

RESEARCH ARTICLE

Open Access

# A survey of green plant tRNA 3'-end processing enzyme tRNase Zs, homologs of the candidate prostate cancer susceptibility protein ELAC2

Lijuan Fan<sup>1,2,3</sup>, Zhikang Wang<sup>1,2,3</sup>, Jinyu Liu<sup>1,2,3</sup>, Weili Guo<sup>1,2,3</sup>, Jie Yan<sup>4</sup> and Ying Huang<sup>1,2,3\*</sup>

## Abstract

**Background:** tRNase Z removes the 3'-trailer sequences from precursor tRNAs, which is an essential step preceding the addition of the CCA sequence. tRNase Z exists in the short (tRNase Z<sup>S</sup>) and long (tRNase Z<sup>L</sup>) forms. Based on the sequence characteristics, they can be divided into two major types: bacterial-type tRNase Z<sup>S</sup> and eukaryotic-type tRNase Z<sup>L</sup>, and one minor type, *Thermotoga maritima* (TM)-type tRNase Z<sup>S</sup>. The number of tRNase Zs is highly variable, with the largest number being identified experimentally in the flowering plant *Arabidopsis thaliana*. It is unknown whether multiple tRNase Zs found in *A. thaliana* is common to the plant kingdom. Also unknown is the extent of sequence and structural conservation among tRNase Zs from the plant kingdom.

**Results:** We report the identification and analysis of candidate tRNase Zs in 27 fully sequenced genomes of green plants, the great majority of which are flowering plants. It appears that green plants contain multiple distinct tRNase Zs predicted to reside in different subcellular compartments. Furthermore, while the bacterial-type tRNase Z<sup>S</sup>s are present only in basal land plants and green algae, the TM-type tRNase Z<sup>S</sup>s are widespread in green plants. The protein sequences of the TM-type tRNase Z<sup>S</sup>s identified in green plants are similar to those of the bacterial-type tRNase Z<sup>S</sup>s but have distinct features, including the TM-type flexible arm, the variant catalytic HEAT and HST motifs, and a lack of the PxKxRN motif involved in CCA anti-determination (inhibition of tRNase Z activity by CCA), which prevents tRNase Z cleavage of mature tRNAs. Examination of flowering plant chloroplast tRNA genes reveals that many of these genes encode partial CCA sequences. Based on our results and previous studies, we predict that the plant TM-type tRNase Z<sup>S</sup>s may not recognize the CCA sequence as an anti-determinant.

**Conclusions:** Our findings substantially expand the current repertoire of the TM-type tRNase Z<sup>S</sup>s and hint at the possibility that these proteins may have been selected for their ability to process chloroplast pre-tRNAs with whole or partial CCA sequences. Our results also support the coevolution of tRNase Zs and tRNA 3'-trailer sequences in plants.

**Keywords:** tRNA precursor (pre-tRNA), endonuclease, tRNase Z, ELAC2, post-transcriptional processing, green plant, phylogenetics

## Background

tRNA 3'-end maturation is a process through which the 3'-trailer sequence of precursor tRNAs (pre-tRNAs) is removed, and processed tRNAs acquire the CCA end which is absolutely essential for tRNA aminoacylation and protein synthesis (for reviews, see [1-3]). In

prokaryotes, this process can be either exonucleolytic or endonucleolytic depending on whether the 3'-CCA sequence is genomically encoded. CCA-containing pre-tRNAs are generally processed by the exonucleases that tend to stop removing nucleotides from the 3'-end upon encountering the transcriptionally encoded CCA, whereas CCA-less pre-tRNAs are processed by a 3'-endonuclease termed tRNase Z (also termed RNase Z or 3'-tRNase; for reviews, see [4-7]) that cleaves immediately after the N<sup>73</sup> discriminator nucleotide (the first

\* Correspondence: yhuang@njnu.edu.cn

<sup>1</sup>Laboratory of Yeast Genetics and Molecular Biology, School of Life Sciences, Nanjing Normal University, 1 Wenyuan Road, Nanjing 210046, China  
Full list of author information is available at the end of the article

unpaired base after the acceptor stem) to allow subsequent addition of the CCA sequence.

Unlike prokaryotic pre-tRNAs, eukaryotic nuclear and organellar pre-tRNAs generally lack the 3'-CCA sequence (which is added post-transcriptionally) and their 3'-trailer sequences are removed by tRNase Z. Also unlike prokaryotic pre-tRNAs, eukaryotic nuclear pre-tRNAs contain oligo (U) at their 3'-ends, which are recognized and bound by the La protein (for reviews see [2,8]). In the budding yeast *Saccharomyces cerevisiae* and fission yeast *Schizosaccharomyces pombe*, the endonucleolytic cleavage of nuclear pre-tRNAs requires the presence of the yeast La protein [9,10]. In the absence of the yeast La protein, the 3'-trailer sequence of nuclear pre-tRNAs is trimmed by 3'-exoribonucleases including Rex1p [11]. However, organellar pre-tRNAs lack terminal oligo (U). Furthermore, unlike nuclear pre-tRNAs which are typically monocistronic, most organellar pre-tRNAs are polycistronic [12,13].

tRNase Z is present in all kingdoms of life. It exists in two forms: tRNase Z<sup>S</sup> [300-400 amino acids (aa)] and tRNase Z<sup>L</sup> (700-800 aa), which are encoded by different genes. It is believed that the tRNase Z<sup>L</sup> gene has evolved from a tandem duplication of the tRNase Z<sup>S</sup> gene, followed by divergence of the sequence [14]. In prokaryotes, only tRNase Z<sup>S</sup> is identified. By contrast, all eukaryotes possess tRNase Z<sup>L</sup>, and some have both forms.

The species distribution of tRNase Z is complex. The majority of eukaryotic species analyzed to date, including *S. cerevisiae*, the fruit fly *Drosophila melanogaster* and the nematode worm *Caenorhabditis elegans* contain a single tRNase Z<sup>L</sup> [15-17]. In contrast, *S. pombe* have two tRNase Z<sup>L</sup>s [18,19]. Interestingly, two tRNase Z<sup>L</sup>s and two tRNase Z<sup>S</sup>s have been experimentally identified in the flowering plant *Arabidopsis thaliana* [20]. In humans, one tRNase Z<sup>S</sup> (also termed ELAC1) and one tRNase Z<sup>L</sup> (also termed ELAC2) are found [14]. Our BLAST searches against public genomic and expressed sequence tag (EST) databases reveal that with few exceptions, vertebrates contain one tRNase Z<sup>L</sup> and one tRNase Z<sup>S</sup> (a detailed description of tRNase Z protein distribution in the animal kingdom will be provided elsewhere).

tRNase Z belongs to the metallo- $\beta$ -lactamase (MBL) superfamily [14,21-24]. The typical MBL domain contains five conserved sequence motifs termed Motifs I-V. Motifs I and IV each harbor an invariant Asp, Motif II (HxHxDH), which is also called the His motif, is the signature motif of the superfamily, whereas Motifs III and V each contain a conserved His residue. Structural studies of tRNase Z<sup>S</sup>s from *E. coli*, *T. maritima* and *B. subtilis* [25-28] and mutation analyses of tRNase Zs from a variety of species [29-35] reveal that the His and Asp

residues of Motifs II-V form the active site for coordination of two catalytic zinc ions. In particular, the Asp residue of Motif II may participate in both zinc ion coordination and act as a general base to generate a hydroxide ion for nucleophilic attack on the scissile phosphodiester bond at the cleavage site [25,29]. The Asp residue of Motif I is also catalytically important and appears to stabilize the catalytic site [33].

Besides tRNase Zs, some nucleic acid processing enzymes are also members of the MBL superfamily. Most of these proteins belong to the  $\beta$ -CASP (MBL-associated CISF Artemis SNM1/PSO2) subfamily of the MBL [23]. This subfamily includes the 73-kD subunit of the cleavage and polyadenylation specificity factor (CPSF-73) and its yeast homolog Ysh1p, which are involved in endonucleolytic cleavage of pre-mRNA, the Intergrator complex subunit 11 (Int11) involved in the 3'-end formation of small nuclear RNAs (snRNA) [36], bacterial RNase J, which participates in rRNA 5'-end maturation [37] and RNA decay [38], and the eukaryotic Pso/Snm1/Artemis proteins, which function in DNA repair and V(D)J recombination [23]. However, unlike tRNase Zs,  $\beta$ -CASP proteins contain conserved  $\beta$ -CASP sequence motifs in place of Motif V.

tRNase Z is distinguished from other MBL members by their unique substrate binding domain termed the flexible arm (also termed the exosite). Based on flexible arm type, there are two major types (bacterial- and eukaryotic-types) and one minor type [*T. maritima* (TM)-type] of tRNase Zs [39]. The bacterial-type tRNase Zs, which are present predominantly in bacteria, possess the bacterial-type flexible arm. The bacterial-type flexible arm is ~55 aa in length and contains the Gly- and Pro-rich GP motif (GxPxGP, sometimes GxPPGP) [39]. The eukaryotic-type tRNase Zs, which are found only in eukaryotes, contain the ELAC2-type flexible arm. This type of flexible arm harbors the GP motif and is ~62 aa long, which is slightly longer than the bacterial-type flexible arm.

The TM-type tRNase Z was believed to be the minor type at the time of discovery since it was found only in *T. maritima* and *A. thaliana* [39]. The flexible arm found in TM-type tRNase Zs appears to be shorter (~30 aa) and lacks the GP motif but instead contains one short basic residue-rich region [39]. In addition, both the bacterial- and eukaryotic-type tRNase Zs contain the PxKxRN, HEAT and HST motifs, which form part of loop structures, whereas the TM-type tRNase Z lacks these motifs [33,40,41]. The PxKxRN motif has been suggested to function in CCA anti-determination (tRNase Z activity is inhibited by 3'-CCA) [25,33], whereas the HEAT and HST motifs have been suggested to play a role in facilitating proton transfer at the final stage of reaction [25,29,40].

tRNase Z has diverse functions besides its primary role in tRNA 3'-end processing. This is perhaps best exemplified by ELAC2, which serves a multitude of functions within cells. Recent studies have shown that ELAC2 is involved in the generation of MALAT1, a cancer-associated long noncoding RNA which participates in regulation of pre-mRNA splicing [42], tRNA-derived small RNAs [43,44], and viral microRNAs (miRNAs) [45,46]. Overexpression of *ELAC2* delays cell cycle progression, suggesting that ELAC2 may be involved in cell cycle control either directly or indirectly via its role as tRNA processing enzyme [47]. ELAC2 also potentiates TGF- $\beta$  (transforming growth factor- $\beta$ /Smad-induced transcription response, indicating a role for ELAC2 in TGF- $\beta$ /Smad signaling mediated growth arrest [48]. Interestingly, a recent study has shown that destruction of human mitochondria through depletion of mitochondrial DNA results in down-regulation of *ELAC2* and a delay in cell cycle progression [49]. Since ELAC2 may be involved in cell cycle regulation, it is likely that ELAC2 may link mitochondrial function and cell cycle control. It is important to note that *ELAC2* is a candidate prostate cancer susceptibility gene as its mutations are associated with prostate cancer [14]. However, the underlying mechanisms are unknown. In *S. cerevisiae*, either inactivating mutations or overexpression of tRNase Z<sup>L</sup> causes a petite phenotype, suggesting that the action of tRNase Z<sup>L</sup> may be related to mitochondrial function [15]. In addition, the *S. cerevisiae* tRNase Z<sup>L</sup> has also been suggested to play a role in 35S rRNA processing [50].

The study of tRNase Z evolution has been facilitated by the increasing availability of genome sequences. A previous study showed that only tRNase Z<sup>S</sup> is found in bacteria and that its presence in bacteria is widespread [6]. We recently reported on a systematic survey of tRNase Zs in fungi [51]. Our analysis reveals that while the majority of fungal species contain one tRNase Z<sup>L</sup>, all four sequenced *Schizosaccharomyces* species contain two distinct tRNase Z<sup>L</sup>s either demonstrated or predicted to be localized to the nucleus and mitochondria, respectively. In addition, the presence of tRNase Z<sup>S</sup> in fungi is restricted to the phylum Basidiomycota and the basal fungal phyla.

Green plants (Viridiplantae) represent a monophyletic group of land plants and green algae that evolved near the base of the tree of eukaryotic life. Flowering plants (angiosperms), which are typically polyploidy, represent the largest, most diverse and most evolutionary advanced phylum of land plants making up 90% of the plant kingdom. It can be divided into two major groups: dicotyledons (dicots), which accounts for the majority of the angiosperm species, and monocotyledons (monocots). At present, there are at least 27 sequenced and

annotated genomes representing the major taxonomic groups within green plants, although the majority of them are those of flowering plants. The public availability of these genome sequences enabled us to identify tRNase Zs in green plants and to study their evolution.

In this study, we undertook a comprehensive survey of candidate tRNase Zs from annotated green plant genomes. To understand the evolutionary relationships among green plant tRNase Zs, we further conducted a phylogenetic analysis of these newly identified candidates. Finally, we presented a detailed sequence analysis of tRNase Zs with the intent of further delineating the distinct features of green plant tRNase Zs.

## Results

### Identification of candidate green plant tRNase Zs

To extend our previous study of tRNase Z diversity and evolution, we searched public genome databases for putative green plant tRNase Zs with significant matches to known bacterial and eukaryotic tRNase Zs. Since most of candidate sequences identified from the databases are computationally generated without subsequent manual annotation, it is likely that many predictions may contain errors. Therefore, we verified each candidate. We first validated each prediction by reciprocal searches against the GenBank. In back-searches, a candidate was confirmed if reverse BLAST also gave tRNase Z hits in the top matches. Accuracy of prediction was further evaluated by multiple sequence alignment. All discordant candidate sequences were checked manually for possible errors including sequencing errors, intron mispredictions and existence of gaps in the genome sequences. We found that many candidate sequences are apparently incomplete or contain annotation errors. For example, the predicted coding sequence of the monkeyflower (*Mimulus guttatus*) MguTRZ2 (Phytozome accession no. mgv1a024577 m.g) in the database was incomplete lacking the N-terminal region. We were able to predict this region from the genomic DNA sequence based on sequence similarity. The predicted full-length coding region of MguTRZ2 has 364 aa. As another example, the sequence annotated as the candidate castor bean (*Ricinus communis*) tRNase Z<sup>L</sup> (Phytozome accession no. 30146.t000117) appears to be mispredicted due to the presence of sequence gaps. Thus, this sequence was excluded from the list.

Several incorrect predictions are apparently caused by the presence of the non-canonical GC-AG splice site pairs. While the GT donor splice site is a canonical 5'-splice site for introns in eukaryotic genes, the GC donor splice sites account for the majority of the non-canonical donor splice site. Thus, for those that cannot be accurately predicted by the conventional FGENESH program, we carried out gene prediction using FGENESH

\_GC, which is a new version of the FGENESH program including noncanonical GC dinucleotide in donor splice sites. Indeed, by doing so, we could predict some exons encoding missing conserved motifs. For example, the HEAT motif was originally missing in several annotated candidate tRNase Z<sup>L</sup>s from flowering plants including papaya (*Carica papaya* CpaTRZ3), cassava (*Manihot esculenta* MesTRZ4), *Medicago* (*Medicago truncatula* MtrTRZ3) and black cottonwood (*Populus trichocarpa* PtrTRZ3). After re-evaluation of intron splice sites using FGENESH\_GC, we were able to recover their HEAT motifs. Because the intron sequences of some candidate sequences appear to be extremely difficult to predict correctly, we could not conclusively rule out the possibility of errors in certain candidates.

