Evolutionary Convergence of C₄ Photosynthesis: A Case Study in the Nyctaginaceae

Roxana Khoshravesh₁,², Matt Stata¹, Shunsuke Adachi³,⁴, Tammy L. Sage¹,⁴ and Rowan F. Sage¹*†

¹Department of Ecology and Evolutionary Biology, The University of Toronto, Toronto, ON, Canada, ²Department of Biology, The University of New Mexico, Albuquerque, NM, United States, ³Institute of Global Innovation Research, Tokyo University of Agriculture and Technology, Fuchu, Japan

C₄ photosynthesis evolved over 65 times, with around 24 origins in the eudicot order Caryophyllales. In the Caryophyllales family Nyctaginaceae, the C₄ pathway is known in three genera of the tribe Nyctagineae: Allionia, Okenia and Boerhavia. Phylogenetically, Allionia and Boerhavia/Okenia are separated by three genera whose photosynthetic pathway is uncertain. To clarify the distribution of photosynthetic pathways in the Nyctaginaceae, we surveyed carbon isotope ratios of 159 species of the Nyctaginaceae, along with bundle sheath (BS) cell ultrastructure, leaf gas exchange, and C₄ pathway biochemistry in five species from the two C₄ clades and closely related C₃ genera. All species in Allionia, Okenia and Boerhavia are C₄, while no C₄ species occur in any other genera of the family, including three that branch between Allionia and Boerhavia. This demonstrates that C₄ photosynthesis evolved twice in Nyctaginaceae. Boerhavia species use the NADP-malic enzyme (NADP-ME) subtype of C₄ photosynthesis, while Allionia species use the NAD-malic enzyme (NAD-ME) subtype. The BS cells of Allionia have many more mitochondria than the BS of Boerhavia. Bundle sheath mitochondria are closely associated with chloroplasts in Allionia which facilitates CO₂ refixation following decarboxylation by mitochondrial NAD-ME. The close relationship between Allionia and Boerhavia could provide insights into why NADP-ME versus NAD-ME subtypes evolve, particularly when coupled to analysis of their respective genomes. As such, the group is an excellent system to dissect the organizational hierarchy of convergent versus divergent traits produced by C₄ evolution, enabling us to understand when convergence is favored versus when divergent modifications can result in a common phenotype.

Keywords: Allionia, Boerhavia, C₄ photosynthesis, convergent evolution, Nyctaginaceae phylogeny, PEP carboxylase

INTRODUCTION

C₄ photosynthesis is a complex trait that arises following modifications to hundreds if not thousands of individual genes within a genome (Gowik et al., 2011). Despite this, it is one of the most convergent of evolutionary phenomena in the biosphere, with over 65 independent origins (Conway-Morris, 2003; Sage, 2016; Heyduk et al., 2019). Evolutionary convergence, however, does not necessarily reflect convergence throughout the hierarchy of traits that give rise to a complex
 phenotype, because multiple mechanisms can support a common function (Losos, 2011). This is well illustrated in the case of C₄ photosynthesis and crassulacean acid metabolism (CAM), each of which have been repeatedly assembled using disparate enzymes and structural modifications (Sage et al., 2012; Christin and Osborne, 2013; Edwards, 2019). While examples of evolutionary convergence are many, the mechanisms of convergence remain a major question in the life sciences, particularly in the cases where complex traits such as C₄ photosynthesis repeatedly evolve (Blount et al., 2018). Because the complexity of the C₄ system is well-understood, as well as the phylogenetic distribution of the many C₄ clades, C₄ photosynthesis represents an excellent system to understand the mechanics of convergent evolution, and its implication for the rise of C₄-dominated biomes over the past 30 million years (Christin and Osborne, 2013; Heyduk et al., 2019).

C₄ photosynthesis first captures CO₂ at low concentration in an outer mesophyll (M) compartment via the activity of phosphoenolpyruvate (PEP) carboxylase (PEPCase), and then concentrates it into an internal compartment, typically a layer of cells around the leaf vasculature termed the bundle sheath (BS; Edwards and Walker, 1983; Hatch, 1987). The C₄ pathway begins with the conversion of CO₂ to bicarbonate (HCO₃⁻) by carbonic anhydrase (CA) in M cells, followed by PEP carboxylation (Figure 1). These two steps occur in all C₄ plants, and thus are universally convergent traits in C₄ photosynthesis; however, different paralogs have been recruited to carry out the PEPCase and CA functions, demonstrating evolutionary flexibility at a lower level of organization (Christin et al., 2010b, 2013a; Ludwig, 2016a; Heyduk et al., 2019). The product of PEP carboxylation, oxaloacetate (OAA), is too labile to safely move between M and BS cells, so it must be converted to a stable metabolite (Edwards and Walker, 1983). Metabolite transport between M and BS cells is by diffusion, which necessitates that metabolites form steep concentration gradients to support rapid flux, and thus must be stable at high concentration (Bräutigam and Weber, 2011). The solution to the challenge presented by OAA instability is to convert it to the stable metabolites malate or aspartate. This highlights another fundamental feature of evolutionary convergence, in that it occurs where there are strict physiochemical constraints such as OAA lability. A common step (OAA conversion) is accomplished via divergent metabolic solutions (formation of malate versus aspartate). The selected transport metabolite, as it turns out, reflects the enzyme that catalyzes the decarboxylation step.

There are three major decarboxylating enzymes co-opted for the decarboxylation step in C₄ photosynthesis: NADP-malic enzyme (NADP-ME), NAD-malic enzyme (NAD-ME) and PEP carboxykinase (PCK; Hatch, 1987). Most C₄ lineages use NADP-ME as the primary decarboxylating enzyme, about 1/3 use NAD-ME and five lineages predominantly use PCK (Sage R. F. et al., 2011). The consequence of the decarboxylase selection affects many aspects of C₄ photosynthesis, to include the pathway biochemistry, BS and M ultrastructure, M to BS transport processes, leaf energetics, and photosynthetic efficiency (Hatch, 1987; Drincovich et al., 2011; Ghannoum et al., 2011). To recognize the suite of traits associated with each decarboxylation mode, three subtypes of C₄ photosynthesis have been delineated – the NADP-ME subtype, the NAD-ME subtype, and the PCK subtype (Figure 1). Although each subtype conducts C₄ photosynthesis, they represent three evolutionary mechanisms to concentrate CO₂ that are derived from a distinct set of biochemical, structural and transport traits (Rao and Dixon, 2016). While each subtype represents a divergent means of concentrating CO₂ around Rubisco, traits associated within each subtype reflect strong convergence in response to constraints imposed by the decarboxylating enzyme. The NADP-ME enzyme co-opted by the C₄ cycle is located in the BS chloroplasts, where it uses malate and NADP⁺ to produce CO₂, NADPH and pyruvate, with the NADPH directly supporting reduction of PGA produced by Rubisco. NAD-ME is mitochondrial, and uses malate and NAD⁺ to produce CO₂, NADH and pyruvate. Because it is active in the BS mitochondria, substantial mitochondrial volume is needed to meet the metabolic requirements of the C₄ pathway, and consistently, species of the NAD-ME subtype often have more and/or larger mitochondria in BS cells than NADP-ME species (Dengler and Nelson, 1999; Edwards and Voznesenskaya, 2011). The PCK enzyme is located in the cytosol, and uses ATP and OAA to produce PEP, CO₂, and ADP + Pᵢ. Large amounts of ATP are needed for the PCK reaction, and hence there is a large investment in BS mitochondrial volume to meet the ATP requirement (Yoshimura et al., 2004; Voznesenskaya et al., 2006). The decarboxylation type also constrains the selection of the transport metabolite. In NADP-ME species, NADPH produced by malate oxidation in the BS chloroplast is rapidly consumed in the metabolism of PGA generated by Rubisco, so there is no feedback onto malate flux. In NAD-ME species, the use of malate as a transport molecule would be problematic. NADH cannot readily exit the mitochondria, and there is an insufficient energy sink in the mitochondria to use the large amount of NADH that would be generated if NAD-ME oxidized malate imported from the M cells. The utilization of aspartate avoids these issues because there is no net import of reducing power into the BS mitochondria (Kanai and Edwards, 1999). Once in the BS mitochondria, aspartate forms OAA, which is reduced to malate via the NADH generated by the PCK reaction (Figure 1). The PCK reaction requires large amounts of ATP to support rapid photosynthesis, and this can be generated in the mitochondria by oxidizing NADH created by NAD-ME using malate directly imported into the BS mitochondria from the M cells (Leegood and Walker, 1999). PEP carboxykinase species are thought to use NAD-ME at about a fourth to a third of the rate of the PCK reaction, in order to supply sufficient NADH for ATP production (Leegood and Walker,
To meet the energy requirements imposed by the decarboxylating enzymes and associated transport systems, a distinct arrangement of chloroplast membranes occurs in each of the three subtypes (Hatch, 1987; Edwards and Voznesenskaya, 2011; Rao and Dixon, 2016). NADP-ME species have low grana stacking in the BS chloroplasts, yet high stacking in the M chloroplasts. More grana stacking in M cells increases the PSII content in the thylakoids, and hence the potential to generate NADPH by linear electron transport. Because NADP-ME species import reducing power into the BS with malate, the demand for NADPH to reduce PGA in the BS chloroplasts of NADP-ME plants is halved; hence less PSII and granal stacking are required (Kanai and Edwards, 1999). Chloroplasts in the M cells of NADP-ME species are well stacked to support malate reduction following PEP carboxylation. NAD-ME species, by contrast, have low grana stacking in M chloroplasts because their primary energetic function is to produce ATP for PEP regeneration. The thylakoids of the BS exhibit pronounced stacking in NAD-ME species, to supply sufficient NADPH to potentially reduce all the PGA generated by the C3 cycle. However, there can be variation on these general chloroplast phenotypes, depending upon the degree to which multiple decarboxylases are employed, whether PGA is exported to the M tissue for reduction, or if aspartate is imported into the chloroplast in lieu of malate, as suggested to occur in NADP-ME *Flaveria* species (Leegood and Walker, 1999; Furbank, 2011).

The theoretical pattern that emerges in C4 evolution is thus one of universal convergence in terms of overall outcome (CO2 concentration into a BS-like compartment), with some steps being universal (CO2 conversion to bicarbonate, PEP carboxylation), and others being divergent (decarboxylation, transport, chloroplast energetics, and possibly how PEP is regenerated). Convergence occurs within the subtype pathways where specific biophysical constraints are present, for example, in

---

**FIGURE 1** A schematic outlining the C4 pathway in NADP-ME type species, NAD-ME species, and PEP carboxykinase species. The diagram represents the classical depiction of the pathway (as per Hatch, 1987 and Furbank, 2011) assuming malate or aspartate predominate in their respective pathways. For clarity, nitrogen exchange reactions between aspartate and pyruvate are not shown in the NAD-ME and PCK subtypes. Thylakoid stacking is indicated by multi-layered rectangles in the chloroplasts. AL, alanine; ASP, aspartate; CA, carbonic anhydrase; MA, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PEP-C, PEP carboxylase; PEP-C, PCK, PEP carboxykinase; PPDK, pyruvate, phosphate dikinase; PVA, pyruvate.
transport metabolites associated with the distinct decarboxylating enzymes. An organizational hierarchy between convergence and divergence can thus be envisioned for complex traits that scales from the level of the genes up through to the composite phenotype. The challenge for research on convergent evolution is to describe this hierarchy and to understand why and under what constraints convergent patterns emerge (Losos, 2011).

