



Independent Recruitment of Duplicated β -Subunit-Coding NAD-ME Genes Aided the Evolution of C₄ Photosynthesis in Cleomaceae

Marcos A. Tronconi^{1†}, Meike Hüdig^{2†}, M. Eric Schranz³ and Veronica G. Maurino^{2*}

¹ Centro de Estudios Fotosintéticos y Bioquímicos (CEFOTBI-CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina, ² Abteilung Molekulare Pflanzenphysiologie, Institut für Molekulare Physiologie und Biotechnologie der Pflanzen, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany, ³ Biosystematics Group, Wageningen University, Wageningen, Netherlands

OPEN ACCESS

Edited by:

Martha Ludwig,
University of Western Australia,
Australia

Reviewed by:

Gerald E. Edwards,
Washington State University,
United States
Sylvain Aubry,
University of Zurich, Switzerland

*Correspondence:

Veronica G. Maurino
vero.maurino@uni-bonn.de

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

Received: 15 June 2020

Accepted: 14 September 2020

Published: 06 October 2020

Citation:

Tronconi MA, Hüdig M,
Schranz ME and Maurino VG (2020)
Independent Recruitment
of Duplicated β -Subunit-Coding
NAD-ME Genes Aided the Evolution
of C₄ Photosynthesis in Cleomaceae.
Front. Plant Sci. 11:572080.
doi: 10.3389/fpls.2020.572080

In different lineages of C₄ plants, the release of CO₂ by decarboxylation of a C₄ acid near rubisco is catalyzed by NADP-malic enzyme (ME) or NAD-ME, and the facultative use of phosphoenolpyruvate carboxykinase. The co-option of gene lineages during the evolution of C₄-NAD-ME has been thoroughly investigated, whereas that of C₄-NAD-ME has received less attention. In this work, we aimed at elucidating the mechanism of recruitment of NAD-ME for its function in the C₄ pathway by focusing on the eudicot family Cleomaceae. We identified a duplication of NAD-ME in vascular plants that generated the two paralogs lineages: α - and β -NAD-ME. Both gene lineages were retained across seed plants, and their fixation was likely driven by a degenerative process of sub-functionalization, which resulted in a NAD-ME operating primarily as a heteromer of α - and β -subunits. We found most angiosperm genomes maintain a 1:1 β -NAD-ME/ α -NAD-ME (β/α) relative gene dosage, but with some notable exceptions mainly due to additional duplications of β -NAD-ME subunits. For example, a significantly high proportion of species with C₄-NAD-ME-type photosynthesis have a non-1:1 ratio of β/α . In the Brassicales, we found C₄ species with a 2:1 ratio due to a β -NAD-ME duplication ($\beta1$ and $\beta2$); this was also observed in the C₃ *Tarenaya hassleriana* and Brassica crops. In the independently evolved C₄ species, *Gynandropsis gynandra* and *Cleome angustifolia*, all three genes were affected by C₄ evolution with α - and $\beta1$ -NAD-ME driven by adaptive selection. In particular, the $\beta1$ -NAD-MEs possess many differentially substituted amino acids compared with other species and the $\beta2$ -NAD-MEs of the same species. Five of these amino acids are identically substituted in $\beta1$ -NAD-ME of *G. gynandra* and *C. angustifolia*, two of them were identified as positively selected. Using synteny analysis, we established that β -NAD-ME duplications were derived from ancient polyploidy events and that α -NAD-ME is in a unique syntenic context in both Cleomaceae and Brassicaceae. We discuss our hypotheses for the evolution of NAD-ME and its recruitment for C₄ photosynthesis. We propose that gene duplications

provided the basis for the recruitment of NAD-ME in C₄ Cleomaceae and that all members of the *NAD-ME* gene family have been adapted to fit the C₄-biochemistry. Also, one of the β -*NAD-ME* gene copies was independently co-opted for its function in the C₄ pathway.

Keywords: C₄-photosynthesis, C₄-evolution, Cleomaceae, gene duplication, NAD-malic enzyme, subfunctionalization, neofunctionalization

INTRODUCTION

C₃ photosynthesis (Bascham et al., 1954) relies exclusively on ribulose-1,5-bisphosphate carboxylase oxygenase (rubisco) for carboxylase activity and evolved early in the history of life (Hayes, 1994). Rubisco is a bifunctional enzyme that catalyzes both the carboxylation and the oxygenation of its substrate ribulose-1,5-bisphosphate. One of the products of the oxygenase activity, 2-phosphoglycolate, is a toxic metabolite (Anderson, 1971; Kelly and Latzko, 1976; González-Moro et al., 1997). As a selective response to rubisco's promiscuity, plants evolved the energetically costly photorespiratory pathway (Maurino and Peterhansel, 2010). Until ~400 million years ago (Mya), the rubisco oxygenase reaction was negligible due to elevated CO₂ and low O₂ levels in the atmosphere (Sage and Monson, 1999; Leakey and Lau, 2012). After this time, the onset of oxygenic photosynthesis introduced changes in atmospheric conditions such as high levels of O₂ which led to significant levels of costly photorespiration. Some land plants evolved a carbon concentrating mechanism, known as the C₄ photosynthetic pathway, which resulted in reduction of the high-levels of photorespiration (Hatch, 1971; Heckmann et al., 2013). Nearly all the independent C₃ to C₄ transitions are dated to the mid-Oligocene (25–30 Mya), a time that was preceded by a massive depletion of atmospheric CO₂ (Christin et al., 2008; Vicentini et al., 2008; Edwards et al., 2010). Under a wide range of environmental conditions, such as high temperatures, dryness, and high light intensities, plants possessing the C₄ biochemical pump are more efficient in terms of water and nitrogen use (Furbank and Hatch, 1987).

The initial step of the C₄ photosynthetic pathway is the fixation of inorganic carbon onto phosphoenolpyruvate (PEP) by PEP-carboxylase (PEPCase) to produce a four-carbon (C₄) acid (Hatch, 1987; Kanai and Edwards, 1999). The C₄ acid moves to the site of rubisco where specific decarboxylases release CO₂ (Drincovich et al., 2011). The decarboxylation reaction also produces a three-carbon acid, which diffuses back to the site of PEPCase where it is recycled to PEP. The C₄ cycle effectively acts as a CO₂ pump, increasing CO₂ levels around rubisco such that it nearly saturates the active site and thus reduces photorespiration to a minimal level (von Caemmerer, 2000).

Two major C₄ photosynthetic metabolic routes, known as the NAD-malic enzyme (ME) and NADP-ME subtypes, are distinguished by the primary ME decarboxylase used (Maier et al., 2011). In grasses, either of these subtypes can also make use of the facultative activity of PEPCase (Wang et al., 2014). C₄ photosynthesis is a complex convergent trait that arose independently in at least 66 plant lineages (Sage et al., 2011, 2012), the majority of which use the NADP-ME subtype

(Sage et al., 2011). The NAD-ME subtype is mostly found in eudicot species, where it is found in approximately 20 C₄-lineages (Sage et al., 2011).

In plants using the NADP-ME subtype of C₄ metabolism, C₄-NADP-ME is present as a unique plastidial isoform with tissue specific expression and special regulatory properties, such as having pH dependent changes in oligomerization and substrate inhibition (Detarsio et al., 2007; Alvarez et al., 2019). The underlying molecular determinants for the distinction of C₄-NADP-ME from the non-photosynthetic isoform (nonC₄-NADP-ME) were recently characterized (Alvarez et al., 2019). The nonC₄-NADP-ME is present as multiple isoforms in higher plants, which are located to plastids and cytosol and show tissue specific expression (Saigo et al., 2004; Gerrard Wheeler et al., 2005; Detarsio et al., 2008; Gerrard Wheeler et al., 2008; Maurino et al., 2009).

In contrast to the well-described C₄-NADP-ME, C₄-NAD-ME has yet to be characterized at the molecular level. Early studies of *Amaranthus hypochondriacus* and various monocot species showed contradictory data for subunit composition and oligomeric states (Murata et al., 1989; Long et al., 1994). In plants, NAD-ME is exclusively present in mitochondria, where its core function is in L-malate respiration, as an associated enzyme of the tricarboxylic acid cycle (Grover et al., 1981; Artus and Edwards, 1985; Tronconi et al., 2008; Fuchs et al., 2020). In *Arabidopsis thaliana*, NAD-ME functions as a homo- and/or heterodimer of two distinct, homologs proteins (~65% sequence identity): known as the α -subunits (to which AtNAD-ME1 belongs) and β -subunits (to which AtNAD-ME2 belong) and with molecular masses in the range of 58 and 63 kDa (Tronconi et al., 2008, 2010a). The α - and β -NAD-ME share only 40% of identity with the NADP-ME isoforms, owing to the fact that the *NAD-ME* and *NADP-ME* genes were acquired in independent evolutionary processes in plants (Tronconi et al., 2018).

Modifications in function and expression are key components of the evolutionary transition of several enzymes involved in the C₄ carbon concentrating mechanism. Gene duplication was proposed as a precondition for the evolution of C₄ activities, as it facilitates C₄-specific adaptive changes in *cis*-regulatory control regions as well as in coding regions (Lynch and Conery, 2000; Moore and Purugganan, 2005). The evolution of C₄-NADP-ME likely followed this process of gene duplication and neofunctionalization, starting from a plastidial non-C₄ isoform-coding gene and including the acquisition of a bundle sheath cell-specific expression pattern (Maurino et al., 2001; Tausta et al., 2002; Saigo et al., 2004; Christin et al., 2009). The C₄ specific isoforms of NADP-ME evolved at least five times independently in this way in grasses (Christin et al., 2009). In contrast, the

evolutionary history of C₄-NAD-ME remains elusive, as to date a gene coding for a C₄-specific isoform has not been identified (Murata et al., 1989; Long et al., 1994).

Here, we aimed at elucidating the mechanism of recruitment of NAD-ME for its function in the C₄ pathway by focusing on the eudicot family Cleomaceae. This family contains species spanning a developmental progression from C₃ to C₄ photosynthesis and at least three separate origins of C₄ lineages (Feodorova et al., 2010; Koteyeva et al., 2011). Cleomaceae belongs to the order Brassicales and is a sister group of the Brassicaceae, which contains one of the best studied plant species, *Arabidopsis thaliana*, for which vast amounts of -omics data are available for comparative analyses. Cleomaceae and Brassicaceae share the At-beta whole-genome duplication (WGD) event, which was estimated to have occurred 75–100 Mya (Edger et al., 2015). The Cleomaceae and Brassicaceae lineages diverged 41 Mya and more recently underwent the independent At-alpha (23–34 Mya) and Cs-alpha (14–20 Mya) WGDs (Schranz and Mitchell-Olds, 2006; Barker et al., 2009). Despite different patterns of gene loss and retention and chromosomal rearrangements after polyploidy, the genomes of Cleomaceae and Brassicaceae species show detectable synteny (Schranz and Mitchell-Olds, 2006). Moreover, there exist no significant differences in gene copy numbers between C₃ and C₄ *Cleome* species (van den Bergh et al., 2014).

