	0.1679
F	2396914	2303946	16239	81.8	0.08037

6. Discussion

6.1 Factors promoting CAM photosynthesis in aquatic conditions

CAM photosynthesis, evolving in highly disparate organisms and ecological settings, represents one of the best examples of complex trait evolution in highly divergent environmental contexts. The extent to which traits respond to similar environmental pressures indicates whether they have the same function, and therefore whether they are actually the same trait at all, with some authors suggesting they are distinct traits (Aulio, 1986a; Bowes and Salvucci, 1989). Whilst this difference is arguably largely semantic (Bowes and Salvucci, 1989), the extent to which they perform a similar role is informative in studying the constraints of evolution. If they are entirely different traits, achieving them via highly similar mechanisms indicates fundamental constraints within evolution.

The function of CAM in submerged aquatic plants has long been considered to be primarily a carbon concentrating mechanism (Keeley, 1981; Keeley, 1998; Silvera et al., 2010), which represents an adaptation to the low levels of carbon dioxide dissolved underwater. However, dynamics of carbon dioxide sources in aquatic systems are significantly more variable than those available to terrestrial plants. with sources including direct dissolving of atmospheric CO_2 into water bodies, organic matter and respiration derived from the sediment. The level of these processes is mediated by temperature, pH and water disturbance affecting boundary layers (Madsen and Sand-Jensen, 1991; Maberly and Madsen, 2002). Only a relatively small proportion of submerged aquatic plants are CAM (Keeley, 1998), and it is likely that only a subset of conditions are favourable to CAM photosynthesis. Submerged aquatic plants with strong CAM activity mostly inhabit oligotrophic lakes and vernal pools (Keeley, 1998), although larger numbers of plants display some level of CO_2 fixation in the dark (Keeley, 1998; Zhang et al., 2014), indicating CAM may represent a continuum in submerged aquatic environments, as in terrestrial plants (Silvera et al., 2010). The plastic nature of CAM in submerged plants further indicates that conditions favouring CAM photosynthesis may be relatively heterogenous in space or time. Indeed, reductions of light level, CO_2 decreases and terrestrialisation have previously been shown to have strong effects on aquatic CAM activity (Aulio, 1985; Madsen, 1987b;

1987c; Baattrup-Pedersen and Madsen, 1999; Keeley, 1998; Klavsen and Maberly, 2010). These effects were however surprisingly not replicated in my experiments (Chapter IV).

Published evidence of environmentally-induced CAM pathways in some aquatic plants was one of the reasons my work focused on *I. lacustris* and *L. uniflora* - comparing transcriptomes in a CAM versus C_3 state was expected to aid the identification of CAM related transcripts, and potentially enable follow-up experimental work. I consequently focused my efforts on getting the plants to switch their photosynthetic type. As described in Chapter IV, I first grew plants in a set of conditions thought to mimic the variation the plants encounter in the wild (low-light vs. high-light and submerged vs. terrestrial). Because these did not switch the CAM cycles on and off as expected, I then grew a new set of plants at higher temperature and low CO_2, which still failed to produce changes in CAM activity between treatments (Chapter IV). After the study presented in Chapter IV, I performed additional experiments using *I.lacustris* and *L.uniflora* to attempt to produce changes in CAM activity. I hypothesised that the acclimation time to low CO_2 in the plants used for Chapter IV may have been insufficient to induce differences in CAM activity, or that light levels may have been reduced by shading from other plants present in the chambers, resulting in reduced CAM activity in both conditions, as was found in *L. uniflora* by Madsen, 1987b. I also hypothesised that high humidity levels may have caused the terrestrial *L. uniflora* in Chapter IV to continue to display CAM activity, as demonstrated in *L. uniflora* by Aulio, 1986b.

These additional experiments were conducted as follows. *Littorella uniflora* and *I. lacustris* were placed in 40 x 30 x 25cm transparent plastic containers in a substrate consisting of 20 parts sand to 1.5 parts Humax Sterilsed Loam (East Riding Horticulture), and filled with deionised water, topped up as necessary. In addition, *L. uniflora* individuals were placed in containers with a substrate of 15 parts sand to 5 parts vermiculite to 1.5 parts Humax Sterlised Loam. These were not filled with water, but watered every few days until the substrate was saturated. Plants were placed in empty Conviron growth chambers with ambient (approximately 500ppm) and low (220ppm) CO_2 concentrations. They were given 16 hours of light (500mol μmol m^{-2} s^{-1} light, 60% humidity, 20°C) and 8 hours of darkness (50% humidity, 15°C). Terrestrial *Littorella* plants were only placed in the low CO_2 chamber and developed terrestrial

leaves prior to harvesting. After 28 days plant leaves were harvested, metabolites were extracted and malate levels were quantified with mass spectrometry, using the same techniques as described in Chapter IV. The diurnal malate fluctuations are shown in Figure 6.1.

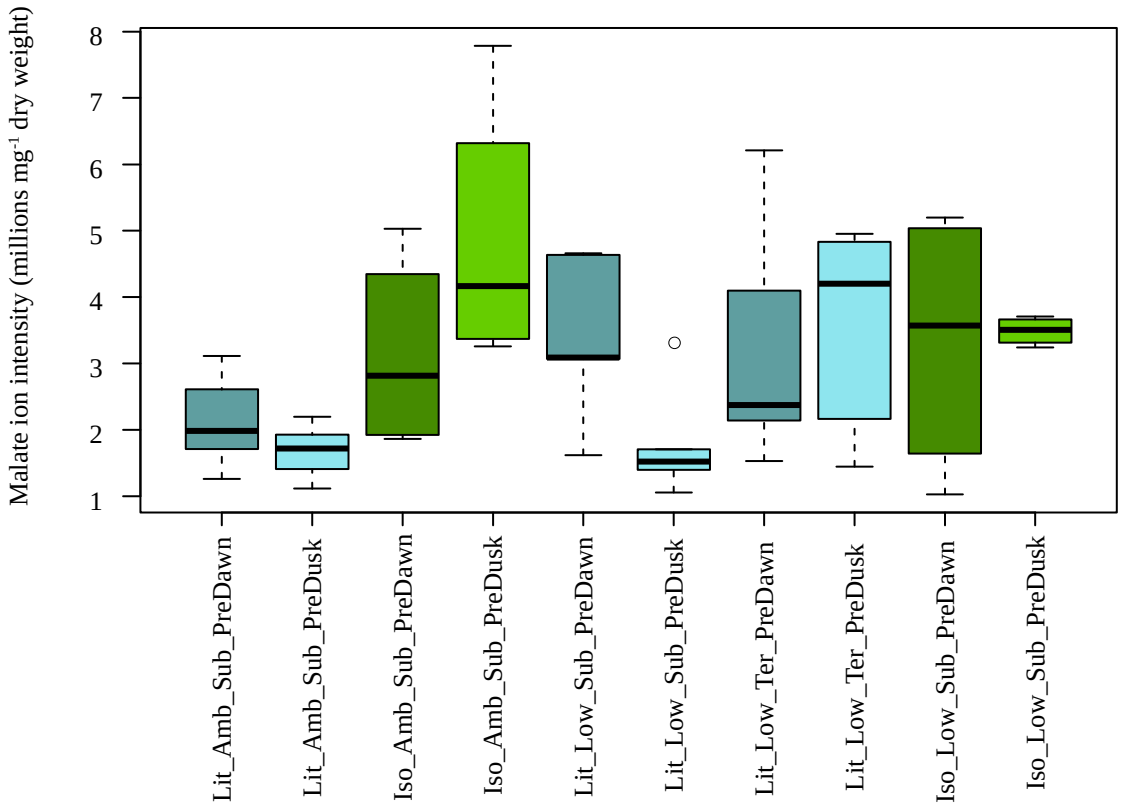


Figure 6.1 Diurnal malate fluctuations in *Littorella uniflora* and *Isoëtes lacustris* in high and low CO_2
Box plots showing levels of malate in millions of ions detected per mg dry weight of material, in samples collected one hour before dawn (PreDawn) and one hour before dusk (PreDusk) in ambient (Amb) and low (Low) CO_2 conditions, in *L. uniflora* (Lit, blue) and *I. lacustris* (Iso, green), in either submerged (Sub) or terrestrial (Ter) form. N = 3-6.

These results do not show diurnal acid fluctuations diagnostic of CAM activity in any of the conditions. Submerged *L. uniflora* in the low CO_2 conditions shows some diurnal acid fluctuations, but these were not consistent across all individuals tested.

Terrestrialisation has previously been shown to reduce CAM activity in *Littorella* and *Isoëtes*, consistent with increased access to atmospheric CO_2. Some studies have shown continued acid fluctuations in terrestrial conditions (Farmer and Spence, 1985; Aulio, 1986b; Nielsen et al., 1991), with fluctuations partly controlled by low levels of humidity (Aulio, 1986b). Reduced levels of diurnal malate fluctuations in

the terrestrial *L. uniflora* in these samples compared with the plants in Chapter IV that were grown at higher humidity levels is consistent with previous results.

Carbon dioxide levels have also previously been shown to affect CAM activity, with increased carbon dioxide levels associated with reduced CAM activity in a variety of aquatic lineages (Madsen, 1987b; Baattrup-Pedersen and Madsen, 1999; Klavsen and Maberly, 2010; Shao et al., 2017). As explained above, this prompted me to try different CO_2 concentrations, but again no clear effect of lowering CO_2 on CAM activity was found. In Chapter IV lower light levels due to shading may have meant carbon was not limiting, consistent with the upregulation of photosynthetic genes in the ambient CO_2 *Isoëtes*. By contrast, CAM activity is absent in the ambient conditions in this experiment, and declining CO_2 levels do not appear to induce them. This is surprising given the positive effects of light on CAM activity as measured in previous experiments (Madsen, 1987b). Considering that changes in CAM capacity can take several weeks to become apparent in *Littorella (*Madsen 1987b), it may be that CAM activity would have similarly been reduced in the high/low CO_2 experiments in Chapter IV if the plants had been left to acclimatise for longer. Prolonged exposure to low nutrient levels may have reduced the capacity for CAM activity, as observed in Madsen (1987b), and although levels of Nitrogen, Potassium and Phosphorous in the substrate were determined to be similar to those in Llyn Idwal sediment (using a LaMotte NPK Soil Testing Kit), it may be that the presence of additional micronutrients may have been required to sustain CAM activity in these plants.

Overall, the distribution and plastic responses of aquatic CAM indicate that it likely functions as a carbon concentrating mechanism linked to CO_2 levels, but that a number of other factors are potentially important in mediating whether CAM is activated. The speed at which CAM activity can be induced (see results of flooding experiment described below) indicates that the conditions favourable to CAM can be highly variable even within the range of a single species. The results of my different experiments do not clearly elucidate the conditions in which *Isoëtes* and *Littorella* perform CAM, and this will require further investigation. In particular, I observed throughout my different experiments significant variation among individuals grown in the same condition. Whether these are linked to genetic variants, the condition of the plants, or random variation is not known. Determining the exact drivers of the levels of CAM activity in different individuals and conditions would require dedicated

experiments. Multiple populations, and clones within populations, should be compared to exclude intraspecific genetic variability of CAM plasticity. Additionally, measurements should be repeated along time series spanning multiple days, to assess random variation. Such experiments would identify more precisely the conditions promoting CAM photosynthesis in these plants, and the environmental triggers of altered CAM activity.

6.2 Global drivers of CAM evolution

The diversity of habitats where CAM plants occur clearly indicates that this adaptation can be advantageous in a number of conditions, although these likely represent only a subset of of habitats found around the globe (Chapter II). However, the extent to which the availability of suitable conditions fluctuated through geological times was unclear. Large fluctuations in atmospheric CO_2 have occurred throughout earth's history (Pagani et al., 2005; Hansen et al., 2008; Foster et al., 2017), and recent declines in CO_2 in the past 30 million years acted as a key enabler of C_4 evolution in terrestrial plants (Christin et al., 2008; Christin and Osborne, 2013). The extent to which atmospheric CO_2 changes are likely to affect CAM plants in submerged aquatic environments is debated. Permanently low CO_2 conditions in oligotrophic lakes are hypothesised to be closely linked to atmospheric CO_2 concentrations (Keeley and Rundel, 2003; Short et al., 2016). Low daytime carbon concentrations in vernal pools, by contrast, are driven mainly by rapid drawdown due to photosynthesis (Witham et al., 1998), which are likely to occur regardless of atmospheric CO_2 levels (Keeley, 1998; Keeley and Rundel, 2003) Nevertheless, the rate at which CO_2 is replenished is proportional to the level of CO2 in the atmosphere, and the correlation of CO_2 levels with depth in vernal pools (Holgerson, 2015) indicates that falling CO_2 levels could reduce the number of habitats with carbon limitation severe enough to prompt the evolution of CAM. Low CO_2 levels are therefore unlikely to act as a hard prerequisite for the evolution of aquatic CAM plants, but may increase the probability of CAM evolution by expanding the range of habitats in which CAM is favourable.

Molecular dating indicates that CAM clades evolved and diversified in terrestrial groups in a period where CO_2 levels were low (Arakaki et al., 2011; Givnish et al., 2014). In most cases, CAM origins are therefore contemporaneous to C_4 origins, but

exceptions were thought to exist. In particular, *Isoëtes* was largely assumed to represent a very ancient CAM lineage (Edwards and Ogburn, 2012). The group certainly diverged from angiosperms a very long time ago, and fossils indicate that *Isoëtes*-like forms existed hundreds of million years ago (Retallack, 1997, McLoughlin et al., 2015). Some molecular dating analysis supported this hypothesis, placing the diversification of extant *Isoëtes* more than 100 million years ago (e.g. Larsén and Rydin, 2015). However, my careful analyses of divergence times based on markers sampled across whole plastid and nuclear genomes indicate that the diversification of CAM *Isoëtes* occurred in the past 45-60 million years (Chapter III). It is therefore contemporaneous to the origins of terrestrial CAM groups (Arakaki et al., 2011) as well as aquatic CAM lineages in the angiosperms (Chapter IIII). The diversification of CAM *Isoëtes* moreover coincided with falling levels of atmospheric CO_2, which is compatible with a large enabling role of falling atmospheric CO_2 in CAM evolution and CAM lineage expansion in aquatic plants (Chapter III). These results do not rule out an ancient origin of CAM in *Isoëtes*, as has previously been hypothesised based on the presence of fossils similar to *Isoëtes* stretching back to the Triassic (Ash and Pigg, 1991; Retallack, 1997; Cantrill and Webb, 1998; Keeley and Rundel, 2003; Lüttge, 2004; Raven et al., 2008; Silvera et al., 2010). The presence of lacunae in these species suggests limits to gas diffusion in submerged aquatic environments occurred in ancient habitats, although many plants lacking CAM share these features (Sculthorpe, 1967). Nevertheless, despite historically low CO_2 levels in the present day, only a small number of aquatic species are CAM (Keeley, 1998) occupying relatively marginal habitats. In the high CO_2 environments of the Mesozoic, if CAM species were present, available habitats may have been relatively small. However, a period of low CO_2 levels existed in the Paleozoic, about 300 million years ago (Osborne and Beerling, 2006; Montañez et al., 2016). This might have triggered an ancient expansion of aquatic CAM lineages, including early members of the *Isoëtes* lineage. This hypothesis however cannot be tested with the data currently available, with extant *Isoëtes* taxa representing a later diversification. Overall, the results of my molecular dating analyses support the idea that the likelihood of aquatic CAM evolution and diversification might be strongly influenced by global environmental factors.

Whilst falling levels of CO_2 are a plausible common factor influencing the evolution of CAM in both terrestrial and aquatic CAM groups (Chapters II and III), it is entirely possible that separate mechanisms explain this convergence. For example,

increasing aridity in terrestrial environments and decreasing temperatures resulting in reduced CO_2 diffusivity in aquatic environments in the past 60 million years could explain the diversification patterns observed. However, the diversity of habitats inhabited by both groups means that a large number of factors is required to explain the coincident increases in their origins and diversification. For example, terrestrial CAM expansion in the past 50-60 million years has been identified in both xerophytic plants such as cacti (Arakaki et al., 2011) and tropical epiphytes such as bromeliads (Givnish et al., 2014), as well as the occurrence of CAM in both temperate oligotrophic lakes and warm vernal pools (Chapter III). Falling levels of atmospheric CO_2 represent a much simpler explanation of this pattern than the coincidental expansion of all these types of habitat. This indicates that falling atmospheric CO_2 levels represent a fundamental challenge to plants in a wide range of habitats, and likely provides part of the explanation as to how a complex trait such as CAM has evolved so many times in diverse lineages of plants (Keeley and Rundel, 2003). Carbon assimilation is a fundamental part of photosynthesis, itself at the heart of plant growth and survival. Atmospheres with low-CO_2 concentrations atmospheres therefore directly and indirectly impact a number of plant processes, ranging from carbon assimilation itself to water-use efficiency (via the regulation of exchanges with the atmospheres) and nitrogen-use efficiency (stemming from the amount of photosynthetic enzymes required). It is therefore likely that changes in the atmospheric conditions over geological times had multifarious consequences on the functional diversification of plants, and drastically impacted the fitness effects of different traits in a variety of environments. This includes CAM photosynthesis in both aquatic and terrestrial settings, which can be used to address a number of the challenges imposed by CO_2 limitations depending on how it is actualised within the plant (Chapter II).