In total, 54 candidate tRNase Z<sup>S</sup> and 32 candidate tRNase Z<sup>L</sup> were identified from 27 green plant species including 21 flowering plants, 1 moss, 1 lycophyte and 4 green algae (Additional file 1). The names of these candidates follow the *A. thaliana* tRNase Z nomenclature [52]. Of these, only tRNase Zs from *A. thaliana* have been experimentally characterized [20]. While most of these sequenced genomes examined belong to the flowering plants, the availability of the genome sequences from two basal land plants (the bryophyte *Physcomitrella patens* and the lycophyte *Selaginella moellendorffii*) and four green algae (*Chlamydomonas reinhardtii*, *Volvox carteri*, *Micromonas pusilla* and *Ostreococcus lucimarinus*) allows for evaluation of differences between flowering plant tRNase Zs and those from basal land plants and green algae.

The flowering plants examined to date appear to contain multiple tRNase Zs. Foxtail millet (*Setaria italica*) contains the largest number of tRNase Zs (5) so far identified in a flowering plant. Most flowering plant species have a single tRNase Z<sup>L</sup>. In contrast, six flowering plants including two *Arabidopsis* species (*A. thaliana* and *Arabidopsis lyrata*) harbor two tRNase Z<sup>L</sup>s. It should be noted that the presence of two tRNase Z<sup>L</sup>s are not species-specific since the plant species containing two tRNase Z<sup>L</sup>s come from diverse taxonomic groups.

Unlike most flowering plants that possess two tRNase Z<sup>S</sup>s, two members of the *Panicoideae* subfamily of grasses, sorghum (*Sorghum bicolor*) and foxtail millet have three tRNase Z<sup>S</sup>s and four tRNase Z<sup>S</sup>s, respectively. The presence of multiple tRNase Z<sup>S</sup>s appears not to be grass-specific, since three other grass plants including rice (*Oryza sativa*), *Brachypodium* (*Brachypodium distachyon*) and maize (*Zea mays*) contain only two tRNase Z<sup>S</sup>s. Calculation of the percentage identity and similarity between candidate tRNase Zs from these two grass species shows strong conservation of the proteins at the amino acid level with the most identity (93%) and

similarity (95%) between *S. bicolor* SbiTRZ2 and *S. italica* SitTRZ2 (Additional files 2 and 3). To our surprise, all candidate tRNase Z<sup>S</sup>s identified in flowering plants are highly similar and belong to the TM-type tRNase Z<sup>S</sup> (see below for a detailed discussion).

The number and type of tRNase Zs appear to be highly variable in the two primitive plant species and four green algae (Table 1 and Additional file 1). The largest number of tRNase Zs is found in the moss *P. patens*, which contains two TM-type and one bacterial-type tRNase Z<sup>S</sup>s, and one tRNase Z<sup>L</sup>. In contrast, two green algae *C. reinhardtii* and *V. carteria* have the least number of tRNase Zs, comprising one TM-type tRNase Z<sup>S</sup> and one tRNase Z<sup>L</sup>. The two other green algae *M. pusilla* and *O. lucimarinus* contain one TM-type and one bacterial-type tRNase Z<sup>S</sup>s and one tRNase Z<sup>L</sup>. The lycophyte *S. moellendorffii*, which has the smallest genome size of any land plant reported, contains two tRNase Z<sup>L</sup>s in addition to one TM-type tRNase Z<sup>S</sup>.

Interestingly, tRNase Z-like proteins (TLP) which apparently lack one or more conserved motifs of tRNase Zs necessary for the enzymatic activity of the protein are widespread in basal land plant and green algal genomes that have been analyzed here (Table 1 and Additional file 1). Many tRNase Z-like proteins appear to lack the flexible arm but contain all other conserved motifs of tRNase Zs (Figure 1 and see below for a more detailed discussion of the motifs). At one extreme, one *S. moellendorffii* tRNase Z-like protein (SmoTLP2) appears to lack all of the conserved motifs.

Based on their sizes, tRNase Z-like proteins can be divided into tRNase Z<sup>S</sup>-like and tRNase Z<sup>L</sup>-like proteins which are comparable in size to tRNase Z<sup>S</sup>s and tRNase Z<sup>L</sup>s, respectively. The number and form of tRNase Z-like proteins vary among the species, being largest in *C. reinhardtii* (two tRNase Z<sup>S</sup>-like and two tRNase Z<sup>L</sup>-like proteins) and next largest in *S. moellendorffii* (two tRNase Z<sup>S</sup>-like and one tRNase Z<sup>L</sup>-like proteins). The basal land plant *P. patens* and the two green algae *M. pusilla* and *O. lucimarinus* contain one tRNase Z<sup>L</sup>-like protein, whereas the green alga *V. carteria* contains one tRNase Z<sup>S</sup>-like and one tRNase Z<sup>L</sup>-like proteins. In contrast, black cottonwood (*P. trichocarpa*) appears to be the only species among the flowering plant genomes examined that contains the tRNase Z<sup>L</sup>-like protein. This species contains one tRNase Z<sup>L</sup>-like protein in which the second His in the His motif is mutated to Gln (data not shown).

#### Prediction of subcellular localization of candidate tRNase Zs from flowering plants

To help understand the function of candidate tRNase Zs, we predicted *in silico* the subcellular localization of each of flowering plant tRNase Zs using different



**Table 1 Distribution of candidate tRNase Zs from representative green plants**

Species	Protein name	Type	Accession number	Database	No. aa <sup>+</sup>
<b>DICOTS</b>					
<i>Aquilegia coerulea</i>	AcoTRZ1	TM-type tRNase Z <sup>S</sup>	AcoGoldSmith_v1.019418 m.g	Phytozome	283
<i>Aquilegia coerulea</i>	AcoTRZ2	TM-type tRNase Z <sup>S</sup>	AcoGoldSmith_v1.007495 m.g	Phytozome	356
<i>Aquilegia coerulea</i>	AcoTRZ3	tRNase Z <sup>L</sup>	AcoGoldSmith_v1.000749 m.g	Phytozome	997*
<i>Aquilegia coerulea</i>	AcoTRZ4	tRNase Z <sup>L</sup>	AcoGoldSmith_v1.024294 m.g	Phytozome	848*
<i>Arabidopsis thaliana</i>	AthTRZ1	TM-type tRNase Z <sup>S</sup>	NP_177608.2	NCBI	280
<i>Arabidopsis thaliana</i>	AthTRZ2	TM-type tRNase Z <sup>S</sup>	NP_178532.2	NCBI	354
<i>Arabidopsis thaliana</i>	AthTRZ3	tRNase Z <sup>L</sup>	NP_175628.2	NCBI	890
<i>Arabidopsis thaliana</i>	AthTRZ4	tRNase Z <sup>L</sup>	NP_188247.2	NCBI	942
<i>Carica papaya</i>	CpaTRZ1	TM-type tRNase Z <sup>S</sup>	evm.TU.supercontig_17.217	Phytozome	306
<i>Carica papaya</i>	CpaTRZ2	TM-type tRNase Z <sup>S</sup>	evm.TU.supercontig_132.51	Phytozome	359*
<i>Carica papaya</i>	CpaTRZ3	tRNase Z <sup>L</sup>	evm.TU.supercontig_34.31	Phytozome	943*
<i>Citrus clementina</i>	CclTRZ1	TM-type tRNase Z <sup>S</sup>	clementine0.9_017657 m.g	Phytozome	288
<i>Citrus clementina</i>	CclTRZ2	TM-type tRNase Z <sup>S</sup>	clementine0.9_014182 m.g	Phytozome	355
<i>Citrus clementina</i>	CclTRZ3	tRNase Z <sup>L</sup>	clementine0.9_001700 m.g	Phytozome	938
<i>Cucumis sativus</i>	CsaTRZ1	TM-type tRNase Z <sup>S</sup>	Cucsa.369470	Phytozome	301*
<i>Cucumis sativus</i>	CsaTRZ2	TM-type tRNase Z <sup>S</sup>	Cucsa.185180	Phytozome	347
<i>Cucumis sativus</i>	CsaTRZ3	tRNase Z <sup>L</sup>	Cucsa.359360	Phytozome	977
<i>Eucalyptus grandis</i>	EgrTRZ1	TM-type tRNase Z <sup>S</sup>	Egrandis_v1_0.020509 m.g	Phytozome	309
<i>Eucalyptus grandis</i>	EgrTRZ2	TM-type tRNase Z <sup>S</sup>	Egrandis_v1_0.021229 m.g	Phytozome	357*
<i>Eucalyptus grandis</i>	EgrTRZ3	tRNase Z <sup>L</sup>	Egrandis_v1_0.001658 m.g	Phytozome	971
<i>Glycine max</i>	GmaTRZ1	TM-type tRNase Z <sup>S</sup>	Glyma20g01480	Phytozome	279
<i>Glycine max</i>	GmaTRZ2	TM-type tRNase Z <sup>S</sup>	ACU18735.1	NCBI	353
<i>Glycine max</i>	GmaTRZ3	tRNase Z <sup>L</sup>	Glyma13g43270	Phytozome	923*
<i>Glycine max</i>	GmaTRZ4	tRNase Z <sup>L</sup>	Glyma15g02070	Phytozome	931*
<i>Mimulus guttatus</i>	MguTRZ1	TM-type tRNase Z <sup>S</sup>	mgv1a010642 m.g	Phytozome	306
<i>Mimulus guttatus</i>	MguTRZ2	TM-type tRNase Z <sup>S</sup>	mgv1a024577 m.g	Phytozome	364*
<i>Mimulus guttatus</i>	MguTRZ3	tRNase Z <sup>L</sup>	mgv1a000815 m.g	Phytozome	976
<i>Vitis vinifera</i>	VviTRZ1	TM-type tRNase Z <sup>S</sup>	XP_002273058.1	NCBI	290
<i>Vitis vinifera</i>	VviTRZ2	TM-type tRNase Z <sup>S</sup>	XP_002277241.1	NCBI	354
<i>Vitis vinifera</i>	VviTRZ3	tRNase Z <sup>L</sup>	XP_002278956.1	NCBI	951
<b>MONOCOTS</b>					
<i>Oryza sativa japonica</i>	OsaTRZ1	TM-type tRNase Z <sup>S</sup>	NP_001046280.1	NCBI	302
<i>Oryza sativa japonica</i>	OsaTRZ2	TM-type tRNase Z <sup>S</sup>	LOC_Os09g30466	Phytozome	365
<i>Oryza sativa japonica</i>	OsaTRZ3	tRNase Z <sup>L</sup>	B9EUI7	UniProt	964
<i>Zea mays</i>	ZmaTRZ1	TM-type tRNase Z <sup>S</sup>	GRMZM2G147727	Phytozome	302
<i>Zea mays</i>	ZmaTRZ2	TM-type tRNase Z <sup>S</sup>	ACG47847.1	NCBI	357
<i>Zea mays</i>	ZmaTRZ3	tRNase Z <sup>L</sup>	GRMZM2G379286	Phytozome	930
<b>MOSESSES</b>					
<i>Physcomitrella patens</i>	PpaTRZ1	TM-type tRNase Z <sup>S</sup>	XP_001785449.1	NCBI	295
<i>Physcomitrella patens</i>	PpaTRZ2	TM-type tRNase Z <sup>S</sup>	Pp1s17_61V6	Phytozome	420
<i>Physcomitrella patens</i>	PpaTRZ3	Bacterial-type tRNase Z <sup>S</sup>	Pp1s409_46V6	Phytozome	336*
<i>Physcomitrella patens</i>	PpaTRZ4	tRNase Z <sup>L</sup>	Pp1s126_39V6	Phytozome	984*
<i>Physcomitrella patens</i>	PpaTLP1	tRNase Z <sup>L</sup> -like	Pp1s3_568V6	Phytozome	923*
<b>LYCOPHYTES</b>					
<i>Selaginella moellendorffii</i>	SmoTRZ1	TM-type tRNase Z <sup>S</sup>	167717	Phytozome	299
<i>Selaginella moellendorffii</i>	SmoTRZ2	tRNase Z <sup>L</sup>	111409	Phytozome	771*
<i>Selaginella moellendorffii</i>	SmoTRZ3	tRNase Z <sup>L</sup>	97243	Phytozome	819*
<i>Selaginella moellendorffii</i>	SmoTLP1	tRNase Z <sup>S</sup> -like	404484	Phytozome	303
<i>Selaginella moellendorffii</i>	SmoTLP2	tRNase-like	XP_002967989	NCBI	346
<i>Selaginella moellendorffii</i>	SmoTLP3	tRNase Z <sup>L</sup> -like	416540	Phytozome	731

**Table 1 Distribution of candidate tRNase Zs from representative green plants (Continued)**

GREEN ALGAE					
<i>Chlamydomonas reinhardtii</i>	CreTRZ1	TM-type tRNase Z <sup>S</sup>	Cre10.g431400	Phytozome	335
<i>Chlamydomonas reinhardtii</i>	CreTRZ2	tRNase Z <sup>L</sup>	Cre01.g039900	Phytozome	880*
<i>Chlamydomonas reinhardtii</i>	CreTLP1	tRNase Z <sup>S</sup> -like	Cre12.g492250	Phytozome	425*
<i>Chlamydomonas reinhardtii</i>	CreTLP2	tRNase Z <sup>S</sup> -like	Cre07.g321900	Phytozome	378
<i>Chlamydomonas reinhardtii</i>	CreTLP3	tRNase Z <sup>L</sup> -like	Cre01.g068350	Phytozome	710*
<i>Chlamydomonas reinhardtii</i>	CreTLP4	tRNase Z <sup>L</sup> -like	Cre09.g394700	Phytozome	1073
<i>Micromonas pusilla</i>	MpuTRZ1	TM-type tRNase Z <sup>S</sup>	5692	JGI	383*
<i>Micromonas pusilla</i>	MpuTRZ2	Bacterial-type tRNase Z <sup>S</sup>	18700	JGI	398*
<i>Micromonas pusilla</i>	MpuTRZ3	tRNase Z <sup>L</sup>	54013	JGI	953*
<i>Micromonas pusilla</i>	MpuTLP1	tRNase Z <sup>L</sup> -like	C1N4P0	UniProt	866
<i>Ostreococcus lucimarinus</i>	OluTRZ1	TM-type tRNase Z <sup>S</sup>	26345	JGI	313
<i>Ostreococcus lucimarinus</i>	OluTRZ2	Bacterial-type tRNase Z <sup>S</sup>	XP_001416263	NCBI	369*
<i>Ostreococcus lucimarinus</i>	OluTRZ3	tRNase Z <sup>L</sup>	A4S9E8	UniProt	781*
<i>Ostreococcus lucimarinus</i>	OluTLP1	tRNase Z <sup>L</sup> -like	A4RR41	UniProt	780
<i>Volvox carteri</i>	VcaTRZ1	TM-type tRNase Z <sup>S</sup>	80918	Phytozome	308*
<i>Volvox carteri</i>	VcaTRZ2	tRNase Z <sup>L</sup>	118548	Phytozome	881*
<i>Volvox carteri</i>	VcaTLP1	tRNase Z <sup>S</sup> -like	103834	Phytozome	445*
<i>Volvox carteri</i>	VcaTLP2	tRNase Z <sup>L</sup> -like	100314	Phytozome	723*

\*The number of amino acids in plant tRNase Z and tRNase Z-like proteins.

\*Indicates that mispredicted sequences obtained from the databases have been corrected.