Our comprehensive analyses indicate that a duplication of the *NAD-ME* gene during the evolution of vascular plants resulted in two paralogs lineages, α - and β -*NAD-ME*, which were retained during seed plant evolution and diversification. We propose that the heteromeric assembly of NAD-ME was established by sub-functionalization of the duplicated NAD-ME genes. Later, neo-functionalization optimized the α - and β -NAD-ME functions and changes in the subunit-specific duplications provided the basis for the recruitment of NAD-ME in C₄ biochemistry. We found that in Cleomaceae all *NAD-ME* genes were affected by C₄ evolution, where one of the β -*NAD-ME* gene copies was co-opted for its function in the C₄ pathway.

MATERIALS AND METHODS

Sequence Retrieval

For species with entire genome information, NAD-ME coding sequences were extracted from primary gene models www.phytozome.net. Sequences from Cleomaceae species were acquired from transcriptome data (Kulahoglu et al., 2014; Mabry et al., 2019); in case of *Gynandropsis gynandra* (C₄), *Cleome angustifolia* (C₄), and *Tarenaya hassleriana* (C₃) the sequences were verified and correctly assembled using cDNA-based sequencing as the transcriptomes showed misassembled transcripts for several NAD-ME genes. Sequences for *Chara braunii*, *Azolla filiculoides*, *Salvinia cucullata*, and *Panicum miliaceum* were identified from their respective genome publications (Li et al., 2018; Nishiyama et al., 2018; Zou et al., 2019). Additional fern sequences were extracted from available large-scale transcriptomic data (Shen et al., 2018). *Ginkgo biloba*, *Amborella trichopoda*, *Taxus baccata*,

Pinus pinaster, *Pinus sylvestris*, and *Pseudotsuga menziesii* sequences were collected via PLAZA 3.0 (Proost et al., 2015). Accession numbers of all NAD-ME coding sequences used in this work are listed in **Supplementary Table 1**. BLASTP with the BLOSUM62 as default scoring matrix and a minimal e-value of 0.0001 was implemented to obtain homologs using AtNAD-ME1 (AT2G13560) and AtNAD-ME2 (AT4G00570) as query. All sequences were manually checked for correct translation start sites and the presence of conserved amino acid regions found in all NAD(P)-ME (Tronconi et al., 2018). Mitochondrial localization was verified using the program Target P (Emanuelsson et al., 2000).

Multiple Sequence Alignments

A data set of the coding sequences was assembled using MEGA X (v.10.0.5) (Kumar et al., 2018). The sequences were then translated into amino acids and aligned using Muscle (Edgar, 2004) with the gap opening penalty value of -2.9 and without penalizing its extension. Once retranslated into nucleotides, the alignment was manually edited to select the most-reliable positions in the alignment, assisted by Gblocks¹ and TrimAl² programs. Since the different phylogenetic methods consider columns with gaps in different ways, we applied a stringent criterion by eliminating codons with coverage less than 95%. The final multiple sequence alignment (MSA) consisted of 240 coding sequences from 118 species with 1,731 nucleotides positions corresponding to 577 codons.

Phylogenetic Analyses

Bayesian inference (BI) was performed using MrBayes 3.1.2 software (Ronquist and Huelsenbeck, 2003). Two parallel runs, each including four Metropolis-coupled Markov chain Monte Carlo (MC3) analyses, were run for 5,000,000 generations and sampled every 1,000 generations. This generated an output of 5,000 trees per run. A site-specific rate model (partition scheme) was used. The characters in the MSA were divided into three sets corresponding to the codon positions. Each position has its own rate labeled m1 in case of the first codons site, m2 in case of the second codons site and m3 in case of the third codons site. For the tree inferred from the third positions we defined a one-partition scheme by excluding the characters represented by the first and second sites in the MSA. For an efficient Metropolis coupling, an incremental heating scheme of three heated chains and one cold chain in each run was used with a temperature parameter setting of 0.1. The final average standard deviation of split frequencies was used as the convergence index (values <0.01 indicated good convergence). The convergence of clade posterior probabilities within and between runs was checked using the potential scale reduction factor. The initial 25% of the sampled trees for each MC3 run were discarded as “burn-in” and the post-burn-in trees from the two runs were integrated to generate a 50% majority-rule consensus tree. The percentage of samples recovering any particular clade in a BI analysis represents the posterior probability (BPP) of a

¹<http://molevol.cmima.csic.es/castresana/Gblocks.html>

²<http://trimal.cgenomics.org/>

clade. In all cases, a GTR (General Time Reversible) model with base frequencies gamma shape parameter (G) and proportion of invariants sites (I) was set. All the active parameters in the GTR + G + I model were optimized separately for each position of the codons. For analyses using third codon positions, the number of chains in each run of was increased from four to five due to convergence conflicts. Nodes with BPP values >90% were considered highly supported.

Maximum likelihood (ML) and neighbor joining (NJ) analyses on the whole data set were conducted using MEGA X (v.10.0.5). The goodness of fit of each model to the data was measured by the Bayesian information criterion (BIC) and the model with the lowest BIC score was considered the best description for a specific substitution pattern. The initial tree for the ML search was generated automatically by applying the NJ and BIONJ algorithms, and its branch lengths were adjusted to maximize the likelihood of the data set for that tree topology under the selected model of evolution. Heuristic searches were conducted with the initial tree based on the nearest neighbor interchange (NNI) search where the alternative trees differ in one branching pattern. Reliability of interior branches was assessed with 2,000 bootstrap (B) re-samplings. Nodes with MLB or NJB values 50–69% were regarded as weakly supported, 70–84% as moderately supported, and 85–100% as strongly supported (Hillis and Bull, 1993). The tree files were saved in Newick format (.nwk) containing all the relevant clade support values and branch length information. The trees were displayed using the FigTree v1.4.4 software and edited by rotating nodes and compressing lineages that were designated by their subdivision, class, order, or family names.

Differential Substitution Analysis

A strictly differentially substituted position is one for which the NAD-ME sequences of the C₄ species, *G. gynandra* and *Cleome angustifolia* (recently reclassified as *Coalisina angustifolia*), contained an identical amino acid, while a second, different amino acid was shared in all other NAD-ME sequences. For the differential substitution analysis, we used the whole data set of NAD-MEs of the Brassicales obtained in this work. MSAs of α - and β -NAD-MEs were computed using the MAFFT algorithm (v7.427) with the iterative refinement method L-INS-i (Katoh et al., 2005; Kuraku et al., 2013). The sequences were aligned using the online tool integrated MKT³. The best amino acid substitution model based on each MSA was estimated using MEGA X (v.10.0.5) (Kumar et al., 2018). We used the MSAs to identify amino acid positions that are strictly differentially conserved in the α -NAD-ME and β -NAD-ME sequences of the C₄ species as previously performed by Alvarez et al. (2019).

Synteny Analysis

To further investigate the evolutionary relationships and duplication of Brassicaceae and Cleomaceae β -NAD-ME and α -NAD-ME genes, syntenic analysis was performed with the SynFind tool using the default parameters (last algorithm, window size set to 40 genes and with a minimum number of

collinear anchors of 4) (Tang et al., 2015). For this analysis, a limited set of representative taxa was used based on phylogenetic breadth (see **Figure 3**), presence of ancient polyploidy (i.e., *Brassica* and *Gynandropsis*) and the availability of high-quality annotated draft genomes. For Cleomaceae, the genomes of *T. hassleriana*, *G. gynandra*, and *Cleome violacea* were included. For Brassicaceae we included *Arabidopsis thaliana*, *Arabis alpina*, *Brassica rapa*, and *Eutrema salsugineum*. As an outgroup representative, we included *Citrus clementina*. For our SynFind analyses, we used either *G. gynandra*, *A. thaliana* or *C. violacea* orthologs of β -NAD-ME and α -NAD-ME genes. The results of our synteny analyses were compared and visualized using GEvo (Tang et al., 2015).

Positive Selection Tests and Statistical Analysis

For testing sites that underwent positive selection during the evolution of C₄ species in the α -NAD-ME and β -NAD-ME coding sequences, different site-class specific models were employed using the software codeml, implemented in the PAML package (Yang, 2007). Because no gene lineages leading to C₄-specific NAD-MEs were identified, we conducted a site-class-specific approach. The models assume that the $\omega = dN/dS$ (non-synonymous to synonymous substitution rates) take a value of 1 under neutral evolution. Positive and purifying (negative) selection are indicated when $\omega > 1$ and $\omega < 1$, respectively. The codon substitution models were: M0 (one-ratio) M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), M8 (beta and $\omega > 1$), and M8a (beta and $\omega = 1$) (Yang et al., 2000). The fit of these models to the sequence data was compared using likelihood-ratio test (LRT). When an LRT test yielded a significant result for any of the pairwise comparisons, the Bayes empirical Bayes (BEB) method was used to identify amino acids residues that have evolved under selection. Posterior probability of 0.90 was selected as the standard threshold for identifying residues under selection (Scheffler and Seoighe, 2005).

Significance of deviations from 1:1 β -NAD-ME/ α -NAD-ME was determined according to the Fisher's exact test using the SigmaStat software (Systat Software, Inc.).

RESULTS

Identification of Two Major Clades of NAD-ME Genes (α - and β) in all Seed Plants

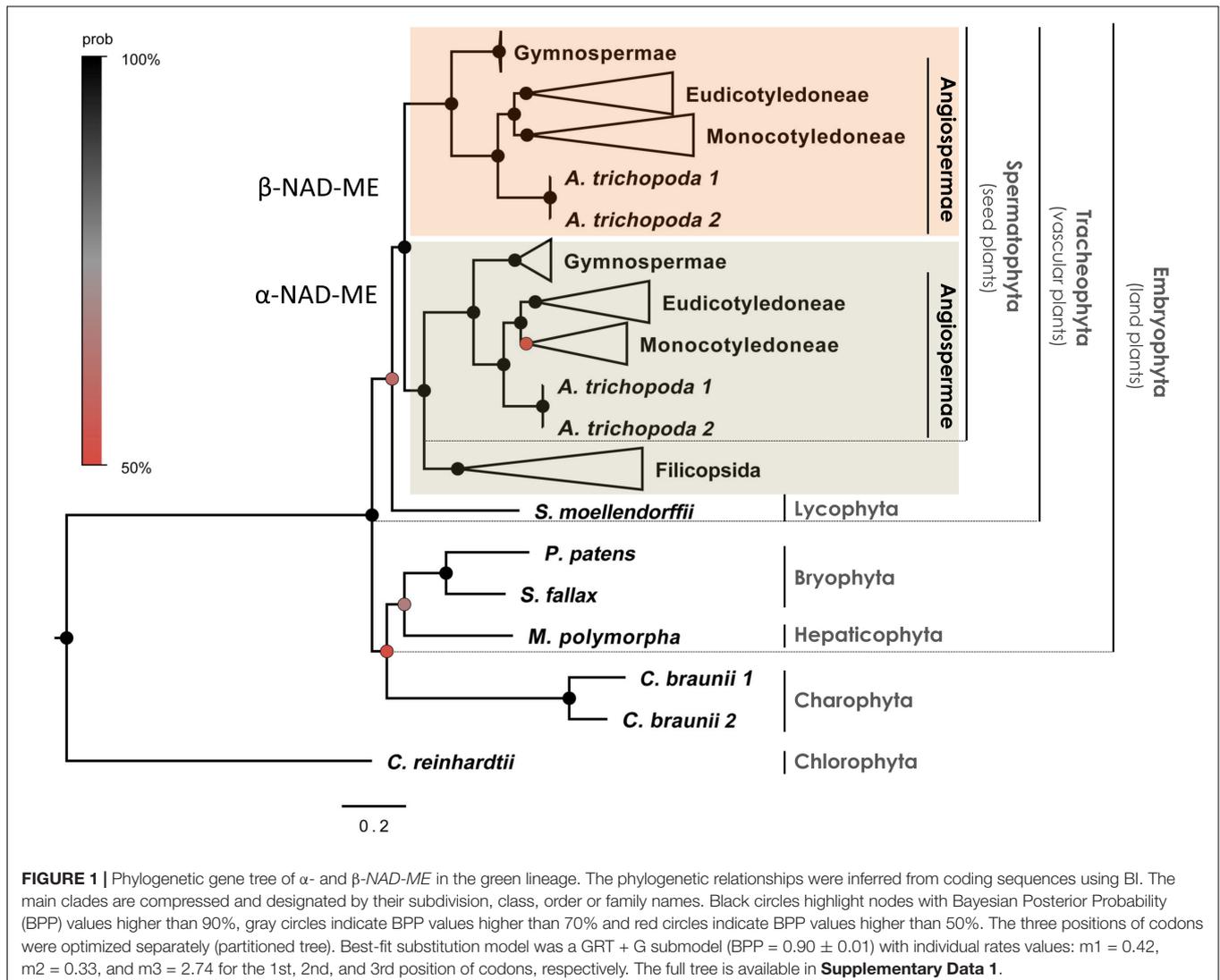
To cover the early and late evolution of NAD-ME proteins of land plants (Embryophyta), we analyzed a dataset of 253 coding sequences (**Supplementary Table 1**) from genomes and assembled transcripts from liverworts (Hepaticophyta), mosses (Bryophyta), early branching vascular plants (Lycophyta), ferns (Filicopsida), gymnosperms, angiosperms, and streptophyte algae (Charophytes), with chlorophyte algae (Chlorophyta) as the out-group. The coding sequences were aligned and used for BI and ML analyses to infer the phylogenetic gene and protein trees.