The origin of lineages of CAM plants appears to take place earlier than C_4 plants, the origins of which are strongly associated with further declines in atmospheric CO_2 levels in the Oligocene, to reach very low levels that persisted over the past 30 million years (Christin et al., 2008; Christin and Osborne, 2013). The continuous and plastic nature of CAM means phylogenetic analyses covering all CAM plants have not been undertaken (Silvera et al., 2010). Whilst obligate CAM plants are relatively easy to identify, facultative CAM may not be detected in all sampling conditions (environment and time). In practice, this means that while the diversification of obligate times can be tracked through time, the origin of the CAM pathway itself is difficult to place on a

phylogeny. For example, whilst the origin of cacti has been dated, the origin of CAM has not been placed among its ancestors, mainly because the extent to which its relatives perform some kind of CAM pathway is still unknown (Arakaki et al., 2011). The same is true of *Isoëtes* (Chapter III), which indeed lacks any close living relatives. In both groups the CAM trait may have evolved a long time prior to diversification of extant species. Therefore, while the major diversification of some CAM and C_4 groups happened in the same period of global changes in the Miocene (Arakaki et al., 2011, Givnish et al., 2014), CAM lineages likely existed before the first C_4 plants appeared. It is probable that the fundamental role of C_4 in alleviating the effects of photorespiration differs from that of CAM and therefore has a different critical threshold of atmospheric CO_2, and that the relatively inflexible nature of C_4 means that evolution may be restricted to when low CO_2 levels affect fitness throughout the lifetime of the plant, in a variety of habitats. The CO_2/temperature combination at which C_4 theoretically gains an advantage over the C_3 state has been evaluated (Ehleringer et al., 1997, Osborne and Beerling, 2006), and whilst models incorporating atmospheric CO_2 and water use efficiency exist for CAM plants (e.g. Comins and Farquhar, 1982, Bartlett et al., 2014), an explicit evolutionary framework is absent. Future modelling efforts integrating global and local environmental variability would be informative in explaining observed patterns of CAM diversification and evolution.

6.3 A new hypothesis for the early steps of CAM evolution

Even with a favourable selective environment present, the assembly of a complex trait such as CAM cannot result from a single mutation – a series of intermediate states are likely required, which may influence both the environments a C_3 to CAM transition occur in and the components likely to be recruited for CAM photosynthesis. The presence of CAM as a continuum from low level acid fluctuations and recycling of respiratory CO_2 at night to total reliance on the pathway with negligible daytime CO_2 uptake provides a plausible pathway of intermediates (Silvera et al. 2010). Selection can act on the rate limiting step in the pathway. For example, increased diurnal expression of decarboxylases, or increased vacuolar storage capacity (Silvera et al., 2010) – until this step is no longer rate limiting, with the next most rate limiting step then being selected for (Newton et al., 2015). However, this does not explain how an initial low level of CAM activity was established (Bräutigam et al., 2017). One

hypothesis is that amino acid metabolism in some C_3 plants involves the night-time synthesis and storage of organic acids, that could result in a weak CAM cycle (Bräutigam et al., 2017). This hypothesis predicts that species lacking night-time organic acid storage are unlikely to evolve towards CAM metabolism (Bräutigam et al., 2017), which could plausibly take place in terrestrial or submerged environments. Alternatively, I hypothesised that the transition to a submerged aquatic environment resulted in perturbations to metabolism over a diurnal cycle which could have provided the required fluxes in malate levels. Hypoxia due to falling oxygen levels is a severe challenge to submerged terrestrial plants, resulting in reductions in respiration and therefore an increase in mortality (Voesenek et al., 2006). Malate accumulation has been hypothesised as a potential response to anoxia, acting as an alternative end-point to glycolysis, avoiding the toxic effects of ethanol build-up (McManmon and Crawford, 1971; Crawford and Zochowski, 1984). Malate indeed accumulates in some flooding tolerant species of plants in response to flooding or hypoxia (Crawford and McManmon, 1968; Joly, 1994; Avadhani et al., 1978). Increased adaptation to flooding could result in increased photosynthetic activity, generating oxygen during the day. Underwater photosynthesis can result in diurnal cycles of oxygen availability in submerged plants (Sand-Jensen et al., 2005; Voesenek et al., 2006), which could in turn lead to malate accumulation at night as a result of hypoxia, followed by decarboxylation during the day. This cycle could then be co-opted as a carbon concentrating mechanism in submerged aquatic plants. This hypothesis would predict that species lacking a diurnal accumulation of malate when flooded would be unlikely to evolve CAM photosynthesis, and would suggest separate mechanisms of CAM evolution in terrestrial and submerged environments.

To test my new hypothesis and that of Brautigam et al., I investigated malate levels over the day-night cycle in two species of *Plantago*, terrestrial relatives of *L. uniflora*, in flooded and terrestrial conditions. Both *Plantago* species tested, *P. lanceolata* and *P. maritima,* exhibit flooding tolerance (Jerling, 1981; Banach et al., 2009). I collected these plants from Sheffield and Cleethorpes, U.K. These plants, in addition to *L. uniflora* collected from Cwm Idwal, Snowdonia (see Chapters III and IV) that had been growing submerged in sand in Sheffield for 1 year were transferred to 40 x 30 x 25 cm transparent plastic boxes. The substrate consisted of 15 parts silica sand to 5 parts perlite to 1.5 parts Humax Sterlised Loam, at a depth of 9 cm. Individuals of

each species were dispersed evenly between 15 boxes (on average three *Littorella uniflora*, four *Plantago lanceolata* and two *Plantago maritima* per box). Plants were grown outside for 1 month, watered weekly with distilled water, before being transferred to a greenhouse chamber with a 16/8 day/night cycle with temperatures of 20/15 degrees. They were left to acclimate for an additional month. In this time, *L. uniflora* developed terrestrial leaves. In the greenhouse, plants were watered to field capacity with distilled water every 2 days. After 1 month, half of the boxes were filled to the top with distilled water. Water lost through evaporation was replenished every 2-3 days, and control boxes were watered with the same watering regime as previously. Entire leaves were harvested from flooded and control plants on 1, 6 and 13 days after flooding, at 1 hour before dusk and 1 hour before dawn and frozen immediately in liquid nitrogen and stored at -80 C. Metabolites were extracted and quantified as described previously (Chapter IV).

Malate levels in the flooded and control plants are indicated in Figure 6.2 for *L. uniflora,* Figure 6.3 for *P. lanceolata* and Figure 6.4 for *P. maritima*. In the control individuals left in terrestrial conditions, malate levels varied between plants but displayed relatively little variation within plants over the day/night cycle. Flooded individuals of *P. maritima* and *P. lanceolata* did not show altered malate levels following flooding, or exhibit diurnal malate fluctuations (Figure 6.3, Figure 6.4). By contrast, after six days of flooding *L. uniflora* showed strong diurnal fluctuations of malate consistent with induced CAM activity (Figure 6.2). These results do not support either of the aforementioned hypotheses. The lack of malate fluctuations in either terrestrial or submerged *Plantago* species suggests that the terrestrial ancestor of *Littorella* lacked diurnal acid fluctuations, either generally or specifically following flooding. This experiment does not comprehensively rule out either hypothesis, being based on only a single origin of CAM and using relatively distantly related species to *L. uniflora.* In addition, the relatively shallow, clear water and fast onset of flooding with mature plants may not be a realistic approximation of the transition to aquatic environments in an ancestor of *L. uniflora*. If either of these hypotheses are correct, however, these results indicate that the respective underlying processes i) are not important in every origin of CAM, ii) are relatively evolutionarily labile, being either lost in *Plantago* or gained in the *Littorella* ancestor, or iii) are sensitive to environmental conditions. Some combination of all three are equally possible.

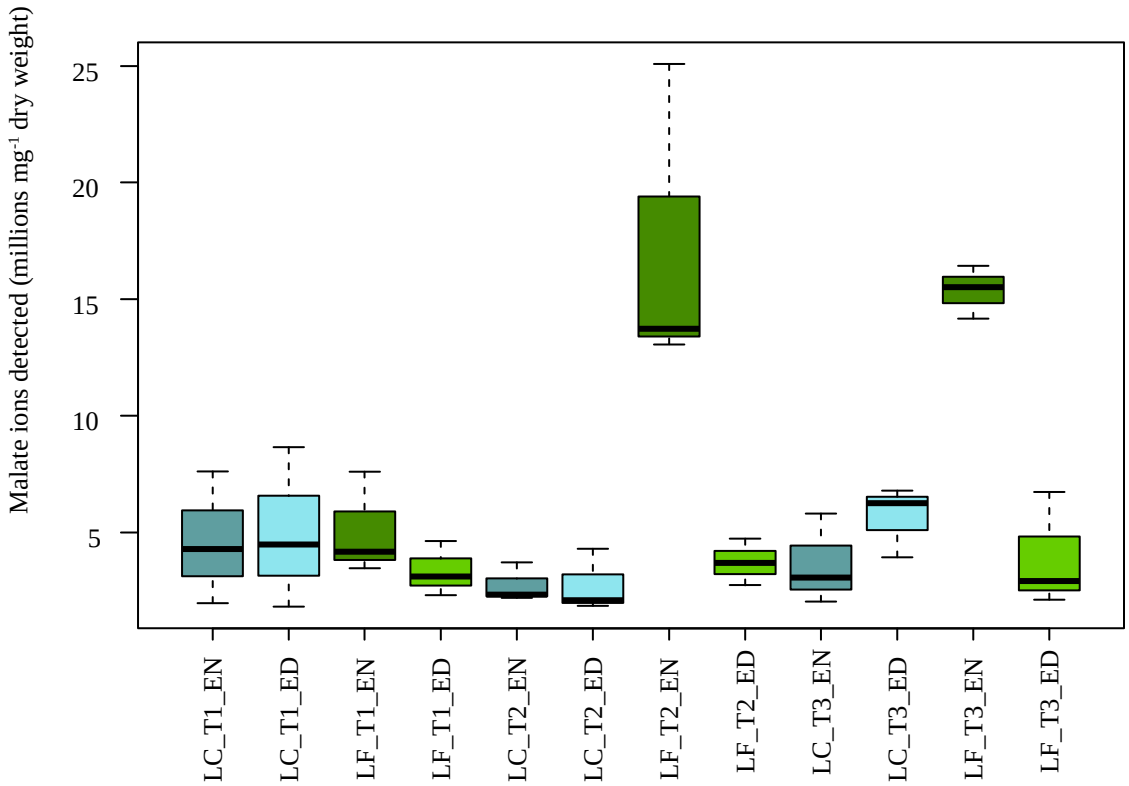


Figure 6.2: Effect of flooding on diurnal malate levels in *Littorella uniflora*

Boxplots for malate levels, displayed in millions of malate ions detected per mg of dry weight, of *L. uniflora* control (C ; blue) and flooded (F; green) plants one hour before dawn (EN) and one hour before dusk (ED) after 1 day (T1), 6 days (T2) and 13 days (T3) of flooding treatment.

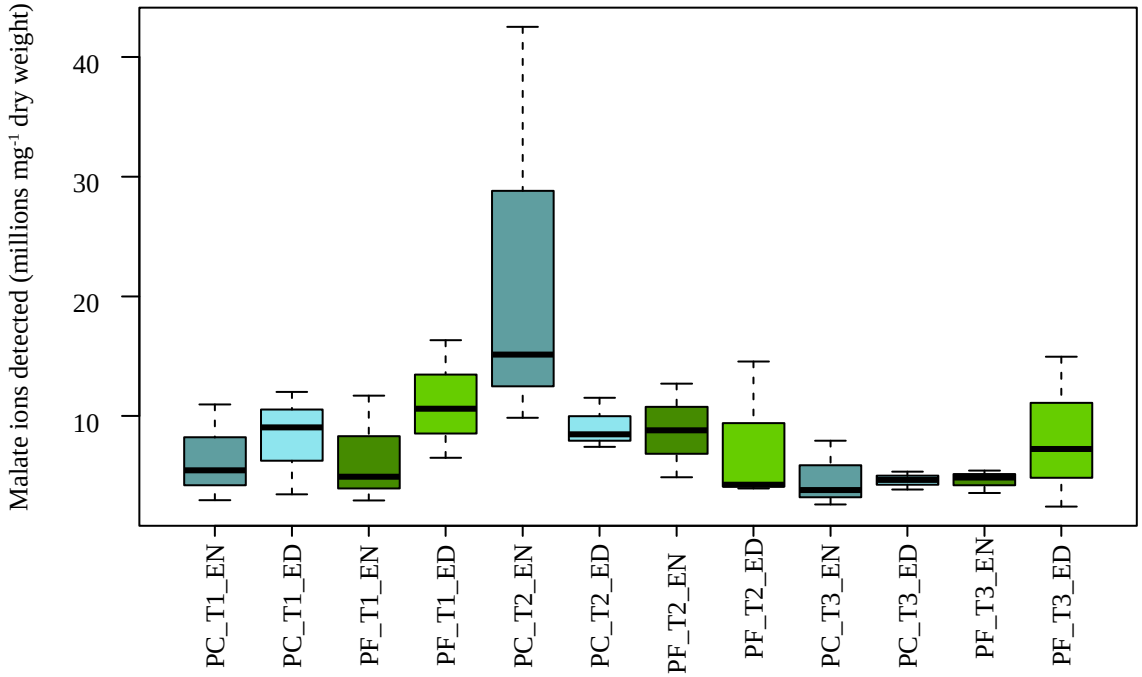


Figure 6.3: Effect of flooding on diurnal malate levels in *Plantago lanceolata*

Boxplots for malate levels, displayed in millions of malate ions detected per mg of dry weight, of *P. lanceolata* control (C ; blue) and flooded (F; green) plants one hour before dawn (EN) and one hour before dusk (ED) after 1 day (T1), 6 days (T2) and 13 days (T3) of flooding treatment.

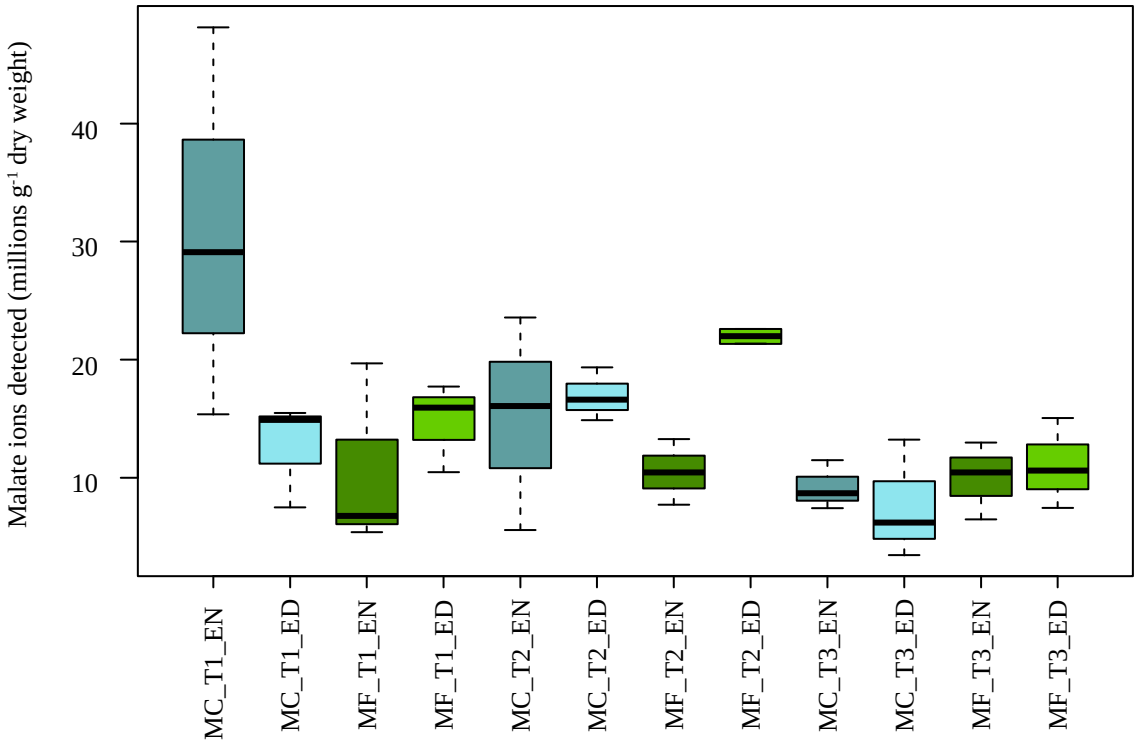


Figure 6.4: Effect of flooding on diurnal malate levels in *Plantago maritima*

Boxplots for malate levels, displayed in millions of malate ions detected per mg of dry weight, of *P. maritima* control (C ; blue) and flooded (F; green) plants one hour before dawn (EN) and one hour before dusk (ED) after 1 day (T1), 6 days (T2) and 13 days (T3) of flooding treatment.