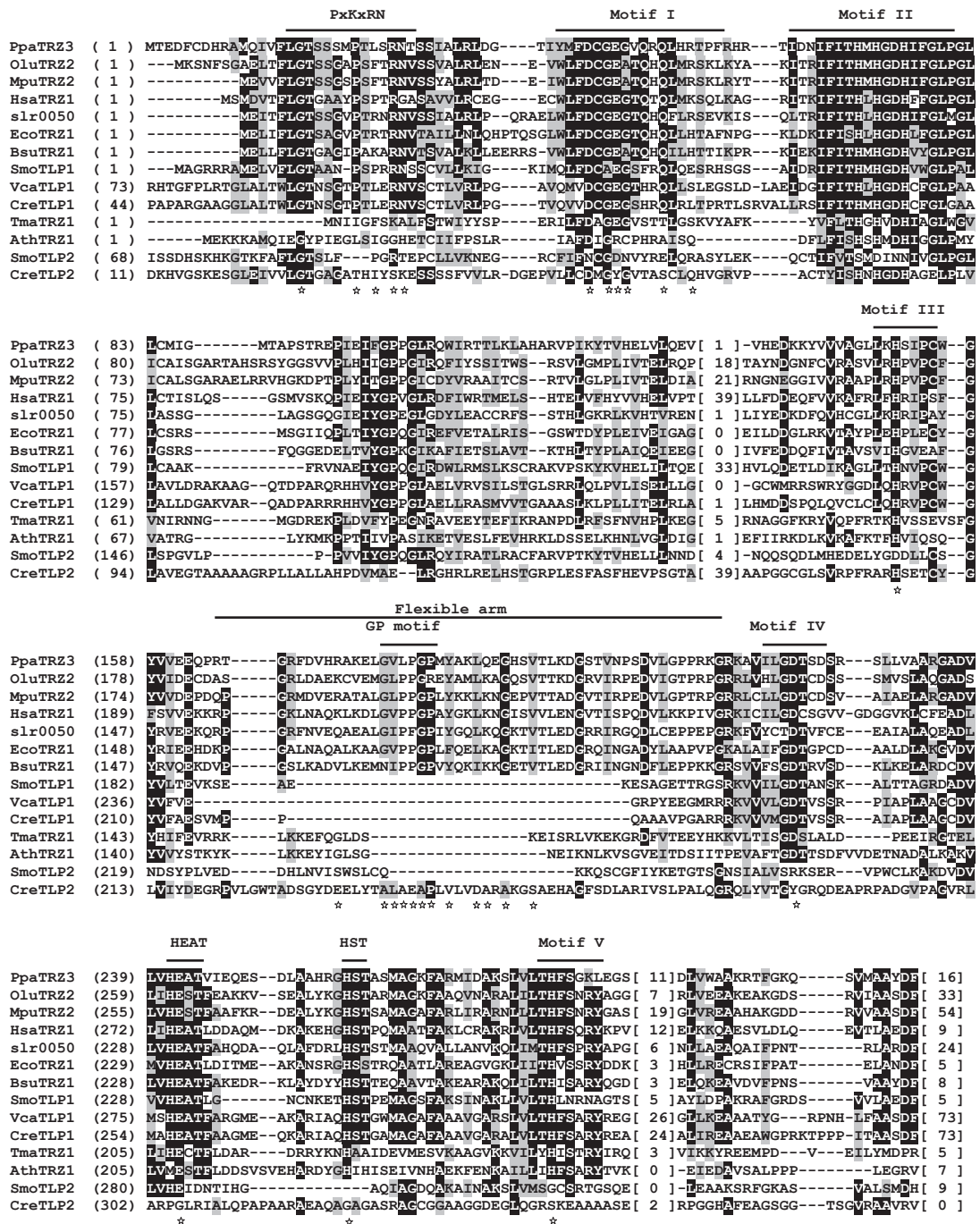
bioinformatic prediction programs. Most flowering plants have two tRNase Z<sup>S</sup>s (tRNase Z<sup>S1</sup> and tRNase Z<sup>S2</sup>). All tRNase Z<sup>S1</sup> proteins apparently lack any predictable signal sequences and therefore predicted to be cytoplasmic proteins, with the exception of *O. sativa* tRNase Z<sup>S1</sup> (OsaTRZ1), which is predicted to be localized in the chloroplasts (Table 2). On the other hand, all tRNase Z<sup>S2</sup> except *P. trichocarpa* tRNase Z<sup>S2</sup> (PtrTRZ2), contain a predicted chloroplast targeting signal (Table 2). The predicted chloroplast targeting signal of PtrTRZ2 is unusual in that it is only seven residues long since the majority of the known chloroplast targeting signals consist of N-terminal 20-80 amino acid residues [53]. In the two flowering plants (*S. bicolor* and *S. italica*) carrying multiple tRNase Z<sup>S</sup>s, only one tRNase Z<sup>S</sup> (tRNase Z<sup>S2</sup>) from each species is predicted to be chloroplast-localized, whereas the rest are predicted to be cytoplasmic (Table 2).

Besides tRNase Z<sup>S</sup>, flowering plants also contain one (tRNase Z<sup>L1</sup>) or two (tRNase Z<sup>L1</sup> and tRNase Z<sup>L2</sup>) tRNase Z<sup>L</sup>s. Most tRNase Z<sup>L1</sup> proteins contain a predicted mitochondrial targeting signal between two putative translational initiation sites at the N-terminus and also have a nuclear localization signal (Table 3). Some tRNase Z<sup>L2</sup> proteins have a putative mitochondrial targeting signal, while others are predicted to have both nuclear and mitochondrial targeting signals. The length of the putative mitochondrial targeting signals found in these candidates are within the expected size-range (20-

80 aa) [54]. Consistent with the prediction, *A. thaliana* tRNase Z<sup>L1</sup> (AthTRZ3) was found in both the nucleus and the mitochondria [20]. However, *A. thaliana* tRNase Z<sup>L2</sup> (AthTRZ4) predicted to have both nuclear and mitochondrial targeting signals is actually found only in the mitochondria [20].

#### Phylogenetic analysis

To gain insights into the evolutionary relationship among plant tRNase Zs, we reconstructed the phylogeny of 86 candidate sequences using Bayesian phylogenetics. Although most of plant species are flowering plants, they are taxonomically diverse. In addition to tRNase Zs from plants, we also include tRNase Z<sup>S</sup> from *T. maritima* as outgroup. Phylogenetic analysis reveals the presences of two well-supported clades: one formed by including all TM-type tRNase Z<sup>S</sup>s, and the other all tRNase Z<sup>L</sup>s and bacterial-type tRNase Z<sup>S</sup>s (Figure 2). The latter clade is further divided into two subclades, of which one contains all tRNase Z<sup>L</sup>s, and the other bacterial-type tRNase Z<sup>S</sup>s. Notably, two tRNase Z<sup>L</sup>s (AlyTRZ4 and AthTRZ4) from the two *Arabidopsis* species (*A. thaliana* and *A. lyrata*) together form a group sister to a group formed by another two tRNase Z<sup>L</sup>s (AlyTRZ3 and AthTRZ3) from the same two species. Likewise, the two tRNase Z<sup>L</sup>s found in some species including *Aquilegia coerulea* (AcoTRZ3 and AcoTRZ4), *Glycine max* (GmaTRZ3 and GmaTRZ4) and *S. moellendorffii* (SmoTRZ2 and SmoTRZ3) are sister to each



**Figure 1** Alignment of bacterial-type and tRNase Z<sup>5</sup>-like candidates in green plants. Bacterial-type tRNase Z<sup>5</sup>s are from *P. patens* (PpaTRZ3), *O. lucimarinus* (OluTRZ2), *M. pusilla* (MpuTRZ2), *Synechocystis* sp. PCC 6803 (slr0050), *E. coli* (EcoTrz1) [72], *B. subtilis* (BsuTrz1) [60] and humans (HsaTrz1) [14]. Plant tRNase Z<sup>5</sup>-like (TLP) candidates are from *S. moellendorffii* (SmoTLP1 and SmoTLP2), *V. carteri* (VcaTLP1), *C. reinhardtii* (CreTLP1 and CreTLP2). TM-type tRNase Z<sup>5</sup>s from *T. maritima* (TmaTrz1) [30] and *A. thaliana* (AthTrz1) [20] are included for comparison. The alignment was constructed using Clustal W [68]. Identical residues are on a black ground and conserved residues shared in gray. Also indicated above the alignment are the conserved motifs of tRNase Zs, which are labeled according to references [33,40,41]. The numbers in brackets indicate the length of the region in the protein, which are species-specific and could not be correctly aligned. Hyphens represent gaps introduced into sequences for maximum alignment. Amino acid residues predicted to be critical for activity are indicated by a star.

**Table 2 Prediction of the chloroplast targeting signals in candidate TM-type tRNase Z<sup>S</sup> from flowering plants**

Protein name	Form	Chl	Chloroplast targeting signal
AcoTrz1	AcoTrz <sup>S1</sup>	-	
AcoTrz2	AcoTrz <sup>S2</sup>	C	MKISLGISTSPILLPFSKPTLKIPIHNRVCIP (31)
AlyTrz1	AlyTrz <sup>S1</sup>	-	
AlyTrz2	AlyTrz <sup>S2</sup>	C	MQLSSFPIPSKIFPFKYHAPPVHIIQLAAQIQSHRSNFVSPVKVSGYFSSISRAIEEEEEYRKAR (67)
AthTrz1	AthTrz <sup>S1</sup>	-	
AthTrz2	AthTrz <sup>S2</sup>	C	MQLSSFPISPPKIFPSTKHHKPPVITHQLAAQIQSNRRHFVSPVKVSGYFSSISRAIEEEEEYRKAR (68)
BdiTrz1	BdiTrz <sup>S1</sup>	-	
BdiTrz2	BdiTrz <sup>S2</sup>	C	MATISLFLPLRLTRGLLSPSSGPASRFQTL (32)
CpaTrz1	CpaTrz <sup>S1</sup>	-	
CpaTrz2	CpaTrz <sup>S2</sup>	C	MLTSVPILPSKLSQSVFPFGHHHYHSISHKPTPIHQFSFLV (40)
CclTrz1	CclTrz <sup>S1</sup>	-	
CclTrz2	CclTrz <sup>S2</sup>	C	MQLSLPTSPSKLPTIFPFHPSSIPKTPQSPHLSLQSHVGPLNALKSAGFLSSISRAIDEEEEYRKAR (68)
CsiTrz1	CsiTrz <sup>S1</sup>	-	
CsiTrz2	CsiTrz <sup>S2</sup>	C	MQLSLPTSPSKLPTIFPFHPSSIPKTPQSPHLSLQSHVGPLNALKSAGFLSSISRAIDEEEEYRKAR (68)
CsaTrz1	CsaTrz <sup>S1</sup>	-	
CsaTrz2	CsaTrz <sup>S2</sup>	C	MQISIPISPLRPPQVFPFHQPLLHTPKPPGVALQSHLNPVNSFRDSSGLLSTIGVE (55)
EgrTrz1	EgrTrz <sup>S1</sup>	-	
EgrTrz2	EgrTrz <sup>S2</sup>	C	MRISIPISPSKSPQIFPFHHHRRRHQIPDLRRRPPALSLPASAVNPLKSSGYLSTIHR (59)
GmaTrz1	GmaTrz <sup>S1</sup>	-	
GmaTrz2	GmaTrz <sup>S2</sup>	C	MQISLSDLAFKTPQLFPIHHPIFPKPPLNHQVST (35)
MesTrz1	MesTrz <sup>S1</sup>	-	
MesTrz2	MesTrz <sup>S2</sup>	C	MQTSFPVSPSKIPSIFPNHPIHNPAPTNNHQLPLQTYLKRPIINSTSLKSSGFLSAIGRAIEEEEEYKAR (73)
MtrTrz1	MtrTrz <sup>S1</sup>	-	
MtrTrz2	MtrTrz <sup>S2</sup>	C	MQIPLTNPFTKPPQIFPFHHTIPSPKQQLSFPT (33)
MguTrz1	MguTrz <sup>S1</sup>	-	
MguTrz2	MguTrz <sup>S2</sup>	C	MITLNPTSNLLNLHPHPTPATHKHHLPPQTRRSVIACVSDAW (45)
OsaTrz1	OsaTrz <sup>S1</sup>	C	MANSKGSSPAATSTTAPPGRPKAKAPPLTVEGYVVEGISIGGQETCVIFPT (52)
OsaTrz2	OsaTrz <sup>S2</sup>	C	MAATSLLSLPSRLRHLVPASSAPASRSQFQTL (36)
PtrTrz1	PtrTrz <sup>S1</sup>	-	
PtrTrz2	PtrTrz <sup>S2</sup>	C	MQISIPL (7)
PpeTrz1	PpeTrz <sup>S1</sup>	-	
PpeTrz2	PpeTrz <sup>S2</sup>	C	MQIFLTISPCKAPLILPFHPPISKTPKTQRTSLKTLSPYRLARSQVLRKGV (52)
RcoTrz1	RcoTrz <sup>S1</sup>	-	
RcoTrz2	RcoTrz <sup>S2</sup>	C	MQTSLPISHSKFPSIFPNHPIHSHKPTTTR (30)
SitTrz1	SitTrz <sup>S1</sup>	-	
SitTrz2	SitTrz <sup>S2</sup>	C	MATASLFSPPSLRLLSRTTARLSRFQTLAARKPP (34)
SitTrz3	SitTrz <sup>S3</sup>	-	
SitTrz4	SitTrz <sup>S4</sup>	-	
SbiTrz1	SbiTrz <sup>S1</sup>	-	
SbiTrz2	SbiTrz <sup>S2</sup>	C	MATASLFLPSLRALSRTSARCSRFQTLAARKPVESSTSTATS (43)
SbiTrz3	SbiTrz <sup>S3</sup>	-	
VviTrz1	VviTrz <sup>S1</sup>	-	
VviTrz2	VviTrz <sup>S2</sup>	C	MQISLPFSTSKVPYLSPLNPPTQPPLTIKPKPKSYIT (39)
ZmaTrz1	ZmaTrz <sup>S1</sup>	-	
ZmaTrz2	ZmaTrz <sup>S2</sup>	C	MVTASLFLPSLRVLSRTSAHGPRFQILAARKPVESSTSTATSGSRRGGSKGAGLLSVLDR (59)

Identification of the putative chloroplast targeting signals was made by using ChloroP <http://www.cbs.dtu.dk/services/ChloroP/>. Chl stands for the chloroplast. C indicates chloroplast localization. "-", the localization could not be predicted. The numbers refer to amino acid positions starting from the N-terminus.



