Isolation and Characterization of a Conserved Domain in the Eremophyte H⁺-PPase Family

Yanqin Wang¹,², Shuangxia Jin¹, Maojun Wang¹, Longfu Zhu¹*, Xianlong Zhang¹

¹ National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China, ² College of Life Science, Tarim University, Alar, Xinjiang, China

Abstract
H⁺-translocating inorganic pyrophosphatases (H⁺-PPase) were recognized as the original energy donors in the development of plants. A large number of researchers have shown that H⁺-PPase could be an early-originated protein that participated in many important biochemical and physiological processes. In this study we cloned 14 novel sequences from 7 eremophytes: Sophra alopecuroid (Sa), Glycyrrhiza uralensis (Gu), Glycyrrhiza inflata (Gi), Sueda salsa (Ss), Sueda rigida (Sr), Halostachys caspica (Hc), and Karelinia caspia (Kc). These novel sequences included 6 ORFs and 8 fragments, and they were identified as H⁺-PPases based on the typical conserved domains. Besides the identified domains, sequence alignment showed that there still were two novel conserved motifs. A phylogenetic tree was constructed, including the 14 novel H⁺-PPase amino acid sequences and the other 34 identified H⁺-PPase protein sequences representing plants, algae, protozoans and bacteria. It was shown that these 48 H⁺-PPases were classified into two groups: type I and type II H⁺-PPase. The novel 14 eremophyte H⁺-PPases were classified into the type I H⁺-PPase. The 3D structures of these H⁺-PPase proteins were predicted, which suggested that all type I H⁺-PPases from higher plants and algae were homodimers, while other type I H⁺-PPases from bacteria and protozoans and all type II H⁺-PPases were monomers. The 3D structures of these novel H⁺-PPases were homodimers except for SaVP3, which was a monomer. This regular structure could provide important evidence for the evolutionary origin and study of the relationship between the structure and function among members of the H⁺-PPase family.

Citation: Wang Y, Jin S, Wang M, Zhu L, Zhang X (2013) Isolation and Characterization of a Conserved Domain in the Eremophyte H⁺-PPase Family. PLoS ONE 8(7): e70099. doi:10.1371/journal.pone.0070099

Introduction
A basic property of life is the ability of an organism to regulate cellular pH and ion homeostasis for its normal growth and development. The concerted action of H⁺-translocating enzymes (H⁺-pumps) and cation/H⁺ exchangers are vital to establish and maintain optimal ion and pH gradients. These gradients exist between the cytoplasm and vacuole and between the cytoplasm and rhizosphere and are essential for cell function and plant development [1,2,3]. Since the 1940s, it has been proposed that the family of vacuolar membrane H⁺-PPases generate proton gradients in endomembrane compartments by using pyrophosphate (PPi) instead of ATP to act as a biological energy donor [4,5,6]. As the energy-rich phosphate product of early bacterial photophosphorylation, PPI was assumed to arise earlier than ATP in the origin and evolution of life as determined by changing the growth conditions of R. rubrum from aerobic/dark to anaerobic/light [7,8,9,10,11]. The H⁺-pump-pumping inorganic pyrophosphatase (H⁺-PPase) family uses PPI rather than ATP for energy coupling and utilization in biological membranes. It was proved that H⁺-PPase couples formation of PPI from Pi in light, hydrolyzes PPI to Pi and transfers energy in dark in the photosynthetic bacterium Rhodopseudomonas rubrum [12,13]. It was inferred that H⁺-PPase could be the early origin of the acidocalcisomes [14], which are characterized by their acidic nature, high electron density, high concentration of calcium, magnesium, and other elements in addition to pyrophosphate (PPi) and poly P [7]. The acidocalcisome may have appeared earlier than the divergence of the superkingdoms of life (Archaea, Bacteria and Eukarya) based on the analysis of function and the evolutionary dynamics of its domains [9,15]. H⁺-PPases use PPI-energy to pump H⁺ into the acidocalcisome and produce electrochemical gradients, so H⁺-PPase was regarded as the specific protein that promotes accumulation of Ca²⁺ and other ions in acidocalcisome [16]. Seufferheld et al. [9] investigated the divergence of protein domains in the H⁺-PPase molecules, and domain PF03030 was found to be shared by 31 species in Eukarya, 231 in Bacteria, and 17 in Archaea. This domain is associated with the function of H⁺-PPase, namely, to hydrolyze diphosphate to phosphate and H₂O or synthesize diphosphate using phosphate as a substrate. This suggests that the domain and the enzyme were already present in the Last Universal Common Ancestor (LUCA). So it was inferred that H⁺-PPase could be an early origin protein related to the acidocalcisome because it was demonstrated that the H⁺-PPase is ubiquitous in some algae, protozoans, bacteria and archaeabacteria [17].

