Bioinformatic Analysis Reveals Archaeal tRNA$^{\text{Tyr}}$ and tRNA$^{\text{Trp}}$ Identities in Bacteria

Takahito Mukai [1], Noah M. Reynolds [1], Ana Crnković [1] and Dieter Söll [1,2,*]

[1] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA;
     takahito.mukai@yale.edu (T.M.); noah.reynolds@yale.edu (N.M.R.); ana.crnkovic@yale.edu (A.C.)
[2] Department of Chemistry, Yale University, New Haven, CT 06520, USA
[*] Correspondence: dieter.soll@yale.edu; Tel.: +1-203-432-6200; Fax: +1-203-432-6202

Abstract: The tRNA identity elements for some amino acids are distinct between the bacterial and archaeal domains. Searching in recent genomic and metagenomic sequence data, we found some candidate phyla radiation (CPR) bacteria with archaeal tRNA identity for Tyr-tRNA and Trp-tRNA synthesis. These bacteria possess genes for tyrosyl-tRNA synthetase (TyrRS) and tryptophanyl-tRNA synthetase (TrpRS) predicted to be derived from DPANN superphylum archaea, while the cognate tRNA$^{\text{Tyr}}$ and tRNA$^{\text{Trp}}$ genes reveal bacterial or archaeal origins. We identified a trace of domain fusion and swapping in the archaeal-type TyrRS gene of a bacterial lineage, suggesting that CPR bacteria may have used this mechanism to create diverse proteins. Archaeal-type TrpRS of bacteria and a few TrpRS species of DPANN archaea represent a new phylogenetic clade (named TrpRS-A). The TrpRS-A open reading frames (ORFs) are always associated with another ORF (named ORF1) encoding an unknown protein without global sequence identity to any known protein. However, our protein structure prediction identified a putative HIGH-motif and KMSKS-motif as well as many α-helices that are characteristic of class I aminoacyl-tRNA synthetase (aaRS) homologs. These results provide another example of the diversity of molecular components that implement the genetic code and provide a clue to the early evolution of life and the genetic code.

Keywords: tRNA; aaRS; genetic code; evolution; lateral gene transfer

1. Introduction

Bacteria, archaea and eukarya share the standard genetic code, which suggests that they share a universal common ancestor (LUCA). However, the molecular systems underlying the standard genetic code are not completely conserved between all domains of life. In aminoacyl-tRNA synthesis, several elements of tRNA such as the anticodon sequence, other nucleotide residues, post-transcriptional modifications, and local and global tertiary structures are recognized by the cognate aminoacyl-tRNA synthetase (aaRS) [1–3]. While it is known that in all domains of life the anticodon sequences of tRNA$^{\text{Tyr}}$ and tRNA$^{\text{Trp}}$ are recognized by their cognate aaRSs, the other major identity elements of tRNA$^{\text{Tyr}}$ and tRNA$^{\text{Trp}}$ are distinct between the bacterial domain and the archaeal and eukaryotic domains [1,2,4,5] (Figure 1A). In bacteria, tRNA$^{\text{Tyr}}$ contains a G1-C72 base pair and a variable arm (V-arm) that is recognized by the additional *C*-terminal S4-like domain of bacteria-type TyrRS (Figure 1A,B) [6]. On the other hand, archaea and eukaryotes encode a tRNA$^{\text{Tyr}}$ lacking the V-arm and containing a C1-G72 base pair (Figure 1A) [6]. For tRNA$^{\text{Trp}}$, bacteria encode a tRNA$^{\text{Trp}}$ with G73, whereas archaea and eukaryotes have a tRNA$^{\text{Trp}}$ with A73 and a G1-C72 base pair (Figure 1A) [7]. Thus, unlike the other aaRS species, archaeal and eukaryotic TyrRS and TrpRS have not been found in the bacterial domain [6,8–15].
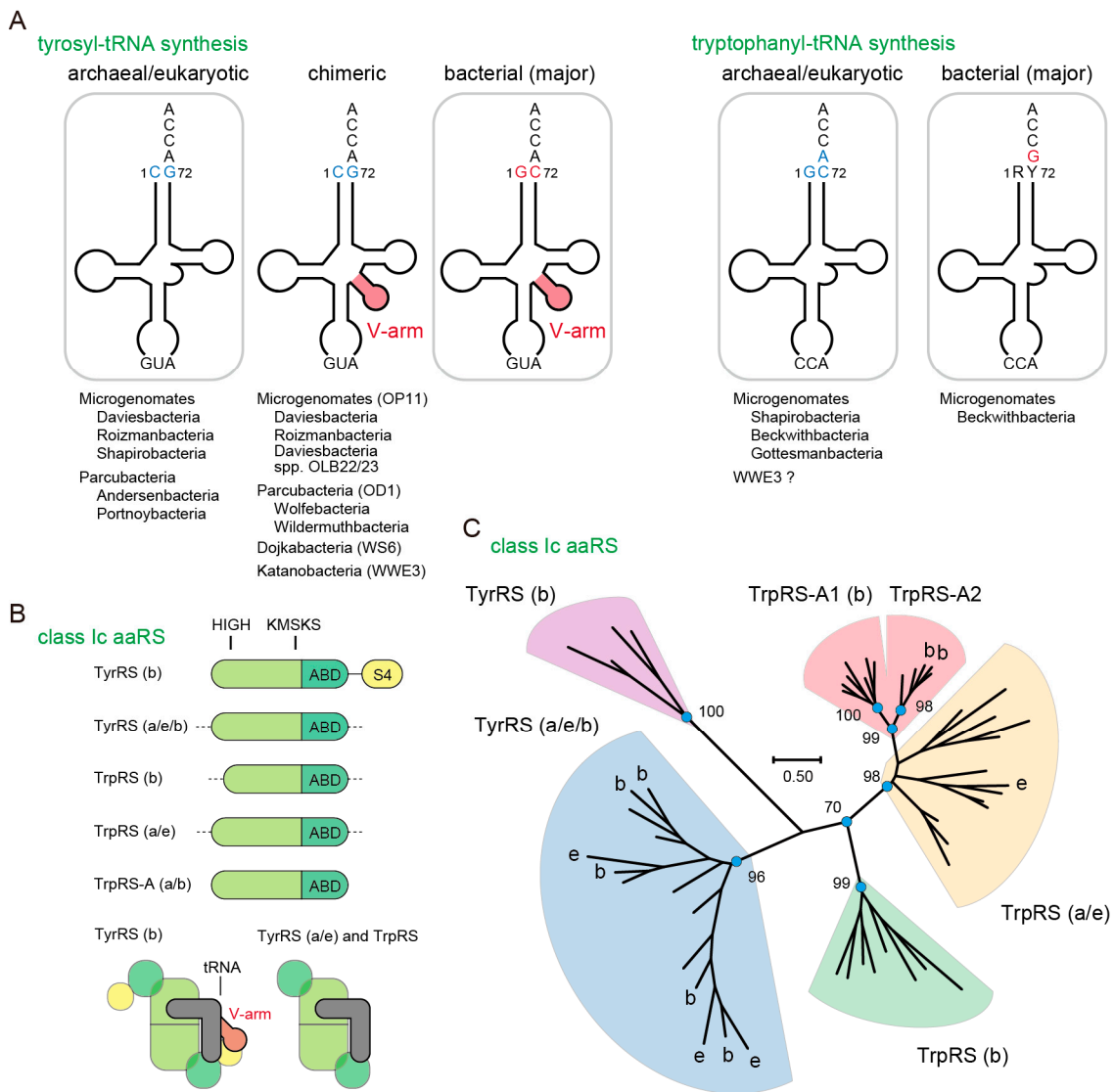
Figure 1. Bacteria with an archaeal tRNA$^{\text{Tyr}}$ or tRNA$^{\text{Trp}}$ identity. (**A**) The major identity elements for tyrosyl-tRNA synthetase (TyrRS) and tryptophanyl-tRNA synthetase (TrpRS) are shown in blue and red. Diverse subgroups of the Candidate Phyla Radiation (CPR) have archaea-like tRNA$^{\text{Tyr}}$ and tRNA$^{\text{Trp}}$ genes, as shown below the tRNA structures. Ca. Beckwithbacteria bacterium RBG_13_42_9 has both bacterial and archaeal pairs of TrpRS•tRNA$^{\text{Trp}}$; (**B**) Domain structures of the class Ic aminoacyl-tRNA synthetase (aaRS) family (homodimer). The S4-like domain binds to the V-arm of bacterial tRNA$^{\text{Tyr}}$. ABD denotes anticodon-binding domain. TrpRS-A is a newly identified TrpRS homolog. Class Ic aaRS is known to form a homodimer (in a few cases pseudo-homodimer) and binds to one tRNA molecule at a time (half-of-the-sites). Bacterial, archaeal and eukaryotic origins are indicated with b, a and e, respectively; (**C**) Phylogenetic analysis of the class Ic aaRS family. Maximum likelihood bootstrap values (100 replicates) are shown. The TrpRS-A species are split into two clades. The TrpRS-A2 proteins may chelate a [4Fe–4S] cluster.

