RNA (2002), *8*:1189–1232. Cambridge University Press. Printed in the USA.
Copyright © 2002 RNA Society.
DOI: 10.1017/S1355838202022021

BIOINFORMATICS

tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features

CHRISTIAN MARCK[1] and HENRI GROSJEAN[2]

[1]Service de Biochimie et de Génétique Moléculaire, Bât 144, CEA/Saclay, 91191 Gif-sur-Yvette, France
[2]Laboratoire d'Enzymologie et Biochimie Structurales, Bât 34, Centre National de la Recherche Scientifique, 91198 Gif-sur-Yvette, France

ABSTRACT

From 50 genomes of the three domains of life (7 eukarya, 13 archaea, and 30 bacteria), we extracted, analyzed, and compared over 4,000 sequences corresponding to cytoplasmic, nonorganellar tRNAs. For each genome, the complete set of tRNAs required to read the 61 sense codons was identified, which permitted revelation of three major anticodon-sparing strategies. Other features and sequence peculiarities analyzed are the following: (1) fit to the standard cloverleaf structure, (2) characteristic consensus sequences for elongator and initiator tDNAs, (3) frequencies of bases at each sequence position, (4) type and frequencies of conserved 2D and 3D base pairs, (5) anticodon/tDNA usages and anticodon-sparing strategies, (6) identification of the tRNA-Ile with anticodon CAU reading AUA, (7) size of variable arm, (8) occurrence and location of introns, (9) occurrence of 3′-CCA and 5′-extra G encoded at the tDNA level, and (10) distribution of the tRNA genes in genomes and their mode of transcription. Among all tRNA isoacceptors, we found that initiator tDNA-iMet is the most conserved across the three domains, yet domain-specific signatures exist. Also, according to which tRNA feature is considered (5′-extra G encoded in tDNAs-His, AUA codon read by tRNA-Ile with anticodon CAU, presence of intron, absence of "two-out-of-three" reading mode and short V-arm in tDNA-Tyr) Archaea sequester either with Bacteria or Eukarya. No common features between Eukarya and Bacteria not shared with Archaea could be unveiled. Thus, from the tRNomic point of view, Archaea appears as an "intermediate domain" between Eukarya and Bacteria.

Keywords: comparative genomics; matching pattern; RNomics; tRNA

Reprint requests to: Christian Marck, Service de Biochimie et de Génétique Moléculaire, Bât. 144, CEA/Saclay, F-91191 Gif-sur-Yvette CEDEX, France; e-mail: christian.marck@cea.fr.

Abbreviations and conventions: 2D base pair: base pair in one of the four stems of the tRNA cloverleaf structure model; 3D base pair: base pair that participates in the stabilization of the L-shaped structure of the tRNA molecule, for example, 18–55 and 19–56; 3′-CCA: 3′-terminal CCA; BHB: bulge-helix-bulge; G(−1): G located at position −1, that is, 5′ to position 1 of the mature tRNA; G:A, A:A, A:C, and so on: mismatched base pairs; tDNA: tRNA gene; V-arm: the variable arm in the tRNA molecule (positions 44 to 48); *UNN: UNN codon family with U modified, and so on; Anticodon-sparing strategy, for example, "C34-sparing strategy": in a given organism, the partial or total avoidance of tRNAs using C at position 34, the first position of the anticodon; Decoding mode: the use of one, two, or three different anticodon-sparing strategies, for example, "A34- or G34-sparing strategy" plus "C34-sparing strategy" corresponding to mode II.

The universal conventional numbering system for tRNA positions is that adopted in the tRNA database (Sprinzl et al., 1998). To report consensus sequences, we use the degenerate one-letter code for bases (Cornish-Bowden, 1985) that allows us to designate each base or combination of bases (see legend to Fig. 2). However, for clarity, in the figures we often use – instead of N to designate any base (A, C, G, or T). The common names of modified nucleosides (Ψ, pseudouridine; m^5C, 5-methylcytosine; k^2C, lysidine) and their chemical structures can be found in Limbach et al. (1994). Because we deal mostly with tRNA genes (tDNA) rather than mature tRNA, we designate the anticodons (with the three bases in parentheses) as they appear in the tRNA genes and similarly the corresponding complementary codons as they appear in the ORFs. We designate tDNAs as in this example: tDNA-Met (CAT). The Met initiator tDNA is written tDNA-iMet (CAT) and the peculiar tDNA-Ile harboring a (CAT) anticodon that decodes the ATA Ile codon in Archaea and Bacteria is referred to as tDNA-Ile (CAT reading ATA).

In the text and figures, genomes are referred to as E01 to E07 for Eukarya, A01 to A13 for Archaea, and B01 to B30 for Bacteria (see Table 3 in Methods for references).

INTRODUCTION

The first nucleotide sequence of a mature tRNA molecule and its putative two-dimensional cloverleaf structure was determined by Holley in 1965 (Holley, 1965; Holley et al., 1965). Since then, more than 4,000 different mature tRNA and tRNA genes (tDNA) originating from a variety of organisms have been sequenced and subsequently compiled in the current release of the tRNA data bank (550 mature tRNA sequences, 500 tRNA genes from various organisms, and 3,700 tRNA genes from fully sequenced genomes of 63 organisms; Sprinzl et al., 1998; http://www.uni-bayreuth.de/departments/biochemie/trna/). This exponentially increasing information not only reinforces the evidence for a universally adopted cloverleaf secondary structure as initially proposed by Holley (reviewed in Dirheimer et al., 1995; Westhof & Auffinger, 2001; see also Rich & RajBhandary, 1976), but also revealed the presence of over 90 different types of modified nucleotides in mature tRNA molecules, as listed in the tRNA Modification Database (http://medlib.med.utah.edu/RNAmods/; for reviews, see Björk, 1995; several chapters in Grosjean & Benne, 1998).

The remarkable progresses in nucleic acid sequencing, especially in large-scale automated DNA sequencing of whole genomes, together with the development of algorithms allowing for identification of tRNA genes within the genomic sequences (Fichant & Burks, 1991; Pavesi et al., 1994; el-Mabrouk & Lisacek, 1996; Lowe & Eddy, 1997; Gautheret & Lambert, 2001) have provided information on many new tRNA genes. To date (July 2002), 7 genomes from Eukarya, 15 from Archaea, and 60 from Bacteria have been fully sequenced, and sequencing of many others is in progress (see GOLD Web site, http://igweb.integratedgenomics.com/GOLD/). These primary sequence data now allow access to more detailed information regarding anticodon usage (in its unmodified version) and the number and the organization of tRNA genes at the genome level, but also allow for searching for characteristic structural features of tRNAs within a given genome as well as according their very diverse genomic origins.

Such a systematic comparison of the primary sequences of all tRNAs from a group of related organisms has been reported for the mammalian mitochondrial tRNAs as found in 31 fully sequenced mammalian mitochondrial genomes (Helm et al., 2000). It revealed much structural variability among this category of tRNAs, with deviations from the classical tRNA cloverleaf secondary structure being mostly concentrated within the D- and the T-loops. Also, classification of the tRNA sequences according to their genomic origin allowed for highlighting of specific features, such as the variable number of mismatches or G-T pairs in stems and the extremely low G or C content in the D- and T-loops. As a result, 22 "typical" mammalian mitochondrial sequences were proposed. As far as cytoplasmic tRNAs are concerned, a genome-wide survey of tRNA genes and retroelements in the yeast *Saccharomyces cerevisiae* was reported (Hani & Feldmann, 1998). Also, for all newly sequenced genomes, raw tDNA sequences found by running tRNAscan-SE (Lowe & Eddy, 1997) were made available on-line (http://rna.wustl.edu/GtRDB/) and aligned versions are available from the tRNA data bank (Sprinzl et al., 1998; http://www.uni-bayreuth.de/departments/biochemie/trna/). More recently, the information content of tDNAs and their upstream sequences derived from five eukaryotic genomes have been systematically explored (Hamada et al., 2001).

In the present work, we selected 50 sequenced genomes (Fig. 1) out of the 84 available to date: 7 of these are of eukaryotic, 13 of archaeal, and 30 of bacterial genomes. Our goal was to investigate and compare nuclear tDNAs data within each genome as well as throughout genomes from the three domains of life (Woese et al., 1990). Archaeal and bacterial genomes were selected to avoid an overrepresentation of phylogenetically too closely related organisms (see Fig. 1) and also to span the widest range of living areas. As several thousand tDNAs were to be examined, we decided to set up numerical procedures able to sort and compare these sequences and also compute relevant statistical data. Regarding the initial step of this quest, namely the identification of tDNA sequences, we devised a simple tDNA search algorithm, solely based on the recognition of the tRNA cloverleaf structure. The 4,000 aligned tDNAs sequences extracted from the 50 genomes are available upon request (send e-mail containing the keyword "tDNAs_50_Genomes" to christian.marck@cea.fr).

RESULTS AND DISCUSSION

Search for cloverleaf structure reveals over 4,000 tRNA genes in 50 fully sequenced genomes

The sequence of each genome was searched for tDNAs on the basis of the standard cloverleaf structure (Dirheimer et al., 1995; Westhof & Auffinger, 2001; see Fig. 2A, B) and these were visually inspected, once sorted out as isoacceptor subfamilies (see definition of subfamily in Methods). Only tDNAs extracted from the four higher eukaryotes had to be checked visually and were cleaned from a few false positives, pseudo-genes, or organelle-derived tDNAs. The number of tDNAs eliminated in this way were: *Caenorhabditis elegans*, 10 from a total of 539; *Drosophila melanogaster*, 5 from 287; *Arabidopsis thaliana*, 30 (that include the tDNA of mitochondrial origin; *Arabidopsis* Genome Initiative, 2000) from 592; *Homo sapiens*, 29 from 186.

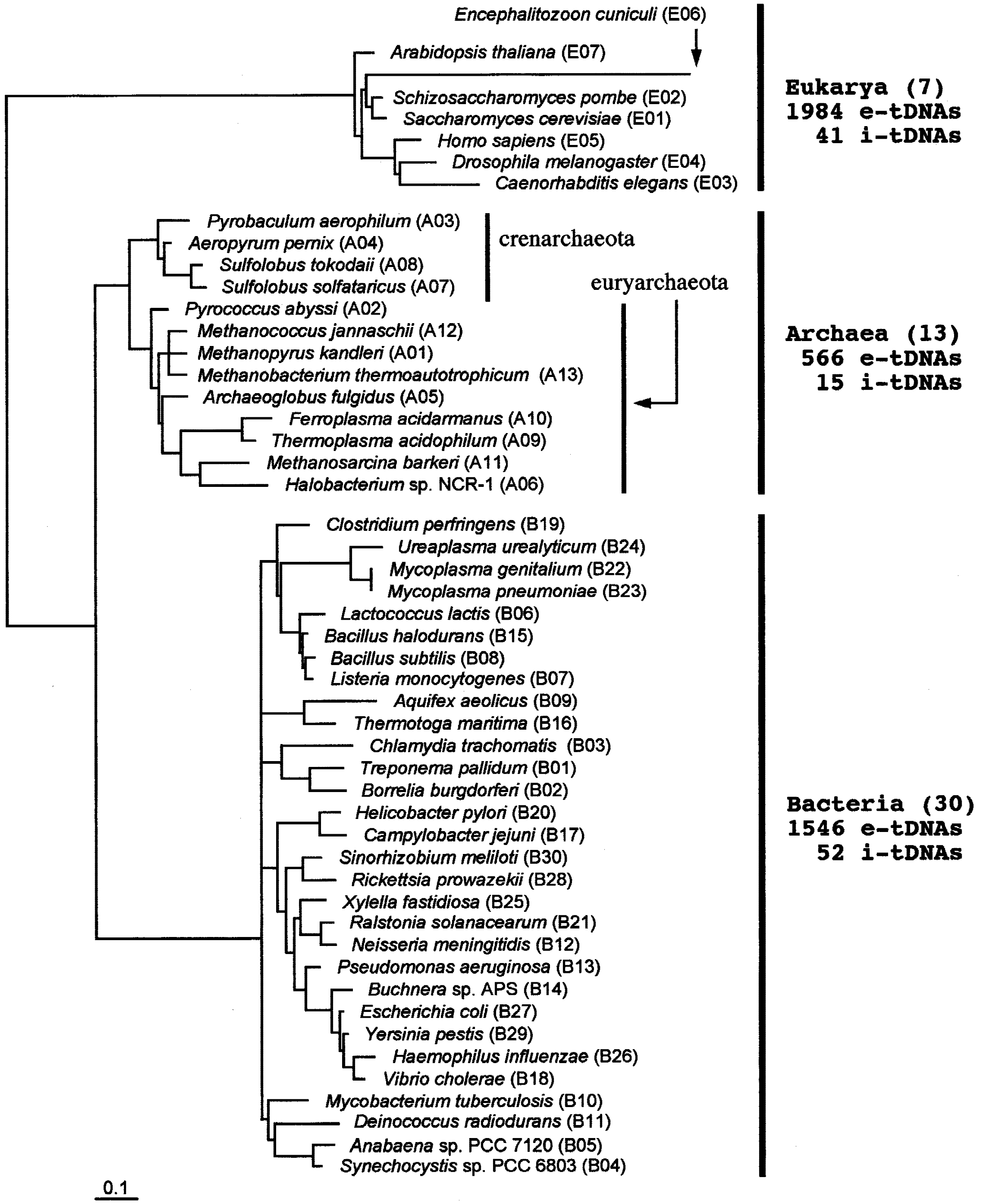
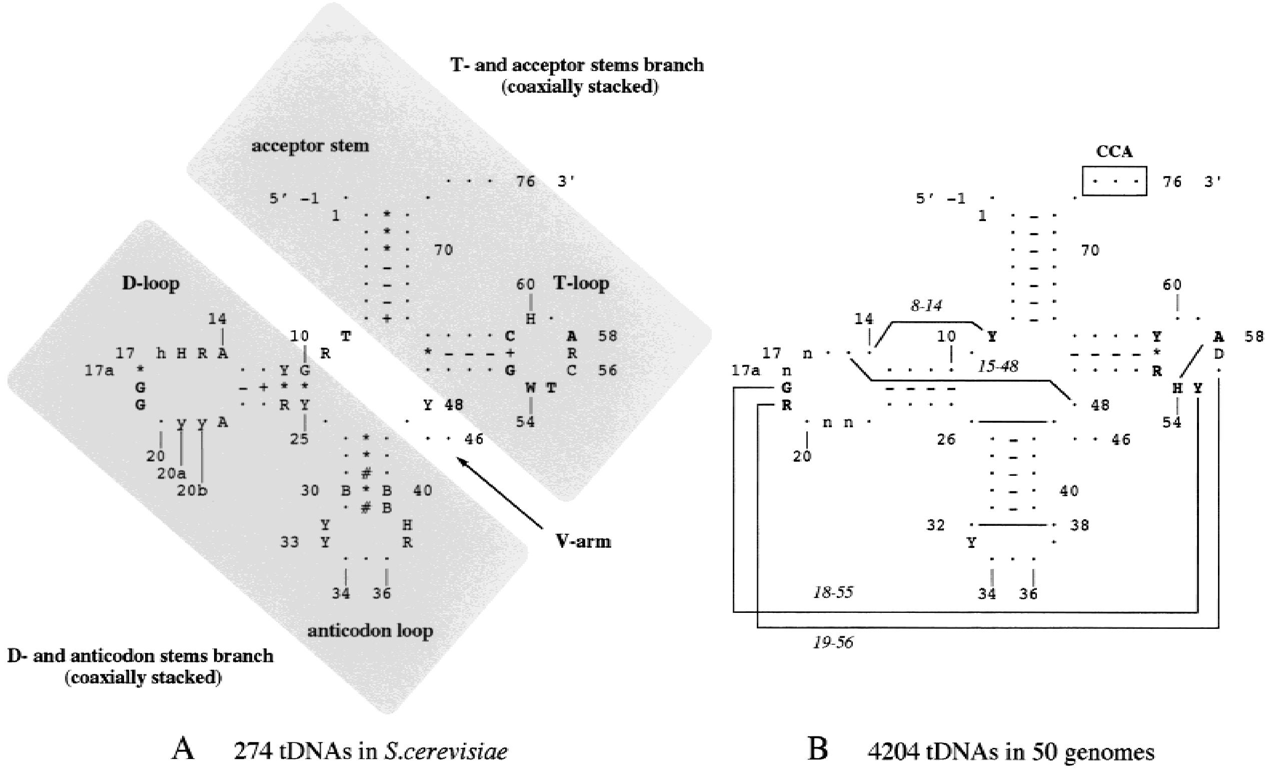

Encephalitozoon cuniculi (E06)
Arabidopsis thaliana (E07)
Schizosaccharomyces pombe (E02)
Saccharomyces cerevisiae (E01)
Homo sapiens (E05)
Drosophila melanogaster (E04)
Caenorhabditis elegans (E03)
Eukarya (7)
1984 e-tDNAs
41 i-tDNAs
Pyrobaculum aerophilum (A03)
Aeropyrum pernix (A04)
Sulfolobus tokodaii (A08)
Sulfolobus solfataricus (A07)
crenarchaeota
euryarchaeota
Pyrococcus abyssi (A02)
Methanococcus jannaschii (A12)
Methanopyrus kandleri (A01)
Methanobacterium thermoautotrophicum (A13)
Archaeoglobus fulgidus (A05)
Ferroplasma acidarmanus (A10)
Thermoplasma acidophilum (A09)
Methanosarcina barkeri (A11)
Halobacterium sp. NCR-1 (A06)
Archaea (13)
566 e-tDNAs
15 i-tDNAs
Clostridium perfringens (B19)
Ureaplasma urealyticum (B24)
Mycoplasma genitalium (B22)
Mycoplasma pneumoniae (B23)
Lactococcus lactis (B06)
Bacillus halodurans (B15)
Bacillus subtilis (B08)
Listeria monocytogenes (B07)
Aquifex aeolicus (B09)
Thermotoga maritima (B16)
Chlamydia trachomatis (B03)
Treponema pallidum (B01)
Borrelia burgdorferi (B02)
Helicobacter pylori (B20)
Campylobacter jejuni (B17)
Sinorhizobium meliloti (B30)
Rickettsia prowazekii (B28)
Xylella fastidiosa (B25)
Ralstonia solanacearum (B21)
Neisseria meningitidis (B12)
Pseudomonas aeruginosa (B13)
Buchnera sp. APS (B14)
Escherichia coli (B27)
Yersinia pestis (B29)
Haemophilus influenzae (B26)
Vibrio cholerae (B18)
Mycobacterium tuberculosis (B10)
Deinococcus radiodurans (B11)
Anabaena sp. PCC 7120 (B05)
Synechocystis sp. PCC 6803 (B04)
Bacteria (30)
1546 e-tDNAs
52 i-tDNAs
0.1

FIGURE 1. Phylogenetic tree of the 50 genomes investigated. The tree was constructed by analyzing the sequences of 16S rRNA genes. Distances are indicated in substitutions per site. The number of tDNAs found in each domain are indicated. e-tDNAs: elongator tDNAs; i-tDNAs: initiator tDNAs-iMet (CAT).

From the 50 selected genomes (see Methods), we extracted 4,204 tDNAs (Eukarya: 2,025; Archaea: 581; Bacteria: 1,598). However, one should keep in mind that: (1) the genome of *D. melanogaster* was available only as "scaffolds"; therefore, the copy number of some tDNAs may be overestimated for this genome; (2) conversely, a few genome sequences were slightly incomplete (of which *Ferroplasma acidarmanus* (A10) and *Methanosarcina barkeri* (A11)) and therefore a few tDNAs are probably missing (see below); (3) human sequences used were largely incomplete (~500 Mb), and were used only tentatively, (4) because of the strict requirement of a standard cloverleaf structure, we may have missed some tDNAs due to sequencing errors; (5) selenocysteine tDNA (anticodon UCA; Hubert et al., 1998), pyrrolysine tDNA (anticodon CUA; Hao et al., 2002; Srinivasan et al., 2002), tmRNA (http://www.indiana.edu/~tmrna; Williams, 2002), or the genes for



A 274 tDNAs in *S.cerevisiae*

B 4204 tDNAs in 50 genomes

FIGURE 2. The tRNA cloverleaf structure. Various features of the tRNA molecule in the form of the cloverleaf model are depicted in the two structures presented. Cloverleaf A is the consensus sequences obtained by adding the 274 tDNA sequences of *S. cerevisiae* (E01) and cloverleaf B by adding the 4,204 tDNA sequences extracted from the 50 genomes listed in Table 3. The conventional IUB/IUPAC degenerate DNA alphabet (Cornish-Bowden, 1985) is used in this and the following figures: **R** (purine), A or G; **Y** (pyrimidine), C or T; **S** (strong), G or C; **W** (weak), A or T; **M** (amino), A or C; **K** (keto), G or T; **B** (not A), C, G, or T; **D** (not C), A, G, or T ; **H** (not G), A, C, or T; **V** (not T), A, C, or G; **N** (any), A, C, G, or T. Although the scanned sequences are DNA, the standard tRNA cloverleaf structure and numbering (Sprinzl et al., 1998) are used to allow the representation of base-pairing consensus in the stems. Key to base-pairing consensus is: +, Watson–Crick base pairing only; ~, G-T (or T-G) base pairing only; *, Watson–Crick pairing or G-T/T-G pairings; # Watson–Crick pairing or mismatch; −, Watson–Crick pairing or G-T/T-G pairings or mismatches. The sequences composing the intron and the extra bases present in the long V-arm (also called variable arm; positions noted 47 to 47k) are not shown for clarity. The bases shown in bold were strictly constrained during the search (see details in Table 4); bases shown as lower case letters (at position 17, 17a, 20a, and 20b) are optional bases not present in all tRNAs; dots are used to indicate "N" (any base), except at the optional positions at which "n" is used. Stems, loops and the two branches of the L-shaped tRNA molecule are indicated in cloverleaf A; 3D base pairs are indicated by black lines in cloverleaf B (bp 8–14, 15–48, 18–55, 19–56, 26–44, 32–38, 54–58) with the base numbers indicated for some of them (in italics). The 3D base pair 45–47k as well as other three-dimensional hydrogen bounds present in tRNA are not shown. Box labeled CCA in cloverleaf B indicates the location of the mature CCA 3′ sequence.

tRNA species that participate in peptidoglycan synthesis (Stewart et al., 1971) were not considered in this study because they deviate from the standard cloverleaf model.

Other sources of abnormalities could arise from the fact that, after transcription of a tRNA gene, some primary transcripts may be edited by enzymatic processes that catalyze insertion/deletion of nucleotides, or convert one canonical base into another or reprogram the nucleotide sequence at the 3′ and/or 5′ extremities (tRNA maturation). However, these so-called tRNA editing processes have been demonstrated to date only in mitochondrial and/or chloroplastic and not in cytoplasmic tRNA of higher Eukarya (from single-celled protozoa to mammals and higher plants; Price & Gray, 1999; Gott & Emeson, 2000; Fey et al., 2001; Schurer et al., 2001). It is also noteworthy that in *Escherichia coli* (B27), mature tRNA-Ser (GGA) exists in two distinguishable forms, differing only by a C or a D (dihydrouridine) at position 20 of the D-loop (Grosjean et al., 1985), whereas on the genome, only one gene in two copies exists, both harboring C20; no gene having T20 was detected (Komine et al., 1990; Blattner et al., 1997). This last observation suggests that, at least for this particular tRNA-Ser (GGA), an enzymatic posttranscriptional editing/modification of C20 to D20 (probably via U20) in the D-loop may take place in *E. coli.* Systematic comparison of tDNA sequences with those of mature tRNA (when available) in other organisms might point out more observations of that sort.

In organisms harboring multiple copies of a given tDNA, as in most eukaryotes and many bacteria (see below), we observed that, only in a few cases, tDNAs with the same anticodon differ in their primary sequence, thus splitting an isoacceptor family into two or more subfamilies. In cases where these changes include a base pair inside a stem (e.g., G-C versus A-T, or G-C versus G-T), these cannot be due to sequencing errors, because a base-pair change requires the coordinated changes of two distantly related bases. In *S. cerevisiae* (E01), for example, base-pair changes were observed in tDNA-Phe (GAA) at bp 6–67, where T-A is found in eight copies (see Dirheimer et al., 1995), C-G in the two other copies; in tDNA-Gln (TTG), bp 5–68 is G-C in seven copies, T-A in the two other copies; in tDNA-Thr (TGT), bp 50–64 is T-A in three copies, G-C in the other copy. Comparable base-pair exchanges are also found in the tDNAs of other Eukarya as well as in bacteria that have multiple copies of the same isoacceptor. Sequence variations were sometimes more drastic, as exemplified by the dividing out of the 60 copies of tDNA-Tyr (GTA) of *A. thaliana* (E07) into seven subfamilies (32, 14, 5, 5, 2, 1, and 1 copies, respectively). Surprisingly, sorting the tDNAs as a function of subfamily also revealed a case of two markedly different tDNA-Ile (TAT) in *C. elegans* (E03, three and four copies, respectively) with an intron in only one of the two subfamilies. As a rule, the higher in evolution a eukaryote lies (from lower unicellular to higher multicellular eukaryotes), the more numerous are the sequence variations for a given isoacceptor.

Comparison of all tDNAs from genomes of the three domains of life reveals domain-specific structural and functional features

The sequences of the 4,204 tDNAs found in the 50 genomes analyzed in this work are presented in the form of linear consensus sequences for each genome (Fig. 3, line E01 to E07 for Eukarya; A01 to A13 for Archaea, and B01 to B30 for Bacteria). Global consensus sequences for all elongator tRNA genes of each domain (noted Ee, Ae, and Be) were computed separately from those of the initiators (noted Ei, Ai, and Bi). Figure 4 shows the corresponding cloverleaf structures with all elongator tDNAs in the upper part and initiator tDNA-iMet (CAT) in the lower part. The frequency of occurrence of a given base (A, G, T, and C) at each position of the cloverleaf structure is given in Table 1, Panel A, for all elongator tDNAs and in Table 1, Panel B, for all initiator tDNA-iMet (CAT). The frequency of occurrence of various base pairings (in stems and 3D base pairs) were also computed and are given in Table 2, Panel A, for all elongator tDNAs and in Table 2, Panel B, for all initiator tDNA-iMet (CAT).

