In-silico, evolutionary, and functional analysis of CHUP1 and its related proteins in *Bienertia sinuspersici*—a comparative study across C_3, C_4, CAM, and SCC_4 model plants

So Youn Won[1,*], Prabhakaran Soundararajan[1,*], Vadivelmurugan Irulappan[1] and Jung Sun Kim[1]

Department of Agricultural Biotechnology, National Institute of Agricultural Sciences, Rural Development Administration, Jeonju-si, Jeollabuk-do, South Korea

[*] These authors contributed equally to this work.

ABSTRACT

Single-cell C_4 (SCC_4) plants with bienertioid anatomy carry out photosynthesis in a single cell. Chloroplast movement is the underlying phenomenon, where chloroplast unusual positioning 1 (CHUP1) plays a key role. This study aimed to characterize CHUP1 and CHUP1-like proteins in an SCC_4 photosynthetic plant, Bienertia sinuspersici. Also, a comparative analysis of SCC4 CHUP1 was made with C_3, C_4, and CAM model plants including an extant basal angiosperm, Amborella. The CHUP1 gene exists as a single copy from the basal angiosperms to SCC_4 plants. Our analysis identified that *Chenopodium quinoa*, a recently duplicated allotetraploid, has two copies of CHUP1. In addition, the numbers of CHUP1-like and its associated proteins such as CHUP1-like_a, CHUP1-like_b, HPR, TPR, and ABP varied between the species. Hidden Markov Model analysis showed that the gene size of CHUP1-like_a and CHUP1-like_b of SCC_4 species, Bienertia, and Suaeda were enlarged than other plants. Also, we identified that CHUP1-like_a and CHUP1-like_b are absent in Arabidopsis and Amborella, respectively. Motif analysis identified several conserved and variable motifs based on the orders (monocot and dicot) as well as photosynthetic pathways. For instance, CAM plants such as pineapple and cactus shared certain motifs of CHUP1-like_a irrespective of their distant phylogenetic relationship. The free ratio model showed that CHUP1 maintained purifying selection, whereas CHUP1-like_a and CHUP1-like_b have adaptive functions between SCC_4 plants and quinoa. Similarly, rice and maize branches displayed functional diversification on CHUP1-like_b. Relative gene expression data showed that during the subcellular compartmentalization process of Bienertia, CHUP1 and actin-binding proteins (ABP) genes showed a similar pattern of expression. Altogether, the results of this study provide insight into the evolutionary and functional details of CHUP1 and its associated proteins in the development of the SCC_4 system in comparison with other C_3, C_4, and CAM model plants.

Subjects Agricultural Science, Bioinformatics, Genomics, Plant Science, Taxonomy

Keywords *Bienertia sinuspersici*, CHUP1 protein, CHUP1-like proteins, Single-Cell C_4 plants, Phylogenetic analysis, Subcellular compartmentalization

Submitted 8 September 2022
Accepted 14 June 2023
Published 11 July 2023

Corresponding author
Jung Sun Kim, jsnkim@korea.kr

Academic editor
Sapna Langyan

Additional Information and Declarations can be found on page 18

DOI **10.7717/peerj.15696**

How to cite this article Won SY, Soundararajan P, Irulappan V, Kim JS. 2023. *In-silico*, evolutionary, and functional analysis of CHUP1 and its related proteins in *Bienertia sinuspersici*—a comparative study across C_3, C_4, CAM, and SCC_4 model plants. *PeerJ* **11**:e15696
http://doi.org/10.7717/peerj.15696

INTRODUCTION

Plants acquired chloroplasts from a cyanobacterial endosymbiont about 1,000 million years ago (MYA) (*Jensen & Leister, 2014*). Plants exploit well-developed chloroplasts for fixing atmospheric CO_2 into organic compounds (*Jarvis & López, 2013*) using C_3, C_4, and crassulacean acid metabolism (CAM) photosynthetic mechanisms (*Sage, 2004*; *Sage, Way & Kubien, 2008*; *Yang et al., 2015*). Approximately 85% of terrestrial plants have adapted the C_3 photosynthetic pathway. It was estimated that C_3 plants could have originated from the Paleozoic and Mesozoic eras preceding C_4 plants (*Sage, Christin & Edwards, 2011*). Exposure of C_3 plants to adverse conditions led to higher photorespiration. To adapt to drought and high-temperature conditions, C_4 pathways have evolved at least 60 times independently in approximately 19 families of angiosperms. Approximately 3% of plants have adapted the C_4 photosynthesis (*Sage, 2004*; *Sage, Christin & Edwards, 2011*). It can be classified into three types based on the type of four carbon acids, such as NAD-dependent malic enzyme (NAD-ME), NADP-dependent malic enzyme (NADP-ME), and PEP carboxykinase (PEPCK) (*Sage, 2004*; *Offermann, Okita & Edwards, 2011*). Apart from the C_3 and C_4 pathways, 10% of plants growing in extreme conditions (deserts and epiphytes) adopted the CAM pathway, which evolved independently in 400 distinct genera in 36 families (*Sage, Christin & Edwards, 2011*; *Ming et al., 2015*; *Yang et al., 2015*). In this pathway, during the night, stomata remain open and fix malate *via* a phosphoenolpyruvate (PEP) reaction. In the daytime, the C_3 pathway takes place with closed stomata (*Ming et al., 2015*; *Yang et al., 2015*). In C_3 plants, photosynthesis occurs only in the mesophyll (MS) cells, whereas almost all C_4 plants require Kranz anatomy, where both MS and bundle sheath (BS) cells participate in primary and secondary carbon fixation, respectively (*Sage, 2004*; *Koteyeva et al., 2016*). In CAM plants, both the C3 and C4 photosynthesis processes take place in the same MS cells (*Ghannoum et al., 2013*; *Yang et al., 2015*). Recent findings indicate that plants have developed an efficient single-cell C_4 (SCC_4) photosynthetic mechanism which can perform C_4 photosynthesis in a single cell in the absence of C_4 Kranz anatomy (*Sage, 2004*). This type was demonstrated in four terrestrial species within the Chenopodiaceae family. SCC_4 plants employ dimorphic chloroplasts to carry out C_4 photosynthesis (*Mai et al., 2019*). In *Suaeda aralocaspica* (hereafter, Suaeda), two types of chloroplasts are arranged at proximal and distal positions in elongated chlorenchyma cells. In contrast, *Bienertia sinuspersici* (hereafter, Bienertia, SCC_4 model plant), *B. kavirense,* and *B. cycloptera* possess bienertioid anatomy, wherein two types of chloroplasts, namely the central chloroplast (CCp) and the peripheral chloroplast (PCp), are localized and compartmentalized in central and peripheral cytoplasmic regions within a single cell, respectively (*Mai et al., 2019*). PCp generates a C_4 organic acid from atmospheric CO_2, which is decarboxylated in the CCp by the Calvin-Benson-Bassham cycle (*Offermann et al., 2015*). Young Bienertia chlorenchyma cells have a uniform distribution of chloroplasts operating in C_3 photosynthesis without partition of the cytoplasm (*Koteyeva et al., 2016*). The partition of two chloroplasts is observed only in mature chlorenchyma cells. As the chlorenchyma cells mature, the vacuoles fuse to create the unusually large central vacuole, which is distinct from most plant cells, and these cells operate C_4 photosynthesis (*Park*

et al., 2009). The distribution of chloroplasts is achieved by moving chloroplasts to the peripheral cytoplasmic compartments and the central cytoplasmic compartments, and yet they are connected by cytoplasmic channels, which limit the gas diffusion between the two compartments. The function of peripheral and central compartments is similar to that of MS cells and BS cells in C_4-type plants with Kranz anatomy, respectively (*Offermann, Okita & Edwards, 2011*).

Chloroplast movement is one of the mechanisms that sessile plants have developed to receive more light in low-light conditions and to reduce photodamage in high-light conditions. Under low light, chloroplasts have a "periclinal" face position to utilize maximum light, whereas, during high light, an "anticlinal" position is taken to minimize photodamage (*Wada, 2016*). Chloroplast movement is carried out by many factors, such as phototropins (PHTs), Chloroplast Unusual Positioning 1(CHUP1), Plastid Movement Impaired 1 (PMI1), Kinesin-like Protein for Actin-based Chloroplast Movement (KAC), and THRUMIN 1, a glutaredoxin-like protein (*Dwyer & Hangarter, 2022*; *Shi et al., 2022*; *Gao et al., 2023*). Among these factors, CHUP1 is crucial for generating and maintaining chloroplast actin (cp-actin) filaments (*Oikawa et al., 2003*; *Oikawa et al., 2008*). In the Arabidopsis model plant (C_3 plant), CHUP1 protein has been reported to be involved in chloroplast movement (*Oikawa et al., 2003*; *Lehmann, Bohnsack & Schleiff, 2011*). It is localized in the chloroplast outer membrane, and it triggers chloroplast movement by polymerizing the cp-actin filaments using profilactins. CHUP1 belongs to the hydroxyproline-rich (HPR) glycoprotein family and is a multi-domain protein consisting of an N-terminal hydrophobic domain (HD), a coiled-coil domain (CCD), an actin-binding domain (ABD), two leucine zipper (LZ) domains, and a proline-rich motif (PRM) (*Von Braun & Schleiff, 2008*). The N-terminal of the CHUP1 protein anchors to the outer envelope membranes of chloroplasts in MS cells, and the CCD of CHUP1 facilitates the anchorage of chloroplasts to the plasma membrane (*Oikawa et al., 2003*; *Oikawa et al., 2008*; *Lehmann, Bohnsack & Schleiff, 2011*). The Arabidopsis *chup1* mutant showed defects in both the chloroplast movement and its anchorage to the plasma membrane (*Von Braun & Schleiff, 2008*; *Kadota et al., 2009*; *Suetsugu et al., 2010*; *Ichikawa et al., 2011*; *Lehmann, Bohnsack & Schleiff, 2011*; *Lehmann et al., 2011*; *Manandhar-Shrestha et al., 2013*; *Suetsugu et al., 2015*). Therefore, CHUP1 is considered a key binding factor between the chloroplast and cytoskeleton filaments. Similarly, in *Zea mays* (maize) and *Eleusine coracana* (finger millet) plants (NAD-malic enzyme-type C_4 plant), it is reported that CHUP1 links the chloroplasts to actin filaments in MS cells (*Kobayashi et al., 2009*). In *Z. mays*, a dramatic increase in phosphorylation of CHUP1 during midday was demonstrated (*Gao et al., 2023*).