A fundamental question in understanding the evolution of the genetic code is whether tRNA identities were established at the time of LUCA, and if so, which tRNA identity set was used [13,16]. The discovery of bacteria with an archaeal tRNA identity would provide support for the hypothesis that archaeal tRNA identity sets may have been used in LUCA. A clue to the answer to this question was provided by two synthetic biology studies [17,18]. The artificial gene transfer of an archaeal or eukaryotic TyrRS or TrpRS gene to *Escherichia coli* was successfully achieved by the simultaneous

transfer of an archaeal or eukaryotic tRNA gene [17,18]. These heterologous aaRS•tRNA pairs functionally replaced the endogenous aaRS•tRNA pairs in *E. coli*. Thus, it can be hypothesized that the archaeal tRNA^{Tyr} and tRNA^{Trp} identities might have been used in LUCA.

Inspired by these studies, we carefully re-investigated the phylogenetic distribution of TyrRS and TrpRS. In the present study, archaea-, eukarya-, and bacteria-type is used to indicate the canonical archaea-, eukarya-, or bacteria-type aaRS, respectively, independent of the organism in which the enzyme is identified. Surprisingly, we found a putative bacterial species annotated to have a eukarya-type TyrRS gene (EKE14628.1) [19]. This Ca. Roizmanbacteria bacterium belongs to the candidate phyla radiation (CPR) composed of diverse uncultured bacteria which are often symbiotic with DPANN archaea [20–24]. Although composite genomes of CPR bacteria and DPANN archaea are sometimes contaminated by DPANN archaeal genomes and CPR bacterial genomes, respectively, a recent study was able to identify archaea-like form II/III RubisCO genes in CPR bacteria [25]. These findings prompted us to search for archaea/eukarya-type Tyr- and Trp-encoding systems in bacteria.

2. Materials and Methods

Bioinformatics

Archaeal and eukaryotic TyrRS genes and non-canonical bacteria-type TyrRS genes were collected in three steps. First, TyrRS genes of representative archaea and eukaryotes were collected by a keyword search (tyrosyl/tryptophanyl-tRNA synthetase; tyrosine/tryptophan–tRNA ligase) and a BLASTp search in the National Center for Biotechnology Information (NCBI) database. Next, archaea/eukarya-type TyrRS genes in the bacterial domain were collected by a BLASTp search in the NCBI database and manually curated. Lastly, TyrRS amino acid sequences which showed about ≥40% similarity with a query sequence (GenBank: KKM02188.1) were collected from all genome, metagenome and metatranscriptome protein sequence datasets in the Integrated Microbial Genomes (IMG) system [26] (last update September 2016). The reason for employing KKM02188.1 was to find bacteria whose TyrRSs resemble opisthokontal (fungal and animal) TyrRSs. The query protein belongs to an unknown fungus in a marine sediment metagenome and shows good similarity (41%–51%) to both opisthokontal and Daviesbacteria GW2011_GWA1_38_7 TyrRS species.

Non-canonical TrpRS genes in the bacterial domain were first identified by a BLASTp search in the NCBI database using the *Pyrococcus horikoshii* TrpRS (UniProtKB: O59584.2) as query. Non-canonical TrpRS genes which showed >40% amino acid similarity with Ca. Beckwithbacteria bacterium RBG_13_42_9 TrpRS-A were collected by BLASTp searches in the NCBI database and IMG's groundwater metagenome datasets. The obtained protein sequences were analyzed by Clustal X 2.1 [27] (for rough alignment), SeaView ver 4.0 [28] (for manual curation), MEGA7 [29] with the default settings (Maximum Likelihood, JTT model, Uniform rates, Use all Gaps/Missing sites, for phylogeny estimation), and BoxShade Server ver. 3.21 (for visualization). Multiple sequence alignment analyses by Clustal X were followed by a manual curation based on the reported structure-based alignments of TyrRS and TrpRS [30–35]. For the phylogenetic analyses of class Ic aaRSs, *N*- and *C*-terminal protein sequences were trimmed and nonconserved insertion sequences removed; up to 13 residues upstream of the HIGH motif and to the end of the anticodon binding domain were included in the analyses. Protein two- and three-dimentional structural prediction was performed using JPred 4 [36] and SWISS-MODEL [37], respectively. tRNA^{Tyr} and tRNA^{Trp} sequences were identified by a BLASTn search using automatically annotated tRNA^{Tyr} and tRNA^{Trp} sequences as queries.

3. Results

3.1. Identification of Non-Canonical Class Ic aaRS Sequences

We found tRNA^{Tyr} with C1-G72 and archaea/eukarya-type TyrRS genes in diverse subgroups of the Parcubacteria (OD1), Microgenomates (OP11), Dojkabacteria (WS6) and Katanobacteria (WWE3) phyla in CPR [22,25] (Figure 1A). In many cases, the CPR tRNA^{Tyr} species with C1-G72 contain a V-arm, indicating that these non-canonical tRNA^{Tyr} species are derived from bacterial tRNA^{Tyr} with a V-arm. Both a V-arm-containing and a V-arm-lacking tRNA^{Tyr} with C1-G72 are found in Ca. Roizmanbacteria bacterium GW2011_GWC2_34_23. On the other hand, tRNA^{Trp} with A73 and archaea/eukarya-type TrpRS genes are found in a few Microgenomates bacteria (Figure 1A). Both a bacterial tRNA^{Trp} with G73 and a tRNA^{Trp} with A73 exist in Ca. Bechwithbacteria bacterium RBG_13_42_9. We named these non-canonical TrpRS species as TrpRS-A (Figure 1B). Interestingly, TrpRS-A is slightly different from the canonical archaeal TrpRS species, but highly similar to minor DPANN archaeal TrpRS species (also named TrpRS-A) found only in the groundwater metagenomes (Figure 1C). We identified only 12 instances of TrpRS-A genes in total, suggesting an infrequent distribution of these genes in nature. These bacterial and archaeal TrpRS-A species form a terminal clade within the archaeal/eukaryotic TrpRS clade and can be grouped into two sub-clades (TrpRS-A1 and TrpRS-A2) (Figure 1C). The TrpRS-A1 proteins appear to be restricted within CPR bacteria. The TrpRS-A2 proteins are predicted to chelate a [4Fe–4S] cluster through their four cysteine residues, like some bacterial and archaeal TrpRS proteins having a C-x22-C-x6-C-x2-C motif [38].

