Re-examining the origin and functions of the urea cycle

Xiliang Song (sxl0424@126.com)
Shandong Agricultural University
Linhong Teng
Dezhou University
Ximei Zhao
Binzhou University

Research Article

Keywords: urea cycle, SAR supergroup, evolution, carbamoyl phosphate synthase, arginase

Posted Date: April 13th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1532969/v1

Abstract

Urea cycle (UC) was first discovered in vertebrates and functioned to remove excess ammonia. To better understand the origination and evolutionary history of UC, we performed a comprehensive investigation on UC genes distribution among various lineages. We found that complete UC genes are widely distributed from prokaryotes to eukaryotes. These genes underwent purifying selection during evolution. The first enzyme carbamoyl phosphate synthase (CPS) fused and duplicated in the common ancestor of metazoans and SAR, formed pyrimidine specific and arginine/urea specific CPS. Prokaryotes and green plants retain one generalized CPS composed of separated small and large subunits. A rather unusual finding was that ARG in plants has closer relationship with agmatinase (AGM) than non-plant ARG. in silico expression analysis shows that UC genes in Arabidopsis are up-regulated in seedlings and stress conditions. ARG in Arabidopsis exhibited coordinately expression with arginine decarboxylase, the first enzyme to produce putrescine. Taken together, UC has increased in importance in vertebrates and SAR supergroup through duplication of the first enzyme. ARG in plants more resemble the paralogous AGM of prokaryotes. In plant species lacking ornithine decarboxylase, ARG were predicted to work as AGM, which participate in polyamine biosynthesis with urea as a byproduct.

Introduction

Urea cycle (UC) was first discovered in terrestrial vertebrates in 1932 by Hans Krebs (Haines et al. 2011). The most significant function of UC is to remove excess nitrogen, which is produced in most living cells through amino acids oxidation (Blair et al. 2015). In marine elasmobranchs, excreted urea to body fluids is utilized to maintain osmotic balance (Nawata et al. 2015). Five enzymes are directly involved in the cycle. The enzyme carbamoyl phosphate synthase (CPS) drives the first step in which carbamoylphosphate is formed from CO_2, ATP and ammonia or glutamine (Cima et al. 2015). Strictly speaking, CPS is not a committed UC member. It catalyzes the common step in pyrimidine synthesis, arginine or urea synthesis (Guo et al. 2015). One generalized or two specific CPS enzymes may exist in different organisms (Lawson et al. 1996). The enzyme consists of one small subunit and one large subunit, corresponding to the amidotransferase domain and synthase domain respectively. In different organisms, the enzymes can be heterodimeric, monomeric, or form a multifunctional protein with other domains, all of which reflects the complex evolutionary history (Holden et al. 1999). The remaining four enzymes of UC include ornithine transcarbamoylase (OTC), argininosuccinate synthetase (ASS), argininosuccinate lyase (ASL) and arginase (ARG). Among them, ASS plays as the rate—limiting step in the conversion of citrulline to argininosuccinate, the immediate precursor of arginine (Haines et al. 2011). ASL are highly conserved from prokaryotes to eukaryotes, which splits argininosuccinate to release fumarate and arginine (Cao et al. 2011). The last enzyme ARG, which catalyzes arginine into ornithine and urea, plays a significant role in nitrogen metabolism as well as in defense responses (Labudda et al. 2016).

Although there are earlier reports showing that UC have originated in metazoans (Anderson 1980; Mommsen and Walsh 1989), accumulating researches suggested an ancient occurrence. Arginine

catabolism in the unicellular cyanobacterium Synechocystis involves the UC and arginase pathway, evidenced by the sequential presence of ^{14}C labeled amino acids production in the UC process. Strangely, proteins responsible for the arginase activity were arginase—related enzymes rather than true arginase (Quintero et al. 2000). One genome sequenced Archaea *Haloarcula marismortui* breaks down arginine via the UC pathway to produce ornithine and urea. Ornithine may be converted into arginine or glutamate, accompanied with the conversion of aspartate to fumarate. This energy consuming pathway might aim to convert excess organic nitrogen absorbed from environments into functional intermediates (Baliga et al. 2004). Since genome sequence of marine diatoms *Thalassiosira pseudonana* was uncovered, researchers found the unanticipated integration of UC enzymes in diatom, suggesting UC also exists in eukaryotic photoautotroph, even much earlier than metazoans (Armbrust et al. 2004; Allen et al. 2011; Prihoda et al. 2012). Diatom, together with other members in SAR supergroup are believed to be evolved from secondary endosymbiosis events, in which a non—photosynthetic organism engulfed a photosynthetic organism (Archibald 2015). Molecular phylogenetic data support that their genome contains red algae derived genes as well as endosymbiotic host genes, and increasing green algal derived genes are also discovered (Chan and Bhattacharya 2013). Study on the CPS gene evolutionary history inferred that the UC in Stramenopile and haptophyte evolved in the exosymbiont before plastid acquisition and the pathway was absent in Archaeplastida (Allen et al. 2011). The authors also conducted critical experiments to investigate how diatoms utilized the UC pathway. They proved that through UC *Phaeodactylum tricornutum* can fully utilize replete nitrogen source under the episodic nutrient—rich conditions in upwelling event. Moreover, many intermediates in UC are important precursor metabolites in TCA cycles and cell wall formation, all of which are essential for diatom growth and prosperity. Besides, proteomics of T. *pseudonana* also revealed that only under iron—replete condition the algae cell can express a complete UC enzymes, and when iron—limited, the last enzyme ARG was not detected, suggesting that urea performed as a N storage molecule under favorable conditions (Brook 2013). In addition to diatoms, haptophyte *Emiliania huxleyi* also uses the ornithine—urea cycle as an efficient metabolic budgeting to rapidly turn over cellular nitrogen (Rokitta et al. 2014). Genes of UC are also present in dinoflagellates, which can perform rampant cell division in high nitrogen concentration (Dagenais-Bellefeuille and Morse 2013). Recently, Horák et al. (2020) proved that metazoan and stramenopiles share the common origins of the OUC enzymes. Taken together, these suggested that UC may be a general adaptation mechanism imposed on SAR species to nutrient scarce environments.

Functional UC has also been reported in many plant species. The tracing experiment of $^{14}CO_2$ produced from ^{14}C labeled arginine in soybean cotyledons indicated that arginine was catabolized by the ARG, giving rise to ornithine and urea, the later was then hydrolyzed by urease (Micallef and Shelp 1989). Afterwards, it was evaluated that approximately 20% of provided exogenous arginine was converted to urea and ornithine in soybean (Goldraij and Polacco 1999). Besides, Arginine degradation by ARG was observed along with the rapid increase in urea in urease—deficient soybean seedlings (Palmieri et al. 2006). ARG gene expression was enhanced in *Arabidopsis* following fungal infection (Brauc et al. 2012). In addition, increased ARG expression was also observed in seedling germination of loblolly pine (Todd et

al. 2001). The multiple roles of ARG in development and defense response were largely determined by its key substrate arginine, which is a precursor for polyamines and proline synthesis (Caldwell et al. 2015).

Taken together, ornithine—urea cycle should play multiple roles in organisms. Its intermediates and derivatives have participated in kinds of biochemical pathways. Triggered by the breakthrough in diatom genome, we hope to better understand the origination and evolutionary history of UC. With the advantage of genome sequencing, we can comprehensively explore the UC enzymes in whole genome level. In the present study, through mining the whole genome—wide protein sequences of various kingdoms, including bacteria, Archaea, metazoan, SAR and Archaeplastida, we identified the various enzymes involved in UC and branching pathways. We try to depict the origination and functional differentiation of UC network in the course of organism evolution. Our work would give new insight into the evolutionary story of urea production.

Materials And Methods

Sequence source

Whole genome—wide protein sequences of representative bacteria, Archaea, metozoan, red algae, and SAR were downloaded from NCBI https://www.ncbi.nlm.nih.gov/. Green algae and plants' genome protein and CDS sequence were obtained from Phytozome V11.0 https://phytozome.jgi.doe.gov/pz/portal.html. Besides, we use the transcriptome sequence of Ulva linza as one representative of macro green algae (Zhang et al. 2012). The detailed species information was listed in Table S1.

UC genes Identification

The five UC enzymes CPS, OTC, ASS, ASL, ARG were downloaded from NCBI's non—redundant protein database. Using them as queries, we performed local BLAST analysis against each organism's proteome, with an E—value cutoff 1*10—1. For every BLAST result, we submit the sequence to online BLAST in NCBI website to verify the protein identity. Meanwhile, to further verify the presence and number of each enzyme, we downloaded the HMM profile of the enzymes from the Pfam website: http://pfam.xfam.org/ (CPS: PF02787, OTC: PF00185, PF02729, ASL:PF14698, ASS: PF00764, ARG and AGM: PF00491). HMMER3.3.2 software (http://hmmer.org/) with the default parameters was used to search for HMM domain in the proteomes of each species. The results of enzyme number for each species was listed in Table S1. In addition, UC—related enzymes were also identified in each organism, including ADC: arginine decarboxylase (PF17944, PF17810); AIH: agmatine iminohydrolase (PF04371); NCPAH: N—carbamoylputrescine amidohydrolase (PF00795); OCD: ornithine cyclodeaminase (PF02423); ODC: ornithine decarboxylase (PF00278, PF02784); OAT: ornithine aminotransferase (PF00202); P5CR: pyrroline—5—carboxylate reductase (PF14748).

Phylogenetic analysis

To explore the evolutionary history of UC enzymes, phylogenetic trees were constructed using the full—length protein sequences of each enzyme. Three methods were applied. Firstly, proteins were aligned using the Clustal W implemented in MEGA7 (Kuma et al. 2016) using the default parameters, that is, Gonnet protein weight matrix, pairwise gap opening/extension penalties of 10/0.1, and multiple alignment gap opening/extension penalties of 10/0.2/. Phylogenetic relationships within each enzyme were reconstructed using the maximum likelihood (ML) procedure performed in MEGA 7 under the best models found by MEGA with 1000 replicates of bootstraps to obtain the confidence support. The partial deletion method with 10% cut off was chosen to remove the gaps with less than 10% coverage. Second, the proteins sequences were aligned using the MAFFT in Phylosuite (Zhang et al. 2020), then the alignment were trimmed under “gappyout” mode. The modelfinder in Phylosuite was performed to search the model used in IQtree and MrBayes respectively based on the trimmed alignment. Then the models were imported into IQtree with 1000 standard bootstrap. The MrBayes tree was constructed using the model chosen by modelfinder, with four different chains and more than 100000 generations until they converged, that is, the average standard deviation of split frequencies was lower than 0.01, and the effective sample size of model parameters reached 200. Posterior probabilities represent statistical support.