The movement of chloroplasts into the two internal cytoplasmic compartments separated into proximal-distal and central-peripheral locations in SCC_4-type plants is fascinating and is a recent trending topic in chloroplast research. Though studies were carried out on the investigation of the response of chloroplasts under different light conditions, such as low light and high light, where they move to "periclinal" and "anticlinal" for maximum utilization and avoidance responses, respectively (*Wada, 2016*), no research was carried out on the involvement of CHUP1 in the chloroplast movement in SCC_4-type plants during the development process. This work was carried out to identify the copy numbers,

conserved and diverged structural patterns, and natural selection between the different plant systems such as C_3, C_4, CAM, and SCC_4 in both monocot and dicot plants for *CHUP1* and its related genes. The genome analysis mentioned that Amborella is the single living plant that could be the sister lineage to all extant flowering plants (*Albert et al., 2013*). Therefore, we have included Amborella as an extant lineage for both monocots and dicots. Our *in-silico* analysis using AtCHUP1 as a query showed that CHUP1-like and other proteins shared functional domains such as tetratricopeptide repeat (TPR) and actin-binding protein (ABP). In addition, hydroxyproline-rich glycoprotein (HRGP) was identified only in specific species. We considered that these proteins could be an interactive partner with CHUP1. Therefore, we included CHUP1-like proteins and their associated proteins in this genome-wide analysis. Also, the studies on *CHUP1* genes and isoforms at the genome level are sparse. Hence, in this study, we attempt to characterize *CHUP1* and its paralogs among the lineages of C_3, C_4, and CAM in monocot and dicot orders, with the main focus on the SCC_4 model plant. The results of this study could be very useful in understanding the evolutionary and functional adaptation of *CHUP1* and its homologs among plant lineages, particularly the SCC_4 species Bienertia.

MATERIALS AND METHODS

Plant material

Initial identification of the sample was done by Prof. Gerald Edwards(School of Biological Sciences and Center for Integrated Biotechnology, Washington State University, USA) (*Akhani et al., 2005*). *Bienertia sinsuspersici* (accession PRJNA273351) was received from one of our collaborators, Prof. Sascha Offermann (Institute for Botany, Leibniz University Hannover, Germany) as seeds. Seeds were germinated, and tissue culture was performed *in vitro* for propagation and maintenance at the National Institute of Agricultural Sciences, Rural Development Administration, Jeonju, Republic of Korea.

Identification of CHUP1 proteins

A well-characterized Arabidopsis CHUP1 protein (NCBI ID: NP_189197.2) (*Oikawa et al., 2003*; *Lehmann, Bohnsack & Schleiff, 2011*) was used as a query to BLAST against the Arabidopsis whole proteome (*Lamesch et al., 2012*). After the selection of proteins, according to a previous method (*Lamesch et al., 2012*), CHUP1-like and other associated proteins of Arabidopsis were taken as queries to BLAST against Amborella (*Amborella trichopoda*) (*Albert et al., 2013*), maize (*Zea mays*) (*Schnable et al., 2009*), sorghum (*Sorghum bicolor*) (*McCormick et al., 2018*), rice (*Oryza sativa*) (*Ouyang et al., 2007*), pineapple (*Anana cosmosus*) (*Ming et al., 2015*), Cactus *(Cactus gigantea)* (*Copetti et al., 2017*), quinoa (*Chenopodium quinoa*) (*Jarvis et al., 2017*), Amaranthus (*Amaranthus hypochondriacis*) (*Sunil et al., 2014*), Suaeda (*Suaeda aralocaspica*) (*Wang et al., 2019*), and Bienertia (*Bienertia sinuspersici*). For Bienertia, the genome project was launched in 2016, and genome assembly has since been completed with about 97.5% coverage of the anticipated genome size of 3.7 Gb (*Soundararajan et al., 2019*). For the identification of CHUP1 and its associated genes in Bienertia, whole genome protein sequences have been used. Recently, we published a comparative analysis of the YABBY gene in Bienertia with

our existing data (*Soundararajan et al., 2019*). As the first step in genome sequencing, cytogenetic analysis of Bienertia has been published (*Soundararajan et al., 2019*; *Sevilleno et al., 2020*). All genome links have been given in Table S3.

Phylogenetic tree, functional domain, motif, gene structural analysis, and subcellular localization

Phylogenetic trees were constructed using MEGA 6.0 with ClustalW using the maximum likelihood method. The nearest neighbor-interchange and partial deletion (95) were chosen with a 1,000 bootstrap value (*Tamura et al., 2013*). CHUP1 is a multifunctional domain-containing protein. Therefore, functional domains have been mapped using different databases and tools. CCD has been identified using the COILS program (https://bio.tools/coils), with Windows 28 showing more than a 0.95 value and other settings defaulted (*Lupas, Van Dyke & Stock, 1991*). LZ domain position was determined with the 2ZIP server (http://2zip.molgen.mpg.de/index.html) (*Lupas, Van Dyke & Stock, 1991*; *Bornberg-Bauer, Rivals & Vingron, 1998*). The ABD position was recognized with a motif in CHUP1 and a pattern in CHUP1_like proteins. PRM is mapped based on the UniProt (https://www.uniprot.org/) curation of homologous sequences (Table S4). Residues involved in interaction with other proteins on the N-terminal of CHUP1 proteins (HD) were predicted based on linear interacting peptides (LIPs) analysis (https://mobidb.bio.unipd.it/) (*Piovesan et al., 2018*). For CHUP1-like homologs, N-terminal sequence homology with CHUP1 was also considered to predict HD.

Conserved protein motifs were predicted using the MEME (*v.* 4.9.1) online tool with options such as zero or one occurrence and 20 motifs with 6 to 200 widths (*Bailey et al., 2009*). Coding sequences (CDS) of Bienertia, Suaeda, and cactus were obtained by tBLASTn of proteins against their genomes (*Soundararajan et al., 2019*). CHUP1-like and its associated proteins were identified using BLASTp with a 10^{-5} e-value of at least 30% homology against their CHUP1 proteins in all genomes. Proteins with the same chromosomal positions, truncation, ambiguity, and highly diverged sequences were removed. Complete CDS and genome regions were extracted using the FGENESH+ Hidden Markov Model (HMM) profile and BEDtools, respectively. For other species, the genomic region of CHUP1 and its associated proteins were obtained from the respective databases. Exon-intron arrangements were constructed with Gene Structure Display Server 2.0 (GSDS) (*Hu et al., 2015*). The subcellular localization of CHUP1 and its related proteins was analyzed using the DeepLoc 2.0 Prediction of eukaryotic protein subcellular localization tool (https://services.healthtech.dtu.dk/services/DeepLoc-2.0/#:~:text=The%20DeepLoc%202.0%20server%20predicts,localizations%20inside%20the%20eukaryotic%20cel). The amino acid sequences in FASTA format were used as input (*Almagro Armenteros et al., 2017*).

Positive Selection-Selectivity pressure analysis

The codeml program of a phylogenetic analysis by maximum likelihood(PAML) was used for identifying the dN/dS ratio of CHUP1 and CHUP1-like proteins. The PAL2NAL program was used to generate codon alignment (*Suyama, Torrents & Bork, 2006*). A free-ratio model (NSsites = 0, model = 1), which allows different dN/dS ratios for each branch

of the tree, was used to detect adaptive selection (*Yang, 2007*). The unrooted tree has been constructed using Phylip.

RNA isolation, cDNA synthesis, and qRT-PCR analysis

RNA was isolated using the CTAB method from three stages of Bienertia leaves: young (1∼3 mm) with yellowish leaves, intermediate(3∼6 mm) containing green tips with yellowish leaves, and mature upper one cm sections of whole green leaves. Finely ground samples (100 mg) in liquid nitrogen were used for RNA isolation by following the CTAB protocol.

For cDNA synthesis, 2000 ng of RNA was used (amfiRivert II cDNA Synthesis Master Mix, GenDEOPT, USA). Ten-fold diluted cDNA was used for quantitative real-time PCR (qRT-PCR) analysis (iQ™SYBR Green® Supermix; Bio-Rad, Hercules, CA, USA) with the primers mentioned in Table S5. The two-step SYBR method with 58 °C amplification was used in the thermocycler (CFX96 Touch™ Real-Time PCR Detection System; Bio-Rad). Denaturation was performed at 95 °C for 3 min in 40 cycles with 95 °C/15s, 58 °C/30s, and a melting curve with 95 °C/10s, 65 °C/5s, and 60 °C/50s. All samples were analyzed with three individual biological replications. The expression of *CHUP1*, *CHUP1-like*, and its associated genes was calculated by their relative expression to internal control (*GAPDH*). The analysis of qRT-PCR was done with three biological and two technical repeats. Statistical analysis was done using one-way ANOVA followed by Duncan's multiple range tests ($p \leq 0.05$) in SAS software (Statistical Analysis System, V. 6.12; SAS, Inc, Cary, NC, USA).

Microscopy

Protoplasts were prepared from young (0.3–0.5 cm long), intermediate (0.5–1.0 cm long), and mature (between 1.0–1.5 cm long) stage leaf tissues according to the protocol published previously (*Wimmer et al., 2017*). Images were captured under the 20X objective lens of a Zeiss (Jena, Germany) Axioplan fluorescence microscope with a cooled CCD camera and the final images were processed using Photoshop CS6.

RESULTS

Identification of CHUP1 and CHUP1-like proteins in C_3, C_4, CAM, and SCC_4 model plants

In the present study, we identified CHUP1, CHUP1-like proteins, and their associated proteins, including actin-binding proteins (ABP), tetratricopeptide repeat (TPR), and other HPR glycoproteins, across the model plants from C_3, C_4 and CAM of both monocot and dicot taken for the study using BLAST with AtCHUP1 as reference (Table 1). Phylogenetic trees showed that all CHUP1, CHUP1-like, and its associated proteins other than ABP of Arabidopsis were present as monophyletic between monocots and dicots. Similarly, CHUP1-like_a was recognized as monophyletic. The recently duplicated *Chenopodium quinoa* (hereafter, quinoa) has two copies of *CHUP1* and *CHUP1-like_b,* whereas *CHUP1-like_a* was a single-copy gene. Contrasting to the C_3 monocot rice and its close C_4 relatives' sorghum, maize possessed two copies of CHUP1-like_b. One is located on chromosome 1, and the other is on chromosome 9 (Fig. 1, Table 1, Table S1).

