



# Chimeric Structure of Plant Malic Enzyme Family: Different Evolutionary Scenarios for NAD- and NADP-Dependent Isoforms

Marcos A. Tronconi\*, Carlos S. Andreo and María F. Drincovich

Centro de Estudios Fotosintéticos y Bioquímicos, Universidad Nacional de Rosario, Rosario, Argentina

## OPEN ACCESS

### Edited by:

Jean Rivoal,  
Université de Montréal, Canada

### Reviewed by:

Jeffrey Charles Waller,  
Mount Allison University, Canada

Dave Speijer,

University of Amsterdam, Netherlands

Sébastien Renaut,

Université de Montréal, Canada

### \*Correspondence:

Marcos A. Tronconi  
[tronconi@cefobi-conicet.gov.ar](mailto:tronconi@cefobi-conicet.gov.ar)

### Specialty section:

This article was submitted to  
Plant Physiology,  
a section of the journal  
Frontiers in Plant Science

Received: 13 December 2017

Accepted: 10 April 2018

Published: 11 May 2018

### Citation:

Tronconi MA, Andreo CS and  
Drincovich MF (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

Malic enzyme (ME) comprises a family of proteins with multiple isoforms located in different compartments of eukaryotic cells. In plants, cytosolic and plastidic enzymes share several characteristics such as NADP specificity (NADP-ME), oxaloacetate decarboxylase (OAD) activity, and homo-oligomeric assembly. However, mitochondrial counterparts are NAD-dependent proteins (mNAD-ME) lacking OAD activity, which can be structured as homo- and hetero-oligomers of two different subunits. In this study, we examined the molecular basis of these differences using multiple sequence analysis, structural modeling, and phylogenetic approaches. Plant mNAD-MEs show the lowest identity values when compared with other eukaryotic MEs with major differences including short amino acid insertions distributed throughout the primary sequence. Some residues in these exclusive segments are co-evolutionarily connected, suggesting that they could be important for enzymatic functionality. Phylogenetic analysis indicates that eukaryotes from different kingdoms used different strategies for acquiring the current set of NAD(P)-ME isoforms. In this sense, while the full gene family of vertebrates derives from the same ancestral gene, plant NADP-ME and NAD-ME isoforms have a distinct evolutionary history. Plant NADP-ME genes may have arisen from the  $\alpha$ -proteobacterial-like mitochondrial ancestor, a characteristic shared with major eukaryotic taxa. On the other hand, plant mNAD-ME genes were probably gained through an independent process involving the Archaeplastida ancestor. Finally, several residue signatures unique to all plant mNAD-MEs could be identified, some of which might be functionally connected to their exclusive biochemical properties. In light of these results, molecular evolutionary scenarios for these widely distributed enzymes in plants are discussed.

**Keywords:** phylogeny, malic enzyme, land plants, molecular evolution, NAD or NADP specificity

## INTRODUCTION

Malic enzymes (MEs) catalyze the oxidative decarboxylation of L-malate, producing pyruvate, CO<sub>2</sub>, and NAD(P)H in the presence of a divalent cation (Mg<sup>2+</sup> or Mn<sup>2+</sup>; Chang and Tong, 2003). MEs are part of a family of structurally related proteins (Malic Enzyme Family, MEF) that also includes malolactic enzymes (MLEs) and soluble oxaloacetate decarboxylases (OADs), which

convert L-malate to L-lactate, and oxaloacetate (OAA) to pyruvate, respectively (Espariz et al., 2011; **Supplementary Figure S1**). MEs can be classified considering different criteria. Based on coenzyme specificity and their ability to decarboxylate OAA, MEs are divided into three categories: EC1.1.1.38 (NAD-dependent; decarboxylates added OAA), EC1.1.1.39 (NAD-dependent; does not decarboxylate added OAA), and EC 1.1.1.40 (NADP-dependent; decarboxylates added OAA; International Union of Biochemistry and Molecular Biology<sup>1</sup>; **Supplementary Figure S1**). Based on the NAD(P) binding domain, MEs can be further divided into two groups: Class I<sup>2</sup> and Class II-MEs<sup>3</sup>. Class I groups MEs typically found in eukaryotes and in some eubacteria. This group also includes MLE of Gram-positive lactic acid bacteria (LAB). Otherwise, the Class II-MEs are almost exclusively constituted by archebacterial and eubacterial MEs, along with the OADs found in Gram-positive bacteria. In eukaryotes, Class II-MEs have been found to date only in the protist *Trichomonas vaginalis* and *Entamoeba histolytica* (Field et al., 2000; Dolezal et al., 2004). Class I and Class II-MEs group in two well-separated branches of a phylogenetic tree of MEF proteins (Espariz et al., 2011), indicating that that they do not share a common evolutionary history.

Malic enzyme is widely distributed in nature, being present in organisms of the major biological divisions. This is due to the extensive participation of its substrates and products in different metabolic pathways. The genome of higher animals (vertebrates) encodes a uniform set of three MEs, one cytosolic NADP-dependent enzyme (cNADP-ME, EC 1.1.1.40) and two mitochondrial isoforms: one NADP-dependent (mNADP-ME, EC 1.1.1.40) and another NAD-dependent (mNAD-ME, EC 1.1.1.38). Animal cNADP-ME and mNADP-ME provide NADPH for the biosynthesis of long-chain fatty acids or steroids (McKeehan and McKeehan, 1982; Goodridge et al., 1996), while mNAD-ME is only expressed in fast-growing and neoplastic tissues where it is implicated in energy metabolism (Moreadith and Lehninger, 1984). On the other hand, the genome of land plants (embryophytes) encodes a variable, species-dependent, ME set with several isoforms distributed among cytosol (cNADP-ME, EC 1.1.1.40), plastids (pLNADP-ME, EC 1.1.1.40), and mitochondria (mNAD-ME, EC 1.1.1.39). Plant cNADP-MEs have been linked to plant defense responses (Schaaf et al., 1995; Alvarez et al., 2013), lignin biosynthesis (Schaaf et al., 1995), cytosolic pH control (Martinoia and Rentsch, 1994), regulation of stomatal closure (Outlaw et al., 1981; Laporte et al., 2002), and photosynthetic metabolism in Crassulacean acid metabolism (CAM) plants (Cushman, 1992). In plastids, pLNADP-MEs are involved in the C<sub>4</sub> photosynthetic process operating in the C<sub>4</sub>-NADP-ME subtype (Drincovich et al., 2010; Saigo et al., 2013) as well as in defense responses (Maurino et al., 2001; Chi et al., 2004) and lipid synthesis (Lai et al., 2002). In plant mitochondria, mNAD-ME participates in the photosynthetic process of C<sub>4</sub>-NAD-ME subtype species, but its constitutive presence in all tissues of C<sub>3</sub> and C<sub>4</sub> species indicates a housekeeping role for

this isoform. In this regard, mNAD-ME could channel the large L-malate reserves commonly present in many plant species to the tricarboxylic acid (TCA) cycle for their complete oxidation (Tronconi et al., 2008) or participate in amino acid and lipid biosynthesis (Sweetlove et al., 2010).

Several biochemical and structural characteristics distinguish plant mNAD-MEs, e.g., the EC1.1.1.39 category is composed only of plant mNAD-MEs, as they are the only isoforms unable to decarboxylate OAA characterized to date. Interestingly, whereas the animal NAD(P)-MEs and plant NADP-MEs are assembled as homo-oligomers of monomers with molecular masses in the range between 60 and 66 kDa, plant mNAD-MEs can be structured as heteromers of two different subunits:  $\alpha$  (~60 kDa, or mNAD-ME1) and  $\beta$  (~58 kDa, or mNAD-ME2) proteins, and also as  $\alpha_2$  and  $\beta_2$  active homodimers (Willeford and Wedding, 1987; Tronconi et al., 2008, 2010b). In addition, these enzymes cannot catalyze the reverse reaction (reductive pyruvate carboxylation) as it has been described for animal NAD(P)-MEs and plant NADP-MEs (Mallick et al., 1991; Gerrard Wheeler et al., 2008; Tronconi et al., 2010a). Finally, the *Arabidopsis*  $\alpha_2$  mNAD-ME homodimer displays a complex allosteric inhibition by L-malate not described for other MEs characterized up to this point (Tronconi et al., 2012, 2015). These properties of plant mNAD-ME, not shared with the other MEs found in higher eukaryotes, are puzzling and could have their origin in a different evolutionary history for these isoforms.

This work addressed the evolutionary relationship of Class I-MEs found in eukaryotes, focusing on the plausible divergent origin of plant mNAD-MEs. Multiple primary sequence alignments and modeling based on three-dimensional structures were explored using bioinformatic approaches to analyze elements exclusively present in plant mNAD-ME. The phylogenetic analysis provides support for different evolutionary scenarios for NADP-ME and NAD-ME genes in plants.

## MATERIALS AND METHODS

### Retrieval of Sequences and Alignments

The sequences of MEs, MLEs, and OADs were recovered from the National Center for Biotechnology Information database<sup>4</sup>. For plants and algal species with entire genome information, MEs sequences were extracted from www.phytozome.net, while animal and fungal MEs were retrieved from www.genomesize.com and http://fungidb.org, respectively. BLASTP or TBLASN with a minimal e-value of 0.0001 was used to obtain homologous MEs in Chromista, Rhodophyta, Excavata, and Amoebozoa kingdoms with the human cNADP-ME (GI:EAW48664), human mNAD-ME (GI:EAW62981), and  $\alpha$  subunit mNAD-ME from *Arabidopsis thaliana* (AT2G13560) used as seeds.

*Cyanophora paradoxa* and *Bigelovia natans* ME sequences were retrieved from draft genomes (Curtis et al., 2012; Price et al., 2012) using TBLASN for searching. The partial nucleotide encoding sequences were translated and putative proteins were aligned by ClustalW for editing before evolutionary analyses.

<sup>1</sup><http://www.chem.qmul.ac.uk/iubmb>

<sup>2</sup><https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrvc.cgi?uid=133454>

<sup>3</sup><https://www.ncbi.nlm.nih.gov/Structure/cdd/cddsrvc.cgi?uid=133453>

<sup>4</sup><http://ncbi.nlm.nih.gov>

For gene identification in *Rhizopus oryzae* and *Mucor circinelloides* with a potential photosynthetic ancestry, *Phytophthora sojae* genes, assumed to have a photosynthetic endosymbiotic origin (Tyler et al., 2006), were used as query to search in the genome database of *M. circinelloides*<sup>5</sup>.

For amino acid comparisons and general editing, sequences were aligned in a ClustalW format. Accession numbers of protein sequences used in the present study are provided in **Supplementary Table S1**.