Selection pressure analysis

To evaluate the selection pressure in the five lineages, i.e., Stramenopiles, red algae, green algae, Streptophyta and metazoan, respectively, rates of evolution and selection pressure were evaluated by calculating the ω (dn/ds ratio, dn, non—synonymous substitution rates, ds, synonymous substitution rates). Firstly, the protein sequences were aligned using clustalW with default parameters. Then the codon sequence were aligned accordingly using pal2nal with the choice that remove gaps, inframe stop codons and mismatches (http://www.bork.embl.de/pal2nal/). PAML format of the aligned CDS sequences was generated for further selection analysis by the PAML v4.8 package (Yang 2007). One—ratio model (model=0) in Codeml program was used to calculate the ω within each group. To compare the ω distribution of the five enzymes within Stramenopiles, the pairwise dn/ds, dn and ds were calculated by the yn00 program. To determine whether positive selection had affected specific amino acid sites in each enzyme, site model in the PAML package were performed under M7 and M8 model.

In Silico Expression analysis of model plant *Arabidopsis thaliana*

To evaluate the UC function in plants, we investigated the gene expression profile data of the model plant *Arabidopsis*. The expression data of different tissues, including root, seedlings, young leaves, old leaves, stems, flowers, vegetative shoot meristem, inflorescence shoot meristem, whole inflorescences and fruits in *Arabidopsis* were downloaded from public database (http://jsp.weigelworld.org) (Laubinger et al. 2008) and they were log2 transformed. Microarray expression data of the abiotic, biotic stress, hormonal

treatments and light treatments were obtained from an earlier study of Ma and Bohnert (2007). Heat map of expression level was constructed in the statistical computing language R using the heatmap.2 function in the gplots package.

Protein—protein interaction (PPI) prediction

To further analyze the possible function involved in the downstream UC, the last critical enzyme ARG sequences from model diatom T. *pseudonana*, model plant *Arabidopsis thaliana* and model animal *Mus musculus* L. were submitted to uniprotKB website (http://www.uniprot.org/help/uniprotkb) to get the interaction networks under STRING 10 section.

Results

UC genes identification and Taxonomic Distribution Analysis

To systematically examine the phylogenetic distribution of UC genes in kinds of kingdom, we performed the local BLAST and HMM search using proteome sequences of 50 genomes, combined with the online search verification. The detailed taxonomic distribution was listed in Figure. 1 and Table S1. UC genes appear to have an ancient origin as evidenced by their presence in Prokaryotes, including Archaea and bacteria. Among the 50 species investigated, completed UC genes are present in most SAR species (except Alveolata species, *Plasmodium falciparum* and *Babesia bovis*, Oomycetes *Phytophthora infestans*), vertebrates, Streptophyta, fungi, two of Archaea (*Natrinema pellirubrum* and *Haloterrigena tukmenica*), three of bacteria *(Deinococcus maricopensis*, *Marinithermus hydrothermalis* and Bacillus subtilis), but absent in invertebrates, green and red algae, cyanobacteria, two of Archaea (*Hyperthermus butylicus* and *Sulfolobus solfataricus*), three of bacteria (*Treponema brennaborense*, *Mycobacterium tuberculosi*s, and *Escherichia coli*).

CPS catalyzes the first step for both pyrimidine and arginine/urea biosynthesis. During the BLAST search, we found that two CPS copies exist in SAR, vertebrates, red algae and fungi. Both the fused protein and the separated large and small subunits were obtained. SAR and vertebrates have two copies of fused CPS genes uCPS and pCPS while red algae and fungi have one separated and one fused genes. The small amidotransferase subunit (carA, 1—376 amino acid residues in *Ectocarpus siliculosus* uCPS) and the large synthase subunit (carB, 392—1445 residues in E. *siliculosus* uCPS) fused into one gene. Likewise, the two subunits of pCPS fused together in SAR and metazoan. For metazoan, fungi and red algae, the pCPS also fused with PyrB domain, which encodes the catalytic chain of aspartate carbamoyltransferase involved in pyrimidine biosynthesis. Invertebrate species have only one fused CPS. Only one copy of CPS occurred in green plants, Archaea and Bacteria, excepted one gram positive bacterium Bacillus, which has two CPS. Notably, CPS in plants and prokaryotes are all composed of separated large and small subunit, represented the ancient state of CPS genes.

The other three UC members, OTC, ASS, and ASL generally appeared as single copy in all species, except several protists and lower metazoan lacking them. OTC is absent in the invertebrates surveyed. Alveolata species P. *falciparum* and B. *bovis* lack all the three inter—enzymes. The issue on last critical enzyme ARG seems to be complicated. Firstly, ARG are completely absent in all red algae and green algae investigated. SAR, except the Alveolata species, have one ARG. The number of ARG in Prokaryotes ranged from 0 to 2. Duplicated ARG are also found in metazoan, fungi, and higher plants. In addition to the five enzymes in UC, we also retrieved other UC related enzymes, including arginine decarboxylase (ADC), agmatine iminohydrolase (AIH, also known as agmatine deininase), N—carbamoylputrescine amidohydrolase (NCPAH) and agmatinase (AGM), all of which initiate the alternate pathway in urea production using arginine as substrate. Ornithine cyclodeaminase (OCD) and ornithine decarboxylase (ODC) are two enzymes catalyzing the formation of proline and polyamine from ornithine. ADC is only found in green plants and bacteria. AGM exist in SAR, cyanobacteria, Archaea, fungi, vertebrates, and some proteobacteria. OCD are only found in brown algae, diatoms, and Archaea, while absent in other groups. ODC exist in most organisms, but absent in *Arabidopsis*, *Zostera marina* and *Spirodela polyrhiza*. Despite that we performed comprehensive BLAST and HMM search, we cannot rule out the possibility that the genes were missed during the genome assembly.

Phylogenetic events of the five enzymes

To further understand the evolutionary events that gave rise to the UC pathway, we constructed the phylogenetic trees using the protein sequences of the five enzymes in 50 species ranging from prokaryotes to eukaryotes. The topology of all the trees was generally consistent with the results reported by Horák et al. (2020). Of the five UC genes, the most intricate evolution events were seen in CPS enzyme. As this enzyme composed of two subunits and undergone fusion or division in different species, in order to make accurate comparison, we performed sequences splice and concatenation before tree construction. Then the trees were constructed using complete CPS sequences, just the small subunit, and just the large subunit, respectively. Trees constructed using only small subunit did not give well resolution, while the trees with complete sequences and large subunits sequences have the same topology and complete sequences tree showed higher bootstrap support (Figure. 2A, Figure. S1 and S2). The tree was divided into two large clades, one comprising green plant, red algae and prokaryotes, while the other clade comprising archaea, fungi, red algae, SAR and metazoa. Green plant, together with prokaryotes were positioned at a primary place, suggested the ancestral heterodimeric CPS enzymes. The gene duplication event occurred in the common ancestor of SAR and metazoa. The duplicated gene passed through gene fusion and sub—functionalization, formed pyrimidine specific and arginine/urea specific enzymes. Gene loss has frequently occurred in the species absent of UC, such as invertebrates, in which only one pyrimidine specific CPS was retained.

The phylogenetic tree of the three enzymes OTC, ASS and ASL were shown in Figure. 2B, C, and D and Figure. S3, S4 and S5. Generally, prokaryotes in these trees are located at more primary place, suggesting an ancient origination. Interestingly, many of the enzymes, as well as CPS of Stramenopiles species are

animal—like, that is, they have closer relationship with metazoans but not green plant or red algae, indicating they are possibly derived from the heterotrophic secondary host. The similar conclusion was drawn by Horák et al. (2020) which found that the stramenopiles—metazoan/opisthokont grouped together.

The tree of the last enzyme, ARG, exhibited an unanticipated picture (Figure. 3A). Genes from SAR, fungi and matazoan clustered together, then grouped with prokaryotes. While ARG from green plants distantly related to them. When added the AGM into the tree, ARG of green plants robustly form a clade with AGM of prokaryotes. When taken the sequence identity into account, the identity within ARG clade, AGM clade ranged from 23.81%—90.32% and 25.64%—90.84%, respectively, while the identity between them ranged from 20.37%—63.88%, suggesting ARG and AGM are two different enzymes or duplicated paralogs and the ARG of green plants are more homologous with AGM. Consistently, when we explored the ADC, which cooperates with AGM to complete the catabolism of arginine, they were only found in green plants (no green algae) and bacteria (Figure. 3B), further indicating that green plants use the ancient ADC and AGM route to perform arginine metabolism.

Selection pressure results

We evaluated the selection pressure of the five enzymes in the above mentioned species through calculating ω (dn/ds). We found that the ω values of all the organisms were far less than 1 ($0.0027<\omega<0.0355$), suggesting that these enzymes have undergone purifying selection during evolution (Figure. 4). Furthermore, the ω value in Stramenoplies is generally higher than other organisms, indicating the relatively relax pressure. Considering that the ratio averaged over all lineages is almost never >1, since positive selection is unlikely to affect all sites over prolonged time, we further performed the site model to detect the positive selection sites. Most enzymes in Stramenoplies have several positive selection sites, suggesting the episodic positive selection might occur during the evolution of UC enzymes in Stramenoplies. To further compare among the five enzymes, we calculated the pairwise dn/ds, dn, and ds values within Stramenoplies. Dn/ds and dn values of ARG are significantly higher than other enzymes, suggesting this enzyme is under more relaxed pressure during evolution. The ds value is much higher than that of dn, and is similar among different genes, therefore ds was suggested to be used as a molecular clock for evaluating the evolutionary time. When comparing the ds values, we found that both pCPS and uCPS are larger than the other four enzymes, indicating CPS is a more ancient enzyme.

