Convergent molecular evolution of phosphoenolpyruvate carboxylase gene family in C$_4$ and crassulacean acid metabolism plants

Jiang-Ping Shu[1,2,3,4], Yue-Hong Yan[2,3,4] and Rui-Jiang Wang[1,4]

[1] Key laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

[2] Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization, The National Orchid Conservation Centre of China and The Orchid Conservation and Research Centre of Shenzhen, Shenzhen, China

[3] Shenzhen Key Laboratory for Orchid Conservation and Utilization, The National Orchid Conservation Centre of China and The Orchid Conservation and Research Centre of Shenzhen, Shenzhen, China

[4] University of Chinese Academy of Sciences, Beijing, China

Corresponding authors
Yue-Hong Yan, yhyan@sibs.ac.cn
Rui-Jiang Wang, wangri@scbg.an.cn

ABSTRACT

Phosphoenolpyruvate carboxylase (PEPC), as the key enzyme in initial carbon fixation of C$_4$ and crassulacean acid mechanism (CAM) pathways, was thought to undergo convergent adaptive changes resulting in the convergent evolution of C$_4$ and CAM photosynthesis in vascular plants. However, the integral evolutionary history and convergence of PEPC in plants remain poorly understood. In the present study, we identified the members of PEPC gene family across green plants with seventeen genomic datasets, found ten conserved motifs and modeled three-dimensional protein structures of 90 plant-type PEPC genes. After reconstructing PEPC gene family tree and reconciled with species tree, we found PEPC genes underwent 71 gene duplication events and 16 gene loss events, which might result from whole-genome duplication events in plants. Based on the phylogenetic tree of the PEPC gene family, we detected four convergent evolution sites of PEPC in C$_4$ species but none in CAM species. The PEPC gene family was ubiquitous and highly conservative in green plants. After originating from gene duplication of ancestral C3-PEPC, C4-PEPC isoforms underwent convergent molecular substitution that might facilitate the convergent evolution of C$_4$ photosynthesis in Angiosperms. However, there was no evidence for convergent molecular evolution of PEPC genes between CAM plants. Our findings help to understand the origin and convergent evolution of C$_4$ and CAM plants and shed light on the adaptation of plants in dry, hot environments.

INTRODUCTION

Plant photosynthesis is the most important biochemical process on the planet: it provides energy, food and oxygen for the survival and reproduction of vast majority though heterotrophic organisms, including humans (*Fischer, Hemp & Johnson, 2016*). Since its

origin, oxygen levels have gradually risen, and carbon dioxide (CO_2) concentrations have decreased in the atmosphere (*Foster, Royer & Lunt, 2017*), which has increased photorespiration and led to bioenergy and carbohydrate losses (*Schlüter & Weber, 2020*). To obtain adequate CO_2 and improve photosynthetic efficiency, plants developed several carbon-concentrating mechanisms (CCMs), such as C_4 photosynthesis and crassulacean acid metabolism (CAM), so it could adapt to the sudden decline of atmospheric CO_2 approximately 350 million years ago (Mya) (*Edwards, 2019*; *Heyduk et al., 2019*). C_4 photosynthesis, an example of convergent evolution, has repeatedly evolved more than 60 times in angiosperms (*Sage, Christin & Edwards, 2011*; *Sage, 2016*); whereas, CAM has independently evolved in approximately 37 families of vascular plants (*Silvera et al., 2010*; *Winter et al., 2021*). Recently, several studies using comparative genomics have provided new insights into the genetics and evolution of CCMs (*Yang et al., 2017*; *Heyduk et al., 2019*; *Wai et al., 2019*; *Yang et al., 2019*; *Jaiswal et al., 2021*). However, the molecular mechanisms underlying convergent evolution in CCMs remain poorly understood.

In plants, phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31) is a key enzyme that catalyzes the primary fixation reaction for CO_2 assimilation in C_4 and CAM photosynthesis (*Lepiniec et al., 1994*; *Nimmo, 2000*; *Svensson, Bläsing & Westhoff, 2003*). PEPC is a highly regulated enzyme that catalyzes the irreversible β-carboxylation of phosphoenolpyruvate in the presence of bicarbonate and magnesium to yield oxaloacetate and inorganic phosphate, a reaction that serves a variety of physiological functions in plants (*Jiao & Chollet, 1991*; *Lepiniec et al., 1994*; *Chollet, Vidal & O'Leary, 1996*). Additionally, PEPC genes play crucial roles in a variety of non-photosynthetic functions, including carbon and nitrogen metabolism, seed development and germination and response to abiotic stresses (*Lepiniec et al., 1994*; *O'Leary, Park & Plaxton William, 2011*; *Shi et al., 2015*; *Ruiz-Ballesta et al., 2016*; *Wang et al., 2016*; *Darabi & Seddigh, 2018*; *Zhao et al., 2019*). As housekeeping genes, PEPC genes are highly conserved, which makes them valuable molecular markers for phylogenetic reconstruction of plants (*Gehrig, Heute & Kluge, 2001*). Although PEPC genes share origins during the evolution of C_4 and CAM photosynthesis, these genes are distinct copies that were generated by whole-genome duplication (WGD) in angiosperms (*Christin et al., 2014*).

Interestingly, PEPC genes are essential for regulation of the circadian clock in CAM photosynthesis (*Boxall et al., 2020*) and share convergent amino acid changes in diverse CAM species (*Yang et al., 2017*). In C_4 grasses, PEPC genes have also undergone parallel, adaptive genetic changes (*Christin et al., 2007*; *Besnard et al., 2009*; *Moreno-Villena et al., 2018*). Therefore, PEPC genes are crucial for elucidating the origin and convergent evolution of C_4 and CAM photosynthesis. Previous studies only examined the convergent evolution of PEPC gene family in a few C_4or CAM photosynthetic lineages, such as grass (*Christin et al., 2007*; *Besnard et al., 2009*; *Moreno-Villena et al., 2018*) and angiosperms (*Yang et al., 2017*), but without *Isoetes*, which was the earliest-diverging lineage of CAM plants (*Keeley, 1981*), and without fern CAM plants such as *Platycerium* (*Rut et al., 2008*), so the origin and evolution of the PEPC gene family could not be clearly elucidated (*Deng et al., 2016*). Furthermore, CAM photosynthesis occurs widely in the major clades of vascular plants, including pteridophytes (lycopods and ferns), gymnosperms and angiosperms, while C_4

photosynthesis is only distributed in angiosperms. Therefore, to understand the convergent molecular evolution of the PEPC gene family in the plant kingdom, species with genomic datasets that represent C_3, C_4 and CAM photosynthesis across the major lineages of plants should be sampled.

Fortunately, more and more plant genomes have been sequenced (https://www.plabipd.de/index.ep), which provides an excellent opportunity to resolve the origin and convergent evolution of C_4 and CAM photosynthesis in plants (*Heyduk et al., 2019*; *Yang et al., 2019*; *Gilman & Edwards, 2020*). However, it is a big challenge to use hundreds of genomic datasets to analyze the convergent evolution of PEPC genes, which have multiple copies in most species. In the present study, we only selected 17 plant genomes, which consisted of C_3, C_4 and CAM species across algae, bryophytes, pteridophytes, gymnosperms and angiosperms, especially included *Isoetes* and *Platycerium*, which represented the earliest-diverging lineages of CAM plants. Then we identified the PEPC genes from the 17 genomes and reconstructed the evolutionary history and molecular convergence of PEPC genes in C_4 and CAM plants. Our study will help to elucidate the origin and evolution of C_4 and CAM plants and shed light on the adaptation of plants in dry, hot environments.

MATERIALS & METHODS

Data acquisition

C_4 and CAM photosynthesis are widely distributed in angiosperms and vascular plants, respectively (*Heyduk et al., 2019*). To obtain representative sampling of these groups and avoid the difficulties in analysis with big datasets, we only selected 17 plant species that exhibited a variety of photosynthetic pathways (including C_3, C_4 and CAM): *Spirogloea muscicola*, which is the closest algae to land plants (*Cheng et al., 2019*), three bryophytes, two lycopods, two ferns, two gymnosperms and seven angiosperms (Table 1). Seventeen genomic datasets were obtained from public databases, such as Figshare (https://figshare.com/), EnsemblPlants (http://plants.ensembl.org), Dryad (https://datadryad.org), Fernbase (https://www.fernbase.org/), GigaDB (http://gigadb.org/), Congenie (https://congenie.org/), and Phytozome (v13, https://phytozome-next.jgi.doe.gov).

Gene family identification

To identify members of the PEPC gene family, we created a local BLAST database with protein sequences from the 17 plant species and then performed BLASTP searches with default parameters using PEPC protein sequences from Arabidopsis (At1g53310, At1g68750, At2g42600 and At3g14940) as queries (*Camacho et al., 2009*). Furthermore, the Pfam seed alignment of the PEPcase domain (PF00311, http://pfam.xfam.org/) was used to build the HMMER profile; then, we searched for candidate PEPC genes in the 17 genomic datasets using HMMER v3.2.1 (*Mistry et al., 2013*). After combining the two PEPC gene sets, we searched for conserved domains from the Conserved Domain Database using Batch CD-Search with default parameters (*Lu et al., 2020*). Only genes with the PEPcase conserved domain were identified as reliable plant-type PEPC genes, and bacterial-type PEPC gene contained the PRK00009 domain.