To do this, one needs effective research systems, such as fast cycling microbes, or in terrestrial plants, multiple lineages where convergence has occurred, such as the dozens of clades that have independently evolved C4 photosynthesis (Blount et al., 2018; Heyduk et al., 2019). However, most closely related C4 clades are of the same subtype (Sage R. F. et al., 2011), which restricts the ability to examine sub-type impacts on convergent versus divergent solutions to creating a C4 pathway. One potentially strong study system occurs in Portulaca, where both NAD-ME and NAD-ME species are present (Voznesenskaya et al., 2010; Ocampo et al., 2013). Another possibility occurs in the Nyctaginaceae tribe of the Nyctaginaceae. Here, three related genera of the tribe Nyctaginaceae - Allionia, Boerhavia and Okenia - contain C4 species while their closest relatives in the genera Anulocaulis, Cyphomeris, Commicarpus, and Nyctaginia are not known to have any C4 species (Sage R. F. et al., 2011). In the phylogeny of Douglas and Manos (2007), Okenia and Boerhavia form a common C4 clade, while Allionia forms a distinct C4 clade that is separated from the Okenia/Boerhavia lineage by the Anulocaulis/Nyctaginia complex. However, there has been no systematic survey of the occurrence of C3 and C4 photosynthesis in the Nyctaginaceae, so it is unclear whether C4 may exist in Anulocaulis, Commicarpus, Cyphomeris, and Nyctaginia, or even whether Allionia, Okenia, and Boerhavia are completely C4, as opposed to also containing C3 and C3–C4 intermediate species. Boerhavia and Allionia are listed as being NAD-ME, although this is based on enzyme assays for Boerhavia only (Muhaidat et al., 2007). Intriguingly, our preliminary TEM images show many mitochondria in the BS ultrastructure of Allionia, suggesting it is NAD-ME. If so, then the Allionia and Okenia/Boerhavia clade could join Portulaca in forming a robust study system for addressing evolutionary convergence as influenced by C4 subtype.

To evaluate the potential of the Nyctaginaceae to become a model for studying evolutionary convergence, we present here a detailed study of the distribution of the C3 and C4 pathways in the Nyctaginaceae. We present a survey of carbon isotope ratios from 560 herbarium specimens, in addition to an anatomical/ultrastructural study using five representative species of Allionia, Anulocaulis, Boerhavia, Commicarpus, and Nyctaginia. We also examine climate data and geographic distributions of species in Allionia, Anulocaulis, Boerhavia, Commicarpus, Cyphomeris, and Nyctaginia to evaluate ecological factors contributing to C4 origins in the Nyctaginaceae. The biochemical sub-types of C4 species in Allionia and Boerhavia were determined, and we present a transcriptome-based phylogeny that updates the phylogenies of Douglas and Manos (2007) and Douglas and Spellenberg (2010). We also use transcriptome data to evaluate whether there has been convergence in the gene sequences of the major C4 pathway enzymes. For comparative purposes, we also present TEM images of leaf tissues from two Portulaca (Portulaceae) species previously shown to be NAD-ME (Portulaca pilosa) or NAD-ME (Portulaca oleracea). Through these efforts, we present an initial hierarchical assessment of how divergence occurs during the convergent evolution of C4 photosynthesis.

MATERIALS AND METHODS

Plant Material and Growth Conditions

Seeds of Allionia incarnata, Anulocaulis gypsogenus, Boerhavia burbigeana, Boerhavia coccinea, Commicarpus scandens, and Nyctaginia capitata were collected from naturally occurring field populations in Southwestern North America and Australia between 2000 and 2016 (see Supplementary Table 1 for collection and voucher information). Portulaca pilosa seeds were a gift from Gerry Edwards and Elena Voznesenskaya (Washington State University), while P. oleracea grew as a weed in our greenhouse. Seeds were sown directly into 10 or 20 L pots containing a sandy-loam mixture and grown in the University of Toronto greenhouse complex housed on the roof of the Earth's Sciences Centre. Plants were watered as needed to avoid soil drying (daily in high summer, two to three times weekly in cooler weather), and fertilized bimonthly with a Miracle-Grow commercial fertilizer mix (All-Purpose Brand, 24-8-16) amended with 4 mM calcium nitrate and 1 mM magnesium sulfate. Greenhouse conditions were 27 to 33°C daytime temperature, 20–24°C night temperature, and a peak photon flux density (PFD) of 1500 μmol photons m−2 s−1 on clear days. We used 400 W sodium vapor lamps to supplement natural daylight to maintain a minimum PFD on cloudy days of 250 μmol m−2 s−1 over a >13 h photoperiod. Unless otherwise indicated, three to five plants were sampled for gas exchange, biochemical assay, leaf structural properties and transcriptomics between May and October of 2010 to 2017. Care was taken to sample tissues under the same environmental conditions in the greenhouse (full sun exposure, a leaf temperature near 27°C, at a time between 9 am and 2 pm) to minimize year to year and month to month variation. For the characteristics examined here, subtle variation that may exist between sampling dates is not known to affect results pertinent to our hypotheses. For example, C3 versus C4 expression patterns are largely constitutive, and sun to shade variation in phenotype would not be present at the bright light intensities present in our greenhouses during summer.

Carbon Isotope Ratio

To determine the distribution of the C3 and C4 pathways in species of the Nyctaginaceae, 560 herbarium specimens representing 23 genera and 159 species were sampled from the herbaria at Kew gardens (Richmond, London, United Kingdom), Missouri Botanical Gardens (St Louis, MO United States), and New York Botanical Gardens (The Bronx, New York, NY, United States; Supplementary Tables 2, 3). Species were sampled from all genera in the Nyctaginaceae with the exception of the monospecific genus Cuscatlainia. Approximately 50% (Mirabilis) to 100% (Okenia, Allionia) of the known species from the
sampled Nyctaginaceae genera are present in the survey. We also sampled species from 12 genera occurring in each of the other six tribes of the Nyctaginaceae (Table 1 and Supplementary Table 2). To measure the carbon isotope ratio ($\delta^{13}C$) of each herbarium specimen, 2–4 mg of leaf or stem material were sampled from herbarium sheets and assayed for $\delta^{13}C$ by the Washington State University Stable Isotope Core.2 $\delta^{13}C$ was determined for at least two distinct plant specimens if available in the herbaria. Carbon isotope ratios between $-10$ and $-16\%e$ correspond to C$_4$ values, while carbon isotope ratios between $-23$ and $-32\%e$ correspond to C$_3$ values. Species that are evolutionary intermediates between C$_3$ and C$_4$ photosynthesis exhibit $\delta^{13}C$ ratios that are similar to C$_3$ values except where a strong C$_4$ metabolic cycle has been engaged; such C$_4$-like plants typically exhibit $\delta^{13}C$ values between $-16$ and $-22\%e$ (Monson et al., 1988; Von Caemmerer, 1992).

### Leaf Gas Exchange and Biochemical Assay

The response of net carbon assimilation rate (A) to intercellular CO$_2$ concentration ($C_i$) was measured for Allionia incarnata, B. coccinea, and N. capitata using a Li-COR 6400 photosynthetic gas analyzer at $33^\circ$C, a vapor pressure differences of 2.0–2.5 kPa and a photosynthetic PFD of 1800 $\mu$mol m$^{-2}$ s$^{-1}$. The most recent, fully mature leaves were first equilibrated in the leaf cuvette at an ambient CO$_2$ concentration of 400 $\mu$mol mol$^{-1}$ and 1800 $\mu$mol photons m$^{-2}$ s$^{-1}$. After equilibration and measurement, the ambient CO$_2$ was increased to 1200 $\mu$mol mol$^{-1}$ and then gas exchange parameters were measured after equilibration. The CO$_2$ concentration was then reduced to near the CO$_2$ compensation point in approximately 13 steps, with steady-state gas exchange values determined at each step. The initial slope of the A versus $C_i$ response was estimated as the linear slope of the measurements below a $C_i$ of 100 $\mu$mol mol$^{-1}$. Intrinsic water use efficiency (WUE) was determined as the ratio of A to stomatal conductance ($g_s$) at an ambient CO$_2$ of 400 ppm, and the CO$_2$-saturated rate of A ($A_{1200}$) was measured at 1200 $\mu$mol mol$^{-1}$ CO$_2$.

The activities of PEPC, NADP-malic enzyme, and NAD-malic enzyme were assayed at 30°C using a coupled-enzyme assay that measured oxidation/reduction rate of NADP(H) or NAD(H) at a wavelength of 340 nm using a Hewitt-Packard 8230 spectrophotometer following procedures in Ashton et al. (1990) as modified by Sage T. L. et al. (2011). Two to three cm$^2$ of recent, fully-mature leaves of A. incarnata, B. coccinea, and N. capitata were sampled under full illumination in the greenhouse and then rapidly ground using a glass tissue homogenizer in an extraction buffer (100 mM HEPES – pH 7.6, 5 mM MgCl$_2$, 10 mM KHCO$_3$, 2 mM EDTA, 10 mM 6-aminocaproic acid, 2 mM benzamide, 1 mM phenylmethylsulfonyl fluoride, 1% (w/v) PVPP, 2% (w/v) PVP, 0.5% Triton X-100, 2% (w/v) BSA, 5 mM DTT, 1% (w/v) casein). After removing two aliquots for chlorophyll assay (in 80% acetone at 645 and 663 nm; Arnon, 1949), the extract was centrifuged 30 s and divided into three aliquots and put on ice. The aliquot used for NAD-ME assay was immediately treated with sufficient MnCl$_2$ to give a 2 mM solution. PEPC was assayed by coupling the production of OAA to NADH oxidation via malate dehydrogenase in an assay buffer containing 50 mM Bicine (pH 8.0), 1 mM EDTA, 5 mM MgCl$_2$, 2 mM NaHCO$_3$, 2 mM DTT, 1 mM glucose 6-phosphate, 2 units ml$^{-1}$ malate dehydrogenase, and 0.2 mM NADH. The reaction was initiated by the addition of PEP to give 5 mM. NAD-ME was assayed by measuring NAD$^+$ reduction at 340 nm. The assay buffer contained 25 mM Tricine (pH 8.2), 1 mM EDTA, 20 mM MgCl$_2$, 2 mM DTT, 0.5 mM NADP. The reaction was initiated by the addition of malic acid to give 5 mM. NAD-ME was assayed by measuring NAD$^+$ reduction at 340 nm. The reaction mixture contained 25 mM Hepes (pH 7.2), 0.2 mM EDTA, 8 mM ammonium sulfate, 1 unit ml$^{-1}$ malate dehydrogenase, 5 mM malic acid, 0.025 mM NADH, and 2 mM NAD$^+$. The reaction was initiated by adding MnCl$_2$ to give 5 mM.

### Imaging and Quantification of Leaf Structure

For all anatomical and ultrastructural imaging, the middle portion of a leaf blade equidistant between the mid-rib and leaf margin were sampled from recent, fully expanded leaves. Leaf pieces approximately 2 mm$^2$ were collected between 9:00–11:00 am and were prepared for light and transmission electron microscopy as described previously (Khoshravesh et al., 2017). Cell and organelle features of BS cells were quantified using Image J software (Schneider et al., 2012). One leaf per plant was sampled from three to five plants. Each mean value per plant was compiled from measurements of 5–10 imaged cells from the adaxial region of the leaf. All measurements were conducted on planar images of cross (=transverse) sections. Cells measured were randomly selected from the pool of cells where the plane of section passed through the central region of a BS cell, rather than the periphery. Parameters measured include BS cell area; number of chloroplasts per BS cell; number of mitochondria per BS cell; area of individual chloroplasts and mitochondria; and% BS cell area covered by all chloroplasts and mitochondria in a BS cell.

### Phylogenomic Analysis

To prepare a phylogenomic tree of the Nyctaginaceae, we used publicly available RNA-sequence data from the NCBI Short Read Archive (Supplementary Table 4).3 FASTQ read data was trimmed using Trimmomatic (Bolger et al., 2014) with minimum leading and trailing quality cutoffs at 30, a 5-base sliding window cutoff of 30, and 70bp minimum trimmed length. Transcriptomes were de novo assembled using Trinity (Grabherr et al., 2011) with default settings including in silico read normalization. Open reading frames were predicted and translated using the getorf program from the EMBOSS package (Rice et al., 2000), and the longest open reading frame for each putative locus identified by Trinity was selected using a Python script. OrthoFinder (Emms and Kelly, 2015) was used to predict groups of orthologous genes using 10 reference proteomes with MapMan annotations (Thimm et al., 2004) and 10 with all other annotation types.