Native H⁺-PPases are divided into two types (type I and type II) according to whether they are K⁺-dependent or K⁺-independent. So far the most K⁺-independent H⁺-PPases were identified in eukaryotes, bacteria and archaea, whereas K⁺-dependent H⁺-
PPases were identified only in eukaryotes [18]. However, this observation was contradicted when H'-PPase was found in the heterotrophic euglenoid A. longa [19], A. atalanta [20] and the Apicomplexan P. falcsomum [21], both of which clearly clustered with the bacterial K'-independent H'-PPases. In the plant and algae, K'-dependent and K'-independent H'-PPases showed different subcellular location; the former is located on tonoplast membrane [22,23], and the latter is located on Golgi-endoplasmic reticulum membranes [6,24]. All the available evidence is consistent with the notion that suggests that most protist H'-PPase constitute a compact cluster and that these organisms are characterized and their conserved domains were inferred, there between eremophytes and glycophytes will be beneficial for tolerance to abiotic stress. Comparing the diversity of genes and strong long duration illumination, and they have evolved a of high saline-alkali soil, extreme drought, extreme temperature inferring correct function and evolution. Evolution is directly predicted. Accurate 3D structure prediction is conducive to bioinformatics, most 3D structures of proteins have been unsuccessful, and the other is the unknown environmental effect on the gene structure of H'-PPase. With the development of bioinformatics, most 3D structures of proteins have been predicted. Accurate 3D structure prediction is conducive to inferring correct function and evolution. Evolution is directly affected by environment. Eremophytes develop under conditions of high saline-alkali soil, extreme drought, extreme temperature and strong long duration illumination, and they have evolved a tolerance to abiotic stress. Comparing the diversity of genes between eremophytes and gymophytes will be beneficial for studying the eremophyte tolerance to abiotic stresses.

In our research, we isolated and characterized the H'-PPases of eremophytes from Xinjiang, a desert region in northwestern China. Six ORFs and 8 fragments of H'-PPase from the eremophytes Sophora alopecuroides L. (Ss), Glycyrrhiza uralsensis Fisch L. (Gu), Glycyrrhiza inflata Batalin L. (Gi), Suarda salsal L. (So), Suarda rigida Kung et G. L. (Sr), Halostachys caspica L. (He), and Karelinia caspia (Pall.) L. (Ke). The first three species belong to Leguminosae, the next three are from Chenopodiaceae, and the last one belongs to Compositae. The seeds of 7 eremophytes were collected from the Alaer environment in the Tarim Basin and were germinated for three weeks. And then the leaves and roots were gathered, respectively, and were put into liquid nitrogen and stored at −80°C till for RNA isolation. Alaer is located at longitudes 80°-30' to 81° 30' east, latitude 40° 22' to 40° 57' north, and belongs to the warm temperate desert and arid climate. The annual sunshine duration is approximately 1,536 to 2,992 hrs, with annual average temperature between 8.9~11.4°C. The highest temperature is above 45°C with annual precipitation of only 42~76 mm, and the average annual evaporation (potential) reaches 1,900~2,800 mm. The relative humidity is less than 5% daytime during the summer, and the soil salt is between 1~10%, pH 7.5~9.5. The 7 species and their habitat are shown in Figure S1. Of these plants, Suarda salsal is a therophyte while the others are perennial plants.

2. RNA Isolation and H'-PPase Genes Cloning

Seeds of the eremophytes were germinated at room temperature under a light period of 16 h/8 h day/night for 3 weeks. Total RNA was extracted from leaves and roots of the 7 eremophytes, using the method described by Zhu et al. [27]. Total RNA (2 μg) was reverse-transcribed into cDNA using the SuperScript III reverse transcriptase (Invitrogen, Carlsbad, CA, USA), the cDNA products were diluted to 200 μL which were used as templates in the following PCR. To clone the H'-PPase genes from the experimental plants, degenerate primers for PCR were designed according to the sequence of the conserved domain region of known H'-PPase genes from NCBI (http://blast.ncbi.nlm.nih.gov/). The RT-PCR mixture contained: 2.5 μL, 10× PCR buffer, 0.3 μL Easy-Taq DNA polymerase (5 U/μL, Transgen, Beijing, China), 0.3 μL dNTP (10 mM), 0.3 μL appropriate paired primers (10 μM), 1 μL cDNA products, and deH2O added to 25 μL. Thermal cycle parameter was: 94°C for 5 min, followed by 35 cycles of 94°C for 40 s, 55°C for 40 s, and 72°C for 90 s. The PCR products were cloned into the pEasy-T3 Cloning vector (Transgen, Beijing, China) and sequenced. The rapid-amplification of cDNA ends (RACE) templates and adapter primers were prepared following the GeneRacer™ Kit user manual (Invitrogen). All primers are listed in Table 1.

