Molecular Phylogenetics and Evolution 105 (2016) 102–113

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Evolution of a CAM anatomy predates the origins of Crassulacean acid metabolism in the Agavoideae (Asparagaceae)

Karolina Heyduk [a,*], Michael R. McKain [b], Falak Lalani [a], James Leebens-Mack [a]

[a] Department of Plant Biology, 2502 Miller Plant Sciences, University of Georgia, Athens, GA 30602, United States
[b] Donald Danforth Plant Science Center, 975 N. Warson Rd., St. Louis, MO 63132, United States

ARTICLE INFO

Article history:
Received 11 April 2016
Revised 24 August 2016
Accepted 25 August 2016
Available online 30 August 2016

Keywords:
CAM
Character evolution
Leaf anatomy
Photosynthesis
Phylogenomics
Sequence capture

ABSTRACT

Crassulacean acid metabolism (CAM) is a modified form of photosynthesis that has arisen independently at least 35 times in flowering plants. The occurrence of CAM is often correlated with shifts to arid, semiarid, or epiphytic habits, as well as transitions in leaf morphology (e.g. increased leaf thickness) and anatomy (e.g. increased cell size and packing). We assess shifts between C_3 and CAM photosynthesis in the subfamily Agavoideae (Asparagaceae) through phylogenetic analysis of targeted loci captured from the nuclear and chloroplast genomes of over 60 species. Carbon isotope data was used as a proxy for mode of photosynthesis in extant species and ancestral states were estimated on the phylogeny. Ancestral character state mapping suggests three independent origins of CAM in the Agavoideae. CAM species differ from C_3 species in climate space and are found to have thicker leaves with densely packed cells. C_3 ancestors of CAM species show a predisposition toward CAM-like morphology. Leaf characteristics in the ancestral C_3 species may have enabled the repeated evolution of CAM in the Agavoideae subfamily. Anatomical changes, including a tendency toward 3D venation, may have initially arisen in C_3 ancestors in response to aridity as a way to increase leaf succulence for water storage.

1. Introduction

When RuBisCO (Ribulose bisphosphate carboxylase oxygenase) evolved some 3 billion years ago, Earth's atmosphere was high in CO_2 and lacking oxygen. Today, however, the 20% of the atmosphere that is oxygen can be fixed by RuBisCO, engaging the photorespiratory pathway in C_3 plants. Photorespiration is energetically costly, and particularly pronounced under high temperatures or where the ratio of CO_2 to O_2 is depleted in leaves. C_3 plants that are subjected to regular water stress will close their stomata to conserve water and draw down the internal leaf carbon concentration. To circumvent photorespiration due to RuBisCO's affinity for O_2, many photosynthetic organisms have developed carbon concentrating mechanisms (CCMs) to enrich the air around RuBisCO with carbon; it's estimated that half of all carbon fixed is acquired through a CCM (Raven et al., 2008). Crassulacean acid metabolism (CAM) is a complex CCM that has evolved from C_3 ancestors at least 35 times in independent plant lineages (Silvera et al., 2010). CAM plants conduct the majority of their gas exchange at night, when open stomata are subjected to much lower vapor pressure deficits, leading to reduced transpiration rates. Incoming CO_2 is fixed by phosphoenolpyruvate carboxylase (PEPC) and stored as malic acid in the vacuole. In the daytime, the stomata close and malic acid is decarboxylated, resulting in increased concentrations of CO_2 around RuBisCO. The higher concentration of carbon within CAM plant cells results in an increased water use efficiency, and allows CAM plants to grow in some of the most water stressed environments on earth. Phylogenetic analyses suggest that CAM evolved in multiple plant lineages 5–10 Mya (e.g. Good-Avila et al., 2006; Ocampo and Columbus, 2010; Bone et al., 2015) and may have allowed species diversification into desert habitats that arose at this time (Axelrod, 1979).