3.2. Archaea/Eukarya-Type TyrRS in the Bacterial Domain

We then investigated the collected TyrRS sequences. A high-resolution phylogeny for archaea/eukarya-type TyrRS suggests that several lineages of CPR bacteria independently obtained an archaea/eukarya-type TyrRS gene from archaea (Figure 2). Alternatively, lateral gene transfer (LGT) might have occurred from bacteria to archaea and other groups of bacteria. Interestingly, one bacterial TyrRS sequence (3300007427.a:Ga0100483_102719) is highly similar to the TyrRS sequences of *Acanthamoeba castellanii* and Pandoraviruses (the Eukarya domain) (see the Ultra-High resolution region in Figure 2). Thus, LGT between bacteria and amoeba or giant viruses can be reasonably assumed. This Pandoravirus-like TyrRS, as well as two other TyrRSs derived from WS6 bacterium GW2011 GWA2_37_6 [24,39] and an active sludge metagenome, possess a B2 domain of bacterial phenylalanyl-tRNA synthetase β-subunit (PheRSβ) [40]. This domain is fused to the *N*-terminus of the TyrRS by a long α-helix (Figure 3A). The B2 domain belongs to the RNA-binding OB folds, but it is missing in many CPR-bacterial PheRSβ (for example, OGE14653.1). We also found that a few Microgenomates lysyl-tRNA synthetases (KKR67068.1 & KKQ91124.1) have an additional *C*-terminal domain that is very similar to this B2 domain and predicted α-helix. It is known that aaRS proteins are often fused with an OB domain [41]. The two Klosterneuburg active sludge metagenomic contigs containing the B2-TyrRS fusion genes showed almost identical gene organization and gene sequences, indicating that these two contigs belong to two closely-related bacteria. However, only the TyrRS 'domain' sequences are different in terms of sequence similarity (Figure 3A,B). While one is Pandoravirus-like, as previously mentioned, the other is most similar to the TyrRS 'domain' sequence of the B2-TyrRS species of WS6 bacterium GW2011 GWA2_37_6 (Figures 2 and 3A). A possible explanation for this is that the TyrRS 'domain' sequence was replaced with the Pandoravirus-like sequence in a bacterial lineage (Figure 3B, the upper contig). This finding will help us understand aaRS evolution through domain fusion and swapping.

Figure 2. A phylogenetic tree for archaea/eukarya-type TyrRS. Bootstrap values (%) are shown for the rooted Maximum Likelihood tree made with 100 replicates using MEGA7. The 'Ultra-High resolution' region shows almost all TyrRS sequences identified by the comprehensive genome/metagenome/metatranscriptome analysis using gi|816604452|gb|KKM02188.1| as query for BLASTp. The archaeal species in the Ultra-High resolution region may belong to the DPANN superphylum according to the Joint Genome Institute's annotation pipeline and our manual annotation. The opisthokontal (fungal and animal) TyrRS clade is marked with a green box. We chose a few representative TyrRS sequences for each major bacterial group (Roizmanbacteria, Daviesbacteria, Shapirobacteria, Wolfebacteria and Andersenbacteria) after confirming the sequence similarity within the same group. In contrast, we identified three orphan TyrRS genes belonging to bacteria in the Ultra-High resolution region. TrpRS sequences of *Thermus thermophilus* (bacteria) and *Pyrococcus horikoshii* (archaea) were used as an outgroup.
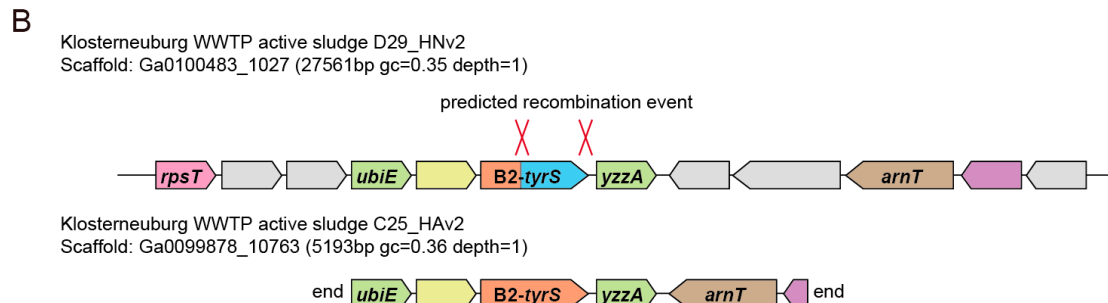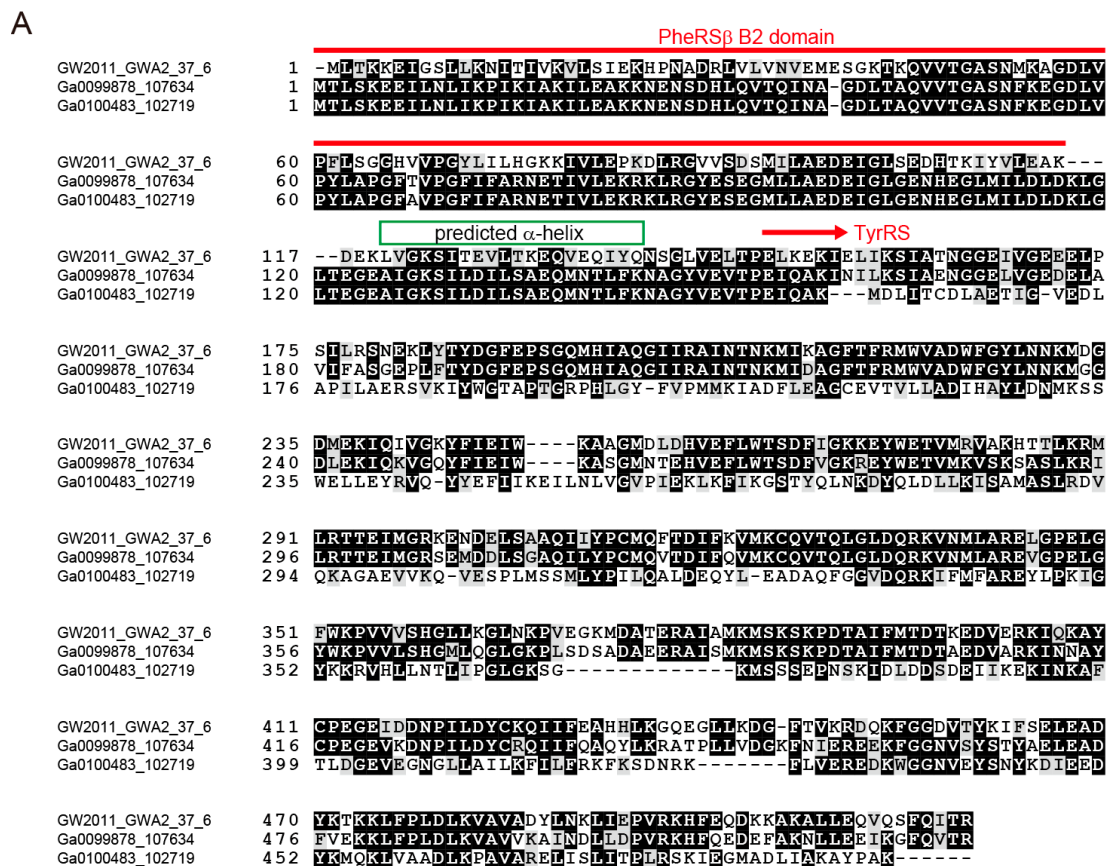
A