3. Homology, Phylogeny and 3D Structural Analysis of H'-PPase Genes

The identified H'-PPase protein sequence were collected from the NCBI (http://blast.ncbi.nlm.nih.gov/), and homology analysis of the sequences was compared using BLASTx of NCBI online. Transmembrane regions of SaVP1 were predicted using TMHMM online (http://www.expasy.ch/tools/). A phylogenetic tree was constructed using the Neighbor-Joining method by CLUSTAL X1.83 and MEGA 4.1 software. Conserved domains and motifs were predicted by CLUSTAL X1.83 and Logo soft online (http://weblogo.berkeley.edu/); 3D structures were predicted using Swiss Model online (http://swissmodel.expasy.org/).
Table 1. Primers for RACE and ORF of the H⁺-PPases from the 7 eremophytes.

| species | Primer name | Primer sequence | Destination |
|---------|-------------|-----------------|-------------|
| For all species | degenerate primer-up | GGGHGGYATTTACABAARGC | H⁺-PPase gene cloning |
| For all species | degenerate primer-down | GCCDGCVAGCRATGGC | H⁺-PPase gene cloning |
| Sa | SaVP1-1 | ATGGGTCATCCATTTCCAGATCTCG | SaVP1 gene cloning |
| Sa | SaVP1-2 | TTAGATCTACAGTAAGACACCTACAA | SaVP1 gene cloning |
| Gu and Gi | GuVP1 | ATGGGAGCAGCGATTTCCAGATCTCG | GuVP1/GiVP gene cloning |
| Gu and Gi | GuVP2 | TTAGATCTGAGTAGGACCACCGT | GuVP1/GiVP gene cloning |
| Kc | KeVP1 | ATGGGTCATCCATTTCCAGATCTCG | KeVP1 gene cloning |
| Kc | KeVP2 | TTAGATCTGAGTAGGACCACCGT | KeVP1 gene cloning |

doi:10.1371/journal.pone.0070099.t001

Results

1. Isolation and Analysis of the Novel H⁺-PPase Gene Sequences
A total of 6 complete ORF sequences and 8 conserved region fragments of H⁺-PPase were cloned from 7 donor eremophytes (Figure S1) by RT-PCR and the RACE method, respectively. Among them, 2 ORFs and 1 fragment were from Sa (named as SaVP1, SaVP2 and SaVP3), 2 ORFs were from Gu (named as GuVP1 and GuVP2) and 1 ORF from Gi (named as GiVP1), while 2 fragments were from Sa (named as SaVP1 and SaVP2), 1 fragment was from Sa (named as SaVP1), 3 fragments were from Kc (named as KcVP1, KcVP2 and KcVP3). These amino acid sequences were aligned by CLUSTAL X (1.83) (Figure 1). In these six ORFs, the cDNA sequence of GuVP1 and KcVP1 contains 2304 bp and codes for 767 amino acids, the SaVP1, SaVP2 and GiVP1 genes contain 2298 bp and code for 765 amino acids, and the SaVP2 gene contains 2292 bp and codes for 763 amino acids. The isoelectric point was 5.4±0.18 and the molecular weight approximately was 80.32±0.17 kD, as analyzed by the Compute pI/Mw tool (http://web.expasy.org/compute/pi/), and 13 transmembrane regions were suggested by TMHMM (http://www.expasy.org/tmhtm/). The other 8 conserved region fragments are approximately 1500 bp and code for 500 amino acids except SaVP1 which has 2223 bp and codes for 741 amino acids. The structures of these genes were analyzed through TMHMM software from Expasy online, using SaVP1 as an example (Figure 2 and Figure 3). It was shown that the sequences from the 7 eremophytes are members of H⁺-PPase family as identified by BLASTx from NCBI online (http://blast.ncbi.nlm.nih.gov/Blast). These sequences have specific conserved domains of H⁺-PPase, namely the GGG, DVGADLVGK, and DNVGDNVGD domains [5,26]. These specific conserved domains of H⁺-PPases are located between the 4th and 5th transmembrane region, which are exactly the same as reported except for SaVP1 with a DGG domain. The second conserved domain in all species is (E/D)VYT from 423–426 sites of SaVP1, located inside the 8th transmembrane region. The third domain is the K⁺-dependent conserved sequence GNTTAAG [6], which is located near the 11th transmembrane region (Figure 4). These conserved domains are exposed to the cytoplasm (Figure 5). Based on the above characteristics, these genes from the 7 eremophytes belong to H⁺-PPase because their conserved domains are similar to the identified H⁺-PPases.