³<https://imkt.uab.cat>

We found all phylogenetic trees to be globally congruent regardless of the phylogenetic approach used (coding sequences-based BI tree in **Figure 1** and **Supplementary Data 1**, coding sequences-based ML tree in **Supplementary Data 2**, and protein-based ML tree in **Supplementary Data 3**). The α - and β -*NAD-ME* genes formed well supported orthologs clades within the seed plants (spermatophytes). Both paralogs genes were retained across all gymnosperms and angiosperms analyzed and orthologs of α - and β -*NAD-ME* were not found in all other non-seed plants. In all the trees we found fern homologs confidently placed as a sister group to the α -*NAD-ME* of seed plants with Bayesian posterior probability (BPP) = 100% (**Figure 1** and **Supplementary Data 1**) and MLB = 92 and 86% (**Supplementary Data 2** and **Supplementary Data 3**, respectively). Similarly, we found no evidence for α - and β -*NAD-ME* duplicates in streptophyte algae, liverwort, mosses, nor the early branching tracheophyte *S. moellendorffii*. Given these data, the duplication of the *NAD-ME* gene and the fixation of both paralogs occurred at the foundation of the seed plants.

Deviations From 1:1 β -*NAD-ME*/ α -*NAD-ME* (β/α) Among *NAD-ME* of C_3 and C_4 -Species

All 92 angiosperm species analyzed have retained orthologs of both *NAD-ME* α - and β coding genes and with most having a 1:1 β -*NAD-ME*/ α -*NAD-ME* (β/α) relative gene dosage. However, changes in the relative genetic dosage could be identified (**Figure 2**). Twenty species have at least two β copies for each α gene, and only one species, *Amaranthus hypochondriacus* (Carophylles), has two α genes for one β (**Figure 2**). The most pronounced change was found in *Glycine max* with a β/α gene ratio of 4:1. In eudicots, the NFC (Nitrogen Fixing Clade including: Rosales, Fabales, Cucurbitales, and Fagales) and COM (Celastrales, Oxalideles, and Malpighiales) clades contain 42% of the species with a non-1:1 ratio. Species having C_4 -*NAD-ME* photosynthesis, with the exception of *Panicum halli* and *Panicum miliaceum* (Poaceae), have a β/α relation deviating from 1:1. This proportion (71.5%) is significantly higher than that observed for



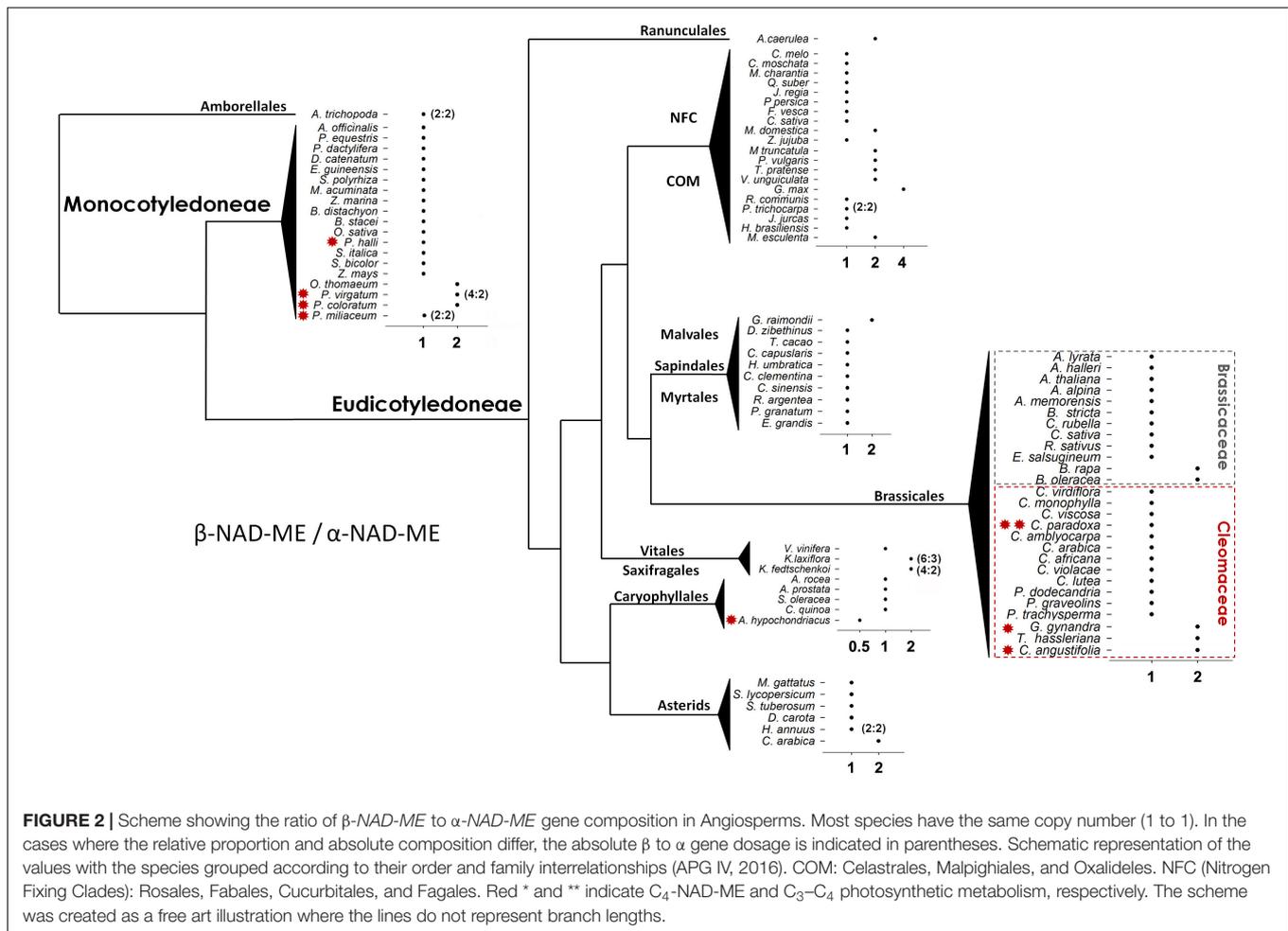


FIGURE 2 | Scheme showing the ratio of β -NAD-ME to α -NAD-ME gene composition in Angiosperms. Most species have the same copy number (1 to 1). In the cases where the relative proportion and absolute composition differ, the absolute β to α gene dosage is indicated in parentheses. Schematic representation of the values with the species grouped according to their order and family interrelationships (APG IV, 2016). COM: Celastrales, Malpighiales, and Oxalidales. NFC (Nitrogen Fixing Clades): Rosales, Fabales, Cucurbitales, and Fagales. Red * and ** indicate C_4 -NAD-ME and C_3 - C_4 photosynthetic metabolism, respectively. The scheme was created as a free art illustration where the lines do not represent branch lengths.

non- C_4 NAD-ME species (18.6%) (Fisher's exact test, $P = 0.006$). The C_3 - C_4 intermediate *Cleome paradoxa* does maintain the 1:1 β/α ratio.

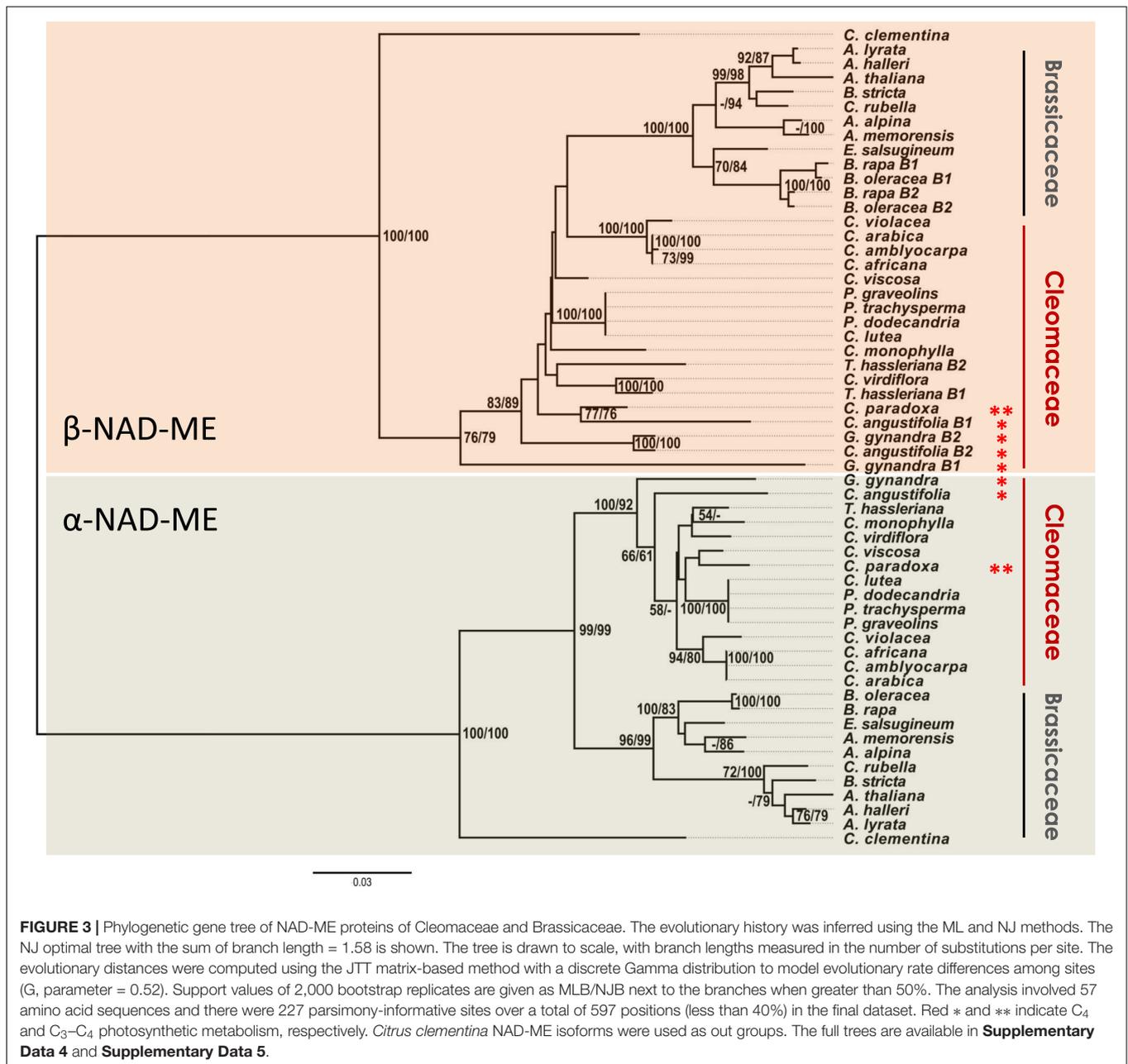
Evolution of α - and β -NAD-ME Proteins in Cleomeaceae and Brassicaceae