## Prediction of Subcellular Localization of Proteins

Protein localization was analyzed using MitoProt II (Claros and Vincens, 1996), TargetP 1.1 (Emanuelsson et al., 2007), and ChloroP 1.1 (Emanuelsson et al., 1999). Sequences were considered as co-localized when at least two out of the three programs predicted matching results.

## Mutual Information Analysis

Co-evolving amino acid pairs in plant mNAD-ME isoforms were identified by the MISTIC (mutual information server to infer coevolution) method (Simonetti et al., 2013). Mature isoforms (without mitochondrial targeting peptide) from angiosperms were aligned and loaded at <http://mistic.leloir.org.ar/index.php> in a FASTA format. *A. thaliana*  $\alpha$  subunit mNAD-ME (AT2G13560) was chosen as reference sequence. MISTIC calculated the Mutual Information (MI) between pairs of columns in the multiple alignments. Briefly, the frequency of appearance for each residue pair was compared with the individual observed frequency for each amino acid at the same positions (Buslje et al., 2009; Simonetti et al., 2013). Next, an MI score was calculated as the number of standard deviations that the observed MI value falls above the mean value obtained from a large set of permuted versions of the alignment for each pair position. MI scores higher than 6.5 provide sufficient specificity and sensitivity being the only ones reported (Buslje et al., 2009). Results were displayed in different formats (i) Circos representation for comparison through the full alignment of position pairs analyzed with MI, (ii) network graphs composed of nodes joined by edges, where nodes represent residues and edges between two nodes indicate an MI value  $> 6.5$ , and (iii) logo representation showing the amino acid enrichment and depletion for a group of MI connected nodes.

## Molecular Evolutionary Analyses

Multiple sequence alignments of amino acid sequences for evolutionary inferences were constructed with MUSCLE 3.2 version (Multiple Sequence Comparison by Log-Expectation; Edgar, 2004) implemented in MEGA6 software (Tamura et al., 2013) and manually edited for missing gaps or poorly aligned regions. The final ME alignment consisted of 142 sequences and 544 characters. Phylogenetic trees were inferred by the maximum likelihood (ML) and Bayesian inference (BI) methods via the MEGA6 (Tamura et al., 2013) and MrBayes 3.1.2

software (Ronquist and Huelsenbeck, 2003), respectively. In ML analysis, data samples were first analyzed to assess the best substitution model describing the observed sequence change. 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. For our data set, the lowest BIC value corresponded to the Whelan and Goldman model (WAG; Whelan and Goldman, 2001) using a discrete gamma distribution with five categories ([+G],  $\alpha$  parameter = 2.0533) under the assumption that a certain fraction of sites are evolutionarily invariable ([+I], 2.6198% sites). Hence, this constituted the evolutionary model: WAG + G (2.0533) + I (2.6198). For ambiguous states and insertion-deletions, we chose a partial-deletion treatment with a site coverage parameter at 95%. The initial tree for the ML search was generated automatically by applying 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. The final tree was the one whose arrangement of branches led to the highest ML value. Reliability of interior branches was assessed with 500 bootstrap re-samplings (MLB). Nodes with MLB values 50–69% were regarded as weakly supported, 70–84% as moderately supported, and 85–100% as strongly supported (Hillis and Bull, 1993). Phylogenetic trees were displayed using the MEGA software version 6.0 (Tamura et al., 2013). Substitutions per site are denoted as a separated bar with a number in each phylogenetic tree.

In the BI analysis, two parallel runs, each including four Metropolis-coupled Markov chain Monte Carlo (MC3) analysis, were run for 5,000,000 generations and sampled every 100 generations for each partition scheme. This generated an output of 50,000 trees. 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 0.08. In order to confirm that the chains achieved stationary, “burn-in” plots were evaluated by plotting log-likelihood against generation number. Also, the final average standard deviation of split frequencies was used as the convergence index, in which values  $< 0.01$  indicated good convergence, and convergence of clade posterior probabilities within and between runs was checked using the potential scale reduction factor (PSRF). After determining convergence, the initial 25% of the sampled trees for each MC3 run were discarded as “burn-in,” and the post-burn-in trees were used 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 clade. Node with BPP values  $\geq 95\%$  were considered highly supported and  $\leq 95\%$  not supported. The tree figures were displayed using the FigTree v1.4 software<sup>6</sup> and visually edited. Substitutions per site are denoted as a separated bar with a number in each phylogenetic tree.

<sup>5</sup><http://genome.jgi-psf.org/Mucci2/Mucci2>

<sup>6</sup><http://tree.bio.ed.ac.uk/software/figtree>

## Identification of Signature Residues

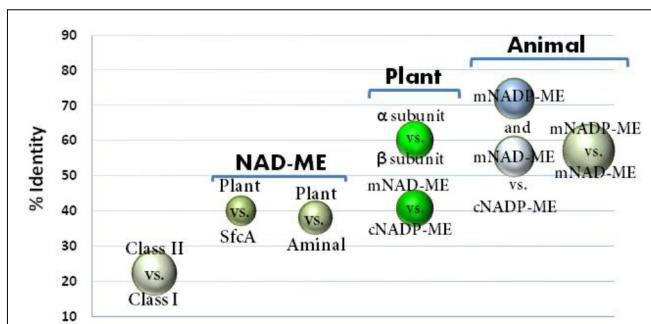
Identification of well-conserved positions in the primary sequences that differ in other groups (SDP, specificity-determining positions) within a group of proteins was performed using SDPpred<sup>7</sup>. ME sequences belonging to A+V and G+R clades (see below) were aligned in a GDE (ClustalW) format before uploading. For each aligned column, the software computed a z-score for positions where the amino acid distribution is more closely associated with grouping than for an average position of the alignment, which is thus likely to be an SDP (Kalinina et al., 2004). Next, SDPpred evaluated the significance of the z-scores in order to decide if it is high enough to indicate an SDP. For this, SDPpred established a z-score threshold “*p*” (or B-cutoff) based on the computation of the Bernoulli estimator. Positions with z-score > *p* were designated as SDPs, as they were non-randomly generated (Kalinina et al., 2004).

## RESULTS

### Amino Acid Sequence Analysis of Animal and Plant MEs

Animal mNADP-ME and mNAD-ME share about 55% identity with each other and they are 72% and 55% identical to animal cNADP-ME, respectively (Figure 1). In plants,  $\alpha$  and  $\beta$  mNAD-MEs are about 63% identical and share only 40% identity with cytosolic or plastidic counterparts (Figure 1). When animal and plant mNAD-MEs are compared, the proteins are nearly 38% identical. These relatively low identity values of plant mNAD-ME with isoforms from higher eukaryotes are similar to those found with the SfcA ME of *Escherichia coli* (Figure 1). Finally, Class I- and Class II-MEs, represented by the sequence from the archeon *Solfolobus solfataricus*, exhibit very low identity (about 20%) with each other (Figure 1). In this top-down analysis of sequences, plant mNAD-ME displays the five conserved sites previously defined for the ME family (Drincovich et al., 2001), including the two phosphate cofactor-binding signature motifs: site II and site V (Figure 2; Chang and Tong, 2003). A multiple protein sequence alignment of animal and plant NAD(P)-MEs reveals that the main differences in plant mNAD-ME lie in short amino acid insertions (indicated as A–F in Figure 2), with the amino (A) and carboxyl terminal (F) insertions being the most significant in extension. Two shorter insertions (B and C) are located near the second dinucleotide-binding signature motif that is part of the cofactor binding site. In addition, some single-amino acid substitutions in this region can be highlighted. For instance, the GAGEAA/G consensus sequence in animal NAD(P)-ME and plant NADP-ME changes to GAGSAG in plant mNAD-ME (Figure 2). Also, amino acid residues implicated in cofactor specificity are different in this enzyme. In NADP-ME, a Lys residue (K362 for human cNADP-ME) is the major determinant for NADP specificity (Hsieh et al., 2011). This amino acid is not present in animal or plant mNAD-ME; it has been substituted by Asn in animal mNAD-ME and by Ala or Val in plant mNAD-ME