---

2www.isotopes.wsu.edu

3http://ncbi.nlm.nih.gov/sra
TABLE 1 | δ¹³C data for sampled species from the Nyctaginaceae showing number of accepted species for each genus and the number of species sampled, the range of δ¹³C for species means and number of C₄ species we identified.

| Tribe and genus      | Accepted species number/sampled species number | δ¹³C range of species means | Number of C₄ species in Supplementary Table 2 |
|---------------------|-----------------------------------------------|-----------------------------|-----------------------------------------------|
| Boldoeae            | Salpianthus Humb. & Bonpl. 5/3                | −28.8 to −27.4              | 0                                             |
|                     | Bougainvilleae                                |                             |                                               |
|                     | Belemia Pres 1/1                              | non-Kranz anatomy           | 0                                             |
|                     | Phaeoptilum Radlk. 1/1                       | −24.4                       | 0                                             |
|                     | Caribeeae                                     |                             |                                               |
|                     | Cryptocarpus H.B.K. 1/1                      | −27.2                       | 0                                             |
|                     | Colignonieae                                  |                             |                                               |
|                     | Colignonia Endl. 6/6                         | −28.7 to −25.3              | 0                                             |
|                     | Leucasterieae                                 |                             |                                               |
|                     | Andradea Allemão 1/1                         | −26.0                       | 0                                             |
|                     | Leucaster Choisy 1/1                         | −28.8                       | 0                                             |
|                     | Ramisia Glaz. ex Baillon 1/1                  | −24.3                       | 0                                             |
|                     | Nyctagineae                                   |                             |                                               |
|                     | Abronia Juss. 24/20                          | −29.5 to −24.5              | 0                                             |
|                     | Acleisanthes A. Gray 17/14                   | −28.0 to −23.9              | 0                                             |
|                     | Allionia L. 2/2                              | −13.3 to −13.2              | 2                                             |
|                     | Anulocaulis Standl. 5/5                     | −28.0 to −24.2              | 0                                             |
|                     | Boerhavia L. 40/43                          | −14.9 to −9.4               | 43                                            |
|                     | Commicarpus Standl. 25/23                    | −28.7 to −25.2              | 0                                             |
|                     | Cyphomeris Standl. 2/2                      | −28.1 to −27.2              | 0                                             |
|                     | Mirabilis L. 54-60/23                        | −29.8 to −22.6              | 0                                             |
|                     | Nyctaginia Choisy 1/1                       | −26.2                       | 0                                             |
|                     | Okenia Schildl. & Cham. 1/4                  | −14.0 to −12.3              | 4                                             |
|                     | Tripterocalyx 4                              | 27.8 to 24.9                | 0                                             |
|                     | Pisonieae                                     |                             |                                               |
|                     | Pisoniella (Heimerl) Standl. 1/1             | −25.8                       | 0                                             |
|                     | Grajalesia Miranda 1/1                      | −25.7                       | 0                                             |
|                     | Cephalotomandra Karst. & Triana 1             | −28.8                       | 0                                             |

C₄ genera are highlighted in bold. See Supplementary Tables 2,3 for herbarium specimens, collection information and individual δ¹³C values. Accepted species numbers follow Spellenberg, 2003, Tropicos (2020), and the International Plant Names Index (IPNI, 2020). Sampled species number includes species listed in Supplementary Tables 2,3 whose acceptability is uncertain. Belemia fucsiodes is a type specimen examined with a dissecting scope for vein density only.

available from the Phytozome v12 (Supplementary Table 5).¹ Single- or low-copy orthogroups were selected with a Python script based on the following criteria: a minimum of 40 of 53 species present; no more than 10% of the species having multiple sequences; and a minimum alignment length of 100 amino acids. For species with multiple sequences, a Python script was used to concatenate all orthogroup sequences from a given species if none overlapped by more than 10% of the shortest sequence length, or remove them all if they did. Sequences were assumed to represent fragmented assemblies in the former case and paralogs in the latter. This selection process resulted in 1084 genes for phylogenetic analysis. Initial protein alignments were produced with mafft (Katoh and Standley, 2013) and DNA codon alignments were generated from these using pal2nal (Suyama et al., 2006). The codon alignments were trimmed using trimAl (Capella-Gutiérrez et al., 2009) with a gap threshold of 0.5, then concatenated into a partitioned super-matrix using a Python script written by co-author M. Stata.

Phylogenetic inference was conducted using RaxML (Stamatakis, 2014). This was conducted using separate evolutionary models for all partitions, rapid bootstrap analysis with a search for the best tree in one run (command line option -f a), and bootstrap convergence testing (autoMRE). Convergence testing showed that 50 bootstrap replicates were sufficient, but in order to be certain of support values we conducted 200. The species tree was visualized using FigTree.⁵ All Python scripts are available at github.com/MattStata/Nyctaginaceae_Scripts.

C₄ Gene Phylogenies and Analysis

All homologs of PEPC, NADPME, NADME, and PPDK were identified in the Nyctaginaceae transcriptomes in the NCBI Short Read Archive (Supplementary Table 4) using BLASP with Arabidopsis sequences as queries. Matches were aligned using mafft (Katoh and Standley, 2013) and preliminary trees were trimmed.
were inferred with FastTree (Price et al., 2010) in order to identify the number of copies present in Nyctaginaceae. For each copy, sequences from the Nyctagineae clade were extracted from the alignment and a consensus sequence was generated using Geneious (www.geneious.com) to create full-length consensus references for each copy onto which RNA-seq data for all Nyctaginaceae species in the NCBI SRA were mapped using HiSat2 (Kim et al., 2019) with the scoring argument – score-min L,0,-1.4. Read mappings were scrutinized manually using the software Geneious to be certain there were no additional paralogs beyond those we had detected, which would be evident as paralogous reads mapped onto the closest reference. Consensus sequences based on mapped reads were generated in Geneious. Gene trees were inferred using MrBayes (Huelsenbeck and Ronquist, 2001) with two runs, 40 chains, 2 million generations, and a heating factor of 0.05. About 10,000 trees from the end of each run where the average SD of split frequencies remained flat at or below 0.01, were used to generate the final tree and infer posterior probabilities using the consense program included with ExaBayes (Aberer et al., 2014). Trees were visualized using FigTree.5 Gene expression values were calculated as reads per kb of transcript per million reads (RPKM) using the SAM files produced by HiSat and a Python script. Only species from the 1KP project6 submitted by our lab were used for gene expression as these represent both C4 lineages and a closely-related C3 outgroup, and were grown and sampled identically. For these, the newest fully expanded leaves were sampled during a sunny day from plants grown in a glass house at the University of Toronto between 9 am and 1 pm. Because complete mappings to all transcripts were not conducted, numbers of total sequenced reads were used in RPKM calculations rather than total mapped reads. Scripts written by M Stata are available at github.com/MattStata/Nyctaginaceae_Scripts.

Species Distribution Data
Biogeographic distributions for species of the tribe Nyctaginaceae were obtained from the Global Biodiversity Information Facility website (GBIF).7 Duplicate data points and those lacking herbarium records or corresponding to marine coordinates were removed (maptools, Bivand and Lewin-Koh, 2013). The remaining 15,870 observations represented 75 species within Allionia (two species), Boerhavia (40 species), Anulocaulis (five species), Commicarpus (25 species), Cyphomeris (two species), and Nyctaginia (one species). Bioclimatic and monthly minimum and maximum temperature parameters (2.5 min resolution) were downloaded from the WorldClim 2.0 dataset8 (Fick and Hijmans, 2017). Monthly potential evapotranspiration and Global Aridity Indexes (AI) were downloaded from CGIAR-CSI GeoPortal.9 Values per observation for 19 bioclimatic variables in the Worldclim dataset, plus minimum and maximum temperatures and AI were then extracted using the extract function in the R raster package (Hijmans and van Etten, 2012). Median values per variable per species were calculated and normalized to Z-scores for use in subsequent analyses. Climate variables that significantly predicted the occurrence of photosynthetic type at p < 0.05 were selected using stepwise regression. Mixed-effect models (R package lme4, Bates et al., 2015) were built using the selected bioclimatic variables and photosynthetic subtypes (NAD-ME or NADP-ME) as the main effect and genus and species as random effects. These models were compared by ANOVA and Akaike’s Information Criteria (AIC). The best model was selected based on the lowest AIC and p values <0.05. A principal component analysis was performed by R package FactoMineR (Lê et al., 2008) to evaluate species distribution across the multivariate predictors. A subset of the data for which we had phylogenetic data was also used to run a phylogenetically corrected ANOVA using phytool in R (Revell, 2012).

RESULTS
Carbon Isotope Ratios
For the δ13C survey, we sampled herbarium specimens from all genera of the Tribe Nyctaginaceae, except for one monospecific genus from El Salvador (Cuscutalainia vulcanicola; Table 1). All sampled species from the genera Allionia, Boerhavia, and Okenia exhibited C4 δ13C values (~9 to ~15‰) (Table 1 and Supplementary Tables 2,3). All other species exhibited C3 δ13C values (~22 to ~30‰), including all assayed species in the Nyctagineae genera Abronia, Alecisanthes, Anulocaulis, Commicarpus, Cyphomeris, Mirabilis, and Tripterocalyx. The δ13C survey provides little evidence for C3–C4 intermediate species in the Nyctaginaceae. Values from all species were clearly C4 or within the more negative range of δ13C values typical of C3 plants, with two possible exceptions – the arid zone species Alecisanthes angustifolia (δ13C = −23.9) and Mirabilis polyphylla (δ13C = −22.6). While high for a typical C3 δ13C value, these exceptions are within the range of values observed in arid-zone C3 species with high WUE (Farquhar et al., 1989).

Gas Exchange and Biochemistry
Both Allionia incarnata and B. coccinea exhibited typical C4 photosynthetic parameters. The CO2 compensation point of photosynthesis (Γ) was below 5 µmol mol−1 in both species, the ratio of intercellular to ambient CO2 concentration (Ci/Ca) ratio was 0.31–0.35, and their carboxylation efficiency of photosynthesis, measured as the initial slope of the photosynthetic response to intercellular CO2 concentration (A/Ci), response, was 5–7 times greater than the carboxylation efficiency in the C3 N. capitata (Table 2). Allionia exhibited a steeper initial slope of the A/Ci, response than Boerhavia. When relativized by dividing by the maximum net CO2 assimilation rate at high CO2 (A1200) to correct for variation in photosynthetic capacity, the normalized carboxylation efficiency in Allionia was 50% greater than in Boerhavia (Table 2).

The in vitro activity of PEPC was high in the C4 species relative to the C3 N. capitata, and similar on a leaf area basis in the two C4 species; however, on a chlorophyll basis, the PEPC activity is 63% higher in Allionia than Boerhavia (Table 2).
NADP-ME activity was high in *Boerhavia*, while NAD-ME activity was negligible in *Boerhavia* and high in *Allionia* (Table 2). Chlorophyll content was 43% higher in *Boerhavia* than *Allionia*, and chlorophyll a/b ratio was 21% higher in *Boerhavia*. Observed enzyme activities are considered robust because they are equivalent to or greater than the observed $A$ values in each species. From these results, we conclude that *Boerhavia* belongs to the NADP-ME subtype, while *Allionia* is of the NAD-ME subtype.