PheRSβ B2 domain

```
GW2011_GWA2_37_6    1   -MLTKKEIGSLLKNITIVKVLSIEKHPNADRLVLVNVEMESGKTKQVVTGASNMKAGDLV
Ga0099878_107634    1   MTLSKEEILNLIKPIKIAKILEAKKNENSDHLQVTQINA-GDLTAQVVTGASNFKEGDLV
Ga0100483_102719    1   MTLSKEEILNLIKPIKIAKILEAKKNENSDHLQVTQINA-GDLTAQVVTGASNFKEGDLV

GW2011_GWA2_37_6   60   PFLSGGHVVPGYLILHGKKIVLEPKDLRGVVSDSMILAEDEIGLSEDHTKIYVLEAK---
Ga0099878_107634   60   PYLAPGFTVPGFIFARNETIVLEKRKLRGYESEGMLLAEDEIGLGENHEGLMILDLDKLG
Ga0100483_102719   60   PYLAPGFAVPGFIFARNETIVLEKRKLRGYESEGMLLAEDEIGLGENHEGLMILDLDKLG
```

predicted α-helix → TyrRS

```
GW2011_GWA2_37_6  117   --DEKLVGKSITEVLTKEQVEQIYQNSGLVELTPELKEKIELIKSIATNGGEIVGEEELP
Ga0099878_107634  120   LTEGEAIGKSILDILSAEQMNTLFKNAGYVEVTPEIQAKINILKSIAENGGELVGEDELA
Ga0100483_102719  120   LTEGEAIGKSILDILSAEQMNTLFKNAGYVEVTPEIQAK---MDLITCDLAETIG-VEDL

GW2011_GWA2_37_6  175   SILRSNEKLYTYDGFEPSGQMHIAQGIIRAINTNKMIKAGFTFRMWVADWFGYLNNKMDG
Ga0099878_107634  180   VIFASGEPLFTYDGFEPSGQMHIAQGIIRAINTNKMIDAGFTFRMWVADWFGYLNNKMGG
Ga0100483_102719  176   APILAERSVKIYWGTAPTGRPHLGY-FVPMMKIADFLEAGCEVTVLLADIHAYLDNMKSS

GW2011_GWA2_37_6  235   DMEKIQIVGKYFIEIW----KAAGMDLDHVEFLWTSDFIGKKEYWETVMRVAKHTTLKRM
Ga0099878_107634  240   DLEKIQKVGQYFIEIW----KASGMNTEHVEFLWTSDFVGKREYWETVMKVSKSASLKRI
Ga0100483_102719  235   WELLEYRVQ-YYEFIIKEILNLVGVPIEKLKFIKGSTYQLNKDYQLDLLKISAMASLRDV

GW2011_GWA2_37_6  291   LRTTEIMGRKENDELSAAQIIYPCMQFTDIFKVMKCQVTQLGLDQRKVNMLARELGPELG
Ga0099878_107634  296   LRTTEIMGRSEMDDLSGAQILYPCMQVTDIFQVMKCQVTQLGLDQRKVNMLAREVGPELG
Ga0100483_102719  294   QKAGAEVVKQ-VESPLMSSMLYPIILQALDEQYL-EADAQFGGVDQRKIFMFAREYLPKIG

GW2011_GWA2_37_6  351   FWKPVVVSHGLLKGLNKPVEGKMDATERAIAMKMSKSKPDTAIFMTDTKEDVERKIQKAY
Ga0099878_107634  356   YWKPVVLSHGMLQGLGKPLSDSADAEERAISMKMSKSKPDTAIFMTDTAEDVARKINNAY
Ga0100483_102719  352   YKKRVHLLNTLIPGLGKSG------------KMSSSEPNSKIDLDDSDEIIKEKINKAF

GW2011_GWA2_37_6  411   CPEGEIDDNPILDYCKQIIFEAHHLKGQEGLLKDG-FTVKRDQKFGGDVTYKIFSELEAD
Ga0099878_107634  416   CPEGEVKDNPILDYCRQIIFQAQYLKRATPLLVDGKFNIEREEKFGGNVSYSTYAELEAD
Ga0100483_102719  399   TLDGEVEGNGLLAILKFILFRKFKSDNRK------FLVEREDKWGGNVEYSNYKDIEED

GW2011_GWA2_37_6  470   YKTKKLFPLDLKVAVADYLNKLIEPVRKHFEQDKKAKALLEQVQSFQITR
Ga0099878_107634  476   FVEKKLFPLDLKVAVVKAINDLLDPVRKHFQEDEFAKNLLEEIKGEQVTR
Ga0100483_102719  452   YKMQKLVAADLKPAVARELISLITPLRSKIEGMADLIAKAYPAK------
```

B

Klosterneuburg WWTP active sludge D29_HNv2
Scaffold: Ga0100483_1027 (27561bp gc=0.35 depth=1)

predicted recombination event

rpsT | ubiE | B2-tyrS | yzzA | arnT

Klosterneuburg WWTP active sludge C25_HAv2
Scaffold: Ga0099878_10763 (5193bp gc=0.36 depth=1)

end | ubiE | B2-tyrS | yzzA | arnT | end

Figure 3. The TyrRS species with a *N*-terminal B2 domain fusion. (**A**) Multiple sequence alignment of the three B2-TyrRS proteins; (**B**) The genomic loci of the two B2-TyrRS genes in active sludge metagenomes. Predicted recombination sites are indicated.

3.3. Non-Canonical TrpRS Species and Their Associating Proteins

We then investigated the genetic loci of the TrpRS-A genes (Figure 4A). In bacteria, tRNA^{Trp} (A73) and TrpRS-A comprise an operon. Interestingly, this operon is either followed by a bacteria-type TrpRS gene or headed by a *trp* repressor gene (Figure 4A). This is consistent with Ca. Beckwithbacteria bacterium RBG_13_42_9 possessing a bacterial tRNA^{Trp} (G73) somewhere else in the genome. The observed operon structures indicate that the tRNA^{Trp} (A73) and TrpRS-A1 genes may be regulated by tryptophan availability and could coexist with a canonical bacterial TrpRS•tRNA^{Trp} pair. In archaea, one or two tRNA^{Trp} (A73) genes, one or two TrpRS-A2 genes and another gene encoding a TrpRS-A homolog (named TrpRS-A-like) are found in addition to a single canonical archaeal TrpRS gene. TrpRS-A-like has several insertions and deletions compared to archaea/eukarya-type TrpRS and TrpRS-A. Thus, TrpRS-A and TrpRS-A-like coexist with a canonical archaeal TrpRS in archaea.