<sup>7</sup><http://bioinf.fbb.msu.ru/SDPpred>

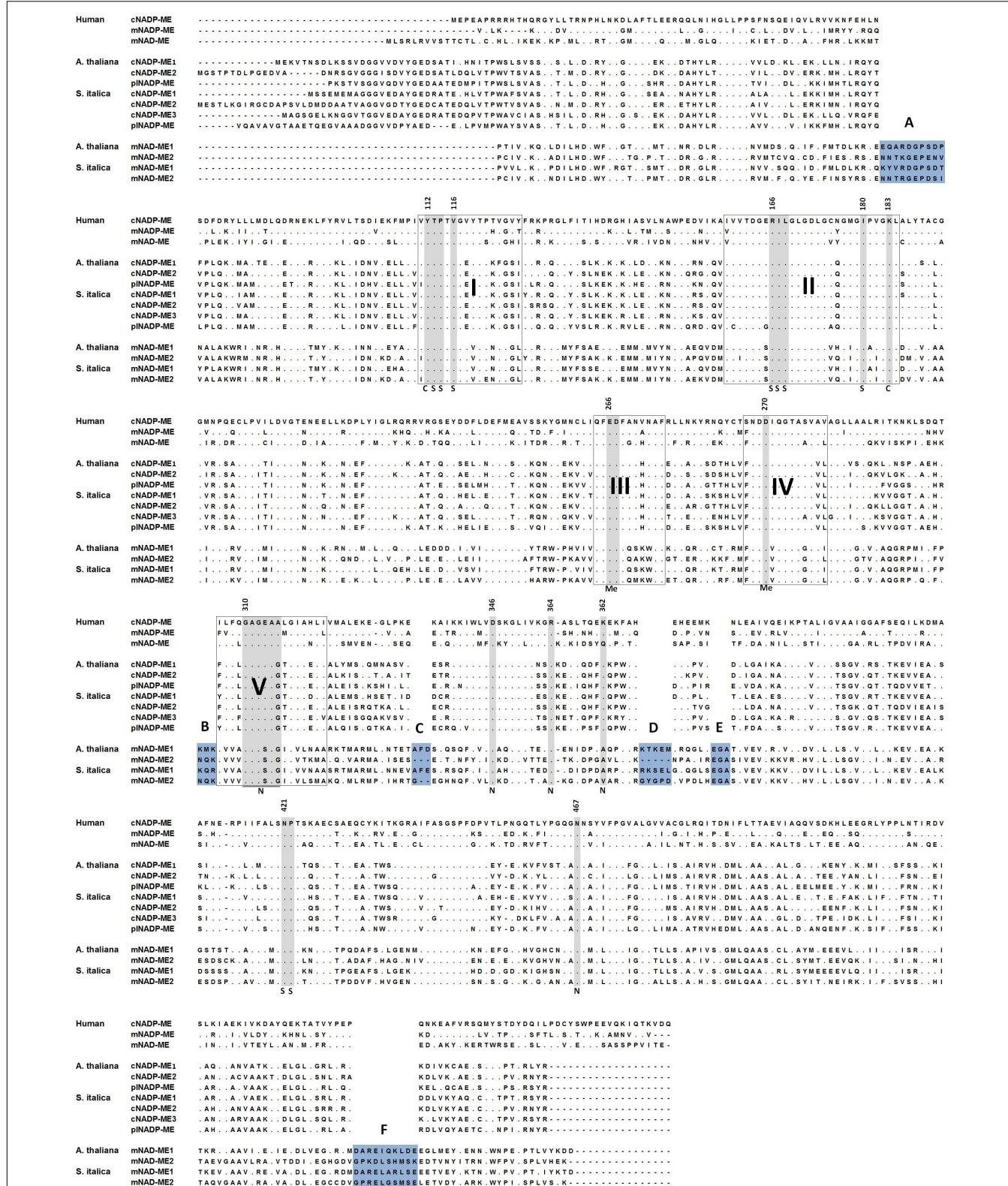


**FIGURE 1 |** Sequence identity among ME isoforms comparing class I- versus class II-MEs; animal MEs; and relevant group isoforms from plant MEs and NAD-MEs. Circles indicate the identity values (%) obtained by ClustalW alignment among the different ME isoforms indicated in the graph. Circle size is proportional to the standard dispersion. For each comparison, 12 animal MEs and 20 plant ME sequences were used. For class I versus II comparison, ME sequences for 10 eubacteria, MaeB from *E. coli*, YtsJ from *Bacillus subtilis* (NC\_000964.3), and the accessed protein WP\_029552528 from *Solfolobus solfataricus* as representative sequences of Class II-ME were also included. The Class I-MEs used for calculations are denoted with (\*) in **Supplementary Table S1**. For clarity, comparison between plant cytosolic and plastidic NADP-ME was not included, which was centered on 85% identity.

(Figure 2). Despite these differences, the residues that have been implicated in substrate binding and catalysis in human mNAD-ME are conserved in plant mitochondrial isoforms (Figure 2; Chang and Tong, 2003).

### Co-evolutionary Residue Positions

In view of the significant changes at the primary sequence level of plant mNAD-ME, we evaluated the extension of the co-evolutionary relationships between amino acid positions through the MI principle. MI assumes that mutations of essential residues may occur only if a compensatory mutation elsewhere in the protein also takes place to preserve activity (Martin et al., 2005). Pairs of residues in a sequence alignment exhibiting high MI score values are postulated to have coevolved and thus are probably functionally connected. Since over 100 aligned sequences are required for an accurate MI prediction and such a number of plant mNAD-MEs are not currently available, our results are only an approximation to the inference of the co-evolutionary scenario. Data sets of 78  $\alpha$  and  $\beta$  plant mNAD-ME mature sequences (590 residues long) were fully aligned and MI calculation was performed by MISTIC (Simonetti et al., 2013). High-scoring MI pairs were accumulated in three main regions of  $\alpha$  and  $\beta$  plant mNAD-ME: 53–71, 317–376, and 547–583 (**Supplementary Figure S2A**). Individual residues with a high number of MI connections can be found within these regions. Interestingly, many of these residues are part of the A, C, D, and F insertion sequences, which are exclusive to plant mNAD-ME (**Supplementary Figure S2A**). For example, residues in the A and F regions, which are far apart in the primary sequence, were highly MI interconnected with each other and with some residues of insertions C and D (**Supplementary Figure S2B**). This suggests that these unique sequence segments could be relevant for plant mNAD-ME



**FIGURE 2 |** Amino acid sequence alignment of representative NADP-MEs and NADP-MEs. NADP-ME and NAD-ME sequences of human, *A. thaliana*, and *S. italica* as representative of animal, eudicot plants, and monocot plants, respectively, are shown. Conserved sequences found in all NAD(P)-ME (sites I–V) are highlighted. Insertions present in plant mNAD-MEs are highlighted in blue. Amino acid residues implicated in catalysis (C) as well as in metal (Me), L-malate (S), or cofactor (N) binding are highlighted in gray. Numbers for each amino acid residue refer to their position in human mNAD-ME primary structure.

functionality. Additionally, we analyzed pairs of  $\alpha$  and  $\beta$  sequences fused in an amino to carboxyl arrangement in both  $\alpha$ - $\beta$  and  $\beta$ - $\alpha$  combinations. A common number of MI connections involving residues belonging to the two subunits were identified for both combinations (**Supplementary Figure S2C**).

## Gene Structure of Animal and Plant MEs

The structure of vertebrate and plant *ME* genes was analyzed considering the number and homology of encoding exons. *ME* genes of four animal species that belong to different taxonomic classes were analyzed: *Homo sapiens* (Mammalia), *Xenopus laevis* (Amphibia), *Danio rerio* (Actinopterygii), and *Columbus livia* (Aves). The structure of *NAD(P)*-*ME* genes is well conserved in the four animal species, with *cNADP-ME* and *mNADP-ME* genes containing 14 encoding exons, and *mNAD-ME* 15 exons (**Supplementary Figure S3A**). In each species, all exons of the three *ME* genes are homologous and the extra exon present in *mNAD-ME* arises from the split of exon number 12 of *cNADP-ME* or *mNADP-ME*. Thus, all the animal *ME* genes seem to be evolutionarily related.

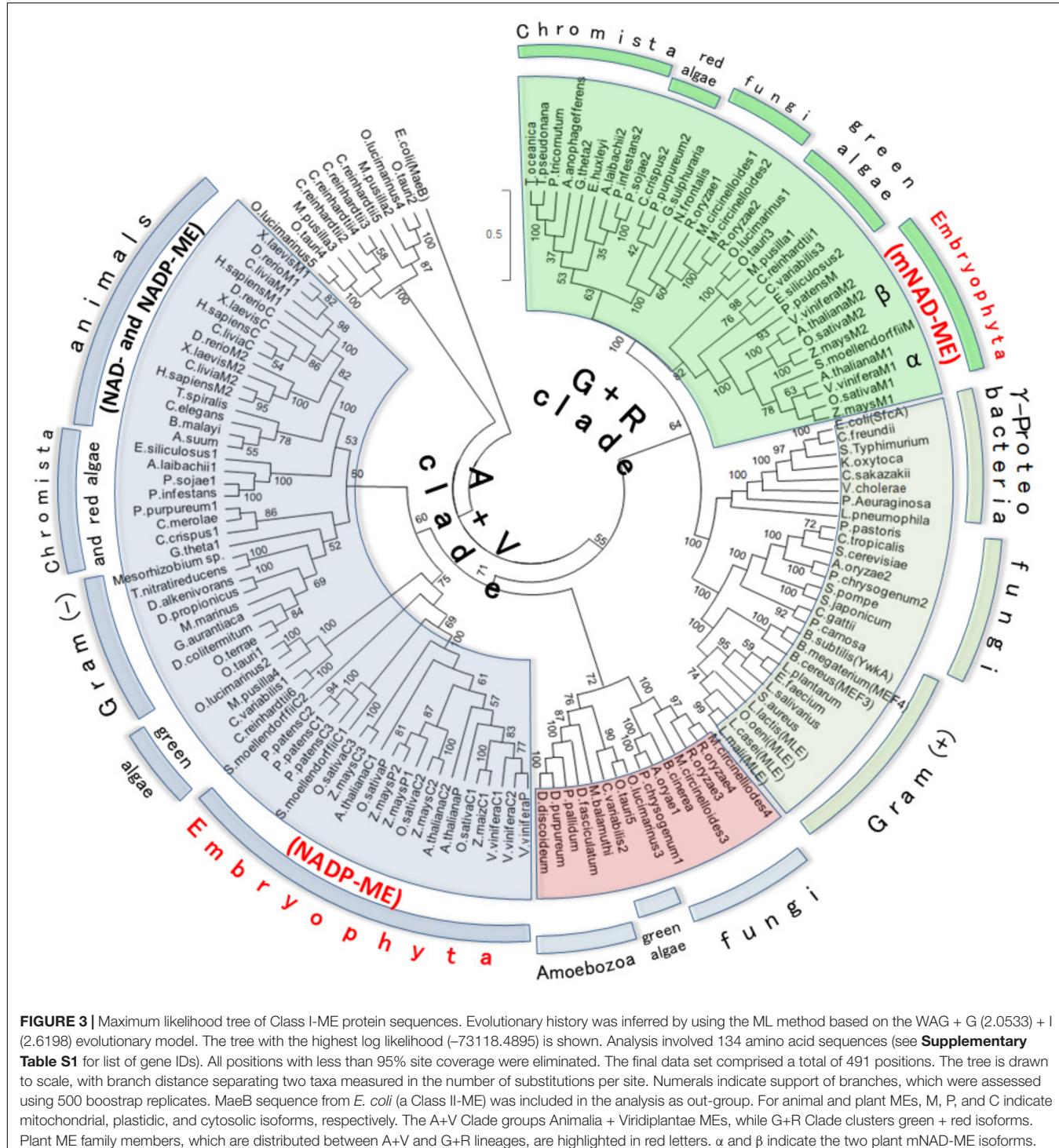
Among plants, we analyzed the complete *ME* gene set in the bryophyte *Physcomitrella patens* and three tracheophytes species: *Selaginella moellendorffii* (Lycopodiophyta), *A. thaliana* (Angiospermae, eudicot), and *Zea mays* (Angiospermae, monocot), which are distant species and show different sets of *ME*. In general, plant *MEs* exhibit higher numbers of encoding exons than the animal genes with some degree of species-dependence (**Supplementary Figure S3B**). In this regard, both *NADP-ME* and *NAD-ME* genes present 19 encoding exons, except *cNADP-ME2* of *Z. mays* with eight, since exons 12–19 are fused. Although there is some variability in the number and structure of exons between animal and plant *NAD(P)*-*ME* genes, two conclusions may be drawn from gene conformation analysis. First, some degree of colinearity between the exons of plant *NADP-ME* genes and those from animals can be noted. Second, each encoding exons from a particular *cNADP-ME* or *pLNADP-ME* gene in plants has its counterpart in other *cNADP-ME* or *pLNADP-ME*, but not in *mNAD-ME* genes, where their exons only show homology to each other (**Supplementary Figure S3B**).