Further testing of multiple closely related C_3 and CAM species, for instance in the Bromeliaceae (Givnish et al., 2014), would provide a more extensive test of the hypothesis outlined by Bräutigam et al., 2017. The *Plantago-Littorella* comparison is probably the best group in which to further investigate my suggested hypothesis, as other lineages of submerged aquatic CAM plants do not have close terrestrial relatives (Chapter III). If CAM did evolve relatively recently in these other groups of plants, as the results of Chapter III suggest, it likely did so after they had evolved a submerged aquatic habit, since they have non-CAM aquatic relatives. This does not rule out a role for this adaptive mechanism in *Litorella*, however, and my hypothesis may be more relevant for plants transitioning to submerged aquatic plants in the low CO_2 atmospheres of the Cenozoic. The current results nevertheless suggest that acid fluctuations and subsequent optimisation leading to CAM evolution arose in the *Littorella* ancestor after the transition to submerged aquatic environments, since its relatives are flood tolerant without any diurnal acid fluctuation.

6.4 Population dynamics in aquatic plants

The evolution of complex traits such as CAM requires the spread of adaptive mutations through populations (Olofsson et al., 2016), a process that is potentially complicated by the fragmented nature of underwater environments. The flooding experiment described above however indicates that aquatic CAM likely evolved in submerged aquatic environments, although there are likely to be exceptions to this, most notably *Crassula*. It is clear that submerged aquatic environments do not prevent evolving a wide range of complex traits such as pressured through-flow systems for gas transport in lilies (Dacey, 1980), radically altered morphologies as in duckweed (Van Hoeck et al., 2015), and underwater pollination systems (Osborn et al., 2001). Nevertheless, the underwater lifestyle potentially creates difficulties in terms of reduced sexual recombination and reduced effective population sizes via habitat fragmentation. My investigation of *I. lacustris* and *L. uniflora* genetic structure in the British Isles (Chapter V) indicates that inherited mating systems can result in differences in gene flow and recombination, with potentially greater levels of sexual reproduction in the spore producing *I. lacustris*. However, putatively greater rates of underwater sex in spore-producing plants such *Isoëtes* clearly did not prevent angiosperms from dominating submerged aquatic habitats (Sculthorpe, 1967) and evolving the complex features mentioned above. However, the diversification of angiosperms in aquatic habitats might have happened despite their mating habits; angiosperms diversified before the secondary colonization of aquatic habitats, and therefore came with a variety of functional traits and an overall genomic diversity. On the other hand, *Isoëtes* diversified only recently (Chapter III), and encompasses a small number of species that are functionally homogeneous. The impact of mating systems might therefore be alleviated by the evolutionary history of the different groups, and targeted studies are needed to disentangle these. The role of sex in the evolution of complexity is not well understood (Luijckx et al., 2017), but I suggest that the variety and plasticity of sexual and asexual reproduction in submerged aquatic plants makes them powerful systems to investigate these effects.

6.5 Genetic enablers of CAM photosynthesis

Whilst shared selective pressures promote the evolution of similar phenotypes in response to similar elements of environmental challenge, the production of appropriate mutations is still necessary for the convergent evolution of traits. The modification of pre-existing components for novel functions is a key source of evolutionary novelty, at many levels of biological organisation (Hanoch and Gerard, 1994). The duplication and subsequent specialisation of gene functions has been thought of as a key step in adaptive novelty (Ohno, 1970; Zhang, 2003), with whole genome and single gene duplications being particularly prevalent in plants (Panchy et al. 2016). Studies of C_4 photosynthesis have identified repeated co-options of certain gene lineages in grasses and Caryophyllales (Christin et al., 2015b) although these differ between the separate groups, with ancestral expression levels being predictive of the genes recruited for C_4 photosynthesis (Moreno-Villena et al., 2018). Despite these differences in suitability of gene recruitment among groups of angiosperms, comparison of transcriptomes across larger evolutionary scales has shown a high level of conservation of the most highly expressed gene lineage for a majority of vascular plant genes despite extensive sequence divergence (Chapter IV). This pattern exists in spite of multiple whole genome duplications and gene family expansions in angiosperms and lycopods in the past 400 million years (Jiao et al., 2011). This suggests that lineage-specific duplications generally do not significantly alter the overall expression pattern of the homolog family, which is consistent with subfunctionalisation of the role of that lineage among descendent gene copies or broad dosage sensitivity of gene copies (Panchy et al., 2016).

The similar expression levels of lineages between divergent lineages is perhaps surprising given the divergent ecologies of the plants, and particularly in the light of large scale transcriptional changes of plants in response to flooding (e.g. Komatsu et al., 2009; Veen et al., 2013; Wu et al., 2017). Gene expression levels are likely to vary significantly between conditions, but the analyses performed in Chapter IV only consider the identity of the most highly abundant leaf transcript within each gene family, without consideration for the absolute expression levels. Transcript abundances are likely to vary through evolutionary time, but my analyses show for the first time that the identity of the most expressed member of each family is conserved across hundreds of millions of years (Chapter IV). Therefore, despite long divergence times, the

differences accumulated between the multiple lineages comprising each gene family do not appear to affect their suitability for expression in different environmental and genetic contexts.

The pattern of conserved gene lineage expression also broadly holds true for genes involved in CAM photosynthesis (Chapter IV). The number of CAM-related genes that is considered is relatively small, because of the relatively small numbers of ancient duplicates of CAM genes and stringent filtering steps. Previous studies of CAM angiosperms show a range of convergent CAM-specific gene properties, such as diurnal expression patterns and convergent sequence evolution (Yang et al., 2017), which may also be present in the instances of CAM in *I. lacustris* and *L. uniflora*. Whilst differences in the expression levels of genes clearly exist between CAM plants, such as the use of different decarboxylating enzymes (Osmond, 1978), our results suggest that these do not involve many changes in the overall expression levels of gene lineages present in the common ancestor of vascular plants (Chapter IV). These results do not rule out significant convergence in the gene expression patterns for some genes of *I. lacustris* and *L. uniflora*, as has been shown for other CAM species (Yang et al., 2017), and indeed the convergent high expression levels of the PCK and the "bacterial-type" PPC indicate likely convergence at the expression level for some individual genes (Chapter IV).

Independently of convergent changes in some genes, the results of Chapter IV suggest that broad patterns of gene expression in plants and their potential for use in new traits such as biochemical carbon concentrating mechanisms were already present in the common ancestor of vascular plants, even if the environmental drivers that produced extant CAM and C_4 plants were not to appear for hundreds of millions of years (Chapters II, III and IV). These results suggest that ancient vascular plants may have been capable of evolving carbon concentrating mechanisms, as has been suggested (Green, 2010; Cowling, 2013). This would likely be contingent on environmental drivers (Chapter III). For example, falling CO_2 conditions in the late Carboniferous (Montañez et al., 2016) may have facilitated the evolution or spread of biochemical carbon concentrating mechansims, with subsequent rising CO_2 levels causing extinction of these lineages (Osborne and Beerling, 2006). Similarly, my results suggest that if more basal vascular lineages had continued their dominance of climates at the expense

of angiosperms, it is likely that pre-existing genetic biases would have prompted the convergent evolution of carbon concentrating mechanisms in groups other than angiosperms in response to falling levels of CO_2 in the past 30 million years.

6.6 On the future of aquatic CAM plants

The increased ecological success of submerged aquatic CAM following historical decreases of CO_2 levels (Chapter III) suggests that rapidly rising CO_2 levels due to human activity constitute a significant threat to submerged aquatic CAM plants. Many *Isoëtes* and isoetid species are increasingly rare, endangered or undergoing range contraction (Brunton and Britton, 1996, Jin-ming et al., 2005, Chen et al., 2007, Free et al., 2009, Kim et al., 2008, Abeli et al. 2017). Whilst their rarefaction is likely due to a number of factors including eutrophication and acidification of oligotrophic lakes (Pedersen et al., 2006; Lucassen et al., 2016) and urban development reducing vernal pool areas (Preisser et al., 2000), rising atmospheric CO_2 levels probably also impact submerged aquatic environments (Short et al., 2016), eroding the competitive advantage of CAM plants over other, faster growing groups of plants (Smolders et al., 2002; Spierenburg et al., 2009, 2010). The insights submerged aquatic CAM plants can provide into the evolution of carbon concentrating mechanisms, and potentially the introduction of these traits into crop plants (Winter and Holtum, 2014), highlight the importance of conserving these systems, and the ecosystems that harbour them.

7. Conclusions

In this dissertation, I adopted aquatic CAM species as a system to understand the environmental and genetic enablers of complex trait evolution. Aquatic CAM evolved independently in the most distantly related vascular plants, therefore enabling comparisons across large evolutionary scales, while allowing comparisons with terrestrial C_3 or CAM relatives. Most CAM research has focused on terrestrial plants, which are widely seen as adapted to the arid or seasonally drought-prone ecosystems that they successfully colonized. The aquatic CAM physiology was therefore seen as an unrelated adaptation to extremely different conditions, with little impact on our understanding of CAM evolution in other groups (Edwards and Ogburn, 2012). At the whole phenotype level these CAM plants are highly divergent, as the succulent CAM species from deserts are obviously functionally distinct from the small isoetid plants thriving in cold, oligotrophic lakes. However, the CAM trait itself is convergent among these groups, and I therefore proposed to consider both aquatic and terrestrial CAM as the same set of adaptations, which can themselves be categorised alongside C_4 plants (Chapter II). Local environmental conditions likely shape the whole organism phenotypes realized in different ecological settings, but based on the literature, I concluded that declining levels of CO_2 were a likely enabler of CAM (and C_4) origins and their later diversification (Chapter II). The most plausible exception to this pattern was *Isoëtes*, which might have been an ancient CAM lineage based on the fossil record and some molecular dating studies. Capitalizing on the latest sequencing technologies, I revisited the age of the group using large numbers of markers from the plastid and nuclear genomes, and showed that extant *Isoëtes* species arose relatively recently, when CO_2 levels were already declining (Chapter III). Therefore, global changes in the composition of the atmosphere likely provided a strong global precondition to the evolution of all carbon concentrating mechanisms.

Environmental drivers are not sufficient to drive the evolution of novel adaptations – suitable components for recruitment are also required. Using transcriptome comparisons across land plants, I demonstrated that the genes co-opted for the CAM cycle were those that were ancestrally the most highly expressed in photosynthetic tissues (Chapter IV). Therefore, genomic changes that occurred more than 400 million years ago determined how novel adaptations would be realized

following global changes in the last 60 million years. This highlights the constraints that dictate evolution. While similar constraints had been reported before in other systems, I provide the first evidence that these can be conserved over hundreds of million years. Besides a pool of suitable mutations, the efficiency of selection, and therefore the ability to acquire novel adaptations, depends on the intraspecific dynamics of accumulation and exchange of mutations. While further studies are needed to study the dynamics of mutations for carbon concentrating mechanisms, my comparisons of two aquatic plants with contrasting mating systems showed that gene flow was more efficient in those reproducing underwater by spores (Chapter V). Based on my results, I further suggest that limited sexual recruitment in the aquatic angiosperms might hamper rapid local adaptation, creating a persistent link between some genetic lineages and environments (Chapter V).

Overall, my work shows that, despite hundred of million years of diversification, all plant lineages are likely to respond to the same environmental challenge in a similar way. Depending on their evolutionary history and genomic background, this will however create diverse whole plant phenotypes, including aquatic CAM plants using a variety of reproductive mechanisms, not to mention the CAM epiphytes and desert succulents, and indeed all the grassland species that use the related C_4 pathway. The fundamental unity of these traits indicates that far from being mere curiosities, consideration of traits in divergent environmental contexts and genetic backgrounds can provide important insights into the evolution of the trait across groups, even if local factors play an important role in integrating the trait into the overall organism phenotype.

References

Abeli, T., Cauzzi, P., Rossi, G., Pistoja, F., and Mucciarelli, M. A gleam of hope for the critically endangered *Isoëtes malinverniana*: Use of small-scale translocations to guide conservation planning (2018). Aquat. Conserv. Mar. Freshw. Ecosyst. 28.2, 501-505.

Abraham, P.E., Yin, H., Borland, A.M., Weighill, D., Lim, S.D., De Paoli, H.C., Engle, N., Jones, P.C., Agh, R., Weston, D.J., et al. (2016). Transcript, protein and metabolite temporal dynamics in the CAM plant Agave. Nat. Plants *2*, 16178.

Ackerman, J.D. (2000). Abiotic pollen and pollination: ecological, functional, and evolutionary perspectives. In Pollen and Pollination, (Springer, Vienna), pp. 167–185.

Aguilar, R., Ashworth, L., Galetto, L., and Aizen, M.A. (2006). Plant reproductive susceptibility to habitat fragmentation: review and synthesis through a meta-analysis. Ecol. Lett. *9*, 968–980.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403–410.

Anderson, C.L., Bremer, K., and Friis, E.M. (2005). Dating phylogenetically basal eudicots using rbcL sequences and multiple fossil reference points. Am. J. Bot. *92*, 1737–1748.

Arakaki, M., Christin, P.-A., Nyffeler, R., Lendel, A., Eggli, U., Ogburn, R.M., Spriggs, E., Moore, M.J., and Edwards, E.J. (2011). Contemporaneous and recent radiations of the world's major succulent plant lineages. Proc. Natl. Acad. Sci. U. S. A. *108*, 8379–8384.

Arendt, J., and Reznick, D. (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends Ecol. Evol. *23*, 26–32.

Arts, G.H.P. (2002). Deterioration of atlantic soft water macrophyte communities by acidification, eutrophication and alkalinisation. Aquat. Bot. *73*, 373–393.

Arts, G.H.P., and van der Heijden, R.A.J.M. (1990). Germination ecology of Littorella uniflora (L.) Aschers. Aquat. Bot. *37*, 139–151.

Ash, S.R., and Pigg, K.B. (1991). A New Jurassic Isoetites (Isoetales) from the Wallowa Terrane in Hells Canyon, Oregon and Idaho. Am. J. Bot. *78*, 1636–1642.

Aubry, S., Brown, N.J., and Hibberd, J.M. (2011). The role of proteins in C_3 plants prior to their recruitment into the C_4 pathway. J. Exp. Bot. *62*, 3049–3059.

Aulio, K. (1985). Differential Expression of Diel Acid Metabolism in Two Life Forms of *Littorella uniflora* (l.) Aschers. New Phytol. *100*, 533–536.

Aulio, K. (1986a). CAM-Like carbon pathway in submerged aquatic plants. Biol. Plant. *28*, 234–236.

Aulio, K. (1986b). CAM-like Photosynthesis in *Littorella uniflora* (L.) Aschers.: The Role of Humidity. Ann. Bot. *58*, 273–275.

Avadhani, P.N., Greenway, H., Lefroy, R., and Prior, L. (1978). Alcoholic Fermentation and Malate Metabolism in Rice Germinating at Low Oxygen Concentrations. Funct. Plant Biol. *5*, 15–25.

Baattrup-Pedersen, A., and Madsen, T.V. (1999). Interdependence of CO_2 and inorganic nitrogen on crassulacean acid metabolism and efficiency of nitrogen use by *Littorella uniflora* (L.) Aschers. Plant Cell Environ. *22*, 535–542.

Bakker, F.T., Lei, D., Yu, J., Mohammadin, S., Wei, Z., van de Kerke, S., Gravendeel, B., Nieuwenhuis, M., Staats, M., Alquezar-Planas, D.E., et al. (2016). Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. Biol. J. Linn. Soc. *117*, 33–43.

von Balthazar, M., Pedersen, K.R., and Friis, E.M. (2005). Teixeiraea lusitanica, a new fossil flower from the Early Cretaceous of Portugal with affinities to Ranunculales. Plant Syst. Evol. *255*, 55–75.