**Table 3 Prediction of the nuclear and mitochondrial targeting sequences in tRNase Z<sup>L</sup>s from flowering plants**

Protein name	From	Nuc	Mito	Predicted mitochondrial targeting sequences
AcoTrz3	AcoTrz <sup>L1</sup>	N	-	
AcoTrz4	AcoTrz <sup>L2</sup>	-	M	MPQISSNLKLFSSKTNHSPFLQFQSFKASFCSSFLSSKTPYKPLSSVTVISSSSSSSSSSSRKGPFPPLRSRS (73)
AlyTrz3	AlyTrz <sup>L1</sup>	N	M	MINSMPYLHKNLRLRLSSKSSPFLSLRPFSPRSFSLSTLFS (44)
AlyTrz4	AlyTrz <sup>L2</sup>	N	M	MLTSSMPHRHVQNLSLGFGSPLKSSSFALFLRPFSLYPIFASSSPSPSRPRRTAGYRRS (62)
AthTrz3	AthTrz <sup>L1</sup>	N	M	MINSMPYLHKNLRLRLSSKSSPFLSLRPFSPRSFSLSTLFS (44)
AthTrz4	AthTrz <sup>L2</sup>	N	M	MLTSSMPQNLSLGFGSPLKSSSFALILRPFSLYPIFASSSPAPSRPRRTAGYRRSGPSPRRK (65)
BdiTrz3	BdiTrz <sup>L1</sup>	-	M	MPQVAAPLRLRLSQTALAPPALLHLSRRLSFCSPASFRRAASLRALAYRRSRHPEPRRG (62)
CpaTrz3	CpaTrz <sup>L1</sup>	N	M	MFLIYPNLRLLLNPLFLFSKPNSTPLSLFTVFASSSHKRHRVSRYRDSFGLHRRRRNFTTFKERD (67)
CclTrz3	CclTrz <sup>L1</sup>	N	M	MPFITPNLRLLFSSSSSLFPLKLSVPLLSTKPTNRHRSLFTILSYSKRQRSTPFPQQNQRRN (63)
CsiTrz3	CsiTrz <sup>L1</sup>	N	M	MPFITPNLRLLFSSSSSLFPLKLSVPLLSTKPTNRHLSLFTILSYSKRQRSTPFPQQNQRRN(64)
CsaTrz3	CsaTrz <sup>L1</sup>	N	M	MPLPHLSTLRFLLFSPSKLFPSPSLYSPKSHSLFTVLASSPPKRRRSATAPPSLNFKRRN (60)
EgrTrz3	EgrTrz <sup>L1</sup>	N	M	MPCVYNLRLLFSSSATAAAATATAAASPFLSPLKLRPSSSSSAFPLLPPLSSLSRS (61)
GmaTrz3	GmaTrz <sup>L1</sup>	N	M	MAQVSKFGYLLHSSLPKPSNIQFRSLLTVLASSSKRHRK (41)
GmaTrz4	GmaTrz <sup>L2</sup>	N	M	MAQVSKFGHLLHSSLPKPSNIQFRSLLTVLASSSKRHRHSIPPFRRKS (50)
MesTrz3	MesTrz <sup>L1</sup>	N	M	MPQISNLRLLSPLKPSLPPFSPKPKPYSLFTVLCSSSSSRHRHTTPNHQSLNFRSRS (58)
MesTrz4	MesTrz <sup>L2</sup>	-	M	MPNVLNFKCLSDLLTQCCKICLQHYILTFSSLSHVLVSGKLQSSNPQTYFSINNPLPRSKNSRT (65)
MtrTrz3	MtrTrz <sup>L1</sup>	N	M	MAQILNFRNFLFLPSYKPTTHFRFLSTLVSSSSRRSNINAPPLHLRRRS (51)
MguTrz3	MguTrz <sup>L1</sup>	N	M	MPQSTNLRLLSSANCHRRHPFAASNFPPKHLFSSSSQFLLKPKQFKTREIPLLFATFSSYSKPYATNNNSNNNNKNSRSFNRN (86)
OsaTrz3	OsaTrz <sup>L1</sup>	-	M	MPQLPSPLRRLPLSQTALAAATPAPLLHLSRRLFSSSSSPSPSPRAACLALAYRGGQAGGGRRGHNNLLRRG (77)
PtrTrz3	PtrTrz <sup>L1</sup>	N	M	MSHISNLRLLSPLNPTLRFPFSSKHRPYSLTILSSSSPYPKRRHRTTPNHPSLNFRRSRS (61)
PpeTrz3	PpeTrz <sup>L1</sup>	N	M	MPQVTNLRLLFFSPFRLSLSLFSKPLKPRFTALASSYRKRHRPIPQSPNTGARN (59)
SitTrz5	SitTrz <sup>L1</sup>	-	M	MPQVAAPLRLRLSQTALAPALLHLSRRLFTSSSPFGRAASLRALAYRRHHHPRRGSSTLRK (64)
SbiTrz4	SbiTrz <sup>L1</sup>	N	M	MPQVAGPLRRLPLSQTALASAPALLHLSRRLSSCSPASFGRAASLRALAYRRRHPEPRRG (63)
VviTrz3	VviTrz <sup>L1</sup>	N	M	MPHLTSFRLLYCSPLSPKSPFLSFTLSKSKSPLLNPPSFTVLSSSSGRYPKLRRHPHHLRRRN (67)
ZmaTrz3	ZmaTrz <sup>L1</sup>	N	M	MPQVAAPLRLRLSQTALAAPALLHLSRRLSSCPASLSRAAGLRALAYRRRHTEPRRG (61)

The putative nuclear localization signals were predicted using PSORT <http://psort.hgc.jp/form.html>, while the putative mitochondrial targeting signals were predicted using MITOPROT <http://ihg2.helmholtz-muenchen.de/ihg/mitoprot.html>. Nuc: stand for the nucleus; Mito: mitochondria. N indicates nuclear localization and M denotes mitochondrial localization. "-", the localization could not be predicted. The numbers refer to amino acid positions starting from the N-terminus.

other with a posterior probability value of 1. These observations suggest that tRNase Z<sup>L</sup> gene duplication took place in certain species.

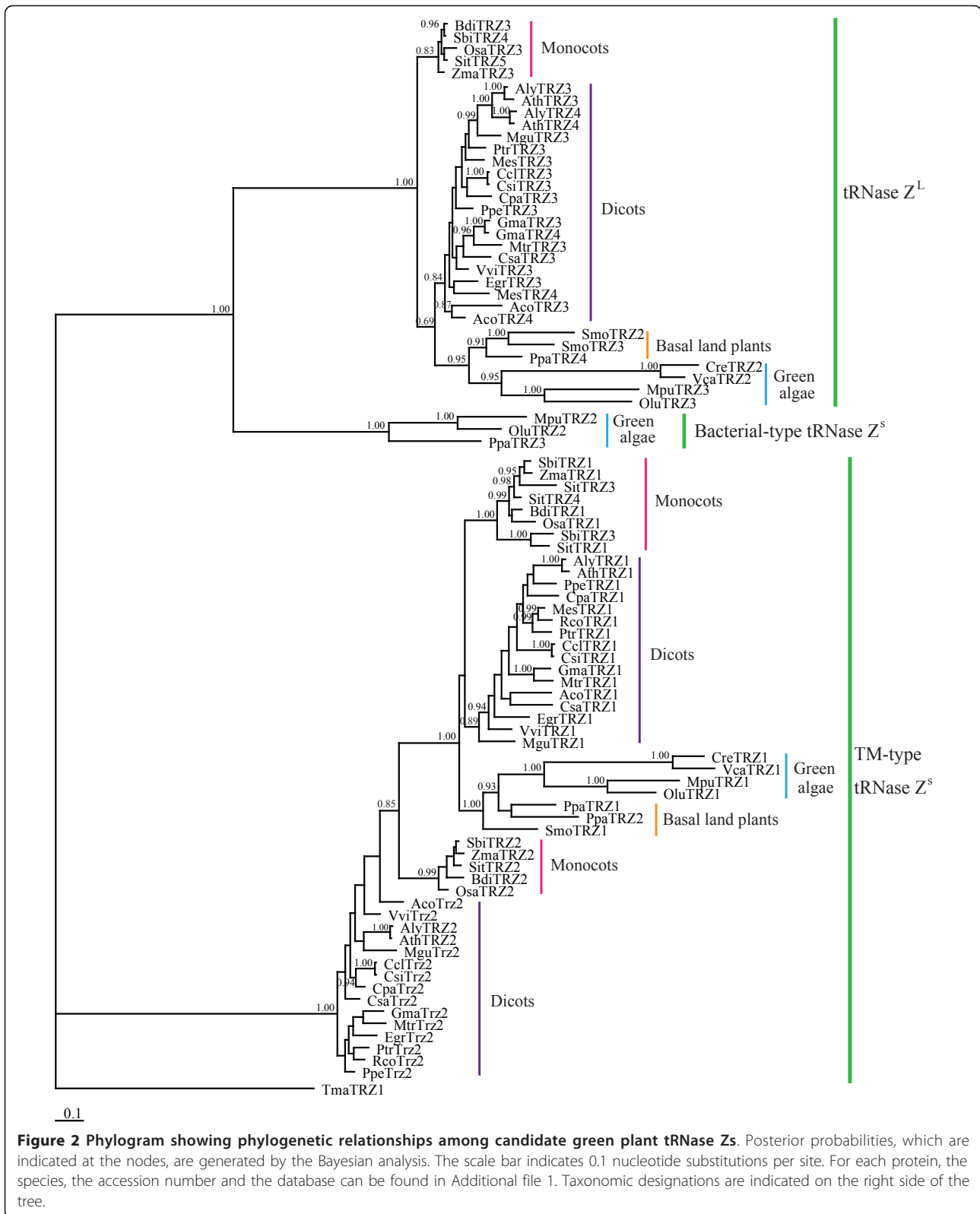
### Conservation of candidate green plant tRNase Z<sup>S</sup>s

To assess the presence and conservation of motifs in candidate green plant tRNase Zs, we performed multiple sequence alignments of identified tRNase Zs sequences. Candidate tRNase Z<sup>S</sup>s are analyzed first. A list of aligned tRNase Z<sup>S</sup>s from representative green plants is shown in Figures 1 and 3. For comparison, tRNase Z<sup>S</sup>s from *T. maritima*, *B. subtilis*, *E. coli*, the cyanobacterium *Synechocystis* sp. PCC 6803 and humans are included as needed. A full list of all aligned green plant tRNase Z<sup>S</sup>s is presented in Additional file 4.

Sequence comparison reveals that except for the N-terminal regions, tRNase Z<sup>S</sup>s from the flowering plants appear to be more similar to each other than to either the basal land plants or green algae (Figure 3). Furthermore, the sequences of the green plant tRNase Z<sup>S</sup>s are highly divergent from those of *T. maritima*, *B. subtilis*,

*E. coli* and human tRNase Z<sup>S</sup>s. For example, *A. thaliana* tRNase Z<sup>S1</sup> (AthTRZ1) and tRNase Z<sup>S2</sup> (AthTRZ2) exhibit only 17% - 24% and 16% - 21% identity (25% - 35% and 27% - 31% similarity), respectively, with those from *T. maritima*, *B. subtilis*, *E. coli* and humans. Overall, the sequence conservation between the green plant tRNase Z<sup>S</sup>s and non-plant species is largely confined to the highly conserved motifs of tRNase Zs.

Unexpectedly, careful examination of the sequences of green plant tRNase Z<sup>S</sup>s reveals that most of the proteins possess several unique features that distinguish them from bacterial-type tRNase Z<sup>S</sup>s and thus justify their classification as TM-type tRNase Z<sup>S</sup>. The sequence logos for motifs unique to TM-type tRNase Z<sup>S</sup>s are presented in Figure 4. First, most green plant tRNase Z<sup>S</sup>s harbor the TM type flexible arm. The plant TM-type flexible arms show only weak protein sequence homology to the bacterial-type flexible arms, and have distinctive features including the absence of the GP motif and the presence of a consensus sequence KLKxxYxxLxGx-xIxxLK, here termed the KL motif (Figures 3 and 4).



This Lys- and Leu-rich motif was previously unappreciated in *A. thaliana*, likely due to the limited number of plant sequences available at the time analysis was performed [39]. As might be expected, there are variations in the consensus sequence (Figures 3 and 4).

Second, most green plant tRNase Z<sup>S</sup>s lack the PxKxRN motif normally present on the amino side of Motif I. Instead, they share a Gly-rich consensus sequence EGxSxxG in an analogous position to the PxKxRN motif (Figure 3). In some cases, variants in the consensus sequence are found (Figures 3 and 4). Notably, there is very weak sequence similarity between this motif and the corresponding region in *T. maritima* tRNase Z<sup>S</sup>.

Third, most green plant tRNase Z<sup>S</sup>s contain conserved xExT and HxH motifs in place of the HEAT and HST motifs, respectively. The xExT motif is related to the HEAT motif except that the invariant His is replaced by a Thr/Met/Val/Leu residue and the invariant Ala is often replaced by Ser, or occasionally Met, Thr and Cys (Figures 3 and 4). In contrast, the HxH motif is related to the HST motif, except that the conserved Ser is mostly replaced by Thr, and the conserved Thr is substituted by His (Figures 3 and 4). Based on the structural and mutagenesis studies, it has been suggested that the Glu of the HEAT motif and His of the HST motif play a role in facilitating proton transfer at the final stage of reaction [25,29,40]. In particular, the His residue is implicated as the proton donor. It is highly likely that the conserved Glu-His pair from the xExT and HxH motifs in the TM-type tRNase Z<sup>S</sup>s may also participate in the terminal proton transfer reaction.

Multiple sequence alignment of bacterial-type tRNase Z<sup>S</sup>s from the basal plant *P. patens* (PpaTRZ3) and green algae reveals that in addition to the well conserved Motifs I-V, these candidates possess a bacterial-type flexible arm containing the GP, HEAT and HST motifs, and the variant PxKxRN motif, in which Lys is replaced with Leu (Figure 1).

### Conservation of tRNase Z<sup>L</sup>s in green plants

Since tRNase Z<sup>L</sup> can be divided into the N-terminal and C-terminal halves, which are related by weak sequence similarity, we aligned these two halves separately. The alignment of the N-terminal and C-terminal halves of representative candidate tRNase Z<sup>L</sup>s from diverse green plant species are shown in Figures 5 and 6, respectively. Two non-plant eukaryotic tRNase Z<sup>L</sup>s from *D. melanogaster* and humans were included for comparison. A complete list of all aligned green plant tRNase Z<sup>L</sup>s identified here is given in Additional file 5.

Except for the N-terminal regions, candidate tRNase Z<sup>L</sup>s from land plants share a high degree of overall sequence similarity to each other. In contrast, candidate

tRNase Z<sup>L</sup>s from the four green algae show relatively low overall sequence similarity to each other and to the land plant tRNase Z<sup>L</sup>s. In particular, their N-terminal halves are highly divergent compared to those from the land plants. Furthermore, sequence similarity between land plant tRNase Z<sup>L</sup>s and their homologs in green algae, *D. melanogaster* and humans is limited to the conserved motifs identified in tRNase Zs.

The N-terminal halves of all green plant tRNase Z<sup>L</sup>s contain a conserved eukaryotic-type flexible arm domain containing the GP motif. In addition, the N-terminal halves of the proteins also contain a pseudo-PxKxRN motif and pseudo-Motifs I and II, which differ from their original motifs in many positions including critical residues for tRNase Z functions (Figure 5). These pseudo-motifs were previously identified in candidate fungi tRNase Z<sup>L</sup>s [51]. In contrast, the C-terminal halves of green plant tRNase Z<sup>L</sup>s harbor conserved Motifs I-V and the PxKxRN, HEAT and HST motifs (Figure 6). These motifs appear in the same relative order in all the sequences in which they are present.