## Phylogeny Pattern of Plant NADP and NAD-ME

The above results suggest that plant *mNAD-MEs* could be evolutionarily unrelated to other eukaryotic *MEs*. Thus, to address the evolutionary relationship of plant *mNAD-ME*, we performed phylogenetic analyses using Class I-*MEs* amino acid sequences from organisms that belong to the major taxonomic divisions: Archaeplastida (Glaucophyta, Rhodophyta [red algae], and Viridiplantae [green algae and Embryophyta]), Opisthokont (Animalia and Fungi), Chromista (Crytophyta, Heterokonta, and Haptophyta), Rhizaria (Chlorarachniophyta), Amoebozoa, and Eubacteria (Gram-negative and Gram-positive bacteria). Archeal *ME* and eubacterial OAD sequences were excluded from our analyses as they are members of the highly divergent Class II-*ME* group. Hence, data sets of 134

eubacterial and eukaryotic *MEs* were retrieved from databases, aligned and, after exclusion of ambiguously aligned regions, 491 amino acid positions were included in ML and BI phylogenetic analyses (see **Supplementary Table S1** for list of gene ID's). When two sequences showed >85% of similarity, one of them was excluded from the alignment. Eukaryotic Class I-*MEs* grouped in two major clades in the ML tree (**Figure 3**). One clade contains all animals *NAD(P)*-*MEs* along with *cNADP-MEs* and *pLNADP-MEs* from plants. This group also contains some algae and chromist sequences and thus constitutes the “Animalia + Viridiplantae clade” (A+V clade). Homologous sequences from the paraphyletic Gram-negative group, comprising Verrucomicrobia (*Opitutus terrae*) and  $\alpha$ -Proteobacteria (*Mesorizobium* sp.), are also included in the A+V clade. Amoeba sequences branch as a subgroup along with fungal and algae enzymes in this clade. On the other hand, plant *mNAD-ME* conform the second major eukaryotic clade together with sequences of green and red algae and chromists (**Figure 3**). We named this grouping the “Green + Red clade” (G+R clade) because in addition to embryophytes, it comprises green and red algae, as well as organisms that are evolutionarily related with them through secondary endosymbiosis. The G+R clade contains a sister group of *MEs* from Gram-negative and Gram-positive bacteria, and some fungal sequences. Interestingly, all Gram-negative sequences in this group belong to  $\gamma$ -proteobacteria class. MLEs are also included in this subclade, as they nest within Gram-positive *MEs* (**Figure 3**). Finally, some green algae *ME* isoforms constitute a cluster separated from the remaining sequences by a very long branch. The analysis of these sequences indicates several mutations at positions of the conserved motifs that characterize the *ME* family (not shown), suggesting that they could be a case of neofunctionalization in the MEF. The BI tree topology was highly congruent with those deduced by ML analysis (**Supplementary Figure S4**) with branches well resolved and supported ( $BPP \geq 95\%$ ). In particular, branches grouping the deep lineages of *ME*, which were weakly/moderately sustained in ML approach, were strongly supported in the BI analysis. When comparing ML and BI trees, two mayor differences can be found: amoeba, fungi, and algae sequences form a third group of eukaryotic *ME*, separated from the A+V clade in the BI tree and Gram-negative sequences do not nest with animal isoforms, constituting a simple cluster within the A+V clade (**Supplementary Figure S4**). It should be mentioned that the highly divergent algae *ME* sequences were omitted in the BI analysis.

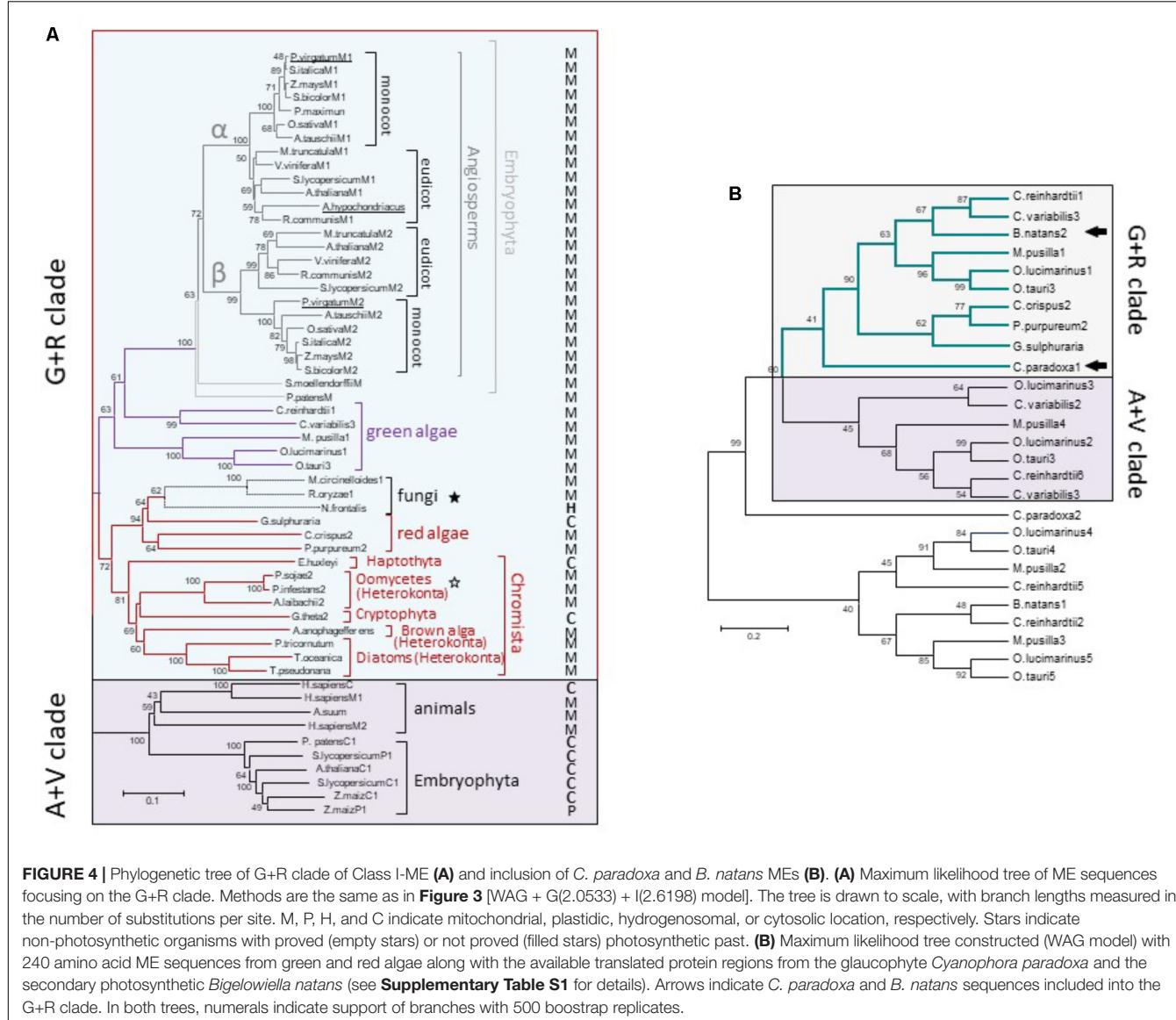
Although the monophyly of the G+R clade is consistently recovered and strongly supported by ML analysis, internal branches corresponding to some chromist sequences were moderately supported (**Figure 3**). This could be due to the existence of long branches in the extensive phylogenetic analyses that destabilize the otherwise robust grouping (Swofford et al., 1996). Longer branches were indentified for green algae *MEs* belonging to the A+V clade and for Gram-positive MLEs (data not shown). Thus, a new ML tree mainly focused on G+R sequences, including a larger number of *mNAD-ME* sequences from C<sub>3</sub> and C<sub>4</sub>-NAD-ME subtype plants, was constructed. In



the absence of long branches, the MLB support for chromists increased significantly (**Figure 4A**).

For  $\alpha$  and  $\beta$  mNAD-ME subunits of Angiosperms, monocot and eudicot isoforms are grouped separately in the G+R clade (**Figure 4A**). Unlike Angiosperms, the moss *P. patens* and the ancient tracheophyte *S. moellendorffii* possess a single mNAD-ME isoform, which constitutes the most immediate

out-group from the monocot/eudicot grouping and completes the Embryophyta cluster in the phylogenetic tree. Some green algae sequences nest closely to embryophytes. Here, the isoforms of both freshwater green algae (*Chlamydomonas reinhardtii* and *Chlorella variabilis*) and planktonic green algae (*Ostreococcus* spp. and *Micromonas pusilla*) were predicted to be located in the mitochondria (**Figure 4A**). Red algae and



chromist sequences constitute a closed and highly supported subgroup in the G+R clade, where mitochondrial allocation was predicted for almost all MEs except for the red alga *Galdieria sulphuraria*, the haptophyte *Emiliania huxleyi*, and the cryptophyte *Guillardia theta* (Figure 4A). Interestingly, a reduced fungal clade composed by *Neocallimastix frontalis*, *R. oryzae*, and *M. circinelloides* was found to nest within red algae sequences with very strong bootstrap support. MEs from *R. oryzae* and *M. circinelloides* are possibly located in the mitochondria while the *N. frontalis* isoform would be targeted to the hydrogenosome, a mitochondria-related compartment present in this anaerobic fungus (van der Giezen et al., 1997). Along with Oomicetes, these fungi constitute a group of non-photosynthetic organisms in the G+R clade (Figure 4A). But, unlike Oomicetes, they lack a proved endosymbiotic evolutionary relationship with photosynthetic organisms. Even so, the position of *R. oryzae* and *M. circinelloides* MEs in the red algae sub-cluster is supported

by the finding that other genes in these species are more related to homologs found in red algae and cyanobacteria than in fungi (Supplementary Figure S5).

Finally, we tracked a G+R-type ME in an ancient Archaeplastida member, the glaucophyte *Cyanophora paradoxa*, as well as in the chlorarachniophyte *B. natans*, a secondary (engulfed green alga) photosynthetic alga. For these species, the ME sequences are not complete and thus, a new tree using 240 aligned amino acid positions along with those corresponding to red and green algae homologs was constructed. Both *C. paradoxa* and *B. natans* possess at least one ME isoform phylogenetically related to the algal proteins that are members of the G+R clade (Figure 4B).

