

Research article

Open Access

## Widespread presence of "bacterial-like" PPP phosphatases in eukaryotes

Alexandra V Andreeva\*<sup>†1,2</sup> and Mikhail A Kutuzov\*<sup>†1,2</sup>

Address: <sup>1</sup>Research School of Biological and Molecular Sciences, Oxford Brookes University, Headington, Oxford OX3 0BP, UK and <sup>2</sup>Present address: University of Illinois, College of Medicine, Department of Pharmacology, 835 S. Wolcott Ave, Chicago, IL 60612, USA

Email: Alexandra V Andreeva\* - aandreeva@brookes.ac.uk; Mikhail A Kutuzov\* - m.kutuzov@usa.net

\* Corresponding authors †Equal contributors

Published: 19 November 2004

Received: 25 August 2004

BMC Evolutionary Biology 2004, 4:47 doi:10.1186/1471-2148-4-47

Accepted: 19 November 2004

This article is available from: <http://www.biomedcentral.com/1471-2148/4/47>

© 2004 Andreeva and Kutuzov; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** In eukaryotes, PPP (protein phosphatase P) family is one of the two known protein phosphatase families specific for Ser and Thr. The role of PPP phosphatases in multiple signaling pathways in eukaryotic cell has been extensively studied. Unlike eukaryotic PPP phosphatases, bacterial members of the family have broad substrate specificity or may even be Tyr-specific. Moreover, one group of bacterial PPPs are diadenosine tetrakisphosphatases, indicating that bacterial PPP phosphatases may not necessarily function as protein phosphatases.

**Results:** We describe the presence in eukaryotes of three groups of expressed genes encoding "non-conventional" phosphatases of the PPP family. These enzymes are more closely related to bacterial PPP phosphatases than to the known eukaryotic members of the family. One group, found exclusively in land plants, is most closely related to PPP phosphatases from some  $\alpha$ -Proteobacteria, including *Rhizobiales*, *Rhodobacterales* and *Rhodospirillaceae*. This group is therefore termed *Rhizobiales / Rhodobacterales / Rhodospirillaceae*-like phosphatases, or Rhilphs. Phosphatases of the other group are found in *Viridiplantae*, *Rhodophyta*, *Trypanosomatidae*, *Plasmodium* and some fungi. They are structurally related to phosphatases from psychrophilic bacteria *Shewanella* and *Colwellia*, and are termed *Shewanella*-like phosphatases, or Shelphs. Phosphatases of the third group are distantly related to ApaH, bacterial diadenosine tetrakisphosphatases, and are termed *ApaH*-like phosphatases, or Alphs. Patchy distribution of Alphs in animals, plants, fungi, diatoms and kinetoplasts suggests that these phosphatases were present in the common ancestor of eukaryotes but were independently lost in many lineages. Rhilphs, Shelphs and Alphs form PPP clades, as divergent from "conventional" eukaryotic PPP phosphatases as they are from each other and from major bacterial clades. In addition, comparison of primary structures revealed a previously unrecognised (I/L/V)D(S/T)G motif, conserved in all bacterial and "bacterial-like" eukaryotic PPPs, but not in "conventional" eukaryotic and archaeal PPPs.

**Conclusions:** Our findings demonstrate that many eukaryotes possess diverse "bacterial-like" PPP phosphatases, the enzymatic characteristics, physiological roles and precise evolutionary history of which have yet to be determined.

## Background

Reversible phosphorylation of proteins is a ubiquitous mechanism, indispensable for regulation of virtually any cellular function. Therefore, protein kinases and phosphatases are of paramount importance for normal functioning of all metabolic and signalling pathways. In eukaryotes, PPP family is one of the two known protein phosphatase families specific for Ser and Thr [1-4]. Unlike eukaryotic (and archaeal [5]) PPP phosphatases, bacterial members of the family have broad substrate specificity [6] or may even be Tyr-specific [7-9]. Moreover, one group of bacterial PPPs are diadenosine tetraphosphatases [10,11]. Unlike eukaryotes, in prokaryotes PPP phosphatases appear to be facultative, since entirely sequenced genomes of some bacteria and archaea do not encode them [5,12]. Nevertheless, when present, they appear to play essential roles [13-15].

Three motifs (GDXHRG, GDXXDRG and GNH(E/D)), highly conserved in the N-terminal subdomains of the catalytic domains of all PPP phosphatases [10,11], contain most residues which coordinate metal ions in the active centre [16] and are considered as the signature of the PPP family.

In a previous work [17] we identified an unusual cDNA fragment from a moss *Physcomitrella patens*, showing no similarities to the known PPP phosphatases beyond the presence of the GDXHRG and GDXXDRG motifs. Detection of homologous cDNA sequences from *Arabidopsis* and rice suggested the presence of an unknown PPP group in plants, distinct from "conventional" eukaryotic PPP phosphatases [17].

We have now taken advantage of a much greater representation (as compared to 1999) of sequence databases for various species to further explore this initial observation. We present the evidence for the existence in eukaryotes of three "non-conventional" branches of the PPP family. We also identify a previously unrecognised conserved motif in the PPP catalytic domain, which can be used as a signature of "bacterial"-type PPP phosphatases.

## Results

### "*Rhizobiales* / *Rhodobacterales* / *Rhodospirillaceae* - like" PPP phosphatases in plants

Two *Arabidopsis* sequences, At3g09960 and At3g09970, were retrieved using the *P. patens* fragment [17] as a query in TblastN searches. They share 85% identity with each other at the protein level (see Figure 1). Both genes are transcribed, since full-length cDNAs have been detected in a large-scale transcription study [18]. They are arranged on chromosome 3 in tandem, suggesting their origin by a recent duplication. A number of ESTs from other plant species (but none from non-plant eukaryotes) were also

detected by TblastN searches, which in most cases provide no evidence for the existence of more than one isoform.

Among prokaryotes, related sequences were detected in some  $\alpha$ -*Proteobacteria*, the closest matches being with *Rhizobiales*, *Rhodobacterales* and *Rhodospirillaceae*. Therefore, we designate this group as *Rhizobiales* / *Rhodobacterales* / *Rhodospirillaceae*-like phosphatases, or Rhilphs (Figure 1; see also Table 1 and Figure 4), and the two *Arabidopsis* genes products At3g09960 and At3g09970 as "*Rhizobiales*-like" phosphatases 1 (RLP1) and 2 (RLP2), respectively.

### Structural features of Rhilphs

All residues that are expected to bind metal ions in the catalytic centre are conserved in Rhilphs (Figure 1). Rhilphs do not have N- or C-terminal extensions beyond their catalytic domains, which in many PPP phosphatases have regulatory function and / or interact with regulatory proteins / subunits. Instead, they have characteristic inserts between the conserved motifs GNH(E/D) and HAG (corresponding to HGG in "conventional" eukaryotic PPPs, see Figure 1). Notably, much shorter inserts are found at a similar position in  $\alpha$ -proteobacterial phosphatases (group I in Figure 1). Inserts in both plant Rhilphs and  $\alpha$ -proteobacterial phosphatases contain a conserved motif LXXAX-PXXH (Figure 1). Similarly to bacteriophage  $\lambda$ phosphatase ( $\lambda$ PP [19]), Rhilphs lack a region corresponding to  $\beta$ 8,  $\beta$ 9 and  $\alpha$ 9 of eukaryotic PPPs [20-22]. Like bacterial PPP phosphatases, Rhilphs do not have a SAPNY motif, conserved in the  $\beta$ 12- $\beta$ 13 loop of eukaryotic PPPs. Analysis of Rhilph sequences did not reveal targeting or signal peptides.

While this work was in progress, a phosphatase encoded by an *Arabidopsis* gene At1g07010 was reported in an independent study [4].

### "*Shewanella*-like" PPP phosphatases in plants, red algae, fungi and unicellular parasites

We undertook further TblastN searches using full-length *Arabidopsis* RLP2 as a query to see whether *Arabidopsis* genome encodes additional "bacterial-like" PPPs. These searches identified two more genes for putative PPP phosphatases, only distantly related to Rhilphs and to any other members of the family (Figure 2).

One of these genes is At1g07010<sup>1</sup>. At least three different predicted products of this gene could be found in protein databases. On the basis of comparison with EST sequences, we consider as the correct structure that of Q8RY10 with Asp and Gly at positions 109 and 208, respectively (see Figure 2). The other detected gene, At1g18480, is also represented in the databases by three



**Table 1: Species, accession numbers (UniProt, EMBL, NCBI or TIGR Gene Index) and common names (where available) of PPP phosphatase sequences shown in Figure 4. For *A. thaliana* sequences, gene numbers are also indicated. Sequence No. 67 is available from *Chlamydomonas reinhardtii* draft genome [65].**