Banach, K., Banach, A.M., Lamers, L.P.M., De Kroon, H., Bennicelli, R.P., Smits, A.J.M., and Visser, E.J.W. (2009). Differences in flooding tolerance between species from two wetland habitats with contrasting hydrology: implications for vegetation development in future floodwater retention areas. Ann. Bot. *103*, 341–351.

Banks, J.A. (2009). Selaginella and 400 Million Years of Separation. Annu. Rev. Plant Biol. *60*, 223–238.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C., Albert, V.A., Aono, N., Aoyama, T., Ambrose, B.A., et al. (2011). The compact Selaginella genome identifies changes in gene content associated with the evolution of vascular plants. Science *332*, 960–963.

Barrett, S.C.H., and Harder, L.D. (2017). The Ecology of Mating and Its Evolutionary Consequences in Seed Plants. Annu. Rev. Ecol. Evol. Syst. *48*, 135–157.

Barrett, S.C.H., Eckert, C.G., and Husband, B.C. (1993). Evolutionary processes in aquatic plant populations. Aquat. Bot. *44*, 105–145.

Bartlett, M.S., Vico, G., and Porporato, A. (2014). Coupled carbon and water fluxes in CAM photosynthesis: modeling quantification of water use efficiency and productivity. Plant Soil *383*, 111–138.

Becks, L., and Agrawal, A.F. (2010). Higher rates of sex evolve in spatially heterogeneous environments. Nature *468*, 89–92.

Behe, M. (1996). Darwin's Black Box: The Biochemical Challenge to Evolution (Simon and Schuster).

Bekker, A., Holland, H.D., Wang, P.-L., Iii, D.R., Stein, H.J., Hannah, J.L., Coetzee, L.L., and Beukes, N.J. (2004). Dating the rise of atmospheric oxygen. Nature *427*, 117-120.

Bell, M.A., Khalef, V., and Travis, M.P. (2007). Directional asymmetry of pelvic vestiges in threespine stickleback. J. Exp. Zoolog. B Mol. Dev. Evol. *308*, 189–199.

Berry, C.M., and Marshall, J.E.A. (2015). Lycopsid forests in the early Late Devonian paleoequatorial zone of Svalbard. Geology *43*, 1043–1046.

Besnard, G., Christin, P.-A., Malé, P.-J.G., Lhuillier, E., Lauzeral, C., Coissac, E., and Vorontsova, M.S. (2014). From museums to genomics: old herbarium specimens shed light on a C_3 to C_4 transition. J. Exp. Bot. *65*, 6711–6721.

Birks, H.H. (2000). Aquatic macrophyte vegetation development in Kråkenes Lake, western Norway, during the late-glacial and early-Holocene. J. Paleolimnol. *23*, 7–19.

Blanquart, F., Kaltz, O., Nuismer, S.L., and Gandon, S. (2013). A practical guide to measuring local adaptation. Ecol. Lett. *16*, 1195–1205.

Blount, Z.D., Borland, C.Z., and Lenski, R.E. (2008). Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. Proc. Natl. Acad. Sci. *105*, 7899–7906.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Bone, R.E., Smith, J.A.C., Arrigo, N., and Buerki, S. (2015). A macro-ecological perspective on crassulacean acid metabolism (CAM) photosynthesis evolution in Afro-Madagascan drylands: Eulophiinae orchids as a case study. New Phytol. *208*, 469–481.

Borgia, G., and Coleman, S.W. (2000). Co–option of male courtship signals from aggressive display in bowerbirds. Proc. R. Soc. Lond. B Biol. Sci. *267*, 1735–1740.

Borland, A.M., Guo, H.-B., Yang, X., and Cushman, J.C. (2016). Orchestration of carbohydrate processing for crassulacean acid metabolism. Curr. Opin. Plant Biol. *31*, 118–124.

Borland, A.M., Griffiths, H., Hartwell, J., and Smith, J.A.C. (2009). Exploiting the potential of plants with crassulacean acid metabolism for bioenergy production on marginal lands. J. Exp. Bot. *60*, 2879–2896.

Borland, A.M., Hartwell, J., Weston, D.J., Schlauch, K.A., Tschaplinski, T.J., Tuskan, G.A., Yang, X., and Cushman, J.C. (2014). Engineering crassulacean acid metabolism to improve water-use efficiency. Trends Plant Sci. *19*, 327–338.

Boston, H.L. (1986). A discussion of the adaptations for carbon acquisition in relation to the growth strategy of aquatic isoetids. Aquat. Bot. *26*, 259–270.

Boston, H.L., and Adams, M.S. (1985). Seasonal diurnal acid rhythms in two aquatic crassulacean acid metabolism plants. Oecologia *65*, 573–579.

Boston, H.L., Adams, M.S., and Pienkowski, T.P. (1987). Utilization of Sediment CO2 by Selected North American Isoetids. Ann. Bot. *60*, 485–494.

Bouchenak-Khelladi, Y., Slingsby, J.A., Verboom, G.A., and Bond, W.J. (2014). Diversification of C_4 grasses (Poaceae) does not coincide with their ecological dominance. Am. J. Bot. *101*, 300–307.

Bousquet, J., Strauss, S.H., Doerksen, A.H., and Price, R.A. (1992). Extensive variation in evolutionary rate of rbcL gene sequences among seed plants. Proc. Natl. Acad. Sci. *89*, 7844–7848.

Bowes, G., and Salvucci, M.E. (1989). Plasticity in the photosynthetic carbon metabolism of submersed aquatic macrophytes. Aquat. Bot. *34*, 233–266.

Bräutigam, A., Schlüter, U., Eisenhut, M., and Gowik, U. (2017). On the Evolutionary Origin of CAM Photosynthesis. Plant Physiol. *174*, 473–477.

Brilhaus, D., Bräutigam, A., Mettler-Altmann, T., Winter, K., and Weber, A.P.M. (2016). Reversible Burst of Transcriptional Changes during Induction of Crassulacean Acid Metabolism in *Talinum triangulare*. Plant Physiol. *170*, 102–122.

Britton, D.M., Brunton, D.F., and Talbot, S.S. (1999). Isoetes in Alaska and the Aleutians. Am. Fern J. *89*, 133–141.

Bromham, L., and Bennett, T.H. (2014). Salt tolerance evolves more frequently in C_4 grass lineages. J. Evol. Biol. *27*, 653–659.

Brower, A.V. (1994). Rapid morphological radiation and convergence among races of the butterfly Heliconius erato inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. *91*, 6491–6495.

Brown, R.H. (1978). A Difference in N Use Efficiency in C_3 and C_4 Plants and its Implications in Adaptation and Evolution. Crop Sci. *18*, 93–98.

Brown, R.H.J. (1963). The Flight of Birds. Biol. Rev. *38*, 460–489.

Brown, N.J., Palmer, B.G., Stanley, S., Hajaji, H., Janacek, S.H., Astley, H.M., Parsley, K., Kajala, K., Quick, W.P., Trenkamp, S., et al. (2010). C_4 acid decarboxylases required for C_4 photosynthesis are active in the mid-vein of the C_3 species *Arabidopsis thaliana*, and are important in sugar and amino acid metabolism. Plant J. *61*, 122–133.

Brown, N.J., Newell, C.A., Stanley, S., Chen, J.E., Perrin, A.J., Kajala, K., and Hibberd, J.M. (2011). Independent and Parallel Recruitment of Preexisting Mechanisms Underlying C_4 Photosynthesis. Science *331*, 1436–1439.

Brunton, D.F. (2001). Quillwort dispersal; which way is the wind blowing? Chinquapin *9*, 20.

Brunton, D.F., and Britton, D.M. (1996). The Status, Distribution, and Identification of Georgia Quillwort (*Isoetes georgiana*; Isoetaceae). Am. Fern J. *86*, 105–113.

Brunton, D.F., and Britton, D.M. (2006). *Isoëtes melanopoda* spp. *silvatica* (subsp. nov.), a New Quillwort (Isoetaceae) From Eastern North America. Castanea *71*, 15–30.

Cantrill, D.J., and Webb, J.A. (1998). Permineralized pleuromeid lycopsid remains from the Early Triassic Arcadia Formation, Queensland, Australia. Rev. Palaeobot. Palynol. *102*, 189–211.

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. Mol. Ecol. *22*, 3124–3140.

Ceballos, G., Ehrlich, P.R., Barnosky, A.D., García, A., Pringle, R.M., and Palmer, T.M. (2015). Accelerated modern human–induced species losses: Entering the sixth mass extinction. Sci. Adv. *1*, e1400253.

Ceusters, J., Borland, A.M., Taybi, T., Frans, M., Godts, C., and De Proft, M.P. (2014). Light quality modulates metabolic synchronization over the diel phases of crassulacean acid metabolism. J. Exp. Bot. *65*, 3705–3714.

Chastain, C.J., Fries, J.P., Vogel, J.A., Randklev, C.L., Vossen, A.P., Dittmer, S.K., Watkins, E.E., Fiedler, L.J., Wacker, S.A., Meinhover, K.C., et al. (2002). Pyruvate,Orthophosphate Dikinase in Leaves and Chloroplasts of C_3 Plants Undergoes Light-/Dark-Induced Reversible Phosphorylation. Plant Physiol. *128*, 1368–1378.

Chen, J., Gituru, W.R., Liu, X., and Wang, Q. (2007). Genetic diversity in *Isoetes yunguiensis*, a rare and endangered endemic fern in China. Front. Biol. China *2*, 46–49.

Christiansen, R., Skovmand Friis, N.J., and Søndergaard, M. (1985). Leaf production and nitrogen and phosphorus tissue content of *Littorella uniflora* (L.) Aschers. In

relation to nitrogen and phosphorus enrichment of the sediment in oligotrophic Lake Hampen, Denmark. Aquat. Bot. *23*, 1–11.

Christin, P.-A., and Osborne, C.P. (2013a). The recurrent assembly of C_4 photosynthesis, an evolutionary tale. Photosynth. Res. *117*, 163–175.

Christin, P.-A., and Osborne, C.P. (2014). The evolutionary ecology of C_4 plants. New Phytol. *204*, 765–781.

Christin, P.-A., Salamin, N., Savolainen, V., Duvall, M.R., and Besnard, G. (2007). C_4 Photosynthesis Evolved in Grasses via Parallel Adaptive Genetic Changes. Curr. Biol. *17*, 1241–1247.

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M.R., Hodkinson, T.R., Savolainen, V., and Salamin, N. (2008). Oligocene CO_2 decline promoted C_4 photosynthesis in grasses. Curr. Biol. CB *18*, 37–43.

Christin, P.-A., Samaritani, E., Petitpierre, B., Salamin, N., and Besnard, G. (2009). Evolutionary Insights on C_4 Photosynthetic Subtypes in Grasses from Genomics and Phylogenetics. Genome Biol. Evol. *1*, 221–230.

Christin, P.-A., Weinreich, D.M., and Besnard, G. (2010). Causes and evolutionary significance of genetic convergence. Trends Genet. *26*, 400–405.

Christin, P.-A., Osborne, C.P., Sage, R.F., Arakaki, M., and Edwards, E.J. (2011a). C_4 eudicots are not younger than C_4 monocots. J. Exp. Bot. *62*, 3171–3181.

Christin, P.-A., Sage, T.L., Edwards, E.J., Ogburn, R.M., Khoshravesh, R., and Sage, R.F. (2011b). Complex Evolutionary Transitions and the Significance of C_3–C_4 Intermediate Forms of Photosynthesis in Molluginaceae. Evolution *65*, 643–660.

Christin, P.-A., Boxall, S.F., Gregory, R., Edwards, E.J., Hartwell, J., and Osborne, C.P. (2013). Parallel Recruitment of Multiple Genes into C_4 Photosynthesis. Genome Biol. Evol. *5*, 2174–2187.

Christin, P.-A., Spriggs, E., Osborne, C.P., Strömberg, C.A.E., Salamin, N., and Edwards, E.J. (2014). Molecular Dating, Evolutionary Rates, and the Age of the Grasses. Syst. Biol. *63*, 153–165.

Christin, P.-A., Arakaki, M., Osborne, C.P., and Edwards, E.J. (2015). Genetic Enablers Underlying the Clustered Evolutionary Origins of C_4 Photosynthesis in Angiosperms. Mol. Biol. Evol. *32*, 846–858.

Clayton, H., Saladié, M., Rolland, V., Sharwood, R., Macfarlane, T., and Ludwig, M. (2017). Loss of the Chloroplast Transit Peptide from an Ancestral C_3 Carbonic

Anhydrase Is Associated with C_4 Evolution in the Grass Genus Neurachne. Plant Physiol. *173*, 1648–1658.

Cock Burn, W. (1981). The evolutionary relationship between stomatal mechanism, crassulacean acid metabolism and C_4 photosynthesis. Plant Cell Environ. *4*, 417–418.

Cockburn, W. (1979). Relationships between Stomatal Behavior and Internal Carbon Dioxide Concentration in Crassulacean Acid Metabolism Plants. Plant Physiol. *63*, 1029–1032.

Cole, J.J., Caraco, N.F., Kling, G.W., and Kratz, T.K. (1994). Carbon Dioxide Supersaturation in the Surface Waters of Lakes. Science *265*, 1568–1570.

Comins, H.N., and Farquhar, G.D. (1982). Stomatal regulation and water economy in crassulacean acid metabolism plants: an optimization model. J. Theor. Biol. *99*, 263–284.

Conte, G.L., Arnegard, M.E., Peichel, C.L., and Schluter, D. (2012). The probability of genetic parallelism and convergence in natural populations. Proc R Soc B *279*, 5039–5047.

Cottrell, J.E., Munro, R.C., Tabbener, H.E., Gillies, A.C.M., Forrest, G.I., Deans, J.D., and Lowe, A.J. (2002). Distribution of chloroplast DNA variation in British oaks (*Quercus robur* and *Q. petraea*): the influence of postglacial colonisation and human management. For. Ecol. Manag. *156*, 181–195.

Cowling, S.A. (2013). Did early land plants use carbon-concentrating mechanisms? Trends Plant Sci. *18*, 120–124.

Cox, P.A. (1988). Hydrophilous Pollination. Annu. Rev. Ecol. Syst. *19*, 261–279.

Crawford, R.M.M., and McManmon, M. (1968). Inductive Responses of Alcohol and Malic Dehydrogenases in Relation to Flooding Tolerance in Roots. J. Exp. Bot. *19*, 435–441.

Crawford, R.M.M., and Zochowski, Z.M. (1984). Tolerance of Anoxia and Ethanol Toxicity in Chickpea seedlings (*Cicer arietinum* L.). J. Exp. Bot. *35*, 1472–1480.

Crayn, D.M., Winter, K., and Smith, J.A.C. (2004). Multiple origins of crassulacean acid metabolism and the epiphytic habit in the Neotropical family Bromeliaceae. Proc. Natl. Acad. Sci. U. S. A. *101*, 3703–3708.

Cressler, W.L., and Pfefferkorn, H.W. (2005). A Late Devonian isoetalean lycopsid, *Otzinachsonia beerboweri*, gen. et sp. nov., from north-central Pennsylvania, USA. Am. J. Bot. *92*, 1131–1140.

Cushman, J.C., and Bohnert, H.J. (1999). Crassulacean Acid Metabolism: Molecular Genetics. Annu. Rev. Plant Physiol. Plant Mol. Biol. *50*, 305–332.\

Cushman, J.C., Tillett, R.L., Wood, J.A., Branco, J.M., and Schlauch, K.A. (2008). Large-scale mRNA expression profiling in the common ice plant, *Mesembryanthemum crystallinum*, performing C_3 photosynthesis and Crassulacean acid metabolism (CAM). J. Exp. Bot. *59*, 1875–1894.

Dacey, J.W. (1980). Internal winds in water lilies: an adaptation for life in anaerobic sediments. Science *210*, 1017–1019.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

Darwin, C. (1859). On the Origin of the Species by Means of Natural Selection: Or, The Preservation of Favoured Races in the Struggle for Life (John Murray).

Decker, J.E., and de Wit, M.J. (2006). Carbon isotope evidence for CAM photosynthesis in the Mesozoic. Terra Nova *18*, 9–17.

Deng, H., Zhang, L.-S., Zhang, G.-Q., Zheng, B.-Q., Liu, Z.-J., and Wang, Y. (2016). Evolutionary history of PEPC genes in green plants: Implications for the evolution of CAM in orchids. Mol. Phylogenet. Evol. *94*, 559–564.

Dengler, N.G., Dengler, R.E., Donnelly, P.M., and Hattersley, P.W. (1994). Quantitative Leaf Anatomy of C_3 and C_4 Grasses (Poaceae): Bundle Sheath and Mesophyll Surface Area Relationships. Ann. Bot. *73*, 241–255.

