

Convergent Molecular Evolution of PEPC Gene Family in C4 and CAM Plants

Jiang-Ping Shu

South China Botanical Garden <https://orcid.org/0000-0001-9714-3064>

Yue-Hong Yan

Key Laboratory of National Forestry and Grassland Administration for Orchid Conservation and Utilization

Rui-Jiang Wang (✉ wangrj@scbg.ac.cn)

University of Chinese Academy of Sciences

Research article

Keywords: Phosphoenolpyruvate carboxylase, convergent evolution, carbon-concentrating mechanisms, photosynthetic convergence, phylogeny

Posted Date: November 18th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-106970/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Phosphoenolpyruvate carboxylase (PEPC), as the key enzyme in initial carbon fixation of C₄ and crassulacean acid mechanism (CAM) pathways, was thought to undergo convergent adaptive changes resulted in the convergent evolution of C₄ and CAM photosynthesis in vascular plants. However, the integral evolutionary history and convergence of PEPC in plants remained lack of understanding.

Results: In present study, we identified the members of PEPC gene family across green plants with genomic data, found ten conserved motifs and modeled three-dimensional protein structures of 90 plant-type PEPC genes. After reconstructed PEPC gene family tree and reconciled with species tree, we found PEPC genes occurred 71 gene duplications and 16 gene losses, which might result from whole-genome duplication events in plants. Based on the integral phylogenetic tree of PEPC gene family, we detected four convergent evolution sites of PEPC in C₄ species but no one in CAM species.

Conclusions: PEPC gene family was ubiquitous and highly conservative in green plants. After originated from gene duplication of ancestral C₃-PEPC, C₄-PEPC isoforms underwent convergent molecular substitution that facilitated the convergent evolution of C₄ photosynthesis in Angiosperms, but PEPC gene did not exist molecular convergence corresponded to the multiple independent evolution of CAM photosynthesis. Our findings help to understand the origin and evolution of C₄ and CAM pathways and shed new light on the adaptation of plants in drought and high-temperature habitats.

Background

Plant photosynthesis is the most important biochemical process on our planet, and it provides energy, food and oxygen for survival and reproduction of all heterotrophic organisms, including our humans [1]. Since its origin, the oxygen levels gradually rose and carbon dioxide (CO₂) concentration decreased in the atmosphere of Earth [2], which enhanced photorespiration and led to losses of bioenergy and carbohydrate [3]. To obtain adequate CO₂ and improve photosynthetic efficiency, plants developed several carbon-concentrating mechanisms (CCMs), such as C₄ photosynthesis and crassulacean acid metabolism (CAM), to adapt the suddenly decline of the atmospheric CO₂ concentration at ~ 350 Mya [4, 5]. C₄ photosynthesis, as the notable examples of convergent evolution, has been repeatedly evolved more than 60 times in flowering plants [6, 7], whereas CAM pathway has been independently occurred in ~ 35 families of vascular plants [8]. Recently, several comparative genomics researches provided new insights into the genetics and evolution of CCMs [4, 9–11]. However, the molecular mechanisms underlying CCMs convergent evolution remained poorly understand.

In plants, phosphoenolpyruvate carboxylase (PEPC; EC 4.1.1.31) is a key enzyme that catalyzes the primary fixation reaction of photosynthesis CO₂ assimilation in C₄ and CAM photosynthesis [12–14]. Additionally, PEPC genes also play crucial roles in a variety of non-photosynthesis functions, such as carbon and nitrogen metabolism [15], seed development and germination [16], abiotic stresses [17, 18] and so on [12, 19, 20]. As the so-called housekeeping functions, PEPC genes are highly conserved, which makes it valuable molecular markers for phylogenetic reconstruction of plants [21]. Although PEPC genes share origins during the evolution of C₄ and CAM photosynthesis, these genes are distinct copies generated by whole genome duplication in angiosperms [22].

Interestingly, PEPC genes were essential for the regulation of core circadian clock in CAM photosynthesis [23], and shared convergent amino acid changes in diverse CAM species [9]. In C₄ grasses, PEPC genes also underwent parallel adaptive genetic changes [10, 24]. Therefore, PEPC genes were crucial for elucidating the origin and convergent evolution of C₄ and CAM photosynthesis. However, PEPC genes were extensively distributed in all plants, previous studies, which only performed in a few lineages of angiosperms, could not sufficiently elucidate the origin and evolution of PEPC gene family [25]. Furthermore, CAM photosynthesis was widely evolved in the major clades of vascular plants, including pteridophytes (lycophods and ferns), gymnosperms and angiosperms, while C₄ photosynthesis was only distributed in angiosperms. Therefore, to understand the convergent molecular evolution of PEPC gene family in plant kingdom, it was urgently required that samples with genomic datasets could sufficiently represent the species of C₃, C₄ and CAM photosynthesis across the major lineages of plants.

Fortunately, up to now, more than 400 plant genomes were sequenced (collected in plaBiPD, <https://www.plabipd.de/index.ep>), which provided us an excellent opportunity to resolve the origin and convergent evolution of C₄ and CAM in plants [4, 11, 26]. In

present study, we extracted PEPC genes from 17 omics data, which consisted of C3, C4 and CAM species across algae, bryophytes, pteridophytes, gymnosperms and angiosperms, and reconstructed the evolutionary history and molecular convergence of PEPC genes in C4 and CAM plants. Our findings help to elucidate the origin and evolution of C4 and CAM plants and shed new lights on adaptation of plants in drought and high temperature environments.

Results And Discussion

PEPC gene family was ubiquitous and conserved in plants

PEPC is a highly regulated enzyme that catalyzes the irreversible β -carboxylation of phosphoenolpyruvate in the presence of bicarbonate and Mg^{2+} to yield oxaloacetate and Pi, a reaction that serves a variety of physiological functions in plants [12, 27, 28]. As the key carboxylase, PEPC genes were widely distributed in green plants (Table 1). In present study, we identified 264 homologous genes using Arabidopsis PEPC genes (At1g53310, At1g68750, At2g42600 and At3g14940) as the BLAST queries and 179 homologous genes by searching with the Pfam seed of PEPCase domain (PF00311) based on 17 genomic data across all of the green plants, respectively. After combination of two gene sets, we obtained 179 common PEPC homologous genes and then searched conserved domains based on the conserved domain database, 109 genes contained conserved domains, in which plant-type PEPC (PTPC) gene domain (PEPCase) was specifically identified in 90 genes and bacterial-type PEPC (BTPC) gene domain (PRK00009) was found in 19 genes, and 70 genes were contained other PEPCase superfamily domains (Table 1, S1).

