SCIENTIFIC REPORTS

OPEN

SUBJECT AREAS:
PLANT EVOLUTION
MOLECULAR EVOLUTION
PHYLOGENETICS
PLANT STRESS RESPONSES

Received
20 May 2013

Accepted
4 October 2013

Published
13 November 2013

Correspondence and requests for materials should be addressed to C.K. (charanpreet06@gmail.com) or A.V. (anchalv@gmail.com)

* These authors contributed equally to this work.

† Current address: Jawaharlal Nehru University, New Delhi-110067 and International Centre for Genetic Engineering and Biotechnology Aruna Asaf Ali Marg 110 067 New Delhi, India.

Episodes of horizontal gene-transfer and gene-fusion led to co-existence of different metal-ion specific glyoxalase I

Charanpreet Kaur[1]*†, Anchal Vishnoi[3]*†, Thilini Udayangani Ariyadasa[1], Alok Bhattacharya[2,3], Sneh Lata Singla-Pareek[1] & Sudhir Kumar Sopory[1]

[1]International Centre for Genetic Engineering and Biotechnology Aruna Asaf Ali Marg 110 067 New Delhi, India, [2]School of Computational and Integrative Sciences, Jawaharlal Nehru University, 110067 New Delhi, India, [3]School of Life Sciences, Jawaharlal Nehru University, New Delhi, India.

Glyoxalase pathway plays an important role in stress adaptation and many clinical disorders. The first enzyme of this pathway, glyoxalase I (GlxI), uses methylglyoxal as a substrate and requires either Ni(II)/Co(II) or Zn(II) for activity. Here we have investigated the origin of different metal ion specificities of GlxI and subsequent pattern of inheritance during evolution. Our results suggest a primitive origin of single-domain Ni dependent GlxI [Ni-GlxI]. This subsequently evolved into Zn activated GlxI [Zn-GlxI] in deltaproteobacteria. However, origin of eukaryotic Zn-GlxI is different and can be traced to GlxI from *Candidatus pelagibacter* and *Sphingomonas*. In eukaryotes GlxI has evolved as two-domain protein but the corresponding Zn form is lost in plants/higher eukaryotes. In plants gene expansion has given rise to multiple two-domain Ni-GlxI which are differentially regulated under abiotic stress conditions. Our results suggest that different forms of GlxI have evolved to help plants adapt to stress.

Glyoxalase I (GlxI) catalyzes conversion of methylglyoxal (MG), a cytotoxic α-oxoaldehyde, to S-D-lactoylglutathione (SLG). The product thus formed, is then converted to D-lactate by the second enzyme of the pathway namely, glyoxalase II (GlxII). This pathway is ubiquitously present in all forms of life, from ancient archaea to bacteria and higher eukaryotes and is the major pathway for detoxification of MG which is generated spontaneously during the course of glycolysis. However, enzyme catalysed synthesis of MG by methylglyoxal synthase, cytochrome P450, and amine oxidase has also been observed in bacteria[1,2] but no such activity has yet been reported in plants. In humans, increased levels of MG have been linked with the development of vascular complications of diabetes, such as nephropathy, retinopathy, neuropathy and cardiovascular disease[3,4] and glyoxalase system is thought to protect against pathogenesis. In plants this pathway is thought to have an important role in conferring tolerance to multiple abiotic stresses[5–8].

The GlxI enzyme has been characterized from several sources, including plants and animals. It is a metalloenzyme, requiring either Ni(II) or Zn(II) for its catalytic activity[9]. In plants, GlxI is present in both dicots and monocots, including *Brassica juncea*[5], *Glycine max*[10], *Oryza sativa*[11], *Triticum aestivum*[12] and *Thlaspi caerulescens*[13]. Glyoxalase enzymes have also been extensively studied in microbial systems[9,14,15]. Based on the metal ion specificity, GlxI proteins have been categorized into two groups, one requires Ni(II) and/or Co(II) for activity and was earlier thought to be mainly of prokaryotic origin. The second group enzymes are dependent on Zn(II) for activation and are thought to be of eukaryotic origin. However, the discovery of newer glyoxalase genes suggests the presence of both the metal dependent forms in prokaryotes as well as in eukaryotes, thereby emphasizing the need for revision in classification. For example; *Pseudomonas aeruginosa*[16] and rice code for enzymes from both groups (our unpublished work).

The GlxI enzymes are usually dimeric in nature but monomeric forms of these enzymes have been reported in few species, such as yeast, *Plasmodium* and wheat. In these organisms the enzymes have two-GlxI domains[17–19]. Moreover, a genome wide analysis of glyoxalase genes in rice and *Arabidopsis*[20] has suggested their presence as a multi-gene family containing 11 GlxI genes.

The present study was taken up with the aim of tracing the origin of metal ion requirement of GlxI and evolutionary pathway leading to expansion of glyoxalase genes as multi-gene family in plants. Our *in silico* studies revealed that the Ni(II) dependent form of GlxI is the primitive form and subsequent evolution has led to the origin of Zn-dependent GlxI. We also investigated the origin of co-existence of both the metal-dependent forms

in an organism. Moreover, an expansion of Ni(II) dependent form was observed in plants, whereas in yeast and animals there was a loss of Ni-GlxI with few a exceptions such as *Bombus impatiens*. Our results thereby suggest multiple pathways for evolution of different forms of GlxI.

Results

Taxonomic distribution of GlxI and origin of different metal ion specificity. The distribution of GlxI in different subdivisions of the "Tree of Life" is not very well explored and has been extensively studied in only few species[16,18,21,22]. Our analysis revealed a mosaic distribution of GlxI when mapped on the species tree (data not shown). In bacteria, we could identify GlxI only in proteobacteria and cyanobacteria. However, a Ni activated GlxI has been recently characterized from *Clostridium acetobutylicum*, a gram positive bacterium[15]. In order to understand the reason for not picking up the gene from gram positive organisms, we examined the corresponding protein sequence (AAK80149) for the type of Pfam domain unique to GlxI. The *C. acetobutylicum* GlxI was found to possess Pfam domain 'PF13669.1' instead of ubiquitous domain for GlxI (PF00903.20) and exhibited only 22% identity with *E. coli* Ni-GlxI protein. Therefore it appears that this gene has undergone high level of divergence in gram positive bacteria. Moreover, GlxI was found to be missing in many eukaryotes, such as Euglenoids, Red Algae, Glaucophytes and Haptophytes. Thus, the gene tree of GlxI was patchy.

Next, the individual distribution of Zn- and Ni-GlxI in different organisms was investigated. The results showed that the Zn dependent form is absent in archaea and cyanobacteria indicating that Zn-GlxI may have originated in proteobacteria. Since deltaproteobacteria is the ancient subdivision of proteobacteria[23], their genomes were analyzed by BLAST algorithm using sequences of Ni- and Zn-GlxI domains from *E. coli* and *P. aeruginosa* (GloA3) respectively. The results revealed the presence of Ni-GlxI and Zn-GlxI in *Myxococcales* and *Bdellovibrionales,* respectively. Interestingly, both these species reside in the same ecological niche. We however, failed to identify the instances of simultaneous presence of both forms of GlxI in the above mentioned subgroups. Further, we compared the protein sequence of GlxI domains from *Myxococcus xanthus* DK 1622 (Ni binding) and *Bdellovibrio bacteriovorus* HD100 (Zn binding). An insert of two conserved modules (A and B) of 16 aa (DVPTENEANARYAFGR) and 9 aa (YHNGNTEPR) separated by a 18 aa spacer was observed in Zn binding GlxI from *B. bacteriovorus* in comparison to the Ni-GlxI (Fig. 1). These extra sequences were not unique to Zn-GlxI of *B. bacteriovorus* but were also present in many other Zn-GlxI proteins with variations in sequence and size. In some organisms B module is further split into two suggesting that these changes may have a bearing on evolution of metal ion requirement and functional adaptation (Fig. 2). We were not able to find the origin of these inserts.

In yeast and animals, reports for the presence of only Zn-GlxI exist[18,21], but our analysis showed that even though there is preponderance of Zn-GlxI in animal kingdom, there are organisms which encode both Ni-GlxI and Zn-GlxI (*Amblyomma maculatum* and *Amphimedon queenslandica*) or only Ni-GlxI (*Bombus impatiens*).

Origin of simultaneous existence of both metal ion binding forms of GlxI. *Pseudomonas aeruginosa*, a gamma proteobacteria, encodes both Ni and Zn forms of the enzyme; GloA1, GloA2 (both Ni binding), and GloA3 (Zn binding)[16]. This motivated us to track the evolutionary origin of both forms of the enzyme together in one organism. This was done by using BLAST algorithm and domain sequences of Ni- and Zn-GlxI forms respectively from *E. coli* and *Pseudomonas aeruginosa.* The sequences of genes, representing ortholog sequences from various subdivisions of three major domains of life, were extracted and cross checked to make sure that these are from respective groups. The two betaproteobacterial species, *Methylibium petroleiphilum PM1* and *Limnobacter sp. MED105* were found to encode both Ni- and Zn-GlxI genes. Both the organisms belong to the order *Burkholderiales*, suggesting that either the Zn or the Ni binding GlxI has been horizontally transferred to the ancestor of *Burkhoderiales*.