No.	Accession	Species name
<b>Rhilphs</b>		
1	Q9SR61; At3g09960	<i>Arabidopsis thaliana</i>
2	Q9SR62; At3g09970	<i>Arabidopsis thaliana</i>
3	BE034080	<i>Mesembryanthemum crystallinum</i>
4	BQ995369	<i>Lactuca sativa</i>
5	AW034786	<i>Lycopersicon esculentum</i>
6	BU875146	<i>Populus balsamifera</i>
7	BG457803	<i>Medicago truncatula</i>
8	AV425727	<i>Lotus japonicus</i>
9	BM731295	<i>Glycine max</i>
10	BF261816	<i>Hordeum vulgare</i>
11	BQ788728	<i>Triticum aestivum</i>
12	AL731641	<i>Oryza sativa</i>
13	CF670562	<i>Pinus taeda</i>
14	BQ039171	<i>Physcomitrella patens</i>
<b>Group I (<math>\alpha</math>-Proteobacteria)</b>		
15	Q987U4	<i>Mesorhizobium loti</i>
16	Q8UA33	<i>Agrobacterium tumefaciens</i>
17	ZP_00054691	<i>Magnetospirillum magnetotacticum</i>
18	ZP_00015226	<i>Rhodospirillum rubrum</i>
19	ZP_00014771	<i>Rhodospirillum rubrum</i>
20	CAE28794	<i>Rhodopseudomonas palustris</i>
21	Q9ABQ8	<i>Caulobacter crescentus</i>
22	ZP_00051041	<i>Magnetospirillum magnetotacticum</i>
23	ZP_00093979	<i>Novosphingobium aromaticivorans</i>
24	Q92V37	<i>Sinorhizobium meliloti</i>
<b>Group VII (heterogeneous)</b>		
25	Q8YZT4	<i>Anabaena</i> sp.
26	Q9WZK1	<i>Thermotoga maritima</i>
27	O34205	<i>Ferribacterium islandicum</i>
28	NZ_AABE01000101	<i>Cytophaga hutchinsonii</i>
<b>Group III (<math>\gamma</math>-Proteobacteria and bacteriophage <math>\lambda</math>)</b>		
29	P03772	Bacteriophage $\lambda$
30	P55798	<i>E. coli</i> PrpA
31	Q8VPE2	<i>Salmonella typhimurium</i> PrpA
32	Q8Z487	<i>Salmonella enterica</i>
33	P55799	<i>E. coli</i> PrpB
<b>Group IV (Firmicutes)</b>		
34	Q81YR3	<i>Bacillus anthracis</i>
35	Q9FB69	<i>Lactococcus lactis</i>
36	Q97FF3	<i>Clostridium acetobutylicum</i>
<b>Alphs</b>		
37	BM291808	<i>Amblyomma variegatum</i>
38	TC9835	<i>Ciona intestinalis</i>
39	BU652795	<i>Chlamydomonas reinhardtii</i>
40	AC091781	<i>Trypanosoma brucei</i>
41	AC084046	<i>Trypanosoma brucei</i>
42	AL499620	<i>Leishmania major</i>
43	BQ143558	<i>Metarhizium anisopliae</i>
44	AC127427	<i>Magnaporthe grisea</i>
45	AA966318	<i>Aspergillus nidulans</i>
46	P40152	<i>Saccharomyces cerevisiae</i>
<b>ApaH</b>		
47	Q8Y1K9	<i>Ralstonia solanacearum</i> ApaH
48	Q9JVF4	<i>Neisseria meningitidis</i> ApaH
49	P05637	<i>Escherichia coli</i> ApaH
<b>Group VI (heterogeneous)</b>		
50	O31614	<i>Bacillus subtilis</i>

**Table 1: Species, accession numbers (UniProt, EMBL, NCBI or TIGR Gene Index) and common names (where available) of PPP phosphatase sequences shown in Figure 4. For *A. thaliana* sequences, gene numbers are also indicated. Sequence No. 67 is available from *Chlamydomonas reinhardtii* draft genome [65]. (Continued)**

51	O69213	<i>Anabaena</i> sp. PrpA
52	Q93JF4	<i>Streptomyces coelicolor</i>
53	Q9RS78	<i>Deinococcus radiodurans</i>
<b>Group II (Cyanobacteria)</b>		
54	O54390	<i>Microcystis aeruginosa</i> PPI-Cyano I
55	P74150	<i>Synechocystis</i> sp.
56	Q8YP31	<i>Anabaena</i> sp.
57	ZP_00072257	<i>Trichodesmium erythraeum</i>
58	Q8DGA2	<i>Thermosynechococcus elongatus</i>
<b>Shelphs</b>		
59	AC119500*	<i>Leishmania major</i>
60	Q8EBN0	<i>Shewanella oneidensis</i>
61	Q9S427	<i>Shewanella</i> sp.
62	TIGR_167879 Contig1731	<i>Colwellia psychrerythraea</i>
63	CF394707	<i>Pinus taeda</i>
64	TC31593	<i>Solanum tuberosum</i>
65	Q944L7; At1g18480	<i>Arabidopsis thaliana</i>
66	BF645180	<i>Medicago truncatula</i>
67	Scaffold_45	<i>Chlamydomonas reinhardtii</i>
68	Q9LMJ5; At1g07010	<i>Arabidopsis thaliana</i>
69	AW266595	<i>Mesembryanthemum crystallinum</i>
70	BG644111	<i>Lycopersicon esculentum</i>
71	BG450922	<i>Medicago truncatula</i>
72	BI787505	<i>Glycine max</i>
73	Q8L676	<i>Oryza sativa</i>
74	TC21958	<i>Hordeum vulgare</i>
75	AC007863	<i>Trypanosoma brucei</i>
76	AL499621	<i>Leishmania major</i>
77	TIGR_246197 Contig433	<i>Myxococcus xanthus</i>
78	EAK84303	<i>Ustilago maydis</i>
79	O74480	<i>Schizosaccharomyces pombe</i>
80	EAK87480	<i>Cryptosporidium parvum</i>
81	Q7RIH8	<i>Plasmodium yoelii</i>
82	Q8IKE5	<i>Plasmodium falciparum</i>
83	Q8I5Y5	<i>Plasmodium falciparum</i>
84	Q7RR22	<i>Plasmodium yoelii</i>
<b>Group V (heterogeneous)</b>		
85	O87639	<i>Streptomyces coelicolor</i>
86	Q9RVT7	<i>Deinococcus radiodurans</i>
<b>Archaea</b>		
87	O28453	<i>Archaeoglobus fulgidus</i> PPA
88	Q8ZW26	<i>Pyrobaculum aerophilum</i>
89	O34200	<i>Methanosarcina thermophila</i> PPI-arch2
90	Y12396	<i>Pyrodictium abyssii</i>
<b>"Conventional" eukaryotic PPP</b>		
91	Q9U493	<i>Plasmodium falciparum</i> PPJ
92	Q8I728	<i>Trypanosoma cruzi</i> PPEF
93	BH900132	<i>Ostreococcus tauri</i>
94	O14829	<i>Homo sapiens</i> PPEF1 (PP7)
95	Q8IDE7	<i>Plasmodium falciparum</i> PP5
96	P53041	<i>Homo sapiens</i> PP5

**Table 1: Species, accession numbers (UniProt, EMBL, NCBI or TIGR Gene Index) and common names (where available) of PPP phosphatase sequences shown in Figure 4. For *A. thaliana* sequences, gene numbers are also indicated. Sequence No. 67 is available from *Chlamydomonas reinhardtii* draft genome [65]. (Continued)**

97	P53043	<i>Saccharomyces cerevisiae</i> PPT
98	P32838	<i>Saccharomyces cerevisiae</i> PPG
99	P05323	<i>Homo sapiens</i> PP2A
100	O00743	<i>Homo sapiens</i> PP6
101	Q08209	<i>Homo sapiens</i> Calcineurin (PP2B)
102	P08129	<i>Homo sapiens</i> PPI
103	P32945	<i>Saccharomyces cerevisiae</i> PPQ
104	O49346	<i>Arabidopsis thaliana</i> PP7

\* This phosphatase is unlikely to be catalytically active due to replacements of essential residues in the active centre.

distinct deduced proteins. Comparison with *A. thaliana* ESTs confirms Q944L7 as the correct structure.

Genomic and EST database searches provided ample evidence for the presence of related phosphatases in a number of green plants, including multiple angiosperm species, pine and a unicellular green alga *Chlamydomonas reinhardtii* (Figure 2; see also Table 1 and Figure 4). Related sequences were also identified in some fungi (several basidiomycetes and an ascomycete *Schizosaccharomyces pombe*, but not other ascomycetes), in *Apicomplexa*, *Trypanosomatidae*, and in a red alga *Porphyra yezoensis* (for available sequence from the latter species, see Figure 5).

The most closely related prokaryotic phosphatases were detected in *Myxococcus xanthus* ( $\delta$ -Proteobacteria) and psychrophilic bacteria *Alteromonadales* ( $\gamma$ -Proteobacteria): uncharacterised phosphatases from *Shewanella oneidensis* and *Colwellia psychrerythraea* and a Tyr-specific phosphatase PPI from *Shewanella* sp. [8]. Therefore, we designate this phosphatase group as "*Shewanella*-like"  $\underline{p}$ hosphatases, or Shelphs, and the products of the two prototype *Arabidopsis* genes At1g07010 and At1g18480 as "*Shewanella*-like"  $\underline{p}$ hosphatases 1 (SLP1) and 2 (SLP2), respectively.

#### Structural features of Shelphs

Like in Rhilphs, all residues that are expected to bind metal ions in the catalytic centre are conserved in Shelphs (see Figure 1). Another feature common with Rhilphs is the presence of inserts (as compared to "conventional" eukaryotic PPPs) between the GNHE and H(A/G)G motifs (Figure 1), which are especially long in plant Shelphs. However, these inserts share no sequence similarity between Rhilphs and Shelphs and probably appeared in the two phosphatase groups independently. Like in Rhilphs, a region corresponding to  $\beta$ 8,  $\beta$ 9 and  $\alpha$ 9 of eukaryotic PPPs is absent in Shelphs, and the primary structure of the region corresponding to the  $\beta$ 12- $\beta$ 13 loop

is similar to that of typical bacterial PPPs. *A. thaliana* SLP1 and corresponding Shelph isoform from rice have chloroplast targeting sequence, which could not be detected in *A. thaliana* SLP2 and corresponding isoform from *Medicago truncatula*.