Table 1
The information of samples used in this study.

ID	Species	Class	Type	PTPC	BTPC	PEPCase superfamily	Data resources
Smu	<i>Spirogloea muscicola</i>	Algae	C3-like	3	2	4	Figshare [29]
Mpo	<i>Marchantia polymorpha</i>	Bryophytes	C3-like	1	1	1	EnsemblPlants [30]
Ppa	<i>Physcomitrella patens</i>	Bryophytes	C3-like	22	0	10	EnsemblPlants [31]
Aan	<i>Anthoceros angustus</i>	Bryophytes	C3-like	1	1	2	Dryad [32]
Smo	<i>Selaginella moellendorffii</i>	Lycopods	C3	4	4	0	EnsemblPlants [33]
Ien	<i>Isoetes sinensis</i>	Lycopods	CAM	2	5	0	Yan <i>et al.</i> unpublished data
Afi	<i>Azolla filiculoides</i>	Ferns	C3	2	0	3	Fernbase [34]
Pbi	<i>Platyserium bifurcatum</i> *	Ferns	CAM	2	0	2	GigaDB [35]
Pab	<i>Picea abies</i>	Gymnosperms	C3	0	0	10	Congenie [36]
Gmo	<i>Gnetum montanum</i>	Gymnosperms	C3	1	1	0	Dryad [37]
Atr	<i>Amborella trichopoda</i>	Angiosperms	C3	1	1	0	EnsemblPlants [38]
Aco	<i>Ananas comosus</i>	Angiosperms	CAM	2	0	1	EnsemblPlants [39]
Osa	<i>Oryza sativa</i>	Angiosperms	C3	9	1	4	EnsemblPlants [40]
Zma	<i>Zea mays</i>	Angiosperms	C4	22	1	33	EnsemblPlants [41]
Ath	<i>Arabidopsis thaliana</i>	Angiosperms	C3	8	1	0	EnsemblPlants [42]
Ahy	<i>Amaranthus hypochondriacus</i>	Angiosperms	C4	3	0	0	Phytozome [43]
Kfe	<i>Kalanchoe fedtschenkoi</i>	Angiosperms	CAM	8	1	0	Phytozome [9]

Note: PTPC: plant-type PEPC; BTPC: bacterial-type PEPC.

The copy number of PEPC genes was various in different clade of green plants. BTPC genes were distributed in 11 of 17 species in this study and retained relatively lower copies than PTPC genes (Table 1). However, due to missing the specific conserved domain of PEPCase, we have not detected PEPC gene in Norway spruce (*Picea abies*), perhaps owing to numerous pseudogenization and insertion of transposable element in conifers [36] and 10 homologs of PEPC in Norway spruce contained the PEPCase superfamily domain, which probably acted as the PEPC-like physiological functions. Thanks to arising numerous whole-genome duplications [44], PEPC gene family had relatively more copies in angiosperms, especially in maize which contained 22 PEPC gene copies. Interestingly, mosses (*Physcomitrella patens*) also retained 22 gene copies but its sister clades, hornworts (*Anthoceros angustus*) and liverworts (*Marchantia polymorpha*), only had one remnant of PEPC gene family (Table 1). This extreme difference of gene content corresponded to their different adaptation strategies of plant terrestrialization, mosses has occurred whole genome duplication events (WGDs) to increase gene family complexity for coordinating multicellular growth and dehydration response [31], however, liverworts have ancient dimorphic sex chromosomes resulted in a lack of WGDs and reduced proliferation of regulatory genes [30], and the genome of *A. angustus* was interestingly simplified and obtained stress-response and metabolic pathways genes through horizontal gene transfer from bacteria or fungi, which probably assisted their survival in a terrestrial environment [32].

PEPC genes displayed highly conserved amino acid sequence in all green plants. Here, we predicted ten conserved motifs without overly similar pairs from 109 PEPC proteins and found that the length of all motifs was more than 29 amino acids and each motif was covered more than 104 of 109 PEPC protein sequences (Table 2). Additionally, most sites of amino acid were exactly the same in all motifs and the linear order of these motifs, especially in PTPC genes, was also identical across all of green plants, although some motifs repeated in various genes (Fig. 1). These results clearly indicated that PEPC genes family was extremely conserved throughout its evolutionary history of more than 500 Mya since origin from algae [14, 20, 28].

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	RKPSGGIESLRAIPWIFAWTQTRFHLPVW
2	50	109	14268	9.6e-5072	[VI]KLTMFHGRGG[TS]VGRGGGPTHLAILSQPPDT[IV][HN]GSLRVT[VI]QGIEVIEQSFG
3	44	108	12398	9.6e-4364	SSWMGGDRDGNPRVTPEVTRDVCLLAR[ML]MAANLYFSQIEDLMFE
4	50	104	13093	2.6e-4568	IQAA[FW]RTDEIRRT[PQ]PTPQDEMAG[ML]SYFHETIW[KN]G[VL]PKFLRRVDTALKNI
5	37	105	10252	2.1e-3574	FSI[DE]WY[RL]NRINGKQEVMIQYSDSGKDAGR[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]NIGSRPSK
9	41	108	10512	6.4e-3599	[TM]L[QR][EA]MYN[EQ]WPFFRVTIDL[VI]EMVFAKGDP[GR]IAALYDKLLVS[ED]
10	29	108	7470	7.5e-2482	EEHLCFRTLQR[FY]TAATLEHGMHPPI[SA]PKP

*LLR: log likelihood rate.