Table 1 ***CHUP1*** **and** ***CHUP1_like*** **genes detail of C_3, C_4, CAM, and SCC_4 model plants.**

Plant type	Type	Species	Gene ID	Gene Name	Genome location	Genomic position	Strand	CDS (bp)	Amino acids
Dicot	SCC_4	Bienertia	*Bsv0100-00004053*	*CHUP1*	000130F	1118154-1124532	+	2949	982
			Bsv0100-00030261	*CHUP1-like_a*	003553F	178764-201267	+	2295	765
			Bsv0100-00002964	*CHUP1-like_b*	000073F	919011-919485	−	1110	370
		Suaeda	*GOSA_00026873*	*CHUP1*	Contig 295	64362- 58799	−	2958	985
			GOSA_00027286	*CHUP1-like_a*	Contig 751	381394-398100	+	2466	821
			GOSA_00025994	*CHUP1-like_b*	Contig 607	272990-279085	+	1197	398
	C_4	Amaranthus	*AHYPO_017349*	*CHUP1*	Scaffold 380	54205-62742	−	2880	959
			AHYPO_012450	*CHUP1-like_a*	Scaffold 149	314501-316325	+	1215	404
			AHYPO_001781	*CHUP1-like_b*	Scaffold 9	174263-178053	+	975	324
	C_3	Quinoa	*XP_021721284.1*	*CHUP1*	US	1550610-1544417	+	2943	980
			XP_021715237.1		US	974091-967433	+	2943	980
			XP_021724805.1	*CHUP1-like_a*	US	1539562-1546012	−	2301	766
			XP_021729089.1	*CHUP1-like_b*	US	3186783-3189139	−	1242	413
			XP_021746130.1		US	3209412-3211923	+	1242	413
		Arabidopsis	*NP_189197.2*	*CHUP1*	Chr 3	9352444-9357953	+	3015	1004
			NP_172192.1	*CHUP1-like_b*	Chr 1	2186627-2184759	+	1179	392
	CAM	Cactus	*Cgig1_04867*	*CHUP1*	Scaffold 1884	35922-42874	−	2958	985
			Cgig1_16952	*CHUP1-like_a*	Scaffold 5088	52262-57326	−	2319	772
	CAM	Pineapple	*XP_020098505.1*	*CHUP1*	Chr 11	10132256-10137677	+	2973	990
			XP_020099723.1	*CHUP1-like_a*	Chr 12	12107512-12111995	−	2343	780
			XP_020085709.1	*CHUP1-like_b*	Chr 4	13960552-13957835	+	1266	421
Monocot	C_3	Rice	*LOC_Os11g01439.1*	*CHUP1*	Chr 11	262011-256827	−	2790	930
			LOC Os08g03560.1	*CHUP1-like_a*	Chr 8	1658274-1653274	−	2397	798
			LOC Os03g18300.1	*CHUP1-like_b*	Chr 3	10255348-10257715	+	1251	416
	C_4	Sorghum	*XP_021303517.1*	*CHUP1*	Chr 9	57875196-57870434	+	3195	1064
			XP_002444993.1	*CHUP1-like_a*	Chr 7	2531140-2526115	+	2394	797
			XP_021319606.1	*CHUP1-like_b*	Chr 1	69005221-69001827	−	1257	418
		Maize	*Zm00001d039104.1*	*CHUP1*	Chr 6	170108921-170113150	−	2826	941
			Zm00001d0351491.1	*CHUP1-like_a*	Chr 6	8696067-8700892	−	2394	797
			Zm00001d028761.1	*CHUP1-like_b*	Chr 1	45864595-45867995	+	1272	423
			Zm00001d047723.1		Chr 9	140000414-140003136	+	1293	430
Outgroup	C_3	Amborella	*XP_020521880.1*	*CHUP1*	US	751159-742747	−	3042	1013
			XP_020523876.1	*CHUP1-like_a*	US	133663-141544	−	2577	858

Notes.
US, Unplaced Scaffold.

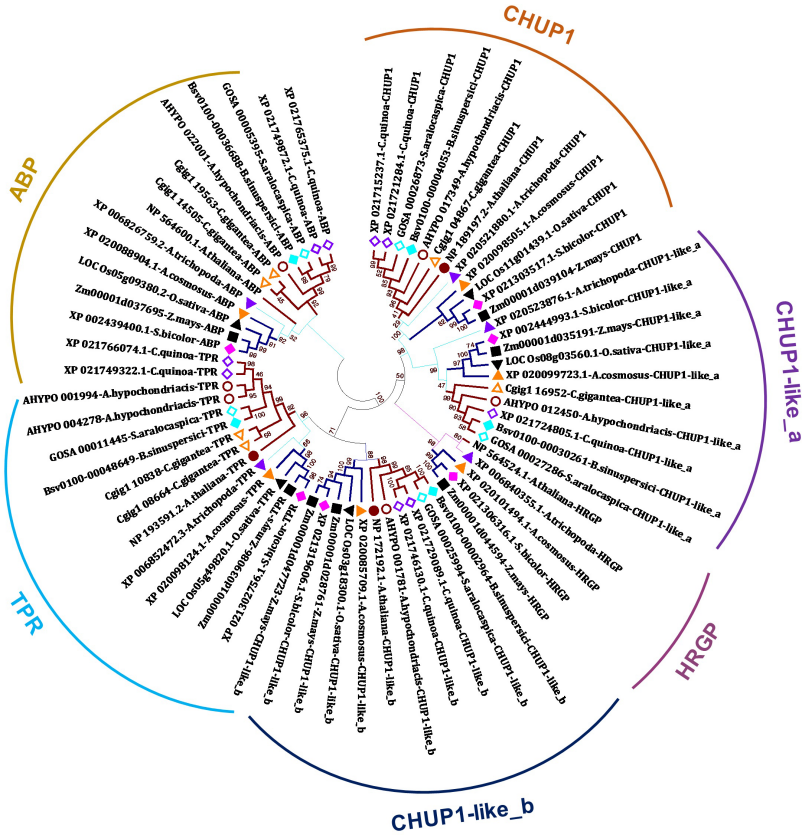

Figure 1 **Phylogenetic tree of CHUP1, CHUP1-like, and its associated proteins.** Am.t (*Amborella trichopoda*, purple up triangle); Zm (*Zea mays*, dark green square); Sb (*Sorghum bicolor*, pink diamond); Os (*Oryza sativa*, light green down triangle); Ac (*Ananas cosmosus*, orange up triangle); At (*Arabidopsis thaliana*, brown circle); Cg (*Carnegiea gigantea*, orange-outlined up triangle); Cq (*Chenopodium quinoa*, purple-outline diamond); Ah (*Amaranthus hypochondriacis*, brown-outlined circle); Sa (*Suaeda aralocaspica*, aqua-outlined diamond); and Bs (*Bienertia sinuspersici*, aqua diamond). The percentage of 1,000 bootstrap replicates was represented in branches. ABP, actin-binding protein; HPR, hydroxyproline-rich glycoproteins; TPR, tetratricopeptide repeat. Dicots, monocots, and Amborella are highlighted in bold dark orange-colored branches, bold dark blue-colored branches, and thin light blue-colored branches, respectively.

Full-size DOI: 10.7717/peerj.15696/fig-1

Functional domain prediction

To identify the functional domains in CHUP1 and CHUP-like proteins, we used the COILs program for CCD, the 2bZIP server for LZ, UniProt for PRM, and LIP analysis for HD predictions. The position of functional domains such as HD, CCD, LZs, ABD, and PRM of CHUP1 and CHUP1-like *proteins* was mapped in the C_3, C_4, CAM, and SCC_4 model plants (Fig. 2). Functional domain results showed that, unlike CHUP1, CHUP1-like_a was deficient in the C-terminal LZ domain. Except for the C_3 photosynthetic plant Arabidopsis, all other model plants had PRM along with CCD and LZ domains in the CHUP1-like_b protein.

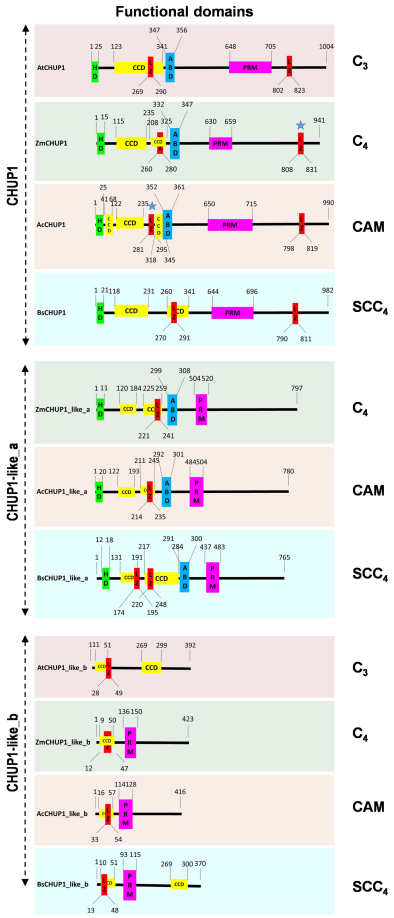

Figure 2 Functional domain prediction of CHUP 1 and CHUP 1_like proteins from the C_3, C_4, CAM, and SCC_4 model plants. HD, hydrophobic domain; CCD, coiled-coil domain; LZ, leucine zipper region; ABD, actin-binding domain; PRM, proline-rich motif. At, *A. thaliana*; Zm, *Z. mays*; Ac, *A. cosmosus*; and Bs, *B. sinuspersici*. The blue star represents the position of the LZ region that has been manually curated based on the motif pattern.

Full-size DOI: 10.7717/peerj.15696/fig-2

Gene structure and motif analysis of *CHUP1* and *CHUP1-like* genes in C_3, C_4, CAM, and SCC_4 plants

We investigated the intron/exon pattern and the presence and positions of motifs in all *CHUP1* and *CHUP1-like* genes across the model plants under study. Gene structure analysis revealed that despite the variation in gene size, the *CHUP1* gene of all species had nine exons (Fig. 3A). A total of 20 motifs were identified, of which motif 19 was observed only in the Caryophyllales and motif 20 was detected as Poales-specific. Interestingly, Amborella shared motifs 9 and 16 with all dicot species and pineapple, and motif 20 with Poales. The pineapple shared motif 7, which was identified only in dicots. All species contained motifs 4 and 18 in their C-terminal region except Amaranthus. As observed in sorghum and maize, C_4 plants had rearrangement displays on motifs 16 and 20 (Fig. 3B).