Though CAM is not as phylogenetically clustered as C_4 photosynthesis (Edwards et al., 2010; Sage et al., 2011), a few large lineages do seem to exhibit repeated origins of CAM. High densities of CAM species are found in the Cactaceae, Euphorbiaceae, Crassulaceae in the eudicots, and in the monocot families Bromeliaceae, Orchidaceae, and Asparagaceae (subfamily Agavoideae). CAM has been found in 35 separate lineages (Silvera et al., 2010), but this is likely an underestimate, as broad-scale surveys are often unable to detect intermediate or weak forms of CAM. For example, surveys of CAM plants across broad phylogenetic scales use carbon isotope ratios ($\delta^{13}C$) as a proxy for mode of photosynthesis: CAM and C_4 plants, via the enzyme PEPC, discriminate less against the heavier ^{13}C isotope of carbon than RuBisCO does, resulting in a ratio of ^{13}C

* Corresponding author.
E-mail address: heyduk@uga.edu (K. Heyduk).

http://dx.doi.org/10.1016/j.ympev.2016.08.018
1055-7903/

and ^{12}C that is more similar to atmospheric ratios in CAM plants relative to C_3. CAM plants have $\delta^{13}C$ ratios between −10‰ and −20‰, while C_3 plants have ratios less than −22‰. Intermediate CAM species, however, act as cryptic CAM plants in that they can have carbon isotope ratio signatures that are within the range of C_3 species (Silvera et al., 2005).

Anatomically, CAM plants from independent lineages have converged on a suite of traits that are directly related to the ability of the CAM pathway to function efficiently. Large cells, and particularly large vacuoles, are required for storing accumulated malic acid, which directly influences the carbon available to RuBisCO during the day. To accommodate large cells, CAM leaves typically exhibit reduced intercellular airspace (IAS), which has the added benefit of limiting diffusion of decarboxylated CO_2 in CAM plant cells during the day (Nelson and Sage, 2008; Nelson et al., 2005; Zambrano et al., 2014). A number of CAM traits represent tradeoffs with the C_3 photosynthetic pathway: for example, a lack of IAS in C_3 plants will limit the ability of CO_2 to diffuse throughout the layers of mesophyll cells. The transition from C_3 photosynthesis to CAM includes the evolution of anatomical modifications that allow for optimal CAM pathway function.

In the C_4 literature, a stepping-stone model of evolution has been proposed in which plants gradually progress from C_3 to C_4 through an intermediate stage known as C_2 (Monson and Rawsthorne, 2000; Sage, 2004). Evolutionary studies of C_3 to C_4 transitions have found that anatomical differences arise before a fully optimized C_4 pathway (Christin et al., 2013, 2011). How CAM lineages evolved from C_3 ancestors is less well described, but a few models have been proposed. Because C_3 plants have the major CAM pathway components functioning in stomatal guard cells, it has been suggested that CAM evolution required shifting these reactions to mesophyll cells (Cockburn, 1981). Alternatively, intermediate or weak CAM species have been implicated in the evolutionary transition from C_3 photosynthesis to full CAM; there is some evidence from the Crassulaceae that suggests CAM cycling or a similar weak form as an evolutionary stepping stone to full constitutive CAM (Teeri, 1982). Sage (2002) suggested that leaf and cell succulence may arise first in a C_3 ancestor, followed by evolution of PEPC function to recapture respired CO_2, eventually leading to circadian control and full-fledged CAM function. Succulent anatomy as a precursor for CAM evolution is an attractive hypothesis, especially for plant lineages where succulence could initially alleviate water stress via water storage before being co-opted for CAM function. However, few studies have examined ancestral characteristics in extant CAM lineages to determine whether ancestral anatomy facilitated CAM evolution. A unified model of C_3 to CAM evolution requires comparative analyses of anatomy and selective forces in closely related C_3 and CAM lineages in order to understand whether certain CAM-like traits appear before the origin of CAM, thus enabling the repeated evolution of a complex trait. Here, we utilize the bimodal karyotype (BK) clade of the Agavoideae (sensu McKain et al., 2012), a subfamily of the Asparagaceae (APG III, Chase et al., 2009), to assess whether CAM anatomical traits arose before or concurrently with the origins of CAM.