2. Homology Analysis of the H⁺-PPase between the 14 Novel H⁺-PPases and the Other H⁺-PPases from NCBI
Blast analysis showed that all of the novel H⁺-PPase genes, except for the SaVP3 that were isolated from 7 eremophytes, have more than 70% sequence identity with the homologous H⁺-PPase genes that were identified from other plant species from NCBI (Figure S2). The highest sequence identity is 95% between SaVP1, SaVP2, GuVP1, GuVP2 and GiVP1 and vacuolar-type H⁺-pyrophosphatase (XM_003609415.1) from Medicago truncatula, and 95% between KcVP1, KcVP2 and Glycine max H⁺-PPase (XP_00342656.1). There was 96% identity between 6 novel genes from the three Chenopodiaceae plants and Salicornia europaea H⁺-PPase (AEI17666.1), while the identity was below 60% between SaVP1 and H⁺-PPase from higher plants and bacteria in the conserved region. The highest sequence identity was 64% between SaVP3 and Oysterichia trifallax H⁺-PPase (EJ773348.1).

3. Comparison of Conserved Domains of H⁺-PPases
To compare the diversity of H⁺-PPases from different origins, we conducted an analysis on the homology of some specific conserved domains. Conserved domains of approximately 240 identified H⁺-PPases from NCBI and the 14 novel H⁺-PPase clones were analyzed. The conserved domains of SaVP1 are as follows: 1. ALFGVRDVGGFVYTKAADVGDVGGKVERNNIPDPRNPNAIADNVGDVGGK (the amino acid sequences underlined were conserved domains); 2. DVGGCDGVGARCRATGGC; 3. DVGADLVGKVERNIPDPRNPNAIADNVGDVGGK; 4. DNVGDNVGD; 5. DVGA. Two nonapeptide sequences are as reported except for SaVP1 with a DGG domain.

Phylogenetic Analysis of H⁺-PPase Family
domains of \( H^+ \)-PPase (Figure 4A). The first nonapeptide sequence, DVGADLVGK, followed by the amino acids VE is similar not only in sequence but also in function to the plant vacuole sequence DX7KXE and bacterian sequence EX7-8KXE. Of all the novel \( H^+ \)-PPases from the 7 eremophytes, \( GuVP2 \) does not have this motif (Figure 1). For the DNVGDNVGD domain of \( GuVP1 \), DNV was repeated between DN and VG (Figure 1). There are rich G, A, D and V 'very early' amino acids [28] in the three motifs of domain 1, especially polar amino acid residues in DNVGDNVGD at positions 1, 5 and 9 of the nonapeptides. These three charge residues would be at the same surface with one turn every 3.6 amino acid residues.

Domain 2 contains EYYT and is located at the end of the 8th transmembrane region of \( SaVP1 \) (Figure 2). The EYYT domain is conserved in all higher plants, while it is variable in every site in algae, protozoans, bacteria and archae (Figure 4B). The EYYT domain had a highly homology in all 14 novel eremophyte \( H^+ \)-PPases.

**Figure 1.** Multiple sequence alignment with \( H^+ \)-PPase sequences isolated from 7 eremophytes. Two ORFs and a fragment were from \( Sa \) (\( SaVP1 \), \( SaVP2 \) and \( SaVP3 \)). The two ORFs were from \( Gu \) (\( GuVP1 \) and \( GuVP2 \)) and one ORF from \( Gi \) (\( GiVP1 \)). The two fragments were from \( Ss \) (\( SsVP1 \) and \( SsVP2 \)), one fragment was from \( Sr \) (\( SrVP1 \)), and three fragments were from \( Hc \) (\( HcVP1 \), \( HcVP2 \) and \( HcVP3 \)). The other ORF and fragment were from \( Kc \) (\( KcVP1 \) and \( KcVP2 \)). The sequences marked with solid underline are the identified conserved domains GGG, DVGADLVGK, DNVGDNVGD, TEYYT, and GNTTAA [5,25,26]. The putative motifs GPISDNAGGIAEM and FLLGGITSLISGFLGM were marked by dotted underline.
Figure 2. The sequence and transmembrane prediction of SaVP1. There are 13 transmembrane regions in SaVP1 which are marked with red Arabic numerals. The yellow marked K is group I, and the green marked K is group II. The purple marked segment consists of identified conserved domains. The grey marked segment consists of supposed novel conserved motifs.

doi:10.1371/journal.pone.0070099.g002
In all of the $H^+\text{-PPases}$ (more than 240 included) there is a greatly conserved GNTTAA domain (domain 3), except where 8 GNSTAA domains are found (Figure 4C). GNTTAA is considered to be an important $H^+\text{-PPase}$ domain because it could be related to potassium-dependence, especially due to the first A site [6,29]. This site has two possible forms: A and K. A is a K+-dependent native and K is a K+-independent variant [6,29]. In each of the 14 novel $H^+\text{-PPases}$ from 7 eremophytes, domain 3 exists as GNTTAA (Figure 1) and is located between the 10th and the 11th transmembrane region of $\text{SaVP1}$ (Figure 2).