However, so conserved gene family performed hyper-diverse housekeeping functions including photosynthesis and non-photosynthesis for survival in the terrestrial environments [19]. To understand the function and molecular mechanisms of PEPC gene family, the three-dimensional (3D) structure of PEPC proteins were elucidated by X-ray crystallographic analysis [45–48], which discovered many structure-function relationships of PEPC catalysis, allosteric control and regulatory phosphorylation [49]. In this study, we modeled 3D structures of 90 PTPC proteins using SWISS-MODEL server (Figure S1) and predicted three templates of protein structures, 5vyj.1.A (Fig. 2A) [48], 3zgb.1.A (Fig. 2B) [46] and 5fdn.1.A (Fig. 2C) [47], with higher seq identify (Table S2). All PTPC proteins were tetrameric enzyme with three kinds of 3D structures, in which, the template of 3zgb.1.A was widely distributed in all species of this study, but 5vyj.1.A template was distributed in seven species and 5fdn.1.A template was only distributed in Arabidopsis (Fig. 3, Table S2). In the widespread template 3zgb.1.A, compared to C3-PEPC, C4-PEPC isoforms carried two amino acid substitution to increase PEP saturation kinetics and reduce inhibitor affinity, respectively [46, 50]. Therefore, the efficiency of photosynthetic carbon fixation was greatly improved in C4 plants.

PEPC was convergent in C4 but not in CAM photosynthesis

The C4 and CAM photosynthetic pathways evolved independently more than 60 times from different phylogenetic lineages in angiosperms and vascular plants, respectively [4, 6, 8]. However, the molecular mechanisms underlying the convergent evolution of these two carbon concentration mechanisms were poorly understood. Recently, comparative genomic analyses indicated that

specific amino acid substitutions at few key sites could lead to highly predictable convergent evolution events [9, 51, 52]. In C4 and CAM plants, PEPC enzyme catalyzed the primary carbon fixation to improve CO₂ concentration at the activate site of the ribulose 1,5-bisphosphate carboxylase oxygenase (RuBisCO), which increased water-use efficiency and utilization of nitrogen and other mineral nutrients [3]. Interestingly, PEPC 1 is very important for regulating the core circadian clock in CAM photosynthetic pathway of *Kalanchoe laxiflora* [23], and PEPC 2 of *Kalanchoe fedtschenkoi* shared convergent amino acid substitution with diverse CAM species [9]. Moreover, PEPC genes did also undergo parallel adaptive genetic changes in C4 grasses [10, 24]. Therefore, PEPC genes probably played crucial roles in the convergent evolution of C4 and CAM photosynthesis [12–14, 19, 25, 28, 49].

Here, we reconstructed the phylogenetic tree of PEPC gene family in green plants with relatively adequate sampling involved the species of C3, C4 and CAM photosynthetic pathways across the major lineages of green plants (Fig. 1). PEPC gene family consisted of two major subfamilies, PTPC and BTPC, the former of which performed the critical roles for initial carbon fixation in C4 and CAM photosynthesis [12, 13, 25, 27]. Therefore, the evolutionary history of PTPC genes with maximum likelihood was reconstructed and reconciled with species tree based on duplication-loss reconciliation (Fig. 3). The reconciliation tree showed PTPC genes underwent, at least, 71 duplications and 16 losses in the evolutionary history of sampled species in present study, which indicated PTPC genes occurred multiple times large-scale duplication events, maybe caused by whole-genome duplication in the evolutionary process of green plants, especially in angiosperms (Fig. 3) [53]. Gene duplication was critical for plant to response new environments through neo-functionalization of gene copies [54, 55]. Previous researches assumed PEPC isoforms in C4 and CAM species were originated from a non-photosynthetic PEPC gene that already exists C3 ancestral species [14, 22], our results however indicated that PEPC gene duplications correlated with presence of C4 but not with CAM photosynthesis (Fig. 3), which was also found in orchids [56]. In other words, PEPC gene duplications were very important for the evolutionary origination of C4 photosynthesis, but not for CAM pathways, in which posttranslational regulation of PEPC possibly played key roles [19, 27, 57].

To test whether convergent molecular evolution at amino acid level existed in PTPC proteins, we performed the comprehensive detection of convergent evolution sites in C4 and CAM photosynthesis based on the PTPC gene tree of green plants with PCOC pipeline, which could detect not only convergent substitutions to exact same amino acid but also convergent shifts correspond to convergent phenotypic changes [58]. Respectively, we detected convergent sites in all phenotypic convergent clades of CAM and C4 photosynthesis. The results showed two convergent shifts existed in PTPC proteins of CAM species and three convergent shifts existed in C4 species, but identical convergent substitution was not detected in clades of both photosynthetic pathways (Fig. 4), which indicated convergent molecular evolution at amino acid level did not exist in all copies of PTPC proteins.

In addition to photosynthetic functions, PEPC genes also performed hyper-diverse non-photosynthetic functions, such as abiotic stress, fruit maturation, seed formation and germination, and so on [12, 15, 17–19, 28, 59]. In other words, different isoforms of PEPC gene family might perform different functions, and only a few of PEPC isoforms was corresponded to convergent evolution of C4 and CAM photosynthesis. Therefore, we further detected convergent evolution sites of different gene groups, in which each convergent phenotypic species retained one clade or one gene copy (one-to-one). Interestingly, four convergent amino acid sites in one gene group (Ahy03/Zma11) were discovered in C4 species, two of which were also confirmed in previous studies [10, 24, 46, 50]. The convergent amino acid mutations in the active site Ala774 (position 898 in Fig. 5) and the inhibitory site Arg884 (position 1010 in Fig. 5) were sufficient to switch the photosynthetic function from C3 to C4 activity [46]. However, when we detected convergent evolution sites of the one-to-one gene groups in CAM species, none of identical convergent sites were found, which indicated PEPC gene did not exist identical convergent sites resulted in photosynthetic conversion from C3 to CAM pathway.

Conclusions

In summary, PEPC gene, as the key enzyme, was a ubiquitous, highly conserved in green plants, and consisted of two major subfamilies, BTPC and PTPC, the latter of which played crucial roles for the initial carbon fixation in C4 and CAM photosynthetic pathways. In the evolutionary history of PEPC gene family, gene duplication events were frequently occurred due to multiple whole-genome duplication events, which resulted in the emergence of C4 photosynthesis after gene neo-functionalization, but not correlated to CAM photosynthesis. Additionally, four convergent substitution sites, two of which were the key functional positions to switch the photosynthetic pathway from C3 to C4 activity, were detected in C4-PEPC isoforms, however, none of convergent sites were detected in CAM-PEPC genes. Our results indicated that convergent molecular substitutions of PEPC gene played key roles for

origin and convergent evolution of C4 photosynthesis, but the convergent evolution of CAM photosynthesis was not caused by PEPC gene convergence at amino acid level.