Gene duplication plays a critical role in generating the diversity needed for the evolution of protein through neo-functionalization. We wanted to investigate the role of gene duplication to the evolution of NAD-ME type C_4 photosynthetic metabolism in Eudicot species. NAD-ME has been independently co-opted for C_4 photosynthesis in Caryophyllales and Brassicales species. Thus, we further investigated the pattern of the NAD-ME protein evolution in the Cleomeaceae (having both C_3 and C_4 species) and in its sister-family the Brassicaceae (including Arabidopsis and Brassica crops). Most species in both families have a 1:1 ratio of β/α relative gene dosage (Figure 2). However, both families possess species carrying an additional copy of the β -NAD-ME gene ($\beta 1$ and $\beta 2$) (Figure 2) that correlate with mesopolyploidy events. In Cleomeaceae, the β -NAD-ME gene duplication is found in the C_3 species *T. hassleriana* and the C_4 species

C. angustifolia and *G. gynandra* (Figure 2). *T. hassleriana* and *G. gynandra* share the Th-alpha ancient polyploidy event, but *C. angustifolia* underwent an independent polyploidy event (Mabry et al., 2019). In Brassicaceae, both *B. rapa* and *B. oleracea* have duplicated β -NAD-ME gene associated with the Br- α genome triplication.

The NJ (Figure 3 and Supplementary Data 4) and ML (Supplementary Data 5) phylogenetic protein tree topologies were very consistent despite the low support values of some branches (MLB and NJB <50%). For closely related species, the protein-based tree has fewer informative sites than the CDS-based tree (Figure 1). However, we were interested in evaluating unusual branch positions and lengths that would suggest evolutionary rate shifts and altered amino acidic sequences. In the Brassicaceae, the relationships of the orthologs in the α -NAD-ME and β -NAD-ME clades agree with recent assessments of the species phylogenetic relationships (Nikolov and Tsiantis, 2017; Bayat et al., 2018). Conversely, α - and β -NAD-ME protein trees of Cleomeaceae are incongruent with current species trees (Feodorova et al., 2010; Patchell et al., 2014).

Regarding the α -NAD-ME orthologs, we recovered the core *Cleome sensu stricto* (*s. str.*) (*C. violacea*, *C. arabica*, *C. africana*, and *C. amblyocarpa*) and the clade containing *Polanisia* species



with well-supported MLB and NJB values (Patchell et al., 2014). However, the α -NAD-ME of the C₄ species *G. gynandra* and *C. angustifolia* were placed as the base of the clade within the Cleomaceae (Figure 3), which disagrees with known species relationships.

In the β -NAD-ME clade, the conflicting pattern of protein relationships from known species relationships is even more pronounced. Here, the β -NAD-ME of Cleomaceae was recovered as a non-monophyletic group with Brassicaceae as a sister group of the Cleome s. str. (Figure 3). Within the Brassicaceae we recovered the same groups as described in the α -NAD-ME clade. Even more so, the β -NAD-ME of the C₄ species *G. gynandra* and *C. angustifolia* appear as first-branching lineages

that precede the formation of Brassicaceae. This whole pattern could be explained by an unrealistic evolutionary scenario in which the β -NAD-ME of the C₄ species *G. gynandra* and *C. angustifolia* are encoded by ancient genes retained only in the genome of these species. Instead, the differences in rates of molecular evolution within the Cleomaceae are most probably due to selective pressures on the NAD-ME genes of the C₄ species. In addition to the incongruent positioning in the protein-based trees, the α - and β -NAD-ME of the C₄ species show branch lengths that do not correlate with those of the C₃ orthologs in Cleomaceae (Supplementary Data 4). At least in *G. gynandra*, such a high number of non-synonymous substitutions cannot be

a consequence of pseudogenization, as α - and β -NAD-ME transcripts are highly abundant in leaves (Brown et al., 2011; Kulahoglu et al., 2014).

Congruency Between NAD-ME Gene-Based Trees and Cleomaceae Species Trees

We hypothesize that evolutionary forces driving the evolution of C₄ in the Cleomaceae could be responsible for the incongruent NAD-ME protein-based phylogenetic trees. To address this hypothesis, we inferred phylogenetic trees from the third position of the codons, which is considered as a nearly neutral marker (Christin et al., 2007; Hilu et al., 2014). We conducted this analysis focused on the NAD-ME coding sequences of the Rosids, a well-defined monophyletic group capable of reconstructing true phylogenetic relationships given the high phylogenetic signals and low homoplasy (Hilu et al., 2014).

The BI (Figure 4, right and Supplementary Data 6) and ML (Figure 4, left and Supplementary Data 7) phylogenetic trees deduced from unconstrained sites were widely congruent with well-supported nodes. In the α -NAD-ME clade, we recovered the expected topology based on published species phylogenies inferred from plastid, mitochondrial and nuclear markers (Feodorova et al., 2010; Patchell et al., 2014). We found that *Polanisia* and the *Cleome s. str.* species cluster as basal lineages of C₃ species. The C₄ species *G. gynandra* and *C. angustifolia* form paraphyletic groups, with the C₃-C₄ intermediate *C. paradoxa* confidently placed in the *Angustifolia* clade and *G. gynandra* nested close to the African species *C. monophylla* (Figure 4). Finally, the C₃ species *T. hassleriana* and *C. virdiflora* confidently branch together as members of the *Tarenaya* clade. For the β -NAD-ME gene tree, the topology of the *Polanisia* and *Cleome s. str.* clades is consistent with species relationships. Again, *Polanisia* and the *Cleome s. str.* cluster together as first-branching C₃ species (Figure 4). However, β -NAD-ME duplicated genes of *T. hassleriana*, *G. gynandra*, and *C. angustifolia* do not group as paralogs gene pairs or a cluster of orthologs. Instead a complex species-dependent arrangement of $\beta 1$ and $\beta 2$ genes is observed. In the C₃ *T. hassleriana*, the $\beta 1$ copy nests with its *C. virdiflora* ortholog, in line with the α -NAD-ME grouping, and *Th* $\beta 2$ -NAD-ME is found in a separate clade with *C. monophylla*. A plausible explanation is that the duplication of β -NAD-ME occurred in a common ancestor to these three C₃ related species and, after speciation, *C. virdiflora* lost the $\beta 1$ gene and *C. monophylla* lost the $\beta 2$ gene. In the C₄ species, the $\beta 1$ gene of *C. angustifolia* groups with the *C. paradoxa* ortholog, in concordance with the α -NAD-ME lineage, but the $\beta 2$ copy groups with $\beta 2$ of *G. gynandra*. Finally, $\beta 1$ of *G. gynandra* is placed basal to the *Gg* $\beta 2$ -NAD-ME–*Ca* $\beta 2$ -NAD-ME cluster (Figure 4). Hence, the β -NAD-ME duplication and subsequent loss of one copy seems to be a possible common phenomenon in Cleomaceae, which overshadows the true gene relationship. However, because of the essential role of NAD-ME in C₄ photosynthesis, we are confident to propose that strong selective pressures accelerated the mutational rate of the α - and β -NAD-ME genes, as the species tree was globally recovered when nearly neutral markers are used.

Synteny Analysis of β -NAD-ME and α -NAD-ME Genes

The analysis of syntenic relationships of β -NAD-ME and α -NAD-ME genes in the genomes of Cleomaceae and Brassicaceae allowed us to clarify the evolutionary dynamics and duplication histories in these two families. We found that all retained copies of β -NAD-ME genes are syntenic across both families and with outgroup species (Figure 5A). Furthermore, it was evident that duplicate copies of β -NAD-ME in *Tarenaya*, *Gynandropsis*, and *Brassica* can be attributed to the known ancient polyploid histories of these species as duplicate copies of the genes were found in larger intra-genomic blocks. We also detected syntenic regions in all species derived from ancient polyploidy events, such as the At-alpha event at the origin the Brassicaceae, where a duplicated copy of β -NAD-ME was lost due to genome fractionation/gene loss processes that are known to have occurred.

We found a very different evolutionary pattern when examining the syntenic relationships for α -NAD-ME homologs. We found that α -NAD-ME genes are in a different genomic context in both families compared to all other angiosperms and in a specific and conserved context in Brassicaceae (Figure 5B) and in Cleomaceae (Figure 5C). The α -NAD-ME homologs are all syntenic with one another within the Brassicaceae, but not with Cleomaceae nor out-group species. Thus, we conclude that the ancestral α -NAD-ME gene transposed to a new genomic context in Brassicaceae after its divergence from the ancestor of Cleomaceae, potentially associated with the dramatic genome repatterning that would have occurred after the At-alpha WGD event. Whereas all three examined Cleomaceae species (*C. violacea*, *G. gynandra*, and *T. hassleriana*) α -NAD-ME orthologs are syntenic. *C. violacea* lacks evidence of the ancient polyploidy event shared between *G. gynandra* and *T. hassleriana*. Thus, the ancestor of all Cleomaceae α -NAD-ME must have transposed before the splitting of these two lineages (perhaps even earlier during the evolution of the Brassicales). The unique transpositions of α -NAD-ME to independent new genomic contexts may have had an impact of the expression and function of the gene in both Brassicaceae and Cleomaceae species compared to other eudicot species. As mentioned above, all Brassicaceae and Cleomaceae species retain only a single copy of α -NAD-ME and consistently we do not detect any retained and syntenic gene duplicates due to ancient genome duplications. Like with β -NAD-ME, we detected syntenic regions derived from ancient polyploidy events where a duplicated copy of α -NAD-ME was lost due to genome fractionation.