### Many chloroplast tRNA genes encode a portion of the 3'-CCA sequence

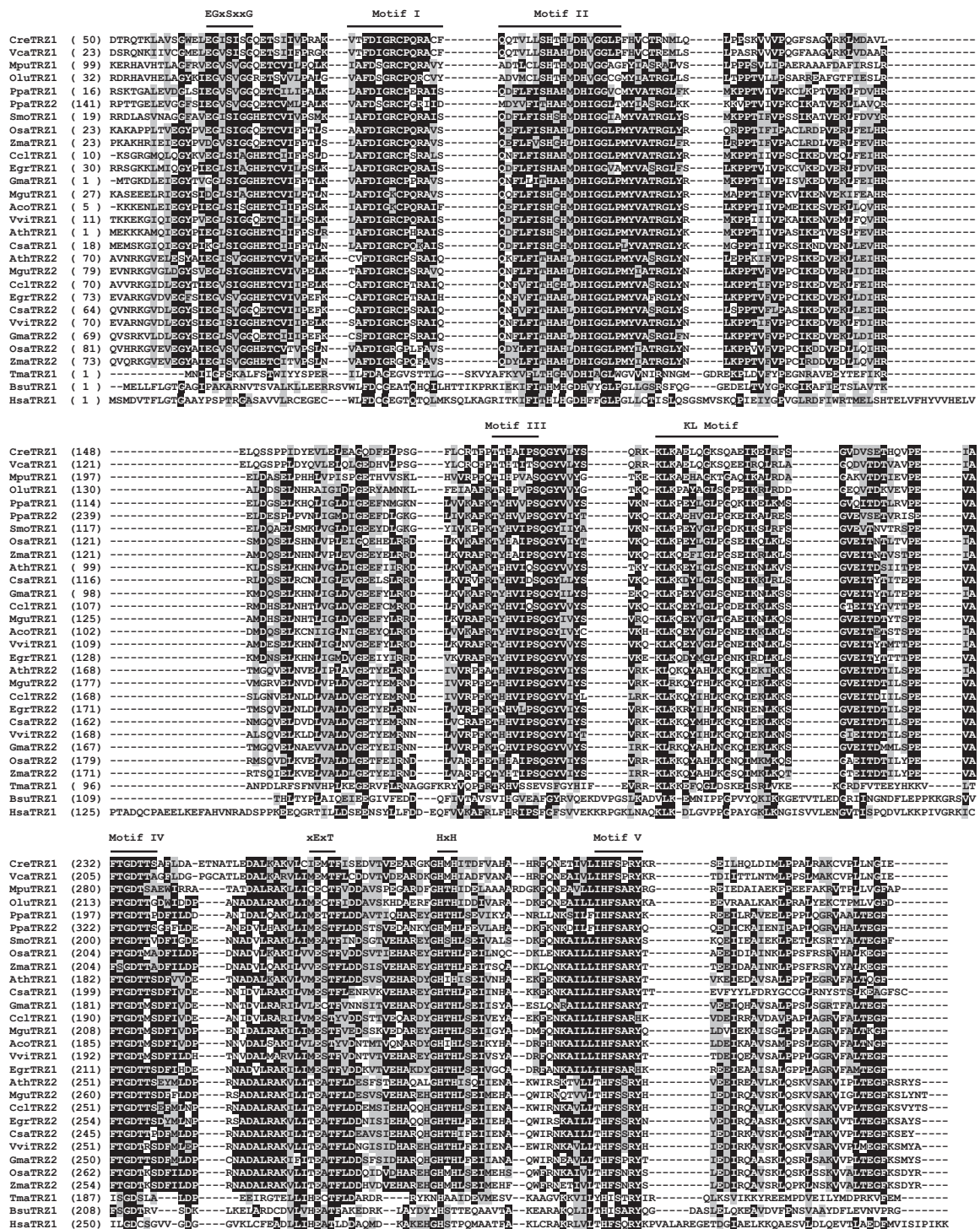
It has been suggested that many chloroplast tRNA genes encode partial CCA sequences [55]. However, this conclusion was primarily based on the examination of the 3'-flanking sequences of tRNA genes from three chloroplast genomes. To determine if this conclusion holds true when more chloroplast genome sequences are now available for analysis, we examined the presence of tRNA genes encoding whole or partial CCA sequences in 15 chloroplast genomes. The results obtained were tabulated in Table 4.

Indeed, we found that many tRNA genes in the chloroplast genomes encode partial CCA sequences, which could serve as part of the CCA sequence (Table 4). For instance, in the *A. thaliana* chloroplast genome, 31% of tRNA genes encode the whole or partial CCA sequences. Moreover, many chloroplast tRNA genes encode the first base of the CCA sequence. For example, of the 37 *A. thaliana* chloroplast tRNA genes examined, 11 had C after the discriminator.

### Discussion

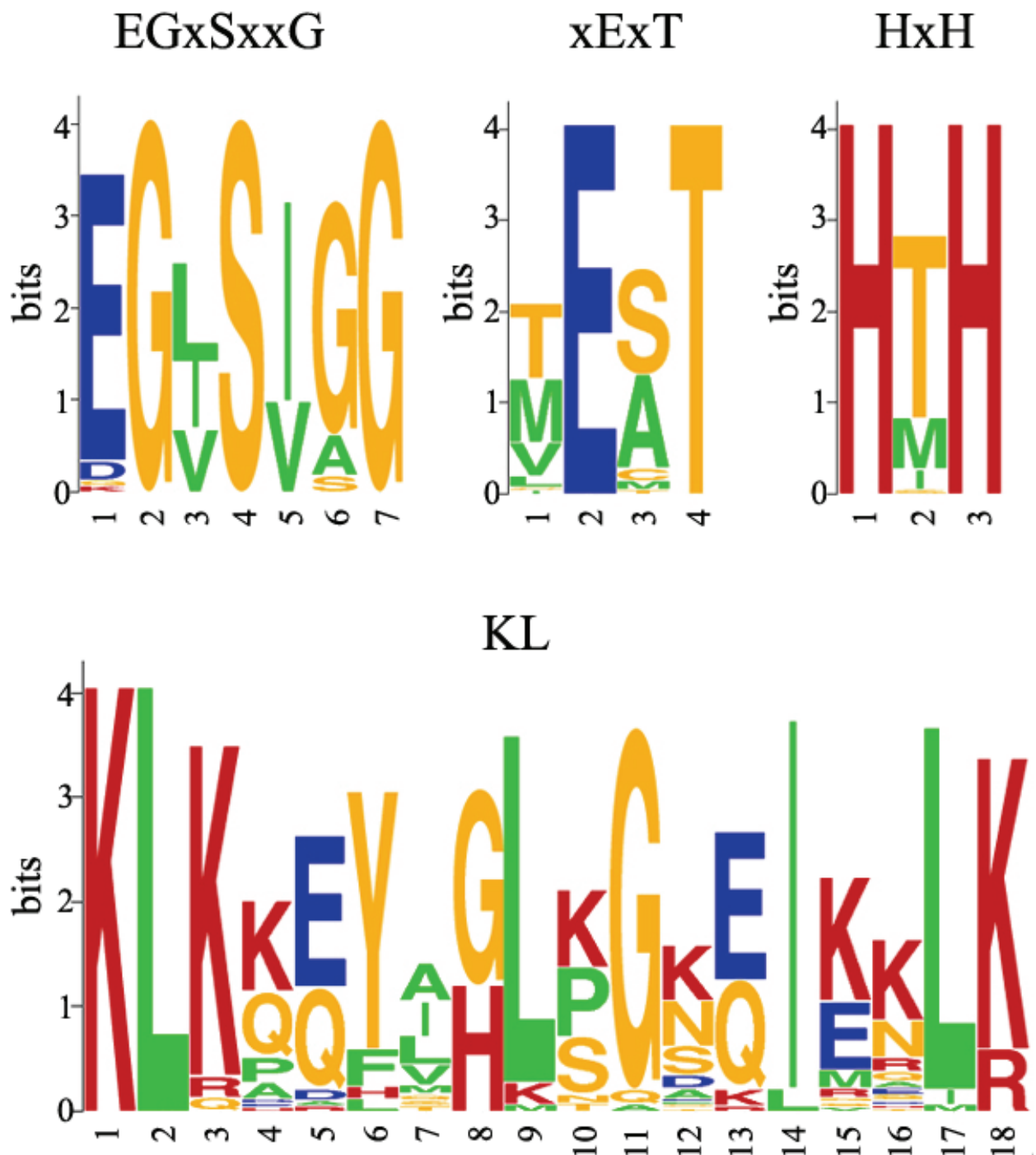
#### The presence of multiple tRNase Zs in green plants

Unlike *C. elegans*, *D. melanogaster*, humans and most fungal species examined to date, green plants are unique in that they possess multiple tRNase Zs. Of the 27 complete green plant genomes analyzed, the majority of them encode two tRNase Z<sup>S</sup>s and one or two tRNase Z<sup>L</sup>s. It is reasonable to expect that the existence of multiple tRNase Zs would be common to green plants. This phenomenon appears to have arisen from genome doubling (polyploidy), which seems to be a driving force in

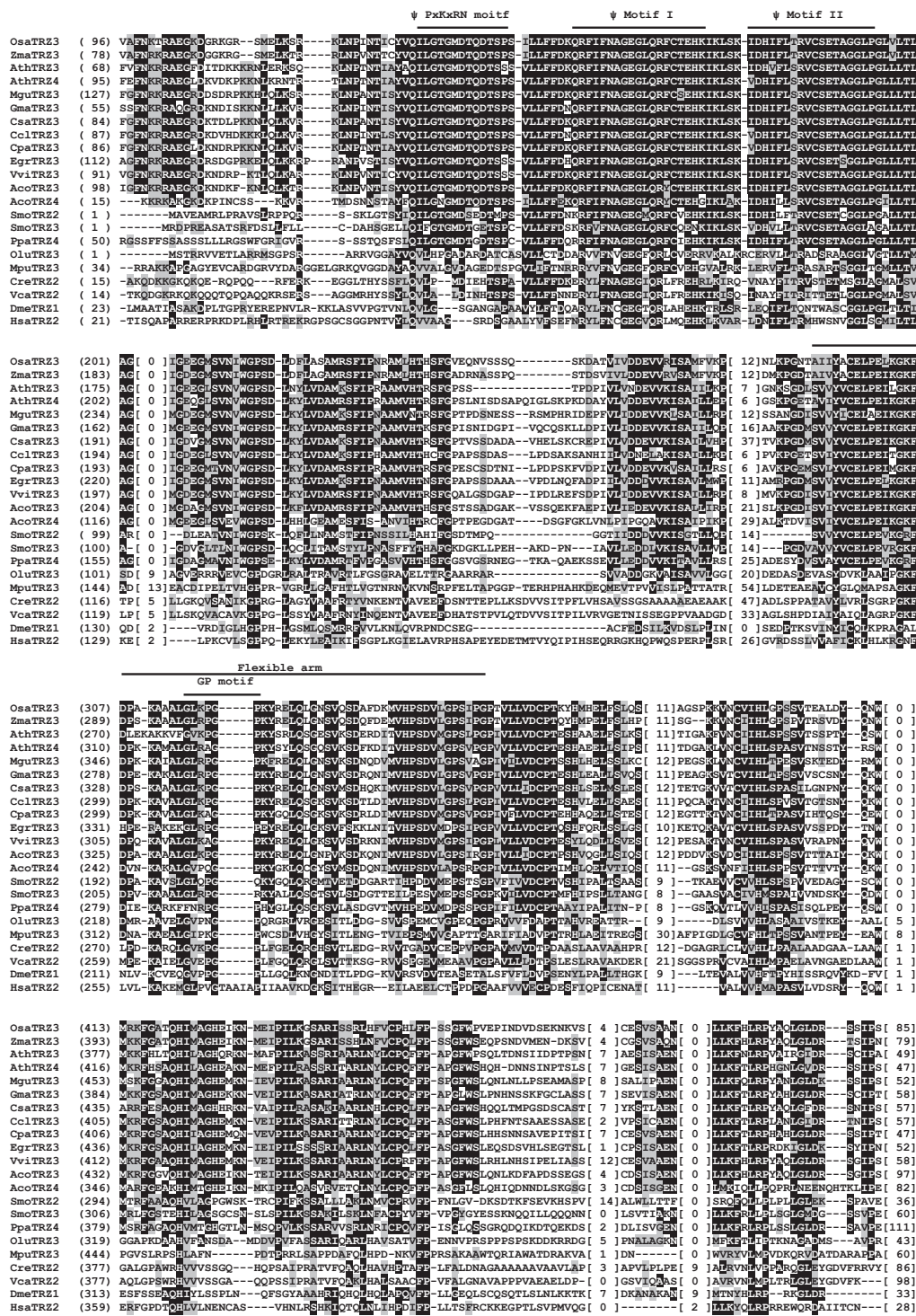


**Figure 3** Amino acid sequence alignment of the representative TM-type green plant and non-green plant and non-green plant tRNase Z<sup>s</sup>. Plant TM-type tRNase Z<sup>s</sup> are from *C. reinhardtii* (CreTRZ1), *V. carteri* (VcaTRZ1), *M. pusilla* (MpuTRZ1), *O. lucimarinus* (OluTRZ1), *P. patens* (PpaTRZ1 and PpaTRZ2), *S. moellendorffii* (SmoTRZ1), *O. sativa japonica* (OsaTRZ1 and OsaTRZ2), *Z. mays* (ZmaTRZ1 and ZmaTRZ2), *A. thaliana* (AthTRZ1 and AthTRZ2) [20], *C. sativus* (CsaTRZ1 and CsaTRZ2), *G. max* (GmaTRZ1 and GmaTRZ2), *C. clementina* (CclTRZ1 and CclTRZ2), *M. guttatus* (MguTRZ1 and MguTRZ2), *A. coerulea* (AcoTRZ1), *V. vinifera* (VviTRZ1 and VviTRZ2) and *E. grandis* (EgrTRZ1 and EgrTRZ2). Non-green plant tRNase Z<sup>s</sup> are from *T. maritima* (TmaTRZ1) [30], *B. subtilis* (BsuTRZ1) [60], humans (HsaTRZ1) [14]. Protein accession numbers are shown in Table 1. Alignment presentations are as described in the legend to Figure 1.





**Figure 4** Sequence logos of unique motifs found in candidate TM-type tRNase Z<sup>s</sup> from green plants. The sequence logos of motifs specific to candidate TM-type tRNase Z<sup>s</sup> from green plants were derived from the 51 alignments. The sequence logos were created using WebLogo <http://weblogo.berkeley.edu>. The height of each amino acid indicates the level of conservation at that position. Amino acids are colored as follows: red, basic; blue, acidic; orange, polar; and green, hydrophobic.



**Figure 5** Sequence alignment of N-terminal halves of representative green plant and non-green plant tRNase Z<sup>L</sup>s. Green plant tRNase Z<sup>L</sup>s are from *O. sativa japonica* (OsaTRZ3), *Z. mays* (ZmaTRZ3), *A. thaliana* (AthTRZ3 and AthTRZ4) [20], *M. guttatus* (MguTRZ3), *G. max* (GmaTRZ3), *C. sativus* (CsaTRZ3), *C. clementina* (CclTRZ3), *C. papaya* (CpaTRZ3), *E. grandis* (EgrTRZ3), *V. vinifera* (VviTRZ3), *A. coerulea* (AcoTRZ3 and AcoTRZ4), *S. moellendorffii* (SmoTRZ2 and SmoTRZ3), *P. patens* (PpaTRZ4), *O. lucimarinus* (OluTRZ3), *M. pusilla* (MpuTRZ3), *Chlamydomonas reinhardtii* (CreTRZ2) and *V. carteri* (VcaTRZ2). Non-green plant *D. melanogaster* (DmeTRZ1) [16] and humans (HsaTRZ2) [14] are included for comparison. The annotation of the alignment is as described in the legend to Figure 1.