## Signature Residues of the G+R Clade

Based on the aforementioned evolutionary relationship, MEs found in higher eukaryotes are included in two groups, the

A+V clade and the G+R clade. We used SDPpred (specificity determining position prediction; Kalinina et al., 2004) to search for residue positions that are well conserved within these two groups, but differ between them, in order to assign a signature amino acid sequence to each clade. Two different SDP predictions were performed: one considering all eukaryote A+V and G+R sequences and another one restricted only to the embryophyte G+R grouping (plant mNAD-MEs) with all eukaryote A+V MEs. The 13 SDPs exhibiting the highest z-scores considering both analyses are shown in **Figure 5A**. Twelve SDPs showed an invariant amino acid residue for MEs that belong to the G+R clade but only five unambiguous positions were predicted for A+V sequences. Using the tridimensional structure of human mNAD-ME (Yang et al., 2000), we observed that the six SDPs with the highest zcores are located near the active site, as the farthest SDP was located at 8 Å from the metallic center (**Figure 5B**). In this regard, using the  $\alpha$  mNAD-ME from *A. thaliana* as a reference sequence, we found that residues S157 and W254 conserved in G+R sequences change to E164 and N261, respectively, in the A+V grouping. These residues are located at the active site in human mNAD-ME (**Figure 5B**). Particularly, S157 is located in the first cofactor-binding signature motif (Site II in **Figure 2**). Also, residues M126, W239, and P240 of the G+R group are almost invariably replaced by L133, Y246, and G247 in the A+V sequences, respectively, and they are part of domain B where the active site is located (**Figure 5B**). In particular, when residue P240 of the G+R group is replaced by G, this site is predicted to be an SDP with the highest confidence

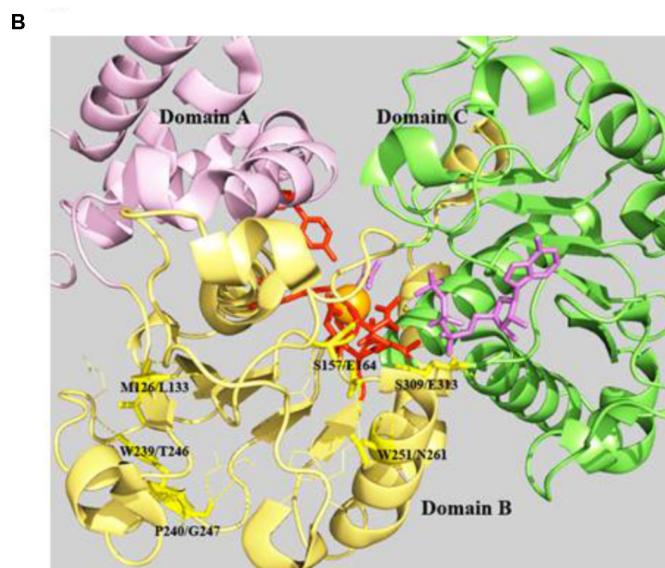
(*p*-value = 0.000; **Figure 5A**). Finally, S309 in the G+R group is replaced by G313 in the A+V grouping, a highly conserved residue located in the second dinucleotide-binding signature motif (Site V in **Figure 2**).

## DISCUSSION

### Molecular Evolution of Plant mNAD-MEs

Plant mNAD-MEs show the lowest identity (38–40%, **Figure 1**) and similarity (57–60%, not shown) values when compared with either plant NADP-MEs or NAD(P)-MEs from distantly related organisms (animal or *Escherichia coli*). These numbers represent the minimum homology values shared among sequences that belong to Class I-MEs. In this regard, when Class I- and Class II-MEs are compared, the identities are roughly 20% (**Figure 1**), and conserved residues are only those implicated in catalysis (not shown). Despite having substantial sequence diversity, MEs share a similar overall tertiary architecture (Chang and Tong, 2003). In fact, the NAD(P)-binding Rossmann fold domain that is part of the MEF structure has been denoted as an extreme case where folding is non-sequence driven (Bhattacharyya et al., 2012). Thus, an important flexibility in sequence seems to exist in the MEF and, as an example of this property, plant mNAD-MEs posses several short amino acid residue insertions distributed throughout the primary structure (**Figure 2**). The presence of such segments can be traced to the lower plants *P. patens* and *S. moellendorffii* as well as to green algae, but

G+R clade	Position in sequence <sup>1</sup>	A+V clade	Position in Sequence <sup>2</sup>	Bernoulli estimator
T(S)*	21	L/M	38	-76.9040
W*	75	Y	82	-58.8242
H*	82	Q(E)	89	-106.1873
A	101	L/M	108	-83.6826
<b>M</b>	<b>126</b>	<b>L(V)</b>	<b>133</b>	<b>-58.5019</b>
<b>S</b>	<b>157</b>	<b>E(G**)</b>	<b>164</b>	<b>-30.2315</b>
W	239	Y(F)	246	-72.5909
P	240	G	247	0.0000
W*	254	N(T)	261	-54.0563
<b>S</b>	<b>309</b>	<b>E(A,Q)</b>	<b>313</b>	<b>-44.4222</b>
F	340	I/V	341	-88.8466
F*	439	Y	435	-84.1795
T*	528	S	523	-104.6614



**FIGURE 5 |** Specificity-determining position in MEs of the A+V and R+G clades. **(A)** <sup>1</sup>Amino acid positions in the primary sequence of  $\alpha$  mNAD-ME from *A. thaliana*. <sup>2</sup>Amino acid positions in human mNAD-ME. Residues between parentheses show a frequency of occurrence lower than 10%. Residues on the right side of bars show a frequency of occurrence higher than 10%. \*Corresponds to SDPs identified only when embryophyte G+R sequences were compared with all A+V groupings. \*\*Corresponds to one residue found only in plastidic NADP-ME from *Setaria italica*. Letters in bold indicate SDPs with the highest z-scores that were defined as signature residues. **(B)** The SDPs with the highest scores are mapped onto the three-dimensional structure of the human mNAD-ME complex with the L-malate analog (oxalate, in magenta), NAD<sup>+</sup> (in magenta), Mn<sup>2+</sup> (in orange), and fumarate (not shown; 1GZ3, Yang et al., 2002). Signature residues are shown as stick yellow lines. Residues involved in L-malate binding and catalysis are shown as stick red lines. Residues contacting signature residues are shown as fine lines. Domains A, B, and C are shown in different colors. For clarity, whole Domain D and the  $\alpha$ B3-helix of Domain B were omitted in the model.

not to red algae or chromists (not shown), suggesting that they were gained after the split of the green and red algal lineages. Co-evolution analysis of residue pairs indicated that these segments exclusive to plant mNAD-ME could be important for enzymatic functionality, since they are located in MI networks (**Supplementary Figure S2**). Crucially, as a common number of MI, connections involve residues belonging to different subunits (**Supplementary Figure S2C**), these inter-subunit co-evolutionary pairs could be the basis of the exclusive heteromeric assembly that characterizes plant mNAD-MEs. Although these regions are far away from each other in the primary structure, co-evolving insertions might be close in folding intermediates or might be part of a transmission signal pathway connecting allosteric and active sites or subunit interfaces (Goodey and Benkovic, 2008; Marino Buslje et al., 2010). In any case, it is clear that a requirement to maintain a particular function constraints structural diversity in these exclusive regions of plant mNAD-ME.

## NADP-ME and NAD-ME Genes Were Independently Acquired in Plants

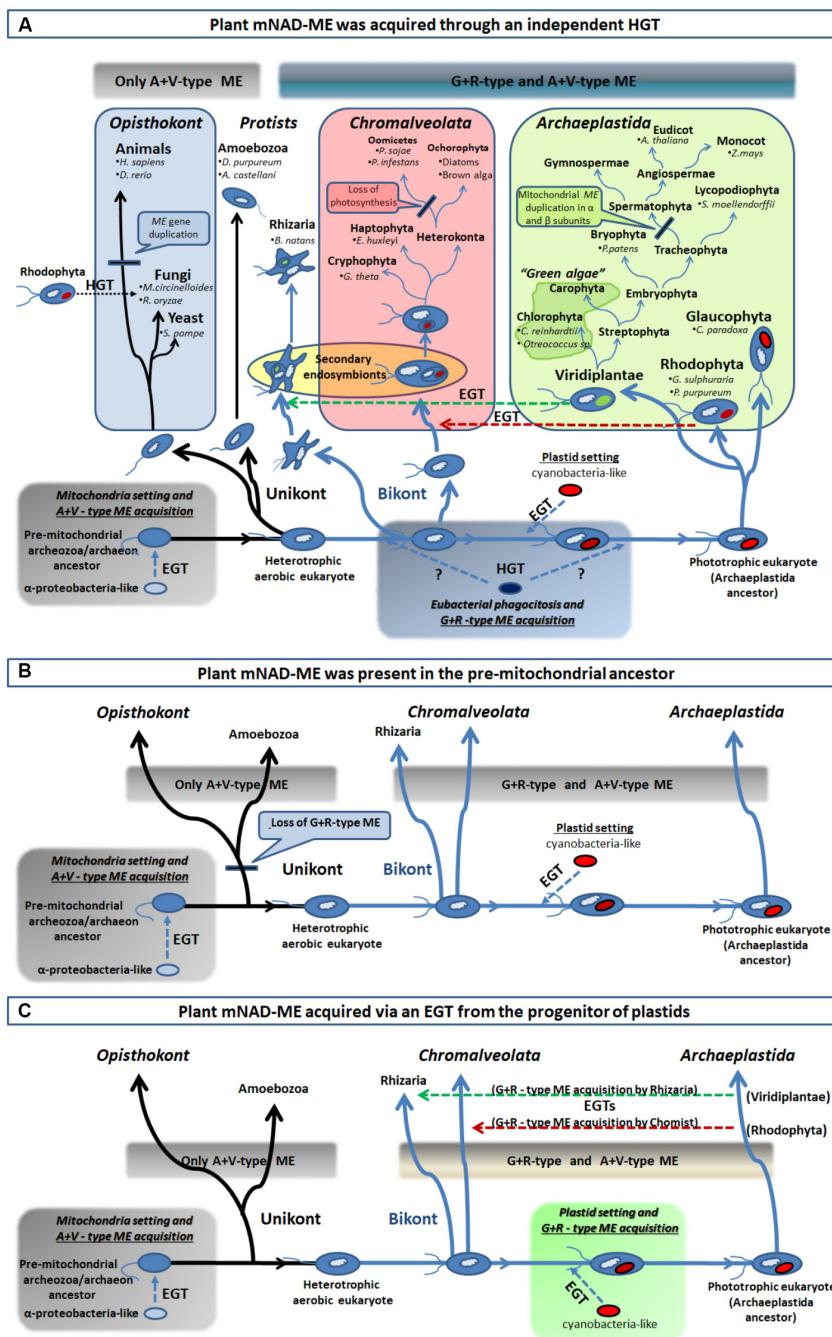
A better understanding of the particular molecular evolution of plant mNAD-MEs, on which their exclusive biochemical properties rely can be achieved through the comprehensive phylogenetic approach performed in this work. ML and BI analyses indicate that Class I-MEs from eukaryotic organisms are separated in two principal branches (**Figure 3** and **Supplementary Figure S4**). The A+V cluster and its immediate out-group include sequences of all major taxa, from Amoebozoa to Embryophyta along with ME from Gram-negative bacteria (**Figure 3**). Because of their presence in unrelated organisms and the inclusion of  $\alpha$ -proteobacteria sequences into the clade, we conclude that eukaryotes acquired A+V isoforms from the  $\alpha$ -proteobacterial-like ancestor of mitochondria through the extensive endosymbiotic gene transfer (EGT) that took place when this organelle was established (Keeling and Palmer, 2008; **Figure 6**).