Amino Acid Substitutions in NAD-ME During the Evolution of C₄ Photosynthesis in Cleomaceae

We found a high number of non-synonymous substitutions in the β -NAD-ME sequences of the C₄ species (Figure 3). We further analyzed the patterns of amino acid positions in the β -NAD-ME sequences that accumulated changes during C₄ evolution in Cleomaceae compared to all Brassicales β -NAD-MEs in a MSA (Supplementary Figure 1). We found that

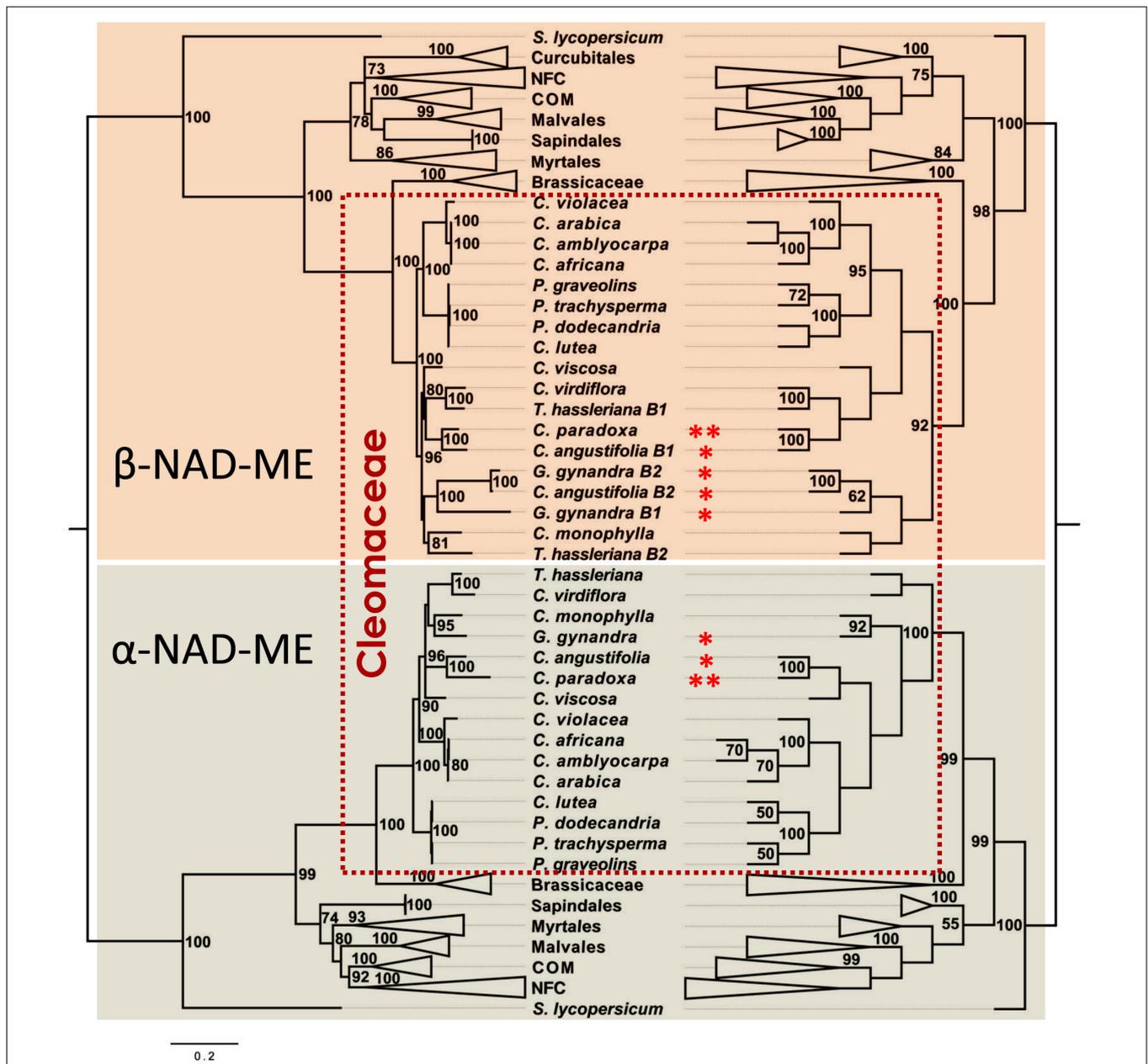
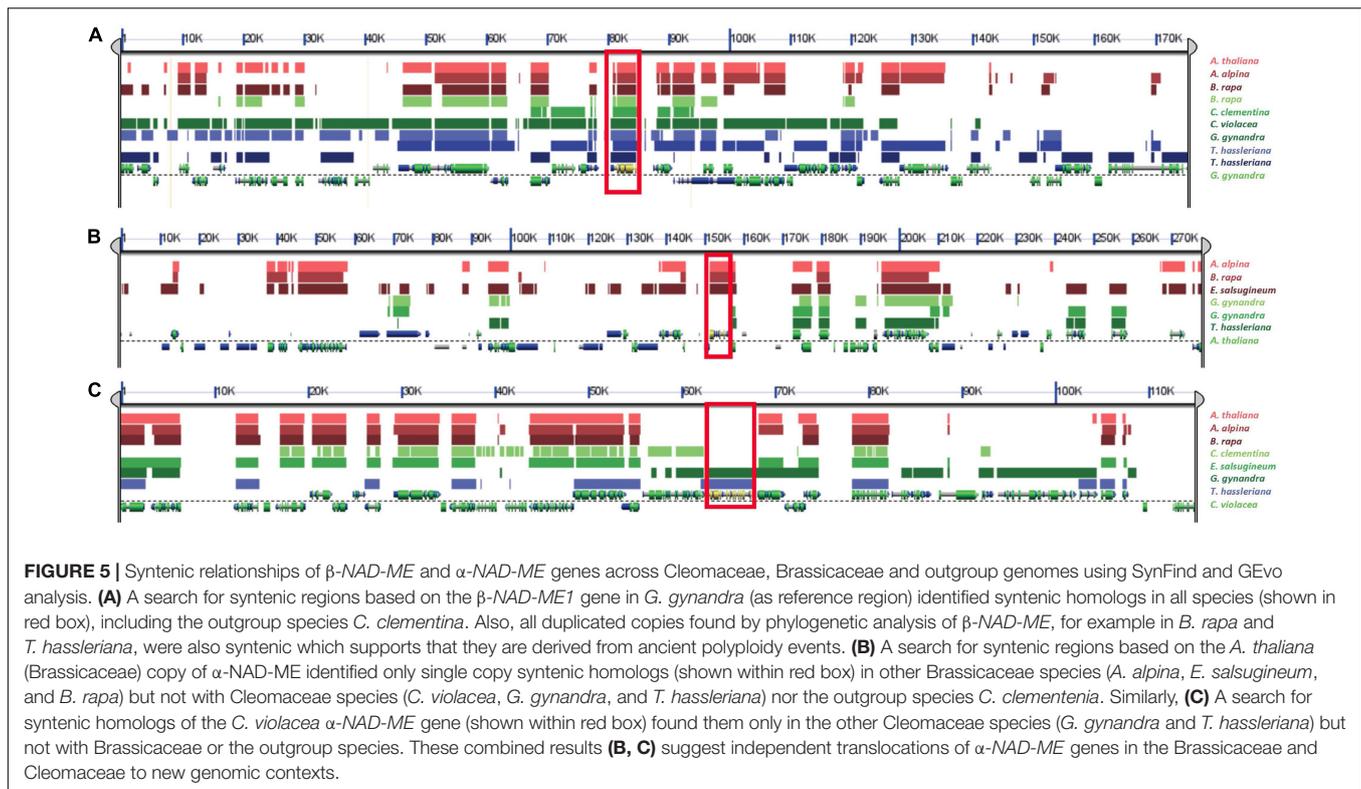


FIGURE 4 | Evolutionary history of *NAD-ME* genes in Cleomaceae deduced from third position of the codons. Bayesian (**left**) and Maximum Likelihood (**right**) phylogenetic tree of *NAD-ME* of the Rosid lineage. Most of the clades are compressed and designated by their Order or Family names. BPP and MLB values higher than 70 and 50%, respectively, are given next to the branches. In the BI analysis, the tree is drawn to scale with branch lengths measured in the number of substitutions per site. In the ML analysis, the bootstrap consensus tree (dendrogram) inferred from 2,000 replicates is taken to represent the evolutionary history of the taxa, in which the partitions reproduced in less than 50% of the bootstrap replicates are collapsed. In both analyses, the best-fit substitution model was a GRT + G (3.70) model involving 124 nucleotide sequences and a total of 497 positions in the final dataset. * and ** indicate C₄ and C₃-C₄ photosynthetic metabolism, respectively. *Solanum lycopersicum* *NAD-ME*s coding sequences were used as out groups. The full trees are available in **Supplementary Data 6** and **Supplementary Data 7**.

393 of 579 amino acids (~68%) are identical in all mature β -NAD-ME protein sequences (**Supplementary Figure 1**). The most diverged protein sequences are the sequences of C₄ species *G. gynandra* β 1-NAD-ME (186 amino acids, ~32%) and *C. angustifolia* β 1-NAD-ME (158 amino acids, ~27%). Interestingly, β 1-NAD-MEs of *G. gynandra* and *C. angustifolia*

share many changed amino acids and both sequences differ only in 83 positions.

To identify amino acid residues potentially involved in C₄-optimization, we compared the NAD-MEs of the Brassicales in the MSA searching for differentially substituted amino acids in the NAD-ME protein sequences of the C₄ species. A differentially



substituted amino acid is a position in the MSA at which a NAD-ME protein of a C₄ species has an amino acid that differs from all other NAD-ME proteins, which all share the same amino acid (Alvarez et al., 2019).

We found three amino acid positions, F127, Q205, and N466 (numbered according to the full-length sequence of *G. gynandra*; **Supplementary Figure 2**), strictly differentially substituted in both *G. gynandra* and *C. angustifolia* α -NAD-ME proteins (**Supplementary Figure 2**, orange amino acids). Interestingly, we found that 36 amino acids are differentially substituted in the β 1-NAD-ME from *G. gynandra* (**Supplementary Figure 1**, light blue + orange amino acids). Similarly, we identified 13 amino acids changes in the β 1-NAD-ME of *C. angustifolia* (**Supplementary Figure 1**, dark blue + orange amino acids). Interestingly, 5 amino acid positions, V131, S132, H195, V297, and K605 (numbered according to the full-length sequence of *G. gynandra*) are identically substituted in both β 1-NAD-ME of the C₄ species *G. gynandra* and *C. angustifolia* (**Supplementary Figure 1**, orange amino acids). This kind of amino acid replacement was not observed in the β 2-NAD-ME isoforms of the C₄ species.