Evolution of two-domain GlxI and eventual loss of a two-domain Zn-GlxI. The GlxI originated as a single domain protein irrespective of their metal ion specificity, as both Ni- and Zn-GlxI are single domain proteins in bacterial species (domain length of 120 aa and 142 aa respectively). However during the evolution of eukaryotes from prokaryotes, gene duplication and gene fusion events might have given rise to two-domain GlxI (Fig. 3). The alignment of the representative sequences of Ni- and Zn-GlxI is depicted in Figure 2 whereas the multiple alignment of all the orthologous sequences under study is shown in Supplementary Fig. S1. Ni-GlxI is present as a two-domain protein in all eukaryotes and among the early branching eukaryotes, algae appears to be the first to encode this gene. However in contrast, only a few organisms have two-domain Zn-GlxI (domain length of 144–146). *Ectocarpus* (Brown Algae) is the only organism that was found to harbor both the single- and the two-domain forms of Zn-GlxI. Other organisms that have two-domain forms of GlxI are diatom *Phaeodactylum tricornutum* and apicomplexan *Toxoplasma gondii*. To our knowledge other eukaryotes do not encode the two-domain form of Zn-GlxI. To address as to why the two-domain Zn-GlxI has not been retained further in the eukaryotes, we analyzed the protein sequences of two-domain Zn-GlxI. The sequences of two-domain Zn-GlxI acquired more number of substitutions as compared to the single domain Zn-GlxI and these substitutions are more significant in *Ectocarpus*. In higher eukaryotes, these substitutions might be unfavorable for gene functionality and thus the two-domain Zn-GlxI has been lost in other

```
                                                                A
Myxococcus      HTMLRVGDLERSLDFYTRIIGMKLLRRHDYPDGKFTLAFVGFGPEDT-------------  51
Bdellovibrio    HTMLRIKDPKASLDFYTRVLGMKLVRKLDFAEWKFSLYFLAYVPEGTDVPTENEANARYA  81
                                          B
Myxococcus      ---HPALELTYNWGVEKYELGT---------AYGHVALGVSDIHGTCEAIRQAGGKVVRE  99
Bdellovibrio    FGREAVLELTHNWGTEEQETTPYHNGNTEPRGFGHICISVPDIQQACARFESLGVNFQKR  141

Myxococcus      PGPMKHGTTVIAFVEDPDGYKVELIQ  125
Bdellovibrio    LG--EGGMKNIAFVKDPDQYWIEVVQ  165
```

Figure 1 | Pairwise alignment of respective Ni and Zn-GlxI domains from delta proteobacterial species, *Myxococcus xanthus* DK 1622 (Mx108756890) and *Bdellovibrio bacteriovorus* HD100 (Bb42524256). The bold letters indicate loop regions (A and B) that exist only in the Zn-GlxI.

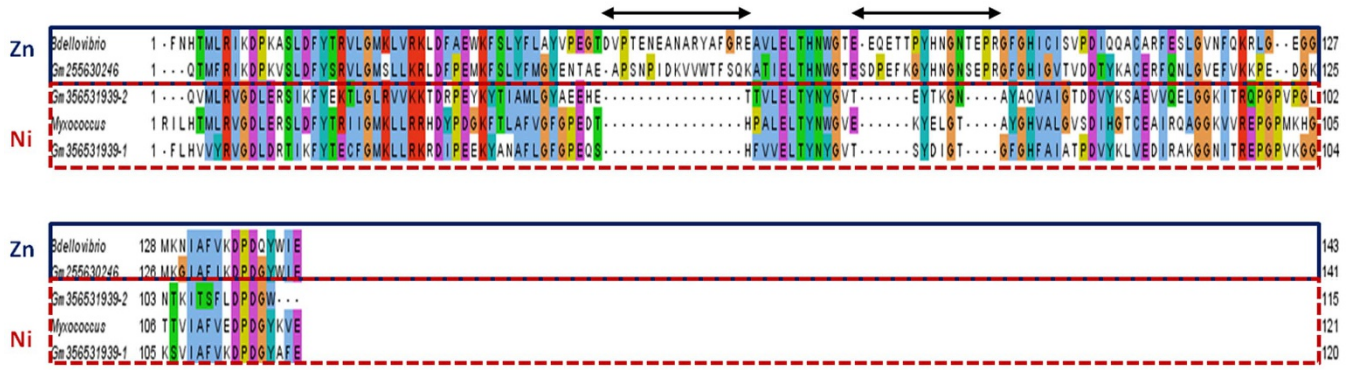

Figure 2 | **Multiple alignment of representative sequences of Ni and Zn-GlxI domains.** Zn-GlxI from *Bdellovibrio* (Bb42524256) and *Glycine max* (Gm255630246) cluster together as shown in blue box whereas Ni-GlxI from *Myxococcus* (Mx108756890) and *Glycine max* (Gm356531939) form another cluster enclosed in red box. The two domains of *Glycine max* (Gm356531939) have been taken as separate sequences. Arrows indicate the regions specific to Zn-GlxI.

eukaryotic genomes with the retention of only the single domain Zn-GlxI proteins.

Sequence motifs and putative sub-cellular localization studies trace the evolutionary mechanism. In order to decipher the mechanism of evolution of metal binding characteristics, we examined the substitution pattern in the metal binding domains of GlxI. For this, we aligned Zn- and Ni-GlxI domain sequences separately. Further, the individual alignments were reordered according to the species tree and each site was examined independently. Proteobacterial Zn-GlxI sequences were placed in a sequential order of evolution as delta, epsilon, alpha, beta and gamma proteobacteria except for *Sphingomonas* and *Candidatus pelagibacter* whose sequence motifs were found to be closer to the eukaryotic proteins than that of bacterial counterparts. In bacterial species the 6 amino acid long motif 'L/FNHTML' named as start motif (see Supplementary Fig. S2) was found to be conserved with one exception, *Sphingomonas* where third position of the motif is substituted to 'Q'. We suggest that Zn-GlxI from *Sphingomonas* was acquired by the ancestor of the brown algae before separation from plants. Many motifs present in plant and other eukaryotic Zn-GlxI domains were found to be similar to *Sphingomonas* (Supplementary Fig. S2), suggesting that the *Sphingomonas* GlxI may be closer to Brown Algae than to other bacterial proteins. This has been further supported by generating Zn-GlxI domain sequence similarity values by pairwise comparison of each of the eukaryotic Zn-GlxI protein under study with *Sphingomonas* and *C. pelagibacter* (Supplementary Table S1). The start motif of Zn-GlxI derived from *Saccharomyces cerevisiae* was however, found to be similar to *C. pelagibacter* and *M. petroleiphilum.* The pairwise comparisons indicate the origin of Zn-GlxI in yeast from *C. pelagibacter* protein (Supplementary Table S1).

In case of two-domain Zn-GlxI, pairwise alignment of both the domains with those of *Sphingomonas* and *C. pelagibacter* showed that the two domains are likely to have different evolutionary origin deduced from sequence closeness to either *Sphingomonas* GlxI domain or to *C. pelagibacter* respectively. But *Toxoplasma gondii* is an exception where both the domains have higher similarity to *C. pelagibacter* than *Sphingomonas* (Supplementary Table S1). Moreover, single domain Zn-GlxI of plant, green and brown algae are more similar to *Sphingomonas* GlxI in contrast to the corresponding sequences from human, yeast and *Amphimedon queenslandica.* The later were found to be closer to *C. pelagibacter*. The only exception to this pattern is *Amblyomma maculatum,* an arthropoda, whose GlxI shares higher degree of sequence similarity with GlxI protein from *Sphingomonas* (a typical of plant GlxI). Taken together we believe there have been two rounds of acquisition of Zn-GlxI. First, an acquisition of a single domain Zn-GlxI from *C. pelagibacter* has taken place in the ancestor of eukaryotes, and subsequently Zn-GlxI from *Sphingomonas* got transferred to the ancestor of plant lineage (see Figure 4 for schematic depiction of evolution of Zn-GlxI). Further, the single domain Zn-GlxI from *Ectocarpus* was found to have higher degree of similarity with the first domain (based

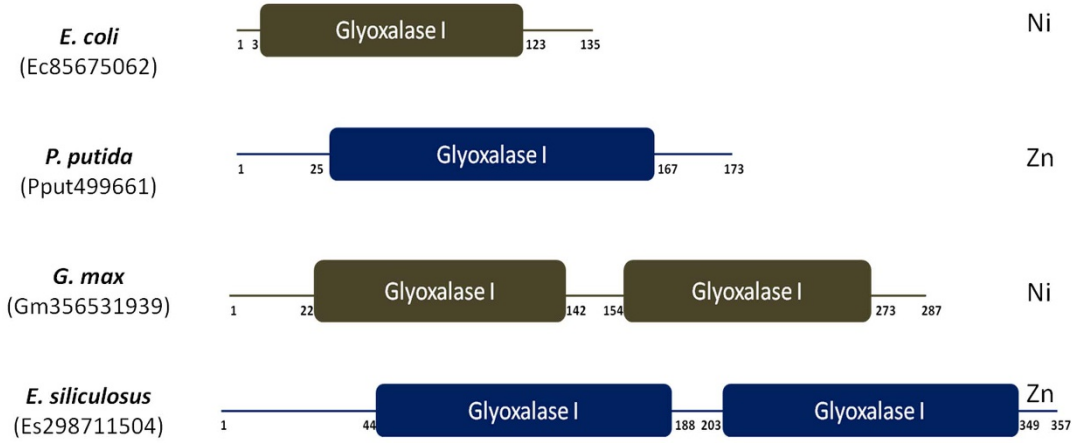

Figure 3 | **Pictorial depiction of observed domain architecture in GlxI proteins.** The Ni- and Zn-GlxI exist as both single- and two-domain proteins. The GlxI proteins from *E. coli* and *P. putida* have been used for representation of single domain GlxI and *G. max* and *E. siliculosus* for two-domain GlxI. Metal ion specificity has been indicated at right side for the respective proteins. The domain start and end positions are also mentioned.