Methods

Data acquisition

C4 and CAM photosynthesis were widely distributed in angiosperms and vascular plants, respectively [4]. To obtain sufficiently representative sampling, we selected 17 plant species with all kind of photosynthetic pathways (included C3, C4 and CAM), including *Spirogloea muscicola*, which was closest algae to land plants [29], three bryophytes, two lycopods, two ferns, two gymnosperms and seven angiosperms. 17 genome datasets were obtained from public database, such as EnsemblPlants (<http://plants.ensembl.org>), Phytozome (v13, <https://phytozome-next.jgi.doe.gov>), Dryad (<https://datadryad.org>) and several specific databases. The detailed information of samples was showed on Table 1.

Gene family identification

To identify the members of PEPC gene family, we created a local BLAST database with protein sequences of 17 plants species, then performed the BLASTp searches with default parameters using Arabidopsis PEPC amino acid sequences, including At1g53310, At1g68750, At2g42600 and At3g14940 as queries. Furthermore, the seed file of the PEPCase domain (PF00311), obtained from the Pfam website (<http://pfam.xfam.org/>), was used to build the HMMER profile, then candidate PEPC genes was searched in 17 genomic datasets by HMMER v3.2.1 [60]. After combined the two PEPC gene sets, the conserved domain search was performed using the Batch CD-search Tool with the default parameters in the conserved domain database [61], and the genes with the PEPCase conserved domain were identified as the reliable PEPC genes of plant type.

Conserved motifs prediction and proteins structures modeling

After identified the reliable PEPC genes with strict pipeline, ten conserved motifs were predicted by MEME v5.1.0 with the default parameters [62]. Motif alignment was performed by MAST v5.1.0 with default parameters [63] and conserved motifs were visualized by TBtools v1.046 [64]. Three-dimensional structures of PEPC proteins were predicted by SWISS-MODEL [65], the optimal protein model was selected with highest value of GMQE score and seq identify, which indicated highest reliability.

Phylogenetic reconstruction and gene-species tree reconciliation

To understand the evolution of PEPC gene family, the sequence alignment of PEPC amino acid sequences was performed by MAFFT v7.453 with most accurate strategy L-INS-i and the maximum number of iterative refinements was 1000 [66]. The conserved blocks were selected by GBLOCKS 0.91b [67] with the parameters (-b4 = 5 -b5 = h), then we reconstructed the PEPC gene family tree with maximum likelihood by IQ-TREE v1.6.11 [68] using the best-fit aa model JTTDCMut + R5 detected by ModelFinder [69] based on Bayesian information criterion, and the ultrafast bootstrap approximation was calculated by 1000 random replicates [70]. Phylogenetic reconciliation of gene tree and species tree was performed by Treerecs [71] with the default parameters, the species tree was based on recently phylogenomic reconstruction of green plants [72].

Convergent substitution detection

Due to the multiple copies of PEPC gene family, putative convergent clades or gene copies were labeled on phylogenetic tree with three kinds of gene combinations, including (1) labeled all gene copies of species with the convergent phenotypes (C4 or CAM), (2) labeled only one clade within each convergent species, (3) labeled only one gene copy within each convergent species. For each combination, convergent substitution detection was performed to identify convergent amino acid sites of PEPC genes in C4 and CAM plants based on the Profile Change with One Change (PCOC) model with 0.8 posterior probability threshold [58].

Abbreviations

PEPC: Phosphoenolpyruvate carboxylase; CAM: Crassulacean acid mechanism; CCMs: Carbon-concentrating mechanisms; PTPC: Plant-type PEPC; BTPC: Bacterial-type PEPC; WGD: Whole-genome duplication; 3D: Three-dimensional; RuBisCO: Ribulose 1,5-bisphosphate carboxylase oxygenase.

Declarations

Author Contributions

JPS performed all analysis, made figures and wrote the manuscript. RJW and YHY conceived the study and revised the manuscript. RJW obtained the funding to support this study. All authors read and approved the final manuscript.

Funding

This research was funded by the “Strategic Priority Research Program” of the Chinese Academy of Sciences (Grant No. XDA13020602).

Availability of data and materials

All data generated or analyzed during this study are included in this published article.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

1. Fischer WW, Hemp J, Johnson JE. Evolution of Oxygenic Photosynthesis. *Annual Review of Earth and Planetary Sciences* 2016, 44(1):647-683.
2. Foster GL, Royer DL, Lunt DJ. Future climate forcing potentially without precedent in the last 420 million years. *Nature Communications* 2017, 8(1):14845.
3. Schlüter U, Weber APM. Regulation and Evolution of C4 Photosynthesis. *Annu Rev Plant Biol* 2020, 71(1):183-215.
4. Heyduk K, Moreno-Villena JJ, Gilman IS, *et al.* The genetics of convergent evolution: insights from plant photosynthesis. *Nature Reviews Genetics* 2019, 20(8):485-493.
5. Edwards EJ, Evolutionary trajectories, accessibility and other metaphors: the case of C4 and CAM photosynthesis. *New Phytol* 2019, 223(4):1742-1755.
6. Sage RF, Christin P-A, Edwards EJ. The C4 plant lineages of planet Earth. *J Exp Bot* 2011, 62(9):3155-3169.
7. Sage RF. A portrait of the C4 photosynthetic family on the 50th anniversary of its discovery: species number, evolutionary lineages, and Hall of Fame. *J Exp Bot* 2016, 67(14):4039-4056.