On the other hand, the G+R clade is almost solely constituted by primary and secondary photosynthetic organisms and species displaying photosynthetic metabolism in the past (**Figures 3, 4A**). Since the presence of G+R isoforms can be traced to the primitive photosynthetic endosymbiont, the glaucophyte *C. paradoxa* (**Figure 4B**), and considering that most of them are predicted to have a targeting peptide (**Figure 4A**), the incorporation of this isoform in mitochondria must have been an early event during evolution of the Archaeplastida lineage. Unlike the A+V clade, eubacterial MEs do not nest within the G+R group, but  $\gamma$ -proteobacteria sequences branch close to it with moderate (MLB = 64%) and strong (BPP = 98%) support (**Figure 3** and **Supplementary Figure S4**). The  $\alpha$ -proteobacterial MEs were not closely related to the  $\gamma$ -proteobacteria clade, which indicates that it is unlikely that G+R type MEs were gained via mitochondrial endosymbiosis. Also, except for a very small group of related fungi (**Figures 3, 4A**), the lack of G+R MEs in other non-photosynthetic eukaryotes lineages further reduces the likelihood of a mitochondrial origin. Given these

data at hand, one possible scenario would be that G+R-type MEs were established in mitochondria of photosynthetic organisms after unikonts and bikonts split via an independent horizontal gene transfer (HGT), possibly from a  $\gamma$ -proteobacteria-like to the Archaeplastida ancestor genome (**Figure 6A**). At this point, it is possible that the presence of a mitochondrial G+R-type ME was a precondition for plastid establishment or, alternatively, it was a subsequent requirement for the success of the emerging photosynthetic eukaryote (**Figure 6A**). Thus, if the G+R-type ME was implanted in the Archaeplastida ancestor through an HGT event, the presence of this kind of ME isoform in mitochondria of secondary photosynthetic organisms, chromalveolates and rhizarians, should be the result of the implantation of a red or green alga, respectively, in two different heterotrophic ancestors (Qiu et al., 2013; **Figure 6A**).

The acquisition of enzymes by the Archaeplastida ancestor through HGT has been proposed as a common phenomenon in this single-cell phagotrophic organism that allowed the rapid spreading of the lineage (Keeling and Palmer, 2008; Ball et al., 2013; Qiu et al., 2013), resulting in a more rapid change in biochemical diversity and metabolic versatility than evolution of isoenzymes by gene duplication and subsequent neofunctionalization (Ghoshroy et al., 2010). A  $\gamma$ -proteobacterial origin was proposed for a subgroup of Class II plant glutamine synthetase (GSII<sub>B</sub>) and phosphoenolpyruvate carboxylase (PEPC) genes (Sanchez and Cejudo, 2003; Ghoshroy et al., 2010). For GSII<sub>B</sub>, the timing of the HGT event was estimated when establishing the Archaeplastida lineage (Ghoshroy et al., 2010). Hence, it is possible that the three genes could have the same evolutionary history in plants.

It is worth to mention that this is not the only possibility for the origin of mitochondrial G+R-type ME, as our phylogenetic results can also be explained by at least other two, albeit less parsimonious, scenarios (**Figures 6B,C**). Without considering the model for the evolution of eukaryotes we adopt, “the archezoa hypothesis” or “the symbiogenesis hypothesis” (Gray, 2012; Archibald, 2015), one other possibility would be that the G+R-type ME could constitute the original isoform present in the pre-mitochondrial ancestor (**Figure 6B**). Because almost all the extant non-photosynthetic organisms lack the G+R ME (**Figure 3**), this isoform must have been lost very early in eukaryotic evolution, probably involving the Unikot lineage (last eukaryote common ancestor of Opisthokont and Amoebozoa) that only kept the A+V ME derived from the EGT. Contrary, G+R ME was conserved in the Bikont lineage (an early ancestor of Archaeplastida), and then, their photosynthetic descendants present both G+R and A+V MEs (**Figure 6B**). However, the observation that none of the current Archaea species (the prokaryotes most closely related to eukaryotes) exhibit Class I-ME reduces the likelihood of this scenario. The second possibility for the presence of mitochondrial G+R-type ME in all photosynthetic organisms connects with the chloroplast evolution; this isoform could have been gained through EGT between the cyanobacterial progenitor of plastids and the heterotrophic host (**Figure 6C**). However, the genome analysis of extant cyanobacteria has not uncovered Class I-MEs



**FIGURE 6 |** Different scenarios of acquisition of the R+G clade of Class I-ME in eukaryotes. **(A)** From bottom left, an extensive EGT from an engulfed  $\alpha$ -proteobacteria-like ancestor (mitochondria precursor) to the nucleus of a unikont (one flagella) host took place. This resulted in the establishment of eubacterial ME version in early eukaryotes and the foundation of the A+V clade of ME. After unikont and bikont (two flagella) splitting and prior – or post – plastid setting (indicated by question marks), an HGT event from one eubacteria (possibly a  $\gamma$ -proteobacteria-like) into the Archaeplastida ancestor established the G+R clade of Class I-MEs in eukaryotes. This allowed spreading of G+R isoforms in all primary photosynthetic organisms, as well as, in the secondary phototrophs (Chromalveolata and Rhizaria) via an EGT (green and red horizontal broken line arrows) and in some fungi by an HGT event (black horizontal dotted line arrow). **(B)** The G+R ME could be the ancestral isoform that was present in the pre-mitochondrial, and founder of the eukaryotic lineage, ancestor (the “Archezoa,” or an autotrophic methanogenic archaeon, depending on the model for the evolution of eukaryotes). The establishment of mitochondria in this ancestor allowed the acquisition of an additional isoform (the type A+V) and this chimeric structure for the MEF was maintained by the bikont lineage. However, only the A+V-type ME was retained by the unikont ancestor, as it lost the more ancestral isoform in what could have been a genetic replacement. **(C)** The G+R ME acquisition by photosynthetic organisms (containing the A+V ME derived from the mitochondria setting) may have taken place in with the plastid genesis, through the extensive EGT from the engulfed cyanobacteria to the host genome. Hence, the presence of this isoform in secondary phototrophs organisms results from secondary endosymbiosis processes, as is explained in **A**. For clarity, the major taxa of eukaryotic organisms were omitted in **B** and **C**.

sequences (Espariz et al., 2011), which argues against this possibility. Despite their inconsistencies, both scenarios still maintain a certain degree of plausibility, considering that, although neither Archaea nor cyanobacteria genomes encode a G+R-type ME to date, they could have done so in the past.

Besides, our results also suggest that eukaryotes from other kingdoms used independent strategies for acquiring the current set of NAD(P)-ME isoforms. While invertebrates hold a single nuclear *NAD-ME* gene, the gene family coding the full set of the cytosolic and mitochondrial NAD(P)-MEs is derived from a  $\alpha$ -proteobacteria-related ancestral gene in vertebrates (Figure 3). In this case, modifying cofactor affinity (NADP/NAD) must have been a highly plausible evolutionary event, given the limited amino acid determinants required to completely switch the dinucleotide specificity (Hsieh et al., 2011). Even in land plants, the plastidic NADP-ME isoforms were generated by duplications of genes encoding cytosolic isoforms, which in turn are related to  $\alpha$ -proteobacterial progenitors of mitochondria (Figure 3). On the other hand, it seems that fungi adopted uncommon mechanisms for gaining mitochondrial ME, as sequences clustered in several branches of the phylogenetic tree (Figure 3). Some fungi possibly duplicated a pre-existing gene or changed the subcellular location of the genetic product, while others attained it through HGT events with Eubacteria (Figure 3) or even with red algae, such as *M. circinelloides* and *R. oryzae* (Supplementary Figure S4). This evolutionary repertoire of fungal MEs is consistent with recently reported studies (Vorapreeda et al., 2013). Another example of an independent mitochondrial ME acquisition in eukaryotes is the case of the amoeba *E. histolytica*. In this case, it is postulated that the hydrogenosomal Class II-ME isoform was obtained from an archaeon related to *Archaeoglobus fulgidus* (Field et al., 2000).

Therefore, plant mNAD-ME seems to be unrelated to cNADP-ME and pNADP-ME counterparts, having been acquired in an independent evolutionary process. The different intron-exon constitution of *NADP-ME* and *NAD-ME* genes is in good agreement with this notion (Supplementary Figure S2). If both gene classes had been incorporated into the plant genome in different stages of evolution, then they would have been subjected to independent intron gain events. A chimeric structure has also been reported for several gene families encoding enzymes of the TCA cycle, as citrate synthase, aconitase, isocitrate dehydrogenase, and malate dehydrogenase (Cavalcanti et al., 2014). In fact, phylogenetic studies for the enzymes of the glycolysis, TCA cycle, gluconeogenesis, and “Calvin–Benson” cycle, among other, indicated that the pathways evolve as coherent entities of enzymatic activity rather than as protein associations (Schnarrenberger and Martin, 2002). Moreover, the pathway-associated genes can be easily replaced by equivalents through lateral gene transfer to furnish to the pathway new functions to full fit specific physiological demands. This implicates that a gene family, with several isoforms differing in their subcellular location and/or being involved in unrelated processes in different organisms, can exhibit an evolutionary-based organism-specific composition.

## SDP Residues Could Be Associated to the OAD Reaction

We found several SDP residues that could determine the inclusion of an ME sequence into one particular phylogenetic group (Figure 5B). The fact that the more constrained analysis, limited to embryophytic G+R isoforms, resulted in the incorporation of only one high z-score SDP (W254) suggests that signature residues are a footprint of the whole phylogenetic group. In this regard, residues homologous to W239 and S309 are present in current  $\gamma$ -, but not  $\alpha$ -proteobacteria, bolstering our idea that G+R-type ME origin was not associated with mitochondrial biogenesis. Regarding S309, this residue as well as S157 is part of the consensus sequence GXGXXG/A included in the two phosphate cofactor-binding regions (Figure 2). Members of the G+R clade show an invariant consensus sequence GXGSXG/A in both regions, but in the A+V isoforms, the Ser residue is not present, being mainly substituted by Glu (Figures 2, 5B).