#### Eukaryotic PPPs distantly related to bacterial diadenosine tetraphosphatases

Identification of Rhilphs and Shelphs prompted us to perform extensive searches of eukaryotic sequence databases. These searches revealed the existence of other "bacterial-like" PPP phosphatases throughout eukaryotes. Sequences only distantly related to Rhilphs, Shelphs or any other PPP phosphatases were detected in several fungi (including a putative *S. cerevisiae* phosphatase reported previously [23]), in *Trypanosomatidae*, a tick *Amblyomma*, an ascidian *Ciona*, *Chlamydomonas*, pine and diatoms *Fragilariopsis cylindrus* and *Phaeodactylum tricorutum*. Blast searches using these sequences revealed that all of them share higher similarity to bacterial diadenosine tetraphosphatases (ApaH) than to other PPP groups. Therefore, we tentatively designate them as  $\underline{A}$ paH-like  $\underline{p}$ hosphatases, or Alphs (Figures 3 and 4; Table 1; partial sequences available for pine and diatoms are shown in Figure 5).

Alphs share a distinctive common structural feature. In the GDXXDRG motif, absolutely conserved in other PPPs, the second Asp (which stabilises the protonation of a His that directly participates in catalysis [20]) is replaced by a neutral amino acid, and the Arg residue (which coordinates phosphate [24]) is replaced, with one exception, by Lys (Figure 3). The former of these replacements is also found in ApaH, while the latter is unique to Alphs. While higher overall sequence similarity and a common alteration in the GDXXDRG motif are compatible with closer relatedness of Alphs to bacterial diadenosine tetraphosphatases, phylogenetic analysis using full length sequences failed to produce a robust tree due to high sequence diversity (not shown).

<i>AtSLP1</i>	FVSAPARRIVAVGDLHGDLGKARDALQLAGVLSDDGRDQ-----WVGD-----TVLVQV	100
<i>Oryza</i>	FVTAPGRRIVAVGDLHGDLNQTTRAILVMAGLSSSESDGHV-----WTGGQ-----TVLVQV	130
<i>AtSLP2</i>	RFEKP-ERLVAIGDLHGDLKKSREAFKIAGLDSSDR-----WTGGS-----TMVVOI	95
<i>Medicago</i>	RLPSP-SRLTATGDLHGDLKKSKEALSTIAGLDSSGN-----YTGGS-----ATVVOI	96
<i>Chlamydomonas</i>	TYVTATGRITATGDLHGDLKAVEALKLGRVLSVSDGEV-----SIVGGD-----TVVVOI	147
<i>Plasmodium 1</i>	SNIKWEHDFSLSDLHSDLDALKKTLTLENIDEENN-----AIREN-----VFVLIIT	77
<i>Plasmodium 2</i>	DNIQWNHKLITATGDLHGDLVSLKLLLRHSNLDENDE-----WIAED-----VMIVOV	97
<i>Trypanosoma 1</i>	TMRVEGRQLVVVCDLHGDLNQTLLAILKITGLVDDRQH-----WIGGD-----SFFVOL	142
<i>Trypanosoma 2</i>	VIPIEHRIVAVGDLHGDAERFRQILEMSGVLSLRNSSSKQVVKPRAGTKEGNFREYFRTLRITLICT	102
<i>S. pombe</i>	AKIKYPTVVYATGDLHGDFPNALDVLSAAGVSPVPFPE-----WTAGN-----ATIVOT	100
<i>Shewanella</i>	DTYLGIKVVVAVSDVHGQOYDVLTLTKKQKLLDSDGN-----WAFGE-----GHVMT	131
<i>PP1</i>	ILLELEAPLKICGDLHGQOYDILLRLFEYCGFPPE-----SNIYFL	90
	β2 * * α2 β3	
<i>AtSLP1</i>	GDILDRGDEETALLSLRSLDDOAKANGGA--VFQVNGNHEITMVEGDFRYVDARAFDECTDFLDYLEDY	168
<i>Oryza</i>	GDILDRGDEETALLSLRSLNMOAKSQGGA--VFQVNGNHEITMVEGDFRYVDPGAFDECIRFMEYLDC	198
<i>AtSLP2</i>	GDVLDRCGEELKILLYLEKLRKRAERAGGK--ILTMNGNHEITMVEGDFRYVTKKGLEEFQIWDWYCLG	163
<i>Medicago</i>	GDVLDRCGEDEIKILLYLEKLRKRAAHCN--FITMNGNHEITMVEGDFRYVATKNGVEEFKVVWLEWFRQG	164
<i>Chlamydomonas</i>	GDVLDRCGEVEICINLRYLDTEARKXGGA--VYMLNGNHEITMVEGDFRYVATKNGVEEFKVVWLEWFRQG	164
<i>Plasmodium 1</i>	GDVLDRCGEDEETALLSLRSLNMOAKSQGGA--VFQVNGNHEITMVEGDFRYVDPGAFDECIRFMEYLDC	134
<i>Plasmodium 2</i>	GDILDRGVEGTYHYDYLLKQDAIKKNSK--VILIMGNHEITMVEGDFRYVATKNGVEEFKVVWLEWFRQG	155
<i>Trypanosoma 1</i>	GDILDRGVEGTYHYDYLLKQDAIKKNSK--VILIMGNHEITMVEGDFRYVATKNGVEEFKVVWLEWFRQG	200
<i>Trypanosoma 2</i>	GDILDRGVEGTYHYDYLLKQDAIKKNSK--VILIMGNHEITMVEGDFRYVATKNGVEEFKVVWLEWFRQG	162
<i>S. pombe</i>	GDVLDRCGEDEETALLSLRSLNMOAKSQGGA--VFQVNGNHEITMVEGDFRYVDPGAFDECIRFMEYLDC	164
<i>Shewanella</i>	GDVLDRCGEDEETALLSLRSLNMOAKSQGGA--VFQVNGNHEITMVEGDFRYVDPGAFDECIRFMEYLDC	196
<i>PP1</i>	GDVLDRCGEDEETALLSLRSLNMOAKSQGGA--VFQVNGNHEITMVEGDFRYVDPGAFDECIRFMEYLDC	150
	* α3 β4 * α4 α5	
<i>AtSLP1</i>	AQDWDKAFNRWIFESRQWKEDR-RSSQTYWDQWNVKQKGVIAARSVLLRPGGRLACELSRHGVLIRVN-	236
<i>Oryza</i>	DGNWDDAFLNWVNCERWKEEYPMSPNGDWRPWNFVKKQKGIARSSSLFKRGGPLACELARHPVVLNVN-	267
<i>AtSLP2</i>	NK-----MKTLCSSGLDK-PKPDYEGIPMSFPRMRADCFEGIRARIALALRPDGP IAKRFLTKNQTVAVVG	226
<i>Medicago</i>	NK-----MKNLCKGLEETVVDPLENVHVAFRVGRVEEFHDGFRARVAALRPNP I SKRFFTONVTVLVVG	228
<i>Chlamydomonas</i>	QYVDANNAVALKALVHCDG	
<i>Plasmodium 1</i>	KARNKLFKKGEVLYNYLLDLPFVIKVN-	161
<i>Plasmodium 2</i>	FQNNRNRLFSFTNKNYGFKKL-IRLPVIVKI	187
<i>Trypanosoma 1</i>	CGVSGDRLLSNRTSVGMYLRTKAIFFHN-	230
<i>Trypanosoma 2</i>	LTRALRKRFAELDGTGGF ILENFTVAYAVA	193
<i>S. pombe</i>	NRIIDWGHSGEIGNLLSEYNVYKDNIT	193
<i>Shewanella</i>	NKLYGADTEIGQWLRSK-NTTIKIND-	221
<i>PP1</i>	IKLWTFITDCFNCLP IAAIVD-	166
	α6 β5	
<i>AtSLP1</i>	-NWLECHCGLLPH---HVAYCIEIRLNREVSTWMSPTNYEDS-----PQMPFIATRGYDSVWVSRILYS	295
<i>Oryza</i>	-DWLECHCGLLPH---HVEYCIERNREVSVMKSSSGSDSD-----ELDIPFIATRGYDSVWVSRILYS	327
<i>AtSLP2</i>	-DSVLEVHCGLLAE---HIEYCLERTNEEVRGWINGFKGGRYA-----PAYCRGNSVWVLRKFS	281
<i>Medicago</i>	-DSVLEVHCGLLKE---HVDYCLEKINGEVSQWYKGLFGRNFS-----PPYCRGRNALVWLRKFS	283
<i>Chlamydomonas</i>	-RWCR-HGCEQAVAGNAGLQEGRS--GKGEPA SARWRYRRE-----LAKLLLGEGEPVWVSRYYI	
<i>Plasmodium 1</i>	-DILSHASHILPY---YAKRCIDYNDGRSEIKNNCTLLKF-----KRKTGQRFVCCYNGP I FNRYFS	222
<i>Plasmodium 2</i>	NNIIEHCGISKHMSEYDINTINLKTRELQIENCKMFOFEKY-----NYLSKEGVLNNEIS	244
<i>Trypanosoma 1</i>	-EFLBMHCGIESTA-TASIIICIDKIH-EFNKDLRKALNTGLTLP LGSTGLNLAEGGGQEVVNP I LVRSIL	297
<i>Trypanosoma 2</i>	-DILSHASHILPY---YAKRCIDYNDGRSEIKNNCTLLKF-----KRKTGQRFVCCYNGP I FNRYFS	244
<i>S. pombe</i>	-GSHBMHAGLSPE---WAYRE-ETVNELGKELLSHFMSREKI-----PEYLEDFWAIEGPMWYRGLA	250
<i>Shewanella</i>	-VLYMHGCTISSE---WISREI-TLDKANALYRANVDASKKS-----LKADDLINF LFFGNP I WYRGLA	280
<i>PP1</i>	EKIFCCHCGI SPD-----LQSMQIRR-----IMRP TDVPDQGLLCDLLWSDPK	211
	β6 * α7 α8 β7	
<i>AtSLP1</i>	RETSELEDYQI--EQVNKILHDTLEAVGAKAMVVGHTPQL---SGVNCEYCGIWRVDVGMSSGVLDLSDRP	360
<i>Oryza</i>	QGPTMTRHSW--KLSSVAERTLKSVAKGMVVGHTPQL---RGVNSKCDGKVCVVDVGMSPGVLYSRP	392
<i>AtSLP2</i>	EEMAHK-----CDCAALEHALSTIPGVKRMIMGHTIQD---AGINGVCNDKATRIDVGMKGCADGLP	341
<i>Medicago</i>	DGN-----CDCSSLEHVLSTIPGVKRMIMGHTIQD---EGINGVCENKATRIDVGMKGCADGLP	340
<i>Chlamydomonas</i>	QAPNEQ-----VCAQVAVAER-LGVORVVVGHVQVOR---GGRVSSRCGGQLIMVDVGLSRAIAGEMA	
<i>Plasmodium 1</i>	R-AAEMPFR---REVCKSLFKTLNKLAKKLVNNGHTIQD---NRKVNEYCKGGLIMADTCHSKWKYGVIN	285
<i>Plasmodium 2</i>	H---QVKLQ---PKKTCCKHLRNILKKNYAKGLVGHVQVOR---SHEIQTYCNNSFFLIDTCHSLFMNNGQP	306
<i>Trypanosoma 1</i>	---NVRCK-----DLIKVLQNKFAKIKSVVVGHTPQL---SGVNCEYCGIWRVDVGMSSGVLDLSDRP	356
<i>Trypanosoma 2</i>	LDAENGR-----ADTKALASLEVKRIVVGHVQVOR---SGRVETFCGGSVTAIDVGMKGCADGLP	303
<i>S. pombe</i>	Q-LSEEEA---CEVALNVKTL---NVNRLVVGHTPQL---HGIVSRCEGRIILLIDTCHSLFMNNGQP	308
<i>Shewanella</i>	S-ETFTA---ELDTILQHF---NVNHLVVGHTPQL---RVLGLFHNKVIADVSSIKVGVKSGELL	335
<i>PP1</i>	DVQGWGEDRGVSETFGAEVVAKFLHKHDLDFICRAHQVVE---DGYEFFAKROLVTLF SAPNYC---GEFD	277
	β8 β9 α9 β10 * β11 β12	