The strictly differentially substituted amino acids in the α -NAD-ME and β 1-NAD-ME of *G. gynandra* and *C. angustifolia* suggest that these amino acids probably evolved under positive selection. To address this, we carried out model tests to identified residues that underwent adaptive changes in the α -NAD-ME and β 1-NAD-MEs of the C₄ species. The models were optimized using the tree topology inferred from third positions of codons (**Figure 4**, right) and the mature α -NAD-ME and β -NAD-ME

coding sequences of Cleomaceae in the MSA. For both paralogs, one model (M8), allowing a proportion of codons evolving under positive selection, provided better fit than the null model (M7) (**Supplementary Tables 2,3**). We found five amino acid positions having a posterior probability greater than 0.9 in the α -NAD-MEs: 80, 205, 423, 500, and 587 (numbered according to the full-length sequence of *G. gynandra*; **Supplementary Figure 2**). The position 205 corresponds to the differentially substituted Q205 in the α -NAD-ME of *G. gynandra* and *C. angustifolia* (**Supplementary Figure 2**). For the β 1-NAD-MEs, three amino acid positions had a posterior probability greater than 0.9: 132, 297, and 395 (numbered according to the full-length sequence of *G. gynandra*, **Supplementary Figure 1**). The positively selected positions 132 and 297 correspond to the differentially substituted S132 and V297 residues in β 1-NAD-ME of *G. gynandra* and *C. angustifolia*.

DISCUSSION

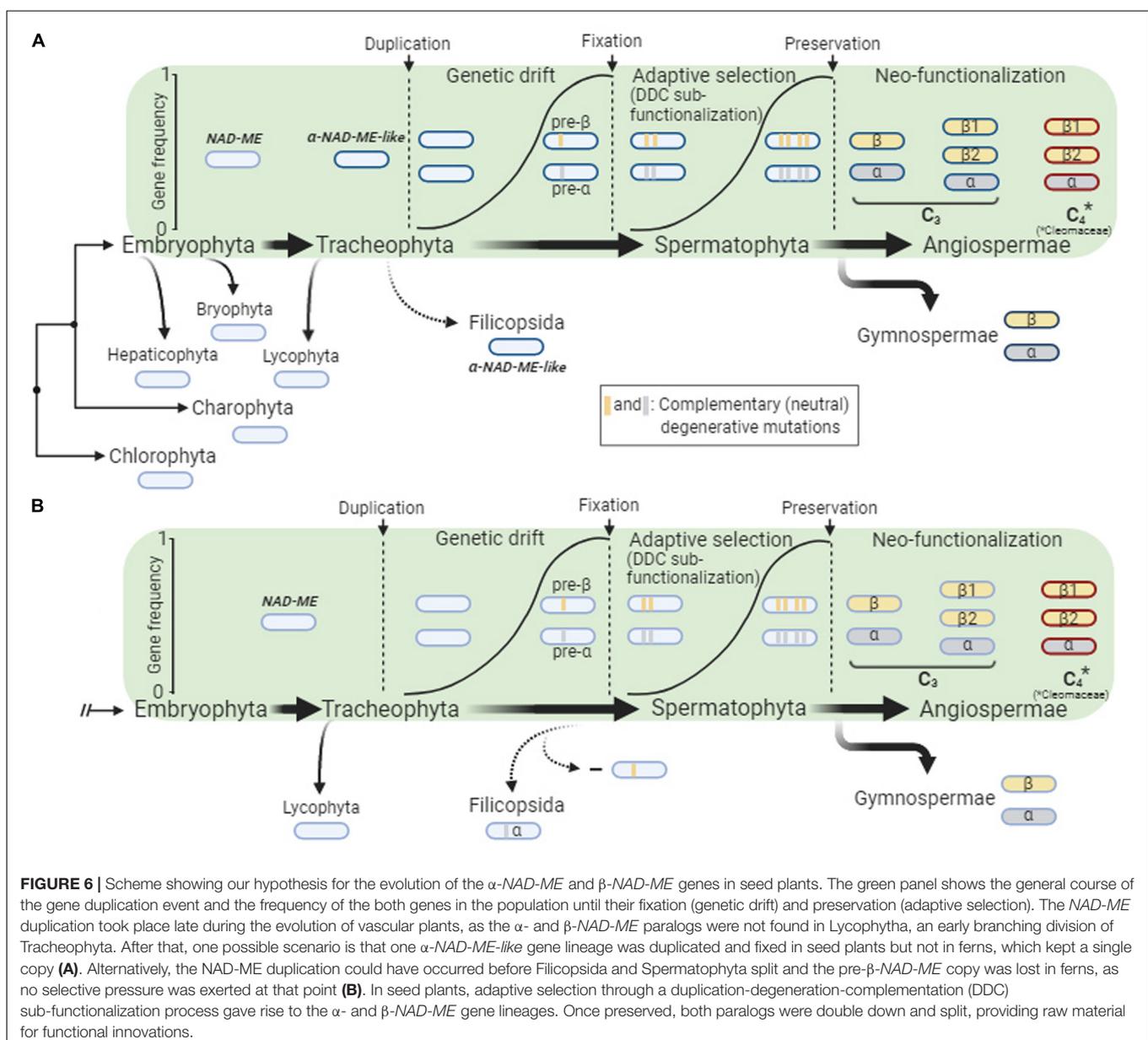
Sub-Functionalization Set Up the Heteromeric Assembly of NAD-ME and Neo-Functionalization Optimized the α - and β -Subunit Functions

We identified a duplication of *NAD-ME* in vascular plants that generated the two paralogs lineages: α - and β -NAD-ME. All seed plants examined maintained at least one α -NAD-ME and one β -NAD-ME homolog (**Figure 1** and **Supplementary Data 1**), with most species maintaining a 1:1 α/β relative gene

dosage (Figure 2). This is a strong indication that the α - and β -NAD-ME homologs diversified and evolved non-redundant, alternative and critical (housekeeping) functions. In higher plants the heteromeric assembly of NAD-ME is most probably a requirement to fully fit into the central carbon metabolism, and specific duplication of the NAD-ME subunit-coding genes are evolutionarily allowed.

The α -NAD-ME and β -NAD-ME gene lineages evolved by duplication of an ancestral NAD-ME gene that occurred late during the evolution of vascular plants, as the paralogs were not found in the early branching tracheophyte *S. moellendorffii* (Figure 1). The well supported positioning of the single NAD-ME genes of ferns as a sister group of spermatophytes α -NAD-ME (Figure 1, Supplementary Data 2 and Supplementary Data 3)

suggests two alternative scenarios for the duplication of NAD-ME (Figure 6): (A) an ancestral α -NAD-ME-like gene likely diverged in Tracheophytes after the Lycophytes separation. This ancestral gene duplicated at the origin of the seed plants giving rise to *pre*- α -NAD-ME and *pre*- β -NAD-ME paralogs sequences, which afterward diverged from the ancestral α -NAD-ME-like gene. Finally, the paralogs were fixed (possibly by genetic drifts) and adaptive selection preserved them in seed plants. In this scenario, a primitive α -NAD-ME-like gene was retained in lycophytes; alternatively, (B) a NAD-ME duplication giving rise to *pre*- α -NAD-ME and *pre*- β -NAD-ME paralogs took place in Tracheophytes directly after the Lycophytes separation. The *pre*- β -NAD-ME was then lost in the ferns before adaptive selection could preserve both genes (Figure 6). Instead, *pre*- α -NAD-ME



and *pre*- β -NAD-ME diverged from each other and were preserved as the α - and β -NAD-ME paralogs in seed plants.

Four major models have been proposed for preservation of duplicated genes: neo-functionalization, specialization, dosage selection and duplication-degeneration-complementation (DDC) sub-functionalization (Conant and Wolfe, 2008; Innan and Kondrashov, 2010). Neo-functionalization and specialization provide an explanation for the cases where after fixation by genetic drift, one or both duplicate genes diverge from the ancestral function (e.g., in expression patterns, substrate specificities, etc.). The dosage selection model supposes that the new duplicated gene evolves by positive selection (e.g., the increase in the amount of protein is beneficial; the new copy emerges with a novel function or by shielding against deleterious mutations). The DDC sub-functionalization is a conservative evolutionary model that assumes that the functions of an ancestral gene have been neutrally divided among the daughter copies due to complementary degenerative mutations: neither copy is able to fulfill the original functions on their own.

Previous findings support the DDC sub-functionalization process by which the duplicated α - and β -NAD-ME copies could be preserved: (i) To fulfill the housekeeping function in L-malate respiration in seed plants, the NAD-ME is predominantly assembled as an heteromer of α - and β -subunits, with both catalyzing the same reaction (Tronconi et al., 2008, 2010b); (ii) both α -NAD-ME and β -NAD-ME genes are constitutively expressed and coordinately regulated and the encoded proteins accumulate at similar levels in the most plant tissues (Tronconi et al., 2008, 2010b; Fuchs et al., 2020); (iii) knockout lines of Arabidopsis lacking either α -NAD-ME or β -NAD-ME show residual NAD-ME activities, which in sum do not reach the activity measured in wildtype. Moreover, the mutants plants lacking β -NAD-ME retain less than 10% of the total NAD-ME activity (Tronconi et al., 2008; Brown et al., 2011); (iv) The identification of coevolved connected amino acidic residues belonging to the α -NAD-ME and β -NAD-ME subunits (Tronconi et al., 2018) indicates that compensatory neutral mutations marked the evolution toward a functional heteromeric NAD-ME in higher plants.

Following the preservation phase, neo-functionalization can occur in one or both duplicated genes (Conant and Wolfe, 2008). In this regard, the comprehensive characterization of *A. thaliana* NAD-ME indicated that the α -NAD-ME and β -NAD-ME homodimers and the α/β -NAD-ME heterodimer behave differently in terms of catalytic mechanism, interaction with the substrates and allosteric regulation (Tronconi et al., 2008, 2010a,b, 2015), pointing out an adaptive advantage in terms of metabolic flexibility. This metabolic flexibility might have played a role in the adaption of a C₄-specific version of the NAD-ME. Moreover, α - and β -NAD-ME differentially accumulate in the separate components of the floral organ (Tronconi et al., 2010b). In sepals, the α -NAD-ME is present at a slightly higher proportion than β -NAD-ME. On the other hand, β -NAD-ME is the only protein present in anthers. All these observations suggest that NAD-ME activity may be regulated by variations of the native association *in vivo*, rendering enzymatic entities with distinct allosteric regulation to fulfill additional

metabolic roles (Maurino et al., 2009; Tronconi et al., 2010b; Maier et al., 2011).

Changes in the Subunit-Specific Duplications Provide the Basis for the Recruitment of NAD-ME in C₄ Biochemistry

In multiple families of the angiosperms we found a higher number of NAD-ME genes coding for β -NAD-ME (20 species) than for α -NAD-ME (one species) (Figure 2). This indicates that if small-scale gene duplications or WGDs occurred, the gene coding for a β -NAD-ME is tolerant of independent copy number variation (possibly fixed by genetic drift) while the α -NAD-ME gene is under evolutionary pressure to return to a 1:1 β/α relative status. It seems that the α -NAD-ME duplication imparts a detrimental effect or does not increase plant fitness and thus, the duplicated gene is not fixed in the population. A new copy can escape the negative effect of the dosage if during the fixation phase (Figure 6) a mutation conferring an adaptive advantage arises (Conant and Wolfe, 2008). This can explain the observation that the C₄-NAD-ME species *A. hypochondriacus* has retained two α -NAD-ME genes (Figure 2). Probably, shortly after the duplication, a copy quickly diverged from the original function by neo-functionalization to fit to the C₄ biochemistry.