### Leaf Structure and Ultrastructure in *Allionia, Boerhavia* and Two *Portulaca* Species

*Allionia incarnata* and *B. coccinea* exhibited typical C$_4$-Atriplicoid Kranz anatomy, which is characterized by large BS cells in planar cross sections with centripetal organelle arrangements (Figures 3, 4; Edwards and Voznesenskaya, 2011). Notably, enlarged BS cells in both *Allionia* and *Boerhavia* do not wrap around the entire vascular bundle, mimicking patterns observed in Atriplex but not many other C$_4$ eudicots classified as having Atriplicoid Kranz anatomy (Muhaidat et al., 2007). *A. gypsogenus, C. scandens, N. capitata,* and *Mirabilis jalapa* have characteristic C$_3$ leaf anatomy (Figures 3, 4 and Supplementary Figure 1). Planar areas of the BS cells are comparable between the C$_3$ and C$_4$ species, indicating similar dimensions of the BS tissue in a radial direction with respect to the vasculature (Table 3). Both C$_4$ species exhibit greater coverage of the BS cell by chloroplasts than the C$_3$ species, because there are more chloroplasts per BS cell and greater mean area per chloroplast in the C$_4$ species (Figure 4; Table 3). In *A. incarnata*, the planar area of BS cells covered by mitochondria is significantly greater than in C$_3$ species and the C$_4$ *B. coccinea*, due to the presence of larger mitochondria in planar section, and more mitochondria per cell area (Table 3; Figures 4A,C). The percent mitochondrial

![Image](https://example.com/image.png)

**FIGURE 2** | The response of the net CO$_2$ assimilation rate to intercellular CO$_2$ concentration in *Allionia incarnata* (C$_4$), *Boerhavia coccinea* (C$_4$) and *Nyctaginea capitata* (C$_3$) at $30^\circ$C and a light intensity of 1800 µmol photons m$^{-2}$ s$^{-1}$. Means ± SE, N = 4 for the C$_4$ species or 1 for the C$_3$ species.

### Table 2 | Summary of gas exchange and biochemical data for three Nyctagineaceae species.

| Parameter | Nyctaginea capitata | Boerhavia coccinea | Allionia incarnata | C$_4$ p value (one-tailed) |
|-----------|---------------------|-------------------|--------------------|--------------------------|
| Net CO$_2$ assimilation rate at 400 µmol CO$_2$ mol$^{-1}$ air, in µmol m$^{-2}$ s$^{-1}$ | 7.0 | 23.0 ± 2.1 | 26.8 ± 0.9 | 0.08 |
| Net CO$_2$ assimilation rate at 1200 µmol CO$_2$ mol$^{-1}$ air, in µmol m$^{-2}$ s$^{-1}$ | 23.5 | 31.3 ± 0.7 | 30.5 ± 1.5 | 0.32 |
| $A_{400/1200}$ | 0.30 | 0.74 ± 0.06 | 0.88 ± 0.03 | 0.04 |
| A/$g_{s}$, mmol CO$_2$ mol$^{-1}$ H$_2$O | 100 | 158 ± 3 | 149 ± 5 | 0.10 |
| C/Chl | 0.58 | 0.31 ± 0.01 | 0.35 ± 0.02 | 0.10 |
| Initial slope of the A vs C$_d$ curve, mol m$^{-2}$ s$^{-1}$ | 0.05 | 0.25 ± 0.03 | 0.36 ± 0.04 | 0.04 |
| Initial slope/A$_{1200}$ | 0.002 | 0.008 ± 0.001 | 0.012 ± 0.001 | 0.03 |
| CO$_2$ compensation point of A, µmol mol$^{-1}$ | 75 | 4.5 ± 0.5 | 2.5 ± 1.0 | 0.06 |
| PEP carboxylase activity, µmol m$^{-2}$ s$^{-1}$ | 15.1 ± 1.9 | 111.8 ± 7.9 | 131.7 ± 22.2 | 0.22 |
| PEP carboxylase activity, mmol mol$^{-1}$ CHL s$^{-1}$ | 27.0 ± 2.7 | 261.9 ± 12.4 | 429.7 ± 53.2 | 0.01 |
| NADP-ME activity, µmol m$^{-2}$ s$^{-1}$ | 0 ± 0 | 50.1 ± 5.6 | 0 ± 0 | < 0.001 |
| NADP-ME activity, mmol mol$^{-1}$ CHL s$^{-1}$ | 0 ± 0 | 116.7 ± 8.9 | 0 ± 0 | < 0.001 |
| NAD-ME activity, µmol m$^{-2}$ s$^{-1}$ | 0 ± 0 | 2.1 ± 0.6 | 27.7 ± 2.8 | < 0.001 |
| NAD-ME activity, mmol mol$^{-1}$ CHL s$^{-1}$ | 0 ± 0 | 5.1 ± 1.4 | 91.5 ± 6.4 | < 0.001 |
| Chlorophyll, mmol m$^{-2}$ | 0.56 ± 0.02 | 0.43 ± 0.02 | 0.30 ± 0.01 | < 0.001 |
| Chlorophyll a/b ratio | 3.46 ± 0.07 | 4.45 ± 0.09 | 3.67 ± 0.08 | < 0.001 |

Means ± SE, N = 4 for the C$_4$ species, 1 for *N. capitata* gas exchange data, and 3 for *N. capitata* biochemical data. The p-values of one-tailed Students T-tests between the two C$_4$ species are indicated, with significant differences highlighted in bold. Gas exchange measurements were conducted at $33^\circ$C and a light intensity of 1800 µmol m$^{-2}$ s$^{-1}$. Biochemical assays were conducted at $30^\circ$C. A, net CO$_2$ assimilation rate; C$_d$, intercellular CO$_2$ concentration; C$_b$, ambient CO$_2$ concentration in the leaf chamber; $A_{400/1200}$, A at a CO$_2$ concentration of 400 µmol CO$_2$ mol$^{-1}$ air; A$_{1200}$, A at a CO$_2$ concentration of 1200 µmol mol$^{-1}$; CHL, leaf chlorophyll content; g$_{s}$, leaf conductance to water vapor; ME, malic enzyme.
coverage of the BS cells is similar in B. coccinea and the C₃ species (Table 3; Figures 4B,D–F).

The BS cells in species with C₃ δ¹³C values have smaller and fewer chloroplasts than the C₄ species of the Nyctagineae, and these are positioned mostly along the outer BS wall exposed to intercellular airspace (Figures 4E,F and Supplementary Figure 1). None of the three C₃ species examined exhibit traits associated with C₃–C₄ intermediacy or even the incipient intermediate state termed “proto-Kranz”; there was not a greater mitochondrial number, nor a repositioning of mitochondria and chloroplasts toward the inner side of the BS cells. Chloroplasts in the BS of Boerhavia and the M tissue of Allionia are largely agranal, whereas distinct levels of grana stacking are apparent in the thylakoids of the BS chloroplasts in Allionia, and the M chloroplasts of Boerhavia (Figure 5).

In P. oleracea (NAD-ME), chloroplasts with pronounced grana stacks cluster in the inner BS (Figure 6A). Many mitochondria are interspersed between the chloroplasts, and exhibit distinct structural connections with nearby chloroplasts (Figure 6B). In P. pilosa (NADP-ME), chloroplasts also occur in the inner BS where they form clusters with no discernable thylakoid stacks (Figures 6C,D). Mitochondria are largely absent between chloroplasts, although some mitochondria occur along the inner wall of the BS in a pattern that is commonly observed in C₃–C₄ intermediate species (Figure 6C).

**Species Phylogeny**

We update the molecular phylogeny of the Nyctagineae using transcriptome sequence data available at the NCBI short read archive (Figure 7). The tree largely replicates the phylogenies of Douglas and Manos (2007) and Douglas and Spellenberg (2010) showing the C₄ and C₃ clades of the Nyctagineae correspond to a clade of xerophytic herbs and shrubs that Douglas and Spellenberg term the North American Xeric (NAX) clade. C₄ Allionia branches with Cyphomeris at the base of a clade that includes the isotopically C₃ Anulocaulis and Nyctaginia species, and the C₄ Okenia and Boerhavia species. Anulocaulis and Nyctaginia form a clade that branches between Allionia/Cyphomeris and Boerhavia/Okenia. C. scandens branches at the base of the clade containing the C₄ species and their immediate non-C₄ sister clades, with Mirabilis species branching just below C. scandens. Unlike Douglas and Spellenberg (2010) who predicted Bougainvilleaeae + Pisoniaeae to be sister to the Nyctagineae, we predict only Bougainvilleaeae branches in a sister position, with maximal support. Also, the branching order of Phytolaccaceae, Sarcobataceae, Gisekiaceae, and Nyctaginaceae is unclear in the literature and in the APG
Caryophyllales tree, where they form a polytomy. Our tree resolves them with 100% support, improving the phylogenetic context for the family.

Gene Phylogenies
We examined the trees of four important C4 cycle genes – PEPCase, NADP-ME, NAD-ME, and PPDK (Figure 8 and Supplementary Figures 2–5). For each gene, we assumed that the functional copy in the C4 pathway was the one with the highest expression in the transcriptome analysis. In this analysis, we numbered gene copies based on branching order from the base of the gene tree; the numbers assigned are not meant to imply direct orthology with gene copies from any other lineage. For Allionia incarnata and two Boerhavia species, the PEPC1 gene is the copy used in the C4 pathway because it shows an expression strength that is orders of magnitude greater than PEPC2 and PEPC3 (Table 4). By similar logic, NAD-ME3 is the main decarboxylase gene in the C4 pathway of Allionia, and NADP-ME2 is the main decarboxylase gene in Boerhavia. Allionia also exhibited significant expression of NADP-ME1, approaching that of NAD-ME3 (Table 4). The expression of the PEP carboxykinase gene was minimally detectable in both C4 clades (not shown). In the gene trees for PEPC1, PPDK and NADP-ME1, the respective orthologs from numerous C3 species of the Nyctaginaceae branch between the two C4 lineages, with strong support (Figure 8 and Supplementary Figures 2–5). This indicates the C4 genes arose independently from an ancestral C3 copy, rather than by lateral transfer between clades. If one or more of the genes had moved laterally between the C4 clades, the orthologs from both C4 lineages would form a common clade.

Positively Selected Amino Acid Substitutions
We examined PEPC1 sequences for evidence of convergence in the C4 isoforms (Table 5). Christin et al. (2008) showed that 21 amino acid sites were under positive selection in the PEPC1 of various C4 grasses, with the best-known example of sequence convergence being a serine for alanine substitution near the maize position 774 in the PEPC1 sequence (position 774 in the Flaveria sequence; Gowik and Westhoff, 2011). We did not find evidence for convergence with other C4 lineages at position 780 in Allionia or Boerhavia, as they both exhibited an alanine, which is typical of C3 isoforms (Table 5 and Supplementary Figure 6). There was consistent convergence at sites 572, 761, and 807, where all C4 species exhibited the same amino acids as those under positive selection in C4 grasses. Convergence was observed in both C4 lineages at site 813, but in only one of the six Boerhavia/Okenia species in the database (Table 5). Boerhavia also converged on the same amino acids as the C4 grass species at site 733 and 863, and partially at position 502 (B. purpureascens and B. torreyana only). In total, of the 21 C3 to C4 amino acid switches repeatedly observed in the grass PEPCase sequences (Christin et al., 2007), about a third were replicated in the Boerhavia lineage and a fifth in the Allionia lineage (Table 5). At one site, position 577, the C3 and C4 Nyctaginaceae species share a serine with the C4 grass Zea mays. In the grasses, most C3 species examined by Christin et al. (2007, 2012b) exhibit an alanine at this site. It is possible that C4 evolution in Nyctaginaceae did not require this substitution as the ancestral C3 state was already a serine.

Ecological Distribution
Nyctaginaceae species from the genera Allionia, Boerhavia, Anilocaulis, Commicarpus, Cyphomeris, and Nyctaginia are distributed in tropical and subtropical regions of similar climate (Supplementary Figures 7–9). Six bioclimatic variables significantly predicted (p < 0.05) the C3 and C4 distribution using a stepwise regression model (Supplementary Figure 8 and Supplementary Table 6). To address whether the variation in distribution of C3 and C4 species could be species or genera dependent, GLMM models were performed with and without considering taxa as random effects. The model was significantly improved (p < 2e-16) when genus was added as the random effect (Supplementary Table 6); however, after this addition, none of the main effects significantly predicted the pattern of photosynthetic pathway distribution.