8. Silvera K, Neubig KM, Whitten WM, *et al.* Evolution along the crassulacean acid metabolism continuum. *Funct Plant Biol* 2010, 37(11):995-1010.
9. Yang X, Hu R, Yin H, *et al.* The *Kalanchoë* genome provides insights into convergent evolution and building blocks of crassulacean acid metabolism. *Nature Communications* 2017, 8(1):1899.
10. Besnard G, Muasya AM, Russier F, *et al.* Phylogenomics of C4 Photosynthesis in *Sedges* (Cyperaceae): Multiple Appearances and Genetic Convergence. *Mol Biol Evol* 2009, 26(8):1909-1919.
11. Yang X, Liu D, Tschaplinski TJ, *et al.* Comparative genomics can provide new insights into the evolutionary mechanisms and gene function in CAM plants. *J Exp Bot* 2019, 70(22):6539-6547.
12. Lepiniec L, Vidal J, Chollet R, *et al.* Phosphoenolpyruvate carboxylase: structure, regulation and evolution. *Plant Sci* 1994, 99(2):111-124.
13. Nimmo HG. The regulation of phosphoenolpyruvate carboxylase in CAM plants. *Trends Plant Sci* 2000, 5(2):75-80.
14. Svensson P, Bläsing OE, Westhoff P. Evolution of C4 phosphoenolpyruvate carboxylase. *Archives of Biochemistry and Biophysics* 2003, 414(2):180-188.
15. Shi J, Yi K, Liu Y, *et al.* Phosphoenolpyruvate Carboxylase in Arabidopsis Leaves Plays a Crucial Role in Carbon and Nitrogen Metabolism. *Plant Physiol* 2015, 167(3):671-681.
16. Ruiz-Ballesta I, Baena G, Gandullo J, *et al.* New insights into the post-translational modification of multiple phosphoenolpyruvate carboxylase isoenzymes by phosphorylation and monoubiquitination during sorghum seed development and germination. *J Exp Bot* 2016, 67(11):3523-3536.
17. Zhao Y, Guo A, Wang Y, *et al.* Evolution of PEPC gene family in *Gossypium* reveals functional diversification and GhPEPC genes responding to abiotic stresses. *Gene* 2019, 698:61-71.
18. Wang N, Zhong X, Cong Y, *et al.* Genome-wide Analysis of Phosphoenolpyruvate Carboxylase Gene Family and Their Response to Abiotic Stresses in Soybean. *Scientific Reports* 2016, 6(1):38448.
19. O'Leary B, Park J, Plaxton WC. The remarkable diversity of plant PEPC (phosphoenolpyruvate carboxylase): recent insights into the physiological functions and post-translational controls of non-photosynthetic PEPCs. *Biochem J* 2011, 436(1):15-34.
20. Darabi M, Seddigh S. Structural, functional, and phylogenetic characterization of phosphoenolpyruvate carboxylase (PEPC) in C4 and CAM plants. *Caryologia* 2018, 71(3):272-288.
21. Gehrig H, Heute V, Kluge M. New partial sequences of phosphoenolpyruvate carboxylase as molecular phylogenetic markers. *Mol Phylogen Evol* 2001, 20(2):262-274.
22. Christin P-A, Arakaki M, Osborne CP, *et al.* Shared origins of a key enzyme during the evolution of C4 and CAM metabolism. *J Exp Bot* 2014, 65(13):3609-3621.
23. Boxall SF, Kadu N, Dever LV, *et al.* *Kalanchoë* PPC1 is Essential for Crassulacean Acid Metabolism and the Regulation of Core Circadian Clock and Guard Cell Signaling Genes. *The Plant Cell* 2020, 32(4):1136-1160.
24. Christin P-A, Salamin N, Savolainen V, *et al.* C4 Photosynthesis Evolved in Grasses via Parallel Adaptive Genetic Changes. *Curr Biol* 2007, 17(14):1241-1247.
25. Deng H, Zhang L-S, Zhang G-Q, *et al.* Evolutionary history of PEPC genes in green plants: Implications for the evolution of CAM in orchids. *Mol Phylogen Evol* 2016, 94:559-564.
26. Gilman IS, Edwards EJ. Crassulacean acid metabolism. *Curr Biol* 2020, 30(2):R57-R62.
27. Jiao J-a, Chollet R. Posttranslational Regulation of Phosphoenolpyruvate Carboxylase in C4 and Crassulacean Acid Metabolism Plants. *Plant Physiol* 1991, 95(4):981-985.
28. Chollet R, Vidal J, O'Leary MH. Phosphoenolpyruvate carboxylase: a ubiquitous, highly regulated enzyme in plants. *Annu Rev Plant Biol* 1996, 47(1):273-298.
29. Cheng S, Xian W, Fu Y, *et al.* Genomes of Subaerial Zygnematophyceae Provide Insights into Land Plant Evolution. *Cell* 2019, 179(5):1057-1067.e1014.
30. Bowman JL, Kohchi T, Yamato KT, *et al.* Insights into Land Plant Evolution Garnered from the *Marchantia polymorpha* Genome. *Cell* 2017, 171(2):287-304.e215.