Interestingly, TrpRS-A2 and TrpRS-A-like genes are also found in putative bacterial metagenomic contigs (Figure 4A), implying multiple LGT events.
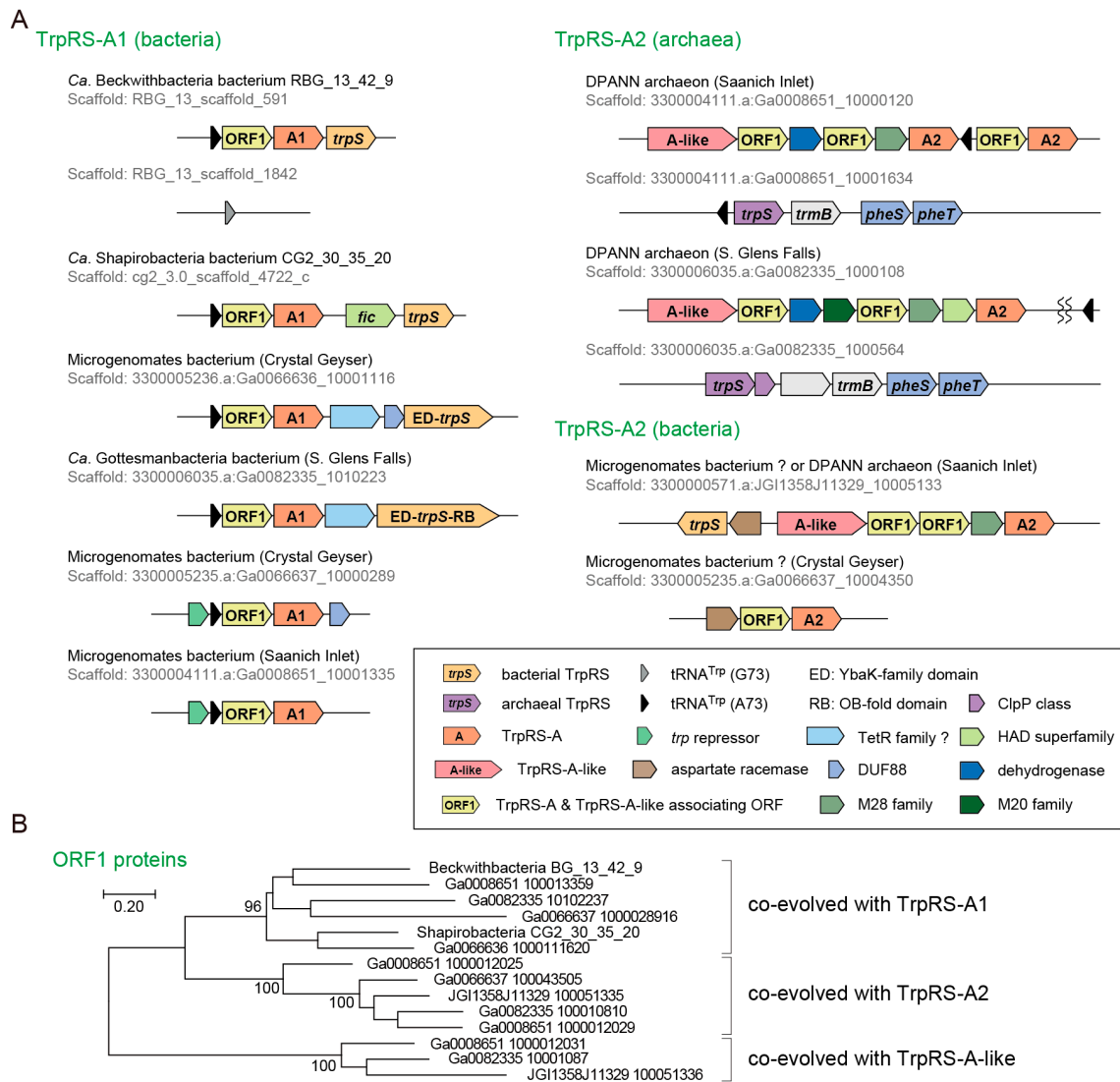

Figure 4. Non-canonical TrpRS species. (**A**) The genetic loci and the operon structures of TrpRS-A genes. The origins of these uncultured organisms are described in the parentheses and indicated with "CG" (Crystal Geyser groundwater) and "RBG" (Rifle BackGround groundwater). In a few Microgenomates species, TrpRS is fused with small proteins; (**B**) Co-evolution of the ORF1 genes with the TrpRS-A and TrpRS-A-like genes. Bootstrap values (%) are shown for the unrooted Maximum Likelihood tree made with 100 replicates using MEGA7.

We found an unusual open reading frame (named as ORF1) between the tRNA^{Trp} (A73) and TrpRS-A1 genes in the bacterial operons and in the gene clusters of the TrpRS-A2 and TrpRS-A-like in archaea (Figure 4A). The ORF1 is completely conserved and appears to have co-evolved with the TrpRS-A1, TrpRS-A2 and TrpRS-A-like ORFs (Figure 4B). ORF1 sequences show no significant similarity with any known protein (less than score 40 in NCBI BLASTp searches). However, the predicted structure of the ORF1 proteins includes HIGH-motif-like and KMSKS-motif-like motifs [2], as well as many α-helices, thereby suggesting that these proteins might be structural homologs of class I aaRSs (Figure 5). Seventy residues from the putative HIGH motif of ORF1 matched the corresponding region of cysteinyl-tRNA synthetase (a class Ia aaRS) with 16.9% sequence identity