10www.mobot.org/MOBOT/research/APweb

FIGURE 4 | Transmission electron micrographs illustrating bundle sheath cells of C3 and C4 Nyctaginaceae species from the tribe Nyctagineae. (A–D) C4 species: (A,C) Allionia incarnata, (B,D) Boerhavia coccinea; (C,D) high magnification images of the bundle sheath cells. Note high numbers of mitochondria in Allionia incarnata (C). (E,F) The C4 species Commicarpus scandens (E), and Nyctaginia capitata (F). C, chloroplast; arrowheads mark mitochondria. Scale bars = 10 μm.
TABLE 3 | Ultrastructural parameters for the bundle sheath cells in cross section of two C₄ species and three C₃ species from the Nyctagineae tribe of the Nyctaginaceae.

|                    | A. incarnata | An. gypsogenus | B. coccinea | C. scandens | N. capitata |
|--------------------|--------------|----------------|-------------|-------------|-------------|
|                    | C₄           | C₃             | C₄          | C₃          | C₃          |
| Cell area, µm²      | 427 ± 169 a  | 727 ± 339 a    | 557 ± 183 a | 505 ± 199 a | 944 ± 389 a |
| % Cell covered by chloroplast, % | 28.3 ± 5.9 b | 2.3 ± 0.9 a   | 35.2 ± 11.5 b | 13.8 ± 4.6 a | 6.5 ± 4.2 a |
| % Cell area covered by mitochondria, % | 3.6 ± 1.7 b | 0.4 ± 0.2 a | 1.2 ± 0.5 a | 0.6 ± 0.3 a | 0.5 ± 0.4 a |
| Number of chloroplasts per cell | 16 (6–24) b | 5 (2–11) a | 18 (3–37) c | 10 (7–15) b | 10 (5–17) b |
| Number of mitochondria per cell | 45 (12–111) c | 14 (4–57) a | 28 (4.8–75) b | 13 (5–30) a | 11 (7–52) a |
| Area per chloroplast, µm² | 8.3 ± 3 c | 2.9 ± 1.1 ab | 11.4 ± 3.8 c | 6.1 ± 2.1 bc | 5.3 ± 1.2 bc |
| Area per mitochondria, µm² | 0.7 ± 0.2 b | 0.3 ± 0.1 a | 0.4 ± 0.1 a | 0.5 ± 0.2 ab | 0.3 ± 0.2 a |

Means ± SE, or median ± range for organelle numbers. N = 3–5. Letters besides values indicate statistically distinct groups by Kruskal Wallis test followed by a Dunn’s test except for organelle numbers which are evaluated using a Poisson regression for count data (number of organelles per cell). Abbreviations: A, Allionia; An, Anulocaulis; B, Boerhavia, C, Commicarpus; N, Nyctaginia.

FIGURE 5 | Transmission electron micrographs illustrating chloroplast fine structure. (A) Boerhavia burbridgeana bundle sheath cell; (B) B. burbridgeana mesophyll cell; (C) Allionia incarnata bundle sheath cell; (D) A. incarnata mesophyll cell. In panels (A,C), thylakoids are lighter stacks and striations against a dark stroma. In panels (B,D), thylakoids are dark stacks and striations against a lighter-staining stroma. Scale bars, 0.5 µm. C, chloroplast; arrows mark stacked thylakoids. Scale bars = 0.5 µm.
A phylogenetically-corrected ANOVA also failed to detect a significant difference between the C₃ and C₄ species in response to six variables (Supplementary Table 7). These analyses show that the distribution of Nyctaginaceae species in climate and geographic space follows taxonomic and phylogenetic affinities rather than photosynthetic pathway.

DISCUSSION

Carbon isotope ratios demonstrate that all examined species in Allionia, Boerhavia and Okenia are C₄, while all examined species in other Nyctaginaceae genera are not, including species in Anulocaulis, Commicarpus, Cyphomeris, and Nyctaginia that...
Photosynthesis in the Nyctaginaceae

FIGURE 7 | Phylotranscriptomic reconstruction of Nyctaginaceae. A maximum likelihood phylogeny of the Nyctaginaceae family and six outgroup families with previously uncertain topology. This tree is based on a concatenated super-matrix of 847 codon-aligned single-copy orthologs totaling 834,018 sites. All bootstrap support values are 100% except for the two nodes with values indicated, based on 200 bootstrap replicates. Within the Nyctagineae (identified as the North American Xerophytic clade by Douglas and Spellenberg, 2010), C_4 lineages are denoted with red branches. The tree is rooted on Hypertelis cerviana and gapped branches here were shortened for visual convenience.

TABLE 5 | Comparison of amino acids in the C_4 PEP carboxylases of the Allionia and Okenia/Boerhavia C_4 clades with PEP carboxylase sites under positive selection in C_4 grasses.

| Sites with posterior probability > 0.999 | Prob > 0.99 | Prob > 0.95 |
|---------------------------------------|------------|------------|
| Clade                                 | Sites      | 466  | 517  | 531  | 560  | 577  | 579  | 625  | 637  | 671  | 780  | 794  | 807  | 572  | 599  | 813  | 665  | 733  | 863  | 866  |
| Boerhavia/Okenia                      | Met        | Thr   | Ala   | Arg   | Ser   | Ala   | Leu   | Met   | Ala   | Ala   | Phe   | Lys   | Gin   | Ile   | Arg   | Gin   | His   | Met   | Lys   | Gin   |      |
| Allionia                              | Met        | Thr   | Ala   | Arg   | Ser   | Ala   | Val   | Met   | Ala   | Ala   | Phe   | Lys   | Gin   | Ile   | Arg   | Gin   | His   | Phe   | Asn   | Gin   |      |
| C_3 Nyctag                            | Met        | Thr   | Ala   | Arg   | Ser   | Ala   | Val   | Met   | Ser   | Ala   | Phe   | Arg   | Gin   | Ile   | Arg   | Gin   | His   | Phe   | Asn   | Gin   |      |
| C_4 grasses (Christin et al. 2007)    | Ile, Val   | Ala   | Cys   | Pro   | Pro   | Ser   | Glu   | Ala   | Leu   | Phe   | Ala   | Ser   | Val   | Lys   | Gin   | Val   | Asn   | Met   | Val   | Lys   | Gin   |

Site numbers correspond to the maize PEP carboxylase gene (PEPC). Sites under positive selection were identified by a screen of grass lineages examined by Christin et al. (2007). Posterior probabilities noted with blue headers are from that study. Sites showing an identical amino acid substitution in either or both C_4 Nyctaginaceae lineages are highlighted in yellow. We list the most frequent positively selected amino acid(s) at each site in C_4 grasses. The species in the C_3 Nyctag clade are the North American Xerophytic C_4 species in the PEPC2 gene tree in Supplementary Figure 2, which exhibited identical amino acids at the sites shown. *Gln is only present in B. burdigeana; the other Boerhavia and Okenia have Arg at position 813.

branch sister to the C_4 clades. The number of C_4 species in the Nyctaginaceae is 43–46 based on recent assessments that conclude there are 40 Boerhavia, one to four Okenia, and two Allionia species (Spellenberg, 2003; Tropicos, 2020). The typically C_3 δ^{13}C values in the 31 examined species of Anulocalais, Commicarpus, Cyphomeris, and Nyctaginia indicate that none of the examined species have a type of C_3–C_4 intermediacy termed C_4-like, where a strong C_4 metabolic cycle has been engaged. As a C_4 cycle becomes engaged, the values of δ^{13}C increase from C_3 toward C_4 values, becoming distinct from typical C_3 values when δ^{13}C rises above −22‰ (Von Caemmerer, 1992; Alonso-Cantabrana and von Caemmerer, 2016). This occurs because PEPCase does not discriminate against the C_3 isotope to the degree that Rubisco does (Farquhar et al., 1989). Carbon isotope screens cannot differentiate between C_3 plants and a type of C_3–C_4 intermediacy utilizing C_2 photosynthesis,
Photosynthesis in the Nyctaginaceae

FIGURE 8 | A Bayesian gene phylogeny for the PEPCase gene copy co-opted for C<sub>4</sub> function in the Nyctaginaceae. Node values are posterior probabilities. Red branches denote C<sub>4</sub> species. A tree with all three PEPC paralogs is presented in Supplementary Figure 2.

a metabolic pathway that concentrates CO<sub>2</sub> into the BS using photorespiratory metabolites to shuttle CO<sub>2</sub> from M to BS cells. In C<sub>2</sub> photosynthesis, the photorespiratory enzyme glycine decarboxylase is expressed only in BS mitochondria, which forces photorespiratory glycine to diffuse from the M cells where it is formed to the BS cell for metabolism (Sage et al., 2012). The released CO<sub>2</sub> from glycine decarboxylation accumulates in BS cells, where it can be refixed by Rubisco in adjacent chloroplasts with high efficiency. C<sub>2</sub> photosynthesis improves carbon gain at low CO<sub>2</sub> concentrations, but because all CO<sub>2</sub> is fixed by Rubisco, the δ<sup>13</sup>C values of C<sub>2</sub> plants reflect those of C<sub>3</sub> plants (Von Caemmerer, 1992). Anatomically, C<sub>2</sub> plants have increased numbers of mitochondria and chloroplasts in BS cells, typically in a centripetal position against the BS wall facing the vascular tissue (Khoshravesh et al., 2016). This characteristic identifies candidate C<sub>2</sub> species, with low CO<sub>2</sub> compensation points of photosynthesis (Γ) confirming the presence of the C<sub>2</sub> physiology. Our examination of the BS structure in Anulocaulis, Commicarpus and Nyctaginia, and leaf gas exchange in Nyctaginia, showed no evidence of C<sub>2</sub> photosynthesis as the structural characteristics and Γ were typically C<sub>3</sub>. Hence, we conclude that the sister clades to Allonia and Boerhavia/Okenia clades are most likely comprised of only C<sub>3</sub> plants. Gas exchange data and anatomical studies demonstrate strong C<sub>4</sub> features in selected species of Allonia and Boerhavia, which along with the consistently C<sub>4</sub> values of δ<sup>13</sup>C in these clades, lead us to conclude they are completely C<sub>4</sub> rather than C<sub>3</sub>-like. C<sub>4</sub>-like species have a fully functional C<sub>4</sub> cycle but retain limited C<sub>3</sub> photosynthesis in M tissues, and exhibit features suggesting they are newly evolved C<sub>4</sub> plants that have not yet optimized the C<sub>4</sub> pathway (Moore et al., 1989).