In addition to the 3 conserved domains mentioned above, Lee et al. [30] divided 18 lysines (K) of $H^+\text{-PPase}$, which are in the cytosolic loop from $\text{Vigna radiate}$, into three groups according to the conserved K. Group I includes K$^{250}$, K$^{261}$, K$^{341}$, K$^{604}$, K$^{605}$, and K$^{730}$; these six K positions are highly conserved in the $H^+\text{-PPases}$ of higher plants and bacteria. Group II consists of moderately conserved lysines, such as K$^{71}$, K$^{181}$, K$^{624}$, K$^{711}$, and K$^{717}$. Among these K sites, K$^{711}$ and K$^{717}$ are highly conserved in higher plants but not in prokaryotes. The lysines in group III are not conserved in the cytosolic loops. In all of the novel eremophyte $H^+\text{-PPases}$, the position of K in group I and group II is similar to those reported by Lee et al. (2011).

4. Hypothetical Convergent Evolutionary Position in the 7 Selected Eremophytes and the Other $H^+\text{-PPase}$ Genes that were Collected from NCBI

In addition to the reported conserved domains, novel conserved motifs were predicted in all the donated species. The first motif is FLLGGTTSLSQFLGM which is located in the 3rd transmembrane end from amino acids 146 to 161 of $\text{SaVP1}$ (Figure 2). The amino acids in positions 4, 8, 12, 15 and 16 of the motif especially are highly conserved in all species (Figure 4D). The underlined amino acids in the sequence GGITSLISG will be at the same surface based on the rule that $\alpha$-helix has 3.6 residues per turn. Amino acids D, G and S are all polar amino acids, and the GGITSLISG structure is similar to the nonapeptide sequence of DNVDNVDGD in the 57 amino acid sequence mentioned above. Thus, this motif could have a vital function.

The other highly conserved motif contains 13 amino acid residues, which are located between the 10th and the 11th transmembrane region from amino acid 503 to 515 of the $\text{SaVP1}$ (Figure 2). They formed GPISDNAGGIAEM, which was highly conserved in all of the species and is underlined (Figure 4E). Furthermore, in the motif GPISDNAGGIAEM, the “very early” amino acids G, D, and A account for 46%, while the polar amino acids account for 61%; this is similar to the 57 amino acid sequence in domain 1 mentioned above. They could be connected with vital structural, functional and evolutionary significance [26].

5. Phylogenetic Analysis between Novel Genes and the $H^+\text{-PPases}$ Identified from NCBI

To explore the phylogenetic relationship between the novel $H^+\text{-PPase}$ genes from the selected 7 eremophytes and the $H^+\text{-PPase}$ identified from NCBI, we constructed a phylogenetic tree using the CLUSTAL X1.83 and NJ methods. This was performed for 34 $H^+\text{-PPase}$ genes from 33 species, which included 23 type I and 9 type II $H^+\text{-PPase}$ genes. Among the 25 type I $H^+\text{-PPase}$ genes, 17 $H^+\text{-PPase}$ genes were from higher plants, including Chenopodiaceae, Leguminosae and Gramineae, and other 8 type I $H^+\text{-PPase}$ genes were from algae, protozoans and bacteria. The genetic

![Figure 3. Tentative transmembrane model of $H^+\text{-PPase}$ from $\text{Sophora alopecuroid}$ generated by TMHMM online. There are 13 putative transmembrane regions. D1, D2 and D3 with solid rectangles are previously identified domains [26], and P1 and P2 with dotted rectangles are putative motifs that we predicted. doi:10.1371/journal.pone.0070099.g003](https://www.plosone.org/doi/10.1371/journal.pone.0070099.g003)
6. Prediction of H^+PPase 3D Structure

In addition to the molecular evolution of H^+PPases, we were interested in the structural similarity among H^+PPase proteins. To understand the structural similarities and variations within the H^+PPase family, we used homology modeling to build 3D structures of H^+PPase members using the Swiss Model. It was shown that all H^+PPases were rich in α-helices. The type I H^+PPase which were from higher plants that were homodimers included SsVP1, SsVP2, GsVP1 and GsVP2, while SaVP3, AYV2, and MgVP1 were monomers (Figure 5). To check the accuracy of the SaVP3 3D structure, SaVP1 was truncated to make it comparable to SaVP3 because SaVP3 was not a complete ORF. It was shown that the truncated SaVP1 was a homodimer as SaVP1-1 (Figure 5), which implied that the length of amino acids did not change the 3D structure of SaVP1. Though we did not obtain the full sequence of SaVP3, the accordance between SaVP1 and the truncated SaVP1 indirectly meant that the predicted 3D structure of SaVP3 monomer should be correct. To confirm the structure of H^+PPase of bacteria, the 3D structures of 10 bacterial H^+PPases were predicted. The predicted structure showed that native H^+PPases consisted of monomers in bacteria (Figure S4). Meanwhile, the predicted 3D structures of H^+PPases showed that all type I H^+PPases in higher plants consisted of homodimers [32], while both type I H^+PPases in bacteria and protozoans and all type II H^+PPases consisted of monomers.