in our SWISS-MODEL prediction. However, in the ORF1 structure prediction, the Rossmann-fold domain is missing [42], indicating a function distinct from class I aaRS or class I aaRS homolog that synthesizes cyclo(L-Trp-L-Trp) using tryptophanyl-tRNA$^{\mathrm{Trp}}$ [43]. Interestingly, both TrpRS-A1 and ORF1 pairs and bacteria-type TrpRSs of these bacteria lack any tryptophan residues (see Figure 5 for the ORF1 cases). This might imply that tryptophan may be limited for these bacteria living in groundwater environments.
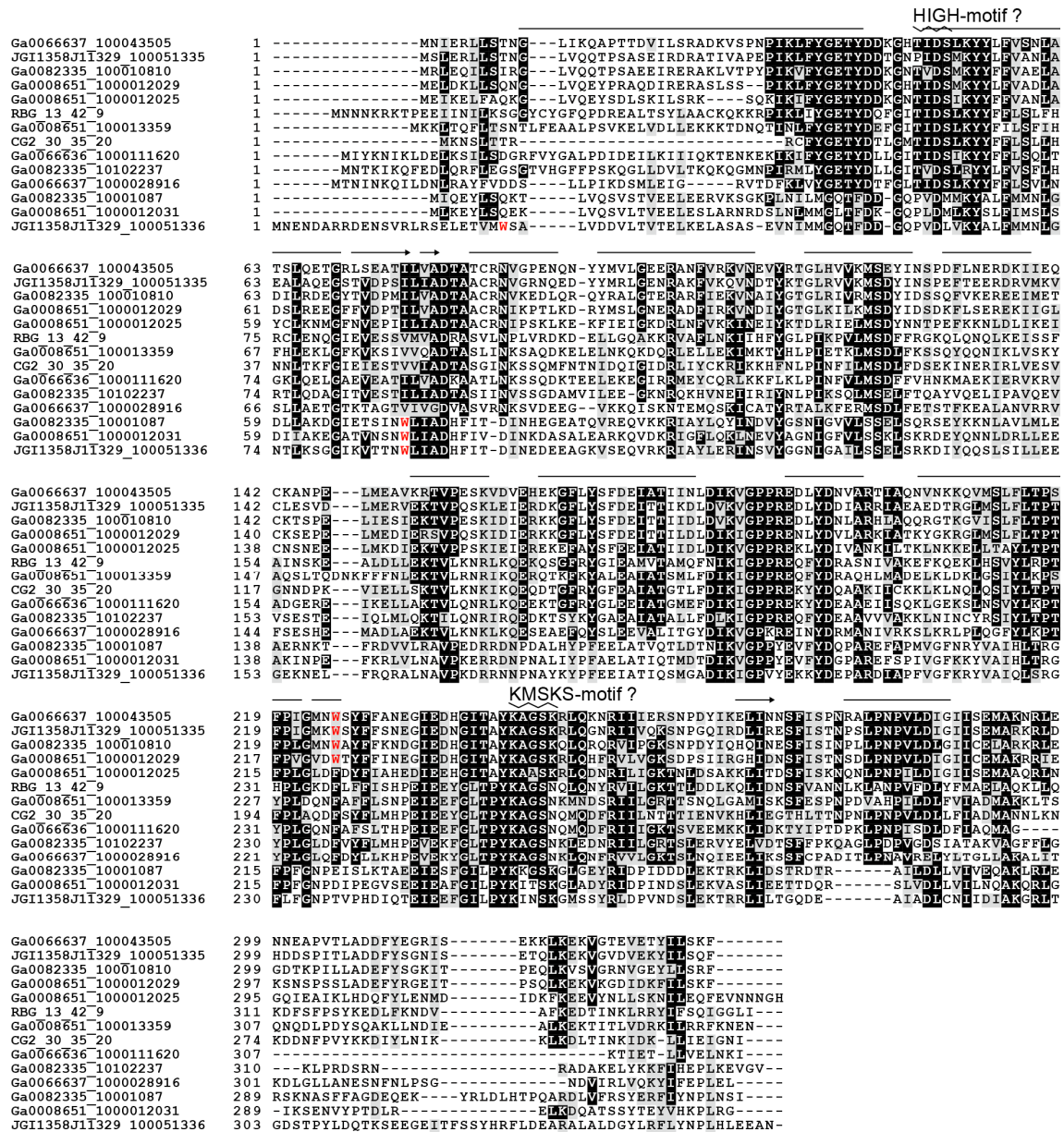
Figure 5. Multiple pairwise alignment of the ORF1 proteins and their structural prediction by SWISS-MODEL. Bars and arrows above the amino acid sequences represent predicted α-helices and β-strands, respectively, whereas zigzags indicate predicted HIGH and KMSKS motifs. The predicted overall structure suggests that the ORF1 protein might be a homolog of class I aaRS. Tryptophan (W) residues are shown as red letters.

4. Discussion

It was thought that LGT of some archaeal/eukaryotic aaRS genes to bacteria would be prevented by the difference in tRNA identity rules. A previous bioinformatic study detected a case of eukaryotic-like histidine (His) tRNA identity (i.e., A73) in certain α-proteobacteria [12]. These α-proteobacterial tRNA^{His} species have A73 and lack G-1 [44]. However, a subsequent biochemical study revealed A73 to be a minor identity element of the *Caulobacter crescentus* histidyl-tRNA synthetase (HisRS) [45]. A recent comprehensive bioinformatic study did not support any LGT event of HisRS from eukaryotes to α-proteobacteria [8], suggesting that the eukaryotic-like tRNA^{His} identity in certain α-proteobacteria might be a result of convergent evolution. Similarly, it is known that some mitochondrial aaRSs violate the bacterial identity rules. For example, human mitochondrial TyrRS charges mitochondrial tRNA^{Tyr} species with G1-C72 (wild-type) or C1-G72 (mutant) with the same efficiency [46]. The V-arm is missing in the human mitochondrial tRNA^{Tyr}, whereas the human mitochondrial TyrRS retains the S4-like domain [47].

In contrast to these previous findings, our results provide the first evidence that archaeal or eukaryotic TyrRS and TrpRS genes exist in a few lineages of bacteria that have tRNA^{Tyr} or tRNA^{Trp} species with archaeal and eukaryotic identity elements. Intimate relationships between CPR bacteria and DPANN archaea may have facilitated LGT of TyrRS and TrpRS genes. Furthermore, our high-resolution TyrRS phylogeny and another comprehensive study [8] revealed that the LGT of TyrRS may have occurred several times among DPANN archaea, CPR bacteria, eukaryotes and giant viruses (Figure 2). Taken together, our data make clear that tRNA identities may not be hardwired to each domain of life.

The functions of TrpRS-A, TrpRS-A-like and the ORF1 proteins remain unclear. Their operon structures suggest that they are involved in tryptophan metabolism and encoding, rather than producing complex antibiotics, plant toxins and peptidoglycans [43,48]. Since TrpRS-A and TrpRS-A-like genes appear to occur in addition to the canonical TrpRS gene in both bacteria and archaea, TrpRS-A and TrpRS-A-like genes might encode additional copies of TrpRS that display a higher or lower specificity for tryptophan than the canonical TrpRS in order to confer antibiotic resistance [49] or to cope with stress [50]. The ORF1 proteins might bind to the homodimers of TrpRS-A and TrpRS-A-like in order to be stabilized in a complex [51]. Another possibility is that TrpRS-A and TrpRS-A-like proteins might form heterodimers with the partner ORF1 protein. It is known that some eukaryotes have a double-length TyrRS species forming a pseudo-dimer in which one of the "subunits" is catalytic but has lost the affinity for the tRNA^{Tyr} anticodon, whereas the other is non-catalytic but still recognizes the anticodon [52]. In addition, trans-oligomerization of duplicated threonyl-tRNA synthetases is known [50]. Future studies should elucidate the biochemical properties and the biological functions of these proteins.

Author Contributions: Takahito Mukai and Dieter Söll designed the project; Takahito Mukai performed bioinformatics; Takahito Mukai, Noah M. Reynolds and Ana Crnković investigated TrpRS-A1 and ORF1 proteins. All authors wrote the manuscript.