In the phylogeny, Anulocaulis and Nyctaginia species occur in a clade that branches sister to the Boerhavia/Okenia clade, while Cyphomeris branches sister to Allonia, supporting a hypothesis of two independent C<sub>4</sub> origins, one in ancestral Allonia and a second in ancestors to the Okenia/Boerhavia clade. An alternative possibility is a single C<sub>4</sub> origin followed by reversions to C<sub>3</sub> in ancestors of Cyphomeris and the Anulocaulis/Nyctaginia clade. This option is less parsimonious, requiring three changes – one acquisition of C<sub>4</sub> and two reversions – rather than two. It is also not favored because reversions are considered unlikely and have never been confirmed (Christin et al., 2010a; Oakley et al., 2014; Bräutigam and Gowik, 2016). In addition, the two C<sub>4</sub> clades exhibit different biochemical and structural subtypes. The high degree of structural, transport, and biochemical specialization associated with each subtype indicates switching subtypes would be difficult, and consistently, no instances of subtype switching have been documented. Gene sequences of C<sub>4</sub> pathway enzymes can also be used to assess reversal possibilities. If a reversion had occurred, the derived ortholog of a C<sub>4</sub>-pathway enzyme in any C<sub>3</sub> progeny might retain sequences from their ancestral C<sub>4</sub> function (Christin et al., 2010a; Khoshravesh et al., 2020). In Anulocaulis, Cyphomeris, and Nyctaginia, the C<sub>4</sub>-type orthologs of PEPCase, PPDK, and both decarboxylases exhibited no signatures of prior C<sub>4</sub> function. For example, the C<sub>3</sub> species in the Nyctagineae share only one positively selected amino acid substitution with the C<sub>4</sub> Nyctaginaceae and Zea mays in the PEPC1 sequence, a serine at position 577. C<sub>3</sub> species outside the clade where an ancestral C<sub>4</sub> transition could have occurred also share this serine,
indicating it is not a relic from a C4-state. Based on these points, we conclude that Allionia and Boerhavia represent independent yet closely related C4 clades that have diverged in terms of biochemical subtype. With other examples of subtype divergence in closely related species of Portulaca and possibly other clades (Salsola and Sesuvioideae; Voznesenskaya et al., 1999; Bohley et al., 2015), comparative methods of evolutionary biology can be used to address questions of convergence and divergence during C4 evolution.

In contrast to a prior conclusion that both Boerhavia and Allionia are NADP-ME (Muhaidat et al., 2007), the biochemical assays demonstrate species in these two clades are different biochemical subtypes of C4 photosynthesis. Consistently, Allionia and Boerhavia showed pronounced differences in BS and M ultrastructure that support their designation as NAD-ME and NADP-ME subtypes, respectively. In both B. coccinea (high NADP-ME activity, low NAD-ME activity) and A. incarnata (high NAD-ME activity, nil NADP-ME activity), chloroplasts and mitochondria occupy the inner half of the BS cells, as is typical in eudicot C4 species (Edwards and Voznesenskaya, 2011). In A. incarnata, there is an abundance of mitochondria scattered among the elongated chloroplasts, while in B. coccinea, BS mitochondria are much less frequent and not commonly interspersed within the chloroplast cluster. In NAD-ME species, the juxtaposition of chloroplasts and mitochondria facilitates rapid movement of CO2 released from mitochondria into adjacent chloroplasts, while the tight packing of chloroplasts and mitochondria reduces the chance of CO2 escape. This tight packing is a major means by which CO2 is trapped in species lacking a suberized layer around the BS wall, as is the case in eudicots (von Caemmerer and Furbank, 2003). In Portulaca species, the same patterns hold put with one important variation that is not reported in the literature, notably, mitochondria in the BS of NAD-ME P. oleracea form distinct attachment structures to adjacent chloroplasts that are not obvious in Allionia and other NAD-ME type C4 species (for example, Anticharis, Cleome, Salsola, and Suaeda; Voznesenskaya et al., 1999, 2018; Khoshravesh et al., 2012). Such structures may facilitate rapid CO2 diffusion between mitochondria and chloroplasts, enhancing refixation efficiency.

The structural characteristics observed here are consistent with patterns in other C4 clades. NADP-ME species among C4 grasses, sedges and eudicots (as shown in Flaveria, Euphorbia, Gymnkhema, Heliotropium, Salsola, and Tribulus) share with Boerhavia and Portulaca pilosa the pattern of enlarged chloroplasts with weakly developed grana stacks, and few BS mitochondria (Carolin et al., 1978; Kim and Fisher, 1990; Ueno, 1996, 2013; Voznesenskaya et al., 1999; Yoshimura et al., 2004; Muhaïdat et al., 2011; Sage T. L. et al., 2011; Lauterbach et al., 2019). NAD-ME eudicots in Amaranthus, Atriplex, Anticharis, Cleome, Gisekia, Salsola, Suaeda, and Tecticornia share with Allionia and P. oleracea the pattern of enlarged chloroplasts with well-developed grana and large numbers of interspersed mitochondria (Carolin et al., 1978; Voznesenskaya et al., 1999, 2007, 2008; Khoshravesh et al., 2012; Bissinger et al., 2014; Oakley et al., 2014). In grasses, Dengler and Nelson (1999) describe NAD-ME species as having 5-to 20-fold more mitochondria than NADP-ME species, and NAD-ME mitochondria are often larger with greater internal membrane surface area in the BS. The differences between C4 subtypes hold even when different tissue layers are co-opted as the site of CO2 concentration, whether it is the mesostome sheath as in grasses, an inner chlorenchyma layer as occurs in Salsola and other succulent chenopods, or single-cell type of C4 photosynthesis as shown in the NAD-ME chenopods Biennertia cycloptera and Saucita aralocapica (Ueno, 1996, 2013; Voznesenskaya et al., 1999; Pyankov et al., 2000; Edwards and Voznesenskaya, 2011; Khoshravesh et al., 2020). In the middle of M cells of Biennertia cycloptera, for example, there are many mitochondria within a ball of Rubisco-containing chloroplasts where CO2 is concentrated (Voznesenskaya et al., 2002). The central chloroplasts in B. cycloptera have well-stacked grana, while peripheral chloroplast that are functionally equivalent to M chloroplasts of typical C4 species do not.

Lateral transfer of genes encoding C4 pathway elements has been observed in closely related grass clades within Alloteropsis and Neurachne (Christin et al., 2012a,b; Dunning et al., 2017; Khoshravesh et al., 2020). Lateral gene transfer is relevant to discussions of evolutionary convergence because the acquisition of previously evolved C4 genes by a non-C4 relative could facilitate C4 evolution in the receptive clade, possibly creating a false impression of convergence. When we examined the phylogenies of C4 pathway genes in the Nyctaginaceae, we found no evidence of lateral transfer. The gene trees consistently showed the two C4 lineages do not share a common gene copy of a C4-adapted isoforms of PEPC, PPDK, or either malic enzyme, because the respective orthologs from related C3 species branch between the C4 lineages with high support (Figure 8 and Supplementary Figures 2–5). This evidence is consistent with a hypothesis that the two C4 clades in the Nyctaginaceae arose de novo from a completely C3 state, rather than via assisted origins involving gene introgression from a pre-existing C4 clade (Christin et al., 2012a,b; Dunning et al., 2017).

With sequencing data becoming widely available, it may be possible to identify C4-subtypes using transcriptomics (Lauterbach et al., 2019). Such an approach, however, could produce errors if not coupled with enzyme activity assays. Here, the transcriptomes consistently showed high expression of the copies we designate as NADP-ME1 in Boerhavia and NADP-ME in Allionia. Allionia also exhibited significant expression of NADP-ME transcripts which approach expression of its NADP-ME gene. This data by itself would suggest co-function of NADP-ME and NAD-ME in Allionia; however, no NADP-ME activity was detected in Allionia, leading us to conclude its NADP-ME transcripts are not translated into functional enzyme. The possibility of error in this conclusion is unlikely because the BS ultrastructure in Allionia is consistent with patterns generally seen in NAD-ME subtypes. The high transcript level of both NAD-ME and NADP-ME may be a relic from ancestral Allionia species that may have utilized both decarboxylases during an early phase of C4 photosynthesis. A chance selection event may have subsequently enhanced NAD-ME activity, after which leaf structure and energetics were optimized for the NAD-ME subtype.
One of the notable features of C<sub>4</sub> photosynthesis is that it is concentrated in relatively few orders of higher plants. The Poideae (grasses, sedges) and Caryophyllales, for example, account for over 90% of all C<sub>4</sub> species and about 50 of the estimated 65 independent origins of C<sub>4</sub> photosynthesis (Sage, 2016). The ability of these clades to repeatedly evolve the C<sub>4</sub> pathway is hypothesized to reflect the presence of enabling traits that facilitate C<sub>4</sub> evolution (Sage, 2004; Christin et al., 2013b).

Wider BS cells in C<sub>3</sub> species, for example, could be a structural enabling trait (Muhiadat et al., 2011; Christin et al., 2013b; Griffiths et al., 2013), while greater numbers of organelles in BS tissues may reflect an activation of photosynthetic physiology which establishes an efficient system of photorespiratory glycine shuttles (Sage et al., 2013; Schulze et al., 2013). Photosynthetic activation of the BS can facilitate the rise of photorespiratory glycine shuttles because mitochondria and some chloroplasts can reposition to the inner BS region, forcing glycine formed in centrifugal chloroplasts to migrate to the inner BS for processing by glycine decarboxylase (Sage et al., 2013). To evaluate whether enabling traits are present in close C<sub>3</sub> relatives of the C<sub>4</sub> Nyctaginaceae clades, we examined their BS structure and organelle characteristics. In the C<sub>3</sub> species of *Anulocaulis*, *Cyphomeris*, and *Nyctaginia*, BS cells had similar cross-sectional areas as the C<sub>4</sub> BS cells, and exhibited numerous chloroplasts along the outer wall, indicating photosynthetic activation. We thus conclude the close C<sub>3</sub> relatives of *Allionia* and *Boerhavia* exhibit numerous enabling traits for C<sub>4</sub> evolution, indicating they were present in ancestral C<sub>3</sub> taxa from which the C<sub>4</sub> clades arose.

### C<sub>4</sub> Selection Environments

The results indicate C<sub>4</sub> photosynthesis evolved in Nyctaginaceae species from hot and dry climates of the New World, most likely in the arid-to-semi-arid regions of Southwestern North America where the center of diversity occurs for the North American Xerophytic clade of the Nyctaginaceae (Douglas and Manos, 2007). Christin et al. (2011) estimate that the C<sub>4</sub> pathway appeared in the *Boerhavia/Okenia* clade about 4.7 million years ago, and 6.1 million years ago in *Allionia*, based on molecular phylogenies. This was during a period of aridification in Southwestern North American and reduced atmospheric CO<sub>2</sub> (Sage et al., 2018). The environmental habitat of *Allionia* versus *Boerhavia* did not differ much from each other or their C<sub>3</sub> ancestors, indicating the C<sub>3</sub> ancestors were adapted to the same kind of hot, dry environments where the C<sub>4</sub> species are common. This supports a hypothesis that C<sub>4</sub> arose in these sorts of environments, consistent with a model that high levels of photorespiration brought on by heat, drought and/or salinity promoted C<sub>4</sub> evolution in C<sub>3</sub> plants pre-adapted to such harsh environments (Sage et al., 2018).

Surveys of the floristic distribution of NADP-ME versus NAD-ME grasses have identified a trend where NADP-ME species predominate in grass floras of wetter climates while NAD-ME species predominate in drier climates (Vogel et al., 1978; Hattersley, 1992; Ghannoum et al., 2011). The reasons for this trend have not been clarified, although it has been hypothesized that the pattern reflects phylogenetic ancestry where NAD-ME species are more likely to arise in clades from drier areas while NADP-ME species evolve in wetter climate zones (Taub, 2000). Explanations for habitat differences between C<sub>4</sub> subtypes can be evaluated using closely related NADP-ME and NAD-ME clades. In *A. incarnata* and *B. coccinea*, we observed similar photosynthetic capacities and intrinsic WUE (A/g<sub>s</sub>), indicating no obvious differences in photosynthetic parameters that might explain subtype segregation along a moisture gradient. *Allionia* did exhibit a steeper initial slope of the A/C<sub>i</sub> response than *Boerhavia*, indicating a stronger C<sub>4</sub> metabolic cycle in the NAD-ME plant. Consistently, the NAD-ME clade in *Allionia* had greater PEPCase activity per unit chlorophyll than *Boerhavia* (NAD-ME clade). If a stronger metabolic cycle is an inherent feature of NAD-ME relative to NADP-ME species, there could be a photosynthetic advantage under low intercellular CO<sub>2</sub> concentrations that commonly occur where drought and low humidity reduce stomatal aperture. A comparative analysis using closely related species of differing subtype in the Nyctaginaceae, *Portulaca* and other lineages could evaluate this possibility.