Inspection of consensus sequences of all elongator tDNAs in a given genome, E01 to E07 for Eukarya (total 1,984 tDNA sequences corresponding to 302 isoacceptor families), A01 to A13 for Archaea (total 566 sequences, 550 isoacceptor families), and B01 to B30 for Bacteria (total 1,546 sequences, 1,111 isoacceptor families) allows to easily visualize the most universally conserved bases (colored in blue) that span the three domains of life. Many of these bases are located in the D- and T-loops and participate in 3D base pairing that stabilizes the tertiary structure of tRNA (Fig. 2A,B; see also Dirheimer et al., 1995). Other conserved bases that appear inside a given domain or only in two domains are shown in red (Figs. 3 and 4). In Bacteria (lines B01 to B30), apart from the universally conserved bases (shown in blue), only a few domain-specific bases emerge: for example, bp 11–24 is predominantly Y-R (T-A or C-G); purine (noted as R) predominates at position 15 and A, C, or T (noted as H) is found at position 32. More positions are specifically conserved in archaeal tDNA (lines A01 to A13), especially at the end of the acceptor stem: for example, bp 1–72 is mostly V-B (where V is any base but T paired or mismatched with B, where B is any base but C), the only exception being for *Halobacterium* sp. NRC-1 (A06), and bp 2–71 is mostly S-S (G-C or C-G). Also, tDNAs of 7 among the 13 archaeons examined harbor a conserved C at position 48. Noteworthy, in the four hyperthermophilic Archaea, *Methanopyrus kandleri* (A01), *Pyrococcus abyssi* (A02), *Pyrobaculum aerophilum* (A03), and *Aeropyrum pernix* (A04), thriving at temperatures above 95 °C, a large number of S bases (G or C) resulting in G-C pairs in either orientation are found in the T- and acceptor stems (in red in Fig. 3). The same is true for two other hyperthermophilic archaeons, the genomes of which have been fully sequenced [*Pyrococcus furiosus* (Robb et al., 2001) and *Pyrococcus horikoshi* (Kawarabayasi et al., 1998; data not shown)]. Such a selective choice of G-C pairs is less evident in the sequences of thermophilic archaeons *Archaeoglobus fulgidus* (A05) or *Sulfolobus solfataricus* (A07) (optimal growth temperature around 80 °C) or other extremophilic archaea such as *Halobacterium* sp. NRC-1 (A06) and *Thermoplasma acidophilum* (A09) thriving at lower temperatures, 37 and 60 °C, respectively. Noteworthy, in the hyperthermophilic bacteria *Aquifex aeolicus* (B09) and *Thermotoga maritima* (B16; thriving at about 85 and 80 °C, respectively), such a G-C-rich base pairing at positions 1–72 and 2–71 does not exist. Apparently, tRNAs of organisms living at extreme temperatures (higher than 90 °C) harbor exceptionally G-C rich and thermostable stems whereas those of organisms living at temperatures around 80 °C or lower do not.

In Eukarya, extra base conservations occur inside and around the D-loop area that add to those common to the three domains (T8, or C8 in one archaeon, A14,

```
a               1   1  1111 222   2       3  33 3 3 3 3  4    4   4      5 5 5   6 6            7 7 7 7        tDNAs per genome (elongators + initiators)
b   12      8   1   4  7789 001   4       0  23 4 6 7 8  0    4   8      3 4 6   0 1            2 3 4 6                  genome size (Mb)       genome number
c                       a   ab                                                                          Nr      G+C %   genome name   abbreviated name
d   >>>>>>>    >>>>    **   **  <<<<   >>>>>    A C     <<<<<    V  >>>> >         < <<<< <<<<<<<   CCA CCA

E01 B------ TR GY-- ARHh*GG-yyA --RY - ---B- YY --- R H BB--- --- Y ---- G TTCRA-Y C ---- ------- - ---    0 274      12.07 Saccharomyces cerevisiae    Sc E01
E02 ------D TR -Y-- AR-y*GGBbtR --R- - ----- YT --- R H ----- --- Y V--D R TTCRA-Y Y HB-B H------ - ---    0 175      13.01 Schizosaccharomyces pombe   Sp E02
E03 -V----- TV KY-- AD-h*GG-bbA D-RH - ---B- YT --- R H -B--D --V Y ---V G WTVRA-Y C B--- ------- - ---    0 529      97.23 Caenorhabditis elegans      Ce E03
E04 B------ T- KY-B AD-y*GG-ytA D-RH - ---B- YT --- R H -B--- --- - ---D G WTCRA-B C H--- ------- - ---    0 282     115.23 Drosophila melanogaster     Dm E04
E05 D------ TV GY-- A--h*GG-bkA D-RY - ----- YY --- R H -B--- -VD - ---D G WTYRA-- C H--- ------H - ---    0 157     477.07 Homo sapiens                Hs E05
E06 ------- TR KY-- ARBh*GGHhhA --RY - ---B- HT --- R H -B--- --- Y ---R G TTCRA-Y C Y--- ------- - ---    0  46       2.48 Encephalitozoon cuniculi    Eu E06
E07 ------- TV BY-B ADHh*GG-hyA D-R- - ----- YT --- R H -B--- --- H ---- G HTYRA-H Y HB-- ------- - ---    0 562     115.55 Arabidopsis thaliana        At E07

Ee  ------- T- -Y-- A--h*GG-nnR --R- - ----- HY --- V H ----- --- - ---- R HT-RA-- Y ---- ------- - ---     1984 elongators (302 families)
Ei  WKCR-VG TR GBRS ARY**GGA**A SYSY G VHGGG CY CAT A A CYCDB AKG Y V--R G WTCKAAW C Y--B CB-YGMW A ---       41 initiators (  7 families)   total  2025
  A box ->  T- -B-- A--hGG cs1                                 B box ->  R HT-BA-- Y
            T- -B-- A--GG  cs2

A01 VSSSSSG YR SBBB AGYhhGG-ynH DVVB B SSSSV HT B-- R V BSSSS -DR C CCSG G TTCAART C CSGG CSSSBSB - CHH   29  34  61   1.69 Methanopyrus kandleri       Mk A01
A02 VSVSBSR TD SBBB AGHhhGG-nw- DVVB V BSVSV YT B-- R V BSBSV --- C SSSG G TTCRAAK C CSSS YSVSBSB - CYA   44  46  45   1.77 Pyrococcus abyssi           Pa A02
A03 VSSSSSG TV SYHB AGYyhGG-ynR DDRB G B-BSV YT B-- R H BSVB- -BV C SBGG G TTCRAAT C CCDB CSSSBSB - Y--   22  46  51   2.22 Pyrobaculum aerophilum      Pe A03
A04 VSSSSSG TV SYHB AGChhGG-ynA RDRB G -SVSV YT B-- R H BSBSD HRV C SBGG G TTCRART C CCVB CSSSBSB - CCA  all  46  57   1.67 Aeropyrum pernix            Ap A04
A05 VS-V-SD TD GB-B AG-hhGG-nwD --VY - ---S- YT B-- R - -B--- --- S B-VV G TTCRAHY C BB-- HS-B-SB - B--    0  46  49   2.18 Archaeoglobus fulgidus      Af A05
A06 -B----D TR G--- RR-yhGG-hhM ---Y - ---S- HT B-- R - -B--- --- Y ---- R TTCRA-H Y BB-- H----V- - Y--    0  47  68   2.01 Halobacterium sp. NCR-1     Ha A06
A07 VS-BVVG TV SYYB AGByhGG-ynV RRRB V ---S- YY B-- R H BS--- -D- C V-RG G TTCRART C CY-B CBBV-SB - B--    1  46  36   2.99 Sulfolobus solfataricus     Ss A07
A08 VS-VSBR TV SYHB AGHhhGG-hwR RDRB V --VS- YY B-- R H -SB-- -D- C SBRG G TTCRARK C CYVS YVSB-SB - Y--    1  46  33   2.69 Sulfolobus tokodaii         St A08
A09 VB----D TD R--- ARHhhGG-ywV ---Y D ---S- YT B-- R - -B--- --- Y --BR G WYCRA-- C YS-- H----VB - Y--    1  46  46   1.56 Thermoplasma acidophilum    Ta A09
A10 VS---BD TD R-B- AAYhhGG-hwR -V-Y D ---B- YT --- R - -B--- --- Y --BD R WYCRADY Y HS-- HV---BB - H--    0  45  37   1.93 Ferroplasma acidarmanus     Fa A10
A11 V-----D TD R--- AG-hhGG-yhD ---Y - ---S- YY B-- R - -S--- --- Y --V- R TTCRADY Y -B-- H-----B - Y--    5  58  40   5.13 Methanosarcina barkeri      Mb A11
A12 VS-B-BD TV GBBH AG-yhGG-ywD -VVY - -B-SV YY B-- D V BS--- --- C ---D G TTCRADY C HB-B HV-V-SB - YYW   25  36  31   1.67 Methanococcus jannaschii    Mj A12
A13 VS----D TR SBBB AR-hhGG-hnR -VVB - ---SV YY B-- R - BS--- -DV Y ---R G TTCRA-Y C YB-B H----SB - Y--    1  39  50   1.75 Methanobacterium            Mt A13
                                                                                                                                thermoautotrophicum

Ae  ------D Y- V--- RR-hhGG-nn- ---B - ---B- HY --- D - -B--- --- B ---- R WYCRA-- Y -B-- H------ - ---      566 elongators (550 families)
Ai  AGCGGSR YR GGVY AGYywGGHvcW KBCC G VBGGG CT CAT A A CCCSB AGR C VV-R G TTCRART C Y-BB YSCCGCT A YH-       15 initiators ( 13 families)   total   581

B01 B-----D T- VB-- W--h*GG-ndD D-VB V ---S- HY B-- R - -B--- -D- - ---V G TTCRA-B C B--- H------ - ---    1  45  53   1.14 Treponema pallidum          Tp B01
B02 -B----- TV VH-- MD-y*GGHhdR --DB - ---B- HT B-- R - -V--- -D- B ---D G TTCRA-H C H--- ------- - H--    0  33  29   0.91 Borrelia burgdorferi        Bb B02
B03 BB----D TV DY-- AR-ytGGHhnD D-RH - ---B- HT --- R H -S--- --- Y V--D G TTCRADY C H--B H------ - ---    0  37  41   1.04 Chlamydia trachomatis       Ct B03
B04 BV----- TV SY-- AR-ytGGHhdD D-RB - ---V- HT --- R - -B--- --- B ---V R TTCRA-Y Y B--- -----B- - YH-    0  42  48   3.57 Synechocystis sp. PCC 6803  Sy B04
B05 BSV---- TV DY-- AR-ttGGHhdD D-RH - ---V- HT --- R H -B--- --- B ---V R TTCRA-Y Y B--- ----BB- - Y--    2  48  41   6.41 Anabaena sp. PCC 7120       An B05
B06 -V----D TV BY-- AR-hyGG-hrR D-R- - ---B- HT --- R H -B--- --- Y V--D G TTYRA-H C H--B H----B- - Y--    0  61  35   2.37 Lactococcus lactis          Ll B06
B07 ------- TR BYB- AR-yyGG-hrR DVR- D ---S- HT --- R - -S--- --- Y V--R G TTCRA-H C Y--B ------- - Y--   16  67  38   2.94 Listeria monocytogenes      Lm B07
B08 ------- TR BY-- ARYhyGG-hrR R-R- - ---S- HT --- R - -S-D- --- Y V--R G TTCRA-Y C Y--B ------- - YH-   59  84  44   4.21 Bacillus subtilis           Bs B08
B19 ---V--S TV GYBB AR-hyGG-hdD RVRC R ----- HY --- R - B---- -D- Y S-VG G TTCRAVT C CB-B S--B--- - Y--   32  43  43   1.55 Aquifex aeolicus            Ae B09
B10 -B----D TV -Y-- AR-ytGG-hdD D-R- - ---S- HT --- R - -S--- --- Y ---V G TTCRADH C B--B H--B--- - C--   16  45  66   4.41 Mycobacterium tuberculosis  Mt B10
B11 -B----- TV KB-- ARByyGG-hdR R-VH R ---S- HT --- D - -S--- --- Y ---R G TTCRA-Y C Y--- ------- - YYW   47  48  67   3.06 Deinococcus radiodurans     Dr B11
B12 -BV---D T- SH-- AR-hyGGHhdD R-DB V ----- HT --- R - -B--- -D- Y V--V G TTCRA-H C B--B H---BV- - YYW   57  59  52   2.27 Neisseria meningitidis      Nm B12
B13 -B----D T- BY-- ARYyyGG-hdD D-R- V ---V- HT --- R H -B--- -D- Y ---D G TTCRA-Y C H--- H----V- - YSM   60  62  57   6.26 Pseudomonas aeruginosa      Pr B13
B14 -B-B--D T- -Y-- ADHhtGG-hwD D-R- - ---B- HT --- R - -S--- --- H ---R G TTCRADY C Y--B H--V-V- - H-W   10  32  26   0.64 Buchnera sp. APS            Bu B14
B15 ------- TR BYB- ARBytGG-hrR DVR- V ---S- HT --- R - -S--- --- Y V--R G TTCRA-Y C Y--B ------- - YH-   14  78  44   4.20 Bacillus halodurans         Bh B15
B16 ------D TV BB-- MR-nyGG-nwR R-V- R ---S- HT B-- R V -B--- DDD Y VB-G G TTCRADY C C-VB H------ - YMR   45  46  46   1.86 Thermotoga maritima         Tm B16
B17 ------D TV BBB- MR-htGG-hrV VVV- V ----- YY B-- V - -B--- -D- Y ---R G TTCRADY C Y--- H------ - CCA  all  43  31   1.64 Campylobacter jejuni        Cj B17
B18 -B----- T- BY-- MR-htGG-hdV D-R- V ---B- HT --- R H -B--- -D- Y ---R G TTCRA-Y C Y--- -----V- - CCA  all  97  48   4.03 Vibrio cholerae             Vc B18
B19 ------- T- -YBV AR-y*GG-hdR RVR- R ---B- YT B-- R V -S--- -D- Y V--R G TTCRA-Y C Y--B ------- - CCA  all  93  29   3.03 Clostridium perfringens     Vp B19
B20 ------D TV -BB- MR-ytGR-hdR VVV- R ---B- YT B-- R - -S--- -D- Y V--R G TTYRA-Y C Y--B H------ - CCA  all  36  39   1.67 Helicobacter pylori         Hp B20
B21 -BV---D TV SH-- AD-htGG-hdV R-DB V ---B- HT --- R - -S--- --- B V--D G TTCRA-H C H--B H---BV- - CCA  all  56  67   5.81 Ralstonia solanacearum      Rs B21
B22 ------- TR SYB- AR-ntGR-hnR RVRB - ---V- -T B-- R H -B--- --- B V--R G TTYRA-H C Y--B ------- - CCA  all  36  32   0.58 Mycoplasma genitalium       Mg B22
B23 ------- TR SYB- AR-ntGR-hnR RVRB - ---V- -T B-- R V -B--- --- B V--R G TTYRA-H C Y--B ------- - CCA  all  36  40   0.83 Mycoplasma pneumoniae       Mp B23
B24 -V----- TR SY-- ARHhyGR-hrR R-RB - ---V- YT --- R H -B--- --- Y ---R G TTYRA-H C Y--B -----B- - CCA  all  30  26   0.75 Ureaplasma urealyticum      Uu B24
B25 -B----D TV BY-- AR-hkGRHhdV --R- - ---V- HT --- R - -B--- -D- Y V--R G TTCRA-H C Y--B H----V- - CCA  all  48  53   2.68 Xilella fastidiosa          Xf B25
B26 -B----- T- BY-- AR-htGG-hdD D-R- V ---B- HT --- R - -B--- -D- B ---R G TTCDA-H C Y--- -----V- - CCA  all  56  38   1.83 Haemophilus influenzae      Hi B26
B27 -B----D T- -Y-- AR-htGG-hvD D-R- V ----- HT --- R H -B--- -D- B ---R G TTCRA-Y C Y--- H----V- - CCA  all  85  51   4.64 Escherichia coli            Ec B27
B28 -V----D TV RY-- AR-ytGGHhdR D-RY - ---S- HT --- R - -B---- -- Y ---R G TTCRA-T C Y--- H----B- - CCA  all  32  29   1.11 Rickettsia prowazekii       Rp B28
B29 -B----D T- BY-- AR-ytGG-hdV D-R- - ---B- HT --- R H -B--- -D- Y ---R G TTCRA-Y C Y--- H-B--V- - CCA  all  67  48   4.83 Yersinia pestis             Yp B29
B30 -V----D TV SB-- ARBhyGGDhdR R-VB - ---S- HT --- R - -S--- -D- Y V--- G TTCRA-Y C ---B H----B- D CCA  all  53  62   6.69 Sinorhyzobium meliloti      Sm B30

Be  ------- T- ---- H--nbGR-nn- ---- - ----- -Y --- - - ----- --- - ---- R TTYDA-- Y ---- ------- - ---     1546 elongators (1111 families)
Bi  YGCRRRR TR GAGY AGYhhGGYy*A KCTC G YYRGG CT CAT A A YCYRR AGR C R--R G TTCRART C Y--Y YYYYGCH A H--       52 initiators (  30 families)  total  1598
```

FIGURE 3. *See caption on facing page.*

G18, and G19, or A19 in a few tDNAs of some bacteria, all colored in blue in Figs. 3 and 4). These Eukarya-specific conservations are: Y11 [C or T, but only B11 (C, G, or T) in initiators], A21, R24 (all colored in red), and a systematic lack of nucleotide at position 17a (red asterisk). Together with the other universally conserved bases of the T-loop (in blue), these bases make up the two regions that participate in the recognition of the tRNA genes by the eukaryotic RNA polymerase III machinery ("A box" and "B box" in Fig. 3; see also below, Distribution of tRNA genes in genomes and peculiarities of their transcription in Eukarya). Base H38 (A, C, or T) is also conserved in Eukarya and because it is located 3′ to the intron, this conservation may result from special requirements of the eukaryotic splicing machinery that is clearly distinct from that of Archaea (reviewed in Abelson et al., 1998; Edgell et al., 2000). As a rule, in Eukarya and Bacteria, many more bases than in Archaea (especially the hyperthermophilic ones) appear to be randomly distributed (as shown by the larger number of N bases, drawn as - in Fig. 3), especially in the stems of elongator tRNA. Evidently, preferences exist in tDNA sequences, specific to each domain of life.

Initiator tDNA-iMet is the most conserved tDNA across the three domains, but nevertheless has domain-specific features

In the set of 50 genomes, a total of 108 tDNAs-iMet were retrieved that represent 50 isoacceptor families. Consensus sequences for the initiator tDNA-iMet within each domain are presented in Figure 3 in line Ei for Eukarya (41 sequences), line Ai for Archaea (15 sequences), and line Bi for Bacteria (52 sequences); the corresponding cloverleaves are presented in the lower row of Figure 4. Clearly, a larger number of positions are more conserved (colored in green) than in the elongator tDNA consensus sequences (compare with lines Ee, Ae, and Be in Fig. 3 and upper row in Fig. 4). This result confirms and extends earlier observations concerning the identity and conservation of characteristic nucleotides in tDNA-iMet and their corresponding mature tRNA-iMet (LaRue et al., 1979; Cedergren et al., 1981), reviewed in RajBhandary and Chow (1995). Among these characteristic invariant or semi-invariant bases are those corresponding to bp 1–72, 11–24, base opposition 54–60, and a large part of the anticodon hairpin (all boxed in Fig. 4, lower part).

FIGURE 3. Consensus sequences of elongator and initiator tDNAs in the three biological domains. Seven eukaryotic genomes (E01 to E07, of which *H. sapiens* (E05) is presently incomplete), 13 archaeal genomes (A01 to A13), and 30 bacterial genomes (B01 to B30) have been examined and the consensus sequences of the tDNA in each genome and domain computed. For archaeal genomes, the organism are listed according to the temperature at which the organism thrives (highest temperatures first); the four hyperthermophilic (A01 to A04) then other extremophiles (*A. fulgidus* (A05), *Halobacterium* sp. NRC-1 (E06), *S. solfataricus* (A07), *S. tokodaii* (A08), and *T. acidophilum* (E09)) and mesophiles (A10 to A13) last. Bacterial genomes have been ordered to emphasize neighboring features in the conservation of their tDNA sequences and the increasing proportion of encoded 3′ CCA-termini. The position of some bases in the tRNA sequence are given in lines a to c (to be read vertically). On line d are indicated the stems (>, direct strand; <, antiparallel strand); the four ∗ indicate the four optional positions of the D-loop (positions 17, 17a, 20a, and 20b); AC refers to the anticodon; V indicates the position 47 (additional bases are sometimes present in the long variable arm). Additional data listed beyond base 73 are as follows. Under CCA is given the consensus sequence of the 3 bases downstream of base 73 in the genomic sequence. A CCA consensus sequence at these positions indicates that all tDNAs encode the 3′ CCA sequence. Under Nr CCA is indicated, for Archaea and Bacteria, the number of tRNA genes in which the CCA sequence is encoded; all is indicated instead, if all genes encode the CCA sequence; Eukarya never encode the 3′ CCA sequence. Consensus sequences Ee (lines E01 to E07, 1,984 sequences), Ae (A01 to A13, 566 sequences), and Be (B01 to B30, 1,546 sequences) were computed by cumulating the sequences of all elongator tDNAs (all but initiator tDNA-iMet (CAT)), within each domain. Below, on lines Ei, Ai, and Bi are given the consensus sequences of initiator tDNA-iMet (CAT) (Ei, 41; Ai, 15; and Bi, 52 sequences). At the four optional positions of the D loop (positions 17, 17a, 20a, and 20b), a lower case letter indicates that this position is sometimes unoccupied in the tDNAs from a given organism (e.g., in line E01, at position 17). ∗ in a genome consensus sequence or in a domain consensus sequence at these optionally occupied positions indicates that the position is never occupied in this genome or domain (e.g., at position 17a in line E01 and in lines Ee and Ei, respectively). Note that all four positions are always unoccupied in eukaryotic initiator tDNAs (line Ei). At the right end of these six consensus lines are indicated the number of sequences used (nb of tRNA genes) and the same value corrected for genes present in multiple copies (number of families). Blue letters indicate conserved bases shared among the three domains of life; many of these participate in 3D base pairings that stabilize the tertiary structure of tRNA (see Fig. 2B). Red bold letters emphasize, in each of the three domains, positions more specifically conserved inside a given domain. Red is also used to enhance the S-S (G-C or C-G), G-C, and C-G base pairs, in the stems of the hyperthermophilic archaeons *M. kandleri* (A01), *P. abyssi* (A02), *P. aerophilum* (A03), and *A. pernix* (A04). In initiator tDNA-iMet (CAT) consensus sequences (lines Ei, Ai, and Bi); green letters indicate bases that are conserved among all the initiator tDNAs-iMet (CAT) in all three biological domains; paired bases 11 and 24, which are differently conserved in the three domains, appear in purple. The two boxes indicated "A box" and "B box" highlight the sequences used as internal promoter elements by the eukaryotic transcription machinery. As position 17a is never occupied in eukaryotic tDNA, the fourth base 3′ to the conserved base A14 is always a G as illustrated in the two consensus sequences "cs1" (position 17 occupied) and "cs2" (position 17 not occupied) written without gaps.

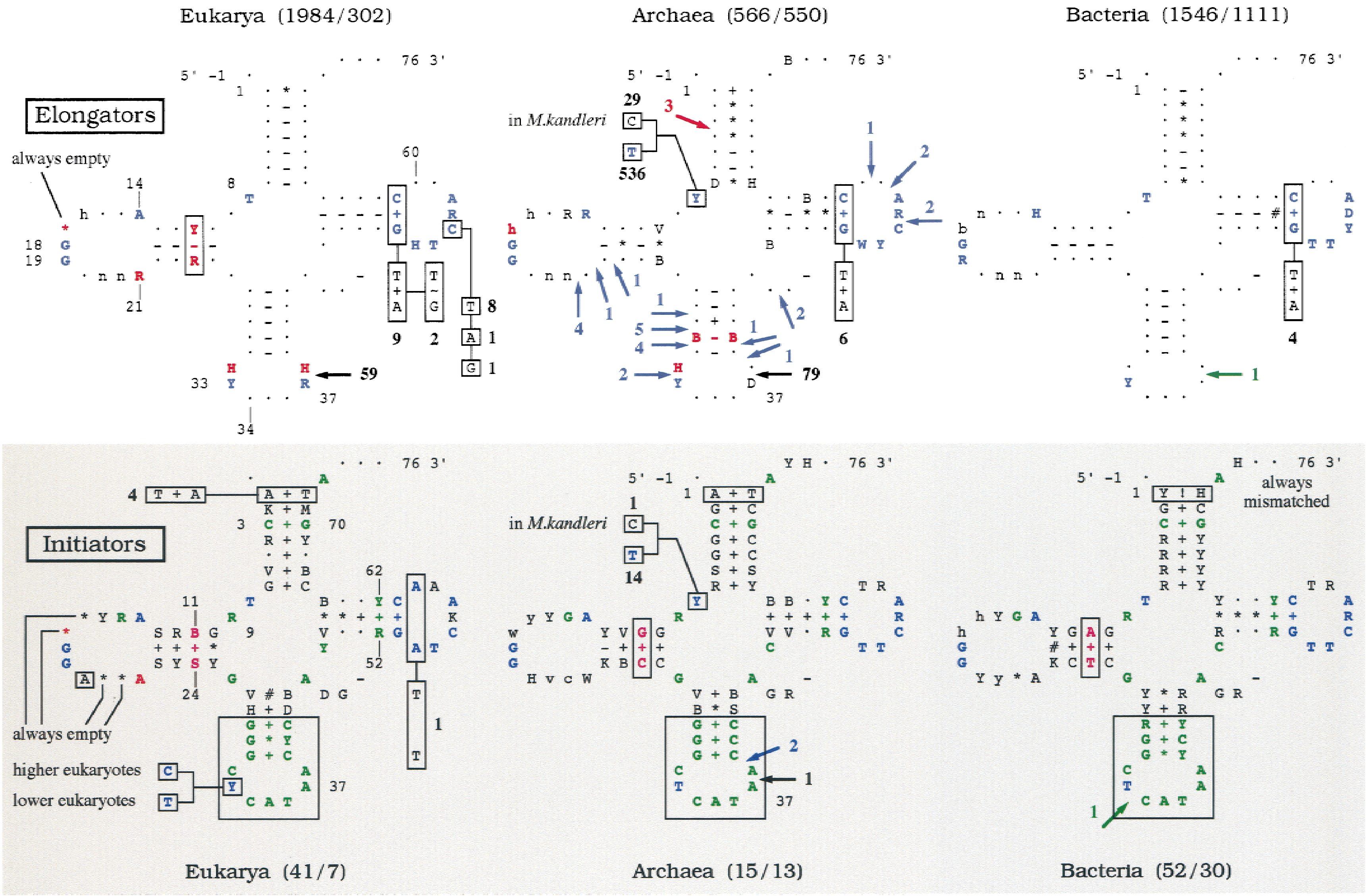


FIGURE 4. *See caption on facing page.*

A highly typical feature of tDNA-iMet is the base pair at positions 1–72, where the Watson–Crick base pair A-T (or T-A in the four copies of the tDNA-iMet of *Schizosaccharomyces pombe* (E02)) is invariably found in all eukaryotic and all archaeal tDNA-iMet. Considering the strong pressure for a G-C base pair and the counterselection of an A-T base pair at these positions in all elongator tDNAs and especially in those from Archaea (Table 3), the A-T base pair at 1–72 of tDNA-iMet is remarkable. In yeast *S. cerevisiae* (E01), it has been demonstrated that A1-U72 is the most important determinant for a tRNA to play the role of initiator tRNA (Åström et al., 1993). Likewise, the systematic occurrence of a mismatched base opposition Y!H (a pyrimidine mismatched with any base but G) is a signature for bacterial tDNA-iMet (RajBhandary & Chow, 1995). Such a characteristic mismatched base pair, as well as other conserved base pairs in the amino acid stem, were also demonstrated to be key determinants for the formylation of the Met moiety of tRNA-iMet by the Met-RNA transformylase that is present only in Bacteria (Lee et al., 1991; Guillon et al., 1992). Moreover, this 1–72 mismatch is, among other features, an antideterminant for the bacterial elongation factor EF-Tu (Rudinger et al., 1996).