For the comparative analysis of *CHUP1-like_a* and *CHUP1-like_b*, Amaranthus was excluded from gene structure and motif analysis because of its smaller size (Table 1). Several

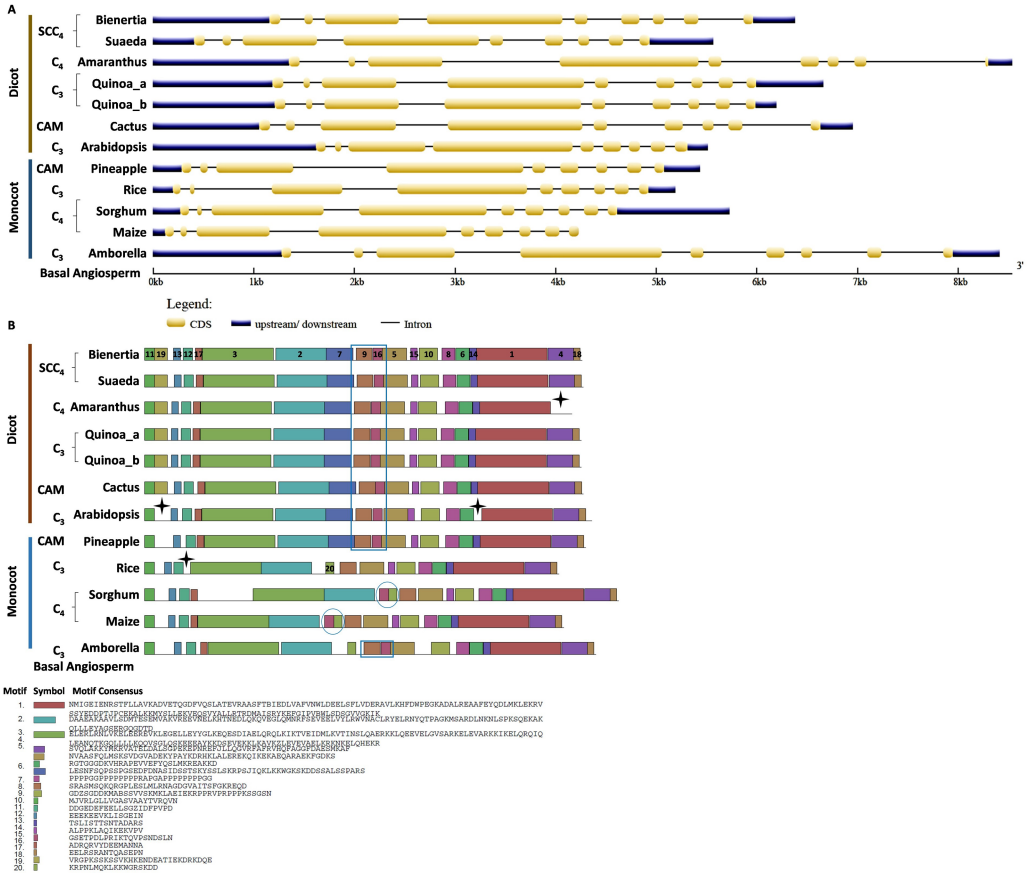

Figure 3 **Gene structure and motif of CHUP1-like_a from C_3, C_4, CAM, and SCC_4 plants.** (A) The gene structure of CHUP1 was analyzed using GENE STRUCTURE DISPLAY SERVER (GSDS) v. 2.0 (http://gsds.gao-lab.org/index.php). The CDS and gene sequences were used as input. (B) The motif of the CHUP1 was identified using the MULTIPLE EM for MOTIF ELICITATION (MEME) server v. 4.9.1 (https://meme-suite.org/meme/tools/meme). The protein sequence was used as input. The motif numbers are labeled in order. The black stars denote the differential motif either between species, lineages, types, or clades. The square box indicates the shared motifs among lineages. The circle (blue) around the motifs specifies the unique motifs present in monocots. Note: The sequences of each motif are provided at the bottom of the figure in a black square box.

Full-size DOI: 10.7717/peerj.15696/fig-3

key differences between C_3, C_4, CAM, and SCC_4 species were found in the gene structure and motif of *CHUP1-like_a* (Fig. 4). Apart from Cactus and Suaeda, all species had seven exons. In the cactus, splits in the first and second exons were observed. Suaeda possessed an extra exon compared to the other species. The *CHUP1-like_a* gene size of all species was between 5kb and 8kb, whereas in SCC_4 species, Bienertia and Suaeda had about 22kb and 17kb, respectively (Fig. 4A). Distinct motifs between the compared species were found in the CHUP1-like_a protein (Fig. 4B). The motifs 17 and 19 were Amaranthaceae-specific whereas motifs 5 and 6 were Poaceae-specific. Interestingly, Amborella and the monocot CAM plant pineapple shared the motifs 8, 9, 10, 11, 12, and 15 with other dicot plants. Additionally, motif 7, which was present only in dicots, was shared with Amborella. Motif

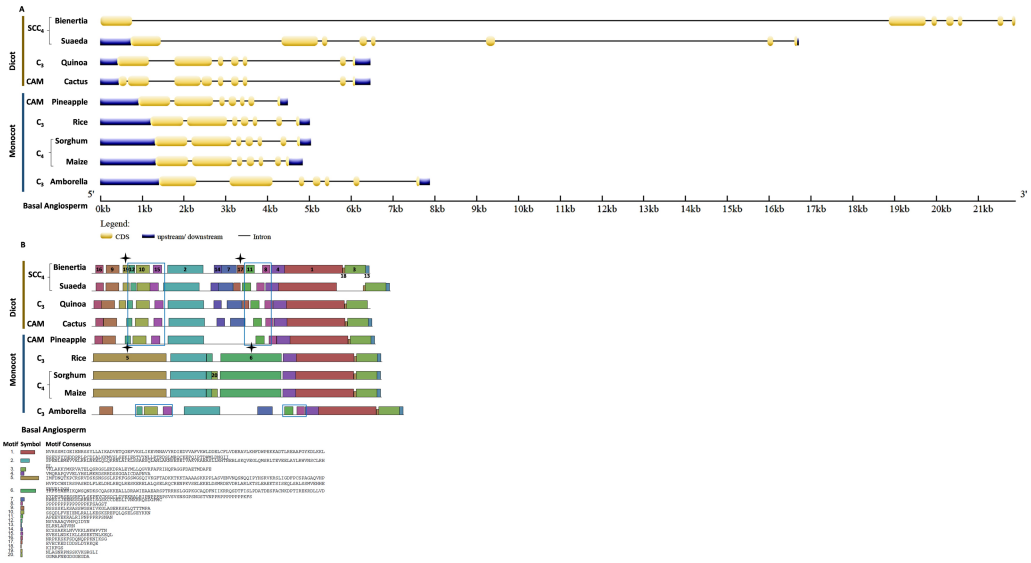

Figure 4 Gene structure and motif of CHUP1-like_a from C_3, C_4, CAM, and SCC_4 plants. (A) The gene structure of CHUP1-like_a was analyzed using GENE STRUCTURE DISPLAY SERVER (GSDS) v. 2.0 (http://gsds.gao-lab.org/index.php). The CDS and gene sequences were used as input. (B) The motif of the CHUP1 was identified using the MULTIPLE EM for MOTIF ELICITATION (MEME) server v. 4.9.1 (https://meme-suite.org/meme/tools/meme). The protein sequence was used as input. The motif numbers are labeled in order. The black stars denote the differential motif either between species, lineages, types, or clades. The square box indicates the shared motifs among lineages. Note: The sequences of each motif are provided at the bottom of the figure in a black square box.

Full-size DOI: 10.7717/peerj.15696/fig-4

14 was dicot-specific. The difference between C_3 and C_4 monocots was the absence of motif 20 in rice.

Bienertia had an extra exon insertion between the first and third exons compared to other plants. The *CHUP1-like_b* gene size of all species was approximately between 2.2 and 3.6kb, whereas Bienertia and Suaeda had 12kb and 6kb, respectively (Fig. 5A). From the identified motifs, 13 and 14 were Amaranthaceae-specific and 8, 9, and 15 were Poaceae-specific (Fig. 5B). Interestingly, two motifs 1 and 3, commonly present in all species, were absent in Bienertia. Similarly, Amaranthaceae-specific motif 13 was also not detected in Bienertia. Among monocots, motif 16 existed only in sorghum and maize, both C_4 plants. We also identified the gene structure and motifs of CHUP1-associated proteins such as ABP (Fig. S1), HPR (Fig. S2), and TPR (Fig. S3).

Adaptive evolution

We calculated the non-synonymous substitution rate (dN) and synonymous substitution rate (dS) to evaluate the divergence and selection pressure in the *CHUP1* gene among the species under study. The analysis showed that the ratio of non-synonymous to synonymous substitutions (dN/dS, ω) was less than 1. So *CHUP1* underwent purifying selection along the lineages. However, adaptive evolution was detected between the Caryophyllales lineages of C_3 (Quinoa) and SCC_4 plants on the CHUP1-like_a protein. Similarly, adaptive selection