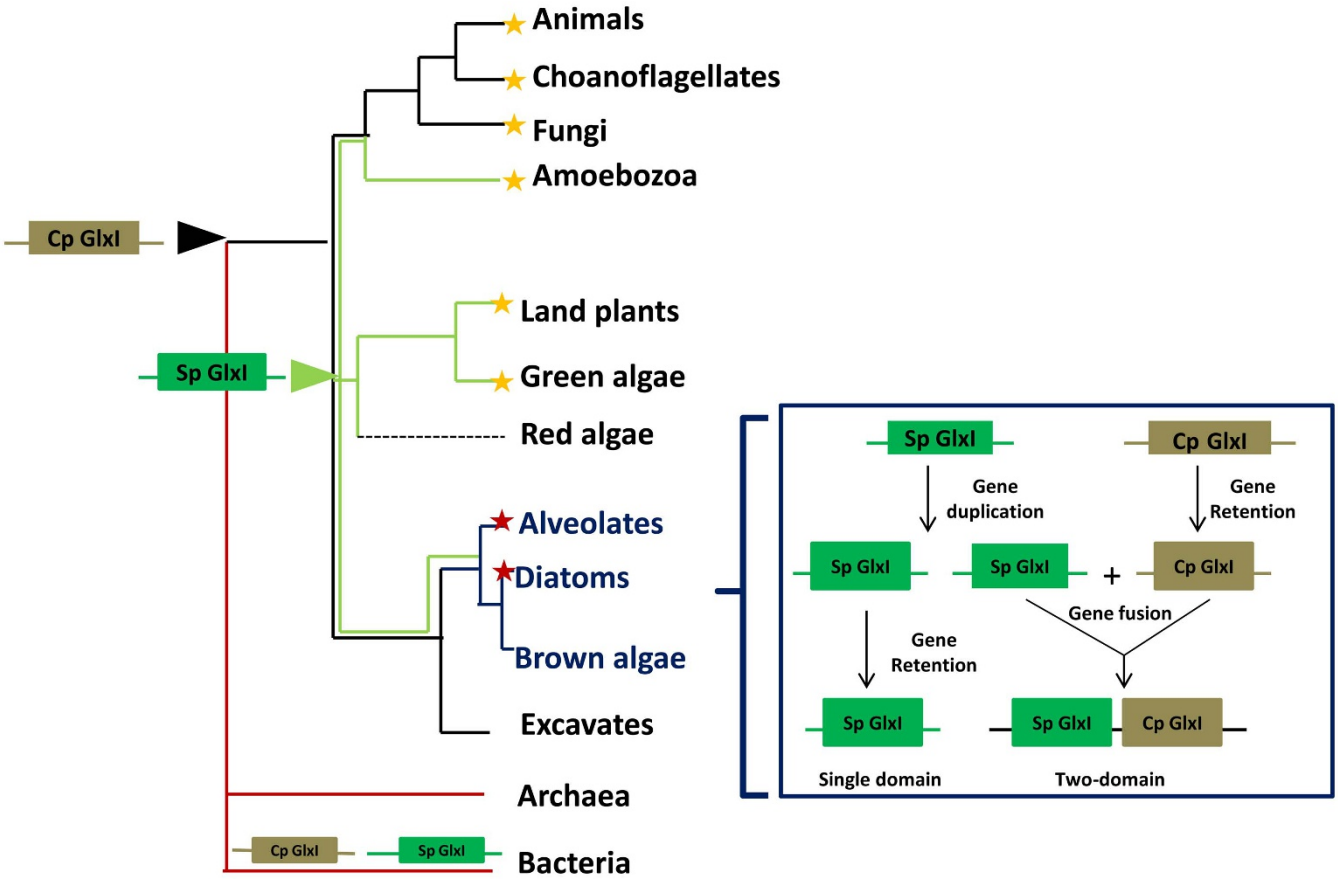

Figure 4 | Proposed mechanism for evolution of Zn-GlxI. The arrows represent point of acquisition of Zn-GlxI from *Candidatus pelagibacter* (Cp) [black] and *Sphingomonas* (Sp) [green] in ancestors of eukaryotes and plants respectively. Loss of single domain Zn-GlxI has been marked as red asterisks whereas loss of two-domain Zn-GlxI has been shown by yellow astericks. The box depicts the scheme of evolution of single and two-domain Zn-GlxI in the ancestor of Brown algae and Diatoms (in blue color). Broken arrow represents absence of GlxI in the corresponding sub-division. Modified from Ref. 50. Reprinted with permission from AAAS.

on location) of two-domain Zn-GlxI as compared to the second domain. It is likely that in the ancestor of brown algae and diatom for two-domain Zn-GlxI, a duplication of *Sphingomonas* originated Zn-GlxI might have occurred. Later one copy of the duplicate might have got fused with the pre-existing *C. peligabacter* Zn-GlxI and hence gave rise to two-domain Zn-GlxI whereas the other copy remained single domain protein (Fig. 4). *T. gondii* is an exception to this. In *Ectocarpus* genome, the single copy Zn-GlxI resides along with two-domain Zn-GlxI. But, in *P. tricornutum* and *T. gondii* the single domain Zn-GlxI has got deleted.

Bacterial Ni-GlxI can also be distinguished from the eukaryotic Ni-GlxI proteins based on the start motif 'LHTML' and 'M/F/LLHA/VVY' respectively (Supplementary Fig. S3). The exception to this rule is the animal GlxI proteins from *Amphimedon queenslandica* and *Bombus impatiens*, which are more similar to bacterial GlxI protein. Ni-GlxI from *Amblyomma maculatum* is placed closer to plant proteins, as seen for Zn-GlxI. We were able to group the proteins from green plants on the basis of first amino acid 'M/F/LLHAVY'. Among these, the proteins with 'M' at the first position of start motif seems to be primitive as this is conserved in *P. patens* (mosses), the base node in the plant phylogeny and *Ectocarpus* (Brown algae). In *P. patens* all the domains have conserved motif 'MLLHVVY' whereas in higher plants in some domains M is substituted to 'F' or 'L'. The second domain of Ni-GlxI does not show any kind of grouping based on the sequence motif. Sequence analysis shows more or less vertical descent of Ni-GlxI from prokaryotes to eukaryotes with only one exception (*Amblyomma maculatum*).

In order to correlate evolutionary information with biological properties an attempt was made to predict intracellular localization of GlxI proteins using experimental data and web based localization prediction tools (see materials and methods). The bacterial Ni- and Zn-GlxI proteins were found to be in the cytosol (Table 1) as also confirmed by *in vivo* localization studies[24,25]. However in eukaryotes, Ni and Zn dependent forms of GlxI were differentially localized. The plant Zn-GlxI and Ni-GlxI proteins were predicted to be in the nucleus and chloroplast/cytosol respectively (Table 1). This suggests different mode of acquisition of Zn and Ni dependent forms of the enzymes in plants. In animals both Zn- and Ni-GlxI proteins were predicted to be in the cytosol.

Phylogenetic analysis. The Maximum Likelihood Tree was constructed from the domain sequences of GlxI using the tool PHYML[26]. The results clearly showed two major clusters, representing Zn- and Ni-GlxI domains respectively (Fig. 5). The bacterial Ni-GlxI proteins and corresponding protein from *Candidatus Nitrosopumilus salaria*, an archaea formed a separate group. This was placed closer to the group formed by Zn-GlxI and not with the other Ni-GlxI sequences. Bacterial GlxI domains formed a monophyletic group for both Zn- and Ni-GlxI clusters. However, Zn-GlxI from *S. cerevisiae* was found to be part of the bacterial Zn-GlxI cluster with *C. Nitrosopumilus salaria* as the root of bacterial cluster. Moreover, Zn-GlxI of *Sphingomonas* was found to be part of the eukaryotic Zn-GlxI cluster close to *Amblyomma maculatum,* a tick and with three other green algae. The single domain Zn-GlxI

Table 1 | Table summarizing the details of GlxI proteins used in the study along with the sub cellular localization site prediction. Abbreviations: Cyto-cytosol; Chlo-chloroplast; Cysk-cytoskeleton; Nucl-nucleus; Mito-mitochondria; Extr-extracellular