**Figure 2**  
**Comparison of the primary structures of "Shewanella-like" phosphatases (Shelphs) and human PP1α as a prototype of eukaryotic PPP phosphatases.** Designations for conserved amino acid residues are as in Figure 1. For *Oryza sativa*, dashed underlined C-terminal sequence has been corrected by comparison with ESTs. Accession numbers: *Plasmodium falciparum* 1, Q815Y5; 2, Q81KE5; *Trypanosoma brucei* 1, AC007863; 2, AC084046.12. *Chlamydomonas reinhardtii* sequence is translation of the EST entries BG855683 and B1995255. Other accession numbers are indicated in Table 1.

Consensus:	GDxxDRG	Accession	EST	
<b>Amblyomma</b>	KLFVGDLVN <b>R</b> GPKE	BM291808	+	Animals
<b>Ciona</b>	VIFVGDVVN <b>R</b> GPKSL	TC9835	+	
<b>Aspergillus</b>	LIFTGD <b>L</b> I <b>R</b> KCPD <b>S</b> L	AA966318	+	Fungi
<b>Botryotinia</b>	LILAGDMI <b>S</b> KGPDS <b>P</b>	AL110714	+	
<b>Magnaporthe</b>	LVLVGD <b>M</b> VN <b>R</b> GPDS <b>A</b>	AC127427	+	
<b>Metarhizium</b>	VIFVGD <b>L</b> V <b>S</b> R <b>G</b> PD <b>S</b> G	BQ143558	+	
<b>Neurospora</b>	LILAGD <b>L</b> V <b>T</b> R <b>G</b> PD <b>S</b> K	EAA35273	+	
<b>Paracoccidioides</b>	LIFTGD <b>L</b> I <b>R</b> GP <b>S</b> I	CA583218	+	
<b>Saccharomyces</b>	MILLG <b>D</b> F <b>I</b> H <b>R</b> GPDS <b>D</b>	P40152	+	
<b>Chlamydomonas</b>	LVLVGD <b>L</b> VN <b>R</b> GPAS <b>L</b>	BU652795	+	Plants
<b>Phaeodactylum</b>	VILVGD <b>L</b> CN <b>R</b> GPDS <b>T</b>	CD375576	+	Diatoms
<b>Leishmania</b>	LFITGD <b>V</b> VN <b>R</b> GPDS <b>I</b>	AL499620	+	Kinetoplasts
<b>Trypanosoma</b>	LIHVGD <b>L</b> VN <b>R</b> GPDS <b>L</b>	AC084046	+	
<i>E. coli</i>	LWLTGD <b>L</b> V <b>R</b> RG <b>S</b> L	P05637		Bacterial ApA
<i>Actinobacillus</i>	LYLTGD <b>L</b> L <b>R</b> GDNS <b>L</b>	O52655		
<i>Magnetococcus</i>	LYFTGD <b>L</b> I <b>S</b> R <b>G</b> PH <b>L</b>	ZP_00044402		
<i>Neisseria</i>	LWLTGD <b>L</b> VN <b>R</b> GP <b>S</b> L	Q9JVF4		
<i>Pseudomonas</i>	LWL <b>V</b> GD <b>L</b> VN <b>R</b> GP <b>S</b> L	Q9I5U7		
<i>Ralstonia</i>	LR <b>F</b> VGD <b>L</b> VN <b>R</b> GPDS <b>L</b>	Q8Y1K9		

**Figure 3**  
**Characteristic modifications (shaded in black) in the conserved PPP signature motif GDXXDRG in bacterial diadenosine tetraphosphatases and eukaryotic Alphas.** Eukaryotic species are shown in bold. Plus signs indicate that gene expression is confirmed by the presence of ESTs.

**Relationship of novel eukaryotic PPP groups to known PPP phosphatases**

In order to better understand the relationship of "bacterial-like" PPP phosphatases in eukaryotes to each other and to bacterial PPPs, we attempted to extend our previous phylogenetic analysis of eukaryotic PPP phosphatases [25] by including PPP sequences from a number of bacteria and archaea. Primary structures of bacterial PPP phosphatases are extremely diverse and, outside the relatively conserved N-terminal subdomain of about 100 amino acids containing the GDXHG, GDXXDRG and GNH(E/D) motifs, they share only a few conserved residues. Moreover, many of the sequences have long insertions at different positions. This leads to the failure to produce informative alignments of full-length catalytic domains. Therefore, we aligned more conserved N-terminal subdomains only, an approach applied previously by Kennelly [6] to a much smaller set of PPP sequences available at that time. Phylogenetic reconstruction was attempted with either neighbor-joining (as implemented in PHYLIP [26] or SplitsTree [27]) or maximum likelihood analysis using quartet puzzling (TreePuzzle [28]); in the latter case a smaller dataset consisting of with consisting of 32 representative sequences was analyzed due to the inability of the algorithm to handle large datasets. Due to the relatively short length of the sequences and their high diversity, some of the major clades did not receive significant bootstrap support and were different depending on the method used, although most major clades, including

Rhilphs and Shelphs, were recovered by both methods. Alphas tended to be grouped together by neighbor-joining but were split into smaller clades when maximum likelihood analysis was used. However, we still tentatively consider Alphas as a single group due to the characteristic replacements in their catalytic centre.