In bacterial tDNA-iMet, bp 11–24 is invariably A-T, while G-C is exclusively found in archaeal tDNA-iMet with strong counterselection of C-G in elongator tDNA from both domains (Tables 1 and 2). It is noteworthy that, in all eukaryotic tDNA-iMet, positions 11–24 are occupied by C-G, with the exception of *S. pombe* (E02, G-C in the four copies) and a single case of T-G (in 1 of the 10 copies of tDNA-iMet of *A. thaliana* (E07), probably a true polymorphism). However, eukaryotic elongators also prefer C-G at these positions. Initiator tDNAs from Archaea and Bacteria use both T54 and T60, whereas those from Eukarya use A54 and A60. Some eukaryotic elongators also use either A54 or A60, but none (with only one exception) uses both. The parasitic eukaryote *Encephalitozoon cuniculi* (E06) is a remarkable exception to this rule, as its tDNA-iMet (CAT) bears the archaeal/bacterial signature T54 and T60. Genetic experiments performed with null alleles of the five elongator tDNA-Met of *S. cerevisiae* have demonstrated that the conserved U54 is an important determinant for an elongator tDNA (Åström et al., 1993); in other words, A54 in eukaryotic initiator tDNA-iMet plays the role of an antideterminant.

Beside these domain-specific features, the whole anticodon loop, together with the lower part of the anticodon stem (both boxed in Fig. 4), are characteristic and similar for the tDNA-iMet of the three biological domains. Indeed, initiator tDNA-iMet (CAT) from all domains display the GGG sequence (Mandal et al., 1996) or, very seldom, the AGG sequence at positions 29 to 31, pairing with the complementary CCC or CCT sequences at positions 39 to 41. In Archaea, the lower part of the anticodon stem is composed exclusively of three G-C pairs, whereas in Bacteria, an A29-T41 base pair is found only in the single copy of tDNA-iMet (CAT) of *Mycoplasma genitalium* (B22), *Mycoplasma pneumoniae* (B23), *Ureaplasma urealyticum* (B24), and *Rickettsia prowazekii* (B28), and in the three copies of the same tDNA in *Sinorhizobium meliloti* (B30). In Eukarya and Bacteria, the second and third Gs are occasionally paired to a T, like in one of the six initiators of *D. melanogaster* (E04) harboring a G30-T40 and in *M. genitalium* (B22), *M. pneumoniae* (B23), and *U. urealyticum* (B24) harboring a G31-T39. A single exception to this RGG rule is the elongator tDNA-Met (CAT) of *M. kandleri* (A01) that bears GGG at positions 29 to 31 of the anticodon stem. However, the initiator tDNA-iMet (CAT)

FIGURE 4. Cloverleaf representation of consensus sequences of elongator and initiator tRNA genes in the three domains. The consensus sequences of all elongator tDNAs (top row, taken from lines Ee, Ae, and Be in Fig. 3) and of all initiator tDNAs (bottom row, taken from lines Ei, Ai, and Bi in Fig. 3) are presented under the cloverleaf model. Between parentheses are indicated (1) the total number of tDNAs in each phylogenetic domain, (2) the same value corrected for the tDNA present in multiple copies inside a given genome (e.g., 15 archaeal initiator tDNA-iMet (CAT) have been discovered but 2 of these archaeons have two copies). See legend of Figure 2 for key to base pairing; "!" is used to indicate that all base pairs are mismatched. Colors have the same meaning as in Figure 3. Blue denotes bases universally conserved, red indicates bases conserved inside a domain. For initiators (lower row and gray background), green indicates bases conserved in all initiators throughout the three domains. Purple is used to emphasize the bp 11–24 of initiator tDNAs, which is differently conserved in the three domains. Boxes are used to enhance remarkable features. Boxes outside the cloverleaf are used to reveal the different configurations making up the consensus at bp 53–61 and position 56 in elongator tDNAs and at base pair 1–72 and base opposition 54–60 in initiator tDNAs. A number close to these boxes indicates the number of times the alternate configurations occur. For example, in the 1,984 elongator tDNAs of Eukarya (upper leftmost cloverleaf), bp A53-T61 is found nine times, G53-T61 twice, and therefore G53-C61 1,973 times. Arrows indicate the positions of introns: black: unique position in Eukarya (between bases 37 and 38), also used in Archaea; blue: alternate introns positions in Archaea (taken from the literature or determined by visual inspection); red: proposed novel location for archaeal introns (see text, Length, localization, and distribution of introns in eukaryotic and archaeal tDNA); green: autocatalytic group I introns present in the tDNA-iMet (CAT) of *Synechocystis* 6803 (B03) (655 bp) and tDNA-Leu (TAA) of *Anabaena* 7120 (B04) (249 bp). A number close to an arrow and in the same color indicates the number of times an intron is found at that position (in all tDNAs of the phylogenetic domain). See Figure 8 for more details on eukaryotic and archaeal introns.

TABLE 1. Base distribution in all tRNA genes.[a]

Requested bases	A. Elongator tDNAs												B. Initiator tDNAs-iMet (CAT)											
	Eukarya 1984/302				Archaea 566/550				Bacteria 1546/1111				Eukarya 41/7				Archaea 15/13				Bacteria 52/30			
Base #	A	G	T	C	A	G	T	C	A	G	T	C	A	G	T	C	A	G	T	C	A	G	T	C
> 1	8	1685	193	98	29	523	1	13	48	1315	97	86	37	—	4	—	15	—	—	—	—	—	3	49
> 2	115	740	213	916	1	255	6	304	35	844	78	589	—	32	9	—	—	15	—	—	—	52	—	—
> 3	169	499	433	883	29	309	25	203	235	772	130	409	—	—	—	41	—	—	—	15	—	—	—	52
> 4	174	556	589	665	35	246	27	258	98	804	243	401	32	9	—	—	—	15	—	—	3	49	—	—
> 5	339	526	400	719	29	152	62	323	317	527	155	547	1	30	1	9	—	15	—	—	1	51	—	—
> 6	208	453	672	651	11	286	35	234	284	394	436	432	23	5	—	13	—	14	—	1	2	50	—	—
> 7	612	877	444	51	47	503	16	—	385	801	327	33	—	41	—	—	1	14	—	—	12	40	—	—
8 Y	—	—	1984	—	—	—	537	29	—	—	1546	—	—	—	41	—	—	—	14	1	—	—	52	—
9	797	1158	12	17	309	199	27	31	1075	363	20	88	5	36	—	—	5	10	—	—	10	42	—	—
> 10	13	1876	50	45	9	524	—	33	17	1438	25	66	—	41	—	—	—	15	—	—	—	52	—	—
> 11	—	—	742	1242	12	49	173	332	3	8	300	1235	—	4	1	36	—	15	—	—	52	—	—	—
> 12	106	485	695	698	13	112	217	224	48	272	969	257	4	37	—	—	1	12	—	2	—	52	—	—
> 13	35	311	773	865	13	77	283	193	190	257	111	988	—	4	—	37	—	—	12	3	—	—	1	51
14	1984	—	—	—	565	1	—	—	1540	—	1	5	41	—	—	—	15	—	—	—	52	—	—	—
15	364	1608	11	1	11	555	—	—	342	1195	7	2	1	40	—	—	—	15	—	—	—	52	—	—
16	89	56	1643	196	30	19	126	391	62	80	932	472	—	—	23	18	—	—	4	11	—	—	25	27
* 17	39	—	457	160	16	—	92	310	28	3	790	174	—	—	—	—	—	—	1	14	1	—	21	28
* 17a	—	—	—	—	99	—	198	64	—	1	67	11	—	—	—	—	10	—	5	—	1	—	32	9
18 G	—	1984	—	—	—	566	—	—	—	1546	—	—	—	41	—	—	—	15	—	—	—	52	—	—
19 R	—	1984	—	—	—	566	—	—	14	1532	—	—	—	41	—	—	—	15	—	—	—	52	—	—
20	124	56	1657	147	118	27	288	133	141	127	1016	262	41	—	—	—	7	—	7	1	—	—	47	5
* 20a	29	29	867	204	9	3	127	204	169	2	512	115	—	—	—	—	2	8	—	2	—	—	2	2
* 20b	1	5	151	12	95	5	56	23	46	90	61	7	—	—	—	—	—	—	—	2	—	—	—	—
21	1981	3	—	—	529	25	7	5	1357	171	11	7	41	—	—	—	14	—	1	—	52	—	—	—
< 22	475	1117	383	9	115	306	130	15	459	1055	28	4	—	37	—	4	—	5	10	—	—	51	1	—
< 23	716	681	103	484	215	226	13	112	965	262	48	271	—	—	4	37	—	2	1	12	—	—	—	52
< 24	727	1257	—	—	168	337	12	49	282	1253	3	8	—	37	—	4	—	—	—	15	—	—	52	—
< 25	49	47	495	1393	—	33	154	379	24	66	139	1317	—	—	31	10	—	—	—	15	—	—	—	52
26	348	1248	319	69	58	452	31	25	900	568	31	47	—	41	—	—	—	15	—	—	—	52	—	—
> 27	126	184	806	868	43	105	72	346	213	262	240	831	5	9	—	27	5	1	—	9	—	—	50	2
> 28	355	203	752	674	49	149	90	278	252	120	352	822	5	—	32	4	—	5	1	9	—	—	4	48
> 29	462	422	576	524	66	374	54	72	349	505	439	253	—	41	—	—	—	15	—	—	7	45	—	—
> 30	10	1681	67	226	—	474	2	90	14	1149	19	364	—	41	—	—	—	15	—	—	—	52	—	—
> 31	786	351	270	577	140	165	7	254	551	190	126	679	—	41	—	—	—	15	—	—	—	52	—	—
32	1	—	594	1389	2	—	116	448	55	2	537	952	—	—	—	41	—	—	—	15	—	—	—	52
33	—	—	1982	2	—	—	560	6	—	—	1542	4	—	—	10	31	—	—	15	—	—	—	52	—
| 34	500	508	520	456	1	215	175	175	46	599	591	310	—	—	—	41	—	—	—	15	—	—	—	52
| 35	451	502	609	422	153	149	126	138	431	367	391	357	41	—	—	—	15	—	—	—	52	—	—	—
| 36	438	442	534	570	114	146	151	155	307	342	454	443	—	—	41	—	—	—	15	—	—	—	52	—

37	1332	652	—	—	269	296	1	—	1278	266	1	1	41	—	—	—	15	—	—	—	52	—	—	—
38	1138	—	379	467	454	25	23	64	1092	34	177	243	41	—	—	—	15	—	—	—	52	—	—	—
< 39	180	631	818	355	6	256	141	163	115	689	546	196	—	—	—	41	—	—	—	15	—	—	3	49
< 40	6	286	90	1602	—	90	7	469	2	378	46	1120	—	—	1	40	—	—	—	15	—	—	—	52
< 41	572	525	464	423	54	72	66	374	436	255	350	505	—	—	—	41	—	—	—	15	—	—	7	45
< 42	693	730	373	188	86	282	63	135	329	834	276	107	32	4	5	—	—	10	—	5	4	48	—	—
< 43	690	980	258	56	59	357	79	71	210	843	345	148	—	27	4	10	—	9	5	1	47	5	—	—
44	1229	105	445	205	197	32	266	71	503	397	381	265	41	—	—	—	15	—	—	—	52	—	—	—
45	103	1704	71	106	39	449	43	35	254	1086	183	23	—	36	5	—	—	15	—	—	—	52	—	—
46	522	1273	59	130	169	322	21	54	180	1094	182	90	—	41	—	—	9	6	—	—	1	51	—	—
48	50	2	297	1635	—	1	14	551	4	11	331	1200	—	—	1	40	—	—	—	15	—	—	—	52
> 49	204	884	74	822	38	198	41	289	343	1051	28	124	1	16	—	24	1	7	—	7	4	48	—	—
> 50	92	337	461	1094	27	108	93	338	204	501	292	549	7	14	6	14	5	5	—	5	1	11	28	12
> 51	310	1221	240	213	11	484	10	61	316	765	180	285	20	1	15	5	2	1	11	1	22	6	11	13
> 52	339	1563	61	21	34	519	6	7	230	1294	10	12	2	39	—	—	2	13	—	—	1	51	—	—
> **53 R**	9	1975	—	—	6	560	—	—	4	1542	—	—	—	41	—	—	—	15	—	—	—	52	—	—
54 H	39	—	1935	10	4	—	562	—	—	—	1546	—	40	—	1	—	—	—	15	—	—	—	52	—
55 Y	—	—	1984	—	—	—	546	20	—	—	1546	—	—	—	41	—	—	—	15	—	—	—	52	—
56	1	1	8	1974	—	—	—	566	—	—	12	1534	—	—	—	41	—	—	—	15	—	—	—	52
57	453	1531	—	—	393	173	—	—	340	1205	1	—	—	40	1	—	12	3	—	—	46	6	—	—
58 A	1984	—	—	—	566	—	—	—	1546	—	—	—	41	—	—	—	15	—	—	—	52	—	—	—
59	639	355	816	174	493	44	24	5	597	536	332	81	41	—	—	—	14	1	—	—	46	6	—	—
60	32	17	1283	652	2	6	539	19	20	1	1300	225	40	—	1	—	—	—	15	—	—	—	52	—
< **61 Y**	—	—	11	1973	—	—	6	560	—	—	4	1542	—	—	—	41	—	—	—	15	—	—	—	52
< 62	61	18	357	1548	4	9	37	516	10	12	220	1304	—	—	2	39	—	—	2	13	—	—	1	51
< 63	80	366	318	1220	—	71	12	483	145	313	367	721	15	5	20	1	11	1	2	1	9	15	22	6
< 64	381	1159	110	334	87	343	38	98	231	585	383	347	6	14	13	8	—	5	5	5	3	37	8	4
< 65	53	842	293	796	26	304	64	172	22	126	738	660	—	24	9	8	—	7	1	7	—	—	7	45
< 66	444	51	645	844	16	—	59	491	326	34	425	761	—	—	—	41	—	—	1	14	—	—	12	40
< 67	406	857	390	331	31	238	11	286	379	483	308	376	—	13	23	5	—	1	—	14	—	—	2	50
< 68	330	810	494	350	42	343	32	149	100	595	365	486	1	9	1	30	—	—	—	15	—	—	1	51
< 69	357	880	272	475	19	266	37	244	195	447	132	772	—	—	32	9	—	—	—	15	—	—	3	49
< 70	406	907	306	365	25	203	73	265	126	413	357	650	—	41	—	—	—	15	—	—	—	52	—	—
< 71	190	941	308	545	5	305	2	254	76	591	77	802	9	—	—	32	—	—	—	15	—	—	—	52
< 72	193	98	42	1651	1	13	29	523	97	85	110	1254	4	—	37	—	—	—	15	—	43	—	7	2
73	1046	572	217	149	350	151	51	14	889	427	194	36	41	—	—	—	15	—	—	—	52	—	—	—
C 74	775	266	750	193	1	2	223	340	41	11	261	1233	10	13	15	3	—	—	7	8	1	—	7	44
C 75	593	170	1002	219	62	35	229	240	115	33	214	1184	18	9	13	1	3	—	5	7	4	1	3	44
A 76	461	184	1104	235	275	52	154	85	1248	76	161	61	13	6	14	8	6	4	3	2	45	2	3	2

[a] This table shows the base distribution (number of A, C, G, or T in each position) found in the 4,204 tDNAs, the consensus sequences of which are presented as linear and cloverleaf consensus in Figs. 3 and 4, respectively. On the top line are indicated the total number of tDNAs in each domain, and the same value corrected for the tDNAs present in multiple copies inside a given genome (e.g., 15 archaeal initiator tDNA-iMet (CAT) have been discovered, but 2 archaeons among the 13 examined have two copies of this tDNA). Base positions are listed at left with > and < indicating the direct and antiparallel strands of the four stems, respectively. ∗ indicates the 4 optional bases 17, 17a, 20a, and 20b. Vertical bars indicate the anticodon (positions 34 to 36). Positions and letters on the left side of the table corresponding to the bases in bold are those constrained in the cloverleaf tDNA-search algorithm. In the array, — is used instead of 0 for clarity. **A:** Elongator tDNAs: 4,096 tDNAs: Eukarya, 1,984 sequences (from line (Ee) in Fig. 3); Archaea (from line (Ae)), 566; Bacteria, 1,546 (from line (Be)). **B:** Initiator tDNAs-iMet (CAT): 108 tDNAs: Eukarya, 41 (from line (Ei) in Fig. 3); Archaea, 15 (from line (Ai); Bacteria, 52 (from line (Bi)). At positions 17, 17a, 20a, and 20b (the four optional positions of the D-loop), the base usage values do not add up to the total number of tDNAs.

TABLE 2. Base-pairing distribution in all tRNA genes.[a]

	A. Elongator tDNAs																				
	Eukarya 1984/302							Archaea 566/550							Bacteria 1546/1111						
Base-pairs	GC	CG	AT	TA	GT	TG	mm	GC	CG	AT	TA	GT	TG	mm	GC	CG	AT	TA	GT	TG	mm
1-72	1651	98	8	193	34	—	—	523	13	29	1	—	—	—	1253	84	47	97	61	—	4
2-71	545	913	112	188	195	25	6	254	304	1	5	1	1	—	802	589	35	76	42	2	—
3-70	365	874	169	400	134	33	9	265	203	29	25	44	—	—	650	409	235	126	122	4	—
4-69	472	650	173	346	75	222	46	244	258	35	19	2	8	—	770	400	98	194	34	47	3
5-68	344	715	332	314	159	86	34	149	322	29	41	3	21	1	482	544	313	98	45	51	13
6-67	330	650	206	405	122	204	67	285	234	10	31	1	4	1	375	432	284	379	19	51	6
7-66	843	51	611	444	34	—	1	491	—	47	16	12	—	—	761	33	385	326	40	1	—
3D 8-14	—	—	—	1984	—	—	—	—	—	—	536	—	1	29	—	—	—	1540	—	—	6
10-25	1393	45	13	48	480	—	5	377	33	7	—	147	—	2	1317	66	17	23	120	—	3
11-24	—	1240	—	725	—	17	2	49	330	12	166	—	7	2	8	1235	3	281	—	18	1
12-23	482	678	102	693	1	2	26	112	224	13	215	—	2	—	271	257	48	964	—	5	1
13-22	—	779	1	99	45	338	722	2	177	2	28	1	129	227	—	982	6	25	6	72	455
3D 15-48	1569	—	285	8	11	1	110	550	—	10	—	4	—	2	1176	—	319	1	12	5	33
3D 18-55	—	—	—	—	1984	—	—	20	—	—	—	546	—	—	—	—	—	—	1546	—	—
3D 19-56	1974	—	—	—	8	—	2	566	—	—	—	—	—	—	1531	—	11	—	1	—	3
3D 26-44	72	23	71	165	345	9	1299	55	19	28	14	229	5	216	1	2	131	3	204	4	1201
27-43	51	860	118	688	133	115	19	67	340	40	54	38	17	10	134	807	204	193	117	35	56
28-42	188	672	355	691	15	58	5	135	277	49	85	14	5	1	105	820	251	329	15	14	12
29-41	421	523	462	571	1	2	4	374	72	66	54	—	—	—	503	253	347	436	2	2	3
30-40	1600	226	9	6	81	60	2	469	90	—	—	5	—	2	1120	364	14	2	29	14	3
31-39	349	575	779	179	2	56	44	163	254	140	6	1	1	1	184	677	539	115	6	10	15
3D 32-38	—	—	—	62	—	—	1922	—	2	1	64	—	23	476	—	9	30	290	—	24	1193
3D 45-47k	1	88	2	—	9	23	188	2	4	2	—	18	5	80	45	3	41	4	36	9	189
49-65	793	820	202	52	91	22	4	172	289	38	26	26	15	—	658	124	341	22	393	2	6
50-64	324	1081	87	379	12	77	24	97	337	26	86	11	6	3	345	542	202	227	156	43	31
51-63	1217	212	309	79	4	154	9	483	61	11	—	1	10	—	702	283	298	144	63	30	26
52-62	1546	18	338	61	17	—	4	516	7	34	4	3	2	—	1294	12	220	10	—	—	10
53-61	1973	—	9	—	2	—	—	560	—	6	—	—	—	—	1542	—	4	—	—	—	—
3D 54-58	—	—	—	1935	—	—	49	—	—	—	562	—	—	4	—	—	—	1546	—	—	—

	B. Initiator tDNAs-iMet (CAT)																				
	Eukarya 41/7							Archaea 15/13							Bacteria 52/30						
Base-pairs	GC	CG	AT	TA	GT	TG	mm	GC	CG	AT	TA	GT	TG	mm	GC	CG	AT	TA	GT	TG	mm
1-72	—	—	37	4	—	—	—	—	—	15	—	—	—	—	—	—	—	—	—	—	52
2-71	32	—	—	9	—	—	—	15	—	—	—	—	—	—	52	—	—	—	—	—	—
3-70	—	41	—	—	—	—	—	—	15	—	—	—	—	—	—	52	—	—	—	—	—
4-69	9	—	32	—	—	—	—	15	—	—	—	—	—	—	49	—	3	—	—	—	—
5-68	30	9	1	1	—	—	—	15	—	—	—	—	—	—	51	—	1	—	—	—	—
6-67	5	13	23	—	—	—	—	14	1	—	—	—	—	—	50	—	2	—	—	—	—
7-66	41	—	—	—	—	—	—	14	—	1	—	—	—	—	40	—	12	—	—	—	—
3D 8-14	—	—	—	41	—	—	—	—	—	—	14	—	—	1	—	—	—	52	—	—	—
10-25	10	—	—	—	31	—	—	15	—	—	—	—	—	—	52	—	—	—	—	—	—
11-24	4	36	—	—	—	1	—	15	—	—	—	—	—	—	—	—	52	—	—	—	—
12-23	37	—	4	—	—	—	—	12	2	1	—	—	—	—	52	—	—	—	—	—	—
13-22	4	37	—	—	—	—	—	—	3	—	—	—	2	10	—	51	—	—	—	—	1
3D 15-48	40	—	1	—	—	—	—	15	—	—	—	—	—	—	52	—	—	—	—	—	—
3D 18-55	—	—	—	—	41	—	—	—	—	—	—	15	—	—	—	—	—	—	52	—	—
3D 19-56	41	—	—	—	—	—	—	15	—	—	—	—	—	—	52	—	—	—	—	—	—
3D 26-44	—	—	—	—	—	—	41	—	—	—	—	—	—	15	—	—	—	—	—	—	52
27-43	9	27	4	—	—	—	1	1	9	5	—	—	—	—	—	2	—	47	—	3	—
28-42	—	4	5	32	—	—	—	5	9	—	—	—	1	—	—	48	—	4	—	—	—
29-41	41	—	—	—	—	—	—	15	—	—	—	—	—	—	45	—	7	—	—	—	—
30-40	40	—	—	—	1	—	—	15	—	—	—	—	—	—	52	—	—	—	—	—	—
31-39	41	—	—	—	—	—	—	15	—	—	—	—	—	—	49	—	—	—	3	—	—
3D 32-38	—	—	—	—	—	—	41	—	—	—	—	—	—	15	—	—	—	—	—	—	52
3D 45-47k	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
49-65	8	24	1	—	8	—	—	7	7	1	—	—	—	—	45	—	4	—	3	—	—
50-64	8	14	7	6	6	—	—	5	5	5	—	—	—	—	4	12	1	3	7	25	—
51-63	1	5	20	15	—	—	—	1	1	2	11	—	—	—	6	13	22	9	—	2	—
52-62	39	—	2	—	—	—	—	13	—	2	—	—	—	—	51	—	1	—	—	—	—
53-61	41	—	—	—	—	—	—	15	—	—	—	—	—	—	52	—	—	—	—	—	—
3D 54-58	—	—	—	1	—	—	40	—	—	—	15	—	—	—	—	—	—	52	—	—	—

[a] This table shows the base-pairing distribution (number of each base-pair type in each position) found in the 4,204 tDNAs the consensus sequences of which are presented in Figures 3 and 4. The top lines in the two arrays have the same meaning as in Table 1. — is used instead of 0 for clarity. **A:** Elongator tDNAs: 4096 tDNAs: Eukarya, 1,984; Archaea, 566; Bacteria, 1,546. **B:** iMet initiator tDNAs-(CAT): 108 tDNAs: Eukarya, 41; Archaea, 15; Bacteria, 52. Base-pairing 45–47k is computed only if the length of the variable arm (bases 44 to 48) is greater than 8 bases; base 47k refers to the base preceding base 48 (Dirheimer et al., 1995). Mismatched base pair (mm) are defined as any base pair different from Watson-Crick or G-T base pairs.

of *M. kandleri* (A01) can be unambiguously recognized by the characteristic A1-T72 base pair of the acceptor stem (see Fig. 4B). In all other cases, the anticodon stem RGG rule allows us to distinguish the initiator tDNA-iMet (CAT) from the elongator tDNA-Met (CAT) (RajBhandary & Chow, 1995) as well as from the tDNA-Ile (CAT reading ATA) (see below, A very peculiar tDNA). Moreover, inside the anticodon loop, a subdomain-specific conservation exists: C33 is found in all higher eukaryotes whereas lower eukaryotes use the usual T33 (boxed in Fig. 4). Another domain-specific feature in all eukaryotic initiators is the systematic nonoccupancy of all optional positions of the D-loop (17, 17a, 20a, and 20b, shown with asterisks in Fig. 4) whereas in elongators, only position 17a is always unoccupied. At position 20, A is strictly conserved in all eukaryotic initiators (boxed in Fig. 4) whereas in bacterial and archaeal initiators, A20 is counterselected for Y20 and H20, respectively.