31. Rensing SA, Lang D, Zimmer AD, *et al.* The *Physcomitrella* Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. *Science* 2008, 319(5859):64-69.
32. Zhang J, Fu X-X, Li R-Q, *et al.* The hornwort genome and early land plant evolution. *Nature Plants* 2020, 6(2):107-118.
33. Banks JA, Nishiyama T, Hasebe M, *et al.* The *Selaginella* Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. *Science* 2011, 332(6032):960-963.
34. Li F-W, Brouwer P, Carretero-Paulet L, *et al.* Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants* 2018, 4(7):460-472.
35. Shen H, Jin D, Shu J-P, *et al.* Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *GigaScience* 2018, 7(2):gix116.
36. Nystedt B, Street NR, Wetterbom A, *et al.* The Norway spruce genome sequence and conifer genome evolution. *Nature* 2013, 497(7451):579-584.
37. Wan T, Liu Z-M, Li L-F, *et al.* A genome for gnetophytes and early evolution of seed plants. *Nature Plants* 2018, 4(2):82-89.
38. Amborella Genome P. The *Amborella* Genome and the Evolution of Flowering Plants. *Science* 2013, 342(6165):1241089.
39. Ming R, VanBuren R, Wai CM, *et al.* The pineapple genome and the evolution of CAM photosynthesis. *Nat Genet* 2015, 47(12):1435-1442.
40. Goff SA, Ricke D, Lan T-H, *et al.* A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*). *Science* 2002, 296(5565):92-100.
41. Schnable PS, Ware D, Fulton RS, *et al.* The B73 Maize Genome: Complexity, Diversity, and Dynamics. *Science* 2009, 326(5956):1112-1115.
42. The Arabidopsis Genome I. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 2000, 408(6814):796-815.
43. Sunil M, Hariharan AK, Nayak S, *et al.* The Draft Genome and Transcriptome of *Amaranthus hypochondriacus*: A C4 Dicot Producing High-Lysine Edible Pseudo-Cereal. *DNA Res* 2014, 21(6):585-602.
44. Soltis PS, Soltis DE. Ancient WGD events as drivers of key innovations in angiosperms. *Curr Opin Plant Biol* 2016, 30:159-165.
45. Matsumura H, Xie Y, Shirakata S, *et al.* Crystal Structures of C4 Form Maize and Quaternary Complex of E. coli Phosphoenolpyruvate Carboxylases. *Structure* 2002, 10(12):1721-1730.
46. Paulus JK, Schlieper D, Groth G. Greater efficiency of photosynthetic carbon fixation due to single amino-acid substitution. *Nature Communications* 2013, 4(1):1518.
47. Connell MB, Lee MJY, Li J, *et al.* Structural and biochemical characterization of citrate binding to AtPPC3, a plant-type phosphoenolpyruvate carboxylase from *Arabidopsis thaliana*. *Journal of Structural Biology* 2018, 204(3):507-512.
48. González-Segura L, Mújica-Jiménez C, Juárez-Díaz JA, *et al.* Identification of the allosteric site for neutral amino acids in the maize C4 isozyme of phosphoenolpyruvate carboxylase: The critical role of Ser-100. *J Biol Chem* 2018, 293(26):9945-9957.
49. Izui K, Matsumura H, Furumoto T, *et al.* PHOSPHOENOLPYRUVATE CARBOXYLASE: A New Era of Structural Biology. *Annu Rev Plant Biol* 2004, 55(1):69-84.
50. Bläsing OE, Westhoff P, Svensson P. 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. *J Biol Chem* 2000, 275(36):27917-27923.
51. Foote AD, Liu Y, Thomas GWC, *et al.* Convergent evolution of the genomes of marine mammals. *Nat Genet* 2015, 47(3):272-275.
52. Hu Y, Wu Q, Ma S, *et al.* Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proceedings of the National Academy of Sciences* 2017, 114(5):1081-1086.
53. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. *Nature Reviews Genetics* 2017, 18(7):411.
54. Cheng F, Wu J, Cai X, *et al.* Gene retention, fractionation and subgenome differences in polyploid plants. *Nature Plants* 2018, 4(5):258-268.
55. Russell K. Monson. Gene Duplication, Neofunctionalization, and the Evolution of C4 Photosynthesis. *Int J Plant Sci* 2003, 164(S3):S43-S54.

56. Zhang L, Chen F, Zhang GQ, *et al.* Origin and mechanism of crassulacean acid metabolism in orchids as implied by comparative transcriptomics and genomics of the carbon fixation pathway. *The Plant Journal* 2016, 86(2):175-185.
57. Chen L-Y, Xin Y, Wai CM, *et al.* The role of cis-elements in the evolution of crassulacean acid metabolism photosynthesis. *Horticulture Research* 2020, 7(1):1-8.
58. Rey C, Guéguen L, Sémon M, Boussau B. Accurate Detection of Convergent Amino-Acid Evolution with PCOC. *Mol Biol Evol* 2018, 35(9):2296-2306.
59. Waseem M, Ahmad F. The phosphoenolpyruvate carboxylase gene family identification and expression analysis under abiotic and phytohormone stresses in *Solanum lycopersicum* L. *Gene* 2019, 690:11-20.
60. Mistry J, Finn RD, Eddy SR, *et al.* Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res* 2013, 41(12):e121-e121.
61. Lu S, Wang J, Chitsaz F, *et al.* CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res* 2019, 48(D1):D265-D268.
62. Bailey TL, Williams N, Misleh C, *et al.* MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 2006, 34(suppl_2):W369-W373.
63. Bailey TL, Gribskov M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* 1998, 14(1):48-54.
64. Chen C, Chen H, Zhang Y, *et al.* TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Molecular Plant* 2020, 13(8):1194-1202.
65. Waterhouse A, Bertoni M, Bienert S, *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* 2018, 46(W1):W296-W303.
66. Katoh K, Standley DM. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol Biol Evol* 2013, 30(4):772-780.
67. Castresana J. Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Mol Biol Evol* 2000, 17(4):540-552.
68. Nguyen L-T, Schmidt HA, von Haeseler A, *et al.* IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 2014, 32(1):268-274.
69. Kalyaanamoorthy S, Minh BQ, Wong TKF, *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017, 14(6):587-589.
70. Hoang DT, Chernomor O, von Haeseler A, *et al.* UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 2017, 35(2):518-522.
71. Comte N, Morel B, Hasic D, *et al.* Treerecs: an integrated phylogenetic tool, from sequences to reconciliations. *Bioinformatics* 2020.
72. Leebens-Mack JH, Barker MS, Carpenter EJ, *et al.* One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 2019, 574(7780):679-685.