Der, J.P., Barker, M.S., Wickett, N.J., dePamphilis, C.W., and Wolf, P.G. (2011). De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. BMC Genomics *12*, 99.

Dierckxsens, N., Mardulyn, P., and Smits, G. (2017). NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. *45*, e18–e18.

DiMario, R.J., Clayton, H., Mukherjee, A., Ludwig, M., and Moroney, J.V. (2017). Plant Carbonic Anhydrases: Structures, Locations, Evolution, and Physiological Roles. Mol. Plant *10*, 30–46.

Dodd, A.N., Borland, A.M., Haslam, R.P., Griffiths, H., and Maxwell, K. (2002). Crassulacean acid metabolism: plastic, fantastic. J. Exp. Bot. *53*, 569–580.

Dong, Y.-H., Chen, J.-M., Gituru, R.W., and Wang, Q.-F. (2007). Gene flow in populations of the endangered aquatic fern *Ceratopteris pteridoides* in China as revealed by ISSR markers. Aquat. Bot. *87*, 69–74.

Doniger, S.W., and Fay, J.C. (2007). Frequent Gain and Loss of Functional Transcription Factor Binding Sites. PLOS Comput. Biol. *3*, e99.

Drummond, A.J., and Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. *7*, 214.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., and Rambaut, A. (2006). Relaxed Phylogenetics and Dating with Confidence. PLOS Biol. *4*, e88.

Duigan, C., Kovach, W., and Palmer, M. (2007). Vegetation communities of British lakes: a revised classification scheme for conservation. Aquat. Conserv. Mar. Freshw. Ecosyst. *17*, 147–173.

Dunning, L.T., Lundgren, M.R., Moreno-Villena, J.J., Namaganda, M., Edwards, E.J., Nosil, P., Osborne, C.P., and Christin, P.-A. (2017). Introgression and repeated co-option facilitated the recurrent emergence of C_4 photosynthesis among close relatives. Evolution *71*, 1541–1555.

Duvall, M.R., Saar, D.E., Grayburn, W.S., and Holbrook, G.P. (2003). Complex Transitions between C_3 and C_4 Photosynthesis during the Evolution of Paniceae: A Phylogenetic Case Study Emphasizing the Position of *Steinchisma hians* (Poaceae), a C_3-C_4 Intermediate. Int. J. Plant Sci. *164*, 949–958.

Eaton, D.A.R. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. Bioinformatics *30*, 1844–1849.

Eckert, C.G., Kalisz, S., Geber, M.A., Sargent, R., Elle, E., Cheptou, P.-O., Goodwillie, C., Johnston, M.O., Kelly, J.K., Moeller, D.A., et al. (2010). Plant mating systems in a changing world. Trends Ecol. Evol. *25*, 35–43.

Edwards, E.J., and Ogburn, R.M. (2012). Angiosperm Responses to a Low-CO_2 World: CAM and C_4 Photosynthesis as Parallel Evolutionary Trajectories. Int. J. Plant Sci. *173*, 724–733.

Edwards, E.J., and Smith, S.A. (2010). Phylogenetic analyses reveal the shady history of C_4 grasses. Proc. Natl. Acad. Sci. *107*, 2532–2537.

Edwards, E.J., Osborne, C.P., Strömberg, C.A.E., Smith, S.A., and Consortium, C.G. (2010). The Origins of C_4 Grasslands: Integrating Evolutionary and Ecosystem Science. Science *328*, 587–591.

Edwards, G.E., Franceschi, V.R., and Voznesenskaya, E.V. (2004). Single-cell C_4 photosynthesis versus the dual-cell (Kranz) paradigm. Annu. Rev. Plant Biol. *55*, 173–196.

Ehleringer, J., and Björkman, O. (1977). Quantum Yields for CO_2 Uptake in C_3 and C_4 Plants: Dependence on Temperature, CO_2, and O_2 Concentration. Plant Physiol. *59*, 86–90.

Ehleringer, J.R., and Monson, R.K. (1993). Evolutionary and Ecological Aspects of Photosynthetic Pathway Variation. Annu. Rev. Ecol. Syst. *24*, 411–439.

Ehleringer, J.R., Cerling, T.E., and Helliker, B.R. (1997). C_4 photosynthesis, atmospheric CO_2, and climate. Oecologia *112*, 285–299.

Ellison, A.M., and Gotelli, N.J. (2001). Evolutionary ecology of carnivorous plants. Trends Ecol. Evol. *16*, 623–629.

Emms, D.M., Covshoff, S., Hibberd, J.M., and Kelly, S. (2016). Independent and Parallel Evolution of New Genes by Gene Duplication in Two Origins of C_4 Photosynthesis Provides New Insight into the Mechanism of Phloem Loading in C_4 Species. Mol. Biol. Evol. *33*, 1796–1806.

Farmer, A.M., and Spence, D.H.N. (1985). Studies of Diurnal Acid Fluctuations in British Isoetid-type Submerged Aquatic Macrophytes. Ann. Bot. *56*, 347–350.

Farmer, A.M., and Spence, D.H.N. (1986). The growth strategies and distribution of isoetids in Scottish freshwater lochs. Aquat. Bot. *26*, 247–258.

Farquhar, G.D., Dubbe, D.R., and Raschke, K. (1978). Gain of the Feedback Loop Involving Carbon Dioxide and Stomata. Theory and Measurement. Plant Physiol. *62*, 406–412.

Felsenstein, J. (2002). {PHYLIP} (Phylogeny Inference Package) version 3.6a3.

Feuk, L., Carson, A.R., and Scherer, S.W. (2006). Structural variation in the human genome. Nat. Rev. Genet. *7*, 85–97.

Fisher, A.E., Hasenstab, K.M., Bell, H.L., Blaine, E., Ingram, A.L., and Columbus, J.T. (2016). Evolutionary history of chloridoid grasses estimated from 122 nuclear loci. Mol. Phylogenet. Evol. *105*, 1–14.

Foster, G.L., Royer, D.L., and Lunt, D.J. (2017). Future climate forcing potentially without precedent in the last 420 million years. Nat. Commun. *8*, 14845.

Free, G., Bowman, J., McGarrigle, M., Caroni, R., Donnelly, K., Tierney, D., Trodd, W., and Little, R. (2009). The identification, characterization and conservation value of isoetid lakes in Ireland. Aquat. Conserv. Mar. Freshw. Ecosyst. *19*, 264–273.

Gacia, E., and Ballesteros, E. (1994). Production of *Isoetes lacustris* in a Pyrenean lake: seasonality and ecological factors involved in the growing period. Aquat. Bot. *48*, 77–89.

Garant, D., Forde, S.E., and Hendry, A.P. (2007). The multifarious effects of dispersal and gene flow on contemporary adaptation. Funct. Ecol. *21*, 434–443.

Garratt, M.J., Tims, J.D., Rickards, R.B., Chambers, T.C., and Douglas, J.G. (1984). The appearance of Baragwanathia (Lycophytina) in the Silurian. Bot. J. Linn. Soc. *89*, 355–358.

Gennidakis, S., Rao, S., Greenham, K., Uhrig, R.G., O'Leary, B., Snedden, W.A., Lu, C., and Plaxton, W.C. (2007). Bacterial- and plant-type phosphoenolpyruvate carboxylase polypeptides interact in the hetero-oligomeric Class-2 PEPC complex of developing castor oil seeds. Plant J. Cell Mol. Biol. *52*, 839–849.

Giordano, M., Beardall, J., and Raven, J.A. (2005). CO_2 concentrating mechanisms in algae: mechanisms, environmental modulation, and evolution. Annu. Rev. Plant Biol. *56*, 99–131.

Giussani, L.M., Cota-Sánchez, J.H., Zuloaga, F.O., and Kellogg, E.A. (2001). A molecular phylogeny of the grass subfamily Panicoideae (Poaceae) shows multiple origins of C_4 photosynthesis. Am. J. Bot. *88*, 1993–2012.

Givnish, T.J., Burkhardt, E.L., Happel, R.E., and Weintraub, J.D. (1984). Carnivory in the Bromeliad *Brocchinia reducta*, with a Cost/Benefit Model for the General Restriction of Carnivorous Plants to Sunny, Moist, Nutrient-Poor Habitats. Am. Nat. *124*, 479–497.

Givnish, T.J., Barfuss, M.H.J., Ee, B.V., Riina, R., Schulte, K., Horres, R., Gonsiska, P.A., Jabaily, R.S., Crayn, D.M., Smith, J.A.C., et al. (2014). Adaptive radiation, correlated and contingent evolution, and net species diversification in Bromeliaceae. Mol. Phylogenet. Evol. *71*, 55–78.

Givnish, T.J., Spalink, D., Ames, M., Lyon, S.P., Hunter, S.J., Zuluaga, A., Iles, W.J.D., Clements, M.A., Arroyo, M.T.K., Leebens-Mack, J., et al. (2015). Orchid phylogenomics and multiple drivers of their extraordinary diversification. Proc. Biol. Sci. *282*.

Godwin (1984). History of the British Flora (Cambridge University Press).

Gompel, N., and Prud'homme, B. (2009). The causes of repeated genetic evolution. Dev. Biol. *332*, 36–47.

Gould, S.J. (1990). Wonderful Life: The Burgess Shale and the Nature of History (W. W. Norton, New York City).

Gould, S.J., and Vrba, E.S. (1982). Exaptation—a Missing Term in the Science of Form. Paleobiology *8*, 4–15.

Granoff, J.A., Gensel, P.G., and Andrews, H.N. (1976). A new species of *Pertica* from the Devonian of Eastern Canada. Palaeontogr. Abt. B 119–128.

Grass Phylogeny Working Group II. New grass phylogeny resolves deep evolutionary relationships and discovers C_4 origins. *New. Phyt.* 193(2). 304-312.

Graur, D., Zheng, Y., and Azevedo, R.B.R. (2015). An Evolutionary Classification of Genomic Function. Genome Biol. Evol. *7*, 642–645.

Greb, S.F., and DiMichele, W.A. (2006). Evolution and importance of wetlands in earth history. In Wetlands through Time, Geological Society of America Special Paper 399, p. 1-40.

Green, W.A. (2010). The function of the aerenchyma in arborescent lycopsids: evidence of an unfamiliar metabolic strategy. Proc. Biol. Sci. *277*, 2257–2267.

Griffiths, H. (1989). Carbon Dioxide Concentrating Mechanisms and the Evolution of CAM in Vascular Epiphytes. In Vascular Plants as Epiphytes, (Springer, Berlin, Heidelberg), pp. 42–86.

Griffiths, H. (1992). Carbon isotope discrimination and the integration of carbon assimilation pathways in terrestrial CAM plants. Plant Cell Environ. *15*, 1051–1062.

Griswold, C.K. (2015). Additive genetic variation and evolvability of a multivariate trait can be increased by epistatic gene action. J. Theor. Biol. *387*, 241–257.

Groenhof, A.C., Smirnoff, N., and Bryant, J.A. (1988). Enzymic Activities Associated with the Ability of Aerial and Submerged Forms of *Littorella uniflora* (L.) Aschers to perform CAM. J. Exp. Bot. *39*, 353–361.

Gross, J.B., Borowsky, R., and Tabin, C.J. (2009). A Novel Role for Mc1r in the Parallel Evolution of Depigmentation in Independent Populations of the Cavefish *Astyanax mexicanus*. PLOS Genet. *5*, e1000326.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence reconstruction from RNA-Seq: reference generation and analysis with Trinity. Nat. Protoc. 8, p1494-1512.

Habets, M.G.J.., Czárán, T., Hoekstra, R.F., and de Visser, J.A.G.. (2007). Spatial structure inhibits the rate of invasion of beneficial mutations in asexual populations. Proc. R. Soc. B Biol. Sci. *274*, 2139–2143.

Hague, M.T.J., Feldman, C.R., Brodie, E.D., and Brodie, E.D. (2017). Convergent adaptation to dangerous prey proceeds through the same first-step mutation in the garter snake *Thamnophis sirtalis*. Evolution *71*, 1504–1518.

Han, H., and Felker, P. (1997). Field validation of water-use efficiency of the CAM plant *Opuntia ellisianain* south Texas. J. Arid Environ. *36*, 133–148.

Hanoch, G., and Gerard, T. (1994). Biology And Computation: A Physicist's Choice (World Scientific).

Hansen, J., Sato, M., Kharecha, P., Beerling, D., Berner, R., Masson-Delmotte, V., Pagani, M., Raymo, M., Royer, D.L., and Zachos, J.C. (2008). Target atmospheric CO_2: Where should humanity aim? Open Atmospheric Sci. J. *2*, 217–231.

Hargreaves, A.D., Swain, M.T., Hegarty, M.J., Logan, D.W., and Mulley, J.F. (2014). Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. Genome Biol. Evol. *6*, 2088–2095.

Hartwell, J., Gill, A., Nimmo, G.A., Wilkins, M.B., Jenkins, G.I., and Nimmo, H.G. (1999). Phosphoenolpyruvate carboxylase kinase is a novel protein kinase regulated at the level of expression. Plant J. Cell Mol. Biol. *20*, 333–342.

Hartwell James, Smith Lucy H., Wilkins Malcolm B., Jenkins Gareth I., and Nimmo Hugh G. (1996). Higher plant phosphoenolpyruvate carboxylase kinase is regulated at the level of translatable mRNA in response to light or a circadian rhythm. Plant J. *10*, 1071–1078.

Harvey, P.H., and Pagel, M.D. (1998). The Comparative Method in Evolutionary Biology (Oxford University Press).

Hatch, M.D. (1987). C_4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. Biochim. Biophys. Acta BBA - Rev. Bioenerg. *895*, 81–106.

Hatch, M.D., and Osmond, C.B. (1976). Compartmentation and Transport in C_4 Photosynthesis. In Transport in Plants III, (Springer, Berlin, Heidelberg), pp. 144–184.

Haydon, M.J., Mielczarek, O., Robertson, F.C., Hubbard, K.E., and Webb, A.A.R. (2013). Photosynthetic entrainment of the *Arabidopsis thaliana* circadian clock. Nature *502*, 689.

Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A.P.M., and Lercher, M.J. (2013). Predicting C_4 photosynthesis evolution: modular, individually adaptive steps on a Mount Fuji fitness landscape. Cell *153*, 1579–1588.

Hedrén, M. (2009). Plastid DNA haplotype variation in *Dactylorhiza incarnata* (Orchidaceae): evidence for multiple independent colonization events into Scandinavia. Nord. J. Bot. *27*, 69–80.

Herrera, A. (2009). Crassulacean acid metabolism and fitness under water deficit stress: if not for carbon gain, what is facultative CAM good for? Ann. Bot. *103*, 645–653.

Heyduk, K., McKain, M.R., Lalani, F., and Leebens-Mack, J. (2016). Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae (Asparagaceae). Mol. Phylogenet. Evol. *105*, 102–113.

Hibberd, J.M., and Quick, W.P. (2002). Characteristics of C_4 photosynthesis in stems and petioles of C_3 flowering plants. Nature *415*, 451.

Hoarau, G., Coyer, J.A., Veldsink, J.H., Stam, W.T., and Olsen, J.L. (2007). Glacial refugia and recolonization pathways in the brown seaweed *Fucus serratus*. Mol. Ecol. *16*, 3606–3616.

Hoffman, L.A., and Tomescu, A.M.F. (2013). An early origin of secondary growth: Franhueberia gerriennei gen. et sp. nov. from the Lower Devonian of Gaspé (Quebec, Canada). Am. J. Bot. *100*, 754–763.

Hofstra, D., and de Winton, M. (2016). Geographic scales of genetic variation amongst *Isoëtes* in New Zealand. Aquat. Bot. *131*, 28–37.

Hoggard, R.K., Kores, P.J., Molvray, M., Hoggard, G.D., and Broughton, D.A. (2003). Molecular systematics and biogeography of the amphibious genus *Littorella* (Plantaginaceae). Am. J. Bot. *90*, 429–435.

Holgerson, M.A. (2015). Drivers of carbon dioxide and methane supersaturation in small, temporary ponds. Biogeochemistry *124*, 305–318.

Holsinger, K.E. (2000). Reproductive systems and evolution in vascular plants. Proc. Natl. Acad. Sci. *97*, 7037–7042.

Hoot, S.B., Taylor, W.C., and Napier, N.S. (2006). Phylogeny and Biogeography of Isoëtes (Isoëtaceae) Based on Nuclear and Chloroplast DNA Sequence Data. Syst. Bot. *31*, 449–460.