Accession No.	Species	Gene ID	Metal ion	Protein length	First domain			Second domain			Subcellular localization
					Start	End	Length	Start	End	Length	
gi|386876271	*Candidatus Nitrosopumilus salaria*	CN386876271	Ni	135	3	123	120	-	-	-	Cyto
gi|75906364	*Anabaena variabilis*	Av75906364	Ni	145	3	123	120	-	-	-	Cyto
gi|1001475	*Synechocystis*	Synecho1001475	Ni	131	3	123	120	-	-	-	Cyto
gi|172037797	*Cyanothece*	C172037797	Ni	143	2	123	121	-	-	-	Cyto
gi|108756890	*Myxococcus xanthus*	Mx108756890	Ni	128	2	123	121	-	-	-	Cyto
gi|114327610	*Granulibacter bethesdensis*	Gb114327610	Ni	138	13	134	121	-	-	-	Cyto
gi|42524256	*Bdellovibrio bacteriovorus*	Bb42524256	Zn	169	20	163	143	-	-	-	Cyto
gi|332186478	*Sphingomonas*	Sp332186478	Zn	169	12	156	144	-	-	-	Cyto
gi|330814357	*Candidatus Pelagibacter*	Cp330814357	Zn	169	23	168	145	-	-	-	Cyto
gi|330814359	*Candidatus Pelagibacter*	Cp330814359	Zn	184	23	168	145	-	-	-	Cyto
gi|124268801	*Methylibium petroleiphilum*	Mp124268801	Zn	180	24	167	143	-	-	-	Cyto
gi|124268109	*Methylibium petroleiphilum*	Mp124268109	Ni	131	3	127	124	-	-	-	Cyto
gi|377819837	*Burkholderia*	Burkhold377819837	Ni	128	3	123	120	-	-	-	Cyto
gi|384426403	*Xanthomonas campestris*	Xc384426403	Zn	174	24	167	143	-	-	-	Cyto
gi|499661	*Pseudomonas putida*	Pput499661	Zn	173	25	167	142	-	-	-	Cyto
gi|386057529	*Pseudomonas aeruginosa*	Pa386057529	Ni	137	11	132	121	-	-	-	Cyto
gi|386060387	*Pseudomonas aeruginosa*	Pa386060387	Ni	131	2	123	121	-	-	-	Cyto
gi|386061283	*Pseudomonas aeruginosa*	Pa386061283	Zn	176	24	164	140	-	-	-	Cyto
gi|22125849	*Yersinia pestis*	Yp22125849	Ni	148	16	136	120	-	-	-	Cyto
gi|319411185	*Neisseria meningitidis*	Nm319411185	Ni	138	3	123	120	-	-	-	Cyto
gi|85675062	*Escherichia coli*	Ec85675062	Ni	135	3	123	120	-	-	-	Cyto
gi|71416475	*Trypanosoma cruzi*	Tc71416475	Ni	141	6	121	115	-	-	-	Cyto/Mito
gi|237835779	*Toxoplasma gondii*	Tg237835779	Zn	335	23	167	144	183	325	142	Nucl
gi|221485282	*Toxoplasma gondii*	Tg221485282	Zn	451	138	282	144	298	440	142	Nucl
gi|219119339	*Phaeodactylum tricornutum*	Pd219119339	Zn	310	5	148	143	165	296	131	Cyto/Nucl
gi|223994749	*Thalassiosira pseudonana*	Thp223994749	Zn	157	16	156	140	-	-	-	Cyto
gi|299116105	*Ectocarpus siliculosus*	Es299116105	Zn	195	32	173	141	-	-	-	Nucl
gi|298708814	*Ectocarpus siliculosus*	Es298708814	Ni	374	93	228	135	241	356	115	Chlo
gi|298711504	*Ectocarpus siliculosus*	Es298711504	Zn	357	44	188	144	203	349	146	Chlo/Nucl
gi|159481167	*Chlamydomonas reinhardtii*	Cr159481167	Zn	184	29	170	141	-	-	-	Cyto/Nucl
gi|307108693	*Chlorella variabilis*	Cv307108693	Ni	280	2	121	119	149	265	116	Cyto
gi|307106959	*Chlorella variabilis*	Cv307106959	Zn	186	28	169	141	-	-	-	Cyto
gi|384253231	*Coccomyxa subellipsoidea*	Cs384253231	Zn	181	30	171	141	-	-	-	Cyto
gi|384250166	*Coccomyxa subellipsoidea*	Cs384250166	Ni	265	2	121	119	132	251	119	Cyto
gi|168025026	*Physcomitrella patens*	Pp168025026	Ni	355	93	213	120	225	341	116	Chlo
gi|168032540	*Physcomitrella patens*	Pp168032540	Ni	309	57	175	118	-	-	-	Cyto
gi|168019816	*Physcomitrella patens*	Pp168019816	Ni	409	147	267	120	280	395	115	Chlo/Mito
gi|168066395	*Physcomitrella patens*	Pp168066395	Zn	178	30	163	133	-	-	-	Mito/Nucl
gi|168035068	*Physcomitrella patens*	Pp168035068	Zn	254	98	239	141	-	-	-	Chlo
gi|7619802	*Triticum aestivum*	Ta7619802	Ni	284	23	143	120	155	270	115	Cyto
gi|115475151	*Oryza sativa*	Os115475151	Ni	292	25	146	121	158	278	120	Cyto
gi|125538981	*Oryza sativa*	Os125538981	Ni	377	116	236	120	248	363	115	Chlo
gi|218196375	*Oryza sativa*	Os218196375	Ni	350	88	208	120	220	335	115	Chlo
gi|218196491	*Oryza sativa*	Os218196491	Zn	237	82	223	141	-	-	-	Nucl
gi|79317307	*Arabidopsis thaliana*	At79317307	Zn	235	80	221	141	-	-	-	Chlo/Nucl
gi|334182487	*Arabidopsis thaliana*	At334182487	Ni	322	57	177	120	189	308	119	Chlo
gi|15220397	*Arabidopsis thaliana*	At15220397	Ni	350	89	209	120	221	336	115	Chlo
Bra019830	*Brassica rapa*	Br019830	Ni	780	92	211	119	224	339	115	Nucl
Bra018654	*Brassica rapa*	Br018654	Zn	237	81	223	142	-	-	-	Chlo
Bra004214	*Brassica rapa*	Br004214	Ni	341	80	200	120	212	327	115	Chlo
Bra016811	*Brassica rapa*	Br016811	Ni	283	150	269	119	-	-	-	Cyto
gi|2113825	*Brassica juncea*	Bj2113825	Zn	185	29	171	142	-	-	-	Nucl
gi|255630246	*Glycine max*	Gm255630246	Zn	233	78	219	141	-	-	-	Nucl/Chlo
gi|356542658	*Glycine max*	Gm356542658	Zn	283	80	221	141	-	-	-	Nucl
gi|351727877	*Glycine max*	Gm351727877	Zn	185	30	171	141	-	-	-	Cysk/Nucl
gi|356541461	*Glycine max*	Gm356541461	Zn	222	67	207	140	-	-	-	Nucl
gi|356520071	*Glycine max*	Gm356520071	Ni	280	15	133	118	147	266	119	Cyto
gi|356531939	*Glycine max*	Gm356531939	Ni	287	22	142	120	154	273	119	Cyto

Table 1 | Continued

Accession No.	Species	Gene ID	Metal ion	Protein length	First domain			Second domain			Subcellular localization
					Start	End	Length	Start	End	Length	
gi|356555674	*Glycine max*	Gm356555674	Ni	287	22	142	120	154	273	119	Cyto
gi|356496416	*Glycine max*	Gm356496416	Ni	345	84	204	120	217	331	114	Chlo
gi|356529638	*Glycine max*	Gm356529638	Ni	346	85	205	120	217	332	115	Chlo
gi|356513161	*Glycine max*	Gm356513161	Ni	362	101	221	120	233	348	115	Chlo
gi|356523753	*Glycine max*	Gm356523753	Ni	356	95	215	120	227	342	115	Chlo
gi|356507913	*Glycine max*	Gm356507913	Ni	346	85	205	120	217	334	117	Chlo
gi|255637721	*Glycine max*	Gm255637721	Ni	280	15	133	118	147	266	119	Cyto
gi|224141755	*Populus trichocarpa*	Pt224141755	Ni	310	49	169	120	181	296	115	Chlo
POPTR_0018s09170	*Populus trichocarpa*	Pt0018s09170	Ni	353	92	212	120	224	339	115	Chlo
gi|224119744	*Populus trichocarpa*	Pt224119744	Ni	355	92	212	120	224	339	115	Chlo
gi|224143607	*Populus trichocarpa*	Pt224143607	Ni	282	21	141	120	153	268	115	Chlo
gi|224078584	*Populus trichocarpa*	Pt224078584	Ni	294	26	146	120	159	276	117	Cyto
gi|224104697	*Populus trichocarpa*	Pt224104697	Zn	184	30	170	140	-	-	-	Nucl
gi|854486	*Saccharomyces cerevisiae*	Sc854486	Zn	326	23	164	141	-	-	-	Cyto
gi|281209167	*Polysphondylium pallidum*	Ppa281209167	Ni	135	3	123	120	-	-	-	Cyto
gi|470392815	*Acanthamoeba castellanii*	Ac470392815	Zn	171	22	162	140	-	-	-	Nucl
gi|346469849	*Amblyomma maculatum*	Am346469849	Zn	179	25	166	141	-	-	-	Cyto/Nucl
gi|346473960	*Amblyomma maculatum*	Am346473960	Ni	238	22	143	121	155	220	65	Cyto
gi|340383181	*Amphimedon queenslandica*	Aq340383181	Ni	133	2	125	123	-	-	-	Cyto/Nucl
gi|340379949	*Amphimedon queenslandica*	Aq340379949	Zn	177	27	170	143	-	-	-	Cyto/Nucl
gi|350427648	*Bombus impatiens*	Bi350427648	Ni	267	135	256	121	-	-	-	Extr
gi|54696838	*Homo sapiens*	Hs54696838	Zn	184	33	174	141	-	-	-	Cyto

present in green algae and plants followed the species tree pattern which thereby suggests a vertical descent of the gene. The two-domain Zn-GlxI was observed only in Brown Algae, Diatoms and Apicomplexa. Each domain of the two-domain Zn-GlxI from the above mentioned phyla separated into a group suggesting the independent evolution of two-domain Zn-GlxI.