Interestingly, we observed that six residues with the highest SDP scores form a cluster of spatial positions surrounding the active site (Figure 5B). Because the evolution-based division separates those animal and plant MEs able to decarboxylate OAA (included in the A+V clade) from those that cannot (plant mNAD-ME in the G+R clade), it is exciting to suppose that some of these catalytic site-linked SDP residues could have a role in the ability of the ME to catalyze the secondary OAD reaction. Also, taking into account the footprint nature of relevant SDP residues in the G+R clade (Figure 5A), the lack of OAA decarboxylation activity would be an emergent property of all G+R isoforms positively selected during evolution of the photosynthetic eukaryotic lineage. A site-directed mutagenesis approach where SDP residues are replaced will shed light on this sequence-function relationship in plant mNAD-MEs.

## CONCLUSION

The MEF exhibits substantial sequence diversity, albeit the related overall tertiary architecture is conserved in the resolved crystallographic structures of several ME members. By adopting a not highly dependence on sequence fold, MEF members have undergone a sub- and/or neo-functionalization during evolution, exemplified by the OAD and MLE activities, as well as the lack of OAD and reverse reaction activities of plant mNAD-MEs. Also, this ability also accounts for the broad range of kinetic parameters, cofactor specificity [NADP(H) and/or NAD(H)], multimodular conformation (e.g., MaeB from *E. coli*), and the remarkable divergence in the nature of allosteric compounds regulating enzymatic activity (Teller et al., 1992; Bologna et al., 2007; Gerrard Wheeler et al., 2008; Tronconi et al., 2010b; Alvarez et al., 2013). Despite this, certain restrictions were imposed on A + V MEs sequence evolution that prevented their co-optation to address the mitochondrial metabolism of photosynthetic organisms. Hence, as in a “shuffle and deal again” scheme, the coupled-photosynthesis evolution molded a particular mNAD-ME structure with exclusive biochemical characteristics. The possibility that some SDP residues of G+R MEs are functionally

connected to the biochemical properties of these isoforms would be an example of such “tailor-made” mechanisms that relied on the flexibility of certain sequence folds in the MEF.

## AUTHOR CONTRIBUTIONS

MT designed the concepts and performed the bioinformatic approaches of this study. MT and MD were involved in the analysis and interpretation of the results. MT, in collaboration with the other authors, drafted the manuscript.

## FUNDING

This study was supported by CONICET.

## ACKNOWLEDGMENTS

MT, CA, and MD are members of the Researcher Career of Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina). We thank Dr. Mariana Saigo for the helpful discussion on sequence analysis.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.00565/full#supplementary-material>

**FIGURE S1 |** Reactions catalyzed by MEF proteins. ME and MLE are involved in the conversion of L-malate into pyruvate and L-lactate, respectively. OAD enzymes catalyze the decarboxylation of OAA to give pyruvate. In the MLE catalysis, NAD acts as cofactor removing two electrons from L-malate and transferring them to pyruvate. Dotted arrow in MLE reaction indicates that the reverse activity has not been measured to date. The presence of a divalent cation ( $\text{Me}^{2+}$ ) is required in all cases.

**FIGURE S2 |** Co-evolutionary residue positions of plant mNAD-MEs. **(A)** Circos representation for co-evolutionary residue positions of  $\alpha$  and  $\beta$  mNAD-MEs. Labels in the first and second circles indicate the amino acid number of the  $\alpha$ -mNAD-ME from *A. thaliana*. Colored square boxes of the third circle indicate the multiple

sequence alignment conservation (from red to cyan, red: the most conserved positions and cyan: the less conserved ones). The fourth circle shows MI scores as histogram. Lines in the center of the circle connect pairs of positions with MI score  $>6.5$ . Red lines represent the top 5%; black lines are between 70 and 95%, while gray lines account for the last 70%. Outer red bars indicate regions with the largest number of MI connections, while sequence insertions A, C, D, and F are highlighted in blue. **(B)** Interactive framework showing part of the MI network between residues surrounding, or located in, the sequence insertions. Ovals with letters indicate residues located in insertions A, C, D, and F. **(C)** Interactive framework showing part of the MI network between residues from different subunits. Ovals include residues exhibiting MIs that are shown in the Logo panel. Asterisks indicate the amino acid residue taken as a reference for the exemplified MI network. Logo panels show the amino acid frequency of occurrence along with the corresponding MI value.

**FIGURE S3 |** Exon structure in plant and animal NAD(P)-MEs. **(A)** Alignment of the predicted amino acid sequences of all ME family members in *H. sapiens*, *X. laevis*, *D. rerio*, and *C. livia*. Fragments encoded by different exons in the amino acid sequence are shown in colors and numbered from E1 to E14. Highlighted sequences indicate mitochondrial targeting peptide. **(B)** Alignment of the predicted amino acid sequences of all ME family members in *P. patens* (three cNADP-ME and one mNAD-ME), *S. moellendorffii* (two cNADP-ME and one mNAD-ME), *A. thaliana* (three cNADP-ME, one pINADP-ME, and two mNAD-ME), and *Z. mays* (three cNADP-ME, two pINADP-ME, and two mNAD-ME). Fragments encoded by different exons in the amino acid sequence are shown in colors and numbered from E0 to E19 for NADP-ME or from e1 to e19 for NAD-ME. Highlighted sequences denote plastidic or mitochondrial targeting peptides.

**FIGURE S4 |** Bayesian 50% majority-rule consensus tree of Class I-ME protein sequences. Evolutionary history was inferred based on the WAG + G + I model. The main lineages are compressed and designated by their name. Nodes with BBP support  $>0.95$  are indicated next to each branch. The area of the triangles representing collapsed clades is not proportional to the number of taxa within the clade. The red section in animal clade represents chromist sequence insertion and the gray section in chromist clade represents fungi sequences.

**FIGURE S5 |** Genes potentially originated from photosynthetic organisms in *R. oryzae* and *M. circinelloides*. Phylogenetic tree using ML method (WAG model) of amino acid sequences. In all cases, numerals indicate bootstrap support of branches using 100 replicates. Branch length is proportional to sequence changes. Color line indicates cyanoacteria, red alga, or fungal sequences. These genes were selected based on the photosynthesis-related homologous present in *Phytophthora sojae* (Tyler et al., 2006).

**TABLE S1 |** Accession number of MEF sequences used in this work. \*Indicate sequences used for identity calculation in **Figure 1**. The sequences of MEs, MLEs and OADs were recovered from the National Center for Biotechnology Information database (<http://ncbi.nlm.nih.gov>). For plants and algal species with entire genome information MEs sequences were extracted from [www.phytozome.net](http://www.phytozome.net), while animal and fungal MEs were retrieved from [www.genomesize.com](http://www.genomesize.com) and <http://fungiadb.org>, respectively.

## REFERENCES

- Alvarez, C. E., Saigo, M., Margarit, E., Andreo, C. S., and Drincovich, M. F. (2013). Kinetics and functional diversity among the five members of the NADP-malic enzyme family from *Zea mays*, a C4 species. *Photosynth. Res.* 115, 65–80. doi: 10.1007/s11120-013-9839-9
- Archibald, J. M. (2015). Endosymbiosis and eukaryotic cell evolution. *Curr. Biol.* 25, R911–R921. doi: 10.1016/j.cub.2015.07.055
- Ball, S. G., Subtil, A., Bhattacharya, D., Moustafa, A., Weber, A. P. M., Gehre, L., et al. (2013). Metabolic effectors secreted by bacterial pathogens: essential facilitators of plastid endosymbiosis? *Plant Cell* 25, 7–21. doi: 10.1105/tpc.112.101329
- Bhattacharyya, M., Upadhyay, R., and Vishveshwara, S. (2012). Interaction signatures stabilizing the NAD(P)-binding Rossmann fold: a structure network approach. *PLoS One* 7:e51676. doi: 10.1371/journal.pone.0051676
- Bologna, F. P., Andreo, C. S., and Drincovich, M. F. (2007). *Escherichia coli* Malic enzymes: two isoforms with substantial differences in kinetic properties, metabolic regulation and structure. *J. Bacteriol.* 189, 5937–5946. doi: 10.1128/JB.00428-07
- Buslje, C. M., Santos, J., Delfino, J. M., and Nielsen, M. (2009). Correction for phylogeny, small number of observations and data redundancy improves the identification of coevolving amino acid pairs using mutual information. *Bioinformatics* 25, 1125–1131. doi: 10.1093/bioinformatics/btp135
- Caivalanti, J. H., Esteves-Ferreira, A. A., Quinhones, C. G., Pereira-Lima, I. A., Nunes-Nesi, A., Fernie, A. R., et al. (2014). Evolution and functional implications of the tricarboxylic acid cycle as revealed by phylogenetic analysis. *Genome Biol. Evol.* 6, 2830–2848. doi: 10.1093/gbe/evu221
- 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+