Although there are high similarities among type I H^+PPases at the primary amino acid structure, there is a difference at the 3D structures. The difference was found at position 1, as noted by
black arrows in Figure 5. The structure noted with black arrow was shown as α-helix and random coil at the position 1 of type I H\(^+\)-PPases from Chenopodiaceae, *Oryza sativa*, *Chlamydomonas reinhardtii* and *Karelinia caspia*, but it was a β-fold in the *Sophora alopecuroides* H\(^+\)-PPase. Monomeric H\(^+\)-PPases are clearly different at position 2 (Figure 5), such as a β-corner in the type II H\(^+\)-PPase of *A. thaliana*, a random coil in *Magnesitoperillum gryphiswoldense*, or a β-fold in *Sophora alopecuroides* *SaVP3*. In summary, for the same protein, different species may have different protein structures.

**Discussion**

H\(^+\)-PPases are important enzymes during PPI-hydrolyzing and PPI-energizing H\(^+\) translocation, with highly conserved sequences. The H\(^+\)-PPase is not only present in the vacuolar membrane of plants but also in the acidocalcisome membrane; it provides H\(^+\), rich PPI and polyP, and indirectly accumulates calcium and other elements [9,15]. However, because the acidocalcisome was regarded as an ancestral organelle which possesses ancestral physiological function in both prokaryotes and eukaryotes, H\(^+\)-PPase was considered to be an ancestral gene [7]. Thus, there is little diversity at the amino acid sequence of H\(^+\)-PPase, implying that these genes play a vital role in the development of the organism and have suffered little selection pressure. In this report, we cloned 14 H\(^+\)-PPases from 7 eremophytes, in which 13 H\(^+\)-PPases shared 70–96% identity with other H\(^+\)-PPase from other representatives plants and have suffered little selection pressure. In this report, we cloned 14 H\(^+\)-PPases from 7 eremophytes, in which 13 H\(^+\)-PPases shared 70–96% identity with other H\(^+\)-PPase from higher plants, all of which are available on NCBI except *SaVP3*, the E values approach zero.

In the amino acid sequence of H\(^+\)-PPase *ALFGRVGGLYIT-KAADVGADLVGKVERNIPEDPRPNPAVPADNVGVGDIAGMGSDL-EEYT* and GNTTAA motifs were regarded as specific domains of the H\(^+\)-PPase family [25,26]. These motifs could be related to Mg-PPI binding, PPI hydrolysis and energy transfer [33]. It was hypothesized that the GGG sequence provides a spacer to allow mimicking of the swing of the lever arm of a...
myosin motor [5,34], and this sequence may change the mechanism for the physiological coupling between the light-induced pumping of protons and either the photophosphorylation of inorganic phosphate (P) to PPI or the hydrolysis of PPI to P under dark conditions [4]. This motif consisted of DGG in SaVP1 and it has other forms, such as AGG and SGG, implying that GGG was substitutable at the first site. DVGADLVGKVE and DNVDNVDNVDG were considered to participate directly in substrate (Mg-PPi) binding and/or hydrolysis [26,35] or Mg2+ or Ca2+ binding based on the 3D structure of soluble PPases and related to the origin of the gene [24]. GaVP1 does not have the DVGADLVGKVE domain (Figure 1), while DN was repeated in the DVNGDNVDG domain of GaVP1 as DNDVGDNDVDG. These could be affected by environment. The other H+PPases were highly conserved in the GGG, DVGADLVGKVE and DNVGDNVDG domains (Figure 4A). These domains could be related to the evolution origin because they are rich in G, A, D and V, which are ‘very early’ amino acids in these three domains [6,26,28]. These characteristics of the H+PPase domains appear to be of vital structural, functional and evolutionary significance [24].