Table 1 The information of samples used in this study.

ID	Species	Class	Type	PTPC	BTPC	PEPCase superfamily	Data resources
Spm	*Spirogloea muscicola*	Algae	C3	3	2	4	Figshare
Mp	*Marchantia polymorpha*	Bryophytes	C3	1	1	1	EnsemblPlants
Pp	*Physcomitrella patens*	Bryophytes	C3	22	0	10	EnsemblPlants
Aa	*Anthoceros angustus*	Bryophytes	C3	1	1	2	Dryad
Sm	*Selaginella moellendorffii*	Lycopods	C3	4	4	0	EnsemblPlants
Is	*Isoetes sinensis*	Lycopods	CAM	2	5	0	This study
Af	*Azolla filiculoides*	Ferns	C3	2	0	3	Fernbase
Pb	*Platycerium bifurcatum*	Ferns	CAM	2	0	2	GigaDB
Pa	*Picea abies*	Gymnosperms	C3	0	0	10	Congenie
Gm	*Gnetum montanum*	Gymnosperms	C3	1	1	0	Dryad
Amt	*Amborella trichopoda*	Angiosperms	C3	1	1	0	EnsemblPlants
Ac	*Ananas comosus*	Angiosperms	CAM	2	0	1	EnsemblPlants
Os	*Oryza sativa*	Angiosperms	C3	9	1	4	EnsemblPlants
Zm	*Zea mays*	Angiosperms	C4	22	1	33	EnsemblPlants
At	*Arabidopsis thaliana*	Angiosperms	C3	8	1	0	EnsemblPlants
Ah	*Amaranthus hypochondriacus*	Angiosperms	C4	3	0	0	Phytozome
Kf	*Kalanchoe fedtschenkoi*	Angiosperms	CAM	8	1	0	Phytozome

Notes.
PTPC, plant-type PEPC; BTPC, bacterial-type PEPC.

Prediction of conserved motifs and modeling of protein structure

After identifying reliable PEPC genes with our pipeline above, conserved motifs were predicted by MEME v5.1.0 (*Bailey et al., 2006*) with default parameters. Motif alignment was performed by MAST v5.1.0 with default parameters, and conserved motifs were visualized by TBtools v1.046 (*Chen et al., 2020a*). The 3D structure homology modelling of PEPC proteins was predicted by SWISS-MODEL (*Waterhouse et al., 2018*), which integrate up-to-date protein sequence and structure database as the structural templates. We selected the optimal protein model with the highest values of Global Model Quality Estimate (GMQE) and sequence identify, which indicated the highest reliability of the homology modelling. To detected the convergence of 3D protein structure, we simply test that if there is a 3D structure that only exists in CAM or C_4 plants, we think this structure is convergent.

Phylogenetic reconstruction and gene tree-species tree reconciliation

To understand the evolution of the PEPC gene family, the alignment of PEPC protein sequences was performed by MAFFT v7.453 (*Katoh & Standley, 2013*) with the accurate L-INS-i method and 1000 maximum iterative refinements. The conserved blocks were selected by GBLOCKS 0.91b (*Castresana, 2000*) with the parameters that minimum length of a block was five and allowed gap positions with half. Then, we reconstructed the PEPC gene family tree with maximum likelihood using IQ-TREE v1.6.11 (*Nguyen et al., 2015*). The best-fit amino acid model, JTTDCMut+R5, was detected by ModelFinder

(*Kalyaanamoorthy et al., 2017*) using the Bayesian information criterion. The ultrafast bootstrap approximation was calculated using 1000 random replicates (*Hoang et al., 2017*). Phylogenetic reconciliation of the gene tree and species tree was performed by Treerecs (*Comte et al., 2020*) with default parameters. The species tree used was based on recent phylogenomic reconstruction of green plants (*Initiative, 2019*).

Convergent site detection

The convergent site definition of *Rey et al. (2018)*, that "a substitution is convergent if it occurred toward the same amino acid preference on every branch where the phenotype also changed toward the convergent phenotype", were employed in this study, because several amino acids with similar biochemical properties may have roughly the same fitness at that site (*Rey et al., 2018*), it indicated that convergent site may be not the exact same amino acid in all species with a convergent phenotype. In addition, only some of PEPC gene copies are possibly involved in CAM or C_4 and others are not, but it is difficult to identify which one is the isoform of CAM or C_4. Therefore, we labeled putative convergent clades or gene copies on the phylogenetic tree with three kinds of gene combinations: (1) including all gene copies of species with convergent phenotypes (C_4 or CAM), (2) only containing one clade within each convergent species and (3) only containing one gene copy within each convergent species. For each combination, convergent amino acid site of PEPC proteins in C_4 and CAM plants was detected based on the PCOC model with a posterior probability threshold of 0.8 (*Rey et al., 2018*).

RESULTS

Identification of PEPC gene family

As the key carboxylase, PEPC genes are widely distributed in green plants (Table 1). In the present study, we identified 264 homologous genes using BLASTP searching with Arabidopsis PEPC genes (At1g53310, At1g68750, At2g42600 and At3g14940) as queries and 179 homologous genes using the Pfam seed alignment of the PEPcase domain (PF00311) from 17 genomic datasets across green plants. After combination of the two gene sets, we obtained 179 common, homologous genes and then searched for conserved domains using the Conserved Domain Database: 109 genes contained conserved domains, of which 90 contained the plant-type PEPC (PTPC) gene domain (PEPcase) and 19 contained the bacterial-type PEPC (BTPC) gene domain (PRK00009); the remaining 70 genes contained other PEPcase superfamily domains (Tables 1, S1).

Conserved motifs and structures of PEPC gene family

We predicted ten conserved motifs from 109 PEPC proteins. Each motif was longer than 29 amino acids and was found in more than 104 of the 109 PEPC proteins (Table 2). Most of the amino acids were conserved across all motifs, and the linear order of these motifs, especially in PTPC genes, was identical across all green plants. Some motifs were repeated in various genes (Fig. 1). In order to test if the protein structures evolved convergent in CAM or C_4 plants, we modeled 3D structures of 90 PTPC proteins using the SWISS-MODEL server (Fig. S1) and obtained three templates of protein structures (Table S2): 5vyj.1.A

Shu et al. (2022), *PeerJ*, DOI 10.7717/peerj.12828

Table 2 Conserved motifs of PEPC gene family in green plants.

Motifs	Width	Sites	LLR	*E*-value	Consensus sequence
1	29	109	8856	4.7e−3069	RKPSGGIESL**R**AIPWIF**A**WTQTRFHLPVW
2	50	109	14268	9.6e−5072	[VI]KLTMFHGRGG[TS]VG**R**GGGPTHLAILSQPPDT[IV][HN]GSLRVT[VI]QGEVIEQSFG
3	44	108	12398	9.6e−4364	SS**W**MGGDRDGNPRVTPEVTRDVCLLAR[ML]MAANLYFSQIEDLMFE
4	50	104	13093	2.6e−4568	IQAA[FW]RTDEIRRT[PQ]PTPQDEMRAG[ML]SYFHETIW[KN]G[VL]PKFLRRVDTALKNI
5	37	105	10252	2.1e−3574	FSI[DE]WY[RL]NRINGKQEV**M**IGYS**D**SG**K**DAGR[LF]SAAWQ[LM][YF]
6	41	109	10936	4.4e−3771	DVL[DG]TF[HRK]V[IL]AELP[SA]D[SC]FGAY[IV]ISMATAPSDVLAVELLQREC
7	41	109	10861	2.7e−3739	DF[LM]RQVS[TC]FGLSLV[KR]LDIRQES[DE]RHT[DE]V[LM]DAIT[TN][HY]LGIGSY
8	50	109	12887	1.6e−4467	WRAL[ML]DE[MI]AVV[AS]T[KE]EYRS[IV]VF[QK][EN]PRFVEYFRLATPELEYGR[ML]NIGS**R**PSK
9	41	108	10512	6.4e−3599	[TM]L[QR][EA]MYN[EQ]WPFFRVTIDL[VI]EMVFA**K**GDP[GR]IAALYDKLLVS[ED]
10	29	108	7470	7.5e−2482	EEHLCFRTLQR[FY]TAATLEHGMHPPI[SA]PKP

Notes.