Abbreviations

The following abbreviations are used in this manuscript:

CPR	candidate phyla radiation
aaRS	aminoacyl-tRNA synthetase
TyrRS	tyrosyl-tRNA synthetase
TrpRS	tryptophanyl-tRNA synthetase
PheRS	phenylalanyl-tRNA synthetase
HisRS	histidyl-tRNA synthetase
ORF	open reading frame
LGT	lateral gene transfer

References

1. Giegé, R.; Sissler, M.; Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **1998**, *26*, 5017–5035. [CrossRef] [PubMed]
2. Ibba, M.; Francklyn, C.; Cusack, S.E. *The Aminoacyl-tRNA Synthetases*; Landes Biosciences: Georgetown, TX, USA, 2005.
3. Chaliotis, A.; Vlastaridis, P.; Mossialos, D.; Ibba, M.; Becker, H.D.; Stathopoulos, C.; Amoutzias, G.D. The complex evolutionary history of aminoacyl-tRNA synthetases. *Nucleic Acids Res.* **2016**, *45*, 1059–1068. [CrossRef] [PubMed]
4. Carter, C.W., Jr. Tryptophanyl-tRNA synthetases. In *The Aminoacyl-tRNA Synthetases*; Ibba, M., Francklyn, C., Cusack, S.E., Eds.; Landes Biosciences: Georgetown, TX, USA, 2005.
5. Bedouelle, H. Tyrosyl-tRNA synthetases. In *The Aminoacyl-tRNA Synthetases*; Ibba, M., Francklyn, C., Cusack, S.E., Eds.; Landes Biosciences: Georgetown, TX, USA, 2005.
6. Bonnefond, L.; Giegé, R.; Rudinger-Thirion, J. Evolution of the tRNA^{Tyr}/TyrRS aminoacylation systems. *Biochimie* **2005**, *87*, 873–883. [CrossRef] [PubMed]
7. Xue, H.; Shen, W.; Giegé, R.; Wong, J.T. Identity elements of tRNA^{Trp}. Identification and evolutionary conservation. *J. Biol. Chem.* **1993**, *268*, 9316–9322. [PubMed]
8. Furukawa, R.; Nakagawa, M.; Kuroyanagi, T.; Yokobori, S.I.; Yamagishi, A. Quest for Ancestors of Eukaryal Cells Based on Phylogenetic Analyses of Aminoacyl-tRNA Synthetases. *J. Mol. Evol.* **2017**, *84*, 51–66. [CrossRef] [PubMed]
9. Andam, C.P.; Gogarten, J.P. Biased gene transfer in microbial evolution. *Nat. Rev. Microbiol.* **2011**, *9*, 543–555. [CrossRef] [PubMed]
10. Shiba, K.; Motegi, H.; Schimmel, P. Maintaining genetic code through adaptations of tRNA synthetases to taxonomic domains. *Trends Biochem. Sci.* **1997**, *22*, 453–457. [CrossRef]
11. Sassanfar, M.; Kranz, J.E.; Gallant, P.; Schimmel, P.; Shiba, K. A eubacterial *Mycobacterium tuberculosis* tRNA synthetase is eukaryote-like and resistant to a eubacterial-specific antisynthetase drug. *Biochemistry* **1996**, *35*, 9995–10003. [CrossRef] [PubMed]
12. Ardell, D.H.; Andersson, S.G. TFAM detects co-evolution of tRNA identity rules with lateral transfer of histidyl-tRNA synthetase. *Nucleic Acids Res.* **2006**, *34*, 893–904. [CrossRef] [PubMed]
13. Fournier, G.P.; Alm, E.J. Ancestral Reconstruction of a Pre-LUCA Aminoacyl-tRNA Synthetase Ancestor Supports the Late Addition of Trp to the Genetic Code. *J. Mol. Evol.* **2015**, *80*, 171–185. [CrossRef] [PubMed]
14. Woese, C.R.; Olsen, G.J.; Ibba, M.; Söll, D. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **2000**, *64*, 202–236. [CrossRef] [PubMed]
15. Wolf, Y.I.; Aravind, L.; Grishin, N.V.; Koonin, E.V. Evolution of aminoacyl-tRNA synthetases—Analysis of unique domain architectures and phylogenetic trees reveals a complex history of horizontal gene transfer events. *Genome Res.* **1999**, *9*, 689–710. [PubMed]
16. Ribas de Pouplana, L.; Frugier, M.; Quinn, C.L.; Schimmel, P. Evidence that two present-day components needed for the genetic code appeared after nucleated cells separated from eubacteria. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 166–170. [CrossRef] [PubMed]
17. Iraha, F.; Oki, K.; Kobayashi, T.; Ohno, S.; Yokogawa, T.; Nishikawa, K.; Yokoyama, S.; Sakamoto, K. Functional replacement of the endogenous tyrosyl-tRNA synthetase-tRNA^{Tyr} pair by the archaeal tyrosine pair in *Escherichia coli* for genetic code expansion. *Nucleic Acids Res.* **2010**, *38*, 3682–3691. [CrossRef] [PubMed]

18. Italia, J.S.; Addy, P.S.; Wrobel, C.J.J.; Crawford, L.A.; Lajoie, M.J.; Zheng, Y.; Chatterjee, A. An orthogonalized platform for genetic code expansion in both bacteria and eukaryotes. *Nat. Chem. Biol.* **2017**. [CrossRef] [PubMed]

19. Yutin, N.; Wolf, Y.I.; Koonin, E.V. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* **2014**, *466–467*, 38–52. [CrossRef] [PubMed]

20. Hug, L.A.; Baker, B.J.; Anantharaman, K.; Brown, C.T.; Probst, A.J.; Castelle, C.J.; Butterfield, C.N.; Hernsdorf, A.W.; Amano, Y.; Ise, K.; et al. A new view of the tree of life. *Nat. Microbiol.* **2016**, *1*, 16048. [CrossRef] [PubMed]

21. Eloe-Fadrosh, E.A.; Ivanova, N.N.; Woyke, T.; Kyrpides, N.C. Metagenomics uncovers gaps in amplicon-based detection of microbial diversity. *Nat. Microbiol.* **2016**, *1*, 15032. [CrossRef] [PubMed]

22. Anantharaman, K.; Brown, C.T.; Hug, L.A.; Sharon, I.; Castelle, C.J.; Probst, A.J.; Thomas, B.C.; Singh, A.; Wilkins, M.J.; Karaoz, U. Thousands of microbial genomes shed light on interconnected biogeochemical processes in an aquifer system. *Nat. Commun.* **2016**, *7*, 13219. [CrossRef] [PubMed]

23. Jaffe, A.L.; Corel, E.; Pathmanathan, J.S.; Lopez, P.; Bapteste, E. Bipartite graph analyses reveal interdomain LGT involving ultrasmall prokaryotes and their divergent, membrane-related proteins. *Environ. Microbiol.* **2016**, *18*, 5072–5081. [CrossRef] [PubMed]

24. Brown, C.T.; Hug, L.A.; Thomas, B.C.; Sharon, I.; Castelle, C.J.; Singh, A.; Wilkins, M.J.; Wrighton, K.C.; Williams, K.H.; Banfield, J.F. Unusual biology across a group comprising more than 15% of domain bacteria. *Nature* **2015**, *523*, 208–211. [CrossRef] [PubMed]

25. Wrighton, K.C.; Castelle, C.J.; Varaljay, V.A.; Satagopan, S.; Brown, C.T.; Wilkins, M.J.; Thomas, B.C.; Sharon, I.; Williams, K.H.; Tabita, F.R.; et al. RubisCO of a nucleoside pathway known from Archaea is found in diverse uncultivated phyla in bacteria. *ISME J.* **2016**, *10*, 2702–2714. [CrossRef] [PubMed]