The significantly higher proportion of C₄-NAD-ME species possessing an additional NAD-ME gene copy (Figure 2) is consistent with the general notion that gene duplication is a precondition for the evolution of the C₄ functions (Lynch and Conery, 2000; Moore and Purugganan, 2005). For other C₄ genes, one copy of a duplicated gene is neo-functionalized without affecting the other members of the genetic family (Christin et al., 2013). However, in the C₄ species of Cleomaceae all three genes were potentially affected by adaptive selection as suggested by the inconsistencies observed in the protein-based tree topology (Figure 3). This could be rectified for the α -NAD-ME genes, albeit partially for the β -NAD-ME genes, when a nearly neutral marker (third codon positions) was used (Figure 4). Higher amino acid substitution rates are associated with an accelerated evolution or potential positive selection due to neo-functionalization of the resulting protein sequence. Neo-functionalization from a duplicated gene is a classic driver of protein evolution.

We detected a high proportion of base substitutions in the Cleomaceae C₄ species β -NAD-ME genes. The β -NAD-ME genes are duplicated in the Cleomaceae C₄ species *C. angustifolia* and *G. gynandra* and the C₃ species *T. hassleriana*, and in the Brassicaceae *B. rapa* and *B. oleracea*. Synteny analysis clearly showed that duplication of β -NAD-ME in both families (Figure 5A) was due to the ancient polyploidy events, such as the Brassica lineage genome triplication and the *Tarenaya* WGD that is shared with *G. gynandra* (Cheng et al., 2014; van den Bergh et al., 2014). Importantly, changes in copy number of genes can maintain dosage-balance relationships if generated by polyploidy events. Nevertheless, only the β 1-NAD-ME of both C₄ species has accumulated a high number of amino acid changes. Intriguingly, five positions show the same amino acid changes in both β 1-NAD-ME of these C₄ (Supplementary Figure 1),

two of which (S132 and V295) evolved under positive selection (**Supplementary Table 3**). We hypothesize that these amino acids play a role in the C₄ function of NAD-ME, as *C. angustifolia* and *G. gynandra* belong to different C₄ lineages and the five amino acids are differentially conserved in the β-NAD-ME of C₃ Cleome species, Brassicales β-NAD-ME or the respective β₂-NAD-ME of *C. angustifolia* and *G. gynandra*. Interesting, four of these five amino acid are involved in substrate coordination or are located in the L-malate binding domain (Chang and Tong, 2003). Finally, we found three differentially conserved amino acid substitutions in the α-NAD-ME of the C₄ species (**Supplementary Figure 2**), one of them (Q205) identified as an adaptive change (**Supplementary Table 2**). This adaptive and differentially conserved position in the α-NAD-ME is neither part of the enzymatic active center nor was shown to participate in the enzymatic mechanism of reaction (Chang and Tong, 2003). Because of the heteromeric NAD-ME assembly, this substitution most likely represents a change necessary to compensate for the evolutionary changes in the β₁-NAD-ME, or enable another kind of function that is necessary in the adaptation of NAD-ME in the C₄ context.

CONCLUSION

It appears that the genes encoding C₄ enzymes evolved by simply duplication of an original metabolic enzyme and further neo-functionalization (Christin et al., 2013; Ludwig, 2016; Alvarez et al., 2019). NAD-ME turns out to be an exceptional case, probably due to its heteromeric structure. NAD-ME followed an intricate molecular mechanism of evolution marked by sub-functionalization and differences in the frequency of α- and β-NAD-ME gene duplication (**Figure 6**). Future work should focus on how NAD-ME in C₄ plant mitochondria has been adapted to perform both housekeeping and C₄-associated functions.

REFERENCES

- Alvarez, C. E., Bovdilova, A., Hoppner, A., Wolff, C. C., Saigo, M., Trajtenberg, F., et al. (2019). Molecular adaptations of NADP-malic enzyme for its function in C₄ photosynthesis in grasses. *Nat. Plants* 5, 755–765. doi: 10.1038/s41477-019-0451-7
- Anderson, L. E. (1971). Chloroplast and cytoplasmic enzymes. II. Pea leaf triose phosphate isomerases. *Biochim. Biophys. Acta* 235, 237–244. doi: 10.1016/0005-2744(71)90051-9
- Artus, N. N., and Edwards, G. E. (1985). NAD-malic enzyme from plants. *FEBS Lett.* 182, 225–233. doi: 10.1016/0014-5793(85)80305-7
- Barker, M. S., Vogel, H., and Schranz, M. E. (2009). Paleopolyploidy in the brassicales: analyses of the cleome transcriptome elucidate the history of genome duplications in *Arabidopsis* and other brassicales. *Genome Biol. Evol.* 1, 391–399. doi: 10.1093/gbe/evp040
- Bassham, J. A., Benson, A. A., Kay, L. D., Harris, A. Z., Wilson, A. T., and Calvin, M. (1954). The path of carbon in photosynthesis. XXI. The cyclic regeneration of carbon dioxide acceptor. *J. Am. Chem. Soc.* 76, 1760–1770. doi: 10.1021/ja01636a012
- Bayat, S., Schranz, M. E., Roalson, E. H., and Hall, J. C. (2018). Lessons from Cleomaceae, the sister of crucifers. *Trend. Plant Sci.* 23, 808–821. doi: 10.1016/j.tplants.2018.06.010

DATA AVAILABILITY STATEMENT

All datasets presented in this study are included in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

VM conceived the project in active discussion with all co-authors. MT and MH performed the phylogenetic analysis. MS performed the syntenic analysis. All authors contributed to data analysis and writing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft under grant MA2379/18-1 to VM and the National Agency for Scientific and Technological Promotion PICT2015-1064 to MT.

ACKNOWLEDGMENTS

We thank J. Chris Pires and Makenzie Mabry for access to Cleomaceae transcriptome data.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2020.572080/full#supplementary-material>

- Brown, N. J., Newell, C. A., Stanley, S., Chen, J. E., Perrin, A. J., Kajala, K., et al. (2011). Independent and parallel recruitment of preexisting mechanisms underlying C-4 photosynthesis. *Science* 331, 1436–1439. doi: 10.1126/Science.1201248
- Chang, G. G., and Tong, L. (2003). Structure and function of malic enzymes, a new class of oxidative decarboxylases. *Biochemistry* 42, 12721–12733. doi: 10.1021/bi035251+
- Cheng, F., Wu, J., and Wang, X. (2014). Genome triplication drove the diversification of *Brassica* plants. *Hortic. Res.* 1:14024. doi: 10.1038/hortres.2014.24
- 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. (2013). Parallel recruitment of multiple genes into c₄ photosynthesis. *Genome Biol. Evol.* 5, 2174–2187. doi: 10.1093/gbe/evt168
- 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 genomics and phylogenetics. *Genome Biol. Evol.* 1, 221–230. doi: 10.1093/gbe/evp020

- Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482
- Detarsio, E., Alvarez, C. E., Saigo, M., Andreo, C. S., and Drincovich, M. F. (2007). Identification of domains involved in tetramerization and malate inhibition of maize C4-NADP-malic enzyme. *J. Biol. Chem.* 282, 6053–6060. doi: 10.1074/jbc.M609436200
- Detarsio, E., Maurino, V. G., Alvarez, C. E., Muller, G. L., Andreo, C. S., and Drincovich, M. F. (2008). Maize cytosolic NADP-malic enzyme (ZmCytNADP-ME): a phylogenetically distant isoform specifically expressed in embryo and emerging roots. *Plant Mol. Biol.* 68, 355–367. doi: 10.1007/s11103-008-9375-8
- Drincovich, M. F., Lara, M. V., Andreo, C. S., and Maurino, V. G. (2011). “Evolution of C4 decarboxylases: different solutions for the same biochemical problem: provision of CO₂ in bundle sheath cells,” in *C4 Photosynthesis and Related CO₂ Concentration Mechanisms; Advances in Photosynthesis and Respiration*, ed. S. Govindjee (Berlin: Springer), 277–300. doi: 10.1007/978-90-481-9407-0_14
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Edger, P. P., Heidel-Fischer, H. M., Bekaert, M., Rota, J., Glockner, G., Platts, A. E., et al. (2015). The butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8362–8366. doi: 10.1073/pnas.1503926112
- Edwards, E. J., Osborne, C. P., Strömberg, C. A. E., and Smith, S. A. (2010). The origins of C4 grasslands: integrating evolutionary and ecosystem science. *Science* 328, 587–591. doi: 10.1126/science.1177216
- Emanuelsson, O., Nielsen, H., Brunak, S., and von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J. Mol. Biol.* 300, 1005–1016. doi: 10.1006/jmbi.2000.3903
- Feodorova, T. A., Voznesenskaya, E. V., Edwards, G. E., and Roalson, E. H. (2010). Biogeographic patterns of diversification and the origins of C-4 in cleome (Cleomeaceae). *Syst. Bot.* 35, 811–826. doi: 10.1600/036364410x539880
- Fuchs, P., Rugen, N., Carrie, C., Elsasser, M., Finkemeier, I., Giese, J., et al. (2020). Single organelle function and organization as estimated from *Arabidopsis* mitochondrial proteomics. *Plant J.* 101, 420–441. doi: 10.1111/tpj.14534
- Furbank, R. T., and Hatch, M. D. (1987). Mechanism of c(4) photosynthesis: the size and composition of the inorganic carbon pool in bundle sheath cells. *Plant Physiol.* 85, 958–964. doi: 10.1104/pp.85.4.958
- Gerrard Wheeler, M. C., Arias, C. L., Tronconi, M. A., Maurino, V. G., Andreo, C. S., and Drincovich, M. F. (2008). *Arabidopsis thaliana* NADP-malic enzyme isoforms: high degree of identity but clearly distinct properties. *Plant Mol. Biol.* 67, 231–242. doi: 10.1007/s11103-008-9313-9
- Gerrard Wheeler, M. C., Tronconi, M. A., Drincovich, M. F., Andreo, C. S., Flugge, U. I., and Maurino, V. G. (2005). A comprehensive analysis of the NADP-malic enzyme gene family of *Arabidopsis*. *Plant Physiol.* 139, 39–51. doi: 10.1104/pp.105.065953
- González-Moro, B., Lacuesta, M., Becerril, J. M., González-Murua, C., and Muñoz-Rueda, A. (1997). Glycolate accumulation causes a decrease of photosynthesis by inhibiting RUBISCO activity in maize. *J. Plant Physiol.* 150, 388–394. doi: 10.1016/s0176-1617(97)80087-9
- Grover, S. D., Canellas, P. F., and Wedding, R. T. (1981). Purification of NAD malic enzyme from potato and investigation of some physical and kinetic properties. *Arch. Biochem. Biophys.* 209, 396–407. doi: 10.1016/0003-9861(81)90297-6
- Hatch, M. D. (1971). “Mechanism and function of C4 photosynthesis,” in *Photosynthesis and Photorespiration*, eds M. D. Hatch, C. B. Osmond, and R. O. Slatyer (New York: Wiley-Interscience), 139–152.
- Hatch, M. D. (1987). C4 photosynthesis: a unique blend of modified biochemistry, anatomy and ultrastructure. *Biochim. Biophys. Acta* 895, 81–106. doi: 10.1016/s0304-4173(87)80009-5
- Hayes, J. M. (1994). “Global methanotrophy at the Archean-Proterozoic transition,” in *Early life on Earth*, ed. S. Bengtson (New York, NY: Columbia University Press), 220–236.
- Heckmann, D., Schulze, S., Denton, A., Gowik, U., Westhoff, P., Weber, A. P. M., et al. (2013). Predicting C-4 photosynthesis evolution: modular, individually adaptive steps on a mount fuji fitness landscape. *Cell* 153, 1579–1588. doi: 10.1016/j.cell.2013.04.058
- Hillis, D. M., and Bull, J. J. (1993). An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42, 182–192. doi: 10.2307/2992540
- Hilu, K. W., Black, C. M., and Oza, D. (2014). Impact of gene molecular evolution on phylogenetic reconstruction: a case study in the rosids (Superorder Rosanae, Angiosperms). *PLoS one* 9:e99725. doi: 10.1371/journal.pone.0099725
- Innan, H., and Kondrashov, F. (2010). The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* 11, 97–108. doi: 10.1038/nrg2689
- Kanai, R., and Edwards, E. J. (1999). “The biochemistry of C4 photosynthesis,” in *C4 Plant Biology*, eds R. F. Sage and R. K. Monson (San Diego, CA: Academic Press), 49–87. doi: 10.1016/b978-012614440-6/50004-5
- Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* 33, 511–518. doi: 10.1093/nar/gki198
- Kelly, G. J., and Lutzko, E. (1976). Inhibition of spinach-leaf phosphofructokinase by 2-phosphoglycollate. *FEBS Lett.* 68, 55–58. doi: 10.1016/0014-5793(76)80403-6
- Koteyeva, N. K., Voznesenskaya, E. V., Roalson, E. H., and Edwards, G. E. (2011). Diversity in forms of C4 in the genus *Cleome* (Cleomeaceae). *Ann. Bot.* 107, 269–283. doi: 10.1093/aob/mcq239
- Kulahoglu, C., Denton, A. K., Sommer, M., Mass, J., Schliesky, S., Wrobel, T. J., et al. (2014). Comparative transcriptome atlases reveal altered gene expression modules between two Cleomeaceae C3 and C4 plant species. *Plant Cell* 26, 3243–3260. doi: 10.1105/tpc.114.123752
- Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* 35, 1547–1549. doi: 10.1093/molbev/msy096
- Kuraku, S., Zmasek, C. M., Nishimura, O., and Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Res.* 41, W22–W28. doi: 10.1093/nar/gkt389
- Leakey, A. D., and Lau, J. A. (2012). Evolutionary context for understanding and manipulating plant responses to past, present and future atmospheric [CO₂]. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 613–629. doi: 10.1098/rstb.2011.0248
- Li, F. W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P. M., et al. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* 4, 460–472. doi: 10.1038/s41477-018-0188-8
- Long, J. J., Wang, J. L., and Berry, J. O. (1994). Cloning and analysis of the C4 photosynthetic NAD-dependent malic enzyme of amaranth mitochondria. *J. Biol. Chem.* 269, 2827–2833.
- Ludwig, M. (2016). Evolution of carbonic anhydrase in C4 plants. *Curr. Opin. Plant Biol.* 31, 16–22. doi: 10.1016/j.pbi.2016.03.003
- Lynch, M., and Conery, J. S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155. doi: 10.1126/science.290.5494.1151
- Mabry, M. E., Brose, J. M., Blischak, P. D., Sutherland, B., Dismukes, W. T., Bottoms, C. A., et al. (2019). Phylogeny and multiple independent whole-genome duplication events in the brassicales. *bioRxiv* [Preprint]. doi: 10.1101/789040
- Maier, A., Zell, M. B., and Maurino, V. G. (2011). Malate decarboxylases: evolution and roles of NAD(P)-ME isoforms in species performing C(4) and C(3) photosynthesis. *J. Exp. Bot.* 62, 3061–3069. doi: 10.1093/jxb/err024
- Maurino, V. G., and Peterhansel, C. (2010). Photorespiration: current status and approaches for metabolic engineering. *Curr. Opin. Plant Biol.* 13, 249–256. doi: 10.1016/j.pbi.2010.01.006
- Maurino, V. G., Saigo, M., Andreo, C. S., and Drincovich, M. F. (2001). Non-photosynthetic ‘malic enzyme’ from maize: a constitutively expressed enzyme that responds to plant defence inducers. *Plant Mol. Biol.* 45, 409–420. doi: 10.1023/A:1010665910095
- Maurino, V. G., Wheeler, M. C. G., Andreo, C. S., and Drincovich, M. F. (2009). Redundancy is sometimes seen only by the uncritical: does *Arabidopsis* need six malic enzyme isoforms? *Plant Sci.* 176, 715–721. doi: 10.1016/j.plantsci.2009.02.012
- Moore, R. C., and Purugganan, M. D. (2005). The evolutionary dynamics of plant duplicate genes. *Curr. Opin. Plant Biol.* 8, 122–128. doi: 10.1016/j.pbi.2004.12.001