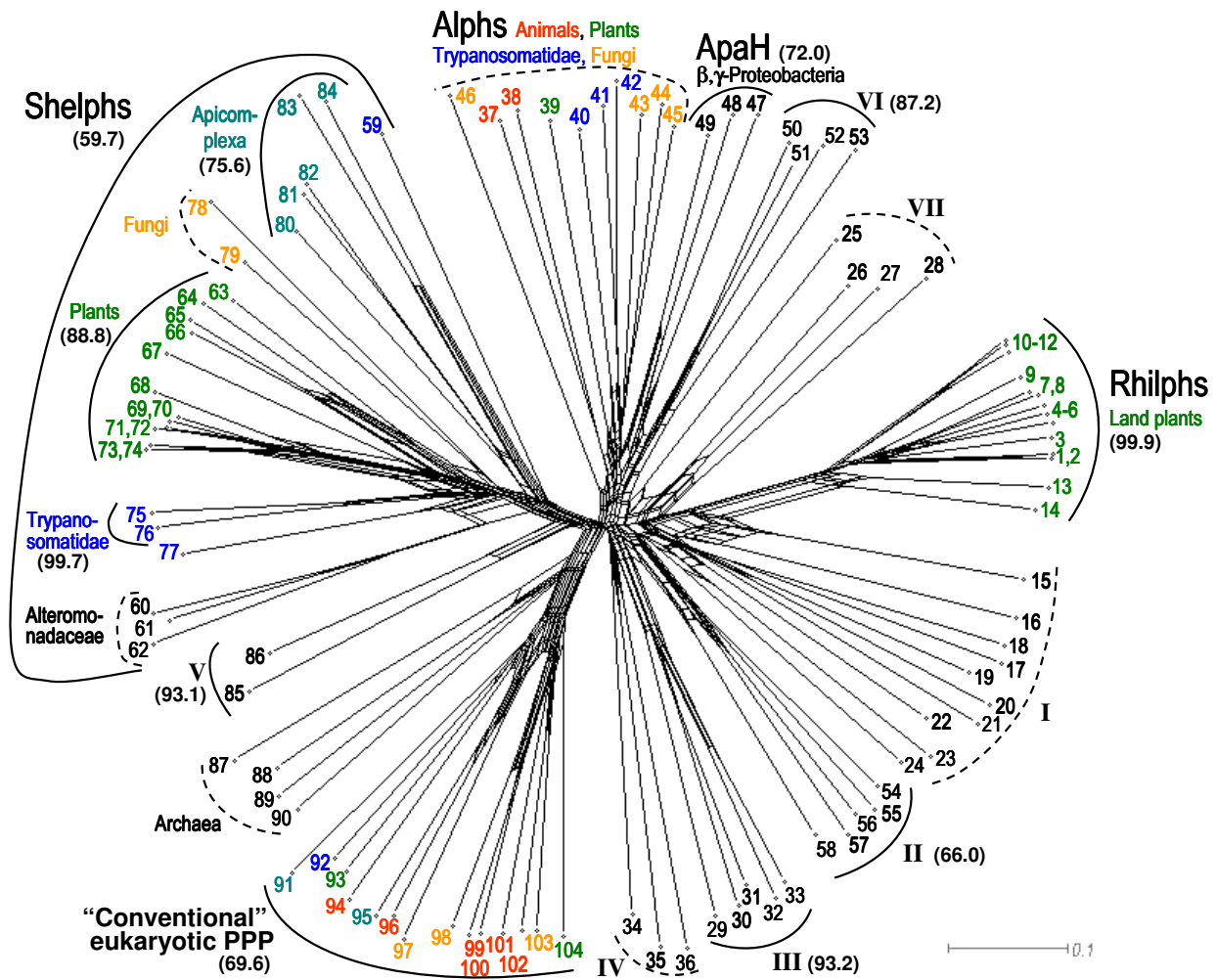
To circumvent the ambiguity of the results, we used Neighbor-Net [29], a neighbor-joining based method that constructs phylogenetic networks rather than trees and thus represents conflicting signals and visualises feasible trees in a single plot (Figure 4; for a high-resolution image, see Additional file 1). The Neighbor-Net analysis accurately identified the major clades such as eukaryotic and archaeal phosphatases, as well as their closer relationship to each other than to bacterial PPPs [6]. Separation of "conventional" eukaryotic PPPs into two branches, suggested previously from the analysis of the full-length catalytic domains [25], was also recovered. As it was suggested by initial sequence similarity searches, Rhilphs, Shelphs and Alphas represent distinct major clades of the PPP family, as divergent from "conventional" eukaryotic and archaeal PPP phosphatases as they are from major bacterial clades (Figure 4; Additional file 1).

**Common structural elements in all "bacterial-like" PPP phosphatases from eukaryotes and bacterial phosphatases**

C-terminal regions of the catalytic domain of all "conventional" eukaryotic PPP phosphatases share a highly conserved (with minor variations) SAPNY motif, located in the β12-β13 loop. This loop and the Tyr residue of the SAPNY motif in particular are implicated in interaction with regulators and inhibitors [21,30-32]. β strands (β9 and β10) corresponding to β12 and β13 are conserved in λPP [19]. However, the sequence on the C-terminal side of β9 is dissimilar to SAPNY in λPP and in bacterial PPPs. A conservative replacement of the first Ser of SAPNY by Thr is found in many bacterial sequences (this Thr is however only moderately conserved and is replaced by Glu or Gln in all Rhilphs and by Val or Phe in most Shelphs). The two adjacent positions are occupied by highly conserved Asp and Gly residues, respectively, thus defining a previously unrecognised motif (I/L/V)D(S/T)G. This motif is present in all examined bacterial PPPs, as well as in all "bacterial-like" phosphatases from eukaryotes described above (Figure 5; see also Additional file 2 for a more complete alignment).

In addition, we note the presence of another characteristic feature of bacterial and "bacterial-like" PPPs: the His residue (H248 in PP1; H186 in λPP) coordinating one of the metal ions in the catalytic centre is preceded by an absolutely conserved Gly; this residue is conserved in some archaeal but not in "conventional" eukaryotic PPPs (Figure 5; Additional file 2).





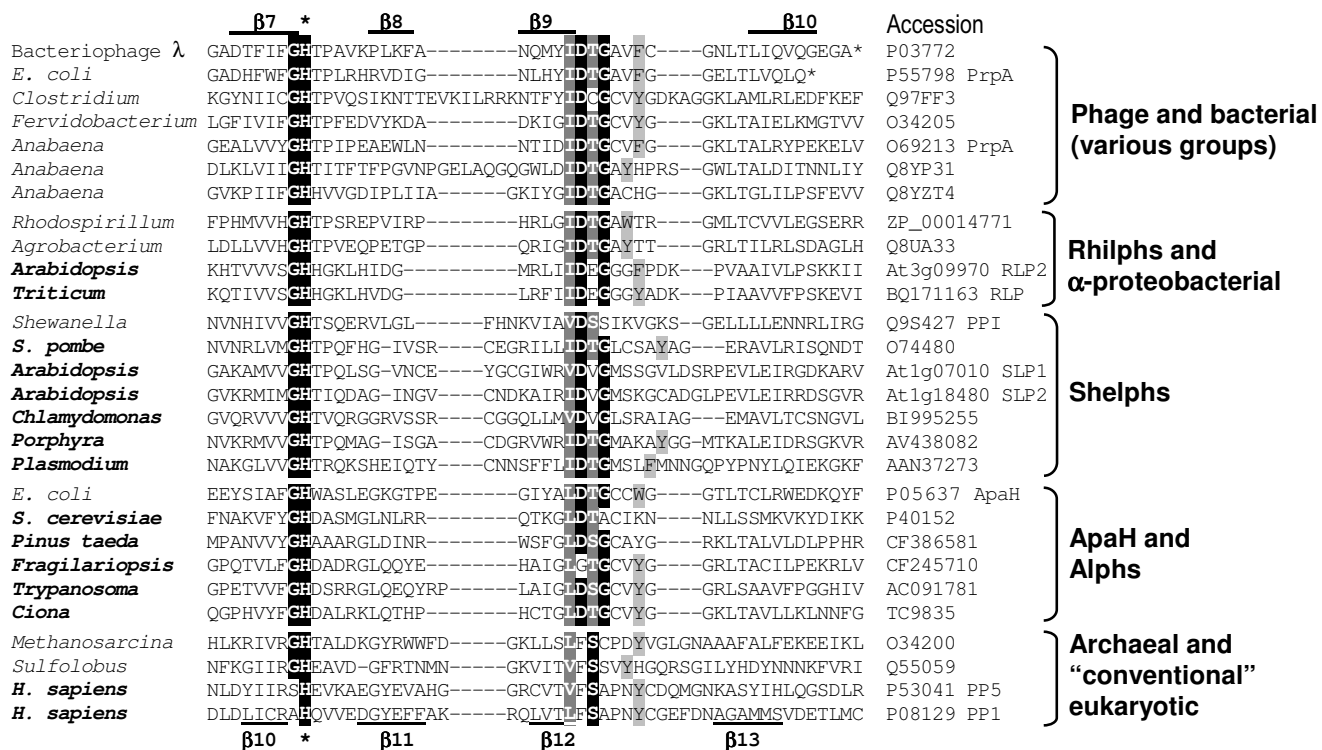
**Figure 4**  
**Neighbor-Net analysis of the conserved N-terminal subdomains (starting 5 amino acid residues before conserved GDXHG and ending 25 residues after GNH(E/D) of 104 bacterial, archaeal and eukaryotic PPP phosphatases.** Bootstrap values exceeding 50% (out of 1000 resamplings) were obtained in a separate neighbour-joining analysis and are shown in brackets. Species and accession numbers are listed in Table I. Note that groups designated as I, IV and VII did not receive significant bootstrap support; corresponding sequences are grouped together for convenience of their representation in Table I. This image (and bootstrap values for the alternative splits) can be viewed at a higher resolution as the Additional File I.

**Discussion**

In this report, we have documented the presence in different eukaryotic lineages of the genes that encode PPP phosphatases resembling those of bacterial origin, rather than "conventional" eukaryotic members of the family. Catalytic domains of these "bacterial-like" phosphatases are characterised by relatively conserved structure of the N-terminal subdomains, but very diverse organisation of the C-terminal subdomains, where conserved motifs and

residues forming the active centre are separated by sequences of various length that share little or no similarity between different clades (Figure 6). In most cases, corresponding EST sequences could be detected, which confirms that these genes are expressed.

The most conspicuous presence of "bacterial-like" PPPs has been detected in plants. Plants possess phosphatases from all three novel groups described in this work.