Horn, J.W., Xi, Z., Riina, R., Peirson, J.A., Yang, Y., Dorsey, B.L., Berry, P.E., Davis, C.C., and Wurdack, K.J. (2014). Evolutionary bursts in *Euphorbia* (Euphorbiaceae) are linked with photosynthetic pathway. Evol. Int. J. Org. Evol. *68*, 3485–3504.

Huang, R., O'Donnell, A.J., Barboline, J.J., and Barkman, T.J. (2016). Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. Proc. Natl. Acad. Sci. *113*, 10613–10618.

Huang, Y.-X., Yin, Y.-G., Sanuki, A., Fukuda, N., Ezura, H., and Matsukura, C. (2015). Phosphoenolpyruvate carboxykinase (PEPCK) deficiency affects the germination, growth and fruit sugar content in tomato (*Solanum lycopersicum* L.). Plant Physiol. Biochem. *96*, 417–425.

Hubbard, J.K., Uy, J.A.C., Hauber, M.E., Hoekstra, H.E., and Safran, R.J. (2010). Vertebrate pigmentation: from underlying genes to adaptive function. Trends Genet. *26*, 231–239.

Hunt, K.D. (1994). The evolution of human bipedality: ecology and functional morphology. J. Hum. Evol. *26*, 183–202.

Hutin, C., Nussaume, L., Moise, N., Moya, I., Kloppstech, K., and Havaux, M. (2003). Early light-induced proteins protect Arabidopsis from photooxidative stress. Proc. Natl. Acad. Sci. *100*, 4921–4926.

Hutsemekers, V., Hardy, O.J., Mardulyn, P., Shaw, A.J., and Vanderpoorten, A. (2010). Macroecological patterns of genetic structure and diversity in the aquatic moss *Platyhypnidium riparioides*. New Phytol. *185*, 852–864.

Hutsemékers, V., Hardy, O.J., and Vanderpoorten, A. (2013). Does water facilitate gene flow in spore-producing plants? Insights from the fine-scale genetic structure of the aquatic moss *Rhynchostegium riparioides* (Brachytheciaceae). Aquat. Bot. *108*, 1–6.

Ibrahim, D.G., Burke, T., Ripley, B.S., and Osborne, C.P. (2009). A molecular phylogeny of the genus *Alloteropsis* (Panicoideae, Poaceae) suggests an evolutionary reversion from C_4 to C_3 photosynthesis. Ann. Bot. *103*, 127–136.

Ikeda, M., and Yamada, Y. (1981). Dark CO_2 fixation in leaves of tomato plants grown with ammonium and nitrate as nitrogen sources. Plant Soil *60*, 213–222.

Jerling, L. (1981). Effects of Microtopography on the Summer Survival of *Plantago maritima* Seedlings. Holarct. Ecol. *4*, 120–126.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. Nature *473*, 97-100.

Jiménez-Mejías, P., Luceño, M., Lye, K.A., Brochmann, C., and Gussarova, G. (2012). Genetically diverse but with surprisingly little geographical structure: the complex history of the widespread herb Carex nigra (Cyperaceae). J. Biogeogr. *39*, 2279–2291.

Jin-ming, C., Jing-yuan, W., Xing, L., Robert, G.W., and Wing-feng, W. (2005). RAPD analysis for genetic variation within the endangered quillwort *Isoetes hypsophila* (Isoetaceae). Wuhan Univ. J. Nat. Sci. *10*, 455–459.

Joly, C.A. (1994). Flooding tolerance: a reinterpretation of Crawford's metabolic theory. Proc. R. Soc. Edinb. Sect. B Biol. Sci. *102*, 343–354.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics *24*, 1403–1405.

Jordan, D.B., and Ogren, W.L. (1981). Species variation in the specificity of ribulose biphosphate carboxylase/oxygenase. Nature *291*, 513-515.

Jump, A.S., and Peñuelas, J. (2005). Running to stand still: adaptation and the response of plants to rapid climate change. Ecol. Lett. *8*, 1010–1020.

Kadereit, G., Borsch, T., Weising, K., and Freitag, H. (2003). Phylogeny of Amaranthaceae and Chenopodiaceae and the Evolution of C_4 Photosynthesis. Int. J. Plant Sci. *164*, 959–986.

Kadereit, G., Ackerly, D., and Pirie, M.D. (2012). A broader model for C_4 photosynthesis evolution in plants inferred from the goosefoot family (Chenopodiaceae s.s.). Proc. R. Soc. Lond. B Biol. Sci. 279(1741) 3304-3311.

Kanai, R., and Edwards, G.E. (1999). The biochemistry of C_4 photosynthesis. In C_4 Plant Biology, (Academic Press, San Diego), 49-80.

Karol, K.G., Arumuganathan, K., Boore, J.L., Duffy, A.M., Everett, K.D., Hall, J.D., Hansen, S.K., Kuehl, J.V., Mandoli, D.F., Mishler, B.D., et al. (2010). Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol. Biol. *10*, 321.

Kasting, J.F. (1993). Earth's Early Atmosphere. Science *259*, 920–926.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–3066.

Kaufman, A.J., and Xiao, S. (2003). High CO_2 levels in the Proterozoic atmosphere estimated from analyses of individual microfossils. Nature *425*, 279-282.

Kawecki, T.J., and Mery, F. (2006). Genetically idiosyncratic responses of *Drosophila melanogaster* populations to selection for improved learning ability. J. Evol. Biol. *19*, 1265–1274.

Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I., and Whitlock, M.C. (2012). Experimental evolution. Trends Ecol. Evol. *27*, 547–560.

Keeley, J.E. (1981). *Isoetes howellii*: A Submerged Aquatic CAM Plant? Am. J. Bot. *68*, 420–424.

Keeley, J.E. (1982). Distribution of Diurnal Acid Metabolism in the Genus Isoetes. Am. J. Bot. *69*, 254–257.

Keeley, J.E. (1998). CAM photosynthesis in submerged aquatic plants. Bot. Rev. *64*, 121–175.

Keeley, J.E. (2014). Aquatic CAM photosynthesis: a brief history of its discovery. Aquat. Bot. *118*, 7.

Keeley, J.E., and Busch, G. (1984). Carbon Assimilation Characteristics of the Aquatic CAM Plant, Isoetes howellii. Plant Physiol. *76*, 525–530.

Keeley, J.E., and Rundel, P.W. (2003). Evolution of CAM and C_4 Carbon-Concentrating Mechanisms. Int. J. Plant Sci. *164*, S55–S77.

Keeley, J.E., Mathews, R.P., and Walker, C.M. (1983a). Diurnal Acid Metabolism in Isoetes howellii from a Temporary Pool and a Permanent Lake. Am. J. Bot. *70*, 854–857.

Keeley, J.E., Walker, C.M., and Mathews, R.P. (1983b). Crassulacean acid metabolism in *Isoëtes bolanderi* in high elevation oligotrophic lakes. Oecologia *58*, 63–69.

Kenrick, P., and Crane, P.R. (1997). The origin and early evolution of plants on land. Nature *389*, 33-39.

Kenrick, P. (2003). Palaeobotany: fishing for the first plants. Nature *425*, 248–249.

Kim, C., and Choi, H.-K. (2016). Biogeography of North Pacific *Isoëtes* (Isoëtaceae) inferred from nuclear and chloroplast DNA sequence data. J. Plant Biol. *59*, 386–396.

Kim, C., Na, H.R., and Choi, H.-K. (2008). Genetic diversity and population structure of endangered *Isoëtes coreana* in South Korea based on RAPD analysis. Aquat. Bot. *89*, 43–49.

Kim, C., Shin, H., and Choi, H. (2009). Genetic Diversity and Population Structure of Diploid and Polyploid Species of *Isoëtes* in East Asia Based on Amplified Fragment Length Polymorphism Markers. Int. J. Plant Sci. *170*, 496–504.

King, B., and Lee, M.S.Y. (2015). Ancestral State Reconstruction, Rate Heterogeneity, and the Evolution of Reptile Viviparity. Syst. Biol. *64*, 532–544.

Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., and Fitch, D.H.A. (2004). Caenorhabditis phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc. Natl. Acad. Sci. U. S. A. *101*, 9003–9008.

Klavsen, S.K., and Maberly, S.C. (2010). Effect of light and CO_2 on inorganic carbon uptake in the invasive aquatic CAM-plant *Crassula helmsii*. Funct. Plant Biol. *37*, 737–747.

Kluge, M., and Ting, I.P. (2012). Crassulacean Acid Metabolism: Analysis of an Ecological Adaptation (Springer Science & Business Media, New York).

Kluge, M., Razanoelisoa, B., and Brulfert, J. (2001). Implications of Genotypic Diversity and Phenotypic Plasticity in the Ecophysiological Success of CAM Plants, Examined by Studies on the Vegetation of Madagascar. Plant Biol. *3*, 214–222.

Komatsu, S., Yamamoto, R., Nanjo, Y., Mikami, Y., Yunokawa, H., and Sakata, K. (2009). A Comprehensive Analysis of the Soybean Genes and Proteins Expressed under Flooding Stress using Transcriptome and Proteome Techniques. J. Proteome Res. *8*, 4766–4778.

Korpelainen, H., von Cräutlein, M., Kostamo, K., and Virtanen, V. (2013). Spatial genetic structure of aquatic bryophytes in a connected lake system. Plant Biol. Stuttg. Ger. *15*, 514–521.

Kott, L.S., and Britton, D.M. (1982). A comparative study of spore germination of some *Isoëtes* species of northeastern North America. Can. J. Bot. *60*, 1679–1687.

Kovach, W.L., and Batten, D.J. (1993). Diversity changes in lycopsid and aquatic fern megaspores through geologic time. Paleobiology *19*, 28–42.

Kozmik, Z., Ruzickova, J., Jonasova, K., Matsumoto, Y., Vopalensky, P., Kozmikova, I., Strnad, H., Kawamura, S., Piatigorsky, J., Paces, V., et al. (2008). Assembly of the cnidarian camera-type eye from vertebrate-like components. Proc. Natl. Acad. Sci. *105*, 8989–8993.

Ku, S.B., and Edwards, G.E. (1978). Oxygen inhibition of photosynthesis : III. Temperature dependence of quantum yield and its relation to O2/CO 2 solubility ratio. Planta *140*, 1–6.

Ku, M.S., Kano-Murakami, Y., and Matsuoka, M. (1996). Evolution and expression of C_4 photosynthesis genes. Plant Physiol. *111*, 949–957.

Kustatscher, E., Wachtler, M., and Van Konijnenburg-Van Cittert, J.H.A. (2010). Lycophytes from the Middle Triassic (Anisian) locality Kühwiesenkopf (Monte Prà della Vacca) in the Dolomites (northern Italy). Palaeontology *53*, 595–626.

Lang, G.I., and Murray, A.W. (2011). Mutation rates across budding yeast chromosome VI are correlated with replication timing. Genome Biol. Evol. *3*, 799–811.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Larsén, E., and Rydin, C. (2015a). Disentangling the Phylogeny of Isoetes (Isoetales), Using Nuclear and Plastid Data. Int. J. Plant Sci. *177*, 157–174.

Larsén, E., and Rydin, C. (2015b). Disentangling the Phylogeny of Isoetes (Isoetales), Using Nuclear and Plastid Data. Int. J. Plant Sci. *177*, 157–174.

Larsson, A. (2014). AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics *30*, 3276–3278.

Laushman, R.H. (1993). Population genetics of hydrophilous angiosperms. Aquat. Bot. *44*, 147–158.

Leimu, R., and Fischer, M. (2008). A Meta-Analysis of Local Adaptation in Plants. PLOS ONE *3*, e4010.

Leon, B., and Young, K.R. (1996). Aquatic plants of Peru: diversity, distribution and conservation. Biodivers. Conserv. *5*, 1169–1190.

Li, W. (2014). Environmental opportunities and constraints in the reproduction and dispersal of aquatic plants. Aquat. Bot. *118*, 62–70.

Li, H., Elphick, M., and Shine, R. (2017). Potential targets for selection during the evolution of viviparity in cold-climate reptiles. Oecologia *183*, 21–30.

Linkies, A., Graeber, K., Knight, C., and Leubner-Metzger, G. (2010). The evolution of seeds. New Phytol. *186*, 817–831.

Logsdon, J.M., and Doolittle, W.F. (1997). Origin of antifreeze protein genes: A cool tale in molecular evolution. Proc. Natl. Acad. Sci. U. S. A. *94*, 3485–3487.

Lokker, C., Susko, D., Lovett Doust, L., and Lovett Doust, J. (1994). Population genetic structure of *Vallisneria americana*, a dioecious clonal macrophyte. Am. J. Bot. 81(8), 1004-1012.

Long, M., Betrán, E., Thornton, K., and Wang, W. (2003). The origin of new genes: glimpses from the young and old. Nat. Rev. Genet. *4*, 865-875.

Loveless, M.D., and Hamrick, J.L. (1984). Ecological Determinants of Genetic Structure in Plant Populations. Annu. Rev. Ecol. Syst. *15*, 65–95.

Lucassen, E.C.H.E.T., Roelofs, J.G.M., Schneider, S.C., and Smolders, A.J.P. (2016). Long-term effects of liming in Norwegian softwater lakes: the rise and fall of bulbous rush (Juncus bulbosus) and decline of isoetid vegetation. Freshw. Biol. *61*, 769–782.

Luijckx, P., Ho, E.K.H., Gasim, M., Chen, S., Stanic, A., Yanchus, C., Kim, Y.S., and Agrawal, A.F. (2017). Higher rates of sex evolve during adaptation to more complex environments. Proc. Natl. Acad. Sci. *114*, 534–539.

Lundgren, M.R., Osborne, C.P., and Christin, P.-A. (2014). Deconstructing Kranz anatomy to understand C_4 evolution. J. Exp. Bot. *65*, 3357–3369.

Lundgren, M.R., Besnard, G., Ripley, B.S., Lehmann, C.E.R., Chatelet, D.S., Kynast, R.G., Namaganda, M., Vorontsova, M.S., Hall, R.C., Elia, J., et al. (2015).

Photosynthetic innovation broadens the niche within a single species. Ecol. Lett. *18*, 1021–1029.

Lupia, R., Lidgard, S., and Crane, P.R. (1999). Comparing palynological abundance and diversity: implications for biotic replacement during the Cretaceous angiosperm radiation. Paleobiology *25*, 305–340.

Lüttge, U. (2002). CO_2-concentrating: consequences in crassulacean acid metabolism. J. Exp. Bot. *53*, 2131–2142.

Lüttge, U. (2004). Ecophysiology of Crassulacean Acid Metabolism (CAM). Ann. Bot. *93*, 629–652.

Lüttge, U. (2008). Clusia: Holy Grail and enigma. J. Exp. Bot. *59*, 1503–1514.

Maberly, S.C., and Gontero, B. (2017). Ecological imperatives for aquatic CO_2-concentrating mechanisms. J. Exp. Bot. *68*, 3797–3814.

Maberly, S.C., and Madsen, T.V. (2002). Freshwater angiosperm carbon concentrating mechanisms: processes and patterns. Funct. Plant Biol. *29*, 393–405.

Macnab, R.M. (2004). Type III flagellar protein export and flagellar assembly. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1694*, 207–217.

Madsen, T.V. (1987a). The effect of different growth conditions on dark and light carbon assimilation in *Littorella uniflora*. Physiol. Plant. *70*, 183–188.

Madsen, T.V. (1987b). Sources of Inorganic Carbon Acquired through CAM in *Littorella uniflora* (L.) Aschers. J. Exp. Bot. *38*, 367–377.

Madsen, T.V. (1987c). Interactions Between Internal and External CO_2 Pools in the Photosynthesis of the Aquatic Cam Plants *Littorella uniflora* (l.) Aschers and *Isoetes lacustris* L. New Phytol. *106*, 35–50.

Madsen, T.V., and Sand-Jensen, K. (1991). Photosynthetic carbon assimilation in aquatic macrophytes. Aquat. Bot. *41*, 5–40.

Madsen, T.V., Olesen, B., and Bagger, J. (2002). Carbon acquisition and carbon dynamics by aquatic isoetids. Aquat. Bot. *73*, 351–371.

Magallón, S., Hilu, K.W., and Quandt, D. (2013). Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. Am. J. Bot. *100*, 556–573.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L.L., and Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. New Phytol. *207*, 437–453.

Maier, A., Zell, M.B., and Maurino, V.G. (2011). Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C_4 and C_3 photosynthesis. J. Exp. Bot. *62*, 3061–3069.

Males, J., and Griffiths, H. (2017). Stomatal Biology of CAM Plants. Plant Physiol. *174*, 550–560.

Mallmann, J.,Heckmann, D., Bräutigam, A., Lercher, M.J., Weber, A.P., Westhoff, P., and Gowik, U. (2014). The role of photorespiration during the evolution of C_4 photosynthesis in the genus *Flaveria*. Elife *3*.