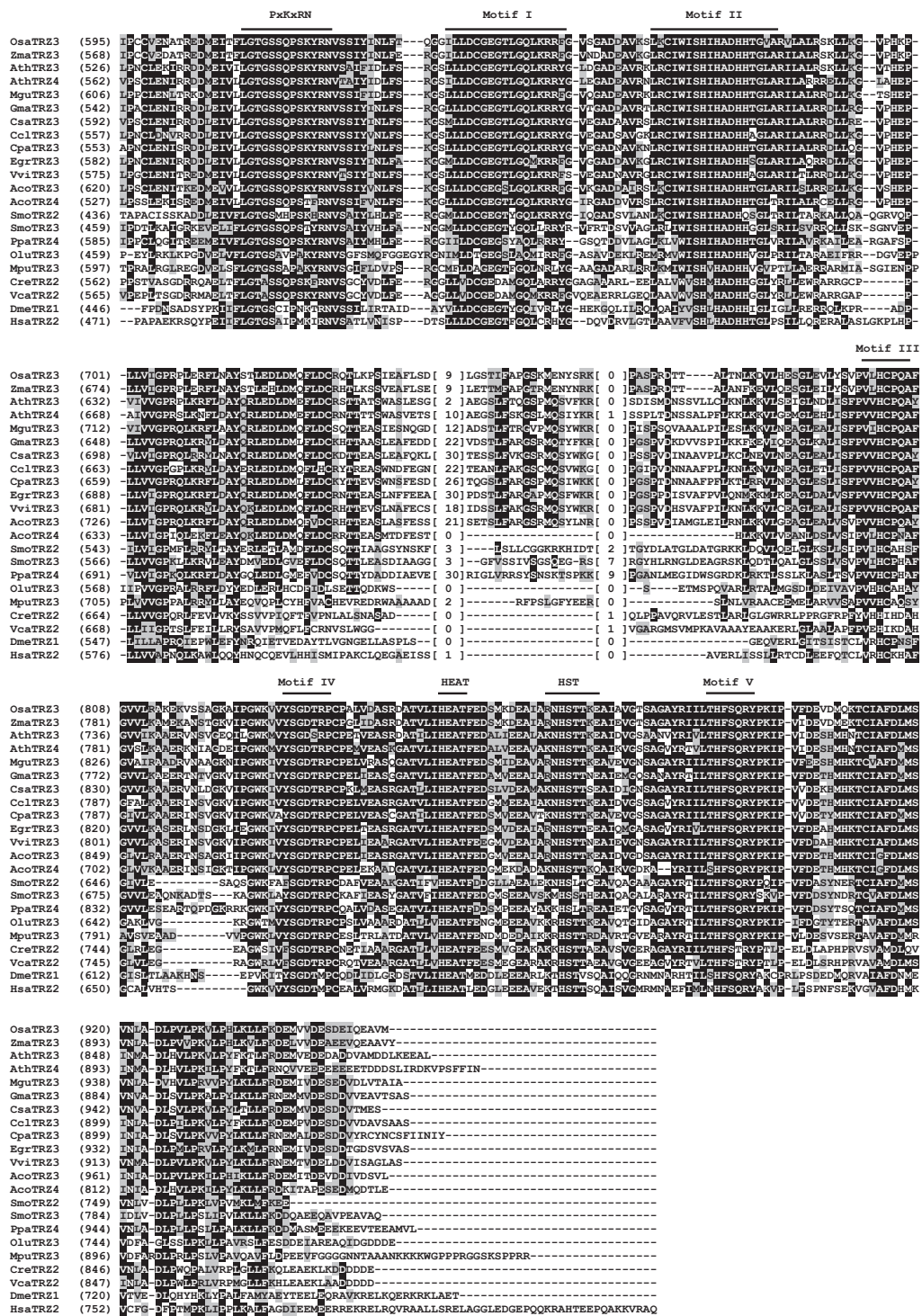


Figure 6 Sequence alignment of C-terminal halves of representative green plant and non-green plant tRNAse ZLs. Same legend as in Figure 5.

**Table 4 The distribution of the trinucleotide sequences immediately after the discriminator nucleotide in chloroplast tRNA genes**

Species	<sup>74</sup> CCA <sup>76</sup>	<sup>74</sup> CCN <sup>76</sup>	<sup>74</sup> CNN <sup>76</sup>	Total genes examined	% <sup>+</sup>
<i>Arabidopsis thaliana</i>	0	1	11	37	32
<i>Brachypodium distachyon</i>	1	1	8	38	26
<i>Carica papaya</i>	0	1	8	37	24
<i>Citrus sinensis</i>	2	1	7	37	27
<i>Cucumis sativus</i>	0	1	7	37	22
<i>Eucalyptus grandis</i>	0	1	8	36	25
<i>Glycine max</i>	0	1	9	37	27
<i>Manihot esculenta</i>	0	1	7	37	22
<i>Medicago truncatula</i>	0	0	5	29	17
<i>Oryza sativa japonica</i>	1	2	7	38	26
<i>Populus trichocarpa</i>	0	0	10	37	27
<i>Prunus persica</i>	0	1	8	37	24
<i>Sorghum bicolor</i>	0	2	7	38	24
<i>Vitis vinifera</i>	0	1	8	37	24
<i>Zea mays</i>	1	2	7	37	27

Nucleotide sequences of chloroplast tRNA genes and their 3' flanking regions were obtained from the GenBank genome database <http://www.ncbi.nlm.nih.gov/genomes/genlist.cgi?taxid=2759&type=4&name=Eukaryotae%20Organelles>, the Chloroplast Genome Database (ChloroplastDB; <http://chloroplast.cbio.psu.edu/organism.cgi>), and the Organelle Genome Database (GOBASE; <http://gobase.bcm.umontreal.ca/searches/gene.php>). The discriminator nucleotide was predicted using tRNAscan-SE 1.21 <http://lowelab.ucsc.edu/tRNAscan-SE/>. <sup>+</sup>Percentage of chloroplast tRNA genes encoding <sup>74</sup>CNN<sup>76</sup>, <sup>74</sup>CCN<sup>76</sup> and <sup>74</sup>CCA<sup>76</sup>.

plant evolution and variation [56]. However, there seems to be no correlation between the number of tRNase Zs and genome size (Table 5).

An unexpected observation in this study is that most green plant tRNase Z<sup>S</sup>s, which are clearly distinct from the bacterial-type tRNase Z<sup>S</sup>, represent the TM-type tRNase Z<sup>S</sup>. A comparison of motifs found in three different types of tRNase Zs from green plants is provided in Figure 7. Our phylogenetic analysis also supports the existence of both the TM- and bacterial-types of tRNase Z<sup>S</sup> in green plants (Figure 2). The TM-type tRNase Z<sup>S</sup> was previously found only in the hyperthermophilic bacterium *T. maritima* and the flowering plant *A. thaliana*,

and thus, it was originally thought to be a minor type. Our data greatly expand the repertoire of this type of tRNase Z. Although it remains to be determined if the TM-type tRNase Z<sup>S</sup> is also widespread in other taxonomic groups, the prevalence of the TM-type tRNase Z<sup>S</sup> in green plants suggests that this type might be plant-specific.

In contrast to the TM-type tRNase Z<sup>S</sup>, which is widespread in the green plants, the bacterial-type tRNase Z<sup>S</sup> is only present in the basal land plants and green algae. Sequence analysis reveals that the two types of tRNase Z<sup>S</sup>s are related to but distinct from each other, suggesting that they likely arose from the same ancestral

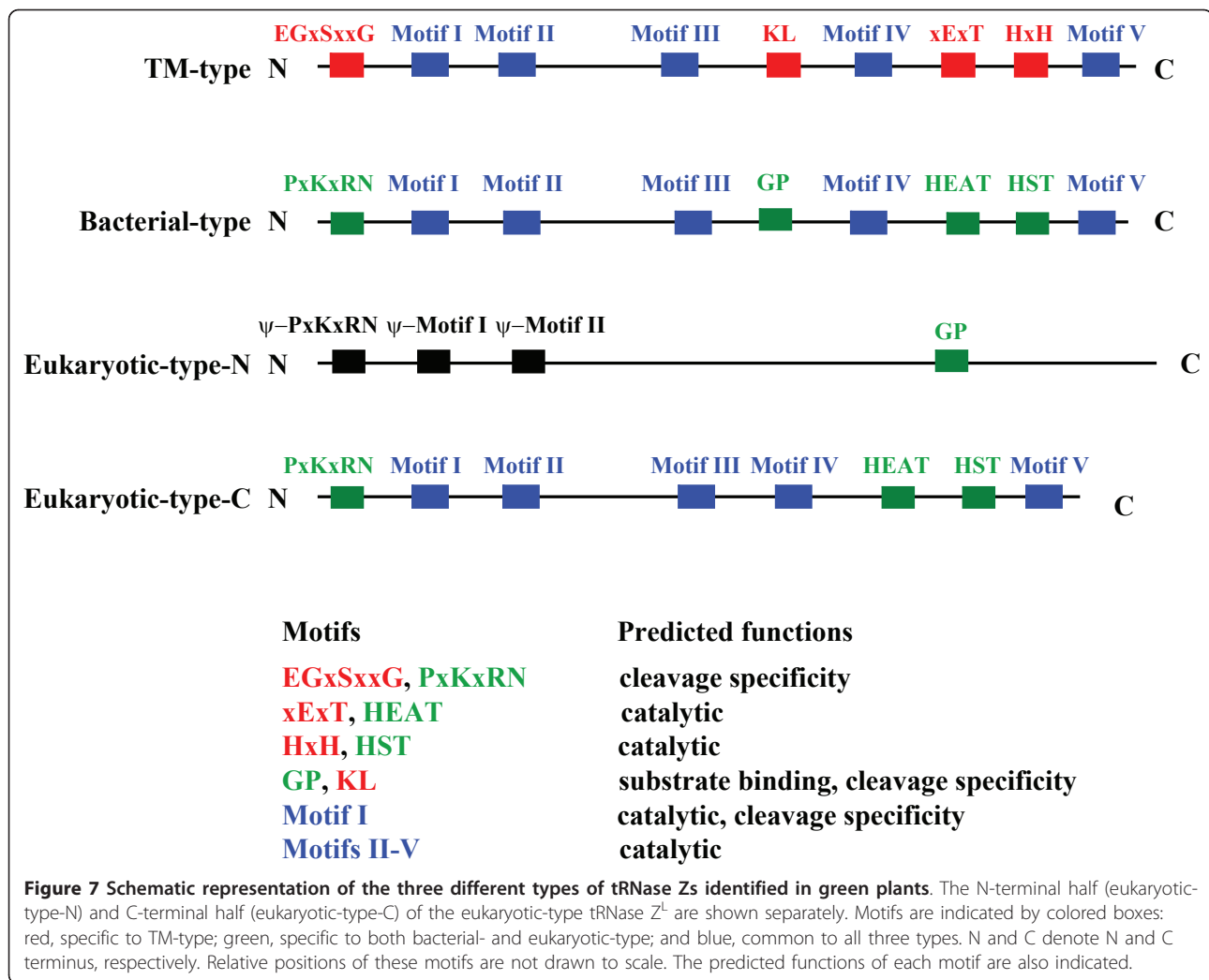
**Table 5 The genome size and ploidy level of representative green plants and the number of tRNase Zs encoded by their genomes**

Species	Common name	No. tRNase Z <sup>#</sup>	Ploidy level	Genome size (Mbp)	Refs
<i>Arabidopsis thaliana</i>	Mouse-ear cress	4	Diploid	125	[73]
<i>Brachypodium distachyon</i>	Purple false brome	3	Diploid	272	[74]
<i>Carica papaya</i>	Papaya	3	Diploid	372	[75]
<i>Cucumis sativus</i>	Cucumber	3	Diploid	367	[76]
<i>Glycine max</i>	Soybean	4	Diploid	1100	[77]
<i>Oryza sativa japonica</i>	Rice	3	Diploid	389	[78]
<i>Populus trichocarpa</i>	Black cottonwood or poplar	3	Diploid	485	[79]
<i>Ricinus communis</i>	Castor bean	3	Diploid	350	[80]
<i>Setaria italica</i>	Foxtail millet	5	Diploid	490	[81]
<i>Sorghum bicolor</i>	Sorghum	4	Diploid	730	[82]
<i>Vitis vinifera</i>	Grape	3	Diploid	505	[83]
<i>Zea mays</i>	Maize	3	Diploid	2300	[84]

The genome size and ploidy level were obtained from the Plant DNA C-values database <http://www.kew.org/cvalues/>.

<sup>#</sup>The number of tRNase Zs





tRNase Z<sup>S</sup> gene, followed by sequence divergence. Since the flowering plants possess only the TM-type tRNase Z<sup>S</sup>, the bacterial-type tRNase Z<sup>S</sup> has apparently been lost during the course of plant evolution whereas the TM-type tRNase Z<sup>S</sup> has been maintained. Interestingly, the bacterial-type tRNase Z<sup>S</sup> is also found in humans. Although the precise function of human tRNase Z<sup>S</sup> is unknown, it may play a role in degradation of a subset of microRNAs (miRNA) in the cytoplasm [57].

It is likely that multiple tRNase Zs found in the flowering plants are localized in different subcellular compartments. In most of the flowering plants examined, the two TM-type tRNase Z<sup>S</sup>s (tRNase Z<sup>S1</sup> and tRNase Z<sup>S2</sup>) are predicted to be either in the cytosol or in the chloroplast, whereas one of tRNase Z<sup>L</sup> (tRNase Z<sup>L1</sup>) is predicted to contain both nuclear and mitochondrial targeting signals. The second tRNase Z<sup>L</sup> (tRNase Z<sup>L2</sup>) found in some plants either contains or lacks a

predicted mitochondrial targeting signal. These predictions are generally supported by recent subcellular localization studies of *A. thaliana* tRNase Zs [20]. *A. thaliana* tRNase Z<sup>S1</sup> (AthTRZ1) and tRNase Z<sup>S2</sup> (AthTRZ2) are localized in the cytoplasm and chloroplast, respectively, whereas *A. thaliana* tRNase Z<sup>L1</sup> (AthTRZ3) and tRNase Z<sup>L2</sup> (AthTRZ4) are targeted to both the nucleus and mitochondria, and the mitochondria, respectively [20].

Based on their predicted subcellular localization, it is most likely that tRNase Z<sup>S2</sup> and tRNase Z<sup>L1</sup> are involved in the 3'-ends processing of pre-tRNAs in the chloroplasts, and both the nucleus and mitochondria, respectively, whereas tRNase Z<sup>L2</sup> either serves as a back-up for mitochondrial pre-tRNA 3'-end processing or plays a specialized, yet to be discovered function. The function of tRNase Z<sup>S1</sup> is unknown. It may play a role in repair of incorrectly processed 3'-ends of tRNAs in the cytosol.

Alternatively, tRNase Z<sup>S1</sup> could also participate in nuclear pre-tRNA 3'-end processing, as suggested by the observation that deletion of both *A. thaliana* tRNase Z<sup>S1</sup> (AthTRZ1) and tRNase Z<sup>L1</sup> (AthTRZ3) results in a lethal phenotype [20].

#### The plant TM-type tRNase Z<sup>S</sup>s may not recognize the CCA sequence as an anti-determinant

The CCA triplet following the discriminator generally inhibits tRNase Z activity, with the first C at nucleotide position 74 having the greatest effect [25,33,58]. This anti-determinant property of the CCA sequence is thought to prevent removal of the 3'-terminal CCA sequence from mature tRNAs, and thus avoids futile cycles of CCA addition and removal [33,58-60]. However, unlike other types of tRNase Zs, the TM-type tRNase Z does not appear to recognize the CCA sequence that is downstream the discriminator and present in the trailer sequence as an anti-determinant. The recombinant *T. maritima* tRNase Z<sup>S</sup> cleaves after the CCA sequence which is encoded by all but one tRNA gene, leaving the CCA sequence intact [30]. Similarly, the recombinant *A. thaliana* tRNase Z<sup>S</sup> can process pre-tRNAs with the 3'-terminal CCA sequences embedded in the trailer sequence, albeit at a different position [61]. Like the recombinant proteins, a partially purified spinach chloroplast tRNase Z fraction can also cleave pre-tRNAs containing complete or partial CCA sequences after the first C<sup>74</sup> regardless of the sequence of the flanking region [55]. Importantly, the remainder of the CCA sequence can be added by chloroplast tRNA nucleotidyltransferase [55].