### Convergence Versus Non-convergence in the C<sub>4</sub> Functional Type

As a highly convergent, complex trait, C<sub>4</sub> photosynthesis represents an excellent system to dissect evolutionary convergence, particularly the degree of convergence versus divergence in the mix of traits that give rise to a composite phenotype (Heyduk et al., 2019). Numerous studies have previously examined convergent properties of C<sub>4</sub> components, such as converge in genes for C<sub>4</sub> enzymes (Christin et al., 2007, 2009, 2010b; Emms et al., 2016); regulatory components (Gowik and Westhoff, 2011; John et al., 2014; Reyna-Llorens and Hibberd, 2017); structural features (Yoshimura et al., 2004; Kadereit et al., 2014; Stata et al., 2014; Danila et al., 2018); and biochemical sub-types (Gutierrez et al., 1974; Hatch et al., 1975; Sage R. F. et al., 2011; Ludwig, 2016b). However, the assembly of multiple datasets into a hierarchical framework has not been attempted in a C<sub>4</sub> context. In Table 6, we present a preliminary hierarchy of convergent and divergent traits observed in comparative studies of the many C<sub>4</sub> clades, thereby enabling deeper assessments of when and why convergence is favored, versus when divergent solutions to CO<sub>2</sub> concentration can occur. Convergence within C<sub>4</sub> photosynthesis is apparent in two key steps: the carboxylation of PEP by PEPCase, and the conversion of CO<sub>2</sub> to bicarbonate by CA. The ubiquitous use of PEPCase is probably due to three factors. First, it is readily available for co-option, because it is widely used in C<sub>3</sub> plants for processes such as pH control, metabolite generation for the Krebs cycle and nitrogen assimilation, and shuttling reducing power between cellular compartments (Aubry et al., 2011; Mallmann et al., 2014). Second, there are no obvious alternatives in vascular plants that may be co-opted to provide the carboxylation function of a C<sub>4</sub> cycle. While numerous carboxylases are active in prokaryotes, few occur in higher plants (Erb, 2011). One obvious candidate to recruit into a C<sub>4</sub> cycle is pyruvate carboxylase, which produces OAA from pyruvate and bicarbonate in bacteria, animals and algae, but it is not known to occur in plants (Tsuij et al., 2012). Third, it is worth considering the chain of events giving rise to C<sub>4</sub> photosynthesis. The evolutionary rise of C<sub>2</sub> photosynthesis in C<sub>3</sub>–C<sub>4</sub> intermediate
species establishes a Kranz-like leaf anatomy and M to BS transport systems that facilitate the subsequent upregulation of PEPCase and a C₄ metabolic cycle (Sage et al., 2012). A leading hypothesis to explain this upregulation is PEPCase provides carbon skeletons to support re-assimilation of photorespiratory ammonia (Mallmann et al., 2014). If so, then the ubiquitous use of PEPCase may arise out of its pre-existing role supporting N metabolism in C₃ leaves, which make it the ready candidate for optimizing the performance of C₂ photosynthesis, which in turn positions it to be further upregulated as the benefits of the nascent C₄ cycle improve fitness (Heckmann et al., 2013). The convergence upon enhanced CA activity then becomes necessary to enable PEPCase to have enough bicarbonate to maintain rapid activity.

As the most heavily studied enzyme in C₄ photosynthesis, the sequence of the PEPCase gene serves as a model of how convergence versus divergence can operate at the gene sequence level. Convergence is apparent in that PEPC1 was separately co-opted for C₄ function in the Allionia and Okenia/Boerhavia lineages, while divergence is apparent in the sequence of PEPC1 in these clades, as indicated by the distinct branches on the gene trees and variation at specific sites in the amino acid sequences. Is this sequence divergence random, or could it reflect divergent solutions for optimizing PEPCase function in C₄ leaves? In C₄ plants, the affinity of PEPCase for its substrate bicarbonate should be increased to compensate for subsaturating bicarbonate concentrations in M tissues (Gowik and Westhoff, 2011; Di Mario and Cousins, 2019). Also, C₄ orthologs of PEPCase require reduced sensitivity to malate and aspartate, because these allosteric inhibitors of PEPCase must accumulate to high concentration in M cells to drive rapid diffusion into the BS (Jacobs et al., 2008; Gowik and Westhoff, 2011). Numerous studies have documented similar changes in the amino acid sequence of PEPCase in a pattern that is associated with changes in sensitivity to malate, aspartate and PEP, supporting hypotheses of convergent optimization of PEPCase for the C₄ leaf (Engelmann et al., 2003; Gowik et al., 2006; Christin et al., 2007). Convergence is indicated by a widely noted substitution of a serine for alanine at a homologous site in PEPC1. The sequence of PEPCase in a pattern that is associated with changes in sensitivity to malate, aspartate and PEP, supporting hypotheses of convergent optimization of PEPCase for the C₄ leaf (Engelmann et al., 2003; Gowik et al., 2006; Christin et al., 2007). Convergence is indicated by a widely noted substitution of a serine for alanine at a homologous site in PEPC1, which is proposed to alter PEP and possibly bicarbonate sensitivity (Christin et al., 2007; Gowik and Westhoff, 2011; Di Mario and Cousins, 2019). Christin et al. (2007) also observed common sequence substitutions in numerous distinct clades of C₄ grasses, some of which occur in gene regions controlling sensitivity to malate. However, transcriptome surveys of numerous chenopods indicate the sequence convergence observed in grasses is less common in eudicots, suggesting divergent solutions to meeting the kinetic and regulatory requirements of a C₄ pathway.

### Table 6

| Convergent trait | Divergent trait |
|------------------|-----------------|
| **Biochemical level** | **Biochemical level** |
| - a) PEPCase upregulated in M cytosol only. | - a) Variation in decarboxylating enzymes and associated ultrastructural, transport and regulatory traits. |
| - b) Reduced malate sensitivity of PEPCase. | - b) PEP-CK species may not require PPDK |
| - c) Carbonic anhydrase expressed in M tissue only. | - c) Transport metabolites are malate or aspartate |
| - d) Functional rubisco restricted to high CO₂ compartment. | |
| **Structural level** | **Structural level** |
| - a) Interior compartment always used for CO₂ concentration. | - a) Cell tissue type for CO₂ concentration (e.g., BS, mestome sheath, inner chlorenchyma, central cluster of chloroplasts as in Bienertia) |
| - b) Reduced M/BS ratio. | - b) Variable M and BS arrangement around veins (e.g., partially or incomplete coverage of vein) |
| - c) High vein density. | - c) Variable plasmodesmata structure (branched, non-branched) |
| - d) More plasmodesmata at M x BS boundary. | |
| **Ultrastructural level** | **Ultrastructural level** |
| - a) More chloroplast volume in Kranz sheath tissue | - a) Chloroplast size, number, shape and position |
| - b) Less chloroplast investment in M cells | - b) Diffusive trap structure (cell wall thickness, suberin presence, vacuole size, chloroplast cluster arrangement) |
| - c) Reduced mitochondria in M cells | - c) Thylakoid stacking and PSII location |
| - d) Variable location of whole-chain versus cyclic electron transport | - d) Variable location of whole-chain versus cyclic electron transport |
| - e) Mitochondria number, size and position in BS | - e) Mitochondria number, size and position in BS |
| - f) Variable CO₂ concentration (e.g., BS, mestome sheath, inner chlorenchyma, central cluster of chloroplasts as in Bienertia) | - Note: most of the variable traits above become convergent within a given C₄ subtype. |

**BS:** bundle sheath; **M:** mesophyll; **PEP-CK:** PEP carboxykinase; **PEPCase:** PEP carboxylase; **PEP carboxykinase.**

**Note:** most of the variable traits above become convergent within a given C₄ subtype.
be able to accomplish the necessary change in sensitivities. A third possibility is the C₄ Nyctaginaceae species may simply have a non-optimized PEPCase for the C₄ function, in which case other mechanisms may compensate for inefficiencies. One compensation mechanism may be environmental, for example, high temperature in the natural habitat may override a need for convergence in PEPCase sequences by stimulating catalytic capacity. Where convergence is beneficial, but not accomplished, the consequences and compensation mechanisms become as interesting as the convergence itself.

We next consider the evolutionary transition at the opposite end of the C₄ pathway, where there is clear divergence in decarboxylase function. The divergence arises because there are enzymatic alternatives to meeting the decarboxylation imperative, with NADP-ME, NAD-ME, and PCK having significant roles in numerous metabolic pathways in C₃ plants (Aubry et al., 2011). Elevated activities of NADP-ME, NAD-ME, and PCK have been detected in vascular tissues and BS cells of C₃ species, where they metabolize organic acids used in long-distance transport from the roots, and may have a role in pH homeostasis, coordinating carbon and nitrogen metabolism, and providing metabolites for numerous biosynthetic pathways (Hibberd and Quick, 2002; Aubry et al., 2011). The mechanism by which one decarboxylase is selected over another is not known, but since all three are active in BS tissues, it is possible that each may function in a nascent C₄ pathway, perhaps to support recovery of photorespired nitrogen in the BS (Mallmann et al., 2014). It may be that chance determines which of the three decarboxylases is upregulated as the C₄ pathway strengthens, with convergence on the distinctive subtype characteristics occurring afterward, as the C₄ pathway is optimized for the selected decarboxylase. Alternatively, pre-existing traits in the C₃ or C₂ ancestors may determine which decarboxylase is selected and also influence the characteristics of each C₄ subtype.

To close, we note that comparative studies of evolutionary convergence and divergence in C₄ plants show that convergence is greatest where constraints are high and biochemical options are limited, while divergence occurs where the constraints are low and multiple alternative solutions can be selected during the evolutionary process. What remains unclear is the influence that chance, ancestry, and environment have over divergent possibilities. By demonstrating multiple sets of closely related clades differing in C₄ subtype, we have identified replicated examples with which to address these issues using comparative methods of evolutionary biology (Harvey and Pagel, 1991). C₄ photosynthesis can thus become a powerful tool to unravel the intricacies of convergence in complex trait evolution.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository.repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

RK led the imaging efforts and their quantification and ecological data analysis. MS conducted the phylotranscriptomic and gene sequence analyses. SA performed gas exchange and enzyme assays. TS and RS directed the labs where the work was conducted and supplied funding. RS wrote the manuscript with input from each co-author. All authors contributed to the article and approved the submitted version.

DEDICATION

This paper is dedicated to the memory of Dr. Udo Gowik (1971–2020), a fun-loving friend, helpful to all, a pioneer in C₄ plant biology and a fearsome political debater.

FUNDING

This research was supported by Canadian Natural Science and Engineering Research Council grants RGPIN-04878-2015 to TS and RGPIN-2017-06476 to RS. The contribution of SA was supported by a travel grant from the Institute of Global Innovation Research, Tokyo University of Agriculture and Technology.

ACKNOWLEDGMENTS

We thank Nina Esmail and Perlina Lim for assistance in plant growth and the fixation and embedding of leaves for microscopic imaging. We also thank Gerry Edwards and Elena Vosnesenskaya for the gift of Portulaca pilosa seeds.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.578739/full#supplementary-material

REFERENCES

Aberer, A. J., Kobert, K., and Stamatakis, A. (2014). ExaBayes: massively Bayesian tree inference for the whole-genome era. Mol. Biol. Evol. 31, 2553–2556. doi: 10.1093/molbev/msu236

Alonso-Cantarabria, H., and von Caemmerer, S. (2016). Carbon isotope discrimination as a diagnostic tool for C₄ photosynthesis in C₃-C₄ intermediate species. J. Exp. Bot. 67, 3109–3121. doi: 10.1093/jxb/erv555

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

AUTHOR CONTRIBUTIONS

RK led the imaging efforts and their quantification and ecological data analysis. MS conducted the phylotranscriptomic and gene sequence analyses. SA performed gas exchange and enzyme assays. TS and RS directed the labs where the work was conducted and supplied funding. RS wrote the manuscript with input from each co-author. All authors contributed to the article and approved the submitted version.