Such a high sequence conservation in tDNA-iMet (CAT) in the three phylogenetic domains is remarkable and is not observed for the two other single isoacceptor tDNAs, that is, elongator tDNA-Met (CAT) and tDNA-Trp (CCA) (data not shown). This clearly points to special functions of this particular tRNA-iMet (CAT) in the process of protein synthesis (reviewed in RajBhandary & Chow, 1995). Moreover, because distinct characteristics exist for tDNA-iMet of each of the three biological domains, the protein synthesis machinery in Eukarya, Bacteria, and Archaea obviously depends on different structural features of their initiator tRNAs, and possibly also of their elongator tRNAs. Despite this unique sequence conservation, polymorphism in tDNA-iMet (CAT) has been demonstrated. In *E. coli* (B27), for example, two tRNA-iMet differ by a single base (G46, three copies or A46, single copy). Gene disruption studies showed that either one or the other is required for cell viability (Kenri et al., 1992). We also observed in *A. thaliana* (E07), that bp 49–65 is G-C in nine copies and A-T in the other one; bp 52–62 is G-C in eight copies and A-T in the other two (data not shown).

A very peculiar tDNA: The archaeal/bacterial tDNA-Ile (CAT) specific for Ile codon ATA

In addition to initiator tDNA-iMet (CAT) and elongator tDNA-Met (CAT), a third type of tDNA bearing a Met CAT anticodon is always present in sequenced archaeal and bacterial genomes (A01 to A13 and B01 to B30, respectively). As far as *E. coli* and *Bacillus subtilis* are concerned, sequencing the corresponding mature tRNA revealed that nucleotide C34 in the wobble position of the anticodon was modified to lysidine (2-lysylcytidine, abbreviated in k^2C; Muramatsu et al., 1988b; Matsugi et al., 1996). This tRNA (k^2CAU) was shown to be specifically charged in vitro by Ile (Muramatsu et al., 1988a). When k^2C was replaced by unmodified C, the resulting tRNA (CAU) became aminoacylated by Met instead of Ile. Together with the fact that none of the bacterial and archaeal genomes sequenced so far harbor a tDNA-Ile (UAU), as is the case for sequenced eukaryotic genomes (see below, Common and domain-specific strategies exist for decoding the genetic information with a minimum number of isoacceptor tRNAs; Fig. 6A), these observations suggest that in both Bacteria and Archaea, the Ile codon AUA is, without exception, translated by a "methionyl" tRNA that is posttranscriptionally modified at its anticodon wobble position. Using the criteria defined in the section above, the initiator tDNA-iMet (CAT) can be easily distinguished from the two other tDNAs that also bear the (CAT) anticodon. However, for each of the 43 archaeal and bacterial genomes examined, the challenge is to decide which one of these two tDNAs is the tDNA-Ile (CAT reading ATA). Fortunately, after a pairwise disjunction analysis of bacterial tDNA-Met (CAT) sets, it was recently pointed out that the nucleotide at position 44 (first position of the V-arm) in tDNA (CAT) may play a major structural discriminatory role for a hypothetical lysinylation enzyme (W. Mitchell, pers. commun.). Testing this hypothesis, we found that, in Archaea, the signature Y44 (C or T) together with the sequence GGG in positions 1 to 3 (start of amino acid stem) clearly identifies the tDNA-Ile (CAT reading ATA), whereas the elongator tDNA-Met-(CAT) harbors A44 (or T44 in *S. solfataricus* (A07) and *Sulfolobus tokodai* (A08)) and the sequence GCC in positions 1 to 3. In fact, in 10 of the 13 archaeons examined, a larger consensus, GGGCCC, is present in positions 1 to 6 of the tDNA-Ile (CAT reading ATA). In 26 of the 30 bacteria examined, the signature Y44 (C or T) also identifies the tDNA-Ile (CAT reading ATA), whereas the elongator tDNA-Met (CAT) harbors R44 (A or G), this base being possibly engaged in 3D base pairing with base 26. In the four remaining bacterial genomes, various situations are encountered and it could be that C34 is modified by another type of enzyme that requires different identity elements. Bases found at position 44 for the two tDNA-Ile (CAT reading ATA) and tDNA-Met (CAT) are: C and G in *Treponema pallidum* (B01) and *Anabaena* 7120 (B05) and G and A in *Chlamydia trachomatis* (B03) and *Clostridium perfringens* (B19). Further experimental work is needed to identify the nature of the modification occurring in most Bacteria, and also in Archaea. The situation in Eukarya is simpler, because a tDNA-Ile (TAT) corresponding to a mature tRNA decoding the AUA Ile codon is always distinct of the elongator tDNA-Met (CAT). Remarkably this tDNA-Ile (TAT) always harbors an intron (except in the case of *A. thaliana*—see below). At least in *S. cerevisiae*, this intron is strictly required for uridines at positions 34 and 37 of the anticodon to be enzymatically modified to pseudouridines (Szweykowska-Kulinska et al., 1994; Motorin et al., 1998).

A guided tour along the cloverleaf structure of elongator tRNAs

Collecting all the tDNA sequences from a given domain (or even from a given organism) into a single consensus (as done in Figs. 3 and 4) leads to a large loss of information because the frequencies of A, C, G, or T at each base position are not indicated, nor is the distribution of the various base-pairing types. Therefore, in Table 1, we also present the base distribution at all positions of the tDNA, arranged by biological domains. Base-pairing distribution in the four stems and also at the eight 3D base pairs (shown as black lines in Fig. 2B) are presented in Table 2, arranged by domain. In these two tables, Panel A deals with the elongator tDNAs and Panel B with the initiator tDNA-iMet (CAT). Below, we have summarized the most outstanding features and exceptions in base and base-pair distribution, together with additional structural data. As far as possible, these peculiarities were related to functional properties of the tRNA molecule. Unless otherwise specified, the tDNAs referred to are each present as a single copy. Three-dimensional base pairs are designated as "3D base pairs."

Base (−1)

All cytoplasmic mature tRNA-His (GTG) from the three biological domains have an additional residue, generally guanosine, at the 5′ end of the acceptor arm. This extra residue, in fact the 5′ phosphate group, is an important identity element for histidyl-tRNA synthetase (reviewed in Giegé et al., 1998; see also Fromant et al., 2000). In Bacteria and Archaea, this G always pairs with C in position 73 thus extending the acceptor stem of 7 bp by 1 extra bp (see Fig. 2A and below, 5′-G addition to tRNA-His in Eukarya versus its systematic encoding in Archaea and Bacteria). In Eukarya, mismatches predominate over Watson–Crick base pairs, whereas G-T/T-G pairings at (−1)–73 and G(−1)–C73 are totally avoided (not shown).

Base pair 1–72

Dominance of Watson–Crick bp 1–72 in all elongator tRNAs is evident (Table 2, Panel A). However, Figure 5A shows that preference for base pairings other than G-C depends on the type of tRNA isoacceptor and on the origin of the tRNA. Beside the prevalence of A-T in all tRNA-iMet from Eukarya and Archaea, or the systematic presence of Y!H (C or T mismatched with any base but G) in all bacterial tDNA-iMet (Fig. 5A, line (b)) as discussed above, only a few selected families of tRNA isoacceptors harbor a characteristic base pair in 1–72: C1-G72 (in red) appears only in eukaryal and archaeal tDNAs-Tyr (GTA) (Fig. 5A, line (d)) and bacterial tDNAs-Pro (NGG family box) (Fig. 5A, line (c)), whereas base pair T1-A72 (in blue) is found in most eukaryal tDNA-Asp (GTC) and all eukaryal tDNAs-Glu (YTC family box; Fig. 5A, line (g)). The most interesting case is that of tRNA-Gln (YTG family box), which contains G1-C72 (in green) in all Eukarya but one (*H. sapiens* (E05)), A-T (in black) in all Archaea, and T-A (or A-T, in blue) in all Bacteria (Fig. 5A, line (e)). In most cases, these sequence constraints, together with the identity of the discriminator base 73 (see below) have been shown to play the role of determinant or antideterminant for the recognition of the tRNA molecules by several enzymes and protein factors, namely the majority of aminoacyl-tRNA synthetases, initiation and elongation factors, Met-tRNA transformylase, and peptidyl-tRNA hydrolase (reviewed in RajBhandary & Chow, 1995; Rudinger et al., 1996; Giegé et al., 1998).

Base pair 3–70

The very first observation that the G3-U70 wobble pair was unique to tRNA-Ala (NGC family box) was made after inspection of a set of 13 sequenced tRNA-Ala from Eukarya and Bacteria (Grosjean et al., 1982). Strong selective pressure to use nonisosteric G3-U70 to mark an RNA acceptor for alanyl-tRNA synthetase has now been demonstrated in all three biological domains (reviewed in Chihade et al., 1998; see also Mueller et al., 1999), with only a few exceptions: G3-T70 is also found in tDNA-Arg (TCT) of *Pseudomonas aeruginosa* (B13), *M. genitalium* (B22), *M. pneumoniae* (B23), and *Xilella fastidiosa* (B25); tDNA-Ser (GCT) of *Listeria monocytogenes* (B07, 2 copies), *M. genitalium* (B22), and *M. pneumoniae* (B23); tDNA-Val (GAC) of *Helicobacter pylori* (B20). Noteworthy, the same G3-U70 was shown to be an antideterminant for the *E. coli* threonyl-tRNA synthetase (Nameki, 1995). Likewise, C3-G70 is found in all initiator tRNA-iMet from all domains (see Figs. 3 and 4). This, however, is not a unique characteristic, as many elongator tRNAs also have C3-G70 (see Table 2). Notice that base-pair context of C3-G70 such as bp 2–71 also plays a role for efficient aminoacylation of tRNA-Ala (Beuning et al., 2002).

Base pair 7–66

Despite the high G or C content in the amino acid stem of all archaeal tRNAs, the lack of a C7-G66 pair (but not G7-C66; see Table 2) at the bottom of the amino acid stem is remarkable. Similarly, in both Eukarya and Bacteria, C7-G66 is clearly avoided over the other types of Watson–Crick base pairs. It is also the position where G-T/T-G and mismatched bp 7–66 are absent in Archaea and Bacteria and very rare in Eukarya.

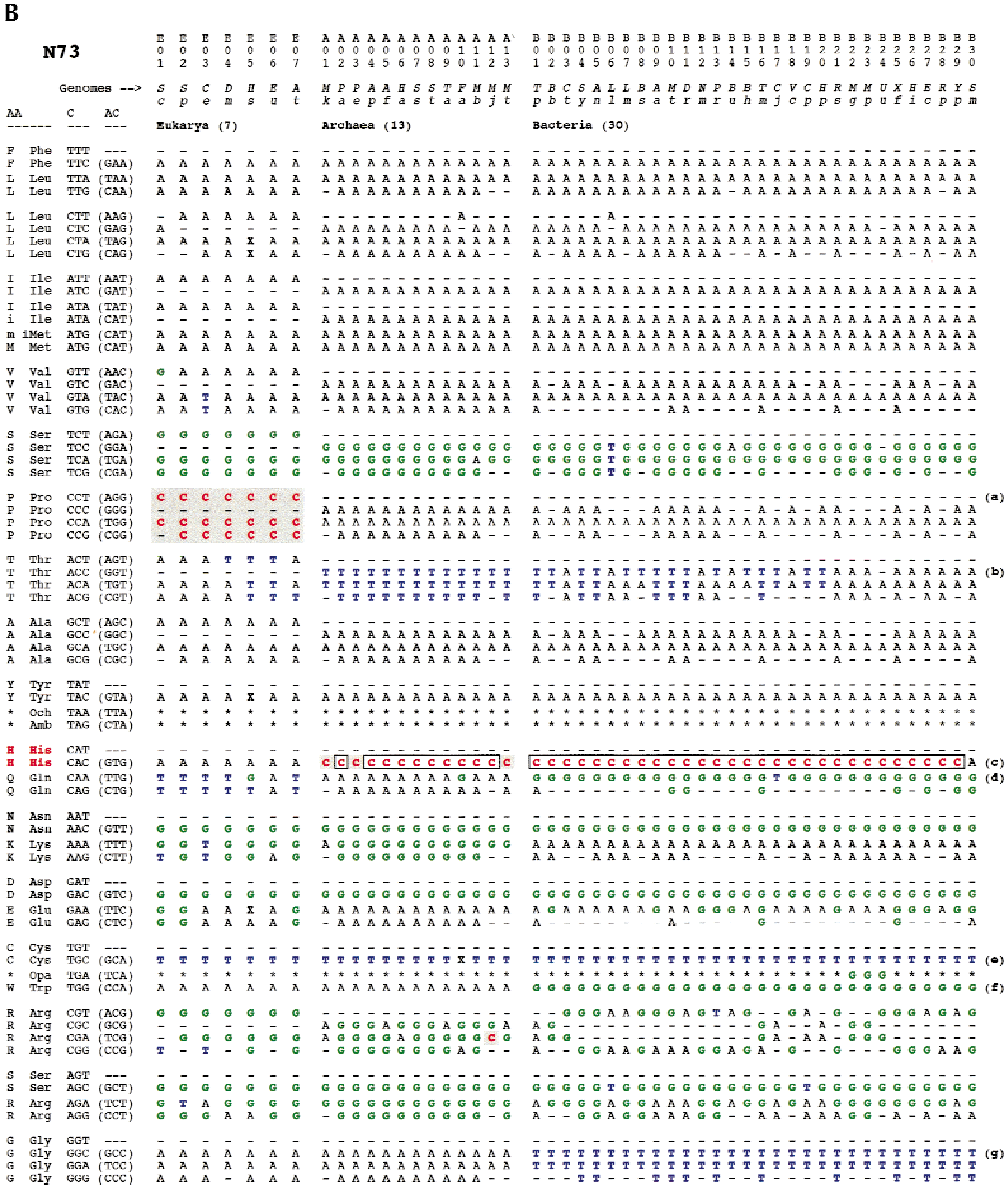
3D base pair 8–14

One of the most conserved tertiary interactions in tRNA involves a *trans*-Hoogsteen U8-A14 tertiary base pair

A

N1				E01	E02	E03	E04	E05	E06	E07	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	B01	B02	B03	B04	B05	B06	B07	B08	B09	B10	B11	B12	B13	B14	B15	B16	B17	B18	B19	B20	B21	B22	B23	B24	B25	B26	B27	B28	B29	B30	
	Genomes -->			Sc	Sp	Ce	Dm	Hs	Eu	At	Mk	Pa	Pe	Ap	Af	Ha	Ss	St	Ta	Fa	Mb	Mj	Mt	Tp	Bb	Ct	Sy	An	Ll	Lm	Bs	Aa	Mt	Dr	Nm	Pr	Bu	Bh	Tm	Cj	Vc	Cp	Hp	Rs	Mg	Mp	Uu	Xf	Hi	Ec	Rp	Yp	Sm	
AA		C	AC	Eukarya (7)							Archaea (13)													Bacteria (30)																														
F	Phe	TTT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
F	Phe	TTC	(GAA)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
L	Leu	TTA	(TAA)	G	G	A	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
L	Leu	TTG	(CAA)	G	G	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	-	G	G	G	G	G	G	G	C	C	G	G	G	G	-	G	G	
L	Leu	CTT	(AAG)	-	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
L	Leu	CTC	(GAG)	G	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	-	G	G	G	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	-	G	G	G	G	G	
L	Leu	CTA	(TAG)	G	G	G	G	X	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	(a)
L	Leu	CTG	(CAG)	-	-	G	G	X	G	G	-	G	G	G	G	G	G	G	G	G	G	-	-	G	-	G	G	G	G	-	G	G	G	G	G	G	-	-	G	-	G	-	-	G	-	-	-	G	-	G	-	G	G	
I	Ile	ATT	(AAT)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
I	Ile	ATC	(GAT)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	
I	Ile	ATA	(TAT)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
i	Ile	ATA	(CAT)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	
m	iMet	ATG	(CAT)	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	T	C	C	C	C	C	C	C	C	T	C	C	T	C	C	(b)
M	Met	ATG	(CAT)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
V	Val	GTT	(AAC)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
V	Val	GTC	(GAC)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	-	G	G	G	-	G	G	A	G	A	A	A	A	G	G	G	G	-	G	A	-	-	-	G	G	G	-	G	G	
V	Val	GTA	(TAC)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
V	Val	GTG	(CAC)	G	A	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	G	G	-	-	-	-	G	-	-	-	-	G	-	-	-	G	-	-	-	-	-	
S	Ser	TCT	(AGA)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S	Ser	TCC	(GGA)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	A	G	G	G	G	G	G	-	G	G	G	G	G	G	
S	Ser	TCA	(TGA)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	G	G	C	G	G	G	G	G	G	G	G	G	G	
S	Ser	TCG	(CGA)	G	G	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	-	G	-	G	G	G	G	G	-	G	G	G	G	G	-	-	G	-	-	-	-	G	G	G	-	G	-	G	-	-	G	
P	Pro	CCT	(AGG)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
P	Pro	CCC	(GGG)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	C	-	C	C	C	-	-	-	C	C	C	C	C	-	-	C	-	C	-	C	C	-	-	-	C	-	C	-	C	C	(c)
P	Pro	CCA	(TGG)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	
P	Pro	CCG	(CGG)	-	G	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	-	C	-	-	C	C	-	-	-	C	C	C	C	C	-	-	C	-	-	-	-	C	-	-	-	C	-	C	-	-	C	
T	Thr	ACT	(AGT)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
T	Thr	ACC	(GGT)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	-	G	G	G	G	G	G	
T	Thr	ACA	(TGT)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
T	Thr	ACG	(CGT)	G	G	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	G	G	-	G	G	G	G	G	-	G	G	G	G	G	-	-	G	-	-	-	-	G	G	G	-	G	-	G	G	-	G	
A	Ala	GCT	(AGC)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
A	Ala	GCC	(GGC)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	-	G	G	G	-	-	G	G	G	G	G	G	G	G	G	G	G	-	G	G	-	-	-	G	G	G	G	G	G	
A	Ala	GCA	(TGC)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
A	Ala	GCG	(CGC)	-	G	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	-	G	-	-	G	G	-	-	-	-	G	G	-	-	-	-	G	-	-	-	-	G	-	-	-	G	-	-	-	-	G	
Y	Tyr	TAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Y	Tyr	TAC	(GTA)	C	C	C	C	X	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	(d)
*	Och	TAA	(TTA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
*	Amb	TAG	(CTA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
H	His	CAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
H	His	CAC	(GTG)	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
Q	Gln	CAA	(TTG)	G	G	G	G	T	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	T	T	T	A	T	A	T	T	T	T	A	T	T	A	T	T	T	A	T	T	T	T	T	(e)
Q	Gln	CAG	(CTG)	G	G	G	G	G	G	G	-	A	A	A	A	A	A	A	A	A	A	-	A	T	-	-	-	-	-	-	-	-	T	T	-	-	-	-	T	-	-	-	-	-	-	-	-	T	-	T	-	T	T	
N	Asn	AAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
N	Asn	AAC	(GTT)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	G	G	G	G	T	T	T	T	T	(f)
K	Lys	AAA	(TTT)	T	T	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
K	Lys	AAG	(CTT)	G	T	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	-	G	G	-	-	G	G	G	-	G	G	G	-	-	-	-	G	-	-	G	-	G	G	G	G	G	G	-	-	G	G	
D	Asp	GAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
D	Asp	GAC	(GTC)	T	T	T	T	T	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	(g)
E	Glu	GAA	(TTC)	T	T	T	T	X	T	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
E	Glu	GAG	(CTC)	T	T	T	T	T	T	T	-	G	G	G	G	G	G	G	G	G	G	-	-	G	-	-	-	-	-	-	-	-	G	-	-	-	-	-	G	-	-	-	-	-	-	-	-	G	-	-	-	-	G	
C	Cys	TGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
C	Cys	TGC	(GCA)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	X	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
*	Opa	TGA	(TCA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	A	A	A	*	*	*	*	*	*	
W	Trp	TGG	(CCA)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	A	A	A	G	A	G	A	A	A	A	A	A	A	A	A	A	G	G	A	G	A	A	A	A	A	
R	Arg	CGT	(ACG)	T	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	G	G	G	C	G	G	G	G	G	G	-	-	G	G	-	G	-	-	G	G	G	G	G	G	G	(h)
R	Arg	CGC	(GCG)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	-	-	G	-	G	G	-	-	-	-	-	-	-	
R	Arg	CGA	(TCG)	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	G	G	-	G	G	-	G	G	G	-	-	-	-	-	-	
R	Arg	CGG	(CCG)	G	-	G	-	G	-	G	-	G	G	G	G	G	G	G	G	G	A	-	-	G	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	-	G	-	-	G	-	-	-	G	G	G	G	G	G	
S	Ser	AGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S	Ser	AGC	(GCT)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
R	Arg	AGA	(TCT)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	C	G	G	G	G	G	G	G	G	G	
R	Arg	AGG	(CCT)	G	T	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	G	G	-	-	G	G	G	G	G	G	G	G	C	G	-	-	G	G	-	G	G	C	C	C	-	G	-	G	-	G	G	
G	Gly	GGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
G	Gly	GGC	(GCC)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
G	Gly	GGA	(TCC)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
G	Gly	GGG	(CCC)	G	G	G	-	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	-	-	-	-	G	G	-	-	-	G	G	G	-	G	-	-	G	-	-	-	-	G	-	-	-	G	-	G	-	G	G	

FIGURE 5. The base at position 1 and the discriminator base (position 73) in all tDNA from all genomes. **A:** Base at position 1: Genomes are listed from left to right in the order Eukarya, Archaea, and Bacteria (i.e., using the same order and genome numbers as in Fig. 3). Abbreviated genome names are to be read vertically. The list obeys the traditional "genetic code" order from codon TTT down to GGG. On each line are listed: the amino acid ("AA", using 3- and 1-letter codes) charged by the tRNA; the codon ("C", as read in the genomic sequence of protein genes); the corresponding anticodon ("AC", as found in the tRNA gene). Anticodons never found in tDNAs of any domains are written as −−− (e.g., (AAA) is never used as a Phe anticodon). The elements of the array indicate which base (A, C, G, or T) is encoded at position 1 in the tRNA gene. - indicates that no tDNA exists for a given anticodon and genome; ∗ indicates stop codons. Five Xs are used to indicate that a tDNA probably exists but has not been discovered or is not present in the sequences examined. Gray background indicates that the base at position 1 is mismatched with that at position 72 (G-T and T-G base pairings are not considered as mismatches). All bacterial initiator tDNA-iMet harbor a mismatched base pair at positions 1–72 (line (b)). Two single cases of 1–72 mismatches are also found (gray background, lines (a) and (h)). Other notes are referred to in text. (*Figure continues on following page.*)