Width, the length of motif; Sites, the number of sequences including the motif; LLR, log likelihood rate.
Amino acid residues of experimentally proven function in *Wang et al. (2016)* are indicated by red majuscules.
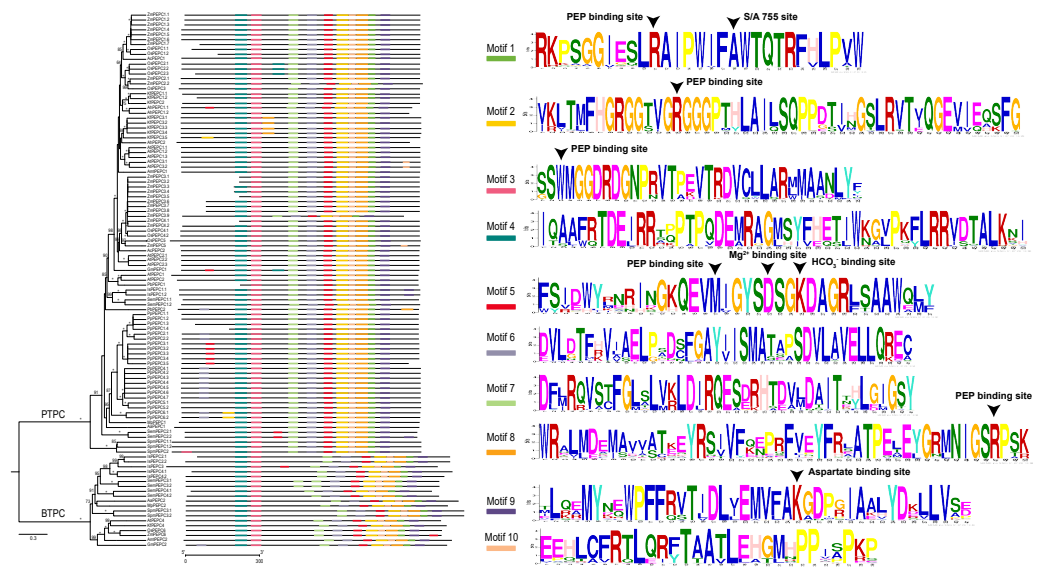
Figure 1 The phylogenetic tree and ten conserved motifs of PEPC gene family in plants. The asterisk in the maximum likelihood tree indicated 100 bootstrap support value, and the value lower than 60 was not showed in gene family tree. Arrowhead is indicated the amino acid residues of experimentally proven function in *Wang et al. (2016)*. PTPC: plant-type PEPC; BTPC: bacterial-type PEPC. Aa, *Anthoceros angustus*; Ac, *Ananas comosus*; Af: *Azolla filiculoides*; Ah, *Amaranthus hypochondriacus*; Amt, *Amborella trichopoda*; At, *Arabidopsis thaliana*; Gm, *Gnetum montanum*; Is, *Isoetes sinensis*; Kf, *Kalanchoe fedtschenkoi*; Mp, *Marchantia polymorpha*; Os, *Oryza sativa*; Pa, *Picea abies*; Pb, *Platycerium bifurcatum*; Pp, *Physcomitrella patens*; Sm, *Selaginella moellendorffii*; Spm, *Spirogloea muscicola*; Zm, *Zea mays*.
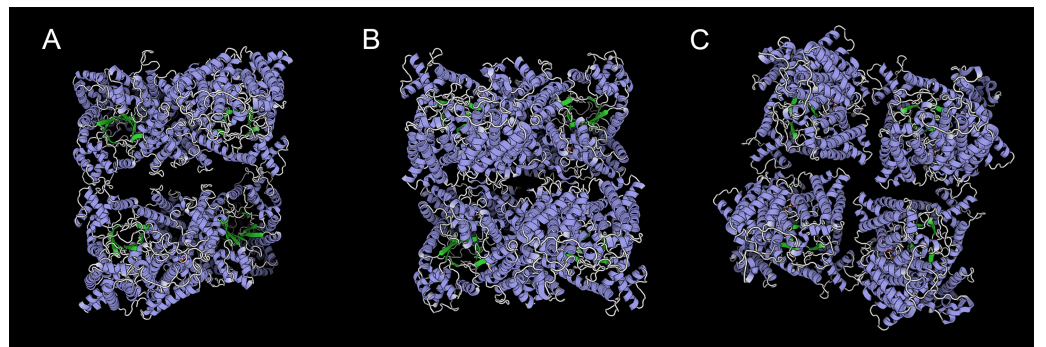
Figure 2 Three templates of three-dimensional structure in PTPC proteins,. (A) 5vyj.1.A; (B) 3zgb.1.A; (C) 5fdn.1.A. The three pictures are viewed from the same angles.

(Fig. 2A), 3zgb.1.A (Fig. 2B) and 5fdn.1.A (Fig. 2C). The 3zgb.1.A template was widely distributed in all sampled species. The 5vyj.1.A template was distributed in seven species. The 5fdn.1.A template was only distributed in Arabidopsis (Fig. 3, Table S2).
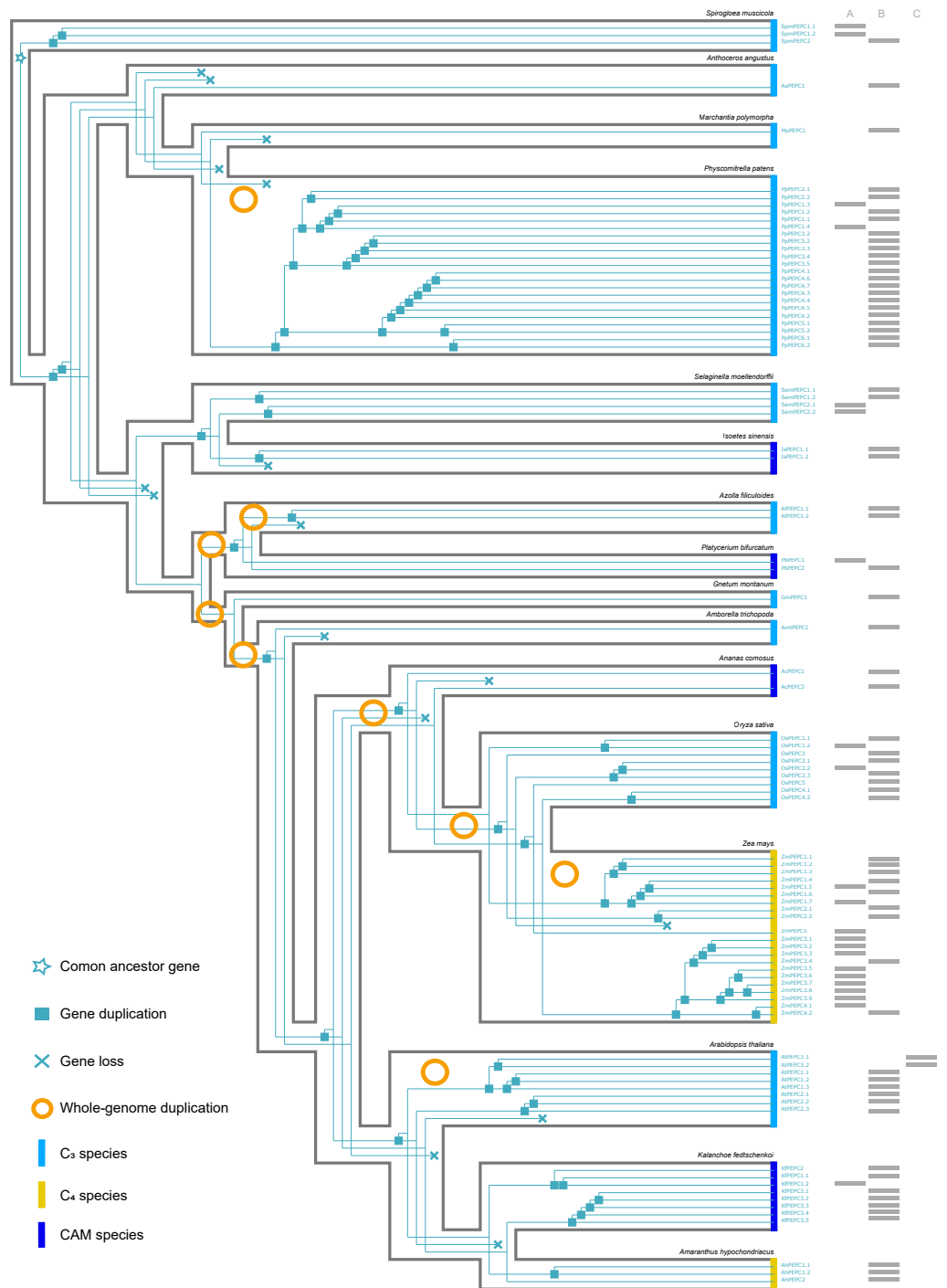
Figure 3 The evolutionary history of PTPC gene family in green plants. (A, B, C) respectively corresponded to three protein structures (A, B, C) in Fig. 2. The information of whole-genome duplication events was based on previous studies (*Cheng et al., 2018*; *Li et al., 2018*). Aa: *Anthoceros angustus*; Ac, *Ananas comosus*; Af, *Azolla filiculoides*; Ah, *Amaranthus hypochondriacus*; Amt, *Amborella trichopoda*; At, *Arabidopsis thaliana*; Gm, *Gnetum montanum*; Is, *Isoetes sinensis*; Kf, *Kalanchoe fedtschenkoi*; Mp, *Marchantia polymorpha*; Os, *Oryza sativa*; Pa, *Picea abies*; Pb, *Platycerium bifurcatum*; Pp, *Physcomitrella patens*; Sm, *Selaginella moellendorffii*; Spm, *Spirogloea muscicola*; Zm, *Zea mays*.

Gene duplication and loss

Here, we reconstructed the robust phylogenetic tree of the PEPC gene family with relatively adequate sampling to include C_3, C_4 and CAM plants and the bootstrap support values of all branches were mostly higher than 60 (Fig. 1). And the evolutionary history of PTPC genes was independently reconstructed with maximum likelihood and was reconciled with the species tree based on duplication-loss reconciliation (Fig. 3). The reconciliation tree showed that PTPC genes underwent at least 71 duplications and 16 losses in the evolutionary history of our sampled species.