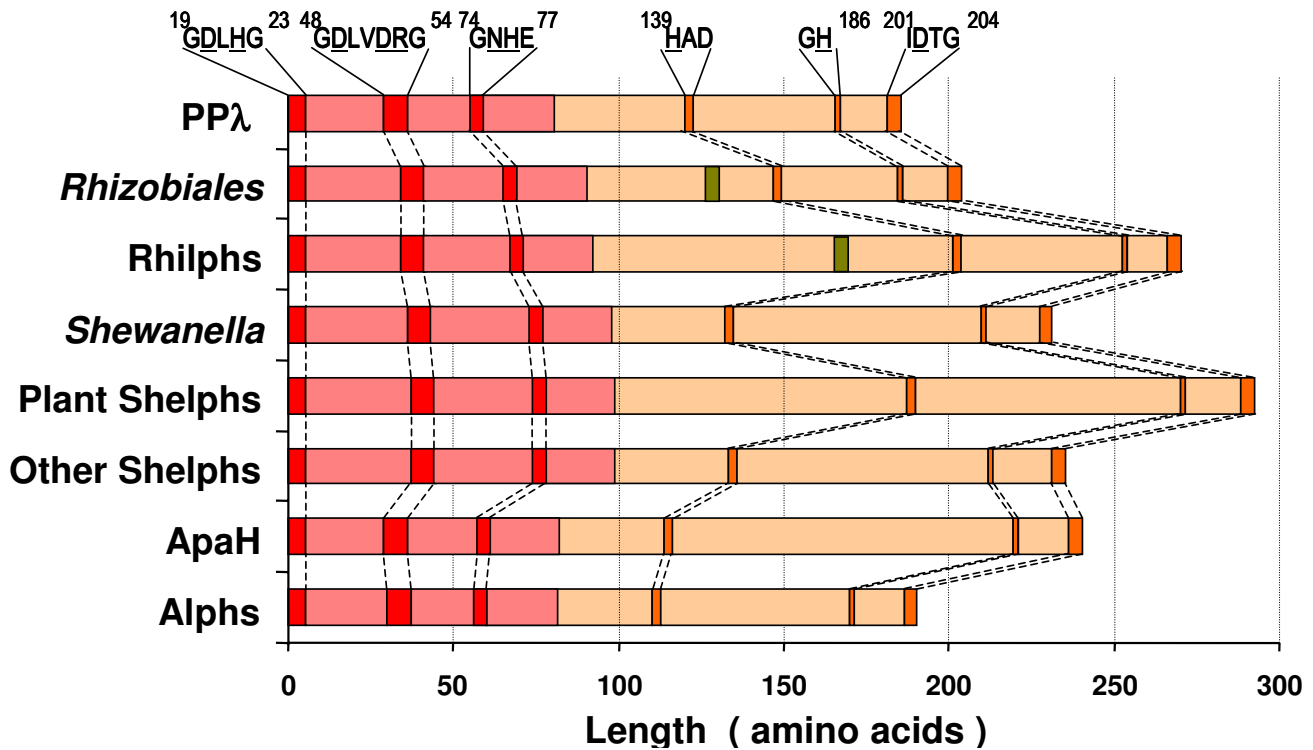


**Figure 5**  
**Distinct conserved motifs in the C-termini of bacterial and "bacterial-like" PPP phosphatases from eukaryotes as opposed to archaeal and eukaryotic PPP phosphatases.** A His residue directly binding a metal ion in the catalytic centre (marked with asterisk), and the elements of secondary structure are shown for bacteriophage λ phosphatase and for human PPI according to ref. [20] and [19], respectively. The (I/L/V)D(S/T)G motif, highly conserved in bacterial and "bacterial-like" phosphatases, is highlighted. An expanded version of this alignment can be viewed as the Additional File 2.

Rhilphs are most closely related to PPP phosphatases from a number of α-proteobacteria, including purple photosynthetic bacteria and *Rhizobiales*. The absence of related sequences in eukaryotes other than land plants suggests that Rhilphs may have been acquired after plants started colonising land. Although bacterial lineage that could be the source for plant Rhilphs could not be unambiguously identified by phylogenetic reconstruction, *Rhizobiales* (or purple photosynthetic bacteria from which *Rhizobiales* are thought to have originated [33,34]) appear to be likely candidates. Indeed, *Rhizobiales* have transmissible chromosomal elements and some, like *Agrobacterium*, are able to integrate their plasmid genes into plant genome [35] or even transform animal cells [36], a situation that would be ideally suited for a horizontal gene transfer to occur. Interestingly, the presence of genes of rhizobial origin has been detected in plant parasitic nematodes [37]. Possible origin of plant Rhilphs from α-proteobacterial phosphatases is also supported by the

presence in the enzymes of both groups of characteristic inserts in similar positions, which share some sequence similarity (see Results).

Phosphatases of another group, designated as Shelphs, are found in green plants, in a red alga, in *Apicomplexa*, *Trypanosomatidae*, as well as in some fungi. The similarity between proteins from *Apicomplexa* and *Trypanosomatidae* and those from plants is well documented. *Trypanosomatidae* are related to photosynthetic euglenoids and are thought to have lost plastids secondarily [38]. *Apicomplexan* parasites have a relict plastid, originated from the engulfment of a red alga [39]. Thus, the presence of phosphatases shared by plants, red algae, *Apicomplexa* and *Trypanosomatidae* is not surprising and probably reflects the presence of Shelphs in a common ancestor of photosynthetic eukaryotes. The presence of chloroplast targeting sequence in SLP1 suggests a possible origin of Shelphs from a bacterial precursor of the chloroplast (however it



**Figure 6**  
**Schematic diagram depicting organisation of the catalytic domains of the phosphatase groups discussed in this study.** N-terminal subdomains (used in the alignment for the Neighbor-Net analysis, Figure 4) and C-terminal subdomains are shown in red and yellow, respectively. Positions of the conserved motifs in PPλ and the residues forming the active centre (underlined; shown according to ref. [19]) are shown. Positions of the LXXAXPXXH motif in plant Rhilphs and related phosphatases from *Rhizobiales* are indicated (green boxes). For more detailed information on the position of inserts in Rhilphs and Shelphs relative to the elements of the secondary structure, see Figures 1 and 2, respectively.

should be noted that Shelphs are absent from cyanobacteria); alternatively, this sequence may have appeared secondarily. Protein Ser/Thr phosphorylation / dephosphorylation is essential for regulation of photosynthesis, and unidentified okadaic acid-insensitive protein phosphatases in chloroplasts have been reported [40,41]. SLP1 appears to be a good candidate for such a phosphatase.

The origin of fungal Shelphs is unclear. Curiously, they are found in basidiomycetes and in an ascomycete *S. pombe*, but not in a number of other ascomycetes, whose genomes have been completed. Current data do not permit to discriminate between (i) the presence of Shelphs in a common ancestor of eukaryotes and their loss in such lineages as animals and many fungi, and (ii) independent acquisition of Shelphs from bacteria by an ancestor of photosynthetic eukaryotes and by fungi. Further sequencing of eukaryotic genomes may shed light on the evolutionary history of this PPP group.

The third group of "bacterial-like" phosphatases detected in eukaryotes, designated here as Alphs, appears to be distantly related to bacterial diadenosine tetraphosphatases ApaH. Patchy distribution in several eukaryotic kingdoms suggests that Alphs were probably present in the common ancestor of eukaryotes, but were independently lost in many lineages, including insects, vertebrates and flowering plants. A characteristic modification of the conserved GDXXDRG motif shared only with ApaH further supports a suggestion that Alphs may represent a divergent branch of diadenosine tetraphosphatases, rather than protein phosphatases. However, relatedness of eukaryotic Alphs to bacterial diadenosine tetraphosphatases remains hypothetical, since Alph sequences are too divergent from ApaH, as well as from each other, to permit a reliable phylogenetic reconstruction.

Diadenosine oligophosphates are considered as emerging signalling molecules in both intra- and intercellular signalling in eukaryotes [42,43]. In particular, human

diadenosine oligophosphate hydrolase FHIT has been identified as a tumor suppressor [44]. It seems plausible that appearance of eukaryotic diadenosine oligophosphate hydrolases (structurally unrelated to the PPP phosphatases) may have made bacterial-type diadenosine tetraphosphatases redundant, leading to their loss in many eukaryotic lineages. It would be interesting to test experimentally whether Alphs are indeed diadenosine oligophosphatases.

More generally, an important implication of our findings is that many eukaryotes possess PPP phosphatases with yet undetermined substrate specificity. Eukaryotic PPP phosphatases are generally considered as Ser/Thr specific *in vivo*, although they may be able to dephosphorylate phosphoTyr-containing substrates *in vitro* (e.g. [45,46]). This is probably true for archaeal PPPs as well [5]. However, Ser/Thr specificity is not a feature of bacterial PPP phosphatases [7-9,13,47-49]. Thus, it would not be possible to predict substrate specificity of uncharacterised "bacterial-like" PPP phosphatases without experimental evidence. In particular, since *Shewanella* PPI is Tyr-specific [8], it would be interesting to determine substrate specificity of eukaryotic Shelpshs. It is also worth noting that interest in tyrosine phosphorylation in plants has recently been stimulated by identification of plant Tyr phosphatase genes and by the finding that Tyr phosphorylation is involved in the regulation of stomatal movement (reviewed by Luan [50]).

Three motifs, GDXHRG, GDXXDRG and GNH(E/D) form the diagnostic signature of all PPP phosphatases [10,11]. We detected a (I/L/V)D(S/T)G motif, which appears to be a characteristic signature of "bacterial"-type PPPs. The existence of such a motif is striking *per se*, taking into account extreme structural diversity of bacterial PPP phosphatases. It indicates that (I/L/V)D(S/T)G was probably present as the fourth "universal" signature motif in the common ancestor of PPP phosphatases, and was lost in the common lineage of archaeal and "conventional" eukaryotic PPPs. An alternative possibility could be that the (I/L/V)D(S/T)G motif was acquired by a bacterium and propagated by lateral gene transfer, replacing the ancestral SAPNY-related motif. However this scenario seems less likely, since the (I/L/V)D(S/T)G motif is present, with minor variations, in virtually all bacterial phosphatases, despite their great diversity, and is replaced by SAPNY-related sequences only in archaeal and "conventional" eukaryotic PPPs.