B

N73				E01	E02	E03	E04	E05	E06	E07	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	B01	B02	B03	B04	B05	B06	B07	B08	B09	B10	B11	B12	B13	B14	B15	B16	B17	B18	B19	B20	B21	B22	B23	B24	B25	B26	B27	B28	B29	B30	
	Genomes -->			*Sc*	*Sp*	*Ce*	*Dm*	*Hs*	*Eu*	*At*	*Mk*	*Pa*	*Pe*	*Ap*	*Af*	*Ha*	*Ss*	*St*	*Ta*	*Fa*	*Mb*	*Mj*	*Mt*	*Tp*	*Bb*	*Ct*	*Sy*	*An*	*Ll*	*Lm*	*Bs*	*Aa*	*Mt*	*Dr*	*Nm*	*Pr*	*Bu*	*Bh*	*Tm*	*Cj*	*Vc*	*Cp*	*Hp*	*Rs*	*Mg*	*Mp*	*Uu*	*Xf*	*Hi*	*Ec*	*Rp*	*Yp*	*Sm*	
AA		C	AC																																																			
------		---	---	**Eukarya (7)**							**Archaea (13)**													**Bacteria (30)**																														
F	Phe	TTT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
F	Phe	TTC	(GAA)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
L	Leu	TTA	(TAA)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
L	Leu	TTG	(CAA)	A	A	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	-	-	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	
L	Leu	CTT	(AAG)	-	A	A	A	A	A	A	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
L	Leu	CTC	(GAG)	A	-	-	-	-	-	-	A	A	A	A	A	A	A	A	A	A	-	A	A	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	A	A	A	A	
L	Leu	CTA	(TAG)	A	A	A	A	X	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
L	Leu	CTG	(CAG)	-	-	A	A	X	A	A	-	A	A	A	A	A	A	A	A	A	A	-	-	A	-	A	A	A	A	-	A	A	A	A	A	A	-	-	A	-	A	-	-	A	-	-	-	A	-	A	-	A	A	
I	Ile	ATT	(AAT)	A	A	A	A	A	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
I	Ile	ATC	(GAT)	-	-	-	-	-	-	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
I	Ile	ATA	(TAT)	A	A	A	A	A	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
i	Ile	ATA	(CAT)	-	-	-	-	-	-	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
m	iMet	ATG	(CAT)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
M	Met	ATG	(CAT)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
V	Val	GTT	(AAC)	G	A	A	A	A	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
V	Val	GTC	(GAC)	-	-	-	-	-	-	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	-	-	-	A	A	A	-	A	A	
V	Val	GTA	(TAC)	A	A	T	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
V	Val	GTG	(CAC)	A	A	T	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	A	A	A	-	-	-	-	-	-	-	-	A	A	-	-	-	-	A	-	-	-	-	A	-	-	-	A	-	-	-	-	-	
S	Ser	TCT	(AGA)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S	Ser	TCC	(GGA)	-	-	-	-	-	-	-	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	G	G	-	G	G	G	G	G	G	
S	Ser	TCA	(TGA)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
S	Ser	TCG	(CGA)	G	G	G	G	G	G	G	-	G	G	G	G	G	G	G	G	G	G	-	-	G	-	G	G	G	T	G	-	G	G	G	G	G	-	-	G	-	-	-	-	G	G	G	-	G	-	G	-	-	G	
P	Pro	CCT	(AGG)	C	C	C	C	C	C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(a)
P	Pro	CCC	(GGG)	-	-	-	-	-	-	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	A	-	-	-	A	A	A	A	A	-	-	A	-	A	-	A	A	-	-	-	A	-	A	-	A	A	
P	Pro	CCA	(TGG)	C	C	C	C	C	C	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
P	Pro	CCG	(CGG)	-	C	C	C	C	C	C	-	A	A	A	A	A	A	A	A	A	A	-	-	A	-	-	A	A	-	-	-	A	A	A	A	A	-	-	A	-	-	-	-	A	-	-	-	A	-	A	-	-	A	
T	Thr	ACT	(AGT)	A	A	A	T	T	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
T	Thr	ACC	(GGT)	-	-	-	-	-	-	-	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	A	T	T	T	T	T	A	T	A	T	T	T	A	T	T	A	A	A	-	A	A	A	A	A	A	(b)
T	Thr	ACA	(TGT)	A	A	A	A	T	T	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	A	T	T	A	A	A	T	T	T	A	A	A	A	T	T	A	T	T	A	A	A	A	A	A	A	A	A	A	
T	Thr	ACG	(CGT)	A	A	A	A	T	T	T	-	T	T	T	T	T	T	T	T	T	T	-	T	T	-	A	T	T	A	A	-	T	T	T	A	A	-	-	T	-	-	-	-	A	A	A	-	A	-	A	A	-	A	
A	Ala	GCT	(AGC)	A	A	A	A	A	A	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
A	Ala	GCC	(GGC)	-	-	-	-	-	-	-	A	A	A	A	A	A	A	A	A	A	A	A	A	A	-	A	A	A	-	-	A	A	A	A	A	A	A	A	A	A	A	-	A	A	-	-	-	A	A	A	A	A	A	
A	Ala	GCA	(TGC)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
A	Ala	GCG	(CGC)	-	A	A	A	A	A	A	-	A	A	A	A	A	A	A	A	A	A	-	-	A	-	-	A	A	-	-	-	-	A	A	-	-	-	-	A	-	-	-	-	A	-	-	-	A	-	-	-	-	A	
Y	Tyr	TAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Y	Tyr	TAC	(GTA)	A	A	A	A	X	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
*	Och	TAA	(TTA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
*	Amb	TAG	(CTA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
H	His	CAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
H	His	CAC	(GTG)	A	A	A	A	A	A	A	C	$\boxed{C}$	C	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	C	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	$\boxed{C}$	A	(c)
Q	Gln	CAA	(TTG)	T	T	T	T	G	A	T	A	A	A	A	A	A	A	A	A	G	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	G	(d)
Q	Gln	CAG	(CTG)	T	T	T	T	T	A	T	-	A	A	A	A	A	A	A	A	A	A	-	A	A	-	-	-	-	-	-	-	-	G	G	-	-	-	-	G	-	-	-	-	-	-	-	-	G	-	G	-	G	G	
N	Asn	AAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
N	Asn	AAC	(GTT)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
K	Lys	AAA	(TTT)	G	G	T	G	G	G	G	A	G	G	G	G	G	G	G	G	G	G	G	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	
K	Lys	AAG	(CTT)	T	G	T	G	G	A	G	-	G	G	G	G	G	G	G	G	G	G	-	-	A	A	-	-	A	A	A	-	A	A	A	-	-	-	-	A	-	-	A	-	A	A	A	A	A	A	-	-	A	A	
D	Asp	GAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
D	Asp	GAC	(GTC)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	
E	Glu	GAA	(TTC)	G	G	A	A	X	A	G	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	A	A	A	A	A	A	G	A	A	G	G	G	A	G	A	A	A	A	G	A	A	A	G	G	G	A	G	G	
E	Glu	GAG	(CTC)	G	G	A	A	A	A	G	-	A	A	A	A	A	A	A	A	A	A	-	-	A	-	-	-	-	-	-	-	-	A	-	-	-	-	-	G	-	-	-	-	-	-	-	-	G	-	-	-	-	A	
C	Cys	TGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
C	Cys	TGC	(GCA)	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	X	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	(e)
*	Opa	TGA	(TCA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	G	G	G	*	*	*	*	*	*	
W	Trp	TGG	(CCA)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	(f)
R	Arg	CGT	(ACG)	G	G	G	G	G	G	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	G	A	A	G	G	G	A	G	T	A	G	-	-	G	A	-	G	-	-	G	G	G	A	G	A	G	
R	Arg	CGC	(GCG)	-	-	-	-	-	-	-	A	G	G	G	A	G	G	G	A	G	G	G	A	A	G	-	-	-	-	-	-	-	-	-	-	-	-	-	G	A	-	-	A	-	G	G	-	-	-	-	-	-	-	
R	Arg	CGA	(TCG)	-	G	G	G	G	G	G	A	G	G	G	G	A	G	G	G	G	G	C	G	A	G	G	-	-	-	-	-	-	-	-	-	-	-	-	G	A	-	A	A	-	G	G	G	-	-	-	-	-	-	
R	Arg	CGG	(CCG)	T	-	T	-	G	-	G	-	G	G	G	G	G	G	G	G	A	G	-	-	A	-	-	G	G	A	A	G	A	A	A	G	G	A	G	A	-	G	-	-	G	-	-	-	G	G	G	A	A	G	
S	Ser	AGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S	Ser	AGC	(GCT)	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	G	T	G	G	G	G	G	G	G	G	G	G	G	
R	Arg	AGA	(TCT)	G	T	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	A	G	G	G	G	A	G	G	A	A	A	G	G	A	G	G	A	G	A	A	G	G	G	G	G	G	G	G	A	G	
R	Arg	AGG	(CCT)	G	G	G	A	A	G	G	-	G	G	G	G	G	G	G	G	G	G	-	G	A	-	-	G	G	A	G	G	A	A	A	G	G	-	-	A	A	-	A	A	A	G	G	-	A	-	A	-	A	A	
G	Gly	GGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
G	Gly	GGC	(GCC)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	(g)
G	Gly	GGA	(TCC)	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	
G	Gly	GGG	(CCC)	A	A	A	-	A	A	A	-	A	A	A	A	A	A	A	A	A	A	-	-	-	-	-	T	T	-	-	-	T	T	T	-	T	-	-	T	-	-	-	-	T	-	-	-	T	-	T	-	T	T	

FIGURE 5 (Continued). B: The discriminator base (position 73): The same presentation is used. The elements of the array indicate which base (A, C, G, or T) is encoded at position 73 in the tRNA gene. Boxed Cs are paired to a G at position −1, those on gray background are not. Features signaled with letters in parenthesis are commented on in the text.

that stacks between the terminal 13–22 bp in the D-stem and the semiconserved purine 15 and pyrimidine 18 (non-Watson–Crick base pairings in RNA are reviewed in Leontis & Westhof, 2001). Altogether, this network of interactions stabilizes the sharp turn from the acceptor stem to the D-stem, which is a crucial constraint for maintaining the overall L-shaped structure of the tRNA molecule (Major et al., 1993). Until recently, base T at position 8 was considered as one of the very few strictly invariant nucleotides, and was used as such in our tDNA search procedure (see Methods). However, during the scanning of the recently sequenced genome of the hyperthermophilic archaeaon *M. kandleri* (A01), C at position 8 had to be allowed in order to detect the full set of tDNAs: indeed, 30 tDNAs harbor C8-A14 and only 4 harbor T8-A14. As far as the base at position 14 is concerned, beside the conserved A14 in Bacteria and Archaea, C14 is also found in tDNA-Trp (CCA) of *H. pylori* (B20), *Borrelia burgdorferi* (B02), *Campylobacter jejuni* (B17), *Vibrio cholerae* (B18), and tDNA-Thr (GGT) of *T. maritima* (B16), T14 only in tDNA-Trp (CCA) of *T. pallidum* (B01), and G14 only in tDNA-Thr (TGT) of the archaeon *Halobacterium* sp. NRC-1 (A06). The exact identity of bp 8–14 may not have much incidence on the function of these tRNAs, at least in *E. coli*. Indeed, it was demonstrated by systematic mutagenesis analysis on *E. coli* amber suppressor tRNA-Ala (CUA) that various combinations of bases at positions 8 and 14 allow the mutated tRNA-Ala (CUA) to suppress the UGA stop codon in vivo, yet with different efficiencies (Sterner et al., 1995).

Base 9

In tRNAs of class I (harboring a short variable arm; see below A nearly consistent and universal rule for the long V-arms of tDNAs-Leu, tDNAs-Ser, and bacterial tDNAs-Tyr), the base at position 9, mostly a purine (noted R), interacts with base 23 (*trans*-Hoogsteen) in the deep groove of bp 12–23 of the D-helix, thus making a base-triple 12–23 … 9, or with base pair 13–22 in tRNAs of Class II (harboring a long variable arm; for details see Dirheimer et al., 1995; Gautheret et al., 1995).

Base pair 10–25

This base pair displays a strong bias for G-C or G-T in all three biological domains. The fact that many elongator tRNAs from Archaea (including the hyperthermophilic ones) select G-T at these positions strongly suggests that this base pair has an important functional role. At least in yeast tRNA-Asp (GUC), this base pair is a positive determinant among others for the aspartyl-tRNA synthetase (reviewed in Giegé et al., 1998), as well as a negative determinant for the posttranscriptional formation of N^2-N^2-dimethylguanosine-26 (reviewed in Edqvist et al., 1995). In the crystal structure of some tRNAs of class I (with a short variable arm), the bp 10–25 interacts in the deep groove of the D-helix with base 45 (mostly G) of the short V-arm, making a base-triple 45 … 10–25 (for details, see Gautheret et al., 1995).

Base pair 11–24

In almost all eukaryotic tDNAs (including the initiator tDNA-iMet; see Initiator tDNA-iMet is the most conserved tDNA across the three domains but nevertheless has domain-specific features; Figs. 3 and 4) this base pair is strictly Y-R (pyrimidine-purine) with a predominance of G-C over T-A. It is less constrained in Bacteria and Archaea, except for a clear avoidance of A-T in Bacteria (a hallmark of bacterial tRNA-iMet) and G-C in Archaea. It is noteworthy that this base pair (in fact purine-24) plays a role in the decoding property of a tRNA at the wobble base-34 of the anticodon loop, at least in *E. coli*. This supports a distal relationship between the structure of the D-stem and the anticodon loop (Smith & Yarus, 1989; Schultz & Yarus, 1994; reviewed in Yarus & Smith, 1995).

Base opposition 13–22

These positions at the end of the D-arm are characterized by the strict avoidance of a G-C base pair, in favor of C-G, mismatched, T-G, or T-A base pairs (see Table 2). In Class I tRNAs, these base oppositions interact in the deep groove of the helix with base 46 (or 45) of the short V-arm, thus making a base-triple 13–22 … 46 that contributes further to the stacking interactions in the D-stem (see Gautheret et al., 1995). The distribution of each type of base pair as well as mismatched base opposition is not random. It clearly depends on both the isoacceptor considered and on the biological domain (data not shown). For example, tDNAs of Class II, bearing a long V-arm, always harbor a mismatched base opposition of the type G:A > A:A > A:C or > C:C. Likewise, tDNAs of Class I bearing short 4-base V-arms most always harbor a wobble T13:G22 base pair (Cermakian et al., 1998). Moreover, base oppositions T:T or T:G are characteristic of eukaryal and archaeal tRNA-Val (NGC family box) and tRNA-Pro (NGG family box), whereas in the homolog tDNA-Val and tDNA-Pro of Bacteria, a Watson–Crick G-C pair is mostly present. The high frequency of such characteristic base pairs or base oppositions at the end of the D-arm in elongator and initiator tDNAs of the hyperthermophilic organisms points out to an important role of this region on the structural and/or functional properties of the tRNA molecule.

3D base pair 15–48

This base pair is usually referred to as the “Levitt base pair” (Levitt, 1969). G15 followed by A15 predominates (especially in Archaea), whereas a pyrimidine Y48 (C or T) is almost always found in the parallel strand forming a *trans*-Watson–Crick base pair that connects the D-loop with the V-loop. The R15-Y48 base pair stacks with the quasi universal *trans*-Hoogsteen U8-A14 pair (see above) and with purine at position 21 (see below). Altogether these tertiary pairs (see Fig. 2B) contribute to the overall three-dimensional organization of tRNAs of both class I and class II into an L-shaped structure. However, as demonstrated with mutated *E. coli* suppressor tRNA-Ala (CUA), positions 15 and 48 taken individually can accommodate all combinations of bases but one (A15:A48), without affecting the suppressor activity in vivo (Hou et al., 1995). Nevertheless, this may not be true in other tRNAs or if another 3D base pair in addition to R15-Y48 were also mutated.

D-loop (bases 14–21)

In this loop, four positions are optionally occupied: bases 17, 17a, 20a, and 20b (see Fig. 2). Position 17 is never occupied by G in Eukarya and Archaea (indicated as h or y in Figs. 3 and 4), whereas in Bacteria it is found only three times among 1,546 elongator tDNAs examined. These three exceptions are: tDNA-Thr (GTT) of *T. maritima* (B16), tDNA-Cys (GCA) of *M. genitalium* (B22), and of *M. pneumoniae* (B23). Remarkably, position 17a is never occupied by any nucleotide in eukaryotic tDNAs as well as in the tDNAs of the human pathogen bacteria *T. pallidum* (B01) and *B. burgdorferi* (B02) and also in *C. perfringens* (B19). Therefore, in eukaryotic tRNAs, the fourth base downstream from the invariant A14 is always a G (the first or the second of the invariant Gs at positions 18 and 19—see “A box” in Fig. 3). This property has to be related to a constraint exerted on the tDNA sequence by the Pol III transcription machinery (see Distribution of tRNA genes in genomes and peculiarities of their transcription in Eukarya). Notice that three copies of an unusual tDNA-Ser (GCT) with an extra “20c” base (T) are found in *S. pombe* (E02) (not shown). Position 20 is mostly T in all domains. Remarkably, in Eukarya, G20 is the exclusive signature of tDNA-Phe (GAA) (reviewed in Giegé et al., 1998), with the exception of A20 in the tDNA-Phe (GAA) of *E. cuniculi* (E06). In Archaea, G20 is also found in all tDNA-Phe (GAA). However G20 also exists in some tDNAs specific for Leu, Ser (NGA family box only), Thr, and Ala. In contrast, all bacterial tDNA-Phe (GAA) bear T20, whereas G20 is only found in bacterial tDNAs specific for Val and Ala. Position 21 is predominantly A over G in Archaea and Bacteria. In Eukarya, only A21 is found except for G21 in each of the three copies of elongator tDNA-Met (CAT) in *S. pombe* (E02). This base stacks with the Levitt *trans*-Watson–Crick base pair R15-Y48 and interacts with U8-A14 to form the A21 … U8-A14 tertiary interaction (see Hou, 1994).

3D base pairs 18–55 and 19–56

Our search procedure requests G at position 18 and R (purine) at position 19. The interloop *trans*-Hoogsteen bp 18–55 is almost always G-T (U55 in the corresponding tRNA transcript being mostly modified into pseudouridine), except in the archaeons *T. acidophilum* (A09, 9 tDNAs) and *F. acidarmanus* (A10, 10 tDNAs) where it is G18-C55. Position 19 is invariably G in Eukarya and Archaea whereas in Bacteria, only 14 As are found in this position, namely in four tDNAs (specific for Ala (TGC), Cys (GCA), Asp (GTC), and Glu (TTC) of *M. genitalium* (B22) and *M. pneumoniae* (B23)); in three tDNAs (specific for Ala (TGC), Asp (GTC), and Glu (TTC) of *U. urealyticum* (B24)); in tDNA-Pro (GGG) and in tDNA-Arg (GCG) of *H. pylori* (B20) and tDNA-Arg (TCT) of *X. fastidiosa* (B25). In these 14 bacterial tDNAs, T is found at position 56 (Y for C or T in Fig. 4) thus forming an A19-T56 interloop *trans*-Hoogsteen pair. Exceptions are the tDNA-Cys (CCA) of *M. pneumoniae* (B23) and of *M. genitalium* (B22) and the tDNA-Arg (TCT) of *X. fastidiosa* (B25), in which base opposition 19–56 is A:C. This implies either an A-C pairing or the existence of a C-to-U posttranscriptional editing process. Base pair G19-T56 is found only once, in one of the two different tDNAs-Ile (CAT reading ATA) of *Lactococcus lactis* (B06). In Eukarya, G19-T56 is found in 5 of the 33 tDNAs-Ser (AGA) of *A. thaliana* (E07), and in 3 tDNAs of *H. sapiens* (E05).

Stretch of bp 27 to 31/39 to 43

These bases make up the anticodon stem. The case is similar to that of the acceptor stem, except that bp 30–40 is mostly G-C or C-G (noted S-S), especially in Archaea. Also, bp 29–41, in the middle of the stem, is obviously the most constrained to *cis*-Watson–Crick pairing (G-T, T-G, and other mismatched base pairs are found only 15 times in 4,204 tDNAs examined and never in Archaea). Base pair 27–43 at the top of the anticodon helix, in combination with other nucleotides in the D-helix (especially purine-24), are critical for accurate codon reading at the wobble position of the anticodon (Schultz & Yarus, 1994; Yarus & Smith, 1995).

Base opposition 32–38

Watson–Crick base pairs, especially C-G or G-C (noted S-S) are rarely encountered at positions 32 and 38. However, in the majority of cases, base 32 can form a variety of isosteric non-Watson–Crick base pairs with base 38 (C:A, U:A, U:C, C:C …) that are essential for the formation of a more rigid anticodon hairpin (for details, see Auffinger & Westhof, 1999).

Base 33

This base is nearly always a T, except in initiator tDNA-iMet of higher Eukarya (see Initiator tDNA-iMet is the most conserved tDNA across the three domains but nevertheless has domain-specific features; Figs. 3 and 4). This T33 (U33 in the mature tRNA) is part of the so-called U-turn (Quigley & Rich, 1976), and is a characteristic signature of the tRNA anticodon loop. It interacts with the sugar–phosphate backbone of bases 36 and 35 as well as with the aromatic ring of base 35, thus helping to expose the anticodon to the solvent region, thereby facilitating the codon–anticodon interaction (Auffinger & Westhof, 2001). At position 33, C is occasionally found in some elongator tDNAs, like in tDNA-Val (GAC) in *A. aeolicus* (B09), tDNA-Leu (GAG) in *T. pallidum* (B01), tDNA-Arg (TCT) in *C. jejuni* (B17), tDNA-Arg (ACG) in *P. aeroginosa* (B13) and six tDNAs in Archaea. In Eukarya, C is also found in the unusual single-copy tDNA-Leu (GAG) of *S. cerevisiae* (E01) and in only one of the six copies of tDNA-Phe (GAA) of *H. sapiens* (E05). An outstanding exception is the tRNA-Ser of *Candida albicans* harboring a Leu anticodon CAG complementary to the Leu codon CTG (Perreau et al., 1999). This tRNA bears the only G33 identified in any of the tRNAs sequenced so far. This peculiarity exists in tRNA-Ser (CUG) of most genus of *Candida* (Yokogawa et al., 1992; Ohama et al., 1993; Santos et al., 1997). The effect of G33 is to alter the anticodon–stem-loop structure, thus lowering the decoding efficiency of the tRNA-Ser (CAG) and preventing binding of the leucyl-tRNA synthetase (Santos et al., 1996; Suzuki et al., 1997). In the tDNAs of the 50 genomes explored in this work, G33 has never been found.

Bases 34, 35, 36

These three bases make up the anticodon. The first base, at position 34, also called the wobble-base (Crick, 1966), is paired to the third base of the codon during mRNA translation. A34 is not found in Archaea with the sole exception of tDNA-Leu (AAG) in *F. acidarmanus* (A10). In Bacteria, aside from tDNA-Arg (AGC), A34 is found in the two copies of tDNA-Leu (AAG) in *L. lactis* (B06). In fully matured tRNA, this A34 is always deaminated into inosine by a specific tRNA A34-deaminase (reviewed in Grosjean et al., 1996; Gerber & Keller, 2001). U34 and to a lesser extent C34 and G34 are often modified after transcription of the tRNA gene (for review, see Björk, 1995; Curran, 1998).

Base 37

This base is always a purine, except in a very few cases (possibly sequencing errors). In fully matured tRNA, this purine is often modified or even hypermodified (for review, see Björk, 1995; Curran, 1998). It stabilizes the codon–anticodon interaction by stacking on the very discriminative base pair formed between the first base of the codon and the third base of the anticodon (position 36; discussed in Houssier & Grosjean, 1985). Together with other bases of the anticodon, this base often plays a role of determinant or antideterminant for interaction with several aminoacy-tRNA synthetases (reviewed in Giegé et al., 1998).

V-arm

The so-called variable arm (V-arm, see below A nearly consistent and universal rule for the long V-arms of tDNA-Leu, tDNA-Ser, and bacterial tDNAs-Tyr) spans positions 45 to 48. In tRNA-Ser, this arm is a major determinant for seryl-tRNA synthetases of the three domains, whereas in tRNA-Leu, the V-arm has no influence on the recognition by the leucyl-tRNA synthetase (reviewed in Giegé et al., 1998).

Base 47

This base belongs to the V-arm and is often missing, thus leaving a V-arm composed of only 4 bases (sometimes referred to as V-4 type of tRNA) instead of 5 as in most tDNAs except all tDNA-Leu and all tDNA-Ser of the three biological domains and also all tDNA-Tyr in Bacteria (see below A nearly consistent and universal rule for the long V-arms of tDNA-Leu, tDNA-Ser, and bacterial tDNAs-Tyr). A correlation between the absence of a nucleotide at position 47 (4-base V-arm) and the presence of a wobble U13:G22 base pair in the D-stem has been noticed and discussed in relation to tRNA evolution (Steinberg & Ioudovitch, 1996; Cermakian et al., 1998).

Stretch of base pairs 49 to 53/61 to 65

These bases form the T-stem that stacks coaxially on the amino acid stem to form the characteristic long T- and acceptor branch that is part of the L-shaped tRNA architecture (see Fig. 2A). A predominance of G-C base pairs appears gradually from base pair 50–64 to the highly conserved G-C base pair at positions 53–61. The search procedure requests R (A or G) at position 53 and Y (C or T) at position 61. The base pair A53-T61 appears very rarely in each of the three domains of life: notable examples are all nine copies of tDNA-Ala (AGC) in *S. pombe* (E02); tDNA-Pro (TGG), tDNA-Pro (CGG), and tDNA-Thr (GGT) in *Halobacterium* sp. NRC-1 (A06); tDNA-Pro (TGG) and tDNA-Pro (GGG) in *F. acidarmanus* (A10); tDNA-Val (CAC) in *M. barkeri* (A11); and tDNA-Ala (GCG) and tDNA-Glu (GAA) in *Synechocystis* 6803 (B04) and *Anabaena* 7120 (B05) (see Fig. 4). The wobble G53-T61 pair is found in only 2 of the 60 copies of tDNA-Tyr (GTA) in *A. thaliana* (E07).

T-loop (bases 54 to 60)

These bases form a seven-membered loop containing the characteristic sequence TTCRA found in a large majority of tRNAs. Base 54, the first base of the TΨC sequence in the mature tRNA, is nearly always modified into m^5U (ribo-T) in mature tRNA except in few archaeal tRNAs where $m^1\Psi$ is found (see Auffinger & Westhof, 1998). Also, A54 is present (probably unmodified) in 4 archaeal elongator tDNAs and 39 eukaryotic elongator tDNAs. A remarkable exception is C54, so far exclusively found in all of the 10 copies of tDNA-His (GTG) in *A. thaliana* (E07). In both Eukarya and Bacteria, base 55 is always a T in the tRNA gene, in fact pseudouridine (Ψ, an isomer of uridine) in the mature tRNA. In Archaea, base 55 is also sometimes C. Base 56 is always C in Archaea whereas in Eukarya, A56 and G56 are each encountered once [in *C. elegans* (E03)] and T56 encountered 8 times [in 5 of the 33 copies of tDNA-Ser (AGA) of *A. thaliana* (E07) and in 3 unique tDNAs of *H. sapien*s (E05)]. Because C56 is absolutely required for the binding of the eukaryotic transcription factor TFIIIC (Dieci et al., 2002; see Distribution of tRNA genes in genomes and peculiarities of their transcription), these tDNAs could be in fact pseudogenes. In Bacteria, T56 is found 12 times, namely in tDNA-Ala (TGC), tDNA-Asp (GTC), and tDNA-Glu (TTC) of *M. genitalium* (B22), *M. pneumoniae* (B23), and *U. urealyticum* (B24); in tDNA-Pro (GGG) and tDNA-Arg (GCG) of *H. pylori* (B20); and in one of the two copies of tDNA Met-(CAT) of *L. lactis* (B06). Base 58, without exception is A (requested as such in the cloverleaf search procedure), often modified into N^1-methyladenosine in mature tRNA of Eukarya and Archaea. At position 59, there is a strong preference for A59 in Archaea and counterselection for C59 in Bacteria. The base at position 60 is variable, yet A60, together with A54, is found in all eukaryotic tDNA-iMet (CAT) (see above Initiator tDNA-iMet is the most conserved tDNA across the three domains but nevertheless has domain-specific features). A few eukaryotic elongator tDNAs have A at either position 54 or 60, but never at both positions (the only exception is in a single copy of tDNA-Val (GTT) in *H. sapiens* (E05)).