EYYT is a highly conserved domain in higher plants. Similarly, this domain was highly conserved in the 14 novel clones of the H+PPases from eukaryotes. EYYT was inferred to play a role in coupling PPI hydrolysis and could be related to H+ translocation and H+PPase activity in plant vacuoles [25,36]. The GNTTAA domain is considered to be a marker of type I/II, according to the first A site [6]. This site has two possible forms, A and K. It was believed that the A form is the K+-dependent H+PPase and the K form is K+-independent H+PPase [6,29,37]. This domain is GNTTAA in all 14 novel eukaryote H+PPases. In the thermophilic bacterium Carboxydothermus hydrogenoformans the A1460/K1460 position is occupied by Ala in the K+-dependent H+PPase and by Lys in the K+-independent H+PPase, while the G463 (Ala)/T463 position is occupied by G or A in the K+-dependent H+PPases and T in the K+-independent H+PPases [6,37]. It was found that an A460K substitution in C. hydrogenoformans H+PPase is sufficient to confer K+-independence to both PPI hydrolysis and PPI-energized H+ translocation. In contrast, the A463T mutation does not affect the K+-dependence of H+PPase [37]. This suggested that the classification of H+PPase could have other evidence. It was believed that type I H+PPase was located in vacuolar membranes and that the type II H+PPase was located in Golgi membranes [25,35,38].

Phylogenetic analysis showed that sequence identity was related to a genetic relationship in the a subgroup, the H+PPases from Chenopodiaceae were clustered together and others from Leguminosae were clustered together (Figure 5). This phenomenon of type I H+PPases from diverse species clustering together indicated that they could be orthologous genes [39]. At the same time, there was convergent evolution in the b subgroup, in which 6 novel clones of H+PPases from eukaryotes, including Chenopodiaceae, Leguminosae and Compositae, were clustered together (Figure 5). This could be the result of environmental selection over the long term. Besides, there were a few genes, such as SaVP3 and AVP2, the identity is 66% between SaVP3 and SaVP1 in the coverage, and it is only 36% between AVP2 and AVP1 [20]. Their sequence identity had little relation to their genetic relationship, type I and type II H+PPases in the same species could be paralogous genes [39].

Using radiation inactivation and gel permeation HPLC, Sato et al. [32] and Chanson and Piilet [40] showed that the H+PPases of pumpkin and maize were dimmers, though previous reports showed the H+PPases in Mung Bean [41] and red beet [42] were a single polypeptide by SDS-PAGE analysis. The reason might be the protein was not native in SDS-PAGE [32]. These conclusions suggested that type I H+PPases were structurally similar and existed as dimers in higher plants. In our study, we predicted the 3D structure of H+PPase from different species and the result showed that type I H+PPases from higher plants and algae were homodimers, which was conform to experimental results in higher plants. Other type I H+PPases from bacteria and protists and all type II H+PPases existed as monomers.

Specially, SaVP3 is type I H+PPase based on the conserved domain GNTTAA. However, compared with SaVP1 and other H+PPases, the amino acid sequence identity is very low. Moreover, phylogenetic analysis showed that SaVP3 was clustered together with ChsVP of Chlamydomonas reinhardtii and PbVP of Plasmodium berghei. 3D structure analysis showed that both SaVP3 and AVP2 were present as monomers, and their amino acid sequence identity were low with SaVP1 and AVP1 in the same species, respectively. It is possible that different evolutionary histories or lateral gene transfers of H+PPase between different (micro) organisms exist [20,25]. Because of the attribution impossibility of SaVP3 into type I/II H+PPase, it was inferred that SaVP3 might be the evolutionary origin of SaVP1 in Sa, or directly was a novel type H+PPase. Our result imply that the 3D structure prediction could be a novel approach for homology protein classification.

Conclusion

In this report, 14 H+PPase genes were cloned from 7 eukaryotes and their sequences were compared with more than 240 other identified H+PPase available in the NCBI. At the same time, a phylogenetic tree was constructed with 14 novel cloned genes and 34 representations of H+PPase sequences. We inferred that H+PPases could have 2 other conserved motifs in addition to the 3 identified conserved domains. These highly conserved motifs indicates that H+PPase plays an important role during the development of the organism and is also characterized by a small amount of influence from the environment during evolution. The 3D structures of some of the H+PPases were also predicted. It was shown that type I H+PPases from higher plants are homodimers, while type I H+PPases from bacteria and protozoans and all type II H+PPases are present as monomers. This regularity motif and structure could provide important evidence on evolutionary origin of H+PPase and in the study of the relationship between its structure and function.

Supporting Information

Figure S1 The selected 7 eremophytes and their natural habitats. These plants were in Alar environs. A: Sophora alopecuroid L. (Sa), B: environment of these plants in winter, C: Glycyrrhiza uralensis Fisch L. (Gu), D: Glycyrrhiza inflata Batalin L. (Gì), E: Soreda salsa L. (Si), F: Soreda rigida Kung et G. L. (Sr), G: Halostachys caspica L. (He), and H: Karelinia caspia (Pall.) L. (Kc).

(TIF)

Figure S2 Multiple sequence alignment of the H+PPase amino acid sequences from 14 novel cloned H+PPases and 34 previously identified H+PPase that were selected from NCBI. Their name and accession number of these identified H+PPases are the same as in Figure 5.