The Asp residue in the 2<sup>nd</sup> position of (I/L/V)D(S/T)G is highly conserved and can only be replaced by Glu, indicating that the negative charge is essential. The presence of a highly conserved Gly in the 4<sup>th</sup> position indicates that flexibility of the polypeptide chain is likely to be impor-

tant. The crystal structure of bacteriophage  $\lambda$  phosphatase (PP $\lambda$ ; [19]) shows that the Asp residue of (I/L/V)D(S/T)G (Asp202) is just downstream of the  $\beta$ 9 strand, which corresponds to the  $\beta$ 12 strand in mammalian PP1. In PP $\lambda$ , Asp202 is hydrogen bonded to a water molecule coordinated to one of the metal ions in the catalytic centre, which probably accounts for its conservation. In eukaryotic or archaeal PPPs, corresponding position is occupied by neutral residues (see Figure 5).

It would be tempting to speculate that this difference in the region just downstream of the  $\beta$ 9 ( $\beta$ 12) may be responsible for a feature that is common to all "bacterial"-type but not to eukaryotic / archaeal PPPs. One such feature is the Ser/Thr specificity of the latter group. The Tyr residue of the SAPNY motif has been suggested to provide a bulky phenol ring in the  $\beta$ 12- $\beta$ 13 loop, sufficient to sterically block access of phosphoTyr-containing substrates to the active site [32]. However this is unlikely to be the sole determinant of Ser/Thr specificity, since residues containing bulky aromatic rings (Tyr, Phe or Trp) are found in the same or adjacent positions in many bacterial phosphatases (Figure 5). Since the (I/L/V)D(S/T)G motif, absent in eukaryotic and archaeal PPPs, is involved in organisation of the catalytic centre [19], it is possible that this difference in the catalytic centre organisation may be one of the determinants of broad substrate specificity *vs.* Ser/Thr specificity.

## Conclusions

So far, eukaryotic PPP phosphatases were considered as a well-defined monophyletic group of enzymes, specifically dephosphorylating phosphoSer and phosphoThr, while a much more structurally and enzymatically diverse PPP phosphatases were known to be present in prokaryotes. Our findings demonstrate that, in addition to "conventional" eukaryotic PPP Ser/Thr-specific protein phosphatases, many eukaryotes possess very diverse "bacterial-like" PPP phosphatases. Enzymatic characteristics, physiological roles and evolutionary history of these phosphatases have yet to be revealed.

## Methods

### Detection of PPP phosphatase-coding sequences

Sequence similarity searches were conducted using BlastP or TblastN [51] at NCBI [52] in the following databases: "non-redundant" (NR), "expressed sequence tags" (EST), "genomic sequence survey" (GSS) and "high-throughput genomic sequences" (HTGS). Additional Blast searches of the following databases were performed: finished and unfinished genomes of eukaryotes at the NCBI [53]; fungal genomes at the Broad Institute [54]; plant genomes at The Arabidopsis Information Resource (TAIR) [55]; Gene Index databases of tentative consensus sequences (EST assemblies) at The Institute for Genomic Research (TIGR)

[56]; *Chlamydomonas reinhardtii* draft genome [57]. In all cases, reciprocal searches were used, i.e. hits retrieved by Blast searches were in their turn used as queries in the following Blast searches. Accuracy of gene prediction was examined by comparison of the retrieved sequences with translations of corresponding EST entries. In the absence of available ESTs, closely related sequences from other species were used. Taxonomy of the species from which the phosphatase sequences is given according to the NCBI taxonomy web site [58].

### Phylogenetic analysis

Multiple alignments were generated using CLUSTAL W [59] at Kyoto University Bioinformatics Centre [60] and edited manually. During manual editing, particular attention was paid to correct alignment of the PPP family signature motifs and other conserved residues known to constitute the catalytic site of PPP phosphatases. Phylogenetic tree construction by the neighbor-joining method [61] and bootstrap analysis were performed using the PHYLIP package, version 3.573 [26]. Maximum likelihood analysis was performed using TreePuzzle [28]. Possible alternative neighbor-joining based phylogenies were visualised using Neighbor-Net [29] as implemented in SplitsTree, version 4.β10 [62]

### Analysis of the primary structure

The presence of signal peptides and targeting sequences was analyzed using TargetP [63] at the the Centre for Biological Sequence Analysis, Technical University of Denmark [64].

### Abbreviations

EST, expressed sequence tag, Alph, ApaH – like phosphatase; PPP, protein phosphatases of the P family; Rhilph (RLP), *Rhizobiales* / *Rhodobacterales* / *Rhodospirillaceae* – like phosphatase; ShelpH (SLP), *Shewanella* – like phosphatase.

### Authors' contributions

Both authors contributed equally to this work.

### Additional material

#### Additional File 1

*Neighbor-Net analysis of the conserved N-terminal subdomains (starting 5 amino acid residues before conserved GDXHRG and ending 25 residues after GNH(E/D) of 104 bacterial, archaeal and eukaryotic PPP phosphatases. This version of Figure 4 is the original SplitsTree file that can be viewed using SplitsTree, freely available for download (see Methods). Bootstrap values (out of 100 resamplings) are shown and can be highlighted by selecting corresponding alternative splits. The file also contains the alignment used for the analysis (Input).*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-4-47-S1.nex>]

#### Additional File 2

*Distinct conserved motifs in the C-termini of bacterial and "bacterial-like" PPP phosphatases from eukaryotes as opposed to archaeal and eukaryotic PPP phosphatases. This is an expanded version of Figure 5.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-4-47-S2.doc>]

### References

- Cohen PTW: **Novel protein serine/threonine phosphatases: Variety is the spice of life.** *Trends Biochem Sci* 1997, **22**:245-251.
- Andreeva AV, Kutuzov MA: **RdgC/PP5-related phosphatases: novel components in signal transduction.** *Cell Signal* 1999, **11**:555-562.
- Ariño J: **Novel protein phosphatases in yeast.** *Eur J Biochem* 2002, **269**:1072-1077.
- Kerk D, Bulgrien J, Smith DW, Barsam B, Veretnik S, Gribskov M: **The complement of protein phosphatase catalytic subunits encoded in the genome of Arabidopsis.** *Plant Physiol* 2002, **129**:908-925.
- Kennelly PJ: **Archaeal protein kinases and protein phosphatases: insights from genomics and biochemistry.** *Biochem J* 2003, **370**:373-389.
- Kennelly PJ: **Protein phosphatases – a phylogenetic perspective.** *Chem Rev* 2001, **101**:2291-2312.
- Tsuruta H, Aizono Y: **Enzymatical properties of psychrophilic phosphatase I.** *J Biochem* 1999, **125**:690-695.
- Tsuruta H, Aizono Y: **Cloning of phosphatase I gene from a psychrophile, Shewanella sp, and some properties of the recombinant enzyme.** *J Biochem* 2000, **127**:143-149.
- Iwanicki A, Herman-Antosiewicz A, Pierechod M, Seror SJ, Obuchowski M: **PrpE, a PPP protein phosphatase from Bacillus subtilis with unusual substrate specificity.** *Biochem J* 2002, **366**:929-936.
- Koonin EV: **Bacterial and bacteriophage protein phosphatases.** *Mol Microbiol* 1993, **8**:785-786.
- Barton GJ, Cohen PTW, Barford D: **Conservation analysis and structure prediction of the protein serine/threonine phosphatases. Sequence similarity with diadenosine tetraphosphate from Escherichia coli suggests homology to the protein phosphatases.** *Eur J Biochem* 1994, **220**:225-237.
- Kennelly PJ: **Protein kinases and protein phosphatases in prokaryotes: a genomic perspective.** *FEMS Microbiol Lett* 2002, **206**:1-8.
- Missiakas D, Raina S: **Signal transduction pathways in response to protein misfolding in the extracytoplasmic compartments of E.coli : role of two new phosphoprotein phosphatases PrpA and PrpB.** *EMBO J* 1997, **16**:1670-1685.
- Umeyama T, Naruoka A, Horinouchi S: **Genetic and biochemical characterization of a protein phosphatase with dual substrate specificity in Streptomyces coelicolor A3(2).** *Gene* 2000, **258**:55-62.
- Shi L: **Manganese-dependent protein O-phosphatases in prokaryotes and their biological functions.** *Front Biosci* 2004, **9**:1382-1397.
- Barford D: **Molecular mechanisms of the protein serine/threonine phosphatases.** *Trends Biochem Sci* 1996, **21**:407-412.
- Andreeva AV, Kutuzov MA: **Physcomitrella patens gene/cDNA fragments related to genes encoding protein Ser/Thr phosphatases.** *J Plant Physiol* 1999, **155**:153-158.
- Haas BJ, Volfovsky N, Town CD, Troukhan M, Alexandrov N, Feldmann KA, Flavell RB, White O, Salzberg SL: **Full-length messenger RNA sequences greatly improve genome annotation.** *Genome Biol* 2002, **3**:research0029..
- Voegtli WC, White DJ, Reiter NJ, Rusnak F, Rosenzweig AC: **Structure of the bacteriophage λ Ser/Thr protein phosphatase with sulfate ion bound in two coordination modes.** *Biochemistry* 2000, **39**:15365-15374.