Expression of UC—related enzymes in model plant *Arabidopsis*

To further understand the potential role of OUC in green plants, we investigated the expression of UC members and related enzymes in different tissues and stress conditions of the plant model *Arabidopsis* (Figure. 5). Generally, the enzymes were differentially expressed in ten tissues, whereas some enzymes

showed tissue—specific expression patterns. ASL was only highly expressed in seedling, and ARGAH1 are just up—regulated in old leaf and flower. ARGAH2 was up—regulated in many tissues but down—regulated in root and stem. On the other hand, the completed UC genes are highly expressed in seedlings, indicating the active arginine and urea metabolism in seedling plants. As for the expression profile in stressed conditions, we found that the first four UC genes are not stress responsive in most conditions, except ASS, which is up—regulated under light and hormone stimuli. However, ARG was obviously influenced by biotic or abiotic stresses, since it was up—regulated under salt, drought, wounding, hormone, pathogen infection or light stimuli, suggesting arginine or ornithine as a precursor for polyamine play important roles in defense response. Notably, as one important arginine catabolism pathway, ADCs are up—regulated in seedlings and most stress conditions, but AIH and NCPAH not. On the contrary, ADC exhibited similar expression profiles with ARG. Taken together, ARG and ADC enzymes play significant roles in both seedling development and stresses response.

Protein interaction network associated with UC enzymes

To further evaluate the UC function in different organisms, we predicted the PPI networks associated with the ARG enzymes. Among the organisms examined, we obtained the networks of diatoms T. *pseudonana*, plant model *Arabidopsis*, and animal model *Mus musculus* (Figure. 6 and Figure. S6). In the network of T. *pseudonana*, ARG was associated with ten proteins, two of which (OTC, ASL) are members of UC. Other proteins are involved in UC derived pathways, such as URE, which degraded urea. ODC, OCD, and Spermine synthase are also connected, which participate in polyamine and proline synthesis. *Arabidopsis* ARGAH1 and ARGAH2 genes show different networks. For ARGAH1, two UC members are involved in, such as OTC and ASL. URE is also directly connected with ARG. Besides, ADC1 and ADC2, which synthesize agmnine from arginine, are involved in. Compared with ARGAH1, ARGAH2 interacted with diverse proteins that participated in wider range of biological processes, such as spermidine synthesis, xylem specification, lysine synthesis, as well as transcription factors interaction. This indicated that diverged paralogs of ARG in *Arabidopsis* has undergone neo—functionalization and developed into various functions. On the other hand, ARG in M. *musculus* interacted with several NO synthase apart from the UC—related enzymes. Despite it is connected with ADC, we noted that ADC in animal actually act as an antizyme inhibitor and prevents degradation of ODC. More detailed description about the interacted proteins was included in Table S2.

Discussion

UC as an ancient metabolism distributed widely from Prokaryotes to Eukaryotes

The ornithine—urea cycle is an enzyme catalyzing system and long known for its significant role of ammonia detoxication in vertebrates, but little is known about its evolution and taxonomic distribution patterns. To elucidate these questions we performed genome—wide exploration of UC genes in various

organisms ranging from prokaryotes to eukaryotes and performed phylogenetic analysis. Distinct distribution patterns are found among different classification lineages. Especially, both prokaryotic and eukaryotic microorganisms have the completed UC genes. Actually, completed UC genes have been reported in the genome of an Archaea, *Haloarcula marismortui* (Baliga et al. 2004). Attention should also be paid that although completed UC genes was not found in the three invertebrate species surveyed, the complete set of UC genes does exist in invertebrates, such as sea urchin (Goel and Mushegian 2006).

CPS catalyzes the formation of carbamoylphosphate from ATP, bicarbonate, and either ammonia or glutamine for both pyrimidine and arginine/urea synthesis (Cima et al. 2015). The evolutionary history of CPS gene gives one way to evaluate nitrogen metabolism along with organism evolution. Despite the presence of CPS across all the 50 species surveyed, varied patterns of gene losses were also observed. For the Apicomplexan species, *Symbiodinium kawagutii* lost one carA subunit, while P. *falciparum* and B. *bovis* lost one entire copy of CPS. Earlier study showed that the CPS duplication occurred after the divergence of Apicomplexan as they are located at the primary branch of the eukaryotic tree (Lawson et al. 1996). Actually, Apicomplexan, together with other members of SAR, formed a robust branch of pCPS, indicating the later loss in Apicomplexan. The phylogenetic tree supported that the duplication event happened before the differentiation between metazoan and SAR, which was similarly concluded in diatom UC evolution (Allen et al. 2011). In addition, duplication likely occurred after the branching off of the plants and before the separation of fungi from other eukaryotes as all the plants studied contain only one CPS while fungi contain two. Gene duplication is a key source to generate new genes with sub/neo—functionalization (Acharya and Ghosh 2015). Since the CPS enzyme can participate in two competing metabolic pathways, it would be an advantage for the metazoan and SAR organisms having two CPS enzymes. Hence, the UC has increasing significance during the evolution of their common ancestor. On the other hand, plants have only one generalized CPS. Red algae *Porphyridium purpureum* lost carA subunit while *Pyropia yezoensis* lost both the carA and carB subunits. In prokaryotes, T. *brennaborense* have only one carA but no carB was detected. All of these suggested their degenerated UC in the course of evolution. In our study, we did not observe completed loss of CPS across 50 species. However, it does occur in some prokaryotes. Symbionts such as *Haemophylus influenza*, and intracellular parasites such as Chlamydiae have lost CPS. Besides, CPS was also absent in some Archaea such as *Pyrococcus horikoshi* (Cammarano et al. 2002).

In those lineages lacking complete UC genes, especially the last critical enzyme ARG, we wonder how the arginine can be utilized. In fact, the first four UC enzymes catalyze the biosynthesis of arginine whereas ARG is usually absent in some lineages. Two cyanobacteria, *Synechocystis* sp. and *Pleurocapsa minor* only have AGM instead of ARG. Consistently, they have one or two ADC, which cooperates with AGM, to complete the catabolism of arginine. Interestingly, it was reported that *Synechocystis* sp. could fulfill the urea cycle and ARG pathway, whereas no ARG homolog was identified (Quintero et al. 2000). In addition, green and red algae are absent of ARG, as well as AGM and ADC, therefore both the ARG pathway and AGM pathway is not operating. It was documented that arginine catabolism in *Chlamydomona*s was conducted through the bacterial—type deiminase pathway (Vallon and Spalding 2009), which produce

citrulline and NH_3. In our study, we confirmed that green and red algae possess arginine deiminase (AD), by which arginine was degraded to produce re—utilizable nitrogen.

SAR possess comprehensive UC branching pathways

The UC of vertebrates evolved following the ammonia detoxication demand (Blair et al. 2015). The roles of UC in adaptation mechanisms of N stress for diatoms, dinoflagellates and *Emiliania huxleyi* have been revealed (Allen et al. 2011; Bender et al. 2011; Brook 2013; Dagenais-Bellefeuille and Morse 2013; Rokitta et al. 2014). Besides, the significant roles of UC could be expanded by several offshoots, which should come from the complex evolutionary history. Arginine is an important precursor for polyamine and proline synthesis. We found that SAR species, at least for diatom and brown algae, have the most comprehensive branching pathways for both polyamine and proline biosynthesis (Figure. 7). Proline is an essential component of proteins and has important roles in adaptation to osmotic stress (Fichman et al. 2015). In all the eukaryotes investigated, proline was synthesized by the action of ornithine aminotransferase (OAT) and pyrroline—5—carboxylate reductase (P5CR). While diatom and brown algae have additional pathway, in which ornithine was converted directly to proline by OCD, which was likely acquired by lateral transfer from Archaea. On the other hand, SAR have three routes to synthesize putrescine, namely ODC, AGM and AIH—NCPAH pathway. Of these, ODC exist in most Eukaryote but absent in some plants. AGM exist in metazoan, but absent in plants. AIH—NCPAH is absent in metazoan. Despite the various routes in SAR, gene loss did occur in some species. For example, AIH—NCPAH is absent in T. *pseudonana* and *oomycetes*. We think that the diverse metabolic pathways present in SAR should be largely attributed to its distinctive evolutionary footprint. Secondary endosymbiosis theory believes that the ancestor of SAR engulfed (or was invaded by) a red alga, which formed the primary plastid (Archibald 2015; Ku et al. 2015). Moreover, green algae genes were also discovered in SAR nuclear genome (Burki et al. 2012; Deschamps and Moreira 2012). All of these endosymbionts, together with exosymbiont constructed the mosaic genomes of SAR. In addition, frequently horizontal gene transfer was considered as an important source of genetic novelty in SAR (Nosenko and Bhattacharya 2007). It was reported that ODC in diatoms was red algal origin whereas OCD and AGM were bacterial origin (Prihoda et al. 2012). The complex evolutionary history may largely contribute to the diversification and prosperity of SAR.

Diversified Functionalization of UC in model plant *Arabidopsis*

Both SAR and metazoan have evolved out UC specialized CPS and have enhanced UC, which seemed to be a futile urea cycle in higher plants. It seems to be paradoxical for plants possessing UC pathway as it is normally employed to excrete excess ammonia in vertebrates, while urea or ammonia is important nitrogen source for plant growth. Nonetheless, UC metabolism was often mentioned in plant species (Micallef and Shelp 1989; Goldraij and Polacco 1999; Palmieri et al. 2006). To better under the UC roles in

plants, we compared the gene expression level in different tissues as well as diverse stress conditions. The up—regulation of UC genes in seedling and stress conditions was consistent with earlier result in soybean or tomato (Micallef and Shelp 1989; Goldraij and Polacco 1999; Chen et al. 2004). Arginine is a nitrogen—rich amino acid and represents a major component of nitrogen pool in seed (Cortés-Giraldo et al. 2016). During seedling growth, arginine was actively broken down to urea and ornithine, which provides amino acids for protein synthesis (Goldraij and Polacco 1999). More importantly, ARG is a key control point linking UC to polyamine synthesis. ARG expression under stress conditions for both animals and plants is largely involved in the production of polyamines, which are required for diverse functions in all eukaryotes, such as cell proliferation, xylem differentiation and normal cellular physiology (Chen et al. 2004; Alessandra Tisi 2011; Liu et al. 2015). *Arabidopsis* and *Solanum* have two ARG, respectively, namely ARGAH and LeARG according to literature (Chen et al. 2004; Brownfield et al. 2008). Expression of ARGAH1 and ARGAH2 in *Arabidopsis* were reported to be tissue—specific and methyl jasmonate induced, respectively (Brauc et al. 2012). When infected with necrotrophic pathogen *Botrytis cinerea*, the putrescine level in *Arabidopsis* increased rather than decreased in the transgenic lines overexpressing ARGAH2, and the overexpressing lines confers increased tolerance (Brauc et al. 2012). ARGAH1 in root is diminished whereas ARGAH2 is enhanced in shoots of *Arabidopsis* when infected with *Heterodera schachtii* (Labudda et al. 2016). Moreover, LeARG2 in tomato was induced by wounding and jasmonate (Chen et al. 2004). The up—regulation of UC genes, particularly for ARG under both abiotic and biotic stress conditions in our results is consistent with previous studies and suggests the diversified functions of this route. Nonetheless, the mechanisms underlying these functions remain elusive.