Figures

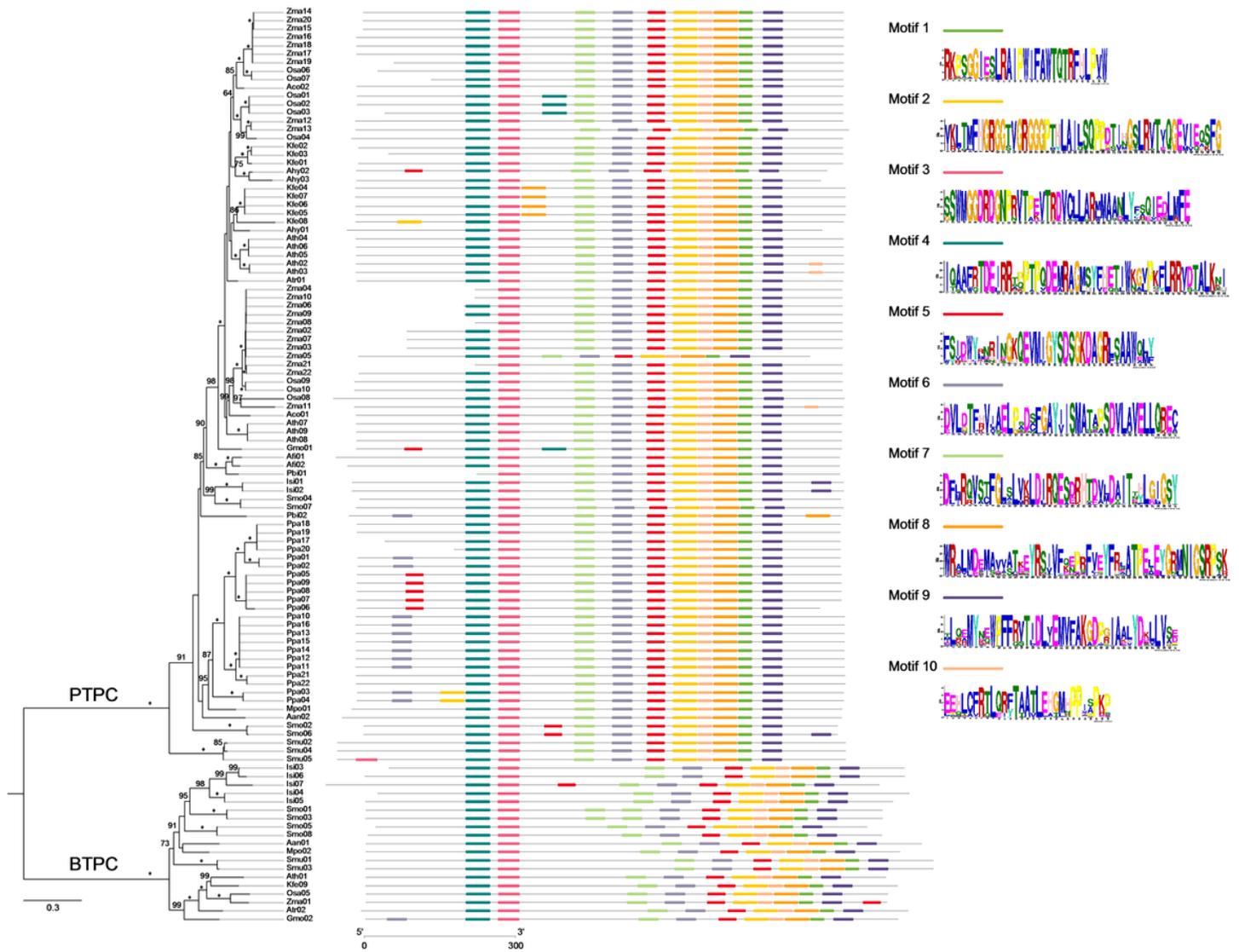


Figure 1

The phylogenetic tree and ten conserved motifs of PEPC gene family in plants. PTPC: plant-type PEPC; BTPC: bacterial-type PEPC. The asterisk in the maximum likelihood tree indicated 100 bootstrap support value, and the value lower than 60 was not shown in gene family tree.

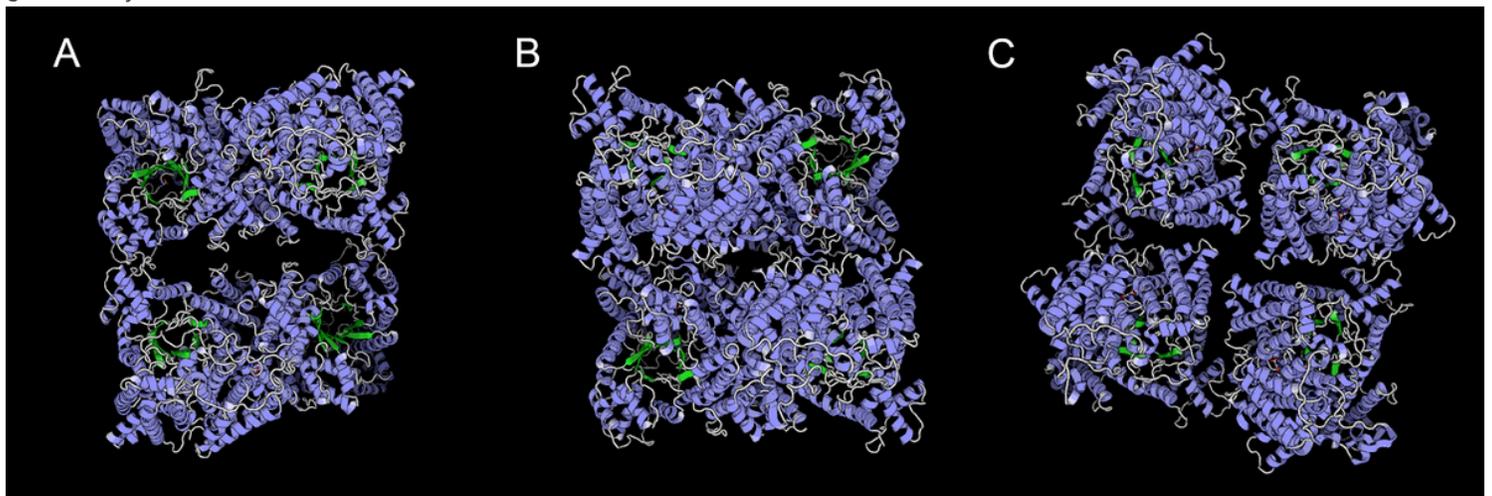


Figure 2

Three templates of three-dimensional structure in PTPC proteins. A: 5vyj.1.A; B: 3zgb.1.A; C: 5fdn.1.A.