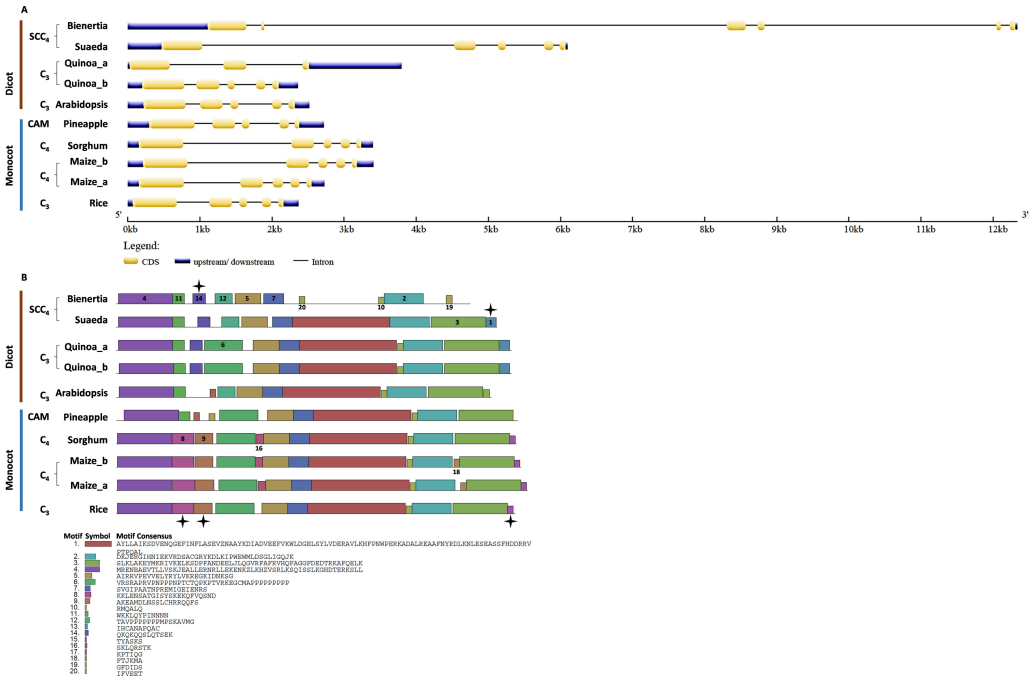

Figure 5 Gene structure and motif of CHUP1-like_b from C_3, C_4, CAM, and SCC_4 plants. (A) The gene structure of CHUP1-like_b was analyzed using GENE STRUCTURE DISPLAY SERVER (GSDS) v. 2.0 (http://gsds.gao-lab.org/index.php). The CDS and gene sequences were used as input. (B) The motif of the CHUP1 was identified using the MULTIPLE EM for MOTIF ELICITATION (MEME) server v. 4.9.1 (https://meme-suite.org/meme/tools/meme). The protein sequence was used as input. The motif numbers are labeled in order. The black star denotes the differential motif either between species, lineages, types, or clades. Note: The sequences of each motif are provided at the bottom of the figure in a black square box.
Full-size DOI: 10.7717/peerj.15696/fig-5

was observed in C_3-C_4-SCC_4 branches of Caryophyllales and also in monocots such as rice and maize (Fig. 6).

Stage-specific morphological development and quantitative expression of *CHUP1* and its associated genes

The young stage cells were sampled from Bienertia leaves less than three mm long, the intermediate stage cells from leaves 4–6 mm long, and the mature stage cells from the tip region of leaves three cm long (Fig. 7A). The scattered starch-like granules with pre-CCps appeared in the youngest stage, and no proper differentiation was observed between CCp and PCp. The progressive transition of PCp was observed in the intermediate stage. The complete subcellular localization and compartmentalization of PCp and CCp occurred in the mature stage (Fig. 7B). The transcript expression analysis of five genes in Bienertia displayed two different patterns during leaf developmental stages. Though the *CHUP1* gene expression was decreased in the intermediate stage compared with that in the young stage, it was almost constitutively expressed in the mature stage, and its expression level was highest among the homologous genes, as shown by the normalized expression value against the internal control, *glyceraldehyde 3-phosphate dehydrogenase* (*GAPDH*). Expression of *ABP* was in a similar pattern to*CHUP1*. The *CHUP1-like* and *TPR* genes showed decreased

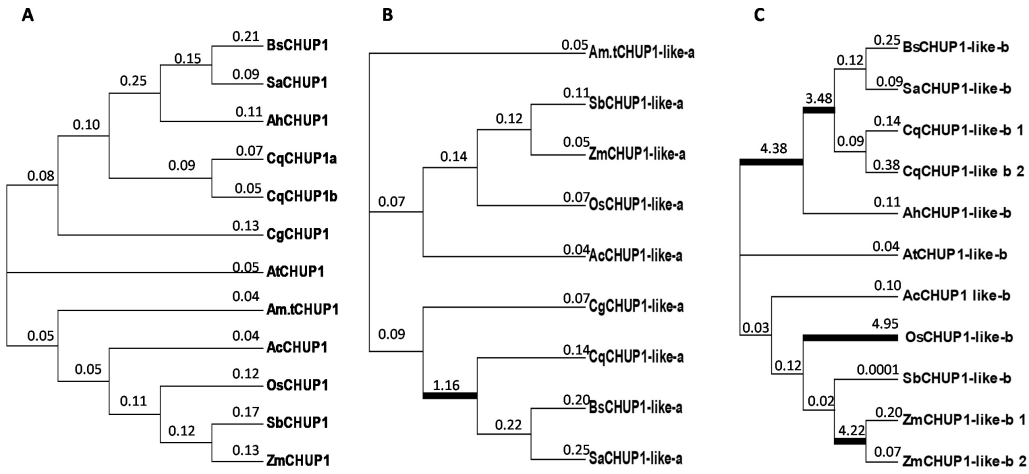

Figure 6 Phylogenetic analysis by maximum likelihood. The unrooted images show the dN/dS ratio of (A) CHUP1, (B) CHUP1-like_a, and (C) CHUP1-like_b proteins from plant species included in the study. A free-ratio model was used to calculate independent ω for each branch of C3, C4, CAM, and SCC4 plants. The ratio of dN/dS was mentioned on the branches and lineages. Branches with an estimated ω ratio > 1 were emphasized in thick black color lines.

Full-size DOI: 10.7717/peerj.15696/fig-6

expression patterns from the young to intermediate to mature stages (Fig. 7C). Subcellular localization analysis showed that BsCHUP1, SaCHUP1, BsABP, and SaABP are predicted to localize at the chloroplast outer membrane, whereas BsCHUP1-like_a, BsCHUP1-like_b, BsTPR, SaCHUP1-like_a, SaCHUP1-like_b, and SaTPR are predicted to be localized in the cytoplasm. This result shows that CHUP1 and CHUP1-like proteins are involved in central and peripheral chloroplast cytoplasmic compartmentation (Table S2).

DISCUSSION

CHUP1 is involved in the chloroplast accumulation and avoidance response to facilitate photosynthetic efficiency. *CHUP1* was extensively studied in the Arabidopsis model plant (C_3), and its orthologs were studied in millet and maize (C_4) (*Kobayashi et al., 2009*). In Arabidopsis, it has been shown that *CHUP1* is involved in the light avoidance mechanism, which was demonstrated in *chup1* mutants. At the chloroplast outer envelope, CHUP1 links chloroplasts with actin (*Oikawa et al., 2003*; *Von Braun & Schleiff, 2008*; *Lehmann, Bohnsack & Schleiff, 2011*). In the case of Bienertia, the role of CHUP1 is not known. Since Bienertia is an SCC_4 plant with bienertioid anatomy, we hypothesized that it may have paralogs of *CHUP1* and its associated evolved proteins. Therefore, this study found CHUP1 and its associated novel proteins such as CHUP1-like_a, CHUP1-like_b, HPR, TPR, and ABP in the genomes included in the study. In this study, genome-wide identification analysis showed the presence of one copy of *CHUP1*, *CHUP1-like_a*, and *CHUP1-like_b* in SCC_4 species, which is similar to the other plants. However, variation in gene size and amino acid number was observed (Fig. 1, Table 1). Variations in *CHUP1*, *CHUP1-like_a*, and *CHUP1-like_b* gene numbers are correlated with gene duplication and deletion of gene architecture between the species (Table 1, Table S1) (*Long & Deutsch, 1999*; *Wang*

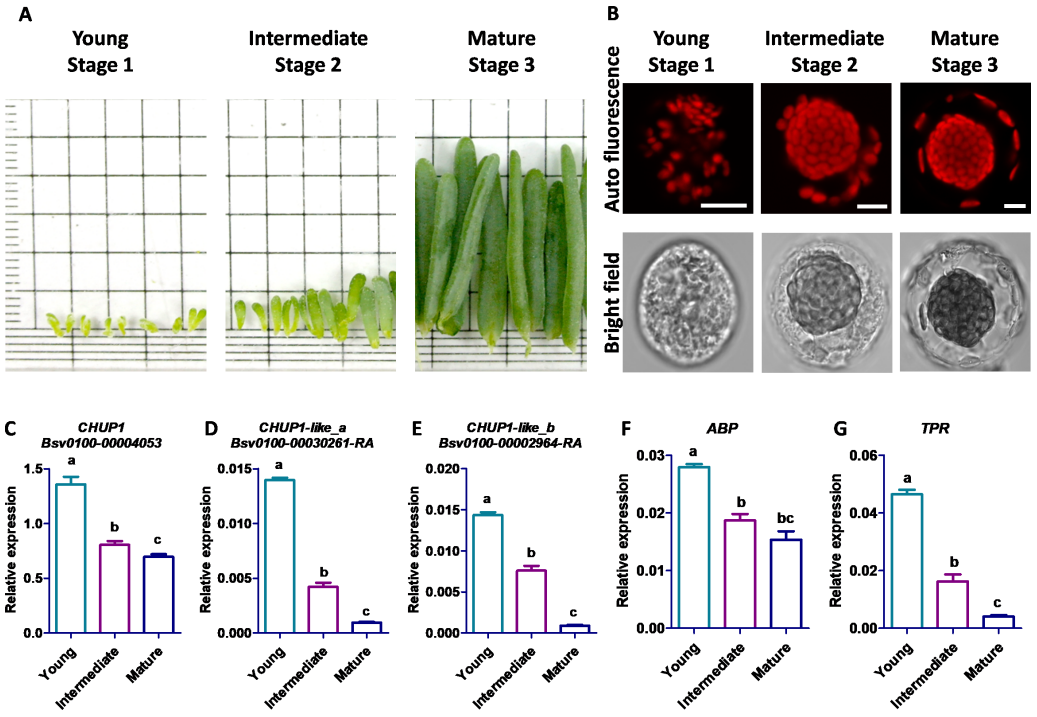

Figure 7 **Expression studies of *CHUP1, CHUP1-like*, and its associated genes in *B. sinuspersici*.** (A) Leaf samples of *B. sinuspersici* from young (Stage 1), intermediate (Stage 2), and mature (Stage 3) stage plants (details in materials and methods) (B) Development of central chloroplasts and peripheral chloroplasts in the different stages of leaves. Chloroplast autofluorescence (top panels) and bright-field (bottom panels) images, scale bar = 10 μm. Relative expression of (C) CHUP1, (D) CHUP1-like_a, (E) CHUP1-like_b, (F) ABP, and (G) TPR was analyzed using qRT-PCR in three different developmental stages. ABP, actin-binding protein; TPR, tetratricopeptide repeat. Different letters indicate that treatments are significantly different at $p \leq 0.05$. The error bar indicates the standard error of the mean.