ARG in Arabidopsis possibly functions as AGM and participates in polyamine biosynthesis

ARG is actually a small gene family, which includes arginase, agmatinase, formiminoglutamase, and proclavaminate amidino hydrolase (Chen et al. 2004). In our study, plants ARG sequences are unexpectedly closer to bacterial AGM (c. 30% sequence identity) than to non—plant ARG from SAR, vertebrates, fungi, and bacteria (c. 24% sequence identity). However, studies on tomato ARG suggested that LeARG show higher affinity with l—arginine compared to agmatine (Chen et al. 2004) and they concluded that LeARGs are authentic ARG. In *Arabidopsis*, ARGAH1 was identified by complementing yeast ARG deficient (car1) mutant, and ARGAH2 was later identified by sequence homology, whereas no direct enzymatic data supports ARGAHs' substrates specificity (Brownfield et al. 2008). In the present study, we inferred that the mechanisms underlying ARGAHs' response may not be the same as LeARG. Ornithine produced by ARG was catalyzed by ODC to form putrescine (Dalton et al. 2016). Notably, it should be emphasized that the ODC gene has been lost from *Arabidopsis*, Z. *marina* and S. *polyrhiza*. It has been proved that plants possess an additional route for putrescine synthesis from arginine. The pathway from arginine to putrescine is comprised of ADC, AIH and NCPAH (Janowitz et al. 2003; Fuell et al. 2010). Therefore, *Arabidopsis* was thought to have this only way to produce putrescine (Hanfrey et al. 2001; Liu et al. 2015). However, we investigated the expression of these genes in different tissues and

various stress conditions. Surprisingly, expression patterns of ADC, AIH and NCPAH were not coordinately regulated. In detail, when ADC is up—regulated in seedlings, AIH and NCPAH are not. Furthermore, ADC was up—regulated under various stress conditions, whereas no obvious change was observed in the other two enzymes. By contrast, ARG showed unexpectedly similar patterns with ADC. Particularly, we observed a perfect match between ARG and ADC1 expression under different light regimens, whereas no AIH or NCPAH was influenced. Here, we proposed that ARG in *Arabidopsis* likely acted as AGM, combined with ADC to enhance the polyamine production, which is necessary for plants to response to kinds of stimuli. Although ARG in tomato has been proved having higher affinity with l—arginine compared to agmatine, caution will be required when extrapolating results obtained from tomato to plants lacking ODC. Compared with the plants possessing ODC, the absence of ODC in *Arabidopsis*, Z. *marina* and S. *polyrhiza* means that ADC is possibly the sole route for putrescine synthesis. ARGAHs in *Arabidopsis*, particularly for ARGAH2 could retain the AGM specificity, which formed the additional route for putrescine synthesis. In this context, we discussed the relative contribution of the AGM pathway compared with AIH —NCPAH pathway. ARGAHs knockout lines exhibited much higher expression levels of AIH and NCPAH, as well as ADC1 and ADC2. Polyamines are still accumulated and *Arabidopsis* enhanced abiotic stress tolerance (Shi and Chan 2013). In our results, ARG was up—regulated accompanied with enhanced ADC expression under many stress conditions, whereas AIH and NCPAH expression almost showed no variation. This observation enabled us to predict such possibility for polyamine biosynthesis in *Arabidopsis*. Normally, putrescine is synthesized by the routine pathway through ADC—AIH—NCPAH. Once confront with higher demand for polyamine, as an anaplerotic pathway, ARG in *Arabidopsis* was up—regulated and could utilize agmatine as substrate to form putrescine. Our prediction is also supported by the fact that ADC and AGM are the primary pathway of polyamine synthesis from arginine in prokaryotes, such as *Synechocystis* sp. and E. *coli* (Iyer et al. 2002; Schriek et al. 2007), in which ARG, AIH and NCPAH are absent. Corresponding to this, both the ADC and ARG in plants are clustered with those of prokaryotes in phylogenetic trees, suggesting plants have acquired this pathway through lateral gene transfer. Future experiment will be needed for confirmation.

Conclusion

In summary, UC genes distributed widely from prokaryote to eukaryotes, except for green and red algae, in which ARG was not found. UC genes underwent purifying selection during evolution. Fusion and duplication of the first enzyme CPS occurred in the common ancestor of metazoan and SAR to form pyrimidine and arginine/urea specific enzyme, respectively. On the other hand, plants have only one generalized CPS and retained the primary state of heterodimeric, which is composed of separated small and large subunits. The other enzymes also have ancient origin from Archaea or eubacteria. The functions of urea cycle have undergone quite a few changes during evolution. SAR acquired multiple offshoots of UC involved in proline and polyamine synthesis. In plants, UC downstream reactions play important roles in seedling development and stress responses. Besides, ARG from plants more closely resemble AGM than currently known ARG, suggesting their potential roles in polyamine synthesis from ADC pathway.

Abbreviations

AD: arginine deiminase; ADC: arginine decarboxylase; AGM: agmatinase; AIH: agmatine iminohydrolase; ARG: arginase; ASL: argininosuccinate lyase; ASS: argininosuccinate synthetase; CPS: carbamoyl phosphate synthase; ML: maximum likelihood; NCPAH: N—cabamoylputrescine amidohydrolase; OAT: ornithine aminotransferase; OCD: ornithine cclodeainase; ODC: ornithine decarboxylase; OTC: ornithine transcarbamoylase; PPI: protein—prtein interaction; P5CR: pyrroline—5—carboxylate reductase; SAR: Stramenopiles, Alveolata, Rhzaria; UC: urea cycle.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Publicly available datasets (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi) were analyzed in this study.

Competing interests

The authors declare no competing interests.

Funding

This research was funded by National Natural Science Foundation of China, grant number 32000402, and Fund for Doctor Research of Dezhou University, grant number 2019xgrc26.

Authors' contributions

LH Teng and XL Song conceived, designed and performed the experiments. LH Teng and XM Zhao participated in the data analysis. XL Song drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors thank Haibin Lin for his assistance in data analysis.

References

1. Acharya D, Ghosh TC (2015) Global analysis of human duplicated genes reveals the relative importance of whole—genome duplicates originated in the early vertebrate evolution. *BMC GENOMICS* 17: 1—14.
2. Allen AE, Dupont CL, Oborník M, Horák A, Nunesnesi A, Mccrow JP, Zheng H, Johnson DA, Hu H, Fernie AR (2011) Evolution and metabolic significance of the urea cycle in photosynthetic diatoms. *NATURE* 473: 203—207.
3. Anderson PM (1980). Glutamine— and N—acetylglutamate—dependent carbamoyl phosphate synthetase in elasmobranchs. *SCIENCE* 208: 291—293.
4. Archibald JM (2015). Endosymbiosis and eukaryotic cell evolution. *CURR BIOL* 25: 911—921.
5. Armbrust EV, Berges JA, Bowler C, Green BR, Martinez D, Putnam NH, Zhou S, Allen AE, Apt KE, Bechner M (2004) The genome of the diatom *Thalassiosira pseudonana*, ecology, evolution, and metabolism. *SCIENCE* 306: 79—86.
6. Baliga N, Bonneau B, Facciotti MT, Pan M, Ng WV (2004) Genome sequence of *Haloarcula marismortui*, a halophilic archaeon from the Dead Sea. *GENOME RES* 14: 2221—2234.
7. Bender SJ, Parker MS, Armbrust EV (2011) Coupled effects of light and nitrogen source on the urea cycle and nitrogen metabolism over a diel cycle in the marine diatom *Thalassiosira pseudonana*. *PROTIST* 163: 232—251.
8. Blair NF, Cremer PD, Tchan MC (2015) Urea cycle disorders; a life—threatening yet treatable cause of metabolic encephalopathy in adults. *PR NEUROL* 15, 45—48.
9. Brauc S, Vooght ED, Claeys M, Geuns JMC, Höfte M, Angenon G (2012) Overexpression of arginase in *Arabidopsis thaliana* influences defence responses against Botrytis cinerea. *PLANT BIOLOGY* 14: 1, 39—45.
10. Brook L, Nunn JFF, Hippmann AA, Maldonado MT, Harvey HR, Goodlett DR, Boyd PW, Strzepek RF (2013) Diatom proteomics reveals unique acclimation strategies to mitigate Fe limitation. *PLOS ONE* 8: 75653.
11. Brownfield D., Todd CD, Deyholos MK (2008) Analysis of *Arabidopsis* arginase gene transcription patterns indicates specific biological functions for recently diverged paralogs. *PLANT MOL BIOL* 67: 429—440.
12. Burki F, Okamoto N, Pombert JF, Keeling PJ (2012) The evolutionary history of haptophytes and cryptophytes, Phylogenomic evidence for separate origins. *P ROY SOC B-BIOL SCI* 279: 2246—2254.