20. Egloff MP, Cohen PTW, Reinemer P, Barford D: **Crystal structure of the catalytic subunit of human protein phosphatase I and its complex with tungstate.** *J Mol Biol* 1995, **254**:942-959.
21. Goldberg J, Huang HB, Kwon YG, Greengard P, Nairn AC, Kuriyan J: **Three-dimensional structure of the catalytic subunit of protein serine/threonine phosphatase-I.** *Nature* 1995, **376**:745-753.
22. Kissinger CR, Parge HE, Knighton DR, Lewis CT, Pelletier LA, Tempczyk A, Kalish VJ, Tucker KD, Showalter RE, Moomaw EW, Gastinel LN, Habuka N, Chen X, Maldonado F, Barker JE, Bacquet R, Villafraña JE: **Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex.** *Nature* 1995, **378**:641-644.
23. Coster F, van Dyck L, Jonniaux J-L, Purnelle B, Goffeau A: **The sequence of a 13.5 kb DNA segment from the left arm of yeast chromosome XIV reveals MER1; RAPI; a new putative member of the DNA replication complex and a new putative serine/threonine phosphatase gene.** *Yeast* 1995, **11**:85-91.
24. Zhang J, Zhang Z, Brew K, Lee EYC: **Mutational analysis of the catalytic subunit of muscle protein phosphatase-I.** *Biochemistry* 1996, **35**:6276-6282.
25. Andreeva AV, Kutuzov MA: **PPP family of protein Ser/Thr phosphatases: Two distinct branches?** *Mol Biol Evol* 2001, **18**:448-452.
26. Felsenstein J: **PHYLIB (phylogeny inference package).** Version 3.5731999 Distributed by the author, Department of Genetics, University of Washington, Seattle .
27. Huson DH, Bryant D: **Estimating phylogenetic trees and networks using SplitsTree4, in preparation.** .
28. Schmidt HA, Strimmer K, Vingron M, von Haeseler A: **TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing.** *Bioinformatics* 2002, **18**:502-504.
29. Bryant D, Moulton V: **Neighbor-Net: An agglomerative method for the construction of phylogenetic networks.** *Mol Biol Evol* 2004, **21**:255-65.
30. Connor JH, Kleeman T, Barik S, Honkanen RE, Shenolikar S: **Importance of the  $\beta$ 12- $\beta$ 13 loop in protein phosphatase-I catalytic subunit for inhibition by toxins and mammalian protein inhibitors.** *J Biol Chem* 1999, **274**:22366-22372.
31. Lee EY, Zhang L, Zhao S, Wei Q, Zhang J, Qi ZQ, Belmonte ER: **Phosphorylase phosphatase: new horizons for an old enzyme.** *Front Biosci* 1999, **4**:D270-D285.
32. Watanabe T, Huang HB, Horiuchi A, da Cruze Silva EF, Hsieh-Wilson L, Allen PB, Shenolikar S, Greengard P, Nairn AC: **Protein phosphatase I regulation by inhibitors and targeting subunits.** *Proc Natl Acad Sci USA* 2001, **98**:3080-3085.
33. Sprent JI: **Evolution and diversity in the legume-rhizobium symbiosis: Chaos theory?** In *Symbiotic Nitrogen Fixation* Edited by: Graham PH, Sadowsky MJ, Vance CP. Dordrecht: Kluwer Academic Publishers; 1994:1-10.
34. Fleischman D, Kramer D: **Photosynthetic rhizobia.** *Biochim Biophys Acta* 1998, **1364**:17-36.
35. Escobar MA, Dandekar AM: **Agrobacterium tumefaciens as an agent of disease.** *Trends Plant Sci* 2003, **8**:380-386.
36. Kunik T, Tzfira T, Kapulnik Y, Gafni Y, Dingwall C, Citovsky V: **Genetic transformation of HeLa cells by Agrobacterium.** *Proc Natl Acad Sci USA* 2001, **98**:1871-1876.
37. Scholl EH, Thorne JL, McCarter JP, Bird DM: **Horizontally transferred genes in plant-parasitic nematodes: a high-throughput genomic approach.** *Genome Biol* 2003, **4**:R39.
38. Hannaert V, Saavedra E, Duffieux F, Szikora JP, Rigden DJ, Michels PA, Opperdoes FR: **Plant-like traits associated with metabolism of Trypanosoma parasites.** *Proc Natl Acad Sci USA* 2003, **100**:1067-1071.
39. Foth BJ, McFadden GI: **The apicoplast: a plastid in Plasmodium falciparum and other Apicomplexan parasites.** *Int Rev Cytol* 2003, **224**:57-110.
40. Hast T, Follmann H: **Identification of two thylakoid-associated phosphatases with protein phosphatase activity in chloroplasts of the soybean (Glycine max).** *J Photochem Photobiol B -Biology* 1996, **36**:313-319.
41. Hammer MF, Markwell J, Sarath G: **Purification of a protein phosphatase from chloroplast stroma capable of dephosphorylating the light-harvesting complex-II.** *Plant Physiol* 1997, **113**:227-233.
42. Kisselev LL, Justesen J, Wolfson AD, Frolova LY: **Diadenosine oligophosphates (Ap(n)A), a novel class of signalling molecules?** *FEBS Lett* 1998, **427**:157-163.
43. Vartanian AA, Suzuki H, Poletaev AI: **The involvement of diadenosine 5',5'''-P-I,P-4-tetrathosphate in cell cycle arrest and regulation of apoptosis.** *Biochem Pharmacol* 2003, **65**:227-235.
44. Pekarsky Y, Zaneni N, Palamarchuk A, Huebner K, Croce CM: **FHIT: from gene discovery to cancer treatment and prevention.** *Lancet Oncology* 2002, **3**:748-754.
45. Chan CP, Gallis B, Blumenthal DK, Pallen CJ, Wang JH, Krebs EG: **Characterization of the phosphotyrosyl protein phosphatase activity of calmodulin-dependent protein phosphatase.** *J Biol Chem* 1986, **261**:9890-9895.
46. MacKintosh C, Garton AJ, McDonnell A, Barford D, Cohen PTW, Tonks NK, Cohen P: **Further evidence that inhibitor-2 acts like a chaperone to fold PPI into its native conformation.** *FEBS Lett* 1996, **397**:235-238.
47. Zhuo S, Clemens JC, Hakes DJ, Barford D, Dixon JE: **Expression, purification, crystallization, and biochemical characterization of a recombinant protein phosphatase.** *J Biol Chem* 1993, **268**:17754-17761.
48. Shi L, Carmichael WW, Kennelly PJ: **Cyanobacterial PPP family protein phosphatases possess multifunctional capabilities and are resistant to microcystin-LR.** *J Biol Chem* 1999, **274**:10039-10046.
49. Shi L, Kehres DG, Maguire ME: **The PPP-family protein phosphatases PrpA and PrpB of Salmonella enterica serovar Typhimurium possess distinct biochemical properties.** *J Bacteriol* 2001, **183**:7053-7057.
50. Luan S: **Protein phosphatases in plants.** *Annu Rev Plant Biol* 2003, **54**:63-92.
51. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
52. **NCBI Blast search** [<http://www.ncbi.nlm.nih.gov/blast/>]
53. **NCBI BLAST with eukaryotic genomes** [[http://www.ncbi.nlm.nih.gov/sutils/genom\\_tree.cgi?organism=euk](http://www.ncbi.nlm.nih.gov/sutils/genom_tree.cgi?organism=euk)]
54. **Fungal Genomes Blast search at the Broad Institute** [[http://www.broad.mit.edu/cgi-bin/annotation/cgi/blast\\_page.cgi](http://www.broad.mit.edu/cgi-bin/annotation/cgi/blast_page.cgi)]
55. **The Arabidopsis Information Resource (TAIR) Blast search** [<http://www.arabidopsis.org/Blast/>]
56. **Gene Index databases of tentative consensus sequences at the Institute for Genomic Research (TIGR)** [<http://www.tigr.org/tdb/>]
57. **Chlamydomonas reinhardtii draft genome BLAST search** [<http://aluminum.igi-psf.org/prod/bin/runBlast.pl?db=chlre1>]
58. **NCBI Taxonomy Homepage** [<http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html>]
59. Higgins DG, Thompson JD, Gibson TJ: **Using CLUSTAL for multiple sequence alignments.** *Methods Enzymol* 1996, **266**:383-402.
60. **CLUSTAL W: Multiple Sequence Alignment at Kyoto University Bioinformatics Centre.** [<http://clustalw.genome.ad.jp/>].
61. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**:406-425.
62. **SplitsTree: Analyzing and Visualizing Evolutionary Data.** [[http://www-ab.informatik.uni-tuebingen.de/software/splits/welcome\\_en.html](http://www-ab.informatik.uni-tuebingen.de/software/splits/welcome_en.html)].
63. Emanuelsson O, Nielsen H, Brunak S, von Heijne G: **Predicting subcellular localization of proteins based on their N-terminal amino acid sequence.** *J Mol Biol* 2000, **300**:1005-1016.
64. **TargetP Server at the Centre for Biological Sequence Analysis, Technical University of Denmark** [<http://www.cbs.dtu.dk/services/TargetP/>]
65. **Chlamydomonas reinhardtii draft genome, scaffold 45** [[http://genome.igi-psf.org/cgi-bin/getScaffold?db=chlre1&scaffold=scaffold\\_45](http://genome.igi-psf.org/cgi-bin/getScaffold?db=chlre1&scaffold=scaffold_45)]