Full-size DOI: 10.7717/peerj.15696/fig-7

et al., 2018). In phylogenetic analysis, it was found that CHUP1, CHUP1-like_a, and CHUP1-like_b were placed in different clades. However, CHUP1 was found to share the monophyletic group with the CHUP1-like_a protein (Fig. 1). In addition, Suaeda and Bienertia were observed with an additional exon on *CHUP1-like_a* (Fig. 4A) and *CHUP1-like_b* (Fig. 5A), respectively. Moreover, exon splits were observed in the 5′ end of *CHUP1-like_a* of cactus (Fig. 5A). These molecular changes in the intron or exon can be attributed to splicing events or elements and genome evolution (*Bondarenko & Gelfand, 2016*). Studies have demonstrated that ND, CCD, ABD, and PRM domains are significant for chloroplast movement by characterizing the mutants and their complements in plants (*Oikawa et al., 2003*; *Oikawa et al., 2008*; *Yamada et al., 2009*; *Li et al., 2010*; *Chotewutmontri & Barkan, 2016*; *Wang et al., 2018*). The absence of the C-terminal LZ domain in CHUP1-like_a proteins was observed in all species in our analysis (Fig. 2), and it showed it might not undergo a dimerization process (*Wang et al., 2018*), unlike CHUP1. In CHUP1, dimerization is one of the key processes for chloroplast anchoring (*Lehmann et al., 2011*). Similarly, unlike the CHUP1 protein, functional domains such as the HD region, ABD, and C-terminal LZ are absent in the CHUP1-like_b protein (Fig. 2). Though the

significance of the absence of conserved motifs in BsCHUP1-like_b is unclear, the presence of a higher number of transposons in its genome may be attributed to the loss of motifs (*Offermann, Okita & Edwards, 2011*; *Panchy, Lehti-Shiu & Shiu, 2016*).

The presence of functional domains was clearly distinguished between CHUP1, CHUP1-like_a, and CHUP1-like_b proteins (Fig. 3). The absence of HD in CHUP1-like_b showed that it is not involved in the chloroplast movement. No chloroplast relocation movement was observed in an Arabidopsis mutant with 300 aa in the N-terminal region of CHUP1 (*Oikawa et al., 2008*). This indicates the importance of ABD and PRM in chloroplast relocation. The CCD and LZ (also called short CCD) regions of CHUP1-like_b showed their involvement in the signaling networks, interaction with filaments, and organization of cellular processes like cell division (*Yamada et al., 2009*). Though the C-terminal LZ role remains elusive in CHUP1, it could have been involved in intra-molecular interaction (*Lehmann, Bohnsack & Schleiff, 2011*). The presence of ABD and PRM (Fig. 6) showed the involvement of CHUP1-like_a in binding with G-actin and profilin (*Von Braun & Schleiff, 2008*). Further, *in vivo* studies are needed to know the role of CHUP1-like_a in chloroplast movement. Well-characterized functional domains for homo-dimerization, anchoring the chloroplast, and the photo-relocation process (*Von Braun & Schleiff, 2008*) are absent in the CHUP1-like_b protein (Fig. 6). This concurs with the fact that BS CHUP1-like_b is not engaged in chloroplast relocation movements (*Li et al., 2010*; *Chotewutmontri & Barkan, 2016*). The existence of differential regulatory motifs between CHUP1 (Fig. 4B) and CHUP1-like_b (Fig. 5B) was enriched with the different composites of the protein-protein interactive network. Intriguingly, no closely related homologs for NP_564524.1, HPR, were identified in other dicots (Fig. 1). *Oikawa et al. (2008)* reported that the CCD of CHUP1 is apparent for oligomerization followed by firm anchorage of the chloroplast on the plasma membrane (*Chotewutmontri & Barkan, 2016*). The MS chloroplasts of maize are structurally similar to those of C_3 plants (*Oikawa et al., 2008*). Usually, the BS chloroplast is not influenced by light intensity. The structure of the BS chloroplast is neither affected when plants are grown under tropical sunlight nor in low-light conditions (*Drozak & Romanowska, 2006*).

Compared with CHUP1, CHUP1-like proteins possessed different motifs (Figs. 3B, 4B, and 5B). For chloroplast movement, the N-terminal has been considered the key translocation component (*Li & Chen, 1996*; *Oikawa et al., 2003*). However, we have found that CHUP1-like_a had a conserved C-terminal region (Fig. 4B). Interestingly, pineapple, a monocot, shared comparable functional motifs of CHUP1-like_a with cactus, a distant CAM-relative, rather than other close grass family members (Fig. 4B). It may be a grass-specific change showing the similarity between pineapple and the dicots (*Kondo et al., 2004*). Since Amborella shared motifs with the dicots and CAM plants (Fig. 4B), an adaptive response could be a possible correlation for the CHUP1-like_a function.

Though CHUP1 proteins exhibited purifying selection, CHUP1-like_a protein showed strong positive selection in Caryophyllales, which contains the SCC_4 plant lineage and C_3 (quinoa). This represents the functional adaptation of CHUP1-like_a in the SCC_4 species, and it could be involved in the unique chloroplast localization mechanisms within the single cell (Fig. 6B). However, positive selection was found commonly in the Caryophyllales clade.

More combinations of evolutionary models need to be conducted to find the possibility of CHUP1-like_a on SCC_4 adaptation. CHUP1-like_b possessing positive selection signifies differential function on BS chloroplast movement in maize and also its functional adaptation in lineages of C_3 (rice) (Fig. 6C). The existence of positiveselection on CHUP1-like_a and CHUP1-like_b denotes the neo-functionalization/sub-functionalization of modified photosynthetic or other regulatory mechanisms as well as independent evolution in certain plant lineages, including SCC_4 plants (*Panchy, Lehti-Shiu & Shiu, 2016*; *Wang et al., 2018*).

Previous studies by *Offermann et al. (2015)* showed that Bienertia exhibits no chloroplast differentiation in the early stages. Organelle position occurs during the developmental stages of leaves (*Offermann, Okita & Edwards, 2011*; *Offermann et al., 2015*). The constitutive expression of *CHUP1* in the intermediate and mature stages showed its involvement in the partitioning of MS chloroplasts towards the plasma membrane (*Li & Chen, 1996*; *Kondo et al., 2004*; *Slewinski, 2013*; *Miyake, 2016*). Increased CHUP1 gene expression in MS was observed during the developmental changes in maize (*Drozak & Romanowska, 2006*; *Li et al., 2010*). Our study shows a similar pattern of *BsCHUP1* expression(Fig. 7C). Similar expression patterns of *CHUP1* and *ABP* from young to mature stage leaf development (Fig. 7C) showed their significance in the involvement of compartmentalization of PCp in Bienertia. *Suetsugu & Wada (2016)* stated that chloroplast movement depends on specialized actin filaments, namely cp-actin filaments (*Suetsugu & Wada, 2016*). Lesser expression of *CHUP1-like* genes is correlated with their localization in the central cytoplasmic compartment. *CHUP1-like* gene expression patterns were found to be similar to those of *TPR* gene (Fig. 7C). Additionally, the TPR protein has been reported to facilitate the interaction between proteins. It is intricate in chloroplast gene expression and essential for signal transduction pathways (*Hu et al., 2014*).

A transcriptomic study on maize leaf detected that 64% and 21% of genes are differentially expressed along with the developmental gradient of BS and MS chloroplasts, respectively (*Li et al., 2010*). Accordingly, *ZmCHUP1* homologs to *AtCHUP1* and *OsCHUP1* are expressed three-fold higher in MS than in BS, and higher phosphorylation of CHUP1 in MS cells was demonstrated in maize at mid-day with high light intensity (*Gao et al., 2023*). In finger millet and maize, BS chloroplasts migrate toward vascular bundles and form the centripetal position during cell maturation (*Li et al., 2010*). The establishment of positioning is related to tissue development and cytoskeletal changes (*Koteyeva et al., 2016*). In maize, extensive partitioning of the photosynthetic process occurs between MS and BS during the leaf developmental changes from the basal zone, transitional zone, maturing zone, and mature zone (tip, +1 cm below the leaf tip), with an active expression of the *ZmCHUP1* gene in MS cells (*Kondo et al., 2004*; *Drozak & Romanowska, 2006*). In most cases, organelle movements are dependent on actin filaments. Similar expression patterns of CHUP1 and ABP from young to mature leaf development (Fig. 7C) showed their significance in the involvement of compartmentalization of PCp in Bienertia. *Suetsugu & Wada (2016)* stated that chloroplast movement depends on specialized actin filaments, namely cp-actin filaments. CHUP1-like gene expression patterns are similar to those of *TPR* gene (Fig. 7C). Additionally, the TPR protein has been reported to be involved in facilitating

structural interactions between proteins. It is involved in chloroplast gene expression and is essential for signal transduction pathways. A mutant of *tpr* in Arabidopsis displayed a slow greening phenotype (*Yang et al., 2011*). Therefore, TPR has been considered an important gene involved in chlorophyll biosynthesis and photosynthetic signaling.

CONCLUSIONS

Overall, CHUP1 and its associated proteins were identified and characterized in SCC_4 with reference to existing C_3, C_4, and CAM models, including the extant plant Amborella. In summary, the phylogeny-based analysis showed CHUP1-like_a shared monophyly with CHUP1 since it might have the property of MS-specific function. Monophyletic clade BsCHUP1-like_b without HD domain showed that it could be expressed in CCp. In contrast, decreased expression of both BsCHUP-like genes in the well-developed leaves was observed. CHUP1 contains domains that anchor the chloroplast to the plasma membrane (*Oikawa et al., 2008*). In MS cells, chloroplasts are anchored to the plasma membrane, whereas in BS cells, chloroplasts are located at the center surrounding the vascular bundle (*Oikawa et al., 2003*). High expression and activation of CHUP1 were reported in MS cells to facilitate the anchorage (*Li et al., 2010*; *Gao et al., 2023*). Similarly, in Bieneritia, higher CHUP1 expression was observed in the younger stage as chloroplasts are evenly distributed. But in the matured stage, chloroplasts are released from the plasma membrane and aggregated at the center. This phenomenon supports the low expression of CHUP1 in the mature stage. In addition, the subcellular localization prediction tool has predicted CHUP1 to be localized in the chloroplast and CHUP1-like proteins in the cytoplasm. Amplification of genes was observed only in SCC_4 species. The CHUP1-like_a and CHUP1-like_b genes showed that these could have undergone evolutionary changes. Finally, positive selection on SCC_4 gives the primary clue about the independent evolution in the Caryophyllales clade. Similarly, positive selection in CHUP1-like_b in rice and maize showed that this species could also vary in the MS-specific chloroplast mechanism.Because these genes are crucial for chloroplast mobility, their involvement in SCC4 photosynthesis does not necessitate further positively selected modifications, at least not at the level of amino acid sequence. Positive selection was observed in gene phylogenies, although it was not always associated with SSC4 function. However, non-coding regulatory sequences could have undergone positive selection. However, establishing this would call for a population genetics-style method, such as looking for indications of a ‘‘selective sweep’’ around a specific locus. The results of this study shed light on the functional and evolutionary details of the CHUP1 protein and its related proteins. Detailed experiments on the CHUP1-regulatory networks are essential to reveal the evolution of chloroplast compartmentalization networks, positive selection, and interacting networks in the SCC_4 system.