13. Caldwell RB, Toque HA, Narayanan SP, Caldwell RW (2015) Arginase: an old enzyme with new tricks. T*RENDS PHARMACOL SCI* 36: 395—405.
14. Cammarano P, Gribaldo S, Johann A (2002) Updating Carbamoylphosphate synthase (CPS) phylogenies, occurrence and phylogenetic identity of archaeal CPS genes. *J MOL EVOL* 55: 153—160.
15. Cao YL, Li GL, Wang KT, Zhang HY, Li LF (2011) Crystallization and preliminary X—ray analysis of argininosuccinate lyase from *Streptococcus mutans*. ACTA CRYSTALLOGR 67: 682—684.
16. Chan CX, Bhattacharya D (2013) Analysis of horizontal genetic transfer in red algae in the post—genomics age. *MOBILE GENET ELEM* 3: 27669—27669.
17. Chen H, Mccaig BC, Melotto M, He SY, Howe GA (2004) Regulation of plant arginase by wounding, jasmonate, and the phytotoxin coronatine. *J BIOL CHEM* 279: 45998—46007.
18. Cima SD, Polo LM, Díezfernández C, Martínez AI, Cervera J, Fita I, Rubio V (2015) Structure of human carbamoyl phosphate synthetase; deciphering the on/off switch of human ureagenesis. *SCI REP* 5: 16950.
19. Cortés—Giraldo I, Megías C, Alaiz M, Girón—Calle J, Vioque J (2016) Purification of free arginine from chickpea (*Cicer arietinum*) seeds. *FOOD CHEM* 192: 114—118.
20. Dagenais—Bellefeuille S, Morse D (2013) Putting the N in dinoflagellates. *FRONT MICROBIOL* 4: 369—369.
21. Dalton HL, Blomstedt CK, Neale AD, Gleadow R, Deboer KD, Hamill JD (2016) Effects of down—regulating ornithine decarboxylase upon putrescine—associated metabolism and growth in *Nicotiana tabacum* L. *J EXP BOT* 67: 3367—3381.
22. Deschamps P, Moreira D (2012) Reevaluating the green contribution to diatom genomes. *GENOME BIOL EVO* 4: 795—800.
23. Fichman Y, Gerdes SY, Kovács H, Szabados L, Zilberstein A, Csonka LN (2015) Evolution of proline biosynthesis, enzymology, bioinformatics, genetics, and transcriptional regulation. *BIO REV CAMBRIDGE* 90: 1065–1099.
24. Fuell C, Elliott KA, Hanfrey CC, Franceschetti M, Michael AJ (2010) Polyamine biosynthetic diversity in plants and algae. *PLANT PHYSIOL BIOCH* 48: 513—520.
25. Goel M, Mushegian AJDB (2006) Intermediary metabolism in sea urchin, The first inferences from the genome sequence. *DEV BIOL* 300: 282—292.
26. Goldraij A, Polacco JC (1999) Arginase is inoperative in developing soybean embryos. *PLANT PHYSIOL* 119: 297—304.
27. Guo J, Song X, Zou LF, Zou HS, Chen GY (2015) The small and large subunits of carbamoyl—phosphate synthase exhibit diverse contributions to pathogenicity in Xanthomonas citri subsp. citri. *J INTEGR AGR* 14: 1338—1347.
28. Haines RJ, Pendleton LC, Eichler DC (2011) Argininosuccinate synthase; at the center of arginine metabolism. *INT J BIOCHEM CELL B* 2: 8—23.

29. Hanfrey C, Sommer S, Mayer MJ, Burtin D, Michael AJ (2001) *Arabidopsis* polyamine biosynthesis, absence of ornithine decarboxylase and the mechanism of arginine decarboxylase activity. *PLANT J* 27: 551—560.
30. Holden HM, Thoden JB, Raushel FM (1999) Carbamoyl phosphate synthetase; an amazing biochemical odyssey from substrate to product. *CELL MOL LIFE SCI* 56: 507—522.
31. Horák A, Allen AE, Reports MOJS (2020) Common origin of ornithine–urea cycle in opisthokonts and stramenopiles. *SCI REP* 10: 1.6687.
32. Iyer RK, Kim HK, Tsoa RW, Grody WW, Cederbaum SD (2002) Cloning and characterization of human agmatinase. *MOL GENET METAB* 75: 209—218.
33. Janowitz T, Kneifel H, Piotrowski M (2006) Identification and characterization of plant agmatine iminohydrolase, the last missing link in polyamine biosynthesis of plants — FEBS Letters. *FEBS LETT* 544: 258—261.
34. Ku C, Nelson—Sathi S, Roettger M, Sousa FL, Lockhart PJ, Bryant D, Hazkani—Covo E, Mcinerney JO, Landan G, Martin WF (2015) Endosymbiotic origin and differential loss of eukaryotic genes. *NATURE* 524: 427—432.
35. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *MOL BIOL EVOL* 33: 1870—1874.
36. Labudda M, Różańska E, Cieśla J, Sobczak M, Dzik JM (2016) Arginase activity in *Arabidopsis thaliana* infected with Heterodera schachtii. *PLANT PATHOL* 65: 1529—1538.
37. Laubinger S, Zeller G, Henz SR, Sachsenberg T, Widmer CK, Naouar N, Vuylsteke M, Schölkopf B, Rätsch G, Weigel D (2008) At—TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in *Arabidopsis* thaliana. *GENOME BIOL* 9: 1—16.
38. Lawson FS, Charlebois RL, Dillon JA (1996) Phylogenetic analysis of carbamoylphosphate synthetase genes; complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *MOL BIOL EVOL* 13: 970—977.
39. Liu JH, Wang W, Wu H, Gong X, Moriguchi T (2015) Polyamines function in stress tolerance, from synthesis to regulation. *FRONT PLANT SCI* 6: 827.
40. Ma S, Bohnert HJ 2007 Integration of *Arabidopsis* thaliana stress—related transcript profiles, promoter structures, and cell—specific expression. *GENOME BIOL* 8: 1—22.
41. Micallef BJ, Shelp BJ (1989) Arginine metabolism in developing soybean cotyledons 1. P*LANT PHYSIOL* 91: 170—174.
42. Mommsen TP, Walsh PJ (1989) Evolution of urea synthesis in vertebrates; the piscine connection. *SCIENCE*. 243: 72—75.
43. Nawata CM, Walsh PJ, Wood CM (2015) Physiological and molecular responses of the spiny dogfish shark (*Squalus acanthias*) to high environmental ammonia; scavenging for nitrogen. *J EXP BIOL* 218: 238—248.

44. Nosenko T, Bhattacharya D (2007) Horizontal gene transfer in chromalveolates. *BMC EVOL BIOL* 7: 6456—6468.
45. Palmieri L, Todd CD, Arrigoni R, Hoyos ME, Santoro A, Polacco JC, Palmieri F (2006) *Arabidopsis* mitochondria have two basic amino acid transporters with partially overlapping specificities and differential expression in seedling development. *BBA-BIOENERGETICS* 1757, 1277—1283.
46. Prihoda J, Tanaka A, de Paula WB, Allen JF, Tirichine L, Bowler C (2012) Chloroplast—mitochondria cross—talk in diatoms. *J EXP BOT* 63: 1543—1557.
47. Quintero MJ, Muropastor AM, Herrero A, Flores E (2000) Arginine catabolism in the cyanobacterium *Synechocystis* sp. strain PCC 6803 involves the urea cycle and arginase pathway. *J BACTERIO*L 182: 1008—1015.
48. Rokitta SD, Dassow PV, Rost B, John U (2014) *Emiliania huxleyi* endures N—limitation with an efficient metabolic budgeting and effective ATP synthesis. *BMC GENOMICS* 15: 1—14.
49. Schriek S, Rückert C, Staiger D., Pistorius E.K., Michel K.P. (2007). Bioinformatic evaluation of L—arginine catabolic pathways in 24 cyanobacteria and transcriptional analysis of genes encoding enzymes of L—arginine catabolism in the cyanobacterium *Synechocystis* sp. PCC 6803. *BMC GENOMICS* 81: 437.
50. Shi H, Chan Z (2013) Manipulation of arginase expression modulates abiotic stress tolerance in *Arabidopsis*, effect on arginine metabolism and ROS accumulation. *J EXP BOT* 64: 1367—1379.
51. Tisi A, Federico R, Moreno S, Lucretti S, Moschou PN, Roubelakis—Angelakis KA, Angelini R, Cona A (2011) Perturbation of polyamine catabolism can strongly affect root development and xylem differentiation. *PLANT PHYSIOL* 157: 200—215.
52. Todd CD, Cooke JEK, Mullen RT, Gifford DJ (2001) Regulation of loblolly pine (*Pinus taeda* L.) arginase in developing seedling tissue during germination and post—germinative growth. *PLANT MOL BIOL* 45: 555—565.
53. Vallon O, Spalding MH (2009) Chapter 4—amino acid metabolism. *The chlamydomonas sourcebook*, 2nd ed., Elsevier Inc,Netherlands. Volume 2, pp. 115—158.
54. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *MOL BIOL EVOL* 24: 1586—1591.
55. Zhang D, Gao F, Jakovli I, Zou H, Resources GWJME (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *MOL ECOL RESOUR* 20: 348—355.
56. Zhang X, Ye N, Liang C, Mou S, Xiao F, Xu J, Dong X, Zhuang Z (2012) De novo sequencing and analysis of the Ulva linza transcriptome to discover putative mechanisms associated with its successful colonization of coastal ecosystems. *BMC GENOMICS* 13: 1—13.