- Murata, T., Ohsugi, R., Matsuoka, M., and Nakamoto, H. (1989). Purification and Characterization of NAD Malic Enzyme from Leaves of *Eleusine coracana* and *Panicum dichotomiflorum*. *Plant Physiol.* 89, 316–324. doi: 10.1104/pp.89.1.316
- Nikolov, L. A., and Tsiantis, M. (2017). Using mustard genomes to explore the genetic basis of evolutionary change. *Curr. Opin. Plant Biol.* 36, 119–128. doi: 10.1016/j.pbi.2017.02.005
- Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K. K., et al. (2018). The chara genome: secondary complexity and implications for plant terrestrialization. *Cell* 174, 448.e24–464.e24. doi: 10.1016/j.cell.2018.06.033
- Patchell, M. J., Roalson, E. H., and Hall, C. H. (2014). Resolved phylogeny of Cleomaceae based on all three genomes. *Taxon* 63, 315–328. doi: 10.12705/632.17
- Proost, S., Van Bel, M., Vanechoutte, D., Van de Peer, Y., Inze, D., Mueller-Roeber, B., et al. (2015). PLAZA 3.0: an access point for plant comparative genomics. *Nucleic Acids Res.* 43, D974–D981. doi: 10.1093/nar/gku986
- Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Sage, R. F., Christin, P. A., and Edwards, E. J. (2011). The C(4) plant lineages of planet Earth. *J. Exp. Bot.* 62, 3155–3169. doi: 10.1093/jxb/err048
- Sage, R. F., and Monson, R. K. (1999). *C4 Plant Biology*. San Diego: Academic.
- Sage, R. F., Sage, T. L., and Kocacinar, F. (2012). Photorespiration and the evolution of C4 photosynthesis. *Annu. Rev. Plant Biol.* 63, 19–47. doi: 10.1146/annurev-arplant-042811-105511
- Saigo, M., Bologna, F. P., Maurino, V. G., Detarsio, E., Andreo, C. S., and Drincovich, M. F. (2004). Maize recombinant non-C4 NADP-malic enzyme: a novel dimeric malic enzyme with high specific activity. *Plant Mol. Biol.* 55, 97–107. doi: 10.1007/s11103-004-0472-z
- Scheffler, K., and Seoighe, C. (2005). A Bayesian model comparison approach to inferring positive selection. *Mol. Biol. Evol.* 22, 2531–2540. doi: 10.1093/molbev/msi250
- Schranz, M. E., and Mitchell-Olds, T. (2006). Independent ancient polyploidy events in the sister families Brassicaceae and Cleomaceae. *Plant Cell* 18, 1152–1165. doi: 10.1105/tpc.106.041111
- Shen, H., Jin, D., Shu, J. P., Zhou, X. L., Lei, M., Wei, R., et al. (2018). Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *Gigascience* 7, 1–11. doi: 10.1093/gigascience/gix116
- Tang, H., Bomhoff, M. D., Briones, E., Zhang, L., Schnable, J. C., and Lyons, E. (2015). SynFind: compiling syntenic regions across any set of genomes on demand. *Genome Biol. Evol.* 7, 3286–3298. doi: 10.1093/gbe/evv219
- Tausta, S. L., Coyle, H. M., Rothermel, B., Stiefel, V., and Nelson, T. (2002). Maize C4 and non-C4 NADP-dependent malic enzymes are encoded by distinct genes derived from a plastid-localized ancestor. *Plant Mol. Biol.* 50, 635–652.
- Tronconi, M. A., Andreo, C. S., and Drincovich, M. F. (2018). Chimeric structure of plant malic enzyme family: different evolutionary scenarios for NAD- and NADP-dependent isoforms. *Front. Plant Sci.* 9:565. doi: 10.3389/fpls.2018.00565
- Tronconi, M. A., Fahnenstich, H., Gerrard Weehler, M. C., Andreo, C. S., Flugge, U. I., Drincovich, M. F., et al. (2008). *Arabidopsis* NAD-malic enzyme functions as a homodimer and heterodimer and has a major impact on nocturnal metabolism. *Plant Physiol.* 146, 1540–1552. doi: 10.1104/pp.107.114975
- Tronconi, M. A., Gerrard Wheeler, M. C., Maurino, V. G., Drincovich, M. F., and Andreo, C. S. (2010a). NAD-malic enzymes of *Arabidopsis thaliana* display distinct kinetic mechanisms that support differences in physiological control. *Biochem. J.* 430, 295–303. doi: 10.1042/BJ20100497
- Tronconi, M. A., Maurino, V. G., Andreo, C. S., and Drincovich, M. F. (2010b). Three different and tissue-specific NAD-malic enzymes generated by alternative subunit association in *Arabidopsis thaliana*. *J. Biol. Chem.* 285, 11870–11879. doi: 10.1074/jbc.M109.097477
- Tronconi, M. A., Wheeler, M. C., Martinatto, A., Zubimendi, J. P., Andreo, C. S., and Drincovich, M. F. (2015). Allosteric substrate inhibition of *Arabidopsis* NAD-dependent malic enzyme 1 is released by fumarate. *Phytochemistry* 111, 37–47. doi: 10.1016/j.phytochem.2014.11.009
- van den Bergh, E., Kùlahoglu, C., Bräutigam, A., Hibberd, J. M., Weber, A. P. M., Zhu, X.-G., et al. (2014). Gene and genome duplications and the origin of C4 photosynthesis: birth of a trait in the Cleomaceae. *Curr. Plant Biol.* 1, 2–9. doi: 10.1016/j.cpb.2014.08.001
- Vicentini, F. C., Denes, F. T., Borges, L. L., Silva, F. A., Machado, M. G., and Srougi, M. (2008). Laparoscopic pyeloplasty in children: is the outcome different in children under 2 years of age? *J. Pediatr. Urol.* 4, 348–351. doi: 10.1016/j.jpuro.2008.03.001
- von Caemmerer, S. (2000). *Biochemical Models of Leaf Photosynthesis*. Collingwood: CSIRO Publishing.
- Wang, Y., Brautigam, A., Weber, A. P., and Zhu, X. G. (2014). Three distinct biochemical subtypes of C4 photosynthesis? A modelling analysis. *J. Exp. Bot.* 65, 3567–3578. doi: 10.1093/jxb/eru058
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088
- Yang, Z., Nielsen, R., Goldman, N., and Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155, 431–449.
- Zou, C. S., Li, L. T., Miki, D., Li, D. L., Tang, Q. M., Xiao, L. H., et al. (2019). The genome of broomcorn millet. *Nat. Commun.* 10:436. doi: 10.1038/s41467-019-08409-5

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Tronconi, Hüdig, Schranz and Maurino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.