Martin, A., and Orgogozo, V. (2013). The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. Evolution *67*, 1235–1250.

Martin, C.E., Higley, M., and Wang, W.-Z. (1988). Ecophysiological Significance of CO_2-Recycling via Crassulacean Acid Metabolism in *Talinum calycinum* Engelm. (Portulacaceae). Plant Physiol. *86*, 562–568.

Martínez-Garrido, J., Bermejo, R., Serrão, E.A., Sánchez-Lizaso, J., and González-Wangüemert, M. (2017). Regional Genetic Structure in the Aquatic Macrophyte *Ruppia cirrhosa* Suggests Dispersal by Waterbirds. Estuaries Coasts *40*, 1705–1716.

Martinoia, E. (2018). Vacuolar transporters - Companions on a longtime journey. Plant Physiol. 176(2). 1384-1407.

McLoughlin, S., Drinnan, A.N., Slater, B.J., and Hilton, J. (2015). *Paurodendron stellatum*: A new Permian permineralized herbaceous lycopsid from the Prince Charles Mountains, Antarctica. Rev. Palaeobot. Palynol. *220*, 1–15.

McManmon, M., and Crawford, R.M.M. (1971). A Metabolic Theory of Flooding Tolerance: The Significance of Enzyme Distribution and Behaviour. New Phytol. *70*, 299–306.

Meyer, M., Seibt, U., and Griffiths, H. (2008). To concentrate or ventilate? Carbon acquisition, isotope discrimination and physiological ecology of early land plant life forms. Philos. Trans. R. Soc. B Biol. Sci. *363*, 2767–2778.

Millikan, R.G. (1989). In Defense of Proper Functions. Philos. Sci. *56*, 288–302.

Ming, R., VanBuren, R., Wai, C.M., Tang, H., Schatz, M.C., Bowers, J.E., Lyons, E., Wang, M.-L., Chen, J., Biggers, E., et al. (2015). The pineapple genome and the evolution of CAM photosynthesis. Nat. Genet. *47*, 1435.

Montañez, I.P., McElwain, J.C., Poulsen, C.J., White, J.D., DiMichele, W.A., Wilson, J.P., Griggs, G., and Hren, M.T. (2016). Climate, p_{CO2} and terrestrial carbon cycle linkages during late Palaeozoic glacial–interglacial cycles. Nat. Geosci. *9*, 824-828.

Moreno-Villena, J.J., Dunning, L.T., Osborne, C.P., and Christin, P.-A. (2018). Highly Expressed Genes Are Preferentially Co-Opted for C_4 Photosynthesis. Mol. Biol. Evol. *35*, 94–106.

Morjan, C.L., and Rieseberg, L.H. (2004). How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. Mol. Ecol. *13*, 1341–1356.

Conway-Morris, S.C. (2003). Life's Solution: Inevitable Humans in a Lonely Universe (Cambridge University Press).

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., and Donoghue, P.C.J. (2018). The timescale of early land plant evolution. Proc. Natl. Acad. Sci. 115(10). 2274-2283.

Muhaidat, R., Sage, T.L., Frohlich, M.W., Dengler, N.G., and Sage, R.F. (2011). Characterization of C_3–C_4 intermediate species in the genus *Heliotropium* L. (Boraginaceae): anatomy, ultrastructure and enzyme activity. Plant Cell Environ. *34*, 1723–1736.

Murphy, K.J. (2002). Plant communities and plant diversity in softwater lakes of northern Europe. Aquat. Bot. *73*, 287–324.

Nagalingum, N.S., Schneider, H., and Pryer, K.M. (2006). Comparative Morphology of Reproductive Structures in Heterosporous Water Ferns and a Reevaluation of the Sporocarp. Int. J. Plant Sci. *167*, 805–815.

Nagalingum, N.S., Marshall, C.R., Quental, T.B., Rai, H.S., Little, D.P., and Mathews, S. (2011). Recent Synchronous Radiation of a Living Fossil. Science *334*, 796–799.

Neander, K. (1991). Functions as Selected Effects: The Conceptual Analyst's Defense. Philos. Sci. *58*, 168–184.

Nelson, E.A., and Sage, R.F. (2008). Functional constraints of CAM leaf anatomy: tight cell packing is associated with increased CAM function across a gradient of CAM expression. J. Exp. Bot. *59*, 1841–1850.

Nelson, E.A., Sage, T.L., and Sage, R.F. (2005). Functional leaf anatomy of plants with crassulacean acid metabolism. Funct. Plant Biol. *32*, 409–419.

Newton, M.S., Arcus, V.L., and Patrick, W.M. (2015). Rapid bursts and slow declines: on the possible evolutionary trajectories of enzymes. J. R. Soc. Interface *12*.

Nielsen, S.L., Gacia, E., and Sand-Jensen, K. (1991). Land plants of amphibious *Littorella uniflora* (L.) Aschers. maintain utilization of CO<Subscript>2</Subscript> from the sediment. Oecologia *88*, 258–262.

Niklas, K.J., and Kutschera, U. (2010). The evolution of the land plant life cycle. New Phytol. *185*, 27–41.

Nisbet, E.G., Grassineau, N.V., Howe, C.J., Abell, P.I., Regelous, M., and Nisbet, R.E.R. (2007). The age of Rubisco: the evolution of oxygenic photosynthesis. Geobiology *5*, 311–335.

Nobel, P.S. (1996). High Productivity of Certain Agronomic CAM Species. In Crassulacean Acid Metabolism, (Springer, Berlin, Heidelberg), pp. 255–265.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302*, 205–217.

Ogren, W.L. (1984). Photorespiration: Pathways, Regulation, and Modification. Annu. Rev. Plant Physiol. *35*, 415–442.

Ohno, S. (1970). Evolution by gene duplication (Springer-Verlag, Berlin Heidelberg).

O'Leary, B., Park, J., and Plaxton, W.C. (2011). The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs. Biochem. J. *436*, 15–34.

Olofsson, J.K., Bianconi, M., Besnard, G., Dunning, L.T., Lundgren, M.R., Holota, H., Vorontsova, M.S., Hidalgo, O., Leitch, I.J., Nosil, P., et al. (2016). Genome biogeography reveals the intraspecific spread of adaptive mutations for a complex trait. Mol. Ecol. *25*, 6107–6123.

Osborn, J.M., El-Ghazaly, G., and Cooper, R.L. (2001). Development of the exineless pollen wall in *Callitriche truncata* (Callitrichaceae) and the evolution of underwater pollination. Plant Syst. Evol. *228*, 81–87.

Osborne, C.P., and Beerling, D.J. (2006). Nature's green revolution: the remarkable evolutionary rise of C_4 plants. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *361*, 173–194.

Osborne, C.P., and Freckleton, R.P. (2009). Ecological selection pressures for C_4 photosynthesis in the grasses. Proc. R. Soc. Lond. B Biol. Sci. 276(1663) 1753-1760.

Osborne, C.P., and Sack, L. (2012). Evolution of C_4 plants: a new hypothesis for an interaction of CO_2 and water relations mediated by plant hydraulics. Phil Trans R Soc B *367*, 583–600.

Osmond, C.B. (1978). Crassulacean Acid Metabolism: A Curiosity in Context. Annu. Rev. Plant Physiol. *29*, 379–414.

Pagani, M., Zachos, J.C., Freeman, K.H., Tipple, B., and Bohaty, S. (2005). Marked Decline in Atmospheric Carbon Dioxide Concentrations During the Paleogene. Science *309*, 600–603.

Paley, W. (1802). Natural Theology or Evidences of the Existence and Attributes of the Deity (R. Faulder and J. Morgan).

Panchy, N., Lehti-Shiu, M., and Shiu, S.-H. (2016). Evolution of Gene Duplication in Plants. Plant Physiol. *171*, 2294–2316.

Park, J., Khuu, N., Howard, A.S.M., Mullen, R.T., and Plaxton, W.C. (2012). Bacterial- and plant-type phosphoenolpyruvate carboxylase isozymes from developing castor oil seeds interact in vivo and associate with the surface of mitochondria. Plant J. *71*, 251–262.

Parker, S.T. (1987). A sexual selection model for hominid evolution. Hum. Evol. *2*, 235–253.

Parker, J., Tsagkogeorga, G., Cotton, J.A., Liu, Y., Provero, P., Stupka, E., and Rossiter, S.J. (2013). Genome-wide signatures of convergent evolution in echolocating mammals. Nature *502*, 228-231.

Patel, R.K., and Jain, M. (2012). NGS QC Toolkit: A Toolkit for Quality Control of Next Generation Sequencing Data. PLOS ONE *7*, e30619.

Pedersen, O., Sand-Jensen, K., and Revsbech, N.P. (1995). Diel Pulses of O_2 and CO_2 in Sandy Lake Sediments Inhabited by *Lobelia dortmanna*. Ecology *76*, 1536–1545.

Pedersen, O., Andersen, T., Ikejima, K., Zakir Hossain, M., and Andersen, F.Ø. (2006). A multidisciplinary approach to understanding the recent and historical occurrence of the freshwater plant, *Littorella uniflora*. Freshw. Biol. *51*, 865–877.

Pedersen, O., Rich, S.M., Pulido, C., Cawthray, G.R., and Colmer, T.D. (2011). Crassulacean acid metabolism enhances underwater photosynthesis and diminishes photorespiration in the aquatic plant *Isoëtes australis*. New Phytol. *190*, 332–339.

Pereira, J.B.S., Labiak, P.H., Stützel, T., and Schulz, C. (2017). Origin and biogeography of the ancient genus *Isoëtes* with focus on the Neotropics. Bot. J. Linn. Soc. *185*, 253–271.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. PLOS ONE *7*, e37135.

Petersen, K.B., and Burd, M. (2017). Why did heterospory evolve? Biol. Rev. Camb. Philos. Soc. *92*, 1739–1754.

Philbrick, C.T. (1988). Evolution of Underwater Outcrossing From Aerial Pollination Systems: A Hypothesis. Ann. Mo. Bot. Gard. *75*, 836–841.

Philbrick, C.T., and Les, D.H. (1996). Evolution of Aquatic Angiosperm Reproductive Systems. BioScience *46*, 813–826.

Phillips, P.C. (2008). Epistasis — the essential role of gene interactions in the structure and evolution of genetic systems. Nat. Rev. Genet. *9*, 855.

Pierce, S., Winter, K. and Griffiths, H. (2002). The role of CAM in high rainfall cloud forests: an *in situ* comparison of photosynthetic pathways in Bromeliaceae. Plant, Cell & Env. 25(9). 1181-1189.

Pigg, K.B. (2001). Isoetalean Lycopsid Evolution: From the Devonian to the Present. Am. Fern J. *91*, 99–114.

Preisser, E.L., Kefer, J.Y., Lawrence, J.D., and Clark, T.W. (2000). Vernal pool conservation in connecticut: an assessment and recommendations. Environ. Manage. *26*, 503–513.

Prentice, H.C., Malm, J.U., and Hathaway, L. (2008). Chloroplast DNA variation in the European herb *Silene dioica* (red campion): postglacial migration and interspecific introgression. Plant Syst. Evol. *272*, 23.

Qiu, Y.-L., Taylor, A.B., and McManus, H.A. (2012). Evolution of the life cycle in land plants. J. Syst. Evol. *50*, 171–194.

Raven, J.A. (1970). Exogenous Inorganic Carbon Sources in Plant Photosynthesis. Biol. Rev. *45*, 167–220.

Raven, J.A. (2013). Rubisco: still the most abundant protein of Earth? New Phytol. *198*, 1–3.

Raven, J.A., Cockell, C.S., and De La Rocha, C.L. (2008). The Evolution of Inorganic Carbon Concentrating Mechanisms in Photosynthesis. Philos. Trans. Biol. Sci. *363*, 2641–2650.

Reid, N.M., Proestou, D.A., Clark, B.W., Warren, W.C., Colbourne, J.K., Shaw, J.R., Karchner, S.I., Hahn, M.E., Nacci, D., Oleksiak, M.F., et al. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science *354*, 1305–1308.

Rensing, S.A. (2017). Why we need more non-seed plant models. New Phytol. *216*, 355–360.

Renzaglia, K.S., Duff RJT, Nickrent, D.L., and Garbary, D.J. (2000). Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. Philos. Trans. R. Soc. B Biol. Sci. *355*, 769–793.

Retallack, G.J. (1997). Earliest Triassic origin of *Isoëtes* and quillwort evolutionary radiation. J. Paleontol. *71*, 500–521.

Richardson, K., Griffiths, H., Reed, M.L., Raven, J.A., and Griffiths, N.M. (1984). Inorganic carbon assimilation in the Isoetids, *Isoëtes lacustri*s L. and *Lobelia dortmanna* L. Oecologia *61*, 115–121.

Robe, W.E., and Griffiths, H. (1992). Seasonal variation in the ecophysiology of *Littorella uniflora* (L.) Ascherson in acidic and eutrophic habitats. New Phytol. *120*, 289–304.

Robe, W.E., and Griffiths, H. (1994). The impact of NO_3 loading on the freshwater macrophyte *Littorella uniflora*: N utilization strategy in a slow-growing species from oligotrophic habitats. Oecologia *100*, 368–378.

Robe, W.E., and Griffiths, H. (1998). Adaptations for an amphibious life: changes in leaf morphology, growth rate, carbon and nitrogen investment, and reproduction during adjustment to emersion by the freshwater macrophyte *Littorella uniflora*. New Phytol. *140*, 9–23.

Rørslett, B. (1991). Principal determinants of aquatic macrophyte richness in northern European lakes. Aquat. Bot. *39*, 173–193.

Rosenblum, E.B., Hoekstra, H.E., and Nachman, M.W. (2004). Adaptive reptile color variation and the evolution of the MC1R gene. Evol. Int. J. Org. Evol. *58*, 1794–1808.

Rosenblum, E.B., Parent, C.E., and Brandt, E.E. (2014). The Molecular Basis of Phenotypic Convergence. Annu. Rev. Ecol. Evol. Syst. *45*, 203–226.

Rosenthal, D.M., Ramakrishnan, A.P., and Cruzan, M.B. (2008). Evidence for multiple sources of invasion and intraspecific hybridization in *Brachypodium sylvaticum* (Hudson) Beauv. in North America. Mol. Ecol. *17*, 4657–4669.

Rothschild, L.J. (2008). The evolution of photosynthesis…again? Philos. Trans. R. Soc. Lond. B Biol. Sci. *363*, 2787–2801.

Rubinstein, C.V., Gerrienne, P., de la Puente, G.S., Astini, R.A., and Steemans, P. (2010). Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). New Phytol. *188*, 365–369.

Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., and Burleigh, J.G. (2014). From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. *14*, 23.

Rundel, P.W., Esler, K.J., and Cowling, R.M. (1999). Ecological and phylogenetic patterns of carbon isotope discrimination in the winter-rainfall flora of the Richtersveld, South Africa. Plant Ecol. *142*, 133–148.

Rundle, H.D., and Nosil, P. (2005). Ecological speciation. Ecol. Lett. *8*, 336–352.

Rury, P.M. (1978). A New and Unique, Mat-Forming Merlin's-Grass (Isoëtes) from Georgia. Am. Fern J. *68*, 99–108.

Sagan, L. (1967). On the origin of mitosing cells. J. Theor. Biol. *14*, 225-IN6.

Sage, R.F. (1999). Why C_4 photosynthesis. In C_4 Plant Biology, (Academic Press San Diego), p.

Sage, R.F. (2001). Environmental and Evolutionary Preconditions for the Origin and Diversification of the C_4 Photosynthetic Syndrome. Plant Biol. *3*, 202–213.

Sage, R.F. (2002). Are crassulacean acid metabolism and C_4 photosynthesis incompatible? Funct. Plant Biol. *29*, 775–785.

Sage, R.F. (2004). The evolution of C_4 photosynthesis. New Phytol. *161*, 341–370.

Sage, R.F., and Stata, M. (2015). Photosynthetic diversity meets biodiversity: The C_4 plant example. J. Plant Physiol. *172*, 104–119.

Sage, R.F., and Zhu, X.-G. (2011). Exploiting the engine of C_4 photosynthesis. J. Exp. Bot. *62*, 2989–3000.

Sage, R.F., Christin, P.-A., and Edwards, E.J. (2011). The C_4 plant lineages of planet Earth. J. Exp. Bot. *62*, 3155–3169.

Sage, R.F., Sage, T.L., and Kocacinar, F. (2012). Photorespiration and the Evolution of C_4 Photosynthesis. Annu. Rev. Plant Biol. *63*, 19–47.