Abbreviations

ABP	Actin binding proteins
BS	Bundle sheath
CCD	Coiled-coil domain
CCp	Central chloroplast

CCC	Central compartment chloroplast
CHUP1	Chloroplast unusual protein 1
cp-actin	chloroplast-actin
CDS	Coding sequences
CAM	Crassulacean acid metabolism
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GSDS	GENE STRUCTURE DISPLAY SERVER
HD	Hydrophobic domain
HMM	Hidden Markov Model
HPR	Hydroxyproline-rich glycoprotein
JAC 1	J-domain containing protein 1
KAC 1	Kinesin-like protein for actin-based chloroplast movement 1
LZ	Leucine zipper
MS	Mesophyll
MEME	MULTIPLE EM for MOTIF ELICITATION
PEPCase	Phosphoenolpyruvate carboxylase
PCp	Peripheral chloroplast
PCC	Peripheral compartment chloroplast
PHOT	Phototropins
PMI 1	Plastid movement impairment 1
PRM	Proline-rich motif
qRT-PCR	Quantitative real-time Polymerase Chain Reaction
SCC_4	Single-cell C_4
SAS	Statistical Analysis System
TPR	Tetratrico peptide

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This study was supported by ‘‘The Cooperative Research Program for Rural Development Administration (PJ010953032019), Republic of Korea. 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:
The Cooperative Research Program for Rural Development Administration, Republic of Korea: PJ010953032019.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- So Youn Won conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Prabhakaran Soundararajan conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Vadivelmurugan Irulappan conceived and designed the experiments, performed the experiments, analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Jung Sun Kim conceived and designed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

The raw DNA, cDNA, and protein sequences of CHUP1 and its associated proteins from *Bienertia sinuspersici* and raw data for the qRT-PCR experiment are available at GitHub: https://github.com/Vadivelmi/CHUP1_PeerJ/tree/c084025c53e4267269cbbeedd025f5142f0c6fd0.

The *B. sinuspersici* sequences are available at GenBank: BsCHUP1, MT151911; BsCHUP1-like_a, MT151912; BsCHUP1-like_b, MT151913.

Functional domains in AtCHUP1 have been mapped based on *Lehmann, Bohnsack and Schleiff (2011)* and *Lehmann et al. (2011)*. Actin binding domain (ABD) has been identified in CHUP1 by motif ELVYLRWVNA. For CHUP1-like_a and CHUP1-like_b, motif pattern (ELAYLRW[I/V]N[S/A]) has been used to search for the ABD region. For Zea mays, one protein sequence (Zm00001d028761.1) has been chosen as representative of CHUP1-like_b as both shared similar positions of functional domains. LR repeats were predicted by the 2ZIP server (http://2zip.molgen.mpg.de/index.html). In ZmCHUP1 and AcCHUP1, one leucine-rich repeat (marked in star) prediction was unsuccessful with the 2ZIP server. Therefore, manually curating ("L…..L……L……L", one L residue is missing in this tandem which could be a sequence error) based on the theoretical concept of leucine-rich repeats that is the appearance of Leu residues on every seventh amino acid which is arranged as stretches between helix and sheet in protein's hydrophobic core.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.15696#supplemental-information.

REFERENCES

Akhani H, Barroca J, Koteeva N, Voznesenskaya E, Franceschi V, Edwards G, Ghaffari SM, Ziegler H. 2005. *Bienertia sinuspersici* (Chenopodiaceae): a new species from Southwest Asia and discovery of a third terrestrial C4 plant without Kranz anatomy. *Systematic Botany* **30**:290–301 DOI 10.1600/0363644054223684.

Albert VA, Barbazuk WB, Depamphilis CW, Der JP, Leebens-Mack J, Ma H, Palmer JD, Rounsley S, Sankoff D, Schuster SC. 2013. The Amborella genome and the evolution of flowering plants. *Science* **342(6165)**:1241089 DOI 10.1126/science.1241089.

Almagro Armenteros JJ, Sønderby CK, Sønderby SK, Nielsen H, Winther O. 2017. DeepLoc: prediction of protein subcellular localization using deep learning. *Bioinformatics* **33**:3387–3395 DOI 10.1093/bioinformatics/btx431.

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research* **37**:W202–W208 DOI 10.1093/nar/gkp335.

Bondarenko VS, Gelfand M. 2016. Evolution of the exon-intron structure in ciliate genomes. *PLOS ONE* **11**:e0161476 DOI 10.1371/journal.pone.0161476.

Bornberg-Bauer E, Rivals E, Vingron M. 1998. Computational approaches to identify leucine zippers. *Nucleic Acids Research* **26**:2740–2746 DOI 10.1093/nar/26.11.2740.

Chotewutmontri P, Barkan A. 2016. Dynamics of chloroplast translation during chloroplast differentiation in maize. *PLOS Genetics* **12**:e1006106 DOI 10.1371/journal.pgen.1006106.

Copetti D, Búrquez A, Bustamante E, Charboneau JLM, Childs KL, Eguiarte LE, Lee S, Liu TL, McMahon MM, Whiteman NK, Wing RA, Wojciechowski MF, Sanderson MJ. 2017. Extensive gene tree discordance and hemiplasy shaped the genomes of North American columnar cacti. *Proceedings of the National Academy of Sciences of the United States of America* **114**:12003–12008 DOI 10.1073/pnas.1706367114.

Drozak A, Romanowska E. 2006. Acclimation of mesophyll and bundle sheath chloroplasts of maize to different irradiances during growth. *BBA-Bioenergetics* **1757**:1539–1546 DOI 10.1016/j.bbabio.2006.09.001.

Dwyer ME, Hangarter RP. 2022. Light-induced displacement of PLASTID MOVEMENT IMPAIRED1 precedes light-dependent chloroplast movements. *Plant Physiology* **189**:1866–1880 DOI 10.1093/plphys/kiac193.

Gao ZF, Yang X, Mei Y, Zhang J, Chao Q, Wang BC. 2023. A dynamic phosphoproteomic analysis provides insight into the C4 plant maize (*Zea mays* L.) response to natural diurnal changes. *The The Plant Journal* **113**:291–307 DOI 10.1111/tpj.16047.

Ghannoum O, Von Caemmerer SV, Taylor N, Millar AH. 2013. Carbon dioxide assimilation and respiration. In: Munns R, Schmidt S, Beveridge C, Mathesius U, eds. *Plants in action: a resource for teachers and students of plant science*. 2nd ed. Shenton Park Field Station: Australian Society of Plant Scientists. *Available at https://www.rseco.org/content/chapter-2-carbon-dioxide-assimilation-and-respiration.html*.

Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. 2015. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**:1296–1297 DOI 10.1093/bioinformatics/btu817.

Hu Z, Xu F, Guan L, Qian P, Liu Y, Zhang H, Huang Y, Hou S. 2014. The tetratricopeptide repeat-containing protein slow green1 is required for chloroplast development in Arabidopsis. *Journal of Experimental Botany* **65**:1111–1123 DOI 10.1093/jxb/ert463.

Ichikawa S, Yamada N, Suetsugu N, Wada M, Kadota A. 2011. Red light, phot1 and JAC1 modulate phot2-dependent reorganization of chloroplast actin filaments and chloroplast avoidance movement. *Plant & Cell Physiology* **52**:1422–1432 DOI 10.1093/pcp/pcr087.

Jarvis DE, Ho YS, Lightfoot DJ, Schmöckel SM, Li B, Borm TJ, Ohyanagi H, Mineta K, Michell CT, Saber N, Kharbatia NM. 2017. The genome of *Chenopodium quinoa*. *Nature* **542**:307–312 DOI 10.1038/nature21370.

Jarvis P, López J. 2013. Biogenesis and homeostasis of chloroplasts and other plastids. *Nature Reviews Molecular Cell Biology* **14**:787–802 DOI 10.1038/nrm3702.

Jensen PE, Leister D. 2014. Chloroplast evolution, structure and functions. *Fprime Reports* **6**:40 DOI 10.12703/P6-40.

Kadota A, Yamada N, Suetsugu N, Hirose M, Saito C, Shoda K, Ichikawa S, Kagawa T, Nakano A, Wada M. 2009. Short actin-based mechanism for light-directed chloroplast movement in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **106**:13106–13111 DOI 10.1073/pnas.0906250106.

Kobayashi H, Yamada M, Taniguchi M, Kawasaki M, Sugiyama T, Miyake H. 2009. Differential positioning of C4 mesophyll and bundle sheath chloroplasts: recovery of chloroplast positioning requires the actomyosin system. *Plant & Cell Physiology* **50**:129–140 DOI 10.1093/pcp/pcn177.

Kondo A, Kaikawa J, Funaguma T, Ueno O. 2004. Clumping and dispersal of chloroplasts in succulent plants. *Planta* **219**:500–506.

Koteyeva NK, Voznesenskaya EV, Berry JO, Cousins AB, Edwards GE. 2016. The unique structural and biochemical development of single cell C4 photosynthesis along longitudinal leaf gradients in *Bienertia sinuspersici* and *Suaeda aralocaspica* (Chenopodiaceae). *Journal of Experimental Botany* **67**:2587–2601 DOI 10.1093/jxb/erw082.

Lamesch P, Berardini TZ, Li D, Swarbreck D, Wilks C, Sasidharan R, Muller R, Dreher K, Alexander DL, Garcia-Hernandez M, Karthikeyan AS, Lee CH, Nelson WD, Ploetz L, Singh S, Wensel A, Huala E. 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Research* **40**:D1202–D1210 DOI 10.1093/nar/gkr1090.

Lehmann P, Bohnsack MT, Schleiff E. 2011. The functional domains of the chloroplast unusual positioning protein 1. *Plant Science* **180**:650–654 DOI 10.1016/j.plantsci.2011.01.006.

Lehmann P, Nöthen J, Braun SSchmidtvon, Bohnsack M, Mirus O, Schleiff E. 2011. Transitions of gene expression induced by short-term blue light. *Plant Biology* **13**:349–361 DOI 10.1111/j.1438-8677.2010.00377.x.