Figures

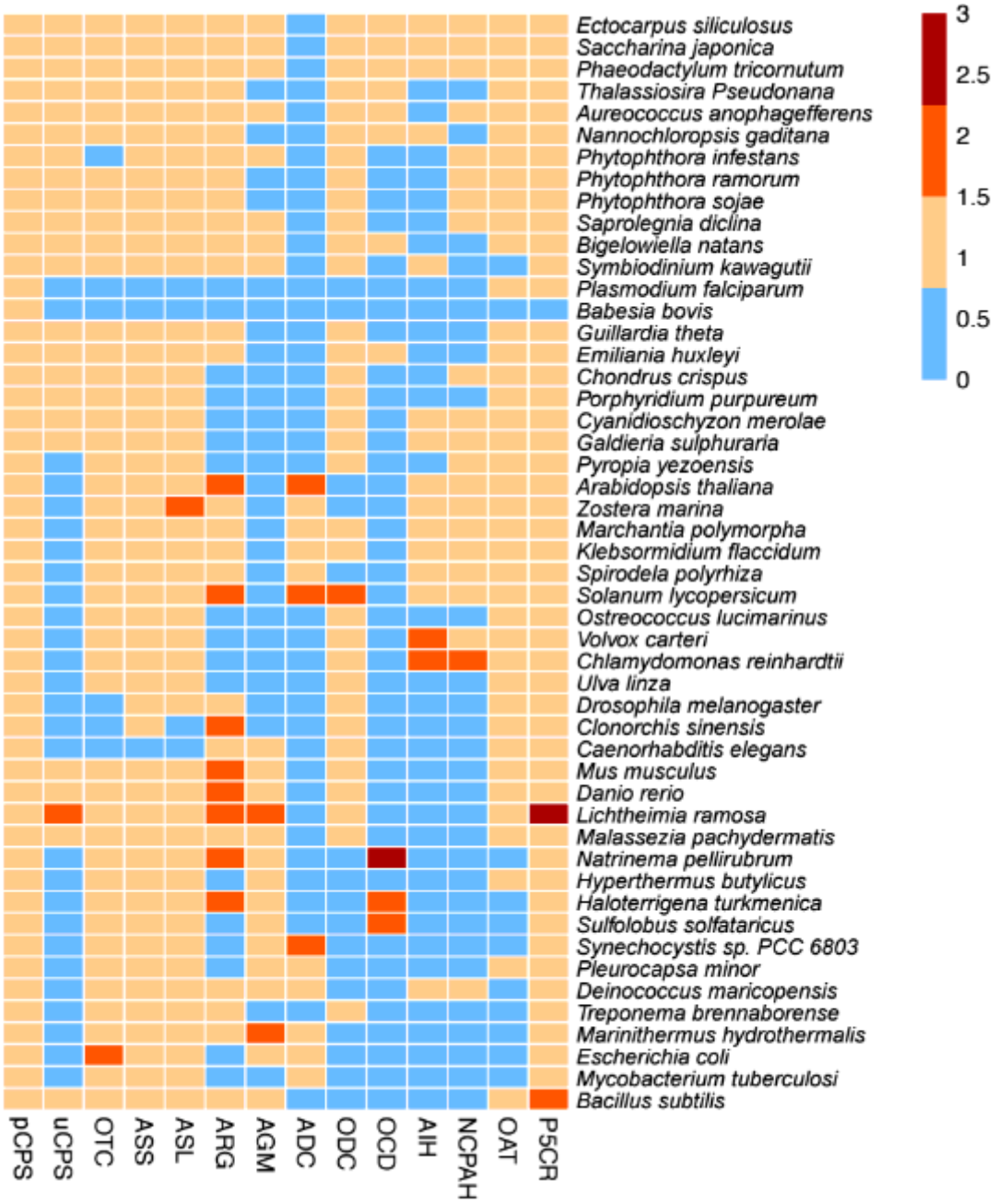

Figure 1

Taxonomic distribution of urea cycle related enzymes in 50 species. Gene number is represented by different color. For the first enzyme carbamoyl phosphate synthases (CPS), when there is only one copy, it is counted as pCPS. The number of splice isoforms is not included, i.e. each count represents all the proteins produced by a single, protein-coding gene. Abbreviations, CPS: carbamoyl phosphate synthase; pCPS: pyrimidine specific CPS; uCPS: urea specific CPS; OTC: ornithine transcarbamoylase; ASS: argininosuccinate synthetase; ASL: argininosuccinate lyase; ARG: arginase; ADC: arginine decarboxylase; AGM: agmatinase; AIH: agmatine iminohydrolase; NCPAH: N-carbamoylputrescine amidohydrolase; OCD: ornithine cyclodeaminase; ODC: ornithine decarboxylase; OAT: ornithine aminotransferase; P5CR: pyrroline-5-carboxylate reductase.

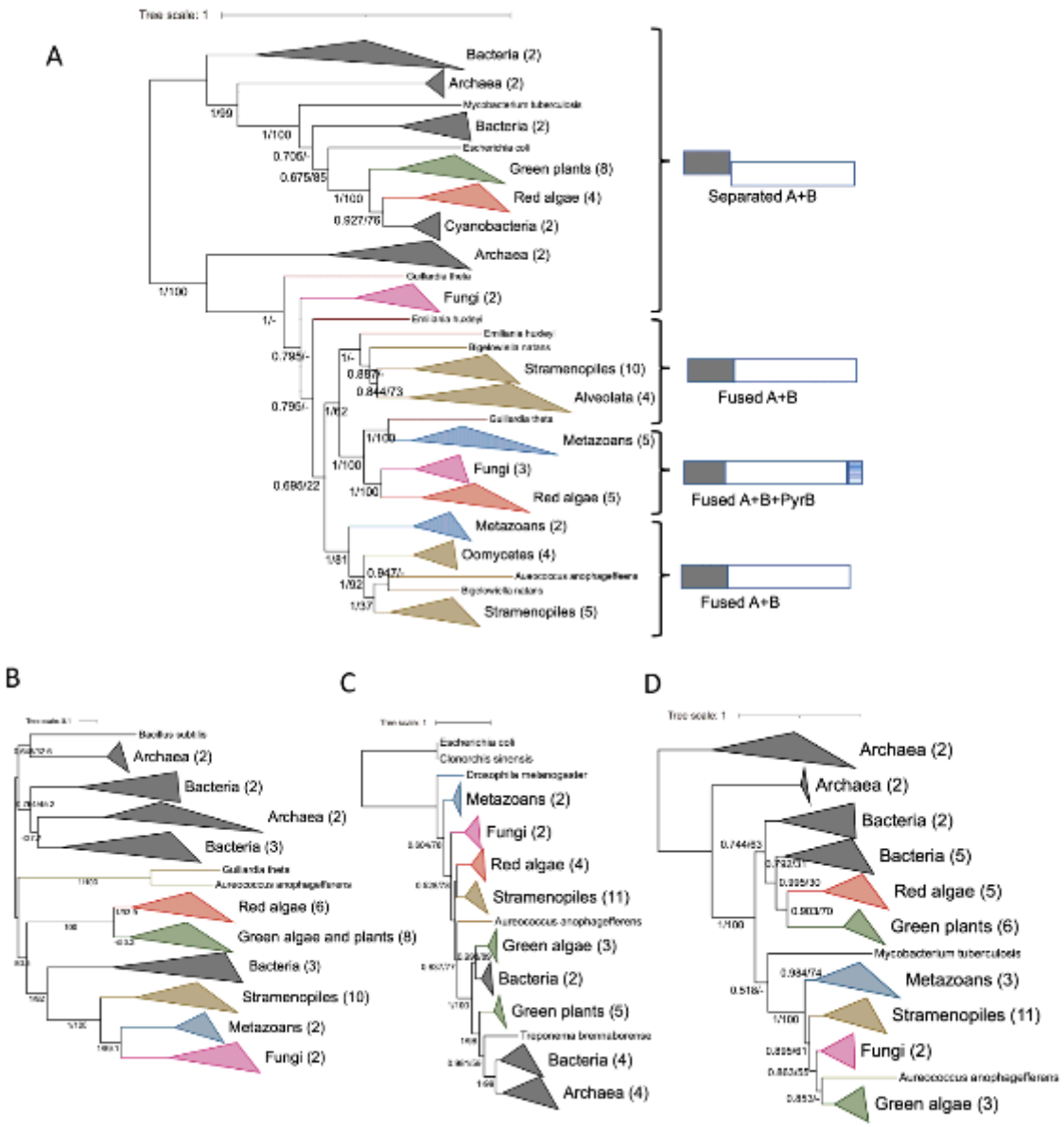

Figure 2

Schematized phylogenetic trees of the four UC enzymes. (A) carbamoyl phosphate synthase (CPS), (B) ornithine transcarbamoylase (OTC), (C) argininosuccinate synthetase (ASS) and (D) argininosuccinate lyase (ASL). The tree was constructed using MrBayes and IQtree method in Phylosuite. Taxonomically homogenous clades with absolute support are collapsed. The numbers in brackets correspond to the number of taxa in the collapsed clade. Different lineages are highlighted in different color. The number on the node was the posterior probabilities and the bootstrap support in MrBayes and IQtree, respectively. the CPS tree is based on the alignment of 896 amino acids from 71 sequences. The model was LG+R7 for IQTree, rtREV+F+I+G4 for MrBayes. The organization of the CPS genes was shown schematically on the right, with the carA and carB subunit shown as gray and white boxes, respectively. Striped boxes denote

PyrB domain which is involved in pyrimidine synthesis. The OTC tree is based on the alignment of 286 amino acids from 43 taxa. The model was WAG+I+G4 for MrBayes, LG+I+G4 for IQTree. The ASS tree is based on the alignment of 387 amino acids from 42 taxa, the model was rtREV+F+I+G4 for MrBayes, LG+I+G4 for IQTree. the ASL tree is based on the alignment of 451 amino acids from 43 sequences, the model was rtREV+F+I+G4 for MrBayes, LG+F+I+G4 for IQTree.