The BK clade of the Agavoideae consists of ~600 species and includes the iconic desert genera *Yucca* and *Agave*, in addition to a number of genera inhabiting mesic environments, like *Camassia* and *Hosta*. The subfamily is widespread but largely restricted to the new world (although *Hosta* is an exception, with an eastern Asian distribution (Rocha et al., 2006)). The deserts of North and Central America have been implicated as the center of diversity for several genera within the Agavoideae. To date, analyses of relationships among Agavoideae lineages have been confounded by low and mixed phylogenetic signals (Archibald et al., 2015; Good-Avila et al., 2006; Smith et al., 2008), particularly within genera. The phylogenetic difficulties largely arise from the young age of the group, as the core Agavoideae are estimated to have diverged 20–26 Mya, while *Agave* and *Yucca* represent much more recent diversification events at 8–10 and 9–17 Mya, respectively (Good-Avila et al., 2006; Smith et al., 2008). Phylogenetic resolution of a young lineage can be further confounded by hybridization and incomplete lineage sorting (both of which can lead to gene tree-species tree incongruence (Degnan and Rosenberg, 2009; Maddison, 1997; Rosenberg, 2002); additionally, some species within the Agavoideae are long-lived perennials whose generation time results in slower mutation accumulation.

Research on mode of photosynthesis within the Agavoideae has focused almost exclusively on the physiology and anatomy of *Agave* species (Nobel, 1976, 1988, but see Smith et al., 1983; Huxman et al., 1998 for examples in *Yucca*), and the related desert species were assumed to be CAM. We assessed the photosynthetic pathway in over 60 species from 12 genera using carbon isotope ratios and ancestral state estimation on a species tree inferred from analysis of 272 nuclear loci and ~70 kb of the plastid genome, sampled by target locus enrichment (Mamanova et al., 2010; Heyduk et al., 2016b). Ancestral state reconstruction suggests multiple origins of CAM and at least one potential reversal from constitutive CAM. In addition, leaf cross sections were examined to assess whether certain CAM promoting traits – such as large cells or decreased IAS – evolved prior to or concurrently with CAM photosynthesis.

2. Materials and methods

2.1. Tissue sampling

For phylogenetic analyses, many of the *Yucca* tissues were collected from natural populations by Olle Pellmyr and collaborators and stored at −80 °C for 10–20 years, while tissue for all *Agave* accessions came from the Desert Botanical Garden's live collection (Phoenix, AZ, USA). Additional sources of tissue are described in Supporting Information Table S1. For each species, the source of tissue for carbon isotope measurements is also indicated in Supporting Information Table S1. Most isotope tissue was sourced from the Missouri Botanical Garden Herbarium, the University of Georgia State (UGA) Herbarium, and the Desert Botanical Garden Herbarium. Tissue for anatomical cross sections was harvested from plants growing at the UGA plant biology greenhouses, with samples taken from central portions of fully mature leaves.

2.2. RNA target enrichment bait design

Following McKain et al., 2012, transcriptome data from *Chlorogalum pomeridianum, Chlorophytum rhizopendulum, Hesperaloe parviflora, Hosta venusta, Yucca brevifolia,* and *Yucca filamentosa* were assembled using Trinity r2011-08-20 (Grabherr et al., 2011), with *C. rhizopendulum* acting as the outgroup to the Agavoideae BK clade. Additionally, available EST sequences from *Agave tequilana* (from Simpson et al., 2011) were included in orthogroup clustering and probe design. Transcripts of each species were translated using the RefTrans pipeline (https://github.com/mrmckain/RefTrans). Protein models for translations were chosen by BLAST (tblastx) against 10 sequenced angiosperm genomes (http://fgp.bio.psu.edu/tribedb/10_genomes/; Wall et al., 2008). The best hit of each transcript, as determined by e-value and overlapping similarity, was then used for protein alignment using GeneWise v.2.2.0 (Birney et al., 2004). The resulting amino acid and nucleotide coding sequences were clustered into gene families using OrthoMCL (Li et al., 2003) as in McKain et al. (2012). Amino acid sequences in each gene family were aligned in MUSCLE v.