3D base pair 54–58

T at position 54 is usually paired with the conserved A at position 58 to form a reverse *trans*-Hoogsteen base pair (Romby et al., 1987). This base pair has been shown to be an essential identity element for the tRNA-pseudouridine synthase that catalyzes the formation of the almost universally conserved Ψ55 (Becker et al., 1997; Gu et al., 1998; Hoang & Ferré-D'Amare, 2001). This intraloop base pair is stacked on bp G53-C61 at the end of the T-stem and forces the two bases at positions 59 and 60 to loop out, thus forming a characteristic T-loop of 4 bases instead of 7. This characteristic T-loop conformation was identified by chemical probing (Romby et al., 1987), NMR spectroscopy (Koshlap et al., 1999), and more recently confirmed by X-ray crystallography (Hoang & Ferré-D'Amare, 2001). It was also shown to be important for recognition by the elongation factors during protein synthesis (Rudinger et al., 1994) and for RNase P, which cleaves the 5′-leader sequence during the tRNA maturation process (Hardt et al., 1993; Loria & Pan, 1997; Levinger et al., 1998; Odell et al., 1998). However, controversial data are reported concerning this last point (Tuohy et al., 1994). Interestingly, inspection of the nucleotide sequences reveals that such a characteristic "4-base T-loop" conformation probably cannot occur within the 7-base anticodon loop of any tRNA. Indeed, the required bp G31-C39 and the reverse *trans*-Hoogsteen bp T32-A36 (homologous of the G53-C61/T54-A58 base pairs in the T-loop) were found only once in the 4204 tDNAs examined (in the tDNA-Ser (GGA) of *L. monocytogenes* (B07)). Evidently, during evolution, the anticodon sequence of tRNAs in all three domains of life have been selected in order to prevent the anticodon-loop from adopting a characteristic 4-base T-loop conformation (see also discussion in Dirheimer et al., 1995). Noteworthy, swapping of the T-loop and anticodon loop revealed that the T-loop as such, rather than its location within the tRNA molecule, is essential for both m^5U54 (ribo-T54) and Ψ55 formation (Becker et al., 1997).

Base 73

The fourth and unpaired nucleotide from the tRNA 3′ end is the so-called "discriminator base" (Crothers et al., 1972), owing to its role in the selection of tRNAs by the aminoacyl-tRNA synthetases (reviewed in Giegé et al., 1998). As shown in Figure 5B and Table 2, a striking constancy of this nucleotide exists in tRNAs specific for a given amino acid and within a given domain of life. For example, in Archaea and Bacteria, C73 is found exclusively in tDNAs-His (GTG), in which it is usually paired to an extra G at position (−1) in the mature tRNA (Cs in red and boxed in Fig. 5B, line (c)). In Eukarya, C73 is, without exception, the hallmark of all tDNA-Pro (NGG family box) and with a single exception, none of these tDNA-Pro has an extra G(−1) (Cs in red, shaded, and not boxed; Fig. 5B, line (a)). Conversely, all tDNA-Pro in Archaea and Bacteria and tDNA-His in Eukarya bear A73, as do almost all tDNA specific for Phe, Leu, Ile, Met, Val, Ala, and Tyr from the three biological domains. Another interesting case is that of tDNA-Gln (YTG family box), in which a G73 is found in almost all tDNA-Gln of Bacteria, a A73 in Archaea, and mostly a T73 in Eukarya (Fig. 5B, line (d)). T73 is also found in all tDNA-Cys (GCA) from the three biological domains (Fig. 5B, line (e)), as well as all tDNA-Thr (NGT family box) of Archaea (Fig. 5B, line (b)) and tDNA-Gly (NCC family box) of Bacteria (Fig. 5B, line (g)); whereas tDNA-Thr (NGT family box) from both Bacteria and Eukarya harbor either T73 or A73 (Fig. 5B, line (b)) and tDNA-Gly (NCC family box) from both Eukarya and Archaea all harbor A73 (Fig. 5B, line (g)). In the case of tRNA-Trp (CCA), A73 is always found in Eukarya and Archaea, whereas G73 prevails in Bacteria (Fig. 5B, line (f)), even in the tDNA-Trp (TCA) corresponding to the tRNA-Trp reading the *opal* stop codon. In the latter case, NMR spectroscopy revealed that mutation of A73 to G73 in bovine tRNA-Trp, elicited a conformational alteration in the G1-C72 base pair (Guo et al., 2002). It is worth mentioning that the tRNA (CAG) of *C. albicans*, which translates the standard Leu CUG codon as Ser (Santos et al., 1996), as well as the same tRNA from other *Candida* species identified so far (Ohama et al., 1993; Santos et al., 1997) harbor a G73 as in all tRNA-Ser (instead of A73 as in all tDNAs-Leu). This G73, together with the characteristic long V-arm, are identity elements for the seryl-tRNA synthetases (reviewed in Giegé et al., 1998). All of these observations are consistent with the fact that the discriminator N73 and the three first base pairs at the end of the acceptor stem are crucial for effective recognition of tRNA by most aminacyl-tRNA synthetases (discussed in Giegé et al., 1998).

Posttranscriptional processing at 3′ and 5′ termini of tRNA

Not all prokaryotic tDNAs encode the universal 3′-terminal CCA

To function in protein synthesis, all cytoplasmic mature tRNAs must have a 3′-terminal CCA terminus (3′-CCA) in positions 74, 75, and 76 to which an amino acid becomes covalently attached at the 2′- or 3′-hydroxyl ribose of the terminal A76 residue. This conserved 3′-CCA sequence is also necessary for the exact positioning of the peptidyl-tRNA at the P-site and of the aminoacyl-tRNA at the A-site on the large ribosomal subunit to facilitate peptide bond formation. In Eukarya, none of the tDNAs encode the 3′-CCA, whereas in many Bacteria and in some Archaea, tDNAs possess the 3′-CCA. Only in a few species do all tDNAs encode the 3′-CCA (see Fig. 3, lines A02, A04, and B17–B30). The physiological significance of such differences in 3′-terminal composition among tDNAs in different organisms is not known. However, in all cells examined so far, a CCA-adding enzyme (ATP(CTP): tRNA nucleotidyltransferase) or a combination of separate CC-adding and A-adding enzymes have been identified, which catalyze the elongation of all tRNAs lacking the CCA-terminus (Deutscher, 1990; Tomita & Weiner, 2001). This “template-free” process (“untemplated” nucleotide addition) occurs on pre-tRNA transcripts after trimming of the 5′ and 3′ ends and on mature tRNA lacking the 3′-CCA because of degradation by ribonuclease. Interestingly, in eukaryotic pre-tRNAs, which always lack the 3′-CCA, it has been shown that the presence of 3′-CCA itself, as in a mature tRNA, is an antideterminant for the 3′ end endonuclease (3′-tRNase), which normally acts before the CCA-adding enzyme (Mohan et al., 1999). Therefore, the “raison d’être” of the nucleotidyltransferase is not only to generate the 3′-CCA on primary transcripts that lack this sequence, but also to constantly maintain the 3′-CCA for proper functioning of tRNAs within the cell. Only in organisms where not every tRNA gene encodes its 3′-CCA are the CCA-adding enzymes essential for cell viability (Aebi et al., 1990). In organisms where all 3′-CCA termini are encoded, inactivation of the CCA-adding enzyme is not fatal and generally results in only decreased growth rate, possibly because of some functional overlap of tRNA nucleotidyltransferase, poly(A) polymerase I, and polynucleotide phosphorylase activities in vivo (Reuven et al., 1997). The rationale for all this may be understood in terms of the emergence of RNA molecules from an RNA world (Maizels & Weiner, 1999).

5′-G addition to tRNA-His in Eukarya versus its systematic encoding in Archaea and Bacteria

All tDNAs are transcribed as tRNA precursors with 5′ extensions that have to be removed posttranscriptionally by the RNase P machinery, leaving the 5′ end of all mature tRNAs with the first nucleotide conventionally numbered 1 (see Fig. 2A,B). In all fully mature tRNAs-His, an extra 5′ GMP is always found (at position −1) that is complementary to the fourth base (C at position 73) from the 3′ terminus, thus creating an acceptor/amino acid stem of 8 bp instead of 7 as in all other tRNAs. In all the 30 Bacteria examined except one, *S. meliloti* (B30), this extra G exists at the appropriate position in the tDNA-His (GTG) (Cs in red and boxed in Fig. 5B, line (c)), thus arguing that the extra G(−1)-C73 base pair may result from unusual processing by RNase P (Orellana et al., 1986; Burkard et al., 1988; Krupp et al., 1991). In the tDNA-His of *S. meliloti* (B30), there is a T at position −1 and a A at position 73, but it is not known whether the corresponding mature tRNA-His (GTC) has an acceptor stem of 8 or 7 bp. In Archaea, G(−1) is also encoded in the tDNA-His of all but 3 of the 13 species examined (*M. kandleri* (A01), *P. aerophilum* (A03) and *Methanobacterium thermoautotrophicum* (A13), see, in Fig. 5B, line (c), the 3 Cs shaded but not boxed). As in Bacteria, such exceptions could possibly result from unusual processing by archaeal RNase P (Pannucci et al., 1999). In contrast to bacterial and archaeal RNase P, the corresponding eukaryotic enzyme does not leave an extra nucleotide, but cleaves the tRNA-His precursors as in all other tRNAs, the extra GMP being added after transcription by a specific tRNA-His guanylyltransferase (Pande et al., 1991). Thus again, as seen above for the CCA-adding enzyme, depending on the organism, the extra G at position (−1) in tRNA is either generated during transcription as part of the tDNA or attached posttranscriptionally by the tRNA-His-guanyltransferase. To date, only the enzyme from yeast has been identified and characterized (Jahn & Pande, 1991; Pande et al., 1991).

The general rule seems to be that, unlike in Eukarya, nearly all Archaea and Bacteria encode both G(−1) and C73, thus making this G-C pair a unique determinant of tRNA-His isoacceptors.

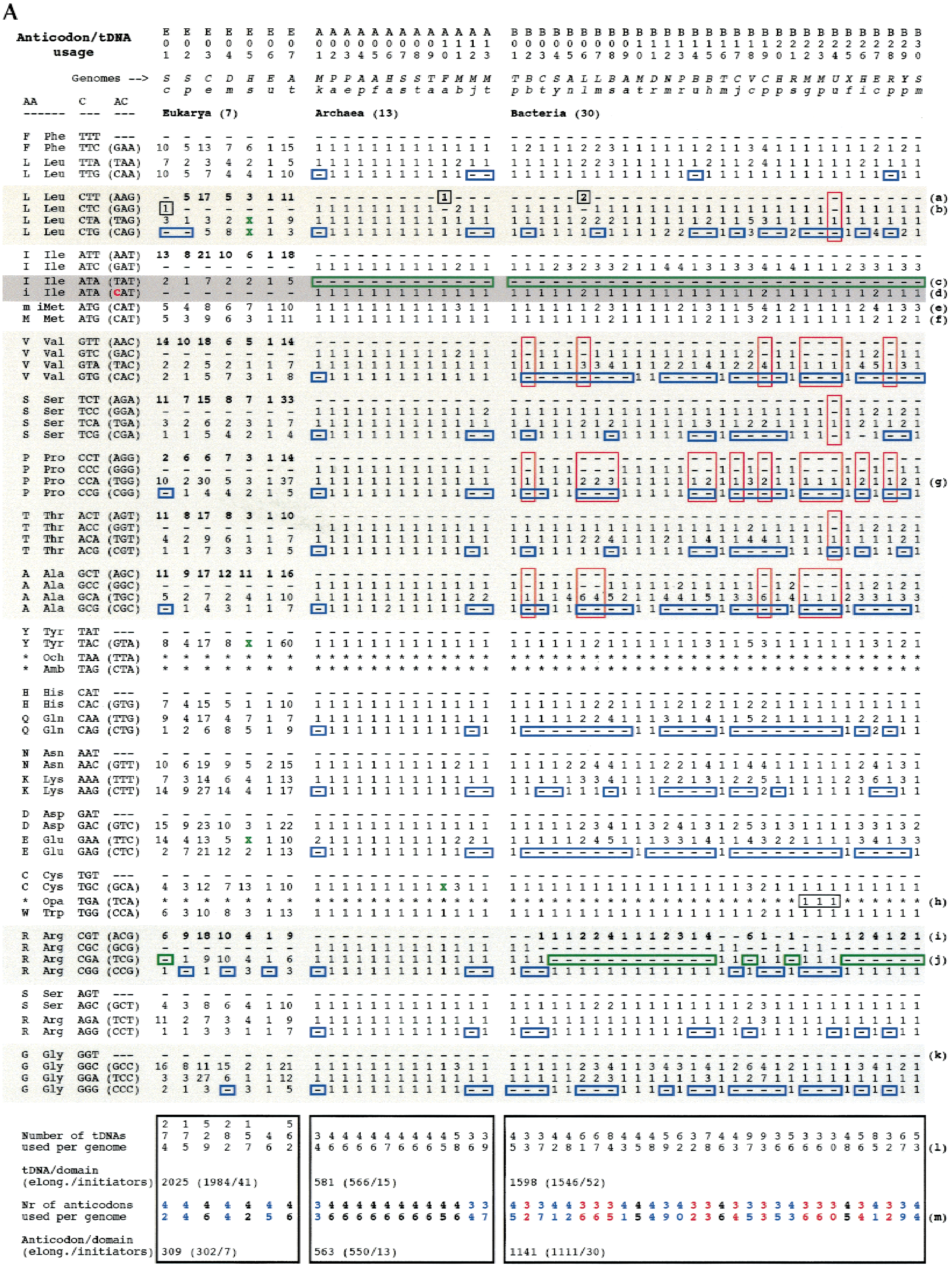
Common and domain-specific strategies exist for decoding the genetic information with a minimum number of isoacceptor tRNAs

During translation on the ribosome, each codon of a messenger RNA has to be read by a tRNA bearing a complementary anticodon. However, because different types of relaxed base pairings are allowed at the "wobble" position 34 of the tRNA anticodon, certain isoacceptors can read two, three, or even four synonymous codons differing by the third base (Crick, 1966; reviewed in Lim & Curran, 2001). Hence, the 62 codons (61 sense codons plus the initiator Met AUG codon) are always read, in any known organism, by far fewer than 62 isoacceptors. On the other hand, the total number of genes encoding the various isoacceptors within a given genome vary significantly from one organism to another and can amount to several hundreds, as in most multicellular eukaryotic cells. These two important features of the decoding system must not be confused and we have to distinguish and define in each organism: (1) the tDNA redundancy that corresponds to all tRNA genes having the same anticodon (also known as the tDNA usage or tRNA gene copy number), and (2) the pattern of anticodons present in the complete tRNA repertoire (also known as the anticodon usage or anticodon choice pattern). We propose to group these two features within a given genome under the name "tRNome" or "tRNomics" (Filipowicz, 2000). Figure 6A) reports both types of information for all genomes examined and Figure 6B summarizes the different anticodon-sparing strategies and "decoding modes" used throughout the three domains.

tDNA redundancy

One general trend that appears from Figure 6A is the obvious high gene redundancy of tDNAs harboring the same anticodon in Eukarya (except in the parasite *E. cuniculi* (E06)), which can be as high as 60 copies, as in tDNA-Tyr (GTA) from *A. thaliana* (E07). As shown in the bottom of Figure 6A, the total number of tDNA per genome in Eukarya can reach 562 as in the case of *A. thaliana* (E07) (see line (l)) and several thousand tDNAs are anticipated to exist in the human genome when completed (Lander et al., 2001). This trend contrasts with the evidently low level of tDNA redundancy in all archaeons as well as in several bacteria, in which only one or occasionally two or three copies of a given tRNA gene exists. A rough correlation exists between the total number tRNA genes (Fig. 6A, line (l)) and the genome size (compare the number of tRNAs per ge-

FIGURE 6. Anticodon/tDNA usages and decoding strategies throughout genomes. **A:** Anticodon and tDNA usages: The presentation of this array is similar to that of Figure 5. Each element of the table is the number of tDNAs bearing the anticodon indicated in the "AC" column. If this number is null, - is used instead for clarity. Gray background is used to indicate four-codon sets in which all codons code for the same amino acid. Darker gray background highlights the regular tDNA-Ile (TAT) (indicated "I Ile ATA (TAT)") used in Eukarya only and the special tDNA-Ile (CAT reading ATA) (indicated "i Ile ATA (CAT)"), with a red C) found in Archaea and Bacteria. The initiator tDNA-iMet (CAT) is indicated "m iMet ATG (CAT)". Black bold numbers are used to highlight eukaryotic tDNAs that use A in the first position of the anticodon (position 34). Such tDNAs are absent from Archaea and Bacteria, with the exceptions of the tDNA-Arg (ACG) in some Bacteria and also two cases of tDNA-Leu (AAG), one in Archaea and the other in Bacteria. Five green Xs are used to indicate eukaryotic or archaeal tDNAs probably missing. Blue boxes highlight the absence of C at position 34 of the anticodon. In this case of the C34-sparing strategy, the corresponding codon is read by the anticodon with U34 (T34 in the tDNA). Red boxes denote cases of "two-out-of-three" sparing in which a single U34 anticodon reads the four codons. Green boxes denote the two special cases of U34-sparing (lines (c) and (j)); the latter sparing cannot be cumulated with the C34 sparing (blue boxes). At the bottom are indicated the total number of tDNAs found in each genome (line (l)) and domain (the number of elongators and initiators are indicated separately between parentheses). The number of anticodons used in each genome (anticodon usage) is indicated in line (m) (to be read vertically). Key to color code for the anticodon: black: organisms using decoding mode I; blue: organisms using decoding mode II; red: organisms using decoding mode III. To compute the anticodon usage, the 3 CAT anticodons belonging to the three tDNA-Ile (CAT reading ATA), initiator tDNA-iMet (CAT), and elongator tDNA-Met (CAT) are considered as being different anticodons. See text for the special features signaled with the letters between parentheses at right. **B:** Decoding strategies: The different anticodon-sparing strategies and decoding modes used in the three biological domains are summarized. Blue, red, and green boxes have the same meaning as in Fig. 6A. Conventions used in the three genetic code arrays at right are as follows. The bases and base pairings in this figure are solely based upon unmodified tRNA sequences. For example, A34 is, in fact, always modified into I34 in the mature tRNA (see Curran, 1998). Bases indicated inside the array (position 34) and solid rectangles indicate the presence of a tRNA bearing the corresponding anticodon; an empty rectangle indicates that the corresponding codon is read by the anticodon to which that empty rectangle is linked through a vertical bar. When two alternate configurations are present side by side (e.g., tDNA-Ser (GGA) and tDNA-Ser (AGA)), only one holds at a same time. Gray background is used to indicate four-codon sets (family box) in which all codons code for the same amino acid. In the decoding boxes at left, horizontal arrows indicate Watson–Crick base pairing, oblique arrows indicate wobble base pairing. In the special box noted Ile/Met/iMet, the oblique dashed green arrow indicates that in Archaea and Bacteria, the AUA codon is read by a tRNA-Ile (CAU reading AUA) due to a modification of the C in the first position of the anticodon (see text).



Anticodon/tDNA usage				E01	E02	E03	E04	E05	E06	E07	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	B01	B02	B03	B04	B05	B06	B07	B08	B09	B10	B11	B12	B13	B14	B15	B16	B17	B18	B19	B20	B21	B22	B23	B24	B25	B26	B27	B28	B29	B30	
		Genomes -->		*Sc*	*Sp*	*Ce*	*Dm*	*Hs*	*Eu*	*At*	*Mk*	*Pa*	*Pe*	*Ap*	*Af*	*Ha*	*Ss*	*St*	*Ta*	*Fa*	*Mb*	*Mj*	*Mt*	*Tp*	*Bb*	*Ct*	*Sy*	*An*	*Ll*	*Lm*	*Bs*	*Aa*	*Mt*	*Dr*	*Nm*	*Pr*	*Bu*	*Bh*	*Tm*	*Cj*	*Vc*	*Cp*	*Hp*	*Rs*	*Mg*	*Mp*	*Uu*	*Xf*	*Hi*	*Ec*	*Rp*	*Yp*	*Sm*	
AA		C	AC	**Eukarya (7)**							**Archaea (13)**													**Bacteria (30)**																														
F	Phe	TTT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
F	Phe	TTC	(GAA)	10	5	13	7	6	1	15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	2	2	3	1	1	1	1	1	1	2	1	1	3	4	1	1	1	1	1	1	1	2	1	2	1	
L	Leu	TTA	(TAA)	7	2	3	4	2	1	5	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	2	3	1	1	1	1	1	1	2	1	1	2	4	1	1	1	1	1	1	2	1	1	1	1	
L	Leu	TTG	(CAA)	10	5	7	4	4	1	10	-	1	1	1	1	1	1	1	1	1	1	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	
L	Leu	CTT	(AAG)	-	5	17	5	3	1	11	-	-	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(a)
L	Leu	CTC	(GAG)	1	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	-	2	1	1	1	1	1	1	1	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	1	1	1	(b)
L	Leu	CTA	(TAG)	3	1	3	2	x	1	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1	1	2	1	1	5	3	1	1	1	1	1	1	1	1	1	1	1	
L	Leu	CTG	(CAG)	-	-	5	8	x	1	3	-	1	1	1	1	1	1	1	1	1	1	-	-	1	-	1	1	1	1	-	1	1	1	1	2	2	-	-	1	-	3	-	-	2	-	-	-	1	-	4	-	2	1	
I	Ile	ATT	(AAT)	13	8	21	10	6	1	18	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
I	Ile	ATC	(GAT)	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	3	2	3	3	2	1	1	4	4	1	3	1	3	3	4	1	4	1	1	1	2	3	3	1	3	3	
I	Ile	ATA	(TAT)	2	1	7	2	2	1	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(c)
i	Ile	ATA	(CAT)	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	2	1	1	1	(d)
m	iMet	ATG	(CAT)	5	4	8	6	7	1	10	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	1	2	3	1	1	1	1	2	1	3	1	1	5	4	1	2	1	1	1	1	2	4	1	3	3	(e)
M	Met	ATG	(CAT)	5	3	9	6	3	1	11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	1	1	3	2	1	1	1	1	1	1	1	2	1	2	1	(f)
V	Val	GTT	(AAC)	14	10	18	6	5	1	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
V	Val	GTC	(GAC)	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	2	1	1	1	-	1	1	1	-	1	1	1	1	1	1	1	1	1	1	1	2	-	1	1	-	-	-	1	1	2	-	1	1	
V	Val	GTA	(TAC)	2	2	5	2	1	1	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	4	1	1	1	2	2	1	4	1	2	2	4	1	1	1	1	1	1	4	5	1	3	1	
V	Val	GTG	(CAC)	2	1	5	7	3	1	8	-	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	1	1	-	-	-	-	1	-	-	-	-	1	-	-	-	1	-	-	-	-	-	
S	Ser	TCT	(AGA)	11	7	15	8	7	1	33	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S	Ser	TCC	(GGA)	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	2	1	2	1	
S	Ser	TCA	(TGA)	3	2	6	2	3	1	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1	1	1	1	1	3	1	1	2	2	1	1	1	1	1	1	2	1	1	2	1	
S	Ser	TCG	(CGA)	1	1	5	4	2	1	4	-	1	1	1	1	1	1	1	1	1	1	-	-	1	-	1	1	1	1	1	-	1	1	1	1	1	-	-	1	-	-	-	-	1	1	1	-	1	-	1	-	-	1	
P	Pro	CCT	(AGG)	2	6	6	7	3	1	14	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
P	Pro	CCC	(GGG)	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	-	-	-	1	1	1	1	1	-	-	1	-	1	-	1	1	-	-	-	1	-	1	-	1	1	
P	Pro	CCA	(TGG)	10	2	30	5	3	1	37	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	1	1	1	1	1	1	2	1	1	3	2	1	1	1	1	1	1	2	1	1	2	1	(g)
P	Pro	CCG	(CGG)	-	1	4	4	2	1	5	-	1	1	1	1	1	1	1	1	1	1	-	-	1	-	-	1	1	-	-	-	1	1	1	1	1	-	-	1	-	-	-	-	1	-	-	-	1	-	1	-	-	1	
T	Thr	ACT	(AGT)	11	8	17	8	3	1	10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
T	Thr	ACC	(GGT)	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	-	1	1	2	1	2	1	
T	Thr	ACA	(TGT)	4	2	9	6	1	1	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	4	1	1	1	2	1	1	4	1	1	4	4	1	1	1	1	1	1	1	1	1	2	1	
T	Thr	ACG	(CGT)	1	1	7	3	3	1	5	-	1	1	1	1	1	1	1	1	1	1	-	1	1	-	1	1	1	1	1	-	1	1	1	1	1	-	-	1	-	-	-	-	1	1	1	-	1	-	1	1	-	1	
A	Ala	GCT	(AGC)	11	9	17	12	11	1	16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
A	Ala	GCC	(GGC)	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	-	-	1	1	1	1	1	2	1	1	1	1	1	-	1	2	-	-	-	1	1	2	1	2	1	
A	Ala	GCA	(TGC)	5	2	7	2	4	1	10	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	4	6	4	5	2	1	1	4	4	1	5	1	3	3	6	1	4	1	1	1	2	3	3	1	3	3	
A	Ala	GCG	(CGC)	-	1	4	3	1	1	7	-	1	1	1	1	2	1	1	1	1	1	-	-	1	-	-	1	1	-	-	-	-	1	1	-	-	-	-	1	-	-	-	-	1	-	-	-	1	-	-	-	-	1	
Y	Tyr	TAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Y	Tyr	TAC	(GTA)	8	4	17	8	x	1	60	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	3	1	1	5	3	1	1	1	1	1	1	1	3	1	2	1	
*	Och	TAA	(TTA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
*	Amb	TAG	(CTA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
H	His	CAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
H	His	CAC	(GTG)	7	4	15	5	1	1	10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	2	1	2	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	
Q	Gln	CAA	(TTG)	9	4	17	4	7	1	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	4	1	1	1	3	1	1	4	1	1	5	2	1	1	1	1	1	1	2	2	1	1	1	
Q	Gln	CAG	(CTG)	1	2	6	8	5	1	9	-	1	1	1	1	1	1	1	1	1	1	-	1	1	-	-	-	-	-	-	-	-	1	1	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	2	-	1	1	
N	Asn	AAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
N	Asn	AAC	(GTT)	10	6	19	9	5	2	15	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	2	2	4	4	1	1	1	2	2	1	4	1	1	4	4	1	1	1	1	1	1	2	4	1	3	1	
K	Lys	AAA	(TTT)	7	3	14	6	4	1	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	3	4	1	1	1	2	2	1	3	1	2	2	5	1	1	1	1	1	2	3	6	1	3	1	
K	Lys	AAG	(CTT)	14	9	27	14	4	1	17	-	1	1	1	1	1	1	1	1	1	1	-	-	1	1	-	-	1	1	1	-	1	1	1	-	-	-	-	1	-	-	2	-	1	1	1	1	1	1	-	-	1	1	
D	Asp	GAT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
D	Asp	GAC	(GTC)	15	9	23	10	3	1	22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	3	4	1	1	3	2	4	1	3	1	2	5	3	1	3	1	1	1	1	3	3	1	3	2	
E	Glu	GAA	(TTC)	14	4	13	5	x	1	10	2	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	2	4	5	1	1	1	3	3	1	4	1	1	4	3	2	2	1	1	1	1	3	4	1	3	3	
E	Glu	GAG	(CTC)	2	7	21	12	2	1	13	-	1	1	1	1	1	1	1	1	1	1	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	1	-	-	-	-	1	
C	Cys	TGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
C	Cys	TGC	(GCA)	4	3	12	7	13	1	10	1	1	1	1	1	1	1	1	1	x	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	3	2	1	1	1	1	1	1	1	1	1	1	1	
*	Opa	TGA	(TCA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	1	1	1	*	*	*	*	*	*	(h)
W	Trp	TGG	(CCA)	6	3	10	8	3	1	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	
R	Arg	CGT	(ACG)	6	9	18	10	4	1	9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	2	2	4	1	1	1	2	3	1	4	-	-	6	1	-	1	-	-	1	1	2	4	1	2	1	(i)
R	Arg	CGC	(GCG)	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	-	1	-	1	1	-	-	-	-	-	-	-	
R	Arg	CGA	(TCG)	-	1	9	10	4	1	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	1	1	-	1	1	-	1	1	1	-	-	-	-	-	-	(j)
R	Arg	CGG	(CCG)	1	-	1	-	3	-	3	-	1	1	1	1	1	1	1	1	1	1	-	-	1	-	-	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	-	-	1	-	-	-	1	1	1	1	1	1	
S	Ser	AGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S	Ser	AGC	(GCT)	4	3	8	6	4	1	10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	1	1	2	3	1	1	1	1	1	1	1	1	1	1	1	
R	Arg	AGA	(TCT)	11	2	7	3	4	1	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	3	1	1	1	1	1	1	1	1	1	1	1	
R	Arg	AGG	(CCT)	1	1	3	3	1	1	7	-	1	1	1	1	1	1	1	1	1	1	-	1	1	-	-	1	1	1	1	1	1	1	1	1	1	-	-	1	1	-	1	1	1	1	1	-	1	-	1	-	1	1	
G	Gly	GGT	---	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(k)
G	Gly	GGC	(GCC)	16	8	11	15	2	1	21	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1	1	1	2	3	4	1	1	3	4	3	1	4	1	2	6	4	1	2	1	1	1	1	3	4	1	2	1	
G	Gly	GGA	(TCC)	3	3	27	6	1	1	12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	3	1	1	1	1	1	1	3	1	1	2	7	1	1	1	1	1	1	1	1	1	1	1	
G	Gly	GGG	(CCC)	2	1	3	-	3	1	5	-	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	1	1	-	-	-	1	1	1	-	1	-	-	1	-	-	-	-	1	-	-	-	1	-	1	-	1	1	
Number of tDNAs used per genome				274	175	529	282	157	46	562	34	46	46	46	46	47	46	46	46	45	58	36	39	45	33	37	42	48	61	67	84	43	45	48	59	62	32	78	46	43	97	93	36	56	36	36	30	48	56	85	32	67	53	(l)
tDNA/domain (elong./initiators)				2025 (1984/41)							581 (566/15)													1598 (1546/52)																														
Nr of anticodons used per genome				42	44	46	44	42	45	46	33	46	46	46	46	46	46	46	46	45	46	34	37	45	32	37	41	42	36	36	35	41	45	44	39	40	32	33	46	34	35	33	35	43	36	36	30	45	34	41	32	39	44	(m)
Anticodon/domain (elong./initiators)				309 (302/7)							563 (550/13)													1141 (1111/30)																														