Sage, R.F., Khoshravesh, R., and Sage, T.L. (2014). From proto-Kranz to C_4 Kranz: building the bridge to C4 photosynthesis. J. Exp. Bot. *65*, 3341–3356.

Sanderson, M.J. (2002). Estimating Absolute Rates of Molecular Evolution and Divergence Times: A Penalized Likelihood Approach. Mol. Biol. Evol. *19*, 101–109.

Sanderson, M.J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics *19*, 301–302.

Sand-Jensen, K., Pedersen, O., Binzer, T., and Borum, J. (2005). Contrasting Oxygen Dynamics in the Freshwater Isoetid *Lobelia dortmanna* and the Marine Seagrass *Zostera marina*. Ann. Bot. *96*, 613–623.

Santamaría, L. (2002). Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. Acta Oecologica *23*, 137–154.

Schenekar, T., Lerceteau-Köhler, E., and Weiss, S. (2014). Fine-scale phylogeographic contact zone in Austrian brown trout *Salmo trutta* reveals multiple waves of post-glacial colonization and a pre-dominance of natural versus anthropogenic admixture. Conserv. Genet. *15*, 561–572.

Schiffers, K., Schurr, F.M., Travis, J.M.J., Duputié, A., Eckhart, V.M., Lavergne, S., McInerny, G., Moore, K.A., Pearman, P.B., Thuiller, W., et al. (2014). Landscape structure and genetic architecture jointly impact rates of niche evolution. Ecography *37*, 1218–1229.

Schuler, M.L., Mantegazza, O., and Weber, A.P.M. (2016). Engineering C4 photosynthesis into C3 chassis in the synthetic biology age. Plant J. Cell Mol. Biol. *87*, 51–65.

Sculthorpe, C.D. (1967). The Biology of Aquatic Vascular Plants (Edward Arnold).

Shao, H., Gontero, B., Maberly, S.C., Jiang, H.S., Cao, Y., Li, W., and Huang, W.M. (2017). Responses of *Ottelia alismoides*, an aquatic plant with three CCMs, to variable CO_2 and light. J. Exp. Bot. *68*, 3985–3995.

Short, F.T., Kosten, S., Morgan, P.A., Malone, S., and Moore, G.E. (2016). Impacts of climate change on submerged and emergent wetland plants. Aquat. Bot. *135*, 3–17.

Silvera, K., Santiago, L.S., Cushman, J.C., and Winter, K. (2009). Crassulacean Acid Metabolism and Epiphytism Linked to Adaptive Radiations in the Orchidaceae. Plant Physiol. *149*, 1838–1847.

Silvera, K., Neubig, K.M., Whitten, W.M., Williams, N.H., Winter, K., and Cushman, J.C. (2010). Evolution along the crassulacean acid metabolism continuum. Funct. Plant Biol. *37*, 995–1010.

Silvera, K., Winter, K., Rodriguez, B.L., Albion, R.L., and Cushman, J.C. (2014). Multiple isoforms of phosphoenolpyruvate carboxylase in the Orchidaceae (subtribe Oncidiinae): implications for the evolution of crassulacean acid metabolism. J. Exp. Bot. *65*, 3623–3636.

Skillman, J.B. (2008). Quantum yield variation across the three pathways of photosynthesis: not yet out of the dark. J. Exp. Bot. *59*, 1647–1661.

Smolders, A.J.P., Lucassen, E.C.H.E.T., and Roelofs, J.G.M. (2002). The isoetid environment: biogeochemistry and threats. Aquat. Bot. *73*, 325–350.

Søndergaard, M., and Sand-Jensen, K. (1979). Carbon uptake by leaves and roots of *Littorella uniflora* (L.) Aschers. Aquat. Bot. *6*, 1–12.

Soria-Carrasco, V., Gompert, Z., Comeault, A.A., Farkas, T.E., Parchman, T.L., Johnston, J.S., Buerkle, C.A., Feder, J.L., Bast, J., Schwander, T., et al. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. Science *344*, 738–742.

Speed, M.P., and Arbuckle, K. (2017). Quantification provides a conceptual basis for convergent evolution. Biol. Rev. *92*, 815–829.

Spierenburg, P., Lucassen, E.C.H.E.T., Lotter, A.F., and Roelofs, J.G.M. (2009). Could rising aquatic carbon dioxide concentrations favour the invasion of elodeids in isoetid-dominated softwater lakes? Freshw. Biol. *54*, 1819–1831.

Spierenburg, P., Lucassen, E.C.H.E.T., Lotter, A.F., and Roelofs, J.G.M. (2010). Competition between isoetids and invading elodeids at different concentrations of aquatic carbon dioxide. Freshw. Biol. *55*, 1274–1287.

Spierenburg, P., Lucassen, E.C.H.E.T., Pulido, C., Smolders, A.J.P., and Roelofs, J.G.M. (2013). Massive uprooting of *Littorella uniflora* (L.) Asch. during a storm event and its relation to sediment and plant characteristics. Plant Biol. *15*, 955–962.

Spriggs, E.L., Christin, P.-A., and Edwards, E.J. (2014). C_4 Photosynthesis Promoted Species Diversification during the Miocene Grassland Expansion. PLOS ONE *9*, e97722.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

Stein, W.E., Berry, C.M., Hernick, L.V., and Mannolini, F. (2012). Surprisingly complex community discovered in the mid-Devonian fossil forest at Gilboa. Nature *483*, 78–81.

Stern, D.L. (2013). The genetic causes of convergent evolution. Nat. Rev. Genet. *14*, 751.

Still, C.J., Berry, J.A., Collatz, G.J., and DeFries, R.S. (2003). Global distribution of C3 and C_4 vegetation: Carbon cycle implications. Glob. Biogeochem. Cycles *17(1)*, 6-1-6-14.

Storz, J.F. (2016). Causes of molecular convergence and parallelism in protein evolution. Nat. Rev. Genet. *17*, 239-250.

Stroud, H., Do, T., Du, J., Zhong, X., Feng, S., Johnson, L., Patel, D.J., and Jacobsen, S.E. (2014). Non-CG methylation patterns shape the epigenetic landscape in Arabidopsis. Nat. Struct. Mol. Biol. *21*, 64–72.

Svensson, P., Bläsing, O.E., and Westhoff, P. (2003). Evolution of C_4 phosphoenolpyruvate carboxylase. Arch. Biochem. Biophys. *414*, 180–188.

Swallow, D.M. (2003). Genetics of lactase persistence and lactose intolerance. Annu. Rev. Genet. *37*, 197–219.

Szövényi, P., Perroud, P.-F., Symeonidi, A., Stevenson, S., Quatrano, R.S., Rensing, S.A., Cuming, A.C., and McDaniel, S.F. (2015). De novo assembly and comparative analysis of the *Ceratodon purpureus* transcriptome. Mol. Ecol. Resour. *15*, 203–215.

Tausta, S.L., Coyle, H.M., Rothermel, B., Stiefel, V., and Nelson, T. (2002). Maize C_4 and non-C_4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. Plant Mol. Biol. *50*, 635–652.

Taylor, T.N. (1981). Paleobotany: an introduction to fossil plant biology (New York ; London: McGraw-Hill).

Taylor, W.C., and Hickey, R.J. (1992). Habitat, Evolution, and Speciation in Isoetes. Ann. Mo. Bot. Gard. *79*, 613–622.

Tcherkez, G.G.B., Farquhar, G.D., and Andrews, T.J. (2006). Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. Proc. Natl. Acad. Sci. *103*, 7246–7251.

Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D., and Gaut, B.S. (2012). The Molecular Diversity of Adaptive Convergence. Science *335*, 457–461.

Tessene, M.F. (1968). Preliminary reports on the flora of Wisconsin no. 59: Plantaginaceae - plantain family. Transcr. Wis. Acad. Sci. Arts Lett. 281–313.

Thorne, R.F. (1972). Major Disjunctions in the Geographic Ranges of Seed Plants. Q. Rev. Biol. *47*, 365–411.

Ting, M.K.Y., She, Y.-M., and Plaxton, W.C. (2017). Transcript profiling indicates a widespread role for bacterial-type phosphoenolpyruvate carboxylase in malate-accumulating sink tissues. J. Exp. Bot. *68*, 5857–5869.

Tomarev, S.I., and Piatigorsky, J. (1996). Lens Crystallins of Invertebrates. Eur. J. Biochem. *235*, 449–465.

Tomescu, A.M.F. (2009). Megaphylls, microphylls and the evolution of leaf development. Trends Plant Sci. *14*, 5–12.

Troia, A. (2016). Dispersal and colonization in heterosporous lycophytes: palynological and biogeographical notes on the genus *Isoëtes* in the Mediterranean region. Webbia *71*, 277–281.

Troia, A., Pereira, J.B., Kim, C., and Taylor, W.C. (2016). The genus *Isoëtes* (Isoetaceae): a provisional checklist of the accepted and unresolved taxa. Phytotaxa *277*, 101–145.

Tsuzuki, M., Miyachi, S., Winter, K., and Edwards, G.E. (1982). Localization of carbonic anhydrase in crassulacean acid metabolism plants. Plant Sci. Lett. *24*, 211–218.

Van Hoeck, A., Horemans, N., Monsieurs, P., Cao, H.X., Vandenhove, H., and Blust, R. (2015). The first draft genome of the aquatic model plant *Lemna minor* opens the route for future stress physiology research and biotechnological applications. Biotechnol. Biofuels *8*.

Veen, H. van, Mustroph, A., Barding, G.A., Eijk, M.V., Welschen-Evertman, R.A.M., Pedersen, O., Visser, E.J.W., Larive, C.K., Pierik, R., Bailey-Serres, J., et al. (2013). Two *Rumex* Species from Contrasting Hydrological Niches Regulate Flooding Tolerance through Distinct Mechanisms. Plant Cell *25*, 4691–4707.

Vermeij, G.J., and Dudley, R. (2000). Why are there so few evolutionary transitions between aquatic and terrestrial ecosystems? Biol. J. Linn. Soc. *70*, 541–554.

Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome Res. *19*, 327–335.

Voesenek, L.A. C.J., Colmer, T.D., Pierik, R., Millenaar, F.F., and Peeters, A.J.M. (2006). How plants cope with complete submergence. New Phytol. *170*, 213–226.

Vogan, P.J., and Sage, R.F. (2011). Water-use efficiency and nitrogen-use efficiency of C_3-C_4 intermediate species of *Flaveria* Juss. (Asteraceae). Plant Cell Environ. *34*, 1415–1430.

Vöge, M. (2006). The reproductive phenology of *Isoëtes lacustris* L.: Results of field studies in Scandinavian lakes. Limnol. - Ecol. Manag. Inland Waters *36*, 228–233.

Wang, Y., and Berry, C.M. (2003). A novel lycopsid from the Upper Devonian of Jiangsu, China. Palaeontology *46*, 1297–1311.

Wellman, C.H. (2010). The invasion of the land by plants: when and where? New Phytol. *188*, 306–309.

Wellman, C.H., and Strother, P.K. (2015). The terrestrial biota prior to the origin of land plants (embryophytes): a review of the evidence. Palaeontology *58*, 601–627.

Weng, J.-K. (2014). The evolutionary paths towards complexity: a metabolic perspective. New Phytol. *201*, 1141–1149.

Westergaard, K.B., Alsos, I.G., Popp, M., Engelskjøn, T., Flatberg, K.I., and Brochmann, C. (2011). Glacial survival may matter after all: nunatak signatures in the rare European populations of two west-arctic species. Mol. Ecol. *20*, 376–393.

White, P.J., and Smith, J.A.C. (1989). Proton and anion transport at the tonoplast in crassulacean-acid-metabolism plants: specificity of the malate-influx system in *Kalanchoë daigremontiana*. Planta *179*, 265–274.

Wilkins, M.B. (1992). Tansley Review No. 37 Circadian Rhythms: Their Origin and Control. New Phytol. *121*, 347–375.

Winter, K., and Holtum, J.A.M. (2014). Facultative crassulacean acid metabolism (CAM) plants: powerful tools for unravelling the functional elements of CAM photosynthesis. J. Exp. Bot. *65*, 3425–3441.

Winter, K., and Holtum, J.A.M. (2015). Cryptic crassulacean acid metabolism (CAM) in *Jatropha curcas*. Funct. Plant Biol. *42*, 711–717.

Winter, K., Garcia, M., and Holtum, J.A.M. (2014). Nocturnal versus diurnal CO2 uptake: how flexible is *Agave angustifolia*? J. Exp. Bot. *65*, 3695–3703.

Winter, K., Holtum, J.A.M., and Smith, J.A.C. (2015). Crassulacean acid metabolism: a continuous or discrete trait? New Phytol. *208*, 73–78.

Wistow, G. (1993). Lens crystallins: gene recruitment and evolutionary dynamism. Trends Biochem. Sci. *18*, 301–306.

Witham, C.W., Bauder, E.T., D Belk, Ferren Jr., W.R., and Ornduff, R. (1998). Ecology, Conservation, and Management of Vernal Pool Ecosystems - Proceedings from a 1996 Conference. (California Native Plant Society, Sacramento, CA.).

Wium-Andersek, S., and Andersen, J.M. (1972). Carbon Dioxide Content of the Interstitial Water in the Sediment of Grane Langsø, a Danish Lobelia Lake. Limnol. Oceanogr. *17*, 943–947.

Wolfe, J.A. (1997). Relations of Environmental Change to Angiosperm Evolution During the Late Cretaceous and Tertiary. In Evolution and Diversification of Land Plants, (Springer, Tokyo), pp. 269–290.

Wolfe, K.H., Sharp, P.M., and Li, W.H. (1989). Mutation rates differ among regions of the mammalian genome. Nature *337*, 283–285.

Wu, J., Zhao, H.-B., Yu, D., and Xu, X. (2017). Transcriptome profiling of the floating-leaved aquatic plant *Nymphoides peltata* in response to flooding stress. BMC Genomics *18*, 119.

Wyman, S.K., Jansen, R.K., and Boore, J.L. (2004). Automatic annotation of organellar genomes with DOGMA. Bioinformatics *20*, 3252–3255.

Xu, H.-H., Wang, Y., and Wang, Q. (2012). A new homosporous, arborescent lycopsid from the Middle Devonian of Xinjiang, Northwest China. Palaeontology *55*, 957–966.

Yang, X., Cushman, J.C., Borland, A.M., Edwards, E.J., Wullschleger, S.D., Tuskan, G.A., Owen, N.A., Griffiths, H., Smith, J.A.C., De Paoli, H.C., et al. (2015). A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. New Phytol. *207*, 491–504.

Yang, X., Hu, R., Yin, H., Jenkins, J., Shu, S., Tang, H., Liu, D., Weighill, D.A., Yim, W.C., Ha, J., et al. (2017). The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. Nat. Commun. *8*, 1899.

Young, A., Boyle, T., and Brown, T. (1996). The population genetic consequences of habitat fragmentation for plants. Trends Ecol. Evol. *11*, 413–418.

Young, J.N., Rickaby, R.E.M., Kapralov, M.V., and Filatov, D.A. (2012). Adaptive signals in algal Rubisco reveal a history of ancient atmospheric carbon dioxide. Philos. Trans. R. Soc. Lond. B Biol. Sci. *367*, 483–492.

Zelitch, I. (1973). Plant Productivity and the Control of Photorespiration. Proc. Natl. Acad. Sci. *70*, 579–584.

Zhang, J. (2003). Evolution by gene duplication: an update. Trends Ecol. Evol. *18*, 292–298.

Zhang, L., Chen, F., Zhang, G.-Q., Zhang, Y.-Q., Niu, S., Xiong, J.-S., Lin, Z., Cheng, Z.-M. (Max), and Liu, Z.-J. (2016). Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. Plant J. *86*, 175–185.

Zhang, Y., Yin, L., Jiang, H.-S., Li, W., Gontero, B., and Maberly, S.C. (2014). Biochemical and biophysical CO_2 concentrating mechanisms in two species of freshwater macrophyte within the genus *Ottelia* (Hydrocharitaceae). Photosynth. Res. *121*, 285–297.

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., Li, X., Ding, Y., Yang, S., and Wang, W. (2008). On the origin of new genes in *Drosophila*. Genome Res. *18*, 1446–1455.

Zhu, J., Yu, D., and Xu, X. (2015). The phylogeographic structure of *Hydrilla verticillata* (Hydrocharitaceae) in China and its implications for the biogeographic history of this worldwide-distributed submerged macrophyte. BMC Evol. Biol. *15*, 95.

Zhu, Y., Chen, L., Zhang, C., Hao, P., Jing, X., and Li, X. (2017). Global transcriptome analysis reveals extensive gene remodeling, alternative splicing and differential transcription profiles in non-seed vascular plant *Selaginella moellendorffii*. BMC Genomics *18*, 1042.