However, *A. thaliana* tRNase Z<sup>S</sup> appears to be unusual in that it can also cleave off the 3'-terminal CCA sequence from mature tRNAs *in vitro* [61]. Since mature tRNAs must be protected from counterproductive cleavage by tRNase Z<sup>S</sup> *in vivo*, it is likely that the plant chloroplast protein may acquire the ability to recognize the 3'-terminal CCA sequence as the mature tRNA 3'-end via cofactors [61].

#### Sequence determinants potentially involved in cleavage specificity

The mechanisms responsible for cleavage site selection and CCA inhibition are not fully understood, but appear to involve a combination of sequence features. First, the flexible arm has been suggested to contribute to cleavage site selection and the inhibitory effect of the CCA sequence [62]. This tRNase Z-specific element is located on the opposite side of the active site and binds primarily the D and T loops of the pre-tRNA [28,39,63]. Notably, the flexible arms of the *T. maritima* and plant tRNase Z<sup>S</sup>s lack the GP motif but contain the KL motif. Additionally, they are significantly smaller in length

relative to those of other tRNase Zs possessing the anti-determinant function. Structural studies reveal that although TM- and bacterial-types of flexible arms share a similar overall structure which is composed of a compact globular domain and an extended two-stranded stalk and protrudes from the protein core, they have different globular domains [25-28,64]. In the TM-type flexible arm, the globular domain consists of one very short  $\alpha$ -helix, one long helix and one  $3_{10}$ -helix, whereas in the bacterial-type flexible arm, it is composed of two  $\alpha$ -helices, two  $\beta$ -strands and one  $3_{10}$ -helix. It would be interesting to know how the differences in the sequence feature, length and topology of the flexible arm may contribute to cleavage specificity by the enzymes.

Another possible motif involved in selection of the cleavage site has been suggested to be Motif I which appears to participate in binding the acceptor stem of pre-tRNA substrates [33,62]. *In vitro* studies using the recombinant *T. maritima* tRNase Z<sup>S</sup> has suggested that the Ser31 and Thr33 residues of Motif I are involved in the cleavage site selection, with the former residue being more critical [30]. However, only individual, but not simultaneous mutations of these two non-Gln residues to Gln (which are found at the corresponding positions in other tRNase Zs with the CCA anti-determinant) in *T. maritima* tRNase Z<sup>S</sup> affects the cleavage site selection [62].

In addition to the flexible arm and Motif I, the PxKxRN motif, in particular, the two basic residues in the motif, has also been suggested to be involved in the cleavage specificity and, by inference, CCA inhibition of tRNase Zs, since this motif is absent from *T. maritima* and *A. thaliana* [33]. Strikingly, tRNase Z<sup>S</sup>s found in all flowering plants examined so far also lack this motif (Figure 3). However, since the archaeon *Methanococcus jannaschii* tRNase Z<sup>S</sup> harbors a PxKxRN motif but cannot recognize the CCA sequence as an anti-determinant *in vitro*, suggesting that other sequence elements may also be involved in the cleavage specificity of tRNase Z [61]. Alternatively, it has been suggested that a high enzyme concentration used in the assay may contribute to the lack of a CCA anti-determinant effect with the *M. jannaschii* tRNase Z<sup>S</sup> [33].

#### Why is the TM-type tRNase Z<sup>S</sup> developed for chloroplast pre-tRNA 3'-end processing?

The discovery that the TM-type tRNase Z<sup>S</sup> is widespread in green plants raises a question as to why plants adopt the TM-type tRNase Z<sup>S</sup> over other types of tRNase Z for the 3'-end processing of chloroplast pre-tRNAs. One possible explanation is that the TM-type tRNase Z<sup>S</sup> identified in plants has evolved to adapt to chloroplast pre-tRNA 3'-end processing. It has previously been suggested that many plant chloroplast

tRNA genes encode  $C^{74}$  based on a limited number of available chloroplast tRNA genes [55]. To see if this conclusion could be extended to more flowering plants, we have examined the 3'-flanking region for tRNA genes in additional chloroplast genomes. Indeed, many chloroplast tRNA genes seem to encode partial CCA sequences (Table 4).

The development of the TM-type tRNase  $Z^S$  may be particularly important for chloroplasts, which have a limited but sufficient number of self-encoded tRNA species and do not import tRNAs [65,66]. As all possible codons are used in the chloroplast protein-encoding genes, all chloroplast tRNA species appear to be used in protein synthesis [66]. Thus, efficient 3'-end processing of each pre-tRNA by tRNase Z may be critical in ensuring maximum efficiency in chloroplast protein synthesis.

#### The presence of candidate tRNase Z-like proteins in green plants

tRNase  $Z^S$ -like proteins have previously been found in cyanobacteria including *Synechocystis* sp. PCC 6803 [67] and fungi [51], whereas tRNase  $Z^L$ -like proteins have not been reported. In *Synechocystis*, one tRNase  $Z^S$ -like protein (sll1036) has been identified. This protein does not exhibit any tRNase Z activity in vitro, consistent with the lack of some of the most conserved motifs of tRNase Zs in the protein sequence [67]. Since candidate tRNase Z-like proteins found in basal land plants and green algae either lack all or some of the conserved motifs of tRNase Zs, they most likely possess no tRNase Z activity.

It has been suggested that tRNase  $Z^L$  has arisen from the fusion of duplicated tRNase  $Z^S$  genes with further sequence diversification. It is possible that tRNase  $Z^S$ -like proteins found in the basal land plants and green algae may represent relics of original tRNase  $Z^S$  that were mutated during diversification of eukaryotic tRNase Z genes. Alternatively, these proteins may play species-specific functions as suggested for cyanobacterial tRNase  $Z^S$ -like proteins [67]. The predominant presence in the basal land plants and green algae of tRNase Z-like proteins also suggests that duplication of ancestral tRNase  $Z^S$  genes may occur early in green plant evolution.

#### Conclusions

This study represents the first large-scale identification and analysis of green plant tRNase Zs. Our survey of current plant genome databases shows that green plants are represented by multiple tRNase Zs, which include one or two tRNase  $Z^L$ s and two tRNase  $Z^S$ s. One tRNase  $Z^L$  is predicted to participate in 3'-end processing of nuclear and mitochondrial pre-tRNA, whereas

the other is likely to provide a backup for mitochondrial pre-tRNA processing. It appears that most tRNase  $Z^S$ s, which is widespread throughout the green plants, belong to a minor but highly distinct type of tRNase  $Z^S$  (TM-type). In contrast, the typical bacterial-type tRNase  $Z^S$  is restricted to basal land plants and green algae. The apparent lack of the bacterial-type tRNase  $Z^S$  in flowering plants suggest that while both types were present in the basal land plants, the bacterial-type tRNase  $Z^S$  was discarded in favor of TM-type during plant evolution. Based on our results and previous studies, we propose that like *T. maritima* tRNase  $Z^S$ , TM-type tRNase  $Z^S$ s found in green plants seem not to recognize the CCA sequence as an anti-determinant and that the rise of this type of tRNase  $Z^S$  appears to accommodate the 3'-end processing of chloroplast pre-tRNAs with partial or whole CCA sequences. This unusual property of green plant tRNase  $Z^S$ s is likely due to multiple sequence determinants including the TM-type-specific flexible arm comprising the KL motif, Motif I and lack of the PxKxRN motif.

However, it should be noted that bioinformatics analysis alone cannot resolve possible differences in cleavage specificity among TM-type tRNase  $Z^S$ s. The complete understanding of the mechanisms of the cleavage specificities of the TM-type tRNase  $Z^S$ s awaits the determination of the structures of these enzymes bound to pre-tRNA and the accumulation of more precise biochemical data.

#### Methods

##### Plant genome database search and protein sequence analysis

Candidate tRNase Zs were identified by BLAST and PSI-BLAST searches against the genome databases using known tRNase Zs as query sequences. The databases used include Phytozome <http://www.phytozome.net/>, the NCBI nonredundant protein sequence database <http://blast.ncbi.nlm.nih.gov/Blast.cgi>, Joint Genome Institute (JGI; <http://www.jgi.doe.gov/>) and Universal Protein Resource (Uniprot; <http://www.uniprot.org/>). An E-value cutoff of 0.001 was used in all searches. With this value, no  $\beta$ -CASP protein or other MBL protein was found. The resulting sequences were subject to validation as described [51]. The splicing pattern was verified using the FGENESH and FGENESH\_GC programs provided at the Softberry website <http://linux1.softberry.com/berry.phtml?topic=fgenes>. Prediction of subcellular localization of proteins was made using web-based prediction programs such as MITOPROT <http://ihg2.helmholtz-muenchen.de/ihg/mitoprot.html>, PSORT <http://psort.hgc.jp/form.html> and ChloroP <http://www.cbs.dtu.dk/services/ChloroP/>. Multiple sequence alignments were done by Clustal W [68].

## Phylogenetic analysis

Phylogenetic analysis was performed using the Bayesian approach, with *T. maritima* tRNase Z<sup>S</sup> (TmaTrz1) as an outgroup as described [51]. Briefly, full-length amino acid sequences of candidate plant tRNase Zs and TmaTrz1 were aligned by using Clustal W implemented in MEGA 5.0 [69]. After excluding gaps and the ambiguous sites, we used ProtTest 2.4 [70] to choose the most appropriate evolutionary model for our data set. The phylogenies were estimated by Bayesian inference with MrBayes 3.1.2 [71] using a mixture of the fixed amino acid models and I + G distribution. Statistical confidence was assessed by using Markov Chain Monte Carlo (MCMC) sampling approaches. Two separate runs including a total of four independent tree searches were conducted. All searches consisted of one 'cold' and three 'heated' Markov chains estimated for 10<sup>7</sup> generations, and every 1000 generations were sampled. The burn-in parameter was estimated by plotting -lnL against the generation number using Tracer 1.4.1 <http://beast.bio.ed.ac.uk/Tracer>, and the retained trees were used to estimate the consensus tree and the Bayesian posterior probabilities.

## Additional material

**Additional file 1: Distribution of candidate tRNase Zs identified in green plants.** Abbreviations for species names are indicated in the parentheses. \*The number of amino acids in plant tRNase Z and tRNase Z-like proteins. \*Indicates that mispredicted sequences obtained from the databases have been corrected. †Indicates the sequence could not be correctly predicted.

**Additional file 2: Pairwise sequence comparisons of *S. bicolor* and *S. italica* tRNase Z<sup>S</sup>s.** The accession numbers for proteins are listed in Additional file 1. The pairwise percent identity (I) and percent similarity (S) between tRNase Z<sup>S</sup>s from *S. bicolor* and *S. italica* were calculated using the Clustal W program [68].

**Additional file 3: Alignment of candidate tRNase Z<sup>S</sup>s from *S. bicolor* and *S. italica*.** The accession numbers for the candidates are listed in Additional file 1. The annotation of the alignment is described in the legend to Figure 1.

**Additional file 4: Alignment of candidate TM-type tRNase Z<sup>S</sup>s identified in green plants.** The accession numbers for the candidates are listed in Additional file 1. Annotation is as described in the legend to Figure 1.

**Additional file 5: Alignment of candidate tRNase Z<sup>L</sup>s identified in green plants.** The accession numbers for the candidates are listed in Additional file 1. The annotation of the alignment is as described in the legend for Figure 1.

## List of abbreviations

pre-tRNAs: tRNA precursors; miRNA: microRNAs; tRNase Z: tRNA 3' endonuclease; tRNase Z<sup>S</sup>: the short form of tRNase Z; tRNase Z<sup>L</sup>: the long form of tRNase Z;  $\beta$ -CASP: MBL-associated CISF Artemis SNM1/PSO2; CPSF-73: the cleavage and polyadenylation specificity factor; MBL: the metallo- $\beta$ -lactamase superfamily; TGF- $\beta$ : transforming growth factor- $\beta$ ; TM: *T. maritima*.

## Acknowledgements

We thank three anonymous reviewers for improving the quality of the manuscript, Jingjing Peng and Wei Zhao for valuable discussion. This work

was supported in part by grants from the National Science Foundation of China (31070703), the Priority Academic Program Development of Jiangsu Higher Education Institutions and Nanjing Normal University (2007104XGQ0148).

## Author details

<sup>1</sup>Laboratory of Yeast Genetics and Molecular Biology, School of Life Sciences, Nanjing Normal University, 1 Wenyuan Road, Nanjing 210046, China.

<sup>2</sup>Jiangsu Key Laboratory for Microbes and Genomics, School of Life Sciences, Nanjing Normal University, 1 Wenyuan Road, Nanjing 210046, China.

<sup>3</sup>Jiangsu Engineering and Technology Research Center for Industrialization of Microbial Resources, School of Life Sciences, Nanjing Normal University, 1 Wenyuan Road, Nanjing 210046, China. <sup>4</sup>Jiangsu Key Laboratory for Biodiversity and Biotechnology, School of Life Sciences, Nanjing Normal University, 1 Wenyuan Road, Nanjing 210046, China.

## Authors' contributions

LF, ZW and JL performed online database searches and sequence analysis. JY carried out a phylogenetic analysis. YH conceived this study, analyzed the data and drafted the manuscript. All authors have read and approved the final version of the manuscript.

## Competing interests

The authors declare that they have no competing interests.

Received: 8 May 2011 Accepted: 23 July 2011 Published: 23 July 2011

## References

1. Phizicky EM, Hopper AK: tRNA biology charges to the front. *Genes Dev* 2010, **24**:1832-1860.
2. Bayfield MA, Yang R, Maraia RJ: Conserved and divergent features of the structure and function of La and La-related proteins (LARPs). *Biochim Biophys Acta* 2010, **1799**:365-378.
3. Maraia RJ, Lamichhane TN: 3' processing of eukaryotic precursor tRNAs. *WIREs RNA* 2011, **2**:362-375.
4. Ceballos M, Vioque A: tRNase Z. *Protein Pept Lett* 2007, **14**:137-145.
5. Hartmann RK, Gossringer M, Spath B, Fischer S, Marchfelder A: Chapter 8 The Making of tRNAs and More - RNase P and tRNase Z. *Prog Nucleic Acid Res Mol Biol* 2009, **85C**:319-368.
6. Redko Y, Li de Lasierra-Gallay I, Condon C: When all's zed and done: the structure and function of RNase Z in prokaryotes. *Nat Rev Microbiol* 2007, **5**:278-286.
7. Spath B, Canino G, Marchfelder A: tRNase Z: the end is not in sight. *Cell Mol Life Sci* 2007, **64**:2404-2412.
8. Wolin SL, Cedervall T: The La protein. *Annu Rev Biochem* 2002, **71**:375-403.
9. Van Horn DJ, Yoo CJ, Xue D, Shi H, Wolin SL: The La protein in *Schizosaccharomyces pombe*: a conserved yet dispensable phosphoprotein that functions in tRNA maturation. *RNA* 1997, **3**:1434-1443.
10. Yoo CJ, Wolin SL: The yeast La protein is required for the 3' endonucleolytic cleavage that matures tRNA precursors. *Cell* 1997, **89**:393-402.
11. Copela LA, Fernandez CF, Sherrer RL, Wolin SL: Competition between the Rex1 exonuclease and the La protein affects both Trf4p-mediated RNA quality control and pre-tRNA maturation. *RNA* 2008, **14**:1214-1227.
12. Schafer B: RNA maturation in mitochondria of *S. cerevisiae* and *S. pombe*. *Gene* 2005, **354**:80-85.
13. Sugita M, Sugiura M: Regulation of gene expression in chloroplasts of higher plants. *Plant Mol Biol* 1996, **32**:315-326.
14. Tavtigian SV, Simard J, Teng DH, Abtin V, Baumgard M, Beck A, Camp NJ, Carillo AR, Chen Y, Dayananth P, Desrochers M, Dumont M, Farnham JM, Frank D, Frye C, Ghaffari S, Gupte JS, Hu R, Iliev D, Janecki T, Kort EN, Laity KE, Leavitt A, Leblanc G, McArthur-Morrison J, Pederson A, Penn B, Peterson KT, Reid JE, Richards S, et al: A candidate prostate cancer susceptibility gene at chromosome 17p. *Nat Genet* 2001, **27**:172-180.
15. Chen Y, Beck A, Davenport C, Chen Y, Shattuck D, Tavtigian SV: Characterization of TRZ1, a yeast homolog of the human candidate prostate cancer susceptibility gene ELAC2 encoding tRNase Z. *BMC Mol Biol* 2005, **6**:12.
16. Dubrovsky EB, Dubrovskaya VA, Levinger L, Schiffer S, Marchfelder A: *Drosophila* RNase Z processes mitochondrial and nuclear pre-tRNA 3' ends in vivo. *Nucleic Acids Res* 2004, **32**:255-262.