FIGURE 6. *See caption on previous page. Figure continues on next page.*

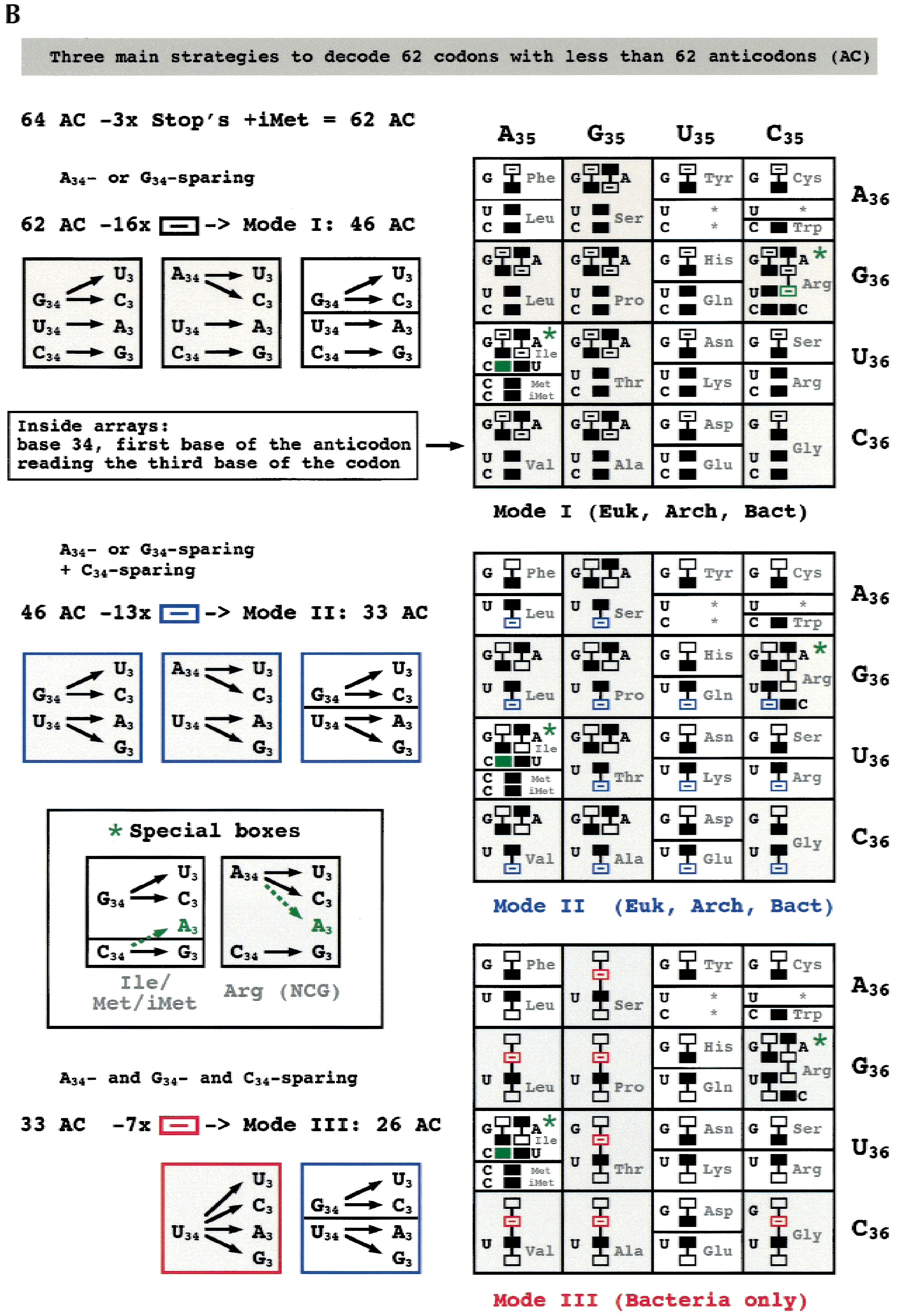

B
Three main strategies to decode 62 codons with less than 62 anticodons (AC)
64 AC -3x Stop's +iMet = 62 AC
A34- or G34-sparing
62 AC -16x ▭ -> Mode I: 46 AC
Inside arrays:
base 34, first base of the anticodon
reading the third base of the codon
A35 G35 U35 C35
A36 G36 U36 C36
Phe Leu Ser Tyr Cys Trp His Gln Arg Pro Ile Met iMet Thr Asn Lys Val Ala Asp Glu Gly
Mode I (Euk, Arch, Bact)
A34- or G34-sparing
+ C34-sparing
46 AC -13x ▭ -> Mode II: 33 AC
Mode II (Euk, Arch, Bact)
* Special boxes
Ile/
Met/iMet
Arg (NCG)
A34- and G34- and C34-sparing
33 AC -7x ▭ -> Mode III: 26 AC
Mode III (Bacteria only)

FIGURE 6. *Continued.*

nome with the genome sizes in Fig. 3; see also Andersson & Kurland, 1995). However, in eukaryotic species like *S. cerevisiae* (E01), *S. pombe* (E02), and *C. elegans* (E03), as well as in bacteria like *E. coli* (B27) and *B. subtilis* (B08), a fairly good correlation exists between the tDNA copy number for individual tRNA species, the corresponding cellular isoacceptors, and the codon usage (see, e.g., Ikemura, 1985; Sharp & Li, 1987; Dong et al., 1996; Percudani et al., 1997; Hani & Feldmann, 1998; Kanaya et al., 1999; Duret, 2000). In higher eukaryotes, as in *D. melanogaster* (E04), this correlation is less evident, and in *H. sapiens* (E05), for example, a systematic bias of CG-containing codon usage has also been revealed (Kanaya et al., 2001). This last bias is probably due to a peculiarity of the transcription regulation process via methylation of CG-dinucleotides (Colot & Rossignol, 1999). A yet unsolved important issue is how tRNA levels are regulated in cells with all tDNAs present as single copies. These include the parasitic eukaryote *E. cuniculi* (E06), the majority of archaeons (A01 to A13, but not A11), several parasitic human pathogen bacteria such as *T. pallidum* (B01), *C. trachomatis* (B03), and *R. prowazekii* (B28), and small organisms such as the wall-less prokaryote *U. urealyticum* (B24) and the two mycoplamas *M. genitalium* (B22) and *M. pneumoniae* (B23). In such type of cells, the concentration of each tRNA isoacceptor is probably solely tuned by transcriptional regulation and obviously not by gene dosage as in the majority of eukaryotic cells (reviewed in Inokuchi & Yamao, 1995; see also Palmer & Daniels, 1995; Ushida et al., 1996). Unfortunately, almost nothing is known about the relative abundance of tRNAs in each of the cells listed above and whether they have developed alternate regulation pathways.

Anticodon usage and decoding modes

Data in Figure 6A also reveal, for each genome, the anticodon-choice pattern. In this figure, each of the 66 lines of the array corresponds to one anticodon. Note that the anticodons of initiator tDNA-iMet (CAT) and elongator tDNA-Met (CAT) are considered to be different and therefore appear on two separate lines (lines (e) and (f)). The regular TAT anticodon of Eukarya and the special CAT anticodon reading the Ile ATA codon in Archaea and Bacteria also appear on two separate lines (lines (c) and (d)). We cannot exclude that a few isoacceptors are missing from this list (five obviously missing tDNAs are indicated with green X signs). However, the most interesting information is that for each genome, the number of anticodons that accounts for the reading of the 62 codons ranges from 30 to no more than 46 (as in most archaea and a few eukarya; line (m)). From the comparison of anticodon usages in the three biological domains, three major anticodon-sparing strategies can be deduced; the first two strategies are common to all domains, whereas the third one exists only in Bacteria.

First strategy—the “A34- or G34-sparing strategy”: Given the two last bases of an anticodon (positions 35 and 36), only anticodon starting with either G34 or A34 (the wobble base) exists (without any exception). The use of A34 (in fact, modified into inosine in the mature tRNA of most organisms; reviewed in Curran, 1998) or G34 (sometimes modified into 2′-*O*-methylguanosine or to derivatives of queuosine; reviewed in Curran, 1998) depends upon the domain and anticodon considered. In Eukarya, tRNAs bearing A34 (see black numbers highlighted in bold in Fig. 6A) are used essentially in the four-codon family boxes corresponding to Leu, Val, Ser, Pro, Thr, Ala, and Arg (except in the Gly-family box in all domains; line (k)), as well as in the three-codon box corresponding to Ile. In Bacteria, A34-containing tDNAs are only found in the Arg-family box, whereas in Archaea, only the tDNA-Leu (AAG) of *F. acidarmanus* (one copy) harbors A34. Notice also the presence of A34 in the same tDNA-Leu (AAG) in the bacterium *L. lactis* (two copies; boxed in Fig. 6A, line (a)). Conversely, in *S. cerevisiae* (E01), the anticodon (GAG), instead of the expected (AAG), is used in a minor tDNA-Ile (one copy; boxed in Fig. 6A, line (b)). Because this tDNA-Leu (GAG) is unique to the *S. cerevisiae* genome (Percudani et al., 1997; Hani & Feldmann, 1998), we rechecked and confirmed its sequence (data not shown). This tDNA is also the only one in *S. cerevisiae* with C33 instead of T33 (see Base 33 above). Despite these unusual features, the corresponding tRNA-Leu was shown to work properly in vivo (Sundararajan et al., 1999). Only one ortholog of this minor tDNA-Leu is known in phylogenetically related hemiascomycetes, that of *Pichia sorbitophila*, which uses the regular (AAG) anticodon (Souciet et al., 2000). Likewise, the genome of each of the two *Mycoplasma* species analyzed (*genitalium*, (B22) and *pneumoniae*, (B23)) does not contain any tDNA harboring A34 (including the tDNA-Arg (NCG); Fig. 6A, line (i)). On the contrary, in *Mycoplasma capricolum* and *mycoides*, a functional matured tRNA-Thr (AGU) bearing an anticodon with unmodified adenine was shown to exist (Andachi et al., 1987; Guindy et al., 1989; Inagaki et al., 1995). Interestingly, in *M. capricolum*, mature tRNA-Arg bears the anticodon ICG (Andachi et al., 1989) attesting that an A34-deaminase specific only for A34-containing tRNA-Arg (ACG) is present in this organism, as is also the case in *E. coli* (Auxilien et al., 1996).

Anticodon GNN (or modified *GNN) reads the two synonymous codons NNU and NNC; anticodon INN reads mostly NNC/NNU codons, inefficiently reads NNA codons, and never reads NNG codons, whereas anticodon ANN reads all four families of codons NNC, NNU, NNA, and NNG, yet with different relative efficiencies (Munz et al., 1981; Boren et al., 1993; Curran, 1995; reviewed in Curran, 1998). As shown in Figure 6A and in the upper part of Figure 6B, this first anticodon-sparing strategy (decoding mode I: “A34- or G34-sparing strategy”) applies to all three biological domains and concerns 16 tDNAs that can be economized, thus lowering the number of anticodons required to decode all 62 sense codons to 62 − 16 = 46. Note that, in this mode I decoding system, the symmetric rule of the A34- or G34-sparing strategy does not hold for U34 or C34: both UNN and CNN anticodons can be used.

Second strategy—the "C34-sparing strategy": U34-containing tRNAs (T34 in the tDNA) read codons ending exclusively with A or A and G depending on the type of modification that occurs at this position during the tRNA maturation process (reviewed in Curran, 1998). The CNN anticodon-sparing strategy can apply in as many as 13 cases (indicated as small blue boxes in Fig. 6A and in the middle part of Fig. 6B). Using this C34-sparing strategy (decoding mode I) in addition to the A34- or G34-sparing strategy, the number of isoacceptor tRNAs can be reduced to $46 - 13 = 33$ (decoding mode II). Clearly, the C34-sparing strategy is rarely used in Eukarya and Archaea whereas it is widely used in Bacteria. In *S. cerevisiae* (E01), the three C34 tDNAs specific for Leu, Pro, and Ala are known to be missing (Percudani et al., 1997; Hani & Feldmann, 1998). Most remarkably, the hyperthermophilic archaeon *M. kandleri* (A01) is the only organism among the 50 studied that systematically uses this C34-sparing strategy (see the 13 blue boxes in the A01 column of Fig. 6A), despite its high G + C content (61%). As other thermophilic archaeons (A02 to A04) do not use this type of anticodon-sparing strategy, this demonstrates that this type of wobbling and the hyperthermophilicity are not antagonistic features.

There are three cases in which the C34-sparing strategy cannot apply (see middle genetic code array in Fig. 6B). First, the initiator and elongator tDNA-Met (CAT) have to read exclusively the ATG codon in all three domains. The same anticodon CAT is also used in Archaea and Bacteria by the special tDNA-Ile (CAT reading ATA) (see above A very peculiar tDNA and "Special boxes" in Fig. 6B). Second, the unique Trp anticodon CCA is never spared. Interestingly, in *M. genitalium* (B22), *M. pneumoniae* (B23), and *U. urealyticum* (B24), an additional tRNA-Trp using the UCA stop anticodon (*opal*) is present (Himmelreich et al., 1996; boxed, Fig. 6A, line (h)). Third, in some bacteria, the tDNA-Arg (CCG) is required to read the CCG codon to insure the decoding pattern of the Arg (NCG family box). This is achieved in concert with the tRNA-Arg (ACG) (in fact, inosine-34 in the mature tRNA) that reads the other three codons, ACG, GCG, and UCG, yet with different efficiencies (Curran, 1995). Among Eukarya, this special reading is occasionally present in the Arg (NCG family box) of *S. cerevisiae* (E01) (Percudani et al., 1997). In *M. capricolum*, this tRNA-Arg (CCG) is absent, and this observation correlates with the fact that codon CGG is an unassigned and untranslatable codon that is also not used for termination (Oba et al., 1991). It is interesting to note that chloroplasts of green plants use 31 anticodons to translate the 62 sense codons of their mRNAs. As a matter of fact, this low number of anticodons mainly results from a systematic adoption of the C34-sparing strategy (Andachi et al., 1989; Jukes & Osawa, 1990; Osawa et al., 1992).

It has recently been proposed that the decoding system of Eukarya has evolved through expansion of tRNA species (up to the maximum value of 46 tRNAs, or 45 if tDNA-Arg (TCG) is absent) as a consequence of a more restricted use of wobbling (use of the A34- or G34-sparing strategy only), rather than an augmented redundancy (Percudani, 2001). However, the parasitic lower eukaryote *E. cuniculi* (E06) as well as many Archaea (A02 to A09) and some Bacteria (B10, B16, B25) also have the same repertoire of 46 (or 45) tRNAs (Fig. 6B), for which U34-G3 wobbling is theoretically not necessary because C34-containing tRNAs are present. However, it might well be that U34-G3 wobbling is present in all Archaea and Bacteria. This could explain the ubiquitous existence of the special tRNA-Ile (CAU reading AUA) (see A very peculiar tDNA, above) instead of a tRNA-Ile (UAU) that could misread AUG Met codons.

Third strategy—the "A34- and G34- and C34-sparing strategy": Only Bacteria uses this ultimate sparing strategy in which U34-containing tRNAs, with U34 mostly unmodified, read all four synonymous codons of a given family box (this type of sparing is also known as the "two-out-of-three" (Lagerkvist, 1986) or "four-way wobbling" (Osawa et al., 1992), as the third base of the codon is irrelevant). The data in Figure 6A indicate that this third strategy applies in only six cases (red boxes in Fig. 6A and in the lower part of Fig. 6B) of which Val-NAC, Pro-NGG (Fig. 6A, line (g)), and Ala-NGC families are the most frequently encountered ones. A single case of this type of codon reading was found for Leu, Ser (NGA family box), and Thr. None was found for the Arg (NCG family box) nor for the Gly (NCC family box). However, two cases of A34- and G34- and C34-sparing strategies have been identified for tRNA-Gly in *M. capricolum* (Andachi et al., 1989) and in *Mycoplasma pulmonis* (Chambaud et al., 2001; not included in the set of 50 genomes examined in this work). If this third strategy is used in these seven cases (together with the two previous sparing strategies), then the number of anticodons requested to decode the whole genetic code is reduced to $33 - 7 = 26$ (decoding mode III), which is possibly the lowest theoretical number of tRNAs an organism can live with. The actual lowest number of tRNAs found among the 50 genomes examined is 30, in the case of *U. urealyticum* (B24). This organism uses the third strategy in six cases out of seven, but does not use the second one in some cases (there are two C34-containing tDNAs, tDNA-Leu (CAA) and tDNA-Lys (CTT)) and an extra tRNA-Trp (TCA) is present. Notice that *M. pulmonis* (Chambaud et al., 2001) uses only 29 tRNA species, but does not decode UGA as a Trp codon as does *U. urealyticum* (B24). The value of 26 tRNAs is very close to that identified from the sequenced genomes of mitochondria of different origins. For example, in human mitochondria, only 22 tDNA species are sufficient to decode mitochondrial mRNAs (Barrell et al., 1980; reviewed in Helm et al., 2000). In these organelles, a systematic decoding mode III (use of all possible sparing strategies exposed above) is used, very similar to the one used in *M. capricolum* (Andachi et al., 1989) and in *U. urealyticum* (B24). The reason why the number of anticodons can be lower than 26 in mitochondria is because: (1) there is generally no tRNA-Ile (CAT reading ATA), this codon being read instead as Met by a unique tRNA-Met harboring an anticodon CAU, which, in mitochondria of certain organisms, is f^5CAU (where f^5C is 5-formyl-cytidine), allowing efficient interaction with codon UAU (reviewed in Yokobori et al., 2001); (2) because this unique tRNA-Met (CAU) also works for the initiation of mitochondrial protein synthesis, there is no need for a particular tRNA-iMet (CAU); (3) in the special Arg box, the four CGM codons can be read by a single tRNA-Arg (anticodon UCG) (Yokobori et al., 2001); (4) in mitochondria of vertebrates, tRNA-Arg (anticodon UCU or CCU) are missing, and the corresponding unassigned codons AGA and AGG are used as stop codons (Osawa et al., 1989). Apart from these cases of mitochondrial anticodon economizing, variations in the genetic code assignments may also occur that do not alter the total number of mitochondrial tRNAs (Ohama et al., 1990b; Osawa et al., 1990; reviewed in Osawa et al., 1992; Yokobori et al., 2001). As a result, the minimal tRNA repertoire needed to read a three-letter code without import of additional tRNAs (see, e.g., Schneider & Marechal-Drouard, 2000; Dorner et al., 2001, and references therein) involves only one isoacceptor for each of the 20 amino acids, except for Leu, Ser, and eventually Arg, which each need two tRNAs, hence a total of 22 or 23 tRNAs. Only in a primordial world, when less than 20 amino acids were used, certainly less than 22 of 23 tRNAs species were used for autonomous protein synthesis (discussed, e.g., in Cedergren et al., 1986; Osawa & Jukes, 1988; Wong, 1988).

Are there other anticodon-sparing strategies? Among a cluster of phylogenetically related bacteria, the mean G + C content of genomic DNA can be very high (Muto & Osawa, 1987; Osawa et al., 1987). For example, in *Micrococcus luteus*, the genomic G + C content is 74% and the codon usage (based on the examination of 5,516 codons) is extremely biased towards use of G and C, especially at the third position (94%). In other words, both NNU and NNA codons are much less frequently used in the mRNA coding regions than the NNC and NNG codons. In particular, out of the 14 NNA sense codons (excluding UAA, UGA stop codons) 6 were reported to be totally absent, and the remaining 8 NNA codons were very seldom used (Ohama et al., 1990a). Moreover, sequencing the anticodon loop region of tRNAs isolated from *M. luteus* revealed that: (1) the six U34-containing anticodons corresponding to the six lacking NNA codons were also missing; (2) tRNAs with UNN (possibly *YNN in the mature tRNAs) corresponding to the other eight NNA codons were very scarce or nonexistent; (3) NNG and NNC codons were the most frequently used codons that are read by abundant tRNAs harboring CNN or GNN anticodons (except tRNA-Arg, for which an anticodon ICG was identified; Kano et al., 1991, 1993). Therefore, the decoding strategy used for translation of the mRNAs in this particular organism corresponds to a combination of the A34- or G34-sparing strategy (mode I; see Fig. 6B) with a novel systematic U34-sparing strategy (novel mode II bis; not shown in Fig. 6B) allowing them to minimize the use of U34-containing tRNAs instead of C34-containing tRNAs (as in the C34-sparing strategy). In none of the 50 genomes analyzed in this work (as well as in any of the remaining fully sequenced genomes publicly available through July 2002) was such a U34-sparing strategy revealed. Even in *M. kandleri* (A01) (G + C content 61%; see Fig. 3) or in *Halobacterium* sp. NRC-1 (A06) (G + C content 68%), the decoding mode was typically that of mode II (A34- or G34-sparing plus C34-sparing strategies; Fig. 6A, column (A01), and middle part of Fig. 6B). It would be of interest to analyze the situation in other organisms with high genomic G + C content, such as the thermophilic bacterium *Thermus thermophilus* HB8 (optimal growth at 82 °C; complete sequence not yet available) for which the genomic G + C content was shown to be 69% whereas that of the third position of the codons was shown to be 89% (Kagawa et al., 1984). In *T. thermophilus*, as in *M. luteus*, the major tRNA species are those carrying anticodons GNN or CNN (Hara-Yokoyama et al., 1986); thus, these two organisms could be examples of a U34-sparing strategy. A similar observation was made for *Halobacterium cutirubrum* (Gu et al., 1983). Therefore, although the U34-sparing strategy is apparently less frequently encountered than the three other types of anticodon sparings revealed above (probably because it results from high G + C content), the identification of other organisms using the U34-sparing strategy, especially in relation to the type of tRNA modification enzymes these organisms use to modify the nucleotides at positions 34 and 37 of their anticodon loops, should be helpful in better understanding how cells can fine-tune the translation of their mRNAs in the different phyla (for examples, see Weiss, 1973; Agris, 1991; Cermakian & Cedergren, 1995).