Convergent evolution of PEPC gene family

To test whether convergent molecular evolution at the amino acid level occurred in PTPC proteins, we detected convergent sites in all phenotypically convergent clades of C_4 and CAM photosynthesis with the Profile Change with One Change (PCOC) pipeline. The results showed two convergent shifts in PTPC proteins that occurred in CAM species and three convergent shifts that occurred in C_4 species, the posterior probabilities (pp) for the PCOC model were greater than 0.9 at all convergent shifts (Fig. 4). In addition, we also detected convergent evolution at sites in different gene groups, in which each convergent phenotypic species retained one clade or one gene copy (one-to-one) (Fig. S2). Four identical convergent amino acid sites in one gene group (AhPEPC1.2/ZmPEPC2) were discovered in C_4 species (Fig. 5). However, no identical convergent sites were found in the one-to-one gene groups of CAM species (Fig. S2).

DISCUSSIONS

PEPC gene family with different copies was conserved in plants

PEPC gene family was ubiquitous but with different copy number in the different lineages of green plants. In the present study, BTPC genes were distributed in 11 of the 17 sampled species and retained relatively fewer copies than PTPC genes (Table 1). Because of missing the conserved PEPcase domain, we did not detect any PTPC genes in Norway spruce (*Picea abies*). This might be resulted from pseudogenization and/or insertion of transposable elements in conifers (*Nystedt et al., 2013*), or incomplete genome assembly because the length of ten homologs proteins (<510 aa) in Norway Spruce was less than that of PTPCs (~900 aa) in other species. Therefore, ten homologs of PEPC in Norway spruce, which contained the PEPcase superfamily domain, probably performed the PTPC-like physiological functions or was the fragments of PTPC genes. Due to numerous WGDs (*Soltis & Soltis, 2016*), the PEPC gene family had relatively more copies in angiosperms, especially in maize that contained 22 PTPC genes. Interestingly, moss (*Physcomitrella patens*) also retained 22 PTPC genes, but its sister clades, hornworts (*Anthoceros angustus*) and liverworts (*Marchantia polymorpha*), only had one PTPC gene (Table 1). This extreme difference in gene content corresponded to different adaptation strategies for plant terrestrialization. Mosses underwent WGDs that increase gene-family complexity for coordinating multicellular growth and responding to dehydration (*Rensing et al., 2008*). However, liverworts have ancient dimorphic sex chromosomes, which may have resulted in a lack of WGDs and reduced proliferation of regulatory genes (*Bowman et al., 2017*).
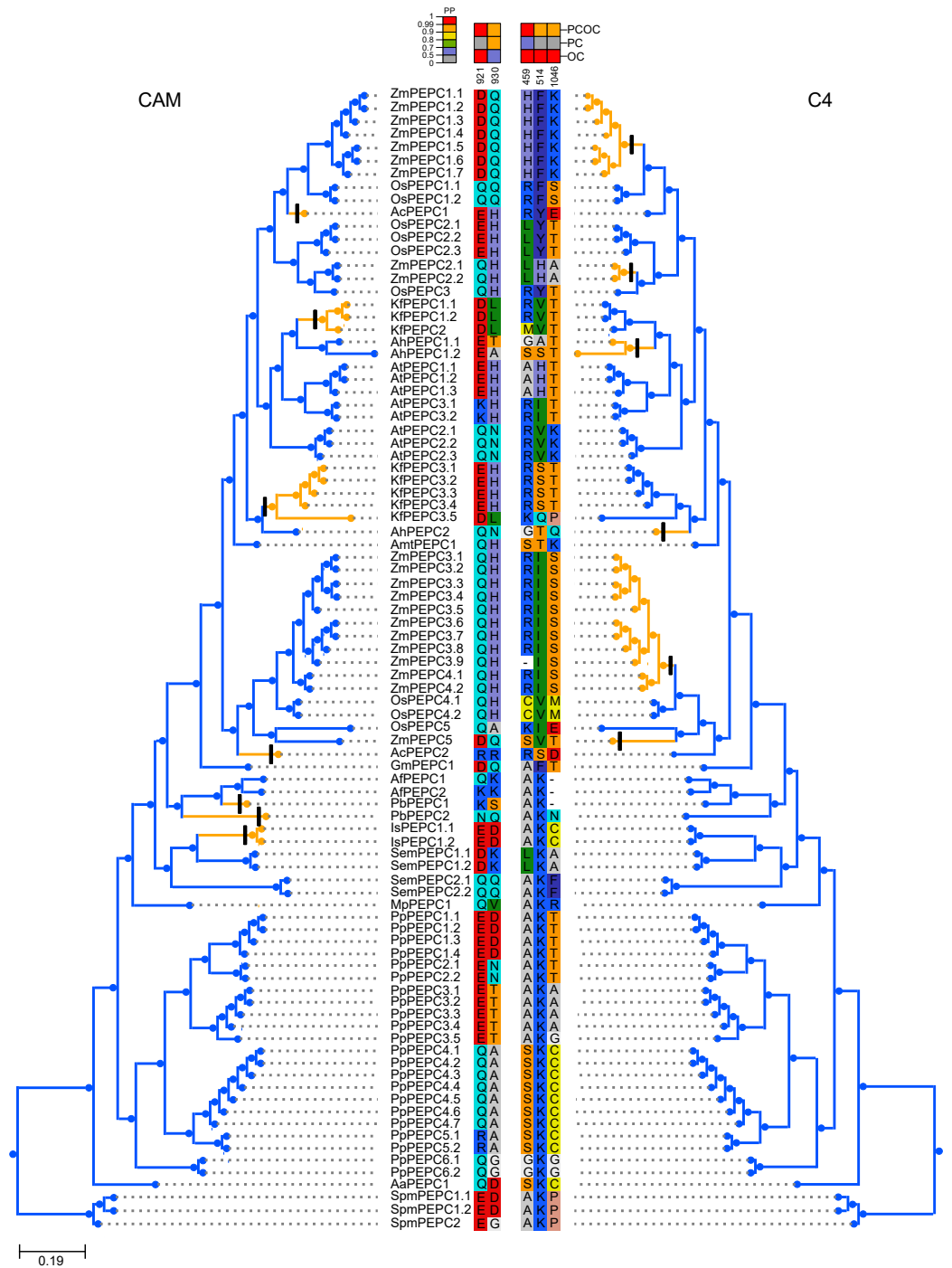
Figure 4 Molecular convergent sites of PEPC gene family in CAM and C₄ species. The clades with orange color indicated the species with convergent phenotypes (CAM and C₄ photosynthesis). PCOC: Profile Change with One Change model; PC: Profile Change model; OC: One Change model, all models were in detail explained by *Rey et al. (2018)*. Posterior probabilities (pp) for the PCOC, PC, and OC models are summarized by top box colors, and the amino acid colors correspond to different amino acid equilibrium frequencies (*i.e.*, different profiles) of the Profile Change with One Change model (PCOC model). (continued on next page…)

Figure 4 (...continued)
Aa, *Anthoceros angustus*; Ac, *Ananas comosus*; Af, *Azolla filiculoides*; Ah, *Amaranthus hypochondriacus*; Amt: *Amborella trichopoda*; At, *Arabidopsis thaliana*; Gm, *Gnetum montanum*; Is, *Isoetes sinensis*; Kf, *Kalanchoe fedtschenkoi*; Mp, *Marchantia polymorpha*; Os, *Oryza sativa*; Pa, *Picea abies*; Pb, *Platycerium bifurcatum*; Pp, *Physcomitrella patens*; Sm, *Selaginella moellendorffii*; Spm, *Spirogloea muscicola*; Zm, *Zea mays*.

The genome of *A. angustus* is interestingly simple and has obtained stress-response and metabolic pathway genes through horizontal gene transfer from bacteria or fungi, which probably assisted its survival in a terrestrial environment (*Zhang et al., 2020*).

PEPC genes displayed highly conserved amino acid sequences in all green plants. Here, all ten motifs were longer than 29 amino acids and was found in more than 104 of the 109 PEPC proteins (Table 2). It is suggested that the PEPC genes were detected reliably, because ultra-conserved motifs indicated similar and/or same function in common (*Bejerano et al., 2004*). Additionally, the linear order of these motifs, especially in PTPC genes, was identical across all green plants (Fig. 1), and all our predicted motifs with the several functional loci, such as PEP binding site, Mg^{2+} binding site, HCO_3^- binding site, S/A 755 site and Asparate binding site, were also reported previously (*Wang et al., 2016*), but the motif 10 reported by *Wang et al. (2016)* was not detected in the present study, which might be only conserved in Angiosperms. These results clearly indicated that the PEPC gene family has been extremely conserved throughout its evolutionary history of more than 500 million years (My), since its origin from algae (*Chollet, Vidal & O'Leary, 1996*; *Svensson, Bläsing & Westhoff, 2003*; *Darabi & Seddigh, 2018*).