17. Smith MM, Levitan DJ: The *Caenorhabditis elegans* homolog of the putative prostate cancer susceptibility gene ELAC2, hoe-1, plays a role in germline proliferation. *Dev Biol* 2004, **266**:151-160.
18. Zhao Z, Su W, Yuan S, Huang Y: Functional conservation of tRNase Z<sup>L</sup> among *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe* and humans. *Biochem J* 2009, **422**:483-492.
19. Gan X, Yang J, Li J, Yu H, Dai H, Liu J, Huang Y: The fission yeast *Schizosaccharomyces pombe* has two distinct tRNase Z<sup>L</sup>s encoded by two different genes and differentially targeted to the nucleus and mitochondria. *Biochem J* 2011, **435**:103-111.
20. Canino G, Bocian E, Barbezier N, Echeverria M, Forner J, Binder S, Marchfelder A: *Arabidopsis* encodes four tRNase Z enzymes. *Plant Physiol* 2009, **150**:1491-1502.
21. Aravind L: An evolutionary classification of the metallo-beta-lactamase fold proteins. *In Silico Biol* 1999, **1**:69-91.
22. Daiyasu H, Osaka K, Ishino Y, Toh H: Expansion of the zinc metallo-hydrolase family of the beta-lactamase fold. *FEBS Lett* 2001, **503**:1-6.
23. Dominski Z: Nucleases of the metallo-beta-lactamase family and their role in DNA and RNA metabolism. *Crit Rev Biochem Mol Biol* 2007, **42**:67-93.
24. Schiffer S, Rosch S, Marchfelder A: Assigning a function to a conserved group of proteins: the tRNA 3'-processing enzymes. *EMBO J* 2002, **21**:2769-2777.
25. de la Sierra-Gallay IL, Pellegrini O, Condon C: Structural basis for substrate binding, cleavage and allostery in the tRNA maturase RNase Z. *Nature* 2005, **433**:657-661.
26. Ishii R, Minagawa A, Takaku H, Takagi M, Nashimoto M, Yokoyama S: Crystal structure of the tRNA 3' processing endoribonuclease tRNase Z from *Thermotoga maritima*. *J Biol Chem* 2005, **280**:14138-14144.
27. Kostecky B, Pohl E, Vogel A, Schilling O, Meyer-Klaucke W: The crystal structure of the zinc phosphodiesterase from *Escherichia coli* provides insight into function and cooperativity of tRNase Z-family proteins. *J Bacteriol* 2006, **188**:1607-1614.
28. Li de la Sierra-Gallay I, Mathy N, Pellegrini O, Condon C: Structure of the ubiquitous 3' processing enzyme RNase Z bound to transfer RNA. *Nat Struct Mol Biol* 2006, **13**:376-377.
29. Minagawa A, Takaku H, Ishii R, Takagi M, Yokoyama S, Nashimoto M: Identification by Mn<sup>2+</sup> rescue of two residues essential for the proton transfer of tRNase Z catalysis. *Nucleic Acids Res* 2006, **34**:3811-3818.
30. Minagawa A, Takaku H, Takagi M, Nashimoto M: A novel endonucleolytic mechanism to generate the CCA 3' termini of tRNA molecules in *Thermotoga maritima*. *J Biol Chem* 2004, **279**:15688-15697.
31. Schilling O, Vogel A, Kostecky B, Natal da Luz H, Spemann D, Spath B, Marchfelder A, Troger W, Meyer-Klaucke W: Zinc- and iron-dependent cytosolic metallo-beta-lactamase domain proteins exhibit similar zinc-binding affinities, independent of an atypical glutamate at the metal-binding site. *Biochem J* 2005, **385**:145-153.
32. Vogel A, Schilling O, Meyer-Klaucke W: Identification of metal binding residues for the binuclear zinc phosphodiesterase reveals identical coordination as glyoxalase II. *Biochemistry* 2004, **43**:10379-10386.
33. Zareen N, Hopkinson A, Levinger L: Residues in two homology blocks on the amino side of the tRNase Z His domain contribute unexpectedly to pre-tRNA 3' end processing. *RNA* 2006, **12**:1104-1115.
34. Zareen N, Yan H, Hopkinson A, Levinger L: Residues in the conserved His domain of fruit fly tRNase Z that function in catalysis are not involved in substrate recognition or binding. *J Mol Biol* 2005, **350**:189-199.
35. Spath B, Kirchner S, Vogel A, Schubert S, Meirlschmidt P, Aymanns S, Nezzar J, Marchfelder A: Analysis of the functional modules of the tRNA 3' endonuclease (tRNase Z). *J Biol Chem* 2005, **280**:35440-35447.
36. Baillat D, Hakimi MA, Naar AM, Shilatfard A, Cooch N, Shiekhatter R: Integrator, a multiprotein mediator of small nuclear RNA processing, associates with the C-terminal repeat of RNA polymerase II. *Cell* 2005, **123**:265-276.
37. Britton RA, Wen T, Schaefer L, Pellegrini O, Uicker WC, Mathy N, Tobin C, Daou R, Szyk J, Condon C: Maturation of the 5' end of *Bacillus subtilis* 16S rRNA by the essential ribonuclease YkqC/RNase J1. *Mol Microbiol* 2007, **63**:127-138.
38. Mathy N, Benard L, Pellegrini O, Daou R, Wen T, Condon C: 5'-to-3' exoribonuclease activity in bacteria: role of RNase J1 in rRNA maturation and 5' stability of mRNA. *Cell* 2007, **129**:681-692.
39. Schilling O, Spath B, Kostecky B, Marchfelder A, Meyer-Klaucke W, Vogel A: Exosite modules guide substrate recognition in the ZIPD/Elac protein family. *J Biol Chem* 2005, **280**:17857-17862.
40. Karkashon S, Hopkinson A, Levinger L: tRNase Z catalysis and conserved residues on the carboxy side of the His cluster. *Biochemistry* 2007, **46**:9380-9387.
41. Levinger L, Hopkinson A, Desetty R, Wilson C: Effect of changes in the flexible arm on tRNase Z processing kinetics. *J Biol Chem* 2009, **284**:15685-15691.
42. Wilusz JE, Freier SM, Spector DL: 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* 2008, **135**:919-932.
43. Haussecker D, Huang Y, Lau A, Parameswaran P, Fire AZ, Kay MA: Human tRNA-derived small RNAs in the global regulation of RNA silencing. *RNA* 2010, **16**:673-695.
44. Lee YS, Shibata Y, Malhotra A, Dutta A: A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev* 2009, **23**:2639-2649.
45. Bogerd HP, Karnowski HW, Cai X, Shin J, Pohlers M, Cullen BR: A mammalian herpesvirus uses noncanonical expression and processing mechanisms to generate viral MicroRNAs. *Mol Cell* 2010, **37**:135-142.
46. Diebel KW, Smith AL, van Dyk LF: Mature and functional viral miRNAs transcribed from novel RNA polymerase III promoters. *RNA* 2010, **16**:170-185.
47. Korver W, Guevara C, Chen Y, Neuteboom S, Bookstein R, Tavtigian S, Lees E: The product of the candidate prostate cancer susceptibility gene ELAC2 interacts with the gamma-tubulin complex. *Int J Cancer* 2003, **104**:283-288.
48. Noda D, Itoh S, Watanabe Y, Inamitsu M, Dennler S, Itoh F, Koike S, Danielpour D, Ten Dijke P, Kato M: ELAC2, a putative prostate cancer susceptibility gene product, potentiates TGF-beta/Smad-induced growth arrest of prostate cells. *Oncogene* 2006, **25**:5591-5600.
49. Mineri R, Pavelka N, Fernandez-Vizarra E, Ricciardi-Castagnoli P, Zeviani M, Tiranti V: How do human cells react to the absence of mitochondrial DNA? *PLoS one* 2009, **4**:e5713.
50. Peng WT, Robinson MD, Mnaimneh S, Krogan NJ, Cagney G, Morris Q, Davierwala AP, Grigull J, Yang X, Zhang W, Mitsakakis N, Ryan OW, Datta N, Jojic V, Pal C, Canadien V, Richards D, Beattie B, Wu LF, Altschuler SJ, Rowley S, Frey BJ, Emili A, Greenblatt JF, Hughes TR: A panoramic view of yeast noncoding RNA processing. *Cell* 2003, **113**:919-933.
51. Zhao W, Yu H, Li S, Huang Y: Identification and analysis of candidate fungal tRNA 3'-end processing endonucleases tRNase Zs, homologs of the putative prostate cancer susceptibility protein ELAC2. *BMC Evol Biol* 2010, **10**:272.
52. Vogel A, Schilling O, Spath B, Marchfelder A: The tRNase Z family of proteins: physiological functions, substrate specificity and structural properties. *Biol Chem* 2005, **386**:1253-1264.
53. Jarvis P, Robinson C: Mechanisms of protein import and routing in chloroplasts. *Curr Biol* 2004, **14**:R1064-1077.
54. Pfanner N, Neupert W: The mitochondrial protein import apparatus. *Annu Rev Biochem* 1990, **59**:331-353.
55. Oommen A, Li XQ, Gegenheimer P: Cleavage specificity of chloroplast and nuclear tRNA 3'-processing nucleases. *Mol Cell Biol* 1992, **12**:865-875.
56. Adams KL, Wendel JF: Polyploidy and genome evolution in plants. *Curr Opin Plant Biol* 2005, **8**:135-141.
57. Takahashi M, Takaku H, Nashimoto M: Regulation of the human tRNase Z<sup>S</sup> gene expression. *FEBS Lett* 2008, **582**:2532-2536.
58. Mohan A, Whyte S, Wang X, Nashimoto M, Levinger L: The 3' end CCA of mature tRNA is an antideterminant for eukaryotic 3'-tRNase. *RNA* 1999, **5**:245-256.
59. Nashimoto M: Distribution of both lengths and 5' terminal nucleotides of mammalian pre-tRNA 3' trailers reflects properties of 3' processing endoribonuclease. *Nucleic Acids Res* 1997, **25**:1148-1154.
60. Pellegrini O, Nezzar J, Marchfelder A, Putzer H, Condon C: Endonucleolytic processing of CCA-less tRNA precursors by RNase Z in *Bacillus subtilis*. *EMBO J* 2003, **22**:4534-4543.
61. Schiffer S, Rosch S, Marchfelder A: Recombinant RNase Z does not recognize CCA as part of the tRNA and its cleavage efficiency is influenced by acceptor stem length. *Biol Chem* 2003, **384**:333-342.
62. Minagawa A, Ishii R, Takaku H, Yokoyama S, Nashimoto M: The flexible arm of tRNase Z is not essential for pre-tRNA binding but affects cleavage site selection. *J Mol Biol* 2008, **381**:289-299.

63. Hopkinson A, Levinger L: **Effects of conserved D/T loop substitutions in the pre-tRNA substrate on tRNase Z catalysis.** *RNA Biol* 2008, **5**:104-111.
64. Ishii R, Minagawa A, Takaku H, Takagi M, Nashimoto M, Yokoyama S: **The structure of the flexible arm of *Thermotoga maritima* tRNase Z differs from those of homologous enzymes.** *Acta Crystallogr Sect F Struct Biol Cryst Commun* 2007, **63**:637-641.
65. Michaud M, Cognat V, Duchene AM, Marechal-Drouard L: **A global picture of tRNA genes in plant genomes.** *Plant J* 2010, **66**:80-93.
66. Pfitzinger H, Guillemaut P, Weil JH, Pillay DT: **Adjustment of the tRNA population to the codon usage in chloroplasts.** *Nucleic Acids Res* 1987, **15**:1377-1386.
67. Ceballos-Chavez M, Vioque A: **Sequence-dependent Cleavage Site Selection by RNase Z from the *Cyanobacterium Synechocystis* sp. PCC 6803.** *J Biol Chem* 2005, **280**:33461-33469.
68. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23**:2947-2948.
69. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.** *Mol Biol Evol* 2011.
70. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21**:2104-2105.
71. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
72. Ezraty B, Dahlgren B, Deutscher MP: **The RNase Z homologue encoded by *Escherichia coli* *elaC* gene is RNase BN.** *J Biol Chem* 2005, **280**:16542-16545.
73. Initiative TAG: **Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*.** *Nature* 2000, **408**:796-815.
74. Initiative TIB: **Genome sequencing and analysis of the model grass *Brachypodium distachyon*.** *Nature* 2006, **443**:763-768.
75. Ming R, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KL, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice DW, Albert H, Wang ML, Zhu YJ, *et al*: **The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus).** *Nature* 2008, **452**:991-996.
76. Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, *et al*: **The genome of the cucumber, *Cucumis sativus* L.** *Nat Genet* 2009, **41**:1275-1281.
77. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, *et al*: **Genome sequence of the *Palaeopolyploid* soybean.** *Nature* 2006, **443**:178-183.
78. Project IRGS: **The map-based sequence of the rice genome.** *Nature* 2005, **436**:793-800.
79. Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, *et al*: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray).** *Science* 2006, **313**:1596-1604.
80. Chan AP, Crabtree J, Zhao Q, Lorenzi H, Orvis J, Puiu D, Melake-Berhan A, Jones KM, Redman J, Chen G, Cahoon EB, Gedil M, Stanke M, Haas BJ, Wortman JR, Fraser-Liggett CM, Ravel J, Rabinowicz PD: **Draft genome sequence of the oilseed species *Ricinus communis*.** *Nat Biotechnol* 2008, **28**:951-956.
81. Dust AN, Kellogg EA, Devos KM, Bennetzen JL: **Foxtail millet: a sequence-driven grass model system.** *Plant Physiol* 2009, **149**:137-141.
82. Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberger G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Ottillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, *et al*: **The *Sorghum bicolor* genome and the diversification of grasses.** *Nature* 2009, **457**:551-556.
83. Velasco R, Zharkikh A, Troggo M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Dematte L, Mraz A, *et al*: **A high quality draft consensus sequence of the genome of a heterozygous grapevine variety.** *PLoS one* 2007, **2**:e1326.
84. Schnable PS, Ware W, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, *et al*: **The B73 maize genome: complexity, diversity, and dynamics.** *Science* 2009, **326**:1112-1115.

doi:10.1186/1471-2148-11-219

**Cite this article as:** Fan *et al*: A survey of green plant tRNA 3'-end processing enzyme tRNase Zs, homologs of the candidate prostate cancer susceptibility protein ELAC2. *BMC Evolutionary Biology* 2011 **11**:219.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