DEDICATION

This paper is dedicated to the memory of Dr. Udo Gowik (1971–2020), a fun-loving friend, helpful to all, a pioneer in C₄ plant biology and a fearsome political debater.

FUNDING

This research was supported by Canadian Natural Science and Engineering Research Council grants RGPIN-04878-2015 to TS and RGPIN-2017-06476 to RS. The contribution of SA was supported by a travel grant from the Institute of Global Innovation Research, Tokyo University of Agriculture and Technology.

ACKNOWLEDGMENTS

We thank Nina Esmail and Perlina Lim for assistance in plant growth and the fixation and embedding of leaves for microscopic imaging. We also thank Gerry Edwards and Elena Vosnesenskaya for the gift of Portulaca pilosa seeds.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2020.578739/full#supplementary-material

Arnon, D. I. (1949). Copper enzymes in isolated chloroplasts: polyphenoloxidase in Beta vulgaris. Plant Physiol. 24, 1–15. doi: 10.1104/pp.24.1.1

Ashton, A. R., Burnell, J. N., Furbank, R. T., Jenkins, C. L. D., and Hatch, M. D. (1990). “Enzymes of C₄ photosynthesis,” in Methods in Plant Biochemistry, Vol. 7A, ed. P. J. Lea, (London: Academic Press), 39–72. doi: 10.1016/B978-0-12-461013-2.50010-1

Aubry, S., Brown, N. J., and Hibberd, J. M. (2011). The role of proteins in C₄ plant photosynthesis in C₄ plants prior to their recruitment into the C₄ pathway. J. Exp. Bot. 62, 3049–3059. doi: 10.1093/jxb/err012

Alonso-Cantarabria, H., and von Caemmerer, S. (2016). Carbon isotope discrimination as a diagnostic tool for C₄ photosynthesis in C₃-C₄ intermediate species. J. Exp. Bot. 67, 3109–3121. doi: 10.1093/jxb/erv555

Frontiers in Plant Science | www.frontiersin.org 18 November 2020 | Volume 11 | Article 578739

Frontiers in Plant Science | www.frontiersin.org 18 November 2020 | Volume 11 | Article 578739

Khoshravesh et al.
Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear mixed-effects models using lme4. *J. Stat. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01

Bisninger, K., Khoshavesh, R., Kotrade, J. P., Oakley, J., Sage, T. L., Sage, R. F., et al. (2014). *Giselia* (Gisieaceae): phylogenetic relationships, biogeography, and eco physiology of a poorly known C₄ lineage in the Caryophyllales. *Am. J. Bot.* 101, 949–959. doi: 10.3732/ajb.1300279

Bivand, R., and Lewin-Koh, N. (2013). maptools: Tools for Reading and Handling Spatial Objects. *R Package Version 0.8–27*. Available online at: https://cran.r-project.org/web/packages/maptools/index.html (accessed August 24, 2020).

Blount, Z. D., Lenski, R. E., and Losos, J. B. (2018). Contingency and determinism in evolution: replaying life’s tape. *Science* 362:eaam5979. doi: 10.1126/science.aam5979

Bohley, K., Joos, O., Hartmann, H., Sage, R. F., Liede-Schumann, S., and Kadereit, G. (2015). phylogeny of sesuvioideae (Aizoaceae)-biogeography, leaf anatomy and the evolution of C₄Photosynthesis. *Perspect. Plant. Ecol. 17*, 116–130.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Bräutigam, A., and Gowik, U. (2016). Photosorption connects C₃ and C₄ photosynthesis. *J. Exp. Bot.* 67, 2953–2962.

Bräutigam, A., and Weber, A. P. M. (2011). “Transport processes: connecting the reactions of C₄ photosynthesis,” in *C₄ Photosynthesis and Related CO₂ Concentrating Mechanisms*, eds A. S. Raghavendra and R. F. Sage (Dordrecht: Springer Netherlands), 277–300.

Capella-Gutierrez, S., Silla-Martinez, J. M., and Gabaldon, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Carolin, R., Jacobs, S., and Veski, M. (1978). Kranz cells and mesophyll in the Chenopodales. *Aust. J. Bot.* 26, 6683–6693. doi: 10.1071/BT97900683

Christin, P.-A., Besnard, G., Samaritani, E., Duvall, M. R., Hodkinson, T. R., Savolainen, V., et al. (2008). Oligocene CO₂ decline promoted C₄ photosynthesis in grasses. *Curr. Biol.* 18, 37–43. doi: 10.1016/j.cub.2007.11.058

Christin, P.-A., Boxall, S. F., Gregory, R., Edwards, E. J., Hartwell, J., and Osborne, C. P. (2013a). Parallel recruitment of multiple genes into C₄ photosynthesis. *Genome Biol. Evol.* 5, 2174–2187. doi: 10.1093/gbe/evt168

Christin, P.-A., Edwards, E. J., Besnard, G., Boxall, S. F., Gregory, R., Kellogg, E. A., et al. (2012a). Adaptive evolution of C₄ photosynthesis through recurrent lateral gene transfer. *Curr. Biol.* 22, 445–449. doi: 10.1016/j.cub.2012.01.054

Christin, P.-A., Freeman, R. T., and Osborne, C. P. (2010a). Can phylogenetics identify C₃ origins and reversals? *Trends Ecol. Evol.* 25, 403–409. doi: 10.1016/j.tree.2010.04.007

Christin, P.-A., and Osborne, C. P. (2013). The recurrent assembly of C₄ photosynthesis, an evolutionary tale. *Photosynth. Res.* 117, 163–175. doi: 10.1007/s11120-013-9852-z

Christin, P.-A., Osborne, C. P., Chatelé, D. S., Columbus, J. T., Besnard, G., Hodkinson, T. R., et al. (2013b). Anatomical enablers and the evolution of C₄ photosynthesis in grasses. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1381–1386. doi: 10.1073/pnas.1216777110

Christin, P. A., Osborne, C. P., Sage, R. F., Arakaki, M., and Edwards, E. J. (2011). C₄ eudicots are not younger than C₃ monocots. *J. Exp. Bot.* 62, 3171–3181. doi: 10.1093/jxb/err041

Christin, P.-A., Salamin, N., Savolainen, V., Duvall, M. R., and Besnard, G. (2007). C₄ photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr. Biol.* 17, 1241–1247. doi: 10.1016/j.cub.2007.06.036

Christin, P.-A., Samaritani, E., Petitpierre, B., Salamin, N., and Besnard, G. (2009). Evolutionary insights on C₄ photosynthetic subtypes in grasses from genomes and phylogenetics. *Genome Biol. Evol.* 1, 221–230. doi: 10.1093/gbe/evp020

Christin, P.-A., Wallace, M. J., Clayton, H., Edwards, E. J., Furbank, R. T., Hattersley, P. W., et al. (2012b). Multiple photosynthetic transitions, polyploidy, and lateral gene transfer in the grass subtribe Neurachninae. *J. Exp. Bot.* 63, 6297–6308. doi: 10.1093/jxb/ers282

Christin, P.-A., Weinreich, D. M., and Besnard, G. (2010b). Causes and evolutionary significance of genetic convergence. *Trends Genet.* 26, 400–405. doi: 10.1016/j.tig.2010.06.005

Conway-Morris, S. (2003). *Life’s Solutions: Inevitable Humans in a Lonely Universe*. Cambridge: Cambridge University Press.
Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things).” Methods Ecol. Evol. 3, 217–223. doi: 10.1111/j.2041-210X.2011.01069

Reyna-Llorens, I., and Hibberd, J. M. (2017). Recruitment of pre-existing networks during the evolution of C4 photosynthesis. Phil. Trans. R. Soc. B 372: 20160386. doi: 10.1098/rstb.2016.0386

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. Trends Genet. 16, 276–277.

Rosnov, J. J., Edwards, G. E., and Roalson, E. H. (2014). Positive selection of Kranz and non-Kranz C4 phosphoenolpyruvate carboxylase amino acids in Suaedoeidae (Chenopodiaceae). J. Exp. Bot. 65, 3595–3607. doi: 10.1093/jxb/eru053

Rosnov, J. J., Evans, M. A., Kapralov, M. V., Cousins, A. B., Edwards, G. E., and Roalson, E. H. (2015). Kranz and single-cell forms of C4 plants in the subfamily Suaedoeidae show kinetic C4 convergence for PEPC and Rubisco with divergent amino acid substitutions. J. Exp. Bot. 66, 7347–7358. doi: 10.1093/jxb/eru431

Sage, R. F. (2004). The evolution of C4 photosynthesis. New Phytol. 161, 341–370. doi: 10.1111/j.1469-8137.2004.00974

Sage, R. F. (2016). A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. J. Exp. Bot. 67, 4039–4056. doi: 10.1093/jxb/erw156

Sage, R. F., Christin, P.-A., and Edwards, E. J. (2011). The C4 plant lineages of planet Earth. J. Exp. Bot. 62, 3155–3169. doi: 10.1093/jxb/erq048

Sage, R. F., Khoshravesh, R., and Sage, T. L. (2014). From proto-Kranz to C4 Kranz: building the bridge to C4 photosynthesis. J. Exp. Bot. 65, 3341–3356. doi: 10.1093/jxb/eru180

Sage, R. F., Monson, R. K., Ehleringer, J. R., Adachi, S., and Pacey, R. W. (2018). Some like it hot: the physiological ecology of C4 plant evolution. Oecologia 187, 941–966. doi: 10.1007/s00442-018-4191-6

Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photospiration and the evolution of C4 photosynthesis. Annu. Rev. Plant Biol. 63, 19–47. doi: 10.1146/annurev-plant-042811-105511

Sage, T. L., Busch, F. A., Johnson, D. C., Friesen, P. C., Stinson, C. R., Stata, M., et al. (2013). Initial events during the evolution of C4 photosynthesis in C3 species of Flaveria. Plant Physiol. 163, 1266–1276. doi: 10.1104/pp.113.221119

Sage, T. L., Sage, R. F., Voznesenskaya, E. V., Oakley, J. C., Rahman, B., Johnson, D. C., et al. (2014). Mesophyll cells of C4 plants have fewer chloroplasts than those of closely related C3 plants: C4 versus C3 mesophyll chloroplasts. Plant Cell Environ. 37, 2522–2535. doi: 10.1111/pce.12334

Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012). NIH Image to Image: 25 years of image analysis. Nat. Methods 9, 671–675. doi: 10.1038/nmeth.2089

Schulze, S., Mallmann, J., Burscheidt, J., Kocacinar, F., Strebrel, M., Bauwe, H., et al. (2013). Evolution of C4 photosynthesis in the genus Flaveria: establishment of a photorespiratory CO2 pump. Plant Cell 25, 2522–2535. doi: 10.1105/tpc.113.114520

Spellenberg, R. W. (2003). Nyctaginaceae. Flora Nort Am. 4, 14–74.

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

Stata, M., Sage, T. L., Rennie, T. D., Khoshravesh, R., Sultmanis, S., Khaikin, Y., et al. (2014). Mesophyll cells of C4 plants have fewer chloroplasts than those of closely related C3 plants: C4 versus C3 mesophyll chloroplasts. Plant Cell Environ. 37, 2587–2600. doi: 10.1111/pce.12331

Suyama, M., Torres, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34, W609–W612. doi: 10.1093/nar/gkl315

Taub, D. R. (2000). Climate and the U.S. distribution of C4 grass subfamilies and decarboxylation variants of C4 photosynthesis. Am. J. Bot. 87, 1211–1215. doi: 10.2307/2656659

Thimm, O., Blasing, O., Gibon, Y., Nagel, A., Meyer, S., Krüger, P., et al. (2004). MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. Plant J. 37, 914–939. doi: 10.1111/j.1365-313X.2004.02016.x

Tropicos (2020). Tropicos.org Missouri Botanical Garden http://www.tropicos.org (accessed October 20, 2020).