However, this conserved gene family has performed hyper-diverse housekeeping functions, including photosynthetic and non-photosynthetic functions, for survival in terrestrial environments (*O'Leary, Park & Plaxton William, 2011*). To understand the function and molecular mechanisms of the PEPC gene family, the three-dimensional (3D) structure of PEPC proteins was elucidated by X-ray crystallographic analysis (*Matsumura et al., 2002*; *Paulus, Schlieper & Groth, 2013*; *Connell et al., 2018*; *González-Segura et al., 2018*), which discovered many structure-function relationships of PEPC catalysis, allosteric control, and regulatory phosphorylation (*Izui et al., 2004*). In order to test if the protein structures evolved convergent in CAM or C_4 plants, we modeled 3D structures of 90 PTPC proteins using the SWISS-MODEL server (Fig. S1) and predicted three templates of protein structures with higher sequence identity (Table S2): 5vyj.1.A (Fig. 2A) (*González-Segura et al., 2018*), 3zgb.1.A (Fig. 2B) (*Paulus, Schlieper & Groth, 2013*) and 5fdn.1.A (Fig. 2C) (*Connell et al., 2018*). All PTPC proteins were tetrameric enzymes with these three kinds of 3D structures, but the four subunits combined differently. In the widespread 3zgb.1.A template, C4-PEPC isoforms had two amino acid substitutions that increased PEP saturation kinetics and reduced inhibitor affinity, respectively, compared to C3-PEPC isoforms (*Bläsing, Westhoff & Svensson, 2000*; *Paulus, Schlieper & Groth, 2013*). Therefore, the efficiency of photosynthetic carbon fixation greatly improved in C_4 plants. However, there are no evidence for convergent evolution of PEPC protein structures in CAM or C_4 plants.
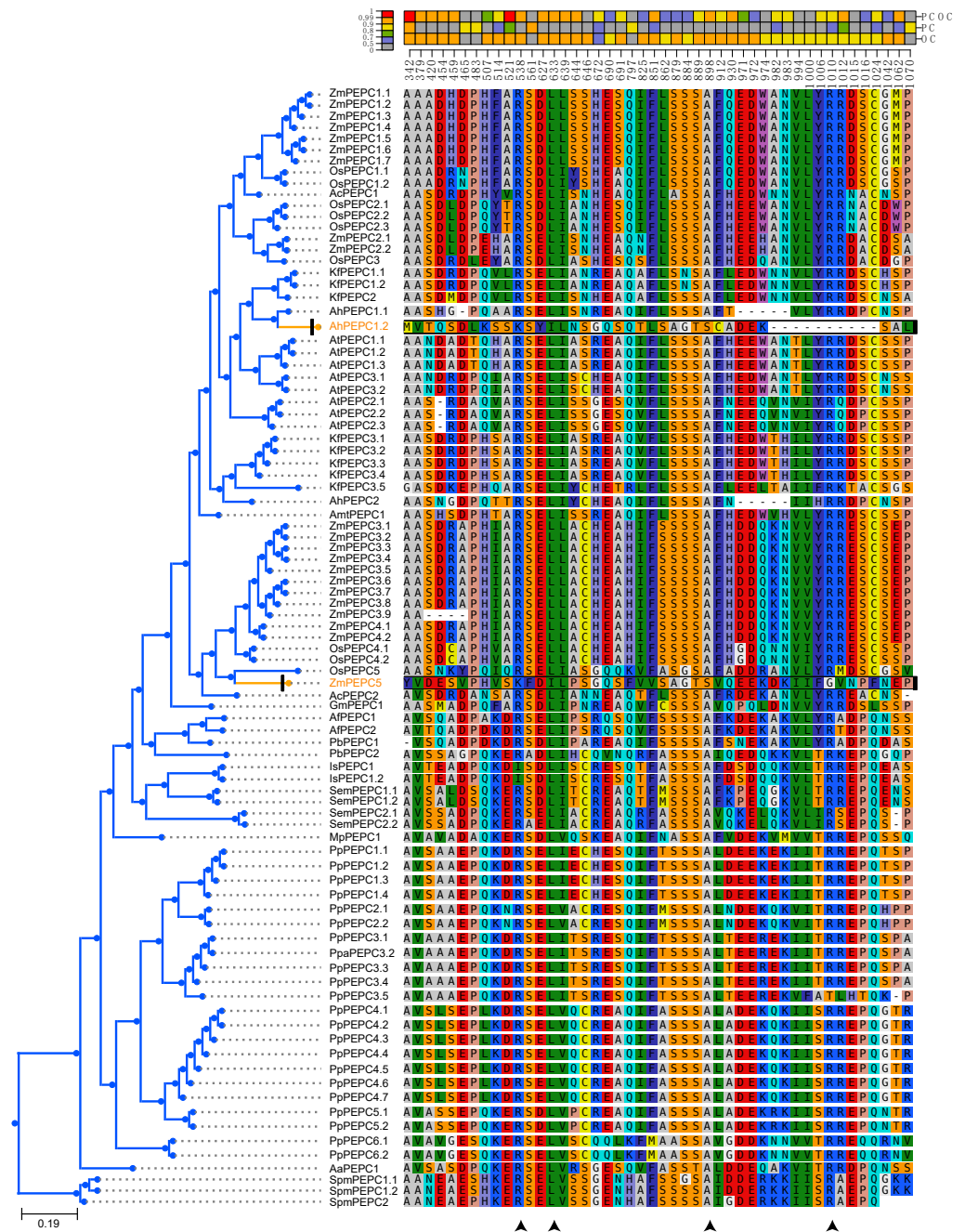
Figure 5 Four convergent substitution sites of C4-PEPC isoforms. The arrowheads at the bottom showed the positions of convergent substitutions. PCOC: Profile Change with One Change model; PC, Profile Change model; OC, One Change model, all models were in detail explained by *Rey et al. (2018)*. Posterior probabilities (pp) for the PCOC, PC, and OC models are summarized by top box colors, and the amino acid colors correspond to different amino acid equilibrium frequencies (*i.e.*, different profiles) of the Profile Change with One Change model (PCOC model). Aa, *Anthoceros angustus*; Ac, *Ananas comosus*; Af, *Azolla filiculoides*; Ah, *Amaranthus hypochondriacus*; Amt, *Amborella trichopoda*; At, *Arabidopsis thaliana*; Gm, *Gnetum montanum*; Is, *Isoetes sinensis*; Kf, *Kalanchoe fedtschenkoi*; Mp, *Marchantia polymorpha*; Os: *Oryza sativa*; Pa, *Picea abies*; Pb, *Platycerium bifurcatum*; Pp, *Physcomitrella patens*; Sm, *Selaginella moellendorffii*; Spm, *Spirogloea muscicola*; Zm, *Zea mays*.

PEPC was convergent in C₄ photosynthesis but not in CAM photosynthesis

According to the robust phylogenetic tree, PEPC gene family consists of two major subfamilies, PTPC and BTPC, which is consistent with the predictions of the conserved domains (Fig. 1, Table S1). PTPC genes perform the critical roles for initial carbon fixation in C_4 and CAM photosynthesis (*Jiao & Chollet, 1991*; *Lepiniec et al., 1994*; *Nimmo, 2000*; *Deng et al., 2016*). Therefore, PTPC gene tree was independently reconstructed and was reconciled with the species tree based on duplication-loss reconciliation (Fig. 3). The reconciliation tree showed that PTPC genes underwent at least 71 duplications and 16 losses in the evolutionary history of our sampled species, which indicated that PTPC genes arose multiple times through frequent duplication events (Fig. 3), potentially caused by WGD in the evolutionary process of green plants, especially in angiosperms (*Van de Peer, Mizrachi & Marchal, 2017*). After gene duplication, plants could respond to new environments through neo-functionalization of gene copies (*Russell, 2003*; *Cheng et al., 2018*). Previous research assumed that PEPC isoforms in C_4 andCAM species were duplicated from a non-photosynthetic PEPC gene that existed in ancestral C3 species (*Svensson, Bläsing & Westhoff, 2003*; *Christin et al., 2014*). Although there are limitations of sampling, our results indicated that no strong association was observed between PEPC gene duplication and CAM/C_4 evolution (Fig. 3), and the similar results were also found in the study of orchids PEPC genes, in which no correlations between the presence of CAM and gene duplication (*Zhang et al., 2016*). In other words, PEPC gene duplications may be important for the evolutionary origin of C_4 and CAM photosynthesis but without clear correlation. In CAM pathway, post-translational regulation of PEPC possibly might play a key role (*Jiao & Chollet, 1991*; *O'Leary, Park & Plaxton William, 2011*; *Chen et al., 2020b*).

To test whether convergent molecular evolution at the amino acid level occurred in PTPC proteins, we performed comprehensive detection of convergent sites in C_4 and CAM photosynthesis using the PTPC gene tree of green plants with the Profile Change with One Change (PCOC) pipeline, which can detect not only convergent substitutions of amino acids but also convergent shifts that correspond to convergent phenotypic changes (*Rey et al., 2018*). The results showed two convergent shifts in PTPC proteins that occurred in CAM species and three convergent shifts that occurred in C4 species, the posterior probabilities (pp) for the PCOC model were greater than 0.9 at all convergent shifts (Fig. 4). However, identical convergent substitutions were not detected in clades from both photosynthetic pathways, which indicated that identical convergent molecular evolution at the amino acid level might not occur in all copies of PTPC proteins.