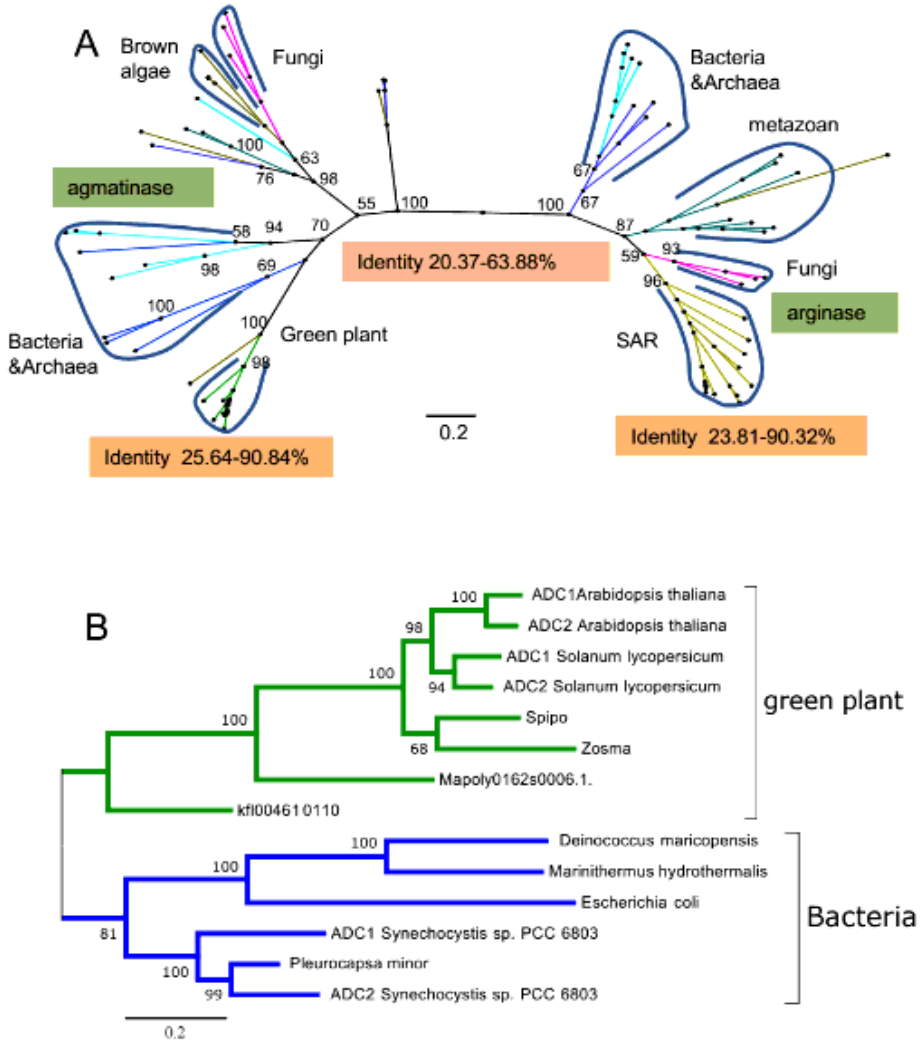

Figure 3

ML trees of arginase (ARG)/agmatinase (AGM) and arginine decarboxylase (ADC). The bootstrap values at each node are shown with 1000 replicates. (A) Phylogenetic relationship of ARG and AGM. The tree was divided into two robust groups. ARG from green plants clustered with AGM from non-plant organisms. The protein sequence identity was calculated by pairwise comparison using local BLASTp (B) Phylogenetic tree of ADC. Of 50 species, ADC was only found in green plants (no green algae) and bacteria, indicating the endosymbiotic lateral gene transfer, form the ancestor of plastid.

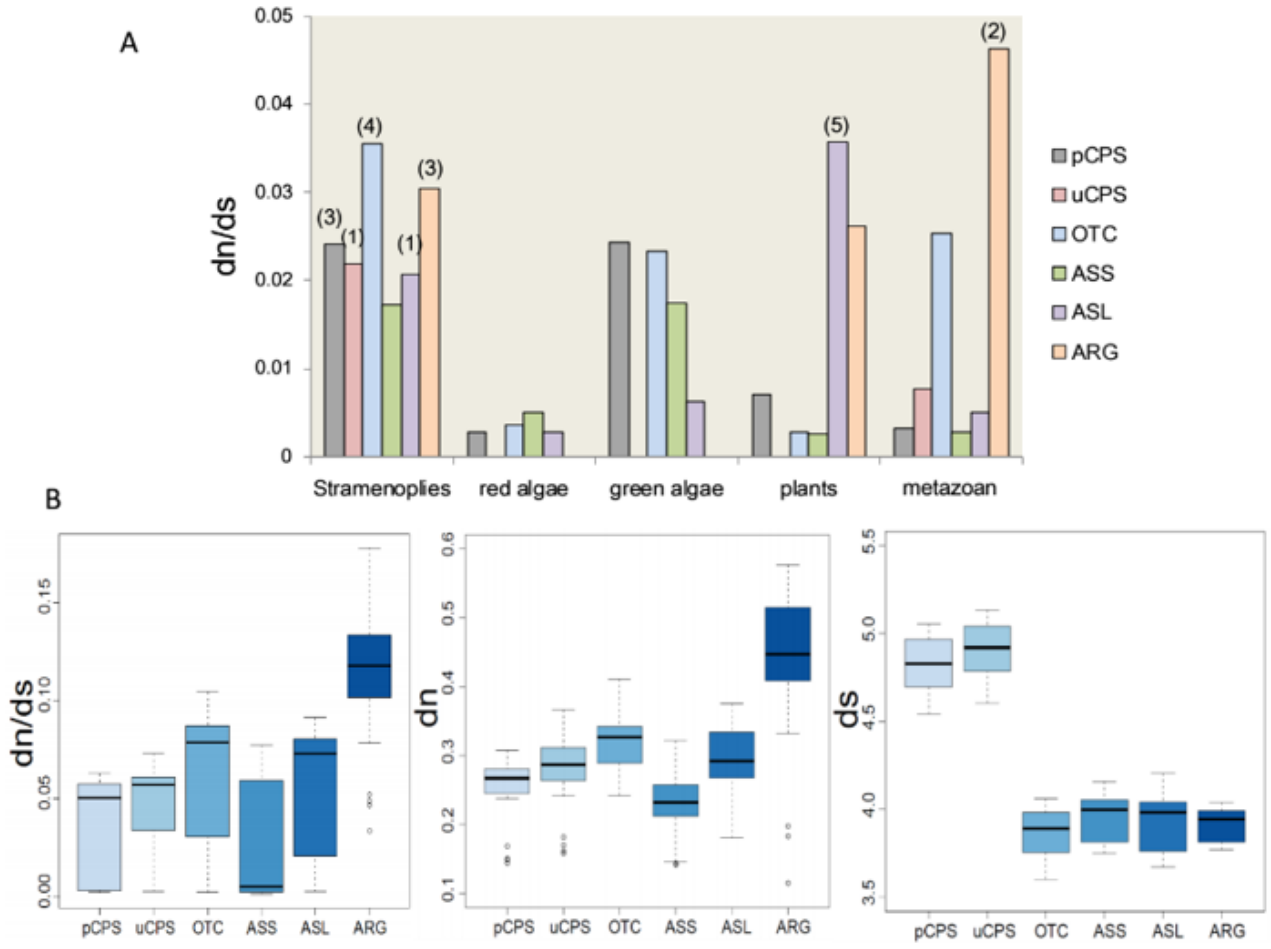

Figure 4

Selection analysis of urea cycle genes. (A) dn/ds values of the five lineages, including Stramenoplies, red algae, green algae, green plants and metazoan. The values were calculated by codeml program in PAML 4.8 package, under branch model (model=0). The number in parentheses was positive selection sites calculated by site model (M7, M8). (B) dn/ds, dn and ds values in Stramenoplies. They were calculated by yn00 program in PAML 4.8 package. '*' means the difference was significant ($p<0.05$) using student t-test.

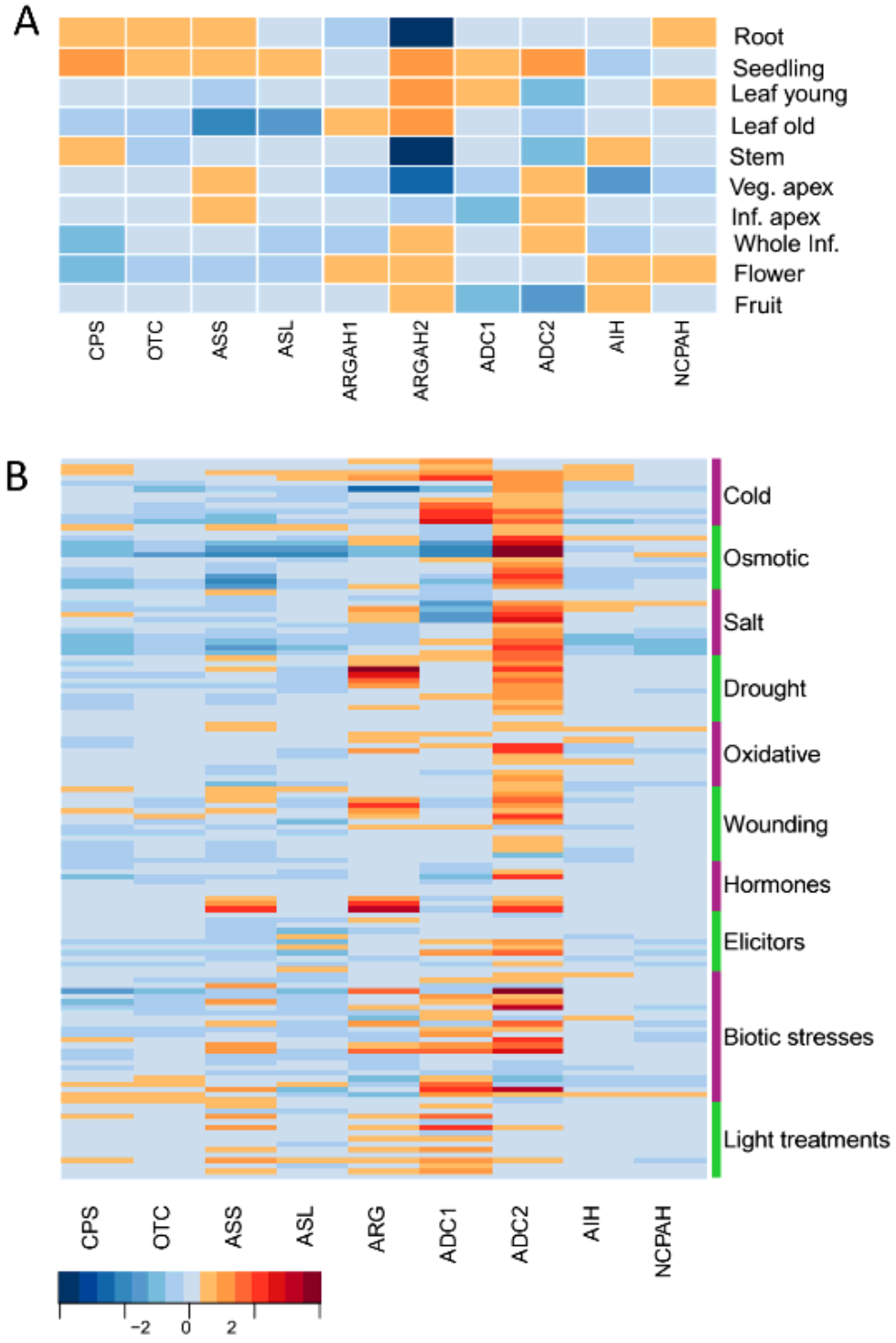

Figure 5

Heat maps of urea cycle-related genes expression in Arabidopsis. (**A**) Tissue-specific expression. Red represents high expression, blue represents low expression. (**B**) Expression under different conditions. Red indicates fold changes (treatment/control) >1, and gene was up-regulated under treatment. Blue indicates fold changes (treatment/control) <1, and gene was down-regulated under treatment. Veg.apex - vegetative shoot meristem; inf.apex - inflorescence shoot meristem; whole inf. - whole inflorescences.