Li H, Chen L-J. 1996. Protein targeting and integration signal for the chloroplastic outer envelope membrane. *Plant Cell* **8**:2117–2126.

Li P, Ponnala L, Gandotra N, Wang L, Si Y, Tausta SL, Kebrom TH, Provart N, Patel R, Myers CR. 2010. The developmental dynamics of the maize leaf transcriptome. *Nature Genetics* **42**:1060–1067 DOI 10.1038/ng.703.

Long M, Deutsch M. 1999. Intron—exon structures of eukaryotic model organisms. *Nucleic Acids Research* **27**:3219–3228 DOI 10.1093/nar/27.15.3219.

Lupas A, Van Dyke M, Stock J. 1991. Predicting coiled coils from protein sequences. *Science* **252**:1162–1164 DOI 10.1126/science.252.5009.1162.

Mai KKK, Yeung W-T, Han S-Y, Cai X, Hwang I, Kang B-H. 2019. Electron tomography Analysis of thylakoid Assembly and fission in Chloroplasts of a Single-Cell C4 plant, Bienertia sinuspersici. *Scientific Reports* **9**:1–12 DOI 10.1038/s41598-018-37186-2.

Manandhar-Shrestha K, Tamot B, Pratt EP, Saitie SM, Braeutigam A, Weber AP, Hoffmann-Benning S. 2013. Comparative proteomics of chloroplasts envelopes from bundle sheath and mesophyll chloroplasts reveals novel membrane proteins with a possible role in C4-related metabolite fluxes and development. *Frontiers in Plant Science* **4**:65 DOI 10.3389/fpls.2013.00065.

McCormick RF, Truong SK, Sreedasyam A, Jenkins J, Shu S, Sims D, Kennedy M, Amirebrahimi M, Weers BD, McKinley B. 2018. The *Sorghum bicolor* reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *The Plant Journal* **93**:338–354 DOI 10.1111/tpj.13781.

Ming R, Van Buren R, Wai CM, Tang H, Schatz MC, Bowers JE, Lyons E, Wang ML, Chen J, Biggers E, Zhang J. 2015. The pineapple genome and the evolution of CAM photosynthesis. *Nature Genetics* **47**:1435–1442 DOI 10.1038/ng.3435.

Miyake H. 2016. Starch accumulation in the bundle sheaths of C3 plants: a possible pre-condition for C4 photosynthesis. *Plant & Cell Physiology* **57**:890–896 DOI 10.1093/pcp/pcw046.

Offermann S, Friso G, Doroshenk KA, Sun Q, Sharpe RM, Okita TW, Wimmer D, Edwards GE, Van Wijk KJ. 2015. Developmental and subcellular organization of single-cell C4 photosynthesis in *Bienertia sinuspersici* determined by large-scale proteomics and cDNA assembly from 454 DNA sequencing. *Journal of Proteome Research* **14**:2090–2108 DOI 10.1021/pr5011907.

Offermann S, Okita TW, Edwards GE. 2011. Resolving the compartmentation and function of C4 photosynthesis in the single-cell C4 species *Bienertia sinuspersici*. *Plant Physiology* **155**:1612–1628 DOI 10.1104/pp.110.170381.

Oikawa K, Kasahara M, Kiyosue T, Kagawa T, Suetsugu N, Takahashi F, Kanegae T, Niwa Y, Kadota A, Wada M. 2003. Chloroplast unusual positioning1 is essential for proper chloroplast positioning. *Plant Cell* **15**:2805–2815 DOI 10.1105/tpc.016428.

Oikawa K, Yamasato A, Kong S-G, Kasahara M, Nakai M, Takahashi F, Ogura Y, Kagawa T, Wada M. 2008. Chloroplast outer envelope protein CHUP1 is essential for chloroplast anchorage to the plasma membrane and chloroplast movement. *Plant Physiology* **148**:829–842 DOI 10.1104/pp.108.123075.

Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, Childs K, Thibaud-Nissen F, Malek RL, Lee Y, Zheng L, Orvis J, Haas B, Wortman J, Buell CR. 2007. The TIGR rice genome annotation resource: improvements and new features. *Nucleic Acids Research* **35**:D883–D887 DOI 10.1093/nar/gkl976.

Panchy N, Lehti-Shiu M, Shiu S-H. 2016. Evolution of gene duplication in plants. *Plant Physiology* **171**:2294–2316 DOI 10.1104/pp.16.00523.

Park J, Knoblauch M, Okita TW, Edwards GE. 2009. Structural changes in the vacuole and cytoskeleton are key to development of the two cytoplasmic domains supporting single-cell C 4 photosynthesis in *Bienertia sinuspersici*. *Planta* **229**:369–382 DOI 10.1007/s00425-008-0836-8.

Piovesan D, Tabaro F, Paladin L, Necci M, Micetic I, Camilloni C, Davey N, Dosztányi Z, Mészáros B, Monzon AM, Parisi G, Schad E, Sormanni P, Tompa P, Vendruscolo M, Vranken WF, Tosatto SCE. 2018. MobiDB 3.0: more annotations for intrinsic disorder, conformational diversity and interactions in proteins. *Nucleic Acids Research* **46**:D471–D476 DOI 10.1093/nar/gkx1071.

Sage RF. 2004. The evolution of C4 photosynthesis. *New Phytologist* **161**:341–370 DOI 10.1111/j.1469-8137.2004.00974.x.

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

Sage RF, Way DA, Kubien DS. 2008. Rubisco, Rubisco activase, and global climate change. *Journal of Experimental Botany* **59**:1581–1595 DOI 10.1093/jxb/ern053.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA. 2009. The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**:1112–1115 DOI 10.1126/science.1178534.

Sevilleno SS, Ju YH, Kim JS, Mancia FH, Byeon EJ, Cabahug RA, Hwang Y-J. 2020. Cytogenetic analysis of *Bienertia sinuspersici* Akhani as the first step in genome sequencing. *Genes & Genomics* **42**:337–345 DOI 10.1007/s13258-019-00908-5.

Shi Y, Ke X, Yang X, Liu Y, Hou X. 2022. Plants response to light stress. *Journal of Genetics and Genomics* **49**:735–747 DOI 10.1016/j.jgg.2022.04.017.

Slewinski TL. 2013. Using evolution as a guide to engineer Kranz-type C4 photosynthesis. *Frontiers in Plant Science* **4**:212 DOI 10.3389/fpls.2013.00212.

Soundararajan P, Won SY, Park DS, Lee Y-H, Kim JS. 2019. Comparative Analysis of the YABBY Gene Family of *Bienertia sinuspersici*, a Single-Cell C(4) Plant. *Plants* **8(12)**:536 DOI 10.3390/plants8120536.

Suetsugu N, Higa T, Kong S-G, Wada M. 2015. PLASTID MOVEMENT IMPAIRED1 and PLASTID MOVEMENT IMPAIRED1-RELATED1 mediate photorelocation movements of both chloroplasts and nuclei. *Plant Physiology* **169**:1155–1167 DOI 10.1104/pp.15.00214.

Suetsugu N, Wada M. 2016. Evolution of the cp-actin-based motility system of chloroplasts in green plants. *Frontiers in Plant Science* **7**:561 DOI 10.3389/fpls.2016.00561.

Suetsugu N, Yamada N, Kagawa T, Yonekura H, Uyeda TQ, Kadota A, Wada M. 2010. Two kinesin-like proteins mediate actin-based chloroplast movement in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **107**:8860–8865 DOI 10.1073/pnas.0912773107.

Sunil M, Hariharan AK, Nayak S, Gupta S, Nambisan SR, Gupta RP, Panda B, Choudhary B, Srinivasan S. 2014. The draft genome and transcriptome of *Amaranthus hypochondriacus*: a C4 dicot producing high-lysine edible pseudo-cereal. *DNA Research* **21**:585–602 DOI 10.1093/dnares/dsu021.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**:W609–W612 DOI 10.1093/nar/gkl315.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**:2725–2729 DOI 10.1093/molbev/mst197.

Von Braun SS, Schleiff E. 2008. The chloroplast outer membrane protein CHUP1 interacts with actin and profilin. *Planta* **227**:1151–1159 DOI 10.1007/s00425-007-0688-7.

Wada M. 2016. Chloroplast and nuclear photorelocation movements. *Proceedings of the Japan Academy* **92**:387–411 DOI 10.2183/pjab.92.387.

Wang L, Ma G, Wang H, Cheng C, Mu S, Quan W, Jiang L, Zhao Z, Zhang Y, Zhang K, Wang X, Tian C, Zhang Y. 2019. A draft genome assembly of halophyte *Suaeda aralocaspica*, a plant that performs C4 photosynthesis within individual cells. *GigaScience* **8**:giz116 DOI 10.1093/gigascience/giz116.

Wang P, Moore BM, Panchy NL, Meng F, Lehti-Shiu MD, Shiu S-H. 2018. Factors influencing gene family size variation among related species in a plant family, Solanaceae. *Genome Biology and Evolution* **10**:2596–2613 DOI 10.1093/gbe/evy193.

Wimmer D, Bohnhorst P, Shekhar V, Hwang I, Offermann S. 2017. Transit peptide elements mediate selective protein targeting to two different types of chloroplasts in the single-cell C4 species *Bienertia sinuspersici*. *Scientific Reports* **7**:41187 DOI 10.1038/srep41187.

Yamada M, Kawasaki M, Sugiyama T, Miyake H, Taniguchi M. 2009. Differential positioning of C4 mesophyll and bundle sheath chloroplasts: aggregative movement of C4 mesophyll chloroplasts in response to environmental stresses. *Plant & Cell Physiology* **50**:1736–1749 DOI 10.1093/pcp/pcp116.

Yang X, Cushman JC, Borland AM, Edwards EJ, Wullschleger SD, Tuskan GA, Owen NA, Griffiths H, Smith JAC, De Paoli HC. 2015. A roadmap for research on crassulacean acid metabolism (CAM) to enhance sustainable food and bioenergy production in a hotter, drier world. *New Phytologist* **207**:491–504 DOI 10.1111/nph.13393.

Yang Y, Sage TL, Liu Y, Ahmad TR, Marshall WF, Shiu S-H, Froehlich JE, Imre KM, Osteryoung KW. 2011. CLUMPED CHLOROPLASTS 1 is required for plastid separation in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America* **108**:18530–18535 DOI 10.1073/pnas.1106706108.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**:1586–1591 DOI 10.1093/molbev/msm088.