However, Ni-GlxI is a two-domain GlxI in eukaryotes unlike Zn-GlxI which is essentially a single domain enzyme. Analysis of animal Ni-GlxI proteins suggested that these have followed a different evolutionary trajectory. The two-domain Ni-GlxI proteins and the species placement pattern for the individual domains was found to be different which suggests that the evolution of the two domains is independent of each other. Moreover, each of the two domains of Ni-GlxI does not follow the species evolution pattern. The animal Ni-GlxI from *A. queenslandica* and *B. impatiens* are placed in the bacterial Ni-GlxI cluster, but here also *A. maculatum* is an exception with its placement in plant cluster.

The tree was constructed by Neighbour Joining method as well. It is similar to one constructed using Maximum Likelihood method except for positioning of some GlxI proteins, for example *Thalassiosira pseudonana* (Supplementary Fig. S4).

Structural comparison of Ni-GlxI and Zn-GlxI. The GlxI belongs to beta-alpha-beta-beta super family of proteins. Each GlxI domain consists of two beta-alpha-beta-alpha subunits, therefore single domain Zn-GlxI has two subunits whereas two-domain Ni-GlxI has four subunits (Fig. 6). The structural topology of the Ni- and Zn-GlxI domains was found to be similar except for the presence of three turn alpha helix in Zn-GlxI. This three turn alpha helix was earlier shown to block the catalytic pocket[21]. The single domain Zn-GlxI is a homodimer[27] whereas two-domain Ni-GlxI remains a monomer. The active sites are present in all the four subunits.

Evolution of role of glyoxalases in stress physiology. The role of glyoxalases in stress physiology is well established where several over-expression studies in plants have determined their ability to confer multiple stress tolerance. Their ubiquitous presence further highlights an important role for these enzymes in biological systems. However, we observed an expansion of glyoxalases as a multi-gene family in plants in contrast to their presence as single genes in most prokaryotes. Specifically, an increase in number of Ni-GlxI was observed. As an approach to understand the need for multiple glyoxalases in plants, a detailed qRT-PCR based expression study was undertaken for the four rice GlxI genes included in the phylogeny analysis to identify the response of each member of the GlxI family to the applied stress treatments (Fig. 7). Two week old IR64 rice seedlings were subjected to different stress treatments for 6 h and 24 h. Expression was analyzed in response to heat, cold, dehydration, wounding, MG, salt and oxidative stress.

OsGLYI11, encoding the cytosol localized Ni-GlxI was found to be the most stress responsive gene amongst all the members of the glyoxalase family. *OsGLYI11* possesses three spliced forms, *OsGLYI11.1*, *OsGLYI11.2* and *OsGLYI11.3* but *OsGLYI11.1* could not be amplified. However, transcript levels of the other two spliced forms varied in response to different stress conditions. *OsGLYI11.2* was induced in response to all applied stress conditions except under wounding stress. Highest up regulation was observed under drought stress at 6 h where it was found to be 4.5-fold upregulated. *OsGLYI11.3* also showed higher expression levels under multiple stress treatments. Other two GlxI genes, *OsGLYI2* and *OsGLYI7,* encoding chloroplast localized Ni-GlxI, were induced 1.7- and 1.3-fold under MG and salinity stress respectively. *OsGLYI8* which is the only Zn-GlxI encoding gene present in the rice glyoxalase family showed a ~2-fold increase in expression in response to MG and oxidative stress at 6 h.

Similarly, multiple stress inducibility of GlxI from *Arabidopsis* has also been documented by Mustafiz et al. (2011). They have reported a high stress inducibility of Ni-GlxI encoding *AtGLYI3* and *AtGLYI6* in response to salt, drought, wounding, cold and heat treatments in both roots and shoots whereas Zn-GlxI encoding *AtGLYI2* was found to be heat inducible. Moreover, microarray data analyzed with respect to different vegetative and reproductive developmental stages

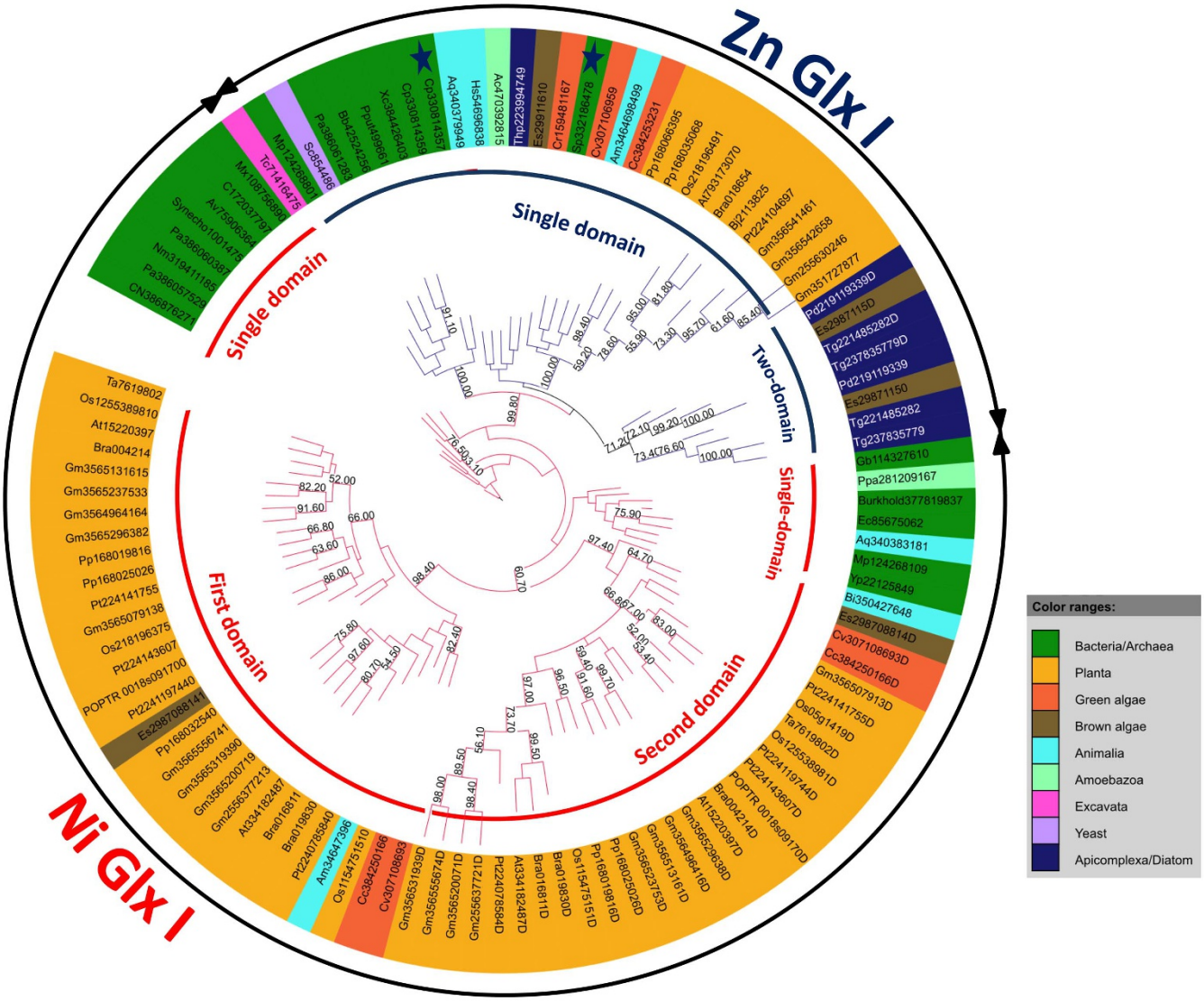

Figure 5 | **Phylogenetic analysis of GlxI family.** Reconstructed phylogenetic tree obtained from domain sequences of 81 putative GlxI sequences showed two major clusters, representing Zn- and Ni-GlxI domains. The Zn-GlxI from *C. pelagibacter* and *Sphingomonas* that have been acquired by horizontal gene transfer events in ancestors of eukaryotes and plants respectively have been marked by asterisk. Bootstrap values above 50% for 1000 bootstrap is depicted on the branches.

revealed expression of both rice and *Arabidopsis* GlxI at all stages of development. Thus, the transcript profiling study revealed an important role of GlxI, particularly Ni(II) activated forms of the enzyme, in stress physiology.