- Chi, W., Yang, J., Wu, N., and Zhang, F. (2004). Four rice genes encoding NADP malic enzyme exhibit distinct expression profiles. *Biosci. Biotechnol. Biochem.* 68, 1865–1874. doi: 10.1271/bbb.68.1865
- Claros, M. G., and Vincens, P. (1996). Computational method to predict mitochondrial imported proteins and their targeting sequences. *Eur. J. Biochem.* 247, 779–786. doi: 10.1111/j.1432-1033.1996.00779.x
- Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., et al. (2012). Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492, 59–65. doi: 10.1038/nature11681
- Cushman, J. (1992). Characterization and expression of a NADP-malic enzyme cDNA induced by salt stress from the facultative crassulacean acid metabolism plant, *Mesembryanthemum crystallinum*. *Eur. J. Biochem.* 208, 259–266.
- Dolezal, P., Vanacova, S., Tachezy, J., and Hrdý, I. (2004). Malic enzymes of *Trichomonas vaginalis*: two enzyme families, two distinct origins. *Gene* 31, 81–92. doi: 10.1111/j.1432-1033.1992.tb17181.x
- Drincovich, M. F., Casati, P., and Andreo, C. S. (2001). NADP-malic enzyme from plants: a ubiquitous enzyme involved in different metabolic pathways. *FEBS Lett.* 490, 1–6. doi: 10.1016/j.gene.2003.12.022
- Drincovich, M. F., Lara, M., Maurino, V. G., and Andreo, C. (2010). “C4 decarboxylases: different solutions for the same biochemical problem, the provision of CO<sub>2</sub> in the bundle sheath cells,” in *C4 Photosynthesis and Related CO<sub>2</sub> Concentrating Mechanisms*, eds A. Raghavendra and R. F. Sage (Heidelberg: Springer), 277–300. doi: 10.1016/S0014-5793(00)02331-0
- 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
- Emanuelsson, O., Brunak, S., Nielsen, H., and von Heijne, G. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* 2, 953–971. doi: 10.1038/nprot.2007.131
- Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein Sci.* 8, 978–984. doi: 10.1110/pst.8.5.978
- Espariz, M., Repizo, G., Blancato, V., Mortera, P., Alarcon, S., and Magni, C. (2011). Identification of malic and soluble oxaloacetate decarboxylase enzymes in *Enterococcus faecalis*. *FEBS J.* 278, 2140–2151. doi: 10.1111/j.1742-4658.2011.08131.x
- Field, J., Rosenthal, B., and Samuelson, J. (2000). Early lateral transfer of genes encoding malic enzyme, acetyl-CoA synthetase and alcohol dehydrogenases from anaerobic prokaryotes to *Entamoeba histolytica*. *Mol. Microbiol.* 38, 446–455. doi: 10.1046/j.1365-2958.2000.02143.x
- 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
- Ghoshroy, S., Binder, M., Tartar, A., and Robertson, D. L. (2010). Molecular evolution of glutamine synthetase II: phylogenetic evidence of a non-endosymbiotic gene transfer event early in plant evolution. *BMC Evol. Biol.* 10:198. doi: 10.1186/1471-2148-10-198
- Goodey, N. M., and Benkovic, S. J. (2008). Allosteric regulation and catalysis emerge via a common route. *Nat. Chem. Biol.* 4, 474–482. doi: 10.1038/nchembio.98
- Goodridge, A. G., Klautky, S. A., Fantozzi, D. A., Baillie, R. A., Hodnett, D. W., Chen, W., et al. (1996). Nutritional and hormonal regulation of expression of the gene for malic enzyme. *Prog. Nucleic Acid Res. Mol. Biol.* 52, 89–122. doi: 10.1016/S0079-6603(08)60965-4
- Gray, M. W. (2012). Mitochondrial evolution. *Cold Spring Harb. Perspect. Biol.* 4:a011403. doi: 10.1101/cshperspect.a011403
- Hillis, M. D., 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: 10.1093/sysbio/42.2.182
- Hsieh, J. Y., Chen, M. C., and Hung, H. C. (2011). Determinants of nucleotide-binding selectivity of malic enzyme. *PLoS One* 6:e25312. doi: 10.1371/journal.pone.0025312
- Kalinina, O. V., Novichkov, P. S., Mironov, A. A., Gelfand, M. S., and Rakhamanova, A. B. (2004). SDPpred: a tool for prediction of amino acid residues that determine differences in functional specificity of homologous proteins. *Nucleic Acids Res.* 32, W424–W428. doi: 10.1093/nar/gkh391
- Keeling, P. J., and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* 9, 605–618. doi: 10.1038/nrg2386
- Lai, L. B., Lin, W., and Nelson, T. M. (2002). Distinct but conserved functions for two chloroplastic NADP-malic isozymes in C3 and C4 *Flaveria* species. *Plant Physiol.* 128, 125–139. doi: 10.1104/pp.010448
- Laporte, M. M., Shen, B., and Tarczynski, M. C. (2002). Engineering for drought avoidance: expression of maize NADP-malic enzyme in tobacco results in altered stomatal function. *J. Exp. Bot.* 3, 699–705. doi: 10.1093/jexbot/53.369.699
- Mallick, S., Harris, B. G., Cook, P. F. (1991). Kinetic mechanism of NAD:malic enzyme from *Ascaris suum* in the direction of reductive carboxylation. *J. Biol. Chem.* 266, 2732–2738.
- Marino Buslje, C., Teppa, E., Di Domenico, T., Delfino, J. M., and Nielsen, M. (2010). Networks of high mutual information define the structural proximity of catalytic sites: implications for catalytic residue identification. *PLoS Comput. Biol.* 6:e1000978. doi: 10.1371/journal.pcbi.1000978
- Martin, L. C., Gloor, G. B., Dunn, S. D., and Wahl, L. M. (2005). Using information theory to search for co-evolving residues in proteins. *Bioinformatics* 21, 4116–4124. doi: 10.1093/bioinformatics/bti671
- Martinoia, E., and Rentsch, D. (1994). Malate compartmentation: responses to a complex metabolism. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 45, 447–467. doi: 10.1146/annurev.pp.45.060194.002311
- Maurino, V. G., Saigo, M., Andreo, C. S., and Drincovich, M. F. (2001). Nonphotosynthetic 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
- McKeehan, W. L., and McKeehan, K. A. (1982). Changes in NAD(P)-dependent malic enzyme and malate dehydrogenase activities during fibroblast proliferation. *J. Cell. Physiol.* 110, 142–148. doi: 10.1002/jcp.1041100206
- Moreadith, R. W., and Lehninger, A. L. (1984). The pathways of glutamate and glutamine oxidation by tumor cell mitochondria. Role of mitochondrial NAD(P)-dependent malic enzyme. *J. Biol. Chem.* 259, 6215–6221.
- Outlaw, W. H., Manchester, J., and Brown, P. H. (1981). High levels of malic enzyme activities in *Vicia faba* L. epidermal tissue. *Plant Physiol.* 68, 1047–1051. doi: 10.1104/pp.68.5.1047
- Price, D. C., Chan, C. X., Yoon, H. S., Yang, E. C., Qiu, H., Weber, A. P., et al. (2012). *Cyanophora paradoxa* genome elucidates origin of photosynthesis in algae and plants. *Science* 333, 843–847. doi: 10.1126/science.1213561
- Qiu, H., Yoon, H. S., and Bhattacharya, D. (2013). Algal endosymbionts as vectors of horizontal gene transfer in photosynthetic eukaryotes. *Front. Plant Sci.* 4:366. doi: 10.3389/fpls.2013.00366
- Ronquist, F., and Huelsenbeck, J. P. (2003). MRBAYES 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 2003, 1572–1574. doi: 10.1093/bioinformatics/btg180
- Saigo, M., Alvarez, C. E., Andreo, C. S., and Drincovich, M. A. (2013). Plastidial NADP-malic enzymes from grasses: unraveling the way to the C4 specific isoforms. *Plant Physiol. Biochem.* 63, 39–48. doi: 10.1016/j.plaphy.2012.11.009
- Sanchez, R., and Cejudo, F. J. (2003). Identification and expression analysis of a gene encoding a bacterial-type phosphoenolpyruvate carboxylase from *Arabidopsis* and rice. *Plant Physiol.* 132, 949–957. doi: 10.1104/pp.102.019653
- Schaaf, J., Walter, M. H., and Hess, D. (1995). Primary metabolism in plant defense (regulation of a bean malic enzyme gene promoter in transgenic tobacco by developmental and environmental cues). *Plant Physiol.* 108, 949–960. doi: 10.1104/pp.108.3.949
- Schnarrenberger, C., and Martin, W. (2002). Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants A case study of endosymbiotic gene transfer. *Eur. J. Biochem.* 269, 868–883. doi: 10.1046/j.0014-2956.2001.02722.x
- Simonetti, F. L., Teppa, E., Chernomoretz, A., Nielsen, M., and Buslje, C. M. (2013). MISTIC: mutual information server to infer coevolution. *Nucleic Acids Res.* 41, W8–W14. doi: 10.1093/nar/gkt427
- Sweetlove, L. J., Beard, K. F. M., Nunes-Nesi, A., Fernie, A. R., and Ratcliffe, R. G. (2010). Not just a circle: flux modes in the plant TCA cycle. *Trends Plant Sci.* 15, 462–470. doi: 10.1016/j.tplants.2010.05.006
- Swofford, D. L., Olsen, G. L., Waddell, P. J., and Hillis, D. M. (1996). “Phylogenetic inference,” in *Molecular Systematics*, 2nd Edn, eds D. M. Hillis, C. Moritz, and B. K. Mable (Sunderland, MA: Sinauer Associates), 407–515.

- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Teller, J. K., Fahien, L. A., and Davis, J. W. (1992). Kinetics and regulation of hepatoma mitochondrial NAD(P) malic enzyme. *J. Biol. Chem.* 25, 10423–10432.
- Tronconi, M. A., Fahnenschlitz, H., Gerrard Wheeler, M. C., Andreo, C. S., Flügge, U. I., Drincovich, M. F., et al. (2008). Arabidopsis NAD-malic enzyme functions as a homodimer and heterodimer and has a major impact during nocturnal metabolism. *Plant Physiol.* 146, 1540–1552. doi: 10.1104/pp.107.114975
- Tronconi, M. A., Gerrard Wheeler, M. C., Drincovich, M. F., and Andreo, C. S. (2012). Differential fumarate binding to Arabidopsis NAD+-malic enzymes 1 and -2 produces an opposite activity modulation. *Biochimie* 94, 1421–1430. doi: 10.1016/j.biochi.2012.03.017
- Tronconi, M. A., Gerrard 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
- Tronconi, M. A., Maurino, V. G., Andreo, C. S., and Drincovich, M. F. (2010b). Three different and tissue-specific NAD-malic enzyme generated by alternative subunit association in *Arabidopsis thaliana*. *J. Biol. Chem.* 285, 11870–11879. doi: 10.1074/jbc.M109.097477
- 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
- Tyler, B. M., Tripathy, S., Zhang, X., Dehal, P., Jiang, R. H., Aerts, A., et al. (2006). Phytophthora genome sequences uncover evolutionary origins and mechanisms of pathogenesis. *Science* 313, 1261–1266. doi: 10.1126/science.1128796
- van der Giesen, M., Rechinger, K. B., Svendsen, I., Durand, R., Hirt, R. P., Fèver, M., et al. (1997). A mitochondrial-like targeting signal on the hydrogenosomal malic enzyme from the anaerobic fungus *Neocallimastix frontalis*: support for the hypothesis that hydrogenosomes are modified mitochondria. *Mol. Microbiol.* 23, 11–21. doi: 10.1046/j.1365-2958.1997.1891553.x
- Vorapreeda, T., Thammarongtham, C., Cheevadhanarak, S., and Laoten, K. (2013). Repertoire of malic enzymes in yeast and fungi: insight into their evolutionary functional and structural significance. *Microbiology* 159, 2548–2557. doi: 10.1099/mic.0.065342-0
- Whelan, S., and Goldman, N. (2001). A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* 18, 691–699. doi: 10.1093/oxfordjournals.molbev.a003851
- Willeford, K. O., and Wedding, R. T. (1987). Evidence for a multiple subunit composition of plant NAD malic enzyme. *J. Biol. Chem.* 262, 8423–8429.
- Yang, Z., Floyd, D. L., Loeber, G., and Tong, L. (2000). Structure of a closed form of human malic enzyme and implications for catalytic mechanism. *Nat. Struct. Biol.* 7, 251–257.
- Yang, Z., Lanks, C. W., and Tong, L. (2002). Molecular mechanism for the regulation of human mitochondrial NAD(P)-dependent malic enzyme by ATP and fumarate. *Structure* 10, 951–960. doi: 10.1016/S0969-2126(02)00788-8

**Conflict of Interest Statement:** 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.

The reviewer SR and handling Editor declared their shared affiliation.

Copyright © 2018 Tronconi, Andreo and Drincovich. 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 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.