Different isoforms of the PEPC gene family might perform different functions. In addition to photosynthetic functions, PEPC genes also perform hyper-diverse non-photosynthetic functions, such as response to abiotic stress, fruit maturation, seed formation and germination (*Lepiniec et al., 1994*; *Chollet, Vidal & O'Leary, 1996*; *O'Leary, Park & Plaxton William, 2011*; *Shi et al., 2015*; *Wang et al., 2016*; *Waseem & Ahmad, 2019*; *Zhao et al., 2019*). Maybe only a few of the PEPC isoforms corresponded to convergent evolution of C_4 and CAM photosynthesis. Therefore, we also detected convergent evolution at sites in different gene groups, in which each convergent phenotypic species retained

one clade or one gene copy (one-to-one) (Fig. S2). Interestingly, four identical convergent amino acid sites in one gene group (AhPEPC1.2/ZmPEPC2) were discovered in C_4 species (Fig. 5), two of which were also reported in previous studies (*Bläsing, Westhoff & Svensson, 2000*; *Christin et al., 2007*; *Besnard et al., 2009*; *Paulus, Schlieper & Groth, 2013*). The convergent amino acid mutations in the active site Ala774 and the inhibitory site Arg884 were sufficient to switch the photosynthetic function from C_3 to C_4 activity (*Paulus, Schlieper & Groth, 2013*). Due to limited sampling, our results maybe overestimate the number of convergent sites in C_4 plants, because increased sample size maybe decrease the number of inferred molecular convergence (*Thomas, Hahn & Hahn, 2017*). Therefore, the two new convergent sites of PEPC gene family in C_4 species should be verified by further studies with adequate sampling. Several convergent sites reported in the previous studies (*Christin et al., 2007*; *Besnard et al., 2009*; *Christin et al., 2014*; *Moreno-Villena et al., 2018*) were not detected in the present study, probably because these sites are only convergent in grass.

Previously, *Yang et al. (2017)* reported a convergent evolution site of PEPC gene in several CAM lineages of angiosperms, except *Ananas comosus*. However, when we detected convergent sites in the one-to-one gene groups of CAM species, no identical convergent sites were found (Fig. S2), which indicated that PEPC genes might not have identical convergent sites that resulted in photosynthetic conversion from the C_3 to CAM pathway (*Wickell et al., 2021*).

CONCLUSIONS

PEPC gene family plays a crucial role in C_4 and CAM photosynthesis and is considered to cause the convergent evolution of these CCMs. In the present study, we detected the convergent amino acid sites of PEPC gene family using relatively limited genomic datasets. In the evolutionary history of the PEPC gene family, gene duplication frequently occurred due to multiple WGD events, but no strong association was observed between PEPC gene duplication and CAM/C_4 evolution. 3D protein structures of PEPC gene family are also not associated with C_4 and CAM evolution. Additionally, four sites with convergent substitutions were detected in C4-PEPC isoforms, two of which were key functional positions to switch the photosynthetic pathway from C_3 to C_4 activity. However, no convergent sites were detected in CAM-PEPC genes. Our results indicated that convergent molecular substitutions of PEPC genes played key roles for the origin and convergent evolution of C_4 photosynthesis, but convergent evolution of CAM photosynthesis maybe not caused by convergence at the amino acid level in PEPC proteins. However, our limited sampling maybe affect the inference of molecular convergence. In future, the clearly evolutionary trajectories of PEPC genes will be clarified by more genomic data, which included more species of CAM, C_4 and C_3 relatives.

ACKNOWLEDGEMENTS

We are very grateful to the Academic Editor Stanislav Kopriva and anonymous reviewers for their helpful comments and modification assistance. and we would like to thank TopEdit for linguistic assistance during preparation of this manuscript.

Funding

This research was funded by the "Strategic Priority Research Program" of the Chinese Academy of Sciences (Grant No. XDA13020602). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Chinese Academy of Sciences: XDA13020602.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Jiang-Ping Shu conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.
- Yue-Hong Yan conceived and designed the experiments, authored or reviewed drafts of the paper, and approved the final draft.
- Rui-Jiang Wang conceived and designed the experiments, authored or reviewed drafts of the paper, obtained the funding to support this study, and approved the final draft.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:
The gene family datasets in the present study are available at FigShare: Shu, Jiang-Ping (2021): PEPC gene family data. figshare. Dataset. https://doi.org/10.6084/m9.figshare.16943797.v1.

Data Availability

The PEPC gene family data is available at FigShare: Shu, Jiang-Ping (2021): PEPC gene family data. figshare. Dataset. https://doi.org/10.6084/m9.figshare.16943797.v1.
The genomes are available as follows:
- Spirogloea muscicola: Chenglab (2019): Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. figshare. Dataset. https://doi.org/10.6084/m9.figshare.9911876.v1.
- Marchantia polymorpha: EnsemblPlants, GCA_003032435.1
https://www.ebi.ac.uk/ena/browser/view/GCA_003032435?show=assembly-stats.

- Physcomitrella patens: EnsemblPlants, GCA_000002425.2
https://www.ebi.ac.uk/ena/browser/view/GCA_000002425?show=assembly-stats.
- Anthoceros angustus: Dryad, Zhang, Jian et al. (2019), The hornwort genome and early land plant evolution, Dryad, Dataset, https://doi.org/10.5061/dryad.msbcc2ftv.
- Selaginella moellendorffii: EnsemblPlants, GCA_000143415.1
https://www.ebi.ac.uk/ena/browser/view/GCA_000143415.1.
- Azolla filiculoides: https://www.fernbase.org.
- Platycerium bifurcatum: NCBI, SRX1098245
https://www.ncbi.nlm.nih.gov/sra/SRX1098245.
- Picea abies assembly results: PlantGDB, http://www.plantgdb.org/download/download.php?dir=/Sequence/ESTcontig/Picea_abies.
- Gnetum montanum: Dryad, Wan, Tao et al. (2018), Data from: A genome for gnetophytes and early evolution of seed plants, Dryad, Dataset, https://doi.org/10.5061/dryad.0vm37.
- Amborella trichopoda: EnsemblPlants, GCA_000471905.1, https://www.ebi.ac.uk/ena/browser/view/GCA_000471905.1.
- Ananas comosus: EnsemblPlants, GCA_902162155.1, https://www.ebi.ac.uk/ena/browser/view/GCA_902162155?show=assembly-stats.
- Oryza sativa: EnsemblPlants, GCA_001433935.1, https://www.ebi.ac.uk/ena/browser/view/GCA_001433935?show=assembly-stats.
- Zea mays: EnsemblPlants, GCA_902167145.1, https://www.ebi.ac.uk/ena/browser/view/GCA_902167145.1.
- Arabidopsis thaliana: EnsemblPlants, GCA_000001735.1, https://www.ebi.ac.uk/ena/browser/view/GCA_000001735.1.
- Amaranthus hypochondriacus: Phytozome, 459, https://phytozome-next.jgi.doe.gov/info/Ahypochondriacus_v2_1.
- Kalanchoe fedtschenkoi: Phytozome, 382, https://phytozome-next.jgi.doe.gov/info/Kfedtschenkoi_v1_1.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.12828#supplemental-information.

REFERENCES

Bailey TL, Williams N, Misleh C, Li WW. 2006. MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Research* **34(suppl_2)**:W369–W373 DOI 10.1093/nar/gkl198.

Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D. 2004. Ultraconserved elements in the human genome. *Science* **304(5675)**:1321–1325 DOI 10.1126/science.1098119.

Besnard G, Muasya AM, Russier F, Roalson EH, Salamin N, Christin P-A. 2009. Phylogenomics of C4 photosynthesis in sedges (Cyperaceae): multiple appearances

and genetic convergence. *Molecular Biology and Evolution* **26**(**8**):1909–1919
DOI 10.1093/molbev/msp103.

Bläsing OE, Westhoff P, Svensson P. 2000. Evolution of C4 phosphoenolpyruvate
carboxylase in Flaveria, a conserved serine residue in the carboxyl-terminal part
of the enzyme is a major determinant for C4-specific characteristics. *Journal of
Biological Chemistry* **275**(**36**):27917–27923 DOI 10.1074/jbc.M909832199.

**Bowman JL, Kohchi T, Yamato KT, Jenkins J, Shu S, Ishizaki K, Yamaoka S, Nishihama
R, Nakamura Y, Berger F, Adam C, Aki SS, Althoff F, Araki T, Arteaga-Vazquez
MA, Balasubrmanian S, Barry K, Bauer D, Boehm CR, Briginshaw L, Caballero-
Perez J, Catarino B, Chen F, Chiyoda S, Chovatia M, Davies KM, Delmans M,
Demura T, Dierschke T, Dolan L, Dorantes-Acosta AE, Eklund DM, Florent SN,
Flores-Sandoval E, Fujiyama A, Fukuzawa H, Galik B, Grimanelli D, Grimwood
J, Grossniklaus U, Hamada T, Haseloff J, Hetherington AJ, Higo A, Hirakawa Y,
Hundley HN, Ikeda Y, Inoue K, Inoue S, Ishida S, Jia Q, Kakita M, Kanazawa T,
Kawai Y, Kawashima T, Kennedy M, Kinose K, Kinoshita T, Kohara Y, Koide E,
Komatsu K, Kopischke S, Kubo M, Kyozuka J, Lagercrantz U, Lin S, Lindquist
E, Lipzen AM, Lu C-W, De Luna E, Martienssen RA, Minamino N, Mizutani
M, Mizutani M, Mochizuki N, Monte I, Mosher R, Nagasaki H, Nakagami H,
Naramoto S, Nishitani K, Ohtani M, Okamoto T, Okumura M, Phillips J, Pollak
B, Reinders A, Rövekamp M, Sano R, Sawa S, Schmid MW, Shirakawa M, Solano
R, Spunde A, Suetsugu N, Sugano S, Sugiyama A, Sun R, Suzuki Y, Takenaka M,
Takezawa D, Tomogane H, Tsuzuki M, Ueda T, Umeda M, Ward JM, Watanabe Y,
Yazaki K, Yokoyama R, Yoshitake Y, Yotsui I, Zachgo S, Schmutz J. 2017.** Insights
into Land plant evolution garnered from the marchantia polymorpha genome. *Cell*
171(**2**):287–304.e215 DOI 10.1016/j.cell.2017.09.030.