Catalytic efficiency of Ni-GlxI and Zn-GlxI characterized from various species. The kinetic properties of characterized GlxI proteins from various species were obtained to compare the catalytic efficiency of enzymes belonging to different metal activated forms (Supplementary Table S2). We found that the K_m value for Ni-GlxI were always lower than respective values for Zn-GlxI in both prokaryotes and eukaryotes except for human GlxI protein which had relatively lower K_m comparable to Ni-GlxI enzymes. Since, K_m value indicates the affinity of the enzyme for the substrate; we found that Ni-GlxI in general possessed higher affinity for their substrate than Zn-GlxI. But K_{cat} values indicated higher turnover number of Zn-GlxI indicating greater efficiency of Zn-GlxI in converting substrate to product once the enzyme-substrate complex is formed. However, inspection of overall catalytic efficiency represented by K_{cat}/K_m value signified similar efficiency for both the forms.

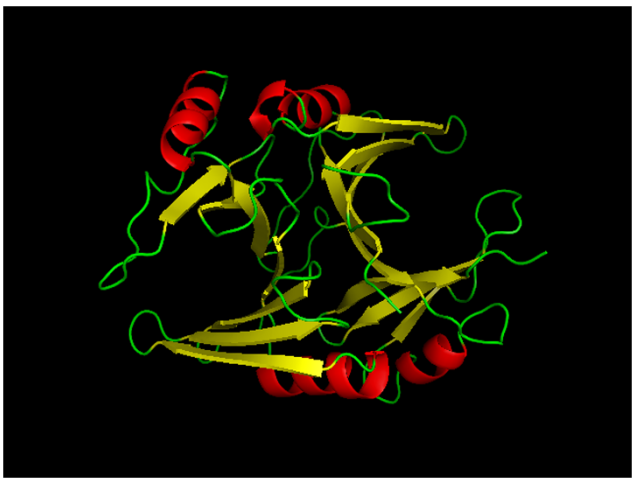

Figure 6 | **Three dimensional predicted structure of the two-domains of Ni-GlxI (Os115475151) from *O. sativa*.** The N and C termini are not shown. Each domain consists of two beta-alpha-beta-beta subunits.

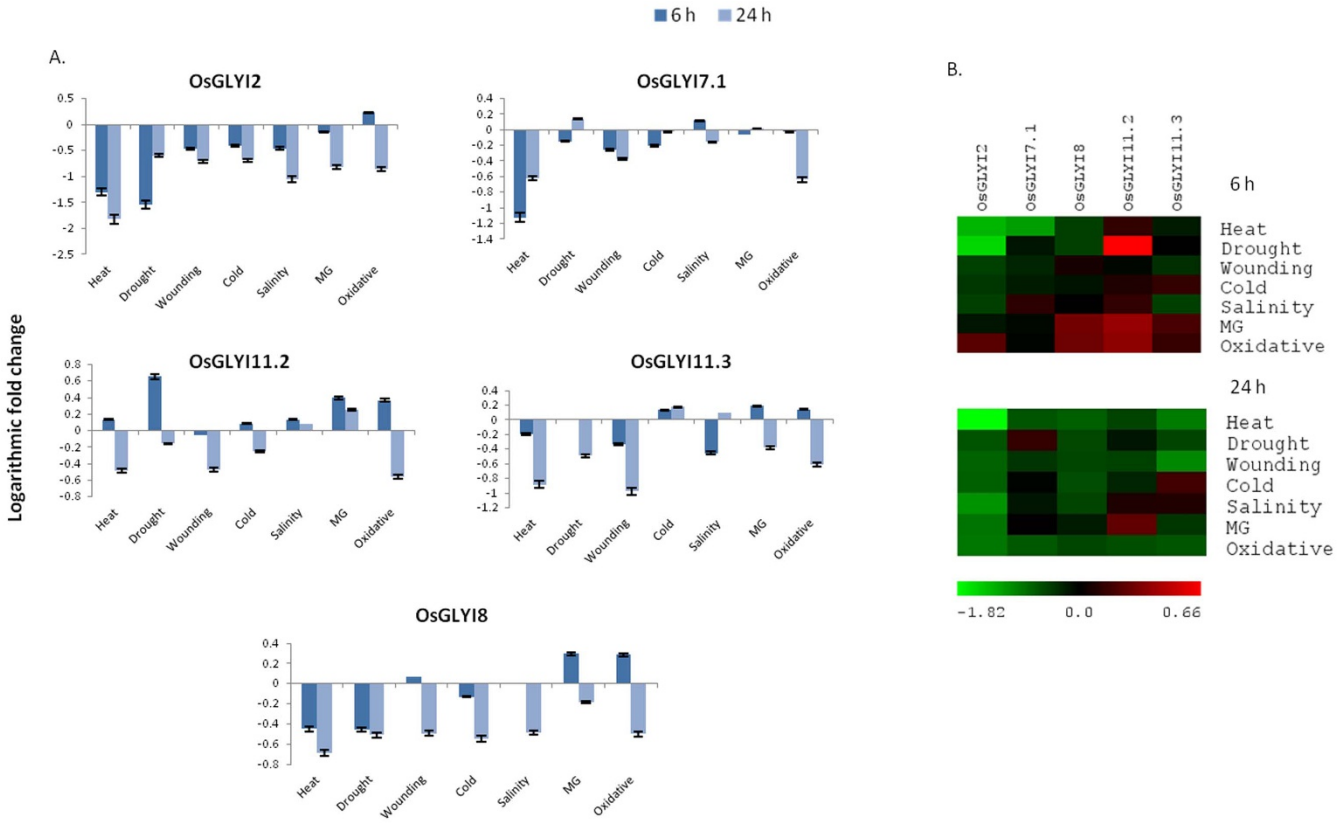

Figure 7 | Transcript profiling of rice GlxI genes in response to various abiotic stress conditions. (A) Histogram depicting logarithmic expression change of rice GlxI genes based on qRT-PCR analysis. Real time PCR analysis was done with cDNA template generated from shoot tissue of 14 day old stressed or control seedlings. As *OsGLYI7.2* and *OsGLYI11.1* could not be amplified they are not included in the real-time PCR analysis. (B) Heat map and hierarchical cluster display of expression profile for GlxI genes showing different levels of expression in response to stress at 6 h and 24 h. Colour bar at the bottom represent expression values in terms of logarithmic fold, green (lowest), black (medium) and red (highest) expression levels.

Discussion

Rules that govern genome evolution are becoming clearer with the use of comparative genomics as more and more genomes are getting sequenced. However, it has become apparent that all parts of genomes do not behave in the same way and different genes/regions follow their own evolutionary trajectory depending on the environmental and ecological history. Multi-gene families are good markers for studying genome evolution as different members of the family may follow their own path. The classical example being hemoglobin genes and evolution of fetal and adult forms[28]. Here we have studied GlxI family, a multi-gene family whose members can be distinguished from each other based on divalent metal ion dependency. We have analyzed both prokaryotes and eukaryotes with special emphasis on the expansion of the gene family in plants.

Our analysis revealed that the Ni dependent form of GlxI has evolved before the Zn form. This conclusion is based on the observation that archaea and cyanobacteria encode only Ni-GlxI. We suggest that *Bdellovibrionales* (deltaproteobacteria) are the first organisms to possess Zn-GlxI. They are known to be the obligate parasites of other gram negative bacteria[29]. However, facultative bacteria such as *Myxococcales*, another sub-division of deltaproteobacteria, encode Ni-GlxI. Absence of Ni-dependent GlxI in obligate parasites is probably due to non-existence of Ni transporters in these organisms[30]. In the absence of Ni(II) homeostasis in these organisms, Zn(II) which possesses similar characteristic properties as of Ni(II) may have led to evolution of Zn specificity for GlxI. This may have been achieved by insertion of short sequences in the enzyme[31]. One of the insert has been shown to form a three-turned α-helix that blocks one side of the catalytic pocket[21] being located next to the active site and thus plays a significant role in dictating the metal ion specificity.

Primitive orders of protoebacteria encode either Ni- or Zn-GlxI gene. However, in many organisms both forms of the enzymes are present. One of the first reported examples is *Pseudomonas aeruginosa*, a gammaproteobacteria[16]. However, our analysis suggested that both forms of GlxI evolved even before gammaproteobacteria, most likely in betaproteobacterial species, such as *M. petroleiphilum* PM1 and *Limnobacter* sp MED105. We believe that the horizontal gene transfer may have helped mobilization of one of the enzymes leading to the presence of both forms[32]. However, their evolutionary history of acquisition may differ from organism to organism. For example, *Pseudomonas* (gammaproteobacteria) may have got the gene from one of the deltaproteobacteria, *Bdellovibrio bacteriovorus* which happens to be a predator of *Pseudomonas*[33]. There are many instances where genes of endosymbiont or predator get incorporated in the host genome. For example, Drosophila is known to accommodate fragments of *Wolbachia* genome[34]. Episodes of horizontal gene transfer occurred not only in prokaryotes but also in eukaryotes during the evolution of GlxI. Our results suggest that different organisms did not follow the same evolutionary pathway. For example, in the ancestor of algae and plants there was a horizontal gene transfer event from *Sphingomonas* and this underwent duplication and fusion leading to the two-domain Zn-GlxI encoding gene. However, the gene is retained only in Brown Algae, Diatoms and Apicomplexa, and deleted thereafter in the ancestor of green algae and plants. We believe that the domain duplication and fusion events confer advantage in enzyme functionality and hence have been preserved during evolution. For example, tandem duplication of a helix-loop-helix polypeptide resulted in a more thermally stable protein[35].