Hormone treatments include abscisic acid, 1-aminocyclopropane-1-carboxylic acid and methyl jasmonate. Elicitors include CaCl2, glutathione S-transferase, hairpin z, GST-necrosis-inducing phytophtora protein 1, flagellin, and lipopolysaccharide. Bacterial stresses include Pseudomonas syringae pv. tomato DC3000, Pseudomonas syringae pv. tomato avrRPM1, Pseudomonas syringae pv. tomato DC3000 hrcC, Pseudomonas syringae pv. phaseolicola, Botrytis cinerea, Erysiphe orontii, and Phytophthora infestans. Light treatments include UV-A light, blue light, far-red light, red light, UV-A/B light, white light.

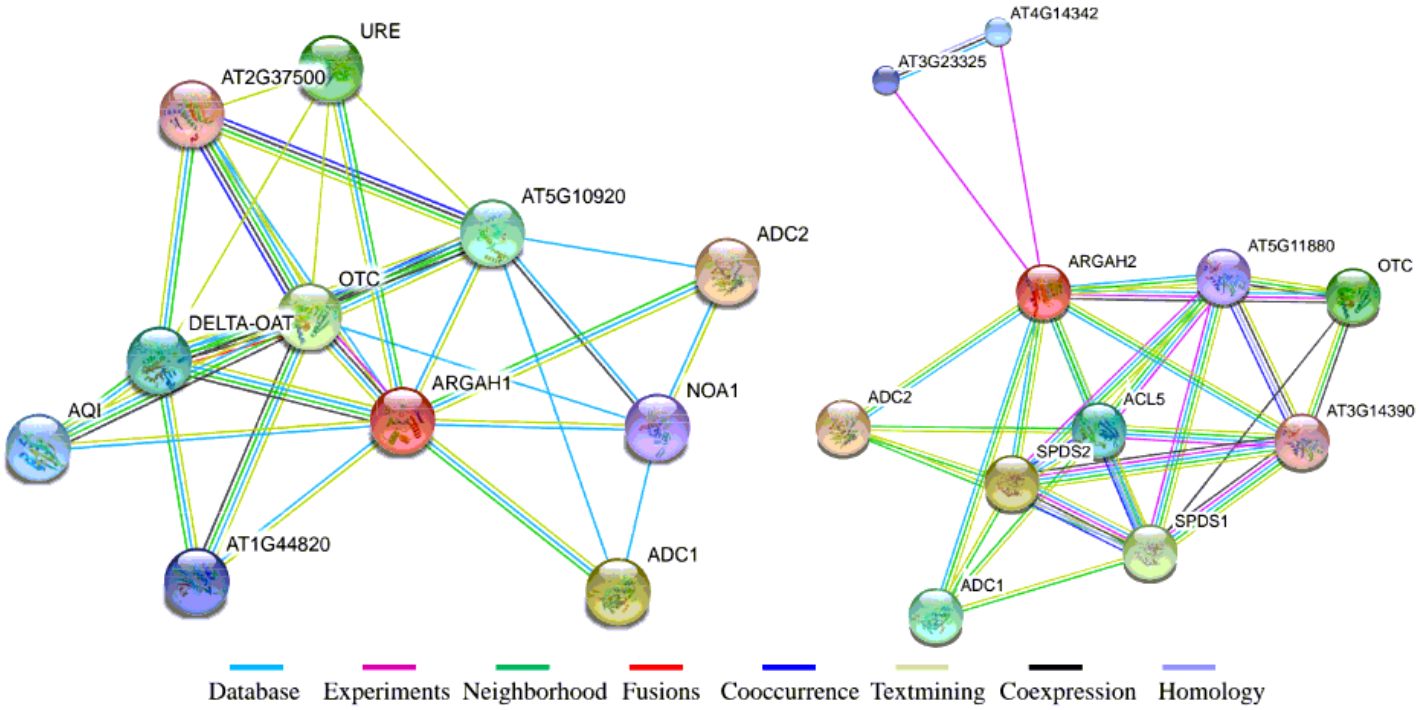

Figure 6

The protein-protein interaction network associated with ARGAH1 (left) and ARGAH2 (right) in Arabidopsis. Each node represents a protein. The red node is the target protein. The colors of lines represent the types of evidence supporting each interaction. The node with figure in it indicates 3D structure is known or predicted. **Table S**2 proviedes detailed description of each node.

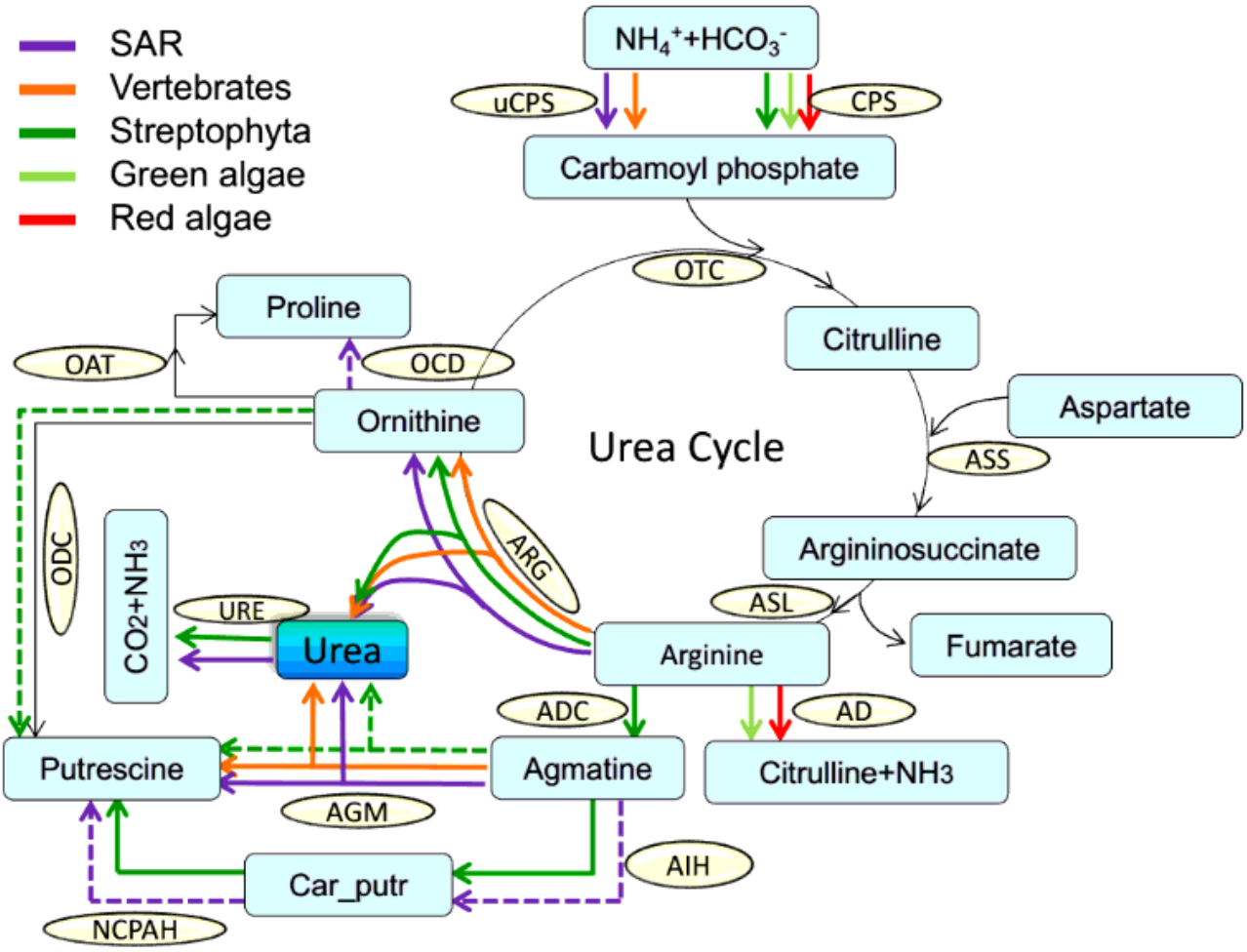

Figure 7

Schematic chart illustrating the occurrence of urea cycle and offshoots in SAR, Archaeplastida and metazoan species used in this study. The black thin arrow represents the pathway exist in all organisms included. Five colorful lines were used to represent each lineage. Dotted lines represent that pathway does not cover all the organisms examined in the lineage. In detail, green dotted lines means AGM pathway but not ODC pathway is present in Arabidopsis, Zostera marina and Spirodela polyrhiza. Purple dotted lines means OCD pathway and AIH-NCPAH pathway are only present in several members of SAR, such as brown algae and diatom. Abbreviations, CPS: carbamoyl phosphate synthase; uCPS: urea specific CPS; OTC: ornithine transcarbamoylase; ASS: argininosuccinate synthetase; ASL: argininosuccinate lyase; ARG: arginase; ADC: arginine decarboxylase; AGM: agmatinase; AIH: agmatine iminohydrolase; NCPAH: N-carbamoylputrescine amidohydrolase; OCD: ornithine cyclodeaminase; ODC: ornithine decarboxylase; OAT: ornithine aminotransferase; AD: arginine deiminase; Car_putr: N-carbamoyl putrescine.

Supplementary Files

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

- TableS1specieslist.xlsx
- TableS2PPIproteinlist.xlsx
- supplementaryfigures.docx