(TIF)

Figure S3 The genetic relationship of the selected species. The species phylogenetic tree was constructed for the 34 identified sequences of H+PPase from 33 representative species, in
which 2 sequences were from Arabidopsis thaliana. The 7 eremophytes marked with red are the selected donor species. (TIF)

Figure S4  Predicted 3D structure of the H+-PPases in bacteria and protozoans. The gene name and accession number of these identified H+-PPases are the same as in Figure 5. (TIF)

References
1. Basile E, Tajima H, Liang YC, Ohlto M, Ushijima K, et al. (2011) The Arabidopsis Na+-H+ antiporters NHX1 and NHX2 control vacuolar pH and K+ homeostasis to regulate growth, flower development, and reproduction. Plant Cell 23: 3482–3497.
2. Haesevoets PM, Bressan RA, Zhu JK, Bohmert HJ (2000) Plant Cellular and Molecular Responses to High Salinity. Annu Rev Plant Physiol Plant Mol Biol 51: 463–499.
3. Yang H, Knapp J, Koirala P, Rajagopal D, Peer WA, et al. (2007) Enhanced photosynthesis in mutants and dicots over-expressing a phosphoruron-responsive type I H+-pyrophosphatase. Plant Biotechnol J 5: 735–745.
4. Balscheffsky H, Von Stedingk J, Helld H, Klingenberg M (1966) Inorganic pyrophosphate: formation in bacterial photophosphorylation. Science 153: 1120–1122.
5. Balscheffsky M, Schultz A, Balscheffsky H (1999) H+-proton-pumping inorganic pyrophosphatase: a tightly membrane-bound family. FEBS Lett 452: 121–127.
6. Serrano A, Perez-Castinea JR, Balscheffsky M, Balscheffsky H (2007) H+-PPases: yesterday, today and tomorrow. IUBMB Life 59: 76–83.
7. Docampo R, Ulrich P, Moreno SN (2010) Evolution of acidocalcisomes and their role in polyphosphate storage and osmoregulation in eukaryotic microbes. Philosophical Transactions of the Royal Society B: Biological Sci 365: 775–784.
8. Lipmann F, Fox S (1965) The origins of pre-biological systems and of their molecular matrices. New York: Academic Press, 361–382 p.
9. Seufferheld MJ, Kim KM, Whitfield J, Valero A, Caetano-Anolles G (2011) Evolution of vacuolar proton pyrophosphatase domains and volutin granules: clues into the early evolutionary origin of the acidocalcisome. Biochim Biophys Acta 1809: 50–64.
10. Lopez-Marques RL, Perez-Castinea JR, Lozada M, Serrano A (2004) Differential regulation of soluble and membrane-bound inorganic pyrophosphatases in the photosynthetic bacterium Rhodospirillum rubrum provides insights into pyrophosphate-based stress bioenergetics. J bacteriology 186: 3418–3426.
11. Balscheffsky NGHH (2011) Links Between Hydrothermal Environments, Pyrophosphate, Na+, and Early Evolution. Orig Life Evol Biosph: 41 (5): 485–495.
12. Balscheffsky M (1967) Inorganic pyrophosphatase and ATP as energy donors in free-living protozoa. Biochem Biophys Res Commun 294: 567–573.
13. Perez-Castinea JR, Alvar J, Ruiz-Perez LM, Serrano A (2002) Evidence for a wide occurrence of proton-translocating pyrophosphatase genes in parasitic and free-living protozoa. Biochem Biophys Res Commun 294: 567–573.
14. Perez-Castinea JR, Lopez-Marques RL, Lozada M, Serrano A (2001) A thermostable K+-stimulated vacuolar-type pyrophosphatase from the hyperthermophile bacterium Thermotoga maritima. FEBS Lett 496: 6–11.
15. Balscheffsky M, Nadanaciva S, Schultz A (1998) A pyrophosphate synthase gene: molecular cloning and sequencing of the cDNA encoding the inorganic pyrophosphate synthase from Rhodospirillum rubrum. Biochem Biophys Acta 1364: 301–306.
16. Drozdowicz YM, Kissing JC, Rea PA (2000) AVP2, a sequence-divergent, K+-insensitive H+-translocating inorganic pyrophosphatase from Arabidopsis. Plant Physiol 123: 353–362.

Acknowledgments
We thank Ling Zhang and Aijun Qiu (College of Plant Science, Tarim University, Xinjiang, China) for their identification and collection of 7 eremophytes and seeds. Professor Lingling Chen (Huazhong Agricultural University, China) helped us to predict the 3D structures.

Author Contributions
Conceived and designed the experiments: XLZ LIFZ. Performed the experiments: YQW. Analyzed the data: YQW SXJ. Contributed reagents/materials/analysis tools: YQW MJW. Wrote the paper: YQW.