Predicted sub cellular distribution of GlxI partly supported results of sequence analysis. In plants, while Ni-GlxI is expected to be in

chloroplast and cytosol, Zn-GlxI is likely to be found in nucleus. Since chloroplasts are believed to be derived from cyanobacterial ancestors, the localization of Ni-GlxI in chloroplast suggests a cyanobacterial origin of Ni-GlxI.

GlxI proteins have been characterized from various plants and their over-expression confers multiple stress tolerance in plants[5–8,12]. In order to find a possible link between their copy number and physiological function, we examined the expression levels of the entire gene family under various stresses. Multiple stress-inducible nature of glyoxalase family was observed. *OsGLYI11*, a Ni-GlxI encoding gene was found to be highly stress inducible amongst the rice glyoxalase family. Similarly, *Arabidopsis* Ni-GlxI encoding genes were found to be more responsive to various stress treatments compared to Zn-GlxI[20]. Additionally, comparison of kinetic parameters of Ni- and Zn-GlxI revealed Ni-GlxI to have lower K_m values compared to that of Zn-GlxI, indicating higher affinity of the Ni-dependent forms for the substrate MG. This probably justifies the existence of multiple Ni-GlxI in plants which due to their high stress-responsiveness and greater affinities towards MG may help in enhanced scavenging of this potent cytotoxin, thereby reducing the deleterious effects. Moreover, since MG can disrupt photosynthetic reactions in chloroplasts by acting as an intrinsic mediator catalyzing the photo reduction of O_2 at photosystem I leading to $O_2^{\cdot-}$ production[36], the presence of multiple GlxI can confer evolutionary advantage. For instance, MG/oxidative stress-induced and chloroplast localized Ni-dependent forms, namely, OsGLYI2 and OsGLYI7, can prevent damage to photosynthetic machinery caused by MG and subsequently generated oxidative stress in chloroplasts.

Taken together, the detailed inspection of evolutionary trajectory of GlxI reveals several interesting facts about these enzymes. We propose that the evolution of multiple metal dependent forms and single/two domains are the result of the adaptation of these genes under different circumstances. The fact that these enzymes are needed for removing a toxin, accumulated under stress conditions may have been a deciding factor in evolution.

Methods

Homologous gene extraction. The orthologs for *E. coli* Ni-GlxI and *P. aeruginosa* Zn-GlxI (GloA3) were initially searched using a less stringent e-value cutoff of 1E-3. One hundred and nineteen proteins were identified by the reciprocal BLAST best hit analysis from all the sub-domains of Tree of Life. Glyoxalase domain (PF00903) was identified from the PFAM database[37] in all the homologs of GlxI. Due to high variation in the N and C termini sequences of the GlxI proteins, only the domain sequences were used for multiple alignment analysis by MUSCLE v3.7[38]. Second domain of each of the two-domain GlxI proteins has been distinguished using the suffix 'D' after the respective accession IDs. We discarded the poorly aligned sequences and finally 81 protein sequences were selected for further analysis, all of which showed >40% sequence identity. In order to retain the two-domain Zn GlxI in our study, the sequence identity criterion was relaxed for the second domain of Ectocarpus GlxI protein (Es298711504) possessing 38% identity. Each orthologous sequence represents a domain in a Tree of Life. Orthologs were then used to extract the paralogous sequences in their respective genomes. The paralogs with greater than 40% sequence identity were kept. JALVIEW was used to edit and view the alignment[39]. The columns which had more than 20% gaps have been removed from the alignment.

Gene tree construction. To construct a robust phylogeny, we used both Maximum Likelihood and Neighbour Joining methods. Both the methods were used in batch mode using 1000 sets of data. The 1000 dataset were generated by "seqboot" of PHYLIP[40]. The Maximum Likelihood tree was generated using PHYML[26] with default LG evolution model and substitution rate category 4. The Neighbor Joining Method was used to construct the tree with evolution model JTT using PHYLIP. The tree has been re-rooted at *C. nitrosopumilus salaria*. The bootstrap values on all the branches above 50% are shown. The phylogenetic tree was visualized using the program iTOL v2.2[41].

Assessment of metal ion specificity of GlxI. A total of 81 GlxI proteins identified were used in the study covering all the subdomains of Tree of Life. The metal ion specificity of GlxI from several prokaryotic species, human and yeast was already known from the available literature. Due to lack of information regarding plant GlxI, the metal ion dependency of the GlxI proteins from plants was experimentally verified using rice as a model plant. A rice GlxI protein from each of the two major clusters observed in the analysis and assumingly representing the two different forms of GlxI was checked for the metal ion activation. Glyoxalase activity was measured in the presence of different divalent metal ions. Predicted Ni- and Zn-GlxI enzymes showed maximal activation on addition of respective metal ions (our unpublished work). Based on the above interpretation, metal ion specificity was assigned to the uncharacterized GlxI according to their position in the phylogenetic tree.

Putative sub cellular localization prediction. Analysis using WoLF PSORT[42] was performed in combination with the available experimental data to identify the putative intracellular localization site of GlxI. The prediction of localization site for prokaryotic GlxI proteins was performed using PSORTb v3.0.2[43].

Structure prediction and comparison of glyoxalase proteins. Modeller[44] has been used to construct the model for OsGLYI11 (Os115475151), a Ni(II) activated GlxI enzyme of *O. sativa*. 1F9z was used as a template to build the 15 model of the gene. Energy minimization has been done in Modeller and structure with minimum energy was obtained. All the models were checked with PROCHECK[45] to test whether they fall in the allowed region in the Ramachandran plot. The domain topology of Ni(II)/Co(II) and Zn(II) activated enzymes have been compared by first identifying the templates of OsGLYI11 (Os115475151) and OsGLYI8 (Os218196491) in PDB[46] and their comparison in PDBsum[47].

Plant material and stress treatment for qRT-PCR analysis. Seedlings of IR64 rice cultivar were grown under control conditions in growth chamber at 28 ± 2°C and 16 h photoperiod. The seeds were surface sterilized with 1% Bavistin for 20 min and allowed to germinate in a hydroponic system. Germinated seeds were supplied with modified Yoshida medium[48]. Fourteen-day-old seedlings were subjected to various stress treatments. For cold or heat treatment, seedlings kept in Yoshida media were transferred to a cold chamber maintained at 4 ± 1°C or an incubator at 42 ± 1°C, respectively. For dehydration stress, seedlings were removed from media and kept on a 3-mm blotting paper. For mechanical wounding, each leaf of seedlings was given a small cut with a blade. For salt, MG and oxidative stress, seedlings were kept in Yoshida medium supplied with 200 mM NaCl, 5 mM MG, or 5 mM hydrogen peroxide, respectively. Tissue samples were harvested after 6 h and 24 h of stress treatment and frozen in liquid nitrogen. Untreated seedlings were used as control.

Real-time PCR. Total RNA was isolated from the shoot tissue of control and stressed plant samples using RaFlex™ solution I and solution II (GeNei, India) as per the manufacturer's protocol. First-strand cDNA synthesis was done using RevertAid™ RNAse H minus cDNA synthesis kit as per manufacturer's instructions (Fermentas Life Sciences, USA). Same primer sequences were used for real-time PCR as described by Mustafiz and co-workers[20]. The nomenclature used for the GlxI genes has also been adopted from Mustafiz et al.[20]. The PCR mixture contained 5 µl of cDNA (50 times diluted), 12.5 µl of 2× SYBR Green PCR Master Mix (Applied Biosystems, USA) and 200 nM of each gene-specific primer in a final volume of 25 µl. The qRT-PCR was performed using ABI Prism 7500 Sequence Detection System and software (PE Applied Biosystems). All the PCRs were performed in 48-well optical reaction plates (Applied Biosystems). The specificity of the amplification was tested by dissociation curve analysis and agarose gel electrophoresis. eEF-1α gene was used as internal control. Three technical replicates were analyzed for each sample. The relative expression ratio of each glyoxalase gene was calculated using delta Ct or comparative Ct value method[49].