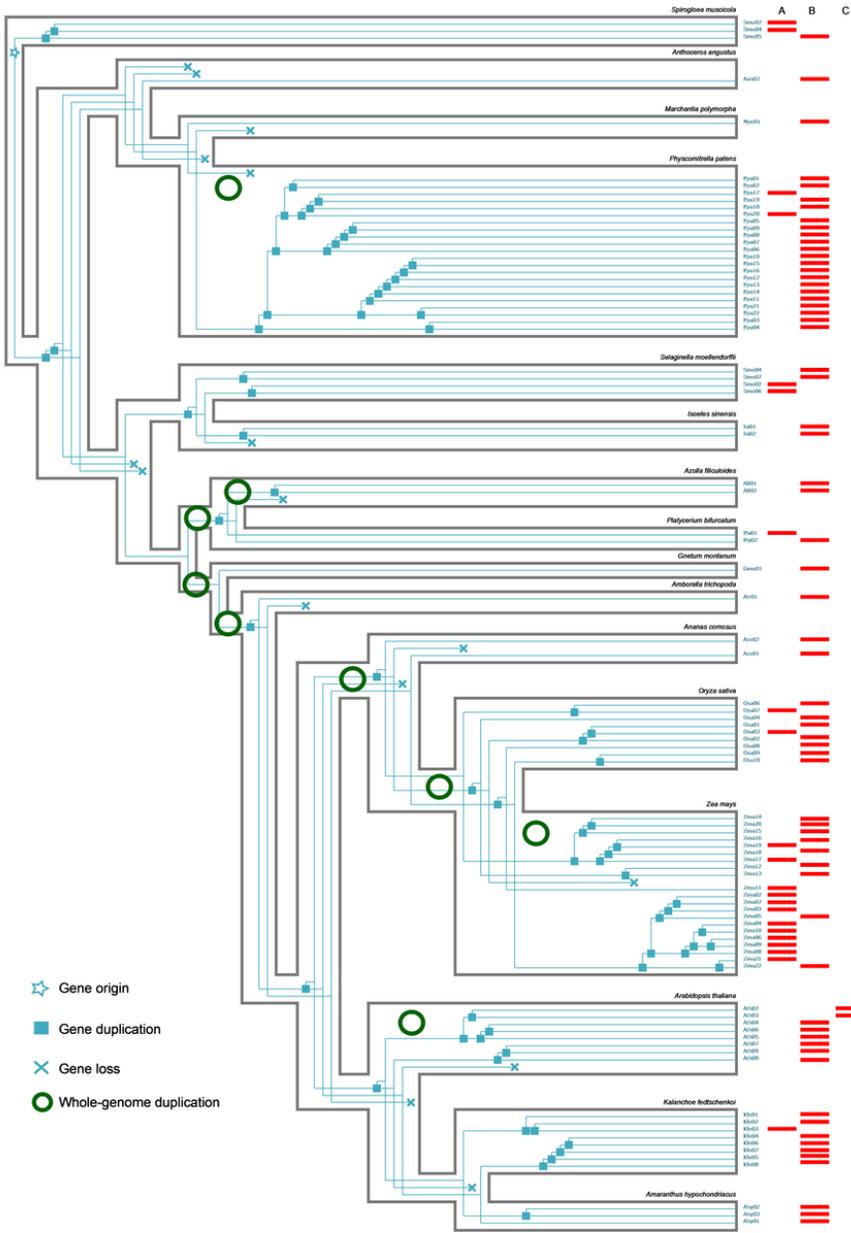


Figure 3

The evolutionary history of PTPC gene family in green plants. A, B, C respectively corresponded to three protein structures (A, B, C) in Figure 2. The information of whole-genome duplication events was based on previous studies [34, 53].

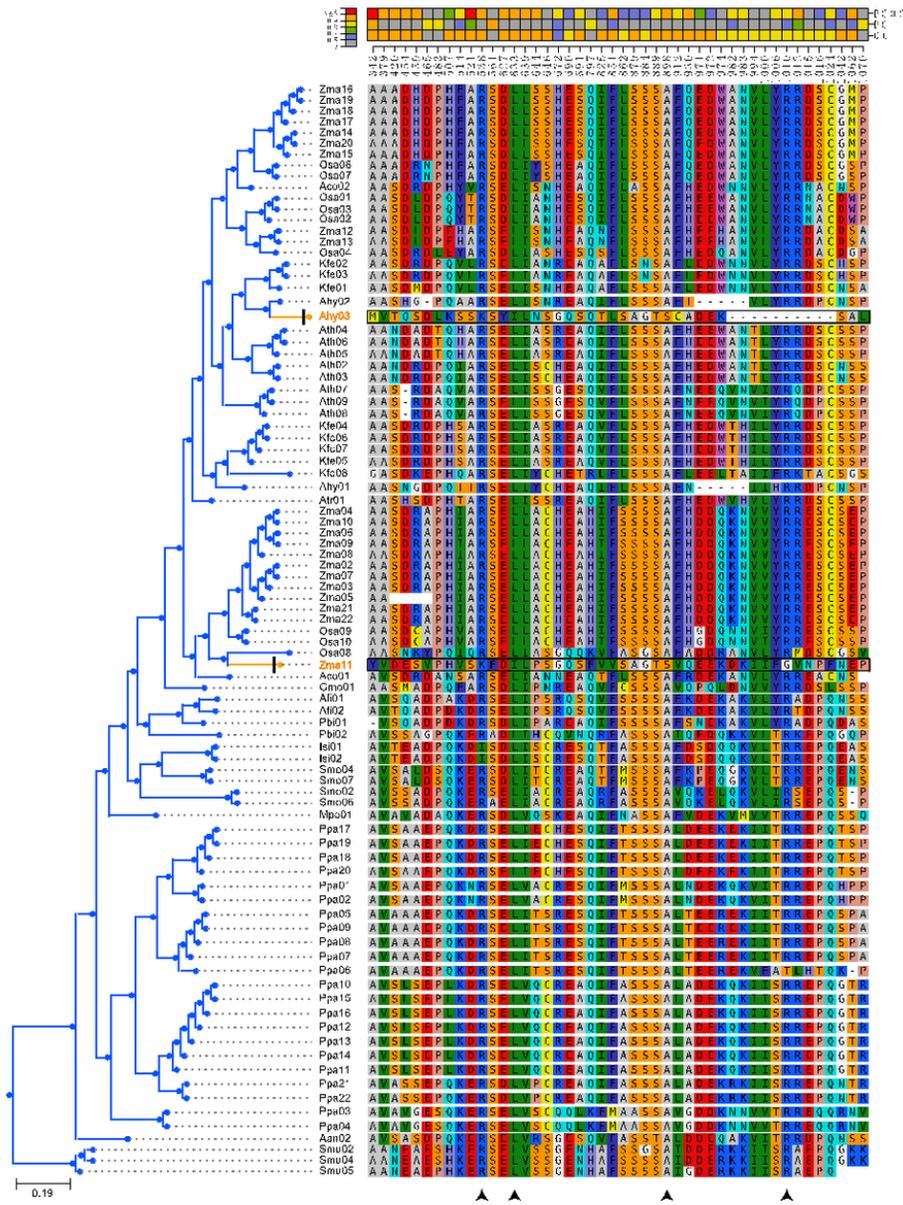


Figure 5

Four convergent substitution sites of C4-PEPC isoforms. The arrowheads at the bottom showed the positions of convergent substitutions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PEPC.zip](#)