26. Markowitz, V.M.; Chen, I.M.; Chu, K.; Szeto, E.; Palaniappan, K.; Pillay, M.; Ratner, A.; Huang, J.; Pagani, I.; Tringe, S.; et al. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* **2014**, *42*, D568–D573. [CrossRef] [PubMed]

27. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef] [PubMed]

28. Gouy, M.; Guindon, S.; Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **2010**, *27*, 221–224. [CrossRef] [PubMed]

29. Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [CrossRef] [PubMed]

30. Jia, J.; Chen, X.L.; Guo, L.T.; Yu, Y.D.; Ding, J.P.; Jin, Y.X. Residues Lys-149 and Glu-153 switch the aminoacylation of tRNA^{Trp} in *Bacillus subtilis*. *J. Biol. Chem.* **2004**, *279*, 41960–41965. [CrossRef] [PubMed]

31. Xu, F.; Chen, X.; Xin, L.; Chen, L.; Jin, Y.; Wang, D. Species-specific differences in the operational RNA code for aminoacylation of tRNA^{Trp}. *Nucleic Acids Res.* **2001**, *29*, 4125–4133. [CrossRef] [PubMed]

32. Yang, X.L.; Otero, F.J.; Skene, R.J.; McRee, D.E.; Schimmel, P.; Ribas de Pouplana, L. Crystal structures that suggest late development of genetic code components for differentiating aromatic side chains. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 15376–15380. [CrossRef] [PubMed]

33. Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S. Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat. Struct. Biol.* **2003**, *10*, 425–432. [CrossRef] [PubMed]

34. Stiebritz, M.T. A role for [Fe$_4$S$_4$] clusters in tRNA recognition—A theoretical study. *Nucleic Acids Res.* **2014**, *42*, 5426–5435. [CrossRef] [PubMed]

35. Abergel, C.; Rudinger-Thirion, J.; Giegé, R.; Claverie, J.M. Virus-encoded aminoacyl-tRNA synthetases: Structural and functional characterization of mimivirus TyrRS and MetRS. *J. Virol.* **2007**, *81*, 12406–12417. [CrossRef] [PubMed]

36. Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G.J. JPred4: A protein secondary structure prediction server. *Nucleic Acids Res.* **2015**, *43*, W389–W394. [CrossRef] [PubMed]

37. Biasini, M.; Bienert, S.; Waterhouse, A.; Arnold, K.; Studer, G.; Schmidt, T.; Kiefer, F.; Cassarino, T.G.; Bertoni, M.; Bordoli, L.; et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res.* **2014**, *42*, W252–W258. [CrossRef] [PubMed]

38. Han, G.W.; Yang, X.L.; McMullan, D.; Chong, Y.E.; Krishna, S.S.; Rife, C.L.; Weekes, D.; Brittain, S.M.; Abdubek, P.; Ambing, E.; et al. Structure of a tryptophanyl-tRNA synthetase containing an iron-sulfur cluster. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **2010**, *66*, 1326–1334. [CrossRef] [PubMed]

39. Zetsche, B.; Gootenberg, J.S.; Abudayyeh, O.O.; Slaymaker, I.M.; Makarova, K.S.; Essletzbichler, P.; Volz, S.E.; Joung, J.; van der Oost, J.; Regev, A.; et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **2015**, *163*, 759–771. [CrossRef] [PubMed]

40. Roy, H.; Ibba, M. Phenylalanyl-tRNA synthetase contains a dispensable RNA-binding domain that contributes to the editing of noncognate aminoacyl-tRNA. *Biochemistry* **2006**, *45*, 9156–9162. [CrossRef] [PubMed]

41. Kapps, D.; Cela, M.; Théobald-Dietrich, A.; Hendrickson, T.; Frugier, M. OB or Not OB: Idiosyncratic utilization of the tRNA-binding OB-fold domain in unicellular, pathogenic eukaryotes. *FEBS Lett.* **2016**, *590*, 4180–4191. [CrossRef] [PubMed]

42. Rao, S.T.; Rossmann, M.G. Comparison of super-secondary structures in proteins. *J. Mol. Biol.* **1973**, *76*, 241–256. [CrossRef]

43. Moutiez, M.; Belin, P.; Gondry, M. Aminoacyl-tRNA-Utilizing Enzymes in Natural Product Biosynthesis. *Chem. Rev.* **2017**. [CrossRef] [PubMed]

44. Wang, C.; Sobral, B.W.; Williams, K.P. Loss of a universal tRNA feature. *J. Bacterial.* **2007**, *189*, 1954–1962. [CrossRef] [PubMed]

45. Yuan, J.; Gogakos, T.; Babina, A.M.; Söll, D.; Randau, L. Change of tRNA identity leads to a divergent orthogonal histidyl-tRNA synthetase/tRNA[His] pair. *Nucleic Acids Res.* **2011**, *39*, 2286–2293. [CrossRef] [PubMed]

46. Bonnefond, L.; Frugier, M.; Giegé, R.; Rudinger-Thirion, J. Human mitochondrial TyrRS disobeys the tyrosine identity rules. *RNA* **2005**, *11*, 558–562. [CrossRef] [PubMed]

47. Bonnefond, L.; Frugier, M.; Touzé, E.; Lorber, B.; Florentz, C.; Giegé, R.; Sauter, C.; Rudinger-Thirion, J. Crystal Structure of Human Mitochondrial Tyrosyl-tRNA Synthetase Reveals Common and Idiosyncratic Features. *Structure* **2007**, *15*, 1505–1516. [CrossRef] [PubMed]

48. Buddha, M.R.; Crane, B.R. Structure and activity of an aminoacyl-tRNA synthetase that charges tRNA with nitro-tryptophan. *Nat. Struct. Mol. Biol.* **2005**, *12*, 274–275. [CrossRef] [PubMed]

49. Kitabatake, M.; Ali, K.; Demain, A.; Sakamoto, K.; Yokoyama, S.; Söll, D. Indolmycin resistance of *Streptomyces coelicolor* A3(2) by induced expression of one of its two tryptophanyl-tRNA synthetases. *J. Biol. Chem.* **2002**, *277*, 23882–23887. [CrossRef] [PubMed]

50. Rubio, M.Á.; Napolitano, M.; Ochoa de Alda, J.A.; Santamaría-Gómez, J.; Patterson, C.J.; Foster, A.W.; Bru-Martínez, R.; Robinson, N.J.; Luque, I. Trans-oligomerization of duplicated aminoacyl-tRNA synthetases maintains genetic code fidelity under stress. *Nucleic Acids Res.* **2015**, *43*, 9905–9917. [CrossRef] [PubMed]

51. Buddha, M.R.; Keery, K.M.; Crane, B.R. An unusual tryptophanyl tRNA synthetase interacts with nitric oxide synthase in *Deinococcus radiodurans*. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15881–15886. [CrossRef] [PubMed]

52. Larson, E.T.; Kim, J.E.; Castaneda, L.J.; Napuli, A.J.; Zhang, Z.; Fan, E.; Zucker, F.H.; Verlinde, C.L.; Buckner, F.S.; Van Voorhis, W.C.; et al. The double-length tyrosyl-tRNA synthetase from the eukaryote *Leishmania major* forms an intrinsically asymmetric pseudo-dimer. *J. Mol. Biol.* **2011**, *409*, 159–176. [CrossRef] [PubMed]