Boxall SF, Kadu N, Dever LV, Kneřová J, Waller JL, Gould PJD, Hartwell J. 2020.
Kalanchoë PPC1 is essential for crassulacean acid metabolism and the regulation of
core circadian clock and guard cell signaling genes. *The Plant Cell* **32**(**4**):1136–1160
DOI 10.1105/tpc.19.00481.

**Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden
TL. 2009.** BLAST+: architecture and applications. *BMC Bioinformatics* **10**(**1**):421
DOI 10.1186/1471-2105-10-421.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their
use in phylogenetic analysis. *Molecular Biology and Evolution* **17**(**4**):540–552
DOI 10.1093/oxfordjournals.molbev.a026334.

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. 2020a. TBtools: an
integrative toolkit developed for interactive analyses of big biological data. *Molecular
Plant* **13**(**8**):1194–1202 DOI 10.1016/j.molp.2020.06.009.

Chen L-Y, Xin Y, Wai CM, Liu J, Ming R. 2020b. The role of cis-elements in the
evolution of crassulacean acid metabolism photosynthesis. *Horticulture Research*
7(**1**):5 DOI 10.1038/s41438-019-0229-0.

Cheng F, Wu J, Cai X, Liang J, Freeling M, Wang X. 2018. Gene retention, fractionation and subgenome differences in polyploid plants. *Nature Plants* **4(5)**:258–268 DOI 10.1038/s41477-018-0136-7.

Cheng S, Xian W, Fu Y, Marin B, Keller J, Wu T, Sun W, Li X, Xu Y, Zhang Y, Wittek S, Reder T, Günther G, Gontcharov A, Wang S, Li L, Liu X, Wang J, Yang H, Xu X, Delaux P-M, Melkonian B, Wong G, Melkonian M. 2019. Genomes of subaerial zygnematophyceae provide insights into land plant evolution. *Cell* **179(5)**:1057–1067.e1014 DOI 10.1016/j.cell.2019.10.019.

Chollet R, Vidal J, O'Leary MH. 1996. Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. *Annual Review of Plant Physiology & Plant Molecular Biology* **47(1)**:273–298 DOI 10.1146/annurev.arplant.47.1.273.

Christin P-A, Arakaki M, Osborne CP, Bräutigam A, Sage RF, Hibberd JM, Kelly S, Covshoff S, Wong GK, Hancock L, Edwards EJ. 2014. Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. *Journal of Experimental Botany* **65(13)**:3609–3621 DOI 10.1093/jxb/eru087.

Christin P-A, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 photo-synthesis evolved in grasses via parallel adaptive genetic changes. *Current Biology* **17(14)**:1241–1247 DOI 10.1016/j.cub.2007.06.036.

Comte N, Morel B, Hasic D, Guéguen L, Boussau B, Daubin V, Penel S, Scornavacca C, Gouy M, Stamatakis A, Tannier E, Parsons DP. 2020. Treerecs: an integrated phy-logenetic tool, from sequences to reconciliations. *Bioinformatics* **36(18)**:4822–4824 DOI 10.1093/bioinformatics/btaa615.

Connell MB, Lee MJY, Li J, Plaxton WC, Jia Z. 2018. Structural and biochemical characterization of citrate binding to AtPPC3, a plant-type phosphoenolpyruvate carboxylase from Arabidopsis thaliana. *Journal of Structural Biology* **204(3)**:507–512 DOI 10.1016/j.jsb.2018.11.003.

Darabi M, Seddigh S. 2018. Structural, functional, and phylogenetic characterization of phosphoenolpyruvate carboxylase (PEPC) in C4 and CAM plants. *Caryologia* **71(3)**:272–288 DOI 10.1080/00087114.2018.1465762.

Deng H, Zhang L-S, Zhang G-Q, Zheng B-Q, Liu Z-J, Wang Y. 2016. Evolution-ary history of PEPC genes in green plants: implications for the evolution of CAM in orchids. *Molecular Phylogenetics and Evolution* **94(Pt B)**:559–564 DOI 10.1016/j.ympev.2015.10.007.

Edwards EJ. 2019. Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytologist* **223(4)**:1742–1755 DOI 10.1111/nph.15851.

Fischer WW, Hemp J, Johnson JE. 2016. Evolution of oxygenic photosynthesis. *Annual Review of Earth and Planetary Sciences* **44(1)**:647–683 DOI 10.1146/annurev-earth-060313-054810.

Foster GL, Royer DL, Lunt DJ. 2017. Future climate forcing potentially without precedent in the last 420 million years. *Nature Communication* **8(1)**:14845 DOI 10.1038/ncomms14845.

Gehrig H, Heute V, Kluge M. 2001. New partial sequences of phosphoenolpyruvate carboxylase as molecular phylogenetic markers. *Molecular Phylogenetics and Evolution* **20(2)**:262–274 DOI 10.1006/mpev.2001.0973.

Gilman IS, Edwards EJ. 2020. Crassulacean acid metabolism. *Current Biology* **30(2)**:R57–R62 DOI 10.1016/j.cub.2019.11.073.

González-Segura L, Mújica-Jiménez C, Juárez-Díaz JA, Güémez-Toro R, Martinez-Castilla LP, Muñoz Clares RA. 2018. Identification of the allosteric site for neutral amino acids in the maize C4 isozyme of phosphoenolpyruvate carboxylase: The critical role of Ser-100. *Journal of Biological Chemistry* **293(26)**:9945–9957 DOI 10.1074/jbc.RA118.002884.

Heyduk K, Moreno-Villena JJ, Gilman IS, Christin P-A, Edwards EJ. 2019. The genetics of convergent evolution: insights from plant photosynthesis. *Nature Reviews Genetics* **20(8)**:485–493 DOI 10.1038/s41576-019-0107-5.

Hoang DT, Chernomor O, Haeseler Avon, Minh BQ, Vinh LS. 2017. UFBoot2: improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35(2)**:518–522 DOI 10.1093/molbev/msx281.

Initiative, OTPT. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **5747780**:679–685 DOI 10.1038/s41586-019-1693-2.

Izui K, Matsumura H, Furumoto T, Kai Y. 2004. Phosphoenolpyruvate carboxylase: a new era of structural biology. *Annual Review of Plant Biology* **55(1)**:69–84 DOI 10.1146/annurev.arplant.55.031903.141619.

Jaiswal SK, Mahajan S, Chakraborty A, Kumar S, Sharma VK. 2021. The genome sequence of Aloe vera reveals adaptive evolution of drought tolerance mechanisms. *iScience* **24(2)**:102079 DOI 10.1016/j.isci.2021.102079.

Jiao J-A, Chollet R. 1991. Posttranslational regulation of phosphoenolpyruvate carboxylase in C4 and crassulacean acid metabolism plants. *Plant Physiology* **95(4)**:981–985 DOI 10.1104/pp.95.4.981.

Kalyaanamoorthy S, Minh BQ, Wong TKF, Haeseler Avon, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14(6)**:587–589 DOI 10.1038/nmeth.4285.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30(4)**:772–780 DOI 10.1093/molbev/mst010.

Keeley JE. 1981. Isoetes Howellii: a submerged aquatic cam plant?. *American Journal of Botany* **68(3)**:420–424 DOI 10.1002/j.1537-2197.1981.tb06380.x.

Lepiniec L, Vidal J, Chollet R, Gadal P, Crétin C. 1994. Phosphoenolpyruvate carboxylase: structure, regulation and evolution. *Plant Science* **99(2)**:111–124 DOI 10.1016/0168-9452(94)90168-6.

Li F-W, Brouwer P, Carretero-Paulet L, Cheng S, Vries Jde, Delaux P-M, Eily A, Koppers N, Kuo L-Y, Li Z, Simenc M, Small I, Wafula E, Angarita S, Barker MS, Bräutigam A, de Pamphilis C, Gould S, Hosmani P.S.Huang, Y-M, Huettel B, Kato Y, Liu X, Maere S, McDowell R, Mueller LA, Nierop KGJ, Rensing SA, Robison T, Rothfels CJ, Sigel EM, Song Y, Timilsena PR, VandePeer Y, Wang H, Wilhelmsson

PKI, Wolf PG, Xu X, Der JP, Schluepmann H, Wong GKS, Pryer KM. 2018. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* **4(7)**:460–472 DOI 10.1038/s41477-018-0188-8.

Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler GH, Song JS, Thanki N, Yamashita RA, Yang M, Zhang D, Zheng C, Lanczycki CJ, Marchler-Bauer A. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Research* **48(D1)**:D265–D268 DOI 10.1093/nar/gkz991.

Matsumura H, Xie Y, Shirakata S, Inoue T, Yoshinaga T, Ueno Y, Izui K, Kai Y. 2002. Crystal structures of C4 form maize and quaternary complex of e. coli phosphoenolpyruvate carboxylases. *Structure* **10(12)**:1721–1730 DOI 10.1016/S0969-2126(02)00913-9.

Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Research* **41(12)**:e121 DOI 10.1093/nar/gkt263.

Moreno-Villena J, Dunning L, Osborne C, Christin P-A. 2018. Highly expressed genes are preferentially co-opted for C4 photosynthesis. *Molecular Biology and Evolution* **35(1)**:94–106 DOI 10.1093/molbev/msx269.

Nguyen LT, Schmidt HA, Haeseler Avon, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32(1)**:268–274 DOI 10.1093/molbev/msu300.

Nimmo HG. 2000. The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends in Plant Science* **5(2)**:75–80 DOI 10.1016/S1360-1385(99)01543-5.

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Å Olson, Rilakovic N, Ritland C, Rosselló JA, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu Z-Q, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Lee Thompson S, VandePeer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S. 2013. The Norway spruce genome sequence and conifer genome evolution. *Nature* **4977451**:579–584 DOI 10.1038/nature12211.

O'Leary B, Park J, Plaxton William C. 2011. The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs. *Biochemical Journal* **436(1)**:15–34 DOI 10.1042/BJ20110078.

Paulus JK, Schlieper D, Groth G. 2013. Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nature Communications* **4(1)**:1518 DOI 10.1038/ncomms2504.

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y, Tanahashi T, Sakakibara K, Fujita

T, Oishi K, Shin IT, Kuroki Y, Toyoda A, Suzuki Y, Hashimoto S, Yamaguchi K, Sugano S, Kohara Y, Fujiyama A, Anterola A, Aoki S, Ashton N, Barbazuk WB, Barker E, Bennetzen JL, Blankenship R, Cho SH, Dutcher SK, Estelle M, Fawcett JA, Gundlach H, Hanada K, Heyl A, Hicks KA, Hughes J, Lohr M, Mayer K, Melkozernov A, Murata T, Nelson DR, Pils B, Prigge M, Reiss B, Renner T, Rombauts S, Rushton PJ, Sanderfoot A, Schween G, Shiu S-H, Stueber K, Theodoulou FL, Tu H, VandePeer Y, Verrier PJ, Waters E, Wood A, Yang L, Cove D, Cuming AC, Hasebe M, Lucas S, Mishler BD, Reski R, Grigoriev IV, Quatrano RS, Boore JL. 2008. The physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* **319**5859:64–69 DOI 10.1126/science.1150646.

Rey C, Guéguen L, Sémon M, Boussau B. 2018. Accurate detection of convergent amino-acid evolution with PCOC. *Molecular Biology and Evolution* **35(9)**:2296–2306 DOI 10.1093/molbev/msy114.

Ruiz-Ballesta I, Baena G, Gandullo J, Wang L, She Y-M, Plaxton WC, Echevarria C. 2016. New insights into the post-translational modification of multiple phospho-enolpyruvate carboxylase isoenzymes by phosphorylation and monoubiquitination during sorghum seed development and germination. *Journal of Experimental Botany* **67(11)**:3523–3536 DOI 10.1093/jxb/erw186.

Russell KM. 2003. Gene duplication, neofunctionalization, and the evolution of C4 photosynthesis. *International Journal of Plant Sciences* **164**:S43–S54 DOI 10.1086/368400.

Rut G, Krupa J, Miszalski Z, Rzepka A, Ślesak I. 2008. Crassulacean acid metabolism in the epiphytic fern Patycerium bifurcatum. *Photosynthetica* **46(1)**:156–160.

Sage RF. 2016. A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *Journal of Experimental Botany* **67(14)**:4039–4056 DOI 10.1093/jxb/erw156.

Sage RF, Christin P-A, Edwards EJ. 2011. The C4 plant lineages of planet Earth. *Journal of Experimental Botany* **62(9)**:3155–3169 DOI 10.1093/jxb/err048.

Schlüter U, Weber APM. 2020. Regulation and evolution of C4 photosynthesis. *Annual Review of Plant Biology* **71(1)**:183–215 DOI 10.1146/annurev-arplant-042916-040915.

Shi J, Yi K, Liu Y, Xie L, Zhou Z, Chen Y, Hu Z, Zheng T, Liu R, Chen Y, Chen J. 2015. Phosphoenolpyruvate carboxylase in arabidopsis leaves plays a crucial role in carbon and nitrogen metabolism. *Plant Physiology* **167(3)**:671–681 DOI 10.1104/pp.114.254474.

Silvera K, Neubig KM, Whitten WM, Williams NH, Winter K, Cushman JC. 2010. Evolution along the crassulacean acid metabolism continuum. *Functional Plant Biology* **37(11)**:995–1010 DOI 10.1071/fp10084.

Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms. *Current Opinion in Plant Biology* **30**:159–165 DOI 10.1016/j.pbi.2016.03.015.

Svensson P, Bläsing OE, Westhoff P. 2003. Evolution of C4 phosphoenolpyruvate carboxylase. *Archives of Biochemistry and Biophysics* **414(2)**:180–188 DOI 10.1016/S0003-9861(03)00165-6.

Thomas GWC, Hahn MW, Hahn Y. 2017. The effects of increasing the number of taxa on inferences of molecular convergence. *Genome Biology and Evolution* **9(1)**:213–221 DOI 10.1093/gbe/evw306.

Van de Peer Y, Mizrachi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* **18(7)**:411–424 DOI 10.1038/nrg.2017.26.

Wai CM, Weise SE, Ozersky P, Mockler TC, Michael TP, Van Buren R. 2019. Time of day and network reprogramming during drought induced CAM photosynthesis in Sedum album. *PLOS Genetics* **15(6)**:e1008209 DOI 10.1371/journal.pgen.1008209.

Wang N, Zhong X, Cong Y, Wang T, Yang S, Li Y, Gai J. 2016. Genome-wide analysis of phosphoenolpyruvate carboxylase gene family and their response to abiotic stresses in soybean. *Scientific Reports* **6(1)**:38448 DOI 10.1038/srep38448.

Waseem M, Ahmad F. 2019. The phosphoenolpyruvate carboxylase gene family identification and expression analysis under abiotic and phytohormone stresses in Solanum lycopersicum L. *Gene* **690**:11–20 DOI 10.1016/j.gene.2018.12.033.

Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* **46(W1)**:W303–W303 DOI 10.1093/nar/gky427.

Wickell D, Kuo L-Y, Yang H-P, Dhabalia-Ashok A, Irisarri I, Dadras A, De Vries S, de Vries J, Huang Y-M, Li Z, Barker MS, Hartwick NT, Michael TP, Li F-W. 2021. Underwater CAM photosynthesis elucidated by *Isoetes* genome. *Nature Communication* **12(1)**:6348 DOI 10.1038/s41467-021-26644-7.

Winter K, Garcia M, Virgo A, Smith JAC. 2021. Low-level CAM photosynthesis in a succulent-leaved member of the Urticaceae, Pilea peperomioides. *Functional Plant Biology* **48(7)**:683–690 DOI 10.1071/FP20151.

Yang X, Hu R, Yin H, Jenkins J, Shu S, Tang H, Liu D, Weighill DA, Yim WCheol, Ha J, Heyduk K, Goodstein DM, Guo HB, Moseley RC, Fitzek E, Jawdy S, Zhang Z, Xie M, Hartwell J, Grimwood J, Abraham PE, Mewalal R, Beltran JD, Boxall SF, Dever LV, Palla KJ, Albion R, Garcia T, Mayer JA, Don Lim S, Man Wai C, Peluso P, Van Buren R, De Paoli HC, Borland AM, Guo H, Chen JG, Muchero W, Yin Y, Jacobson DA, Tschaplinski TJ, Hettich RL, Ming R, Winter K, Leebens-Mack JH, Smith JAC, Cushman JC, Schmutz J, Tuskan GA. 2017. The Kalanchoë genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications* **8(1)**:1899 DOI 10.1038/s41467-017-01491-7.

Yang X, Liu D, Tschaplinski TJ, Tuskan GA. 2019. Comparative genomics can provide new insights into the evolutionary mechanisms and gene function in CAM plants. *Journal of Experimental Botany* **70(22)**:6539–6547 DOI 10.1093/jxb/erz408.

Zhang L, Chen F, Zhang G-Q, Zhang Y-Q, Niu S, Xiong J-S, Lin Z, Cheng Z-M, Liu Z-J. 2016. Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. *The Plant Journal* **86(2)**:175–185 DOI 10.1111/tpj.13159.

Zhang J, Fu X-X, Li R-Q, Zhao X, Liu Y, Li M-H, Zwaenepoel A, Ma H, Goffinet B, Guan Y-L, Xue J-Y, Liao Y-Y, Wang Q-F, Wang Q-H, Wang J-Y, Zhang G-Q,

Wang Z-W, Jia Y, Wang M-Z, Dong S-S, Yang J-F, Jiao Y-N, Guo Y-L, Kong H-Z, Lu A-M, Yang H-M, Zhang S-Z, Van dePeer Y, Liu Z-J, Chen Z-D. 2020. The hornwort genome and early land plant evolution. *Nature Plants* **6**(2):107–118 DOI 10.1038/s41477-019-0588-4.

Zhao Y, Guo A, Wang Y, Hua J. 2019. Evolution of PEPC gene family in Gossypium reveals functional diversification and GhPEPC genes responding to abiotic stresses. *Gene* **698**:61–71 DOI 10.1016/j.gene.2019.02.061.