1. Cooper, R. A. Metabolism of methylglyoxal in microorganisms. *Annu. Rev. Microbiol.* **38**, 49–68 (1984).
2. Kalapos, M. P. Methylglyoxal in living organisms: chemistry, biochemistry, toxicology and biological implications. *Toxicol. Lett.* **110**,145–175 (1999).
3. Miller, A. G., Smith, D. G., Bhat, M. & Nagaraj, R. H. Glyoxalase I is critical for human retinal capillary pericyte survival under hyperglycemic conditions. *J. Biol. Chem.* **281**, 11864–11871 (2006).
4. Rabbani, N. & Thornalley, P. J. Glyoxalase in diabetes, obesity and related disorders. Semin. *Cell Dev. Biol.* **22**, 309–317 (2011).
5. Veena Reddy, V. S. & Sopory, S. K. Glyoxalase I from *Brassica juncea*: molecular cloning, regulation and its over-expression confer tolerance in transgenic tobacco under stress. *Plant J.* **17**, 385–395 (1999).
6. Singla-Pareek, S. L., Reddy, M. K. & Sopory, S. K. Genetic engineering of the glyoxalase pathway in tobacco leads to enhanced salinity tolerance. *Proc. Natl. Acad. Sci. USA* **100**, 14672–14677 (2003).
7. Singla-Pareek, S. L., Yadav, S. K., Pareek, A., Reddy, M. K. & Sopory, S. K. Transgenic tobacco overexpressing glyoxalase pathway enzymes grow and set viable seeds in zinc-spiked soils. *Plant Physiol.* **140**, 613–623 (2006).
8. Singla-Pareek, S. L., Yadav, S. K., Pareek, A., Reddy, M. K. & Sopory, S. K. Enhancing salt tolerance in a crop plant by overexpression of glyoxalase II. *Transgenic Res.* **17**, 171–180 (2008).
9. Sukdeo, N., Clugston, S. L., Daub, E. & Honek, J. F. Distinct classes of glyoxalase I: Metal specificity of the *Yersinia pestis*, *Pseudomonas aeruginosa* and *Neisseria meningitidis* enzymes. *Biochem. J.* **384**, 111–117 (2004).
10. Skipsey, M., Andrews, C. J., Townson, J. K., Jepson, I. & Edwards, R. Cloning and characterization of glyoxalase I from soybean. *Arch. Biochem. Biophys.* **374**, 261–268 (2000).

11. Usui, Y. *et al.* A 33-kDa allergen from rice (*Oryza sativa* L. Japonica): cDNA cloning, expression, and identification as a novel glyoxalase I. *J. Biol. Chem.* **276**, 11376–11381 (2001).
12. Lin, F., Xu, J., Shi, J., Li, H. & Li, B. Molecular cloning and characterization of a novel glyoxalase I gene TaGlyI in wheat (*Triticum aestivum* L.). *Mol. Biol. Rep.* **37**, 729–735 (2010).
13. Tuomainen, M. *et al.* Characterization of the glyoxalase 1 gene TcGLXI in the metal hyperaccumulator plant. *Thlaspi caerulescens. Planta* **233**, 1173–1184 (2011).
14. Sukdeo, N. & Honek, J. F. Microbial glyoxalase enzymes: Metalloenzymes controlling cellular levels of methylglyoxal. *Drug Metabol. Drug Interact.* **23**, 29–50 (2008).
15. Suttisansanee, U. *et al.* Structural variation in bacterial Glyoxalase I enzymes: Investigation of the metalloenzyme Glyoxalase I from *Clostridium acetobutylicum*. *J. Biol. Chem.* **286**, 38367–38374 (2011).
16. Sukdeo, N. & Honek, J. K. *Pseudomonas aeruginosa* contains multiple glyoxalase I-encoding genes from both metal activation classes. *Biochim. Biophys. Acta* **1774**, 756–763 (2007).
17. Johansen, K. S., Svendsen, I. & Rasmussen, S. K. Purification and cloning of the two domain glyoxalase I from wheat bran. *Plant Sci.* **155**, 11–20 (2000).
18. Frickel, E. M., Jemth, P., Widersten, M. & Mannervik, B. Yeast glyoxalase I is a monomeric enzyme with two active sites. *J. Biol. Chem.* **276**, 1845–1849 (2001).
19. Iozef, R., Rahlfs, S., Chang, T., Schirmer, H. & Becker, K. Glyoxalase I of the malarial parasite *Plasmodium falciparum*: evidence for subunit fusion. *FEBS Lett.* **554**, 284–288 (2003).
20. Mustafiz, A., Singh, A. K., Pareek, A., Sopory, S. K. & Singla-Pareek, S. L. Genome-wide analysis of rice and *Arabidopsis* identifies two glyoxalase genes that are highly expressed in abiotic stresses. *Funct. Integr. Genomics* **11**, 293–305 (2011).
21. Cameron, A. D., Olin, B., Ridderstrom, M., Mannervik, B. & Jones, T. A. Crystal structure of human glyoxalase I-evidence for gene duplication and 3 D domain swapping. *EMBO J.* **16**, 3386–3395 (1997).
22. Clugston, S. L. *et al.* Overproduction and characterization of a dimeric non-zinc glyoxalase I from *Escherichia coli*: evidence for optimal activation by nickel ions. *Biochemistry* **37**, 8754–8763 (1998).
23. Gupta, R. S. The phylogeny of proteobacteria: relationships to other eubacterial phyla and eukaryotes. *FEMS Microbiol Rev.* **24**, 367–402 (2000).
24. Chauhan, S. C. & Madhubala, R. Glyoxalase I gene deletion mutants of *Leishmania donovani* exhibit reduced methylglyoxal detoxification. *PLoS ONE* **4**, e6805 doi:10.1371/journal.pone 0006805 (2009).
25. Urscher, M., Przyborski, J., Imoto, M. & Deponte, M. Distinct subcellular localization in the cytosol and apicoplast, unexpected dimerization and inhibition of *Plasmodium falciparum* glyoxalases. *Mol. Microbiol.* **76**, 92–195 (2010).
26. Guindon, S., Lethiec, F., Duroux, P. & Gascuel, O. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* **33**, W557–559 (2005).
27. Kawatani, M. *et al.* The identification of an osteoclastogenesis inhibitor through the inhibition of glyoxalase I. *Proc. Natl. Acad. Sci. USA.* **105**, 11691–11696 (2008).
28. Hardison, R. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J. Exp. Biol.* **201**, 1099–1117 (1998).
29. Stolp, H. The *bdellovibrios*: bacterial parasites of bacteria. *Annu. Rev. Phytopathol.* **11**, 53–76 (1973).
30. Zhang, Y., Rodionov, D. A., Gelfand, M. S. & Gladyshev, V. N. Comparative genomic analysis of nickel, cobalt and vitamin B12 utilization. *BMC Genomics* **10**, 78 (2009).
31. Suttisansanee, U. & Honek, J. F. Bacterial glyoxalase enzymes. *Semin. Cell Dev. Biol.* **22**, 285–292 (2011).
32. Kloesges, T., Popa, O., Martin, W. & Dagan, T. Networks of gene sharing among 329 proteobacterial genomes reveal differences in lateral gene transfer frequency at different phylogenetic depths. *Mol. Biol. Evol.* **28**, 1057–1074 (2011).
33. Korchak, G. I. & Omel'yanets, T. G. Lytic action of *Bdellovibrio bacteriovorus* on bacteria of the genus *Pseudomonas*. *Mikrobiologicheskii Zhurnal.* **42**, 761–765 (1980).
34. Salzberg, S. L. *et al.* Serendipitous discovery of *Wolbachia* genomes in multiple *Drosophila* species. *Genome Biol.* **6**, R23 (2005).
35. Akanuma, S., Matsuba, T., Ueno, E., Umeda, N. & Yamagishi, A. Mimicking the evolution of a thermally stable monomeric four-helix bundle by fusion of four identical single-helix peptides. *J. Biochem.* **147**, 371–379 (2010).
36. Saito, R., Yamamoto, H., Makino, A., Sugimoto, T. & Miyake, C. Methylglyoxal functions as Hill oxidant and stimulates the photoreduction of O(2) at photosystem I: a symptom of plant diabetes. *Plant Cell Environ.* **34**, 1454–1464 (2011).
37. Punta, M. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **40**, D290–D301 (2012).
38. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
39. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview version 2: A multiple sequence alignment and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
40. Felsenstein, J. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).
41. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2006).
42. Horton, P. *et al.* WoLF PSORT: Protein Localization Predictor. *Nucleic Acids Res.* **35**, W585–587 (2007).
43. Yu, N. Y. *et al.* PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**, 1608–1615 (2010).
44. Fiser, A. & Sali, A. Modeller Generation and Refinement of Homology-Based Protein Structure Models. *Methods Enzymol.* **374**, 461–491 (2003).
45. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *PROCHECK*: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
46. Berman, H. M. *et al.* Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
47. Laskowski, R. A. PDBsum new things. *Nucleic Acids Res.* **37** (suppl 1), D355–359 (2009).
48. Yoshida, S., Forno, D. A., Cock, J. H. & Gomez, K. A. *Laboratory Manual for Physiological Studies of Rice.* International Rice Research Institute, Manila (1972).
49. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–8 (2001).
50. Baldauf, S. L. The deep roots of eukaryotes. *Science* **300**, 1703–1706 (2003).

Acknowledgements

The work is supported by internal grants of International Centre for Genetic Engineering and Biotechnology, India. CK received fellowship from Department of Biotechnology, Government of India.

Author contributions

C.K. and T.A. did the experiments. A.V. did the *in silico* analysis. C.K. and A.V. wrote the manuscript. C.K., A.V., S.S.P. and S.S. conceived and designed the experiments. A.B. and S.S. critically revised the manuscript.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Kaur, C. *et al.* Episodes of horizontal gene-transfer and gene-fusion led to co-existence of different metal-ion specific glyoxalase I. *Sci. Rep.* **3**, 3076; DOI:10.1038/srep03076 (2013).