A nearly consistent and universal rule for the long V-arms of tDNAs-Leu, tDNAs-Ser, and bacterial tDNAs-Tyr

In the cloverleaf model, the region called the variable arm (V-arm) lies between the anticodon stem and the T-stem (positions 44 to 48, with position 47 being optionally occupied; see Fig. 2A, B). The length of the V-arm in each of the tDNAs is given in Figure 7 (blue

V-arm length			E01	E02	E03	E04	E05	E06	E07	A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	B01	B02	B03	B04	B05	B06	B07	B08	B09	B10	B11	B12	B13	B14	B15	B16	B17	B18	B19	B20	B21	B22	B23	B24	B25	B26	B27	B28	B29	B30	
		Genomes -->	*Sc*	*Sp*	*Ce*	*Dm*	*Hs*	*Eu*	*At*	*Mk*	*Pa*	*Pe*	*Ap*	*Af*	*Ha*	*Ss*	*St*	*Ta*	*Fa*	*Mb*	*Mj*	*Mt*	*Tp*	*Bb*	*Ct*	*Sy*	*An*	*Ll*	*Lm*	*Bs*	*Aa*	*Mt*	*Dr*	*Nm*	*Pr*	*Bu*	*Bh*	*Tm*	*Cj*	*Vc*	*Cp*	*Hp*	*Rs*	*Mg*	*Mp*	*Uu*	*Xf*	*Hi*	*Ec*	*Rp*	*Yp*	*Sm*	
AA	C	AC	**Eukarya (7)**							**Archaea (13)**													**Bacteria (30)**																														
F Phe	TTT		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
F Phe	TTC	(GAA)	5	5	5	5	$\boxed{6}$	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	(a)
L Leu	TTA	(TAA)	5	3	6	5	5	4	5	6	3	4	4	4	4	4	4	4	4	4	4	4	7	4	5	8	7	7	7	7	6	$\boxed{5}$	3	8	5	5	7	8	7	5	7	5	9	5	5	5	5	5	5	5	4	1	(b)
L Leu	TTG	(CAA)	3	3	5	4	5	3	5	-	4	4	4	4	4	4	4	4	4	4	-	-	4	5	5	3	5	5	5	3	6	$\boxed{5}$	1	3	5	-	4	6	5	3	5	5	3	1	1	1	5	4	3	-	3	3	
L Leu	CTT	(AAG)	-	1	3	3	3	3	1	-	-	-	-	-	-	-	-	-	4	-	-	-	-	-	-	-	-	7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
L Leu	CTC	(GAG)	3	-	-	-	-	-	-	6	4	4	4	4	4	4	4	4	-	4	4	5	3	5	3	2	2	-	5	6	6	8	1	4	4	4	6	5	4	2	5	5	3	7	5	-	3	4	5	5	5	3	
L Leu	CTA	(TAG)	3	1	3	1	X	2	1	6	4	4	4	4	4	4	4	4	4	4	4	4	2	2	3	2	3	3	3	4	5	5	4	3	3	3	4	5	3	3	2	3	5	4	4	3	2	3	3	3	3	3	
L Leu	CTG	(CAG)	-	-	5	4	X	2	1	-	4	4	4	4	4	4	4	4	4	4	-	-	5	-	5	5	3	5	-	5	5	5	3	5	5	-	-	5	-	4	-	-	3	-	-	-	3	-	5	-	5	5	
I Ile	ATT	(AAT)	5	5	5	5	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
I Ile	ATC	(GAT)	-	-	-	-	-	-	-	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
I Ile	ATA	(TAT)	5	5	$\boxed{6}$	5	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	(c)
i Ile	ATA	(CAT)	-	-	-	-	-	-	-	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
m iMet	ATG	(CAT)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
M Met	ATG	(CAT)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
V Val	GTT	(AAC)	5	5	5	5	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
V Val	GTC	(GAC)	-	-	-	-	-	-	-	5	5	5	5	5	5	5	5	5	5	5	5	5	5	-	5	5	5	-	5	5	5	5	5	5	5	5	5	5	5	5	-	5	5	-	-	-	5	5	5	-	5	5	
V Val	GTA	(TAC)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
V Val	GTG	(CAC)	5	5	5	5	5	5	5	-	5	5	5	5	5	5	5	5	5	5	5	5	5	-	-	-	-	-	-	-	-	5	5	-	-	-	-	5	-	-	-	-	5	-	-	-	5	-	-	-	-	-	
S Ser	TCT	(AGA)	5	4	4	4	4	4	4	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S Ser	TCC	(GGA)	-	-	-	-	-	-	-	7	5	6	7	5	5	5	5	6	5	6	6	6	9	8	8	8	8	9	0	0	0	7	8	9	8	6	1	9	6	6	8	6	9	9	9	-	1	8	6	6	8	8	
S Ser	TCA	(TGA)	4	4	4	4	4	4	4	7	5	6	7	5	4	6	5	4	4	$\boxed{5}$	6	6	6	8	6	6	6	9	9	9	1	8	6	9	8	6	9	7	6	9	8	6	8	0	9	9	6	8	6	8	6	8	(d)
S Ser	TCG	(CGA)	4	4	4	4	4	3	4	-	5	6	7	5	5	6	6	4	5	7	-	-	0	-	8	6	8	0	9	-	1	8	8	8	8	-	-	7	-	-	-	-	8	7	7	-	9	-	8	-	-	8	
P Pro	CCT	(AGG)	5	5	5	5	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
P Pro	CCC	(GGG)	-	-	-	-	-	-	-	4	4	5	5	4	4	5	5	4	4	4	4	4	5	-	5	5	5	-	-	-	5	5	5	5	5	-	-	5	-	5	-	5	5	-	-	-	5	-	5	-	5	5	
P Pro	CCA	(TGG)	5	5	5	5	5	5	5	4	4	5	5	4	4	5	5	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
P Pro	CCG	(CGG)	-	5	5	5	5	4	5	-	4	5	5	4	4	5	5	4	4	4	-	-	5	-	-	5	5	-	-	-	5	5	5	5	5	-	-	5	-	-	-	-	5	-	-	-	5	-	5	-	-	5	
T Thr	ACT	(AGT)	5	5	5	5	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
T Thr	ACC	(GGT)	-	-	-	-	-	-	-	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	-	5	5	5	5	5	5	
T Thr	ACA	(TGT)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
T Thr	ACG	(CGT)	5	5	5	5	5	4	5	-	5	5	5	5	5	5	5	5	5	5	-	5	5	-	5	5	5	5	5	-	5	5	5	5	5	-	-	5	-	-	-	-	5	5	5	-	5	-	5	5	-	5	
A Ala	GCT	(AGC)	5	5	5	5	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
A Ala	GCC	(GGC)	-	-	-	-	-	-	-	5	5	5	5	5	5	5	5	5	5	5	5	5	5	-	5	5	5	-	-	5	5	5	5	5	5	5	5	5	5	5	-	5	5	-	-	-	5	5	5	5	5	5	
A Ala	GCA	(TGC)	5	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
A Ala	GCG	(CGC)	-	5	5	5	5	4	5	-	5	5	5	5	5	5	5	5	5	5	-	-	5	-	-	5	5	-	-	-	-	5	5	-	-	-	-	5	-	-	-	-	5	-	-	-	5	-	-	-	-	5	
Y Tyr	TAT		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Y Tyr	TAC	(GTA)	5	5	5	5	X	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	3	3	4	3	4	2	2	3	6	2	4	2	3	3	3	5	4	3	4	3	4	5	5	2	4	3	3	4	3	3	
* Och	TAA	(TTA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
* Amb	TAG	(CTA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	
H His	CAT		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
H His	CAC	(GTG)	4	4	4	4	4	4	4	5	4	$\boxed{8}$	5	4	4	5	5	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	$\boxed{7}$	5	5	5	5	5	5	(e)
Q Gln	CAA	(TTG)	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	$\boxed{6}$	5	5	5	5	5	5	5	5	4	4	5	5	4	5	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	(f)
Q Gln	CAG	(CTG)	4	4	4	4	4	4	4	-	4	4	4	4	4	4	4	4	4	4	-	4	4	-	-	-	-	-	-	-	-	4	4	-	-	-	-	4	-	-	-	-	-	-	-	-	5	-	5	-	5	4	
N Asn	AAT		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
N Asn	AAC	(GTT)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
K Lys	AAA	(TTT)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
K Lys	AAG	(CTT)	5	5	5	5	5	5	5	-	5	5	5	5	5	4	5	5	5	5	-	-	5	5	-	-	5	5	5	-	5	5	5	-	-	-	-	5	-	-	5	-	5	5	5	5	5	5	-	-	5	5	
D Asp	GAT		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
D Asp	GAC	(GTC)	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	5	
E Glu	GAA	(TTC)	4	4	4	4	X	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	5	5	5	4	4	4	4	4	4	
E Glu	GAG	(CTC)	4	4	4	4	4	4	4	-	4	4	4	4	4	4	4	4	4	4	-	-	4	-	-	-	-	-	-	-	-	4	-	-	-	-	-	4	-	-	-	-	-	-	-	-	4	-	-	-	-	4	
C Cys	TGT		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
C Cys	TGC	(GCA)	5	5	5	5	5	5	5	4	5	5	$\boxed{6}$	5	5	$\boxed{6}$	$\boxed{6}$	5	X	5	5	5	5	5	4	4	5	4	4	4	5	4	4	4	4	4	4	5	4	4	5	5	4	4	4	5	4	4	4	5	4	4	(g)
* Opa	TGA	(TCA)	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	5	5	5	*	*	*	*	*	*	
W Trp	TGG	(CCA)	5	5	5	5	5	5	5	4	4	5	5	5	5	5	5	4	4	5	4	4	5	5	5	5	5	4	4	4	5	5	5	5	5	5	4	5	5	5	5	5	5	4	4	5	5	5	5	5	5	5	
R Arg	CGT	(ACG)	5	5	5	5	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5	5	5	5	5	5	5	5	5	5	5	5	5	-	-	5	5	-	5	-	-	5	5	5	5	5	5	5	
R Arg	CGC	(GCG)	-	-	-	-	-	-	-	4	5	5	5	4	4	5	5	4	4	4	$\boxed{7}$	4	5	5	-	-	-	-	-	-	-	-	-	-	-	-	-	5	5	-	-	5	-	5	5	-	-	-	-	-	-	-	(h)
R Arg	CGA	(TCG)	-	5	5	5	5	4	5	6	5	5	5	4	5	5	5	4	4	4	4	5	5	5	5	-	-	-	-	-	-	-	-	-	-	-	-	5	5	-	5	5	-	5	5	5	-	-	-	-	-	-	
R Arg	CGG	(CCG)	4	-	4	-	5	-	5	-	5	5	5	4	5	5	5	4	4	4	-	-	5	-	-	5	5	4	4	$\boxed{7}$	5	5	5	5	5	5	4	5	-	5	-	-	5	-	-	-	5	5	5	5	5	5	(i)
S Ser	AGT		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
S Ser	AGC	(GCT)	4	4	4	4	4	4	4	7	5	5	5	4	3	6	6	5	6	6	6	6	8	0	0	1	$\boxed{3}$	9	9	9	0	0	1	1	9	0	9	8	8	1	0	9	2	8	8	8	1	2	1	0	1	8	(j)
R Arg	AGA	(TCT)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	
R Arg	AGG	(CCT)	5	5	5	5	5	5	5	-	5	5	5	5	5	5	5	5	5	5	-	5	5	-	-	5	5	4	4	5	5	5	5	4	5	-	-	4	5	-	4	5	4	4	4	-	5	-	4	-	4	5	
G Gly	GGT		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
G Gly	GGC	(GCC)	4	4	4	4	4	4	4	4	4	5	5	4	4	5	5	4	4	4	4	4	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	4	4	5	5	5	5	5	5	
G Gly	GGA	(TCC)	4	4	4	4	4	4	4	4	4	5	5	4	4	5	5	4	4	4	4	4	4	4	4	4	4	4	4	4	5	4	4	4	4	4	4	5	5	4	4	5	4	4	4	4	4	4	4	4	4	4	
G Gly	GGG	(CCC)	4	4	$\boxed{5}$	-	4	4	4	-	4	5	5	4	4	5	5	4	4	4	-	-	-	-	-	4	5	-	-	-	5	4	5	-	4	-	-	5	-	-	-	-	4	-	-	-	4	-	4	-	4	5	(k)

FIGURE 7. Length of the V-arm of tRNA throughout genomes. This table reports the length of the extra arm in all tDNAs from all genomes (number of bases present between and including positions 44 and 48 according to the standard cloverleaf model; see Fig. 2). The presentation is similar to that of Figures 5A,B, and 6A. - indicates that no tRNA exists for a given anticodon and genome, and X indicates that a tDNA probably exists but has not been discovered. When several copies of a same tDNA have slightly different extra arm lengths, the longest one was retained. For the sake of compactness, lengths above 9 are shown in color: blue data should be read as augmented by +10 and those in red by +20. Gray background highlights blocks of isoacceptors harboring 4-base V-arms (with a few exceptions). Boxes denote unusual or remarkable values commented on in the text and referred to with the letters between parentheses at right.

color means +10 nt; red, +20 nt). Evidently, tDNAs with a 5-base V-arm are the most frequent, yet a few isoacceptors systematically harbor 4-base V-arms (highlighted with gray background in Fig. 7). In each of the three domains, tDNA-Leu and tDNA-Ser have generally a longer V-arm, 11 to 23 nt instead of 4 or 5 nt in the majority of the other tDNAs, the longest V-arm being found in the tDNA-Ser (GCT) of *Anabaena* 7120 (B05) (Fig. 7, line (j), boxed). The only exceptions are for the two copies of tDNAs-Leu (TAA) and the single copy of tDNA-Leu (CAA) of *Mycobacterium tuberculosis* (B10), and for tDNA-Ser (TGA) of *M. barkeri* (A11), which have an unusually short V-arm (only 5 bases; Fig. 7, lines (b) and (d), numbers boxed). The tDNA-Tyr also has a long V-arm (12 to 16 nt), however only in Bacteria. These three tDNA families with long V-arm belong to the so-called Class II tDNAs, whereas all the other tDNAs with a short V-arm belong to the Class I (Dirheimer et al., 1995). Remarkably, in *C. elegans* (E03), tDNA-Ile (TAT) (four copies, boxed, line (c)) and tDNA-Gly (CCC) (three copies, boxed, line (k)) harbor, respectively, 16- and 15-bases-long V-arms (three copies of a distinct tDNA-Ile (TAT) harbor a regular short V-arm). It is noteworthy that these two tDNAs with long V-arms also harbor a G13:A22 mismatch, characteristic of Class II tDNAs. A few Class I tDNAs with V-arms of 6, 7, or 8 bases are also found (boxed, Fig. 7, lines (a, e, f, g, h, and i)). The importance of the V-arm was demonstrated only in the case of eukaryotic tRNA-Ser (Cusack et al., 1996) and archaeal tRNA-Leu (Soma et al., 1999). Indeed, the V-arms of these tRNAs contain essential determinants for the recognition by their respective synthetases (reviewed in Giegé et al., 1998). It has been suggested that tRNA with a large V-arm arose from those with a small V-arm by retaining a splicing-deficient intron (Kjems et al., 1989). In summary, most tRNAs have short V-arms (4 or 5 bases) although lengths up to 8 are found occasionally. All tDNAs-Leu, tDNAs-Ser, and bacterial tDNAs-Tyr have long arms (up to 23 bases), long V-arms being found occasionally outside the set tDNAs-Leu, tDNAs-Ser, and bacterial tDNAs-Tyr.

Length, localization, and distribution of introns in eukaryotic and archaeal tDNA: No clear rule

Introns in nuclear tDNA were first discovered in yeast, after comparison of the tDNA sequence with the sequence of the corresponding mature gene product (Goodman et al., 1977; Valenzuela et al., 1978). They were subsequently found in the tDNAs of many other eukaryotic organisms, always located at the same position, between nt 37 and 38 of the anticodon loop (location indicated by black arrows in Fig. 4). More recently, introns have also been found in several archaeal pre-tRNAs at the same location (Daniels et al., 1985). Surprisingly, sequencing of archaeal genomes has also revealed some unusually located introns (blue arrows in Fig. 4), as well as tDNAs harboring two introns. Nowadays, the presence of introns in tDNA at positions 37/38 only are easily discovered by a computer search (Lowe & Eddy, 1997; this work). Detecting novel introns as in archaeal tDNAs is a difficult task because of our ignorance about their length as well as their exact location within the tDNA sequence. An uncomplete set of tDNAs within a given genome (required for reading all codons; see Fig. 6A) is generally an

FIGURE 8. Compilation of introns in tRNA genes from Archaea and Eukarya. The introns present in the tRNA genes of Eukarya (top) and Archaea (bottom) are sorted out according to the amino acid encoded. Columns indicate the genome number (e.g. E01) and the genome abbreviated name (e.g., *Sc*; see Table 3 for full names); the number of introns located at position 37 followed, for Archaea only, by the number (in bold blue) of introns located elsewhere; and each of the 21 isoacceptors (initiator and elongator tDNAs-Met are distinguished). Each element of the table indicates the length of the intron and below the anticodon (as found in the tRNA gene, e.g. TAT, instead of UAU in the tRNA). - indicates that none of the tRNA genes, among those coding for a given amino acid, have an intron. In Eukarya, the different copies (genes) of the same isoacceptor tDNA may have a slightly different intron length and these are all indicated (e.g., 18/19). In Archaea, the intron is commonly located after the base 3′ to the anticodon, as in Eukarya (between positions 37 and 38; intron length written in black). The lengths of unusually located introns are written in bold blue as "xxpyy," where "xx" is the intron length and "yy" the position of the nucleotide preceding the intron (e.g., 22p58 means a 22 bp intron between positions 58 and 59). The three proposed novel introns at positions 3/4 in *P. aerophilum* (A03) are shown in red. The bulge-helix-bulge (BHB) motif was found in all introns (only G-C, A-T, and G-T pairings, in either orientation, were accepted; a C:A mismatch was often observed, whereas C:T or T:T mismatches rarely occurred). Boxes signal the six cases of two introns in the same archaeal tDNA. Horizontal light gray stripes emphasize *S. pombe* (E02) and *P. aerophilum* (A03), which are the eukaryote and archaeon with the largest number of intron-bearing tDNAs. Darker gray vertical stripes are used to emphasize amino acids/anticodons for which an intron is present in the tDNA of nearly all genomes of a biological domain: tDNA-Ile (TAT) and tDNA-Leu (various anticodons) in nearly all Eukarya; tDNA-Met (CAT) (with the exception of *M. kandleri* (A01) and *A. fulgidus* (A05)) and tDNA-Trp (CCT) in all archaea (with the exception of *P. aerophilum* (A03)). * indicates tDNA-Trp introns harboring methylation guides (d'Orval et al., 2001). Notes (indicated in red; other notes are referred to in text): (b) This tDNA-Ile (TAT) of *C. elegans* tDNA is present in three copies and another markedly different tDNA-Ile (TAT) is present in four copies and has no intron. (c) No tDNA-Tyr has been discovered in the fraction of human genome examined (about 1/7 of the complete genome, GtRDB reports 15 to 21 bp introns for human tDNA-Tyr). (d) Two copies of the tDNA-Glu (TTC) (differing by a single base change) that bear the same two introns (with slight sequence variations) are present in *M. kandleri* (A01); all other archaeal intron-bearing tDNAs are present as a single copy.

Amino Acid			A Ala	C Cys	D Asp	E Glu	F Phe	G Gly	H His	I Ile	K Lys	L Leu	M iMet	M Met	N Asn	P Pro	Q Gln	R Arg	S Ser	T Thr	V Val	W Trp	Y Tyr
Intron Nb		↓	Length of intron at position 37/38																				
E01	*Sc*	10	–	–	–	–	18/19 GAA	–	–	60 TAT (a)	23 TTT	19 TAG 32/33 CAA	–	–	–	30/31 /33 TGG	–	–	19 GCT 19 TCG	–	–	34 CCA	14 GTA
E02	*Sp*	14	11 CGC	–	–	–	–	–	–	25 TAT	8 CTT	18 TAG 16 TAA 19 CAA	–	7/9 CAT	–	24 CGG	–	30 CCT	15 TGA 16 CGA 11 GCT	–	8/9 AAC	–	10 GTA
E03	*Ce*	3	–	–	–	–	–	–	–	10/11 TAT (b)	–	35/36 /38 CAA	–	–	–	–	–	–	–	–	–	–	11 GTA
E04	*Dm*	3	–	–	–	–	–	–	–	24/26 TAT	–	38/40 /42/44 CAA	–	–	–	–	–	–	–	–	–	–	20/21 /35/38 GTA
E05	*Hs*	3	–	–	–	–	–	–	–	20 TAT	–	22/23 /24 CAA	–	–	–	–	–	12/14 /18 TCT	–	–	–	–	–(c)
E06	*Eu*	2	–	–	–	–	–	–	–	41 TAT	–	–	–	–	–	–	–	–	–	–	–	–	12 GTA
E07	*At*	2	–	–	–	–	–	–	–	–	–	–	–	11/12 CAT	–	–	–	–	–	–	–	–	11/12 /13/16 GTA
A01	*Mk*	7+1	–	21 GCA	–	33p21 21 TTC (d)	32 GAA	–	–	–	–	–	–	36 CAT	27 GTT	15 GGG	–	–	–	–	–	76* CCA	–
A02	*Pa*	2	–	–	–	–	–	–	–	–	–	–	–	31 CAT	–	–	–	–	–	–	–	71* CCA	–
A03	*Pe*	7+19	20 TGC	22p58 GCA	21p3 GTC	21p3 21p58 TTC 21p3 CTC	–	15 TCC	13 GTC	20p29 CAT	–	–	17p38 CAT	20p29 CAT	13 GTT	15 TGG 16 GGG	17p21 CTG	24p30 CCT 16p56 GCG 24p30 CCG	–	20p29 20p59 TGT 19p38 16p56 GGT 20p29 CGT	25 TAC 20p29 GAC	–	42p39 GTA
A04	*Ap*	9+5	–	18 GCA	121 GTC (e)	–	–	–	–	38 CAT	36p45 TTT 36p45 CTT (f)	–	–	48 CAT	–	37 GGG 44p32 CGG	–	44 TCT	36 CGA	49 CGT 19p22 TGT	–	37p30 CCA	39 GTA
A05	*Af*	5	–	–	35 GTC	16 TTC	–	–	–	–	–	15 CAA	–	–	–	–	–	–	–	–	–	62* CCA	24 GTA
A06	*Ha*	3	–	–	–	–	–	–	–	33 CAT	–	–	–	84 CAT	–	–	–	–	–	–	–	102* CCA	–
A07	*Ss*	17+3	–	25p28 14 GCA	–	15p20b CTC 15p20b TTC	–	–	–	12 GAT	23 TTT 22 CTT	15 TTA 16 CAA	25 CAT	17 CAT	14 GTT	21 GGG	–	15 TCT 13 CCT	24 CGA	15 TGT 13 CGT	–	65* CCA	13 GTA
A08	*St*	20+4	–	14 GCA	–	16p20b CTC 16p20b TTC	17 GAA	–	–	16 GAT 19 CAT	25 TTT 27 CTT	16 TAG 16 CAG 13 CAA 19p30 GAG	24p38 CAT	12 CAT	–	18 GGG	–	16 TCT 24 CCT 13 GCG	11 TGA 26 CGA	17 TGT 24 CGT	–	57* CCA	19 GTA
A09	*Ta*	4	–	–	–	–	–	12 GGG	–	–	–	–	–	14 CAT	–	–	–	–	–	–	–	69* CCA	25 GTA
A10	*Fa*	3	–	–	–	–	–	–	–	–	–	–	–	16 CAT	–	–	–	–	–	–	–	69* CCA	27 GTA
A11	*Mb*	4	–	–	–	–	–	–	–	–	–	–	–	37 CAT	–	–	–	20 TCG	–	–	–	118* CCA	44 GTA
A12	*Mj*	2	–	–	–	–	–	–	–	–	–	–	–	35 CAT	–	–	–	–	–	–	–	33 CCA	–
A13	*Mt*	3+1	–	–	–	–	–	–	–	–	–	–	–	35 CAT	–	16 32p32 GGG	–	–	–	–	–	68* CCA	–
Intron Nb		↑	Length of intron at position 37/38 or elsewhere																				
Amino Acid			A Ala	C Cys	D Asp	E Glu	F Phe	G Gly	H His	I Ile	K Lys	L Leu	M iMet	M Met	N Asn	P Pro	Q Gln	R Arg	S Ser	T Thr	V Val	W Trp	Y Tyr

FIGURE 8. *See caption on facing page.*

indication for missing tDNA(s) containing one or two unusually located introns. Such introns have yet to be identified by visual examination, once a new potential tDNA harboring the expected anticodon is located.

All intron-containing tDNA present in the 20 eukaryotic and archaeal genomes analyzed in this work are compiled in Figure 8. The sizes of these introns range from 7 bases for the shortest intron in tDNA-Met (CAT) of *S. pombe* (E02) to 121 for the longest in tDNA-Asp (GTC) of *A. pernix* (A04). No general rule is evident for the occurrence of an intron within a given isoacceptor tDNA in the eukaryotic or archaeal genomes analyzed in this work. In the eukaryotic isoacceptors specific for Ile (TAT) and Leu (TAG, AAG, CAA, TAA), introns are frequently found. In Archaea, tDNA-Met (CAT) and tDNA-Trp (CCA) harbor introns in all cases but one (Fig. 8). The most "intron-rich" species are *S. pombe* (E02) in Eukarya (14 isoacceptors with intron) and *P. aerophilum* (A03) in Archaea (26 introns in 23 isoacceptors—indicated by horizontal light gray stripes in Fig. 8). The number of tDNAs harboring an intron appears to be higher in lower eukaryotes (like *S. cerevisiae* (E01) or *S. pombe* (E02)) than in higher eukaryotes. In Eukarya, the longest intron is found in each of the two tDNA-Ile (TAT) of *S. cerevisiae* (60 bp, probably related by a chromosomal duplication (Fig. 8, note (a); Seoighe & Wolfe, 1999). In Archaea, the longest known intron (121 bases) belongs to tDNA-Asp (GTC) of *A. pernix* (A04) (Fig. 8, note (e); Kawarabayasi et al., 1999).

The exact location of archaeal introns at positions other than 37/38 was verified by looking for the presence of a consensus bulge-helix-bulge (BHB) motif (Lykke-Andersen et al., 1997). In many cases, introns were found at slightly or significantly different locations than those initially reported. For example, in *A. pernix* (A04), the introns of tDNAs-Lys (TTT) and (CTT) are located in the V-arm (at positions 45/46) and not at 37/38 as previously reported (Kawarabayasi et al., 1999; Fig. 8, see note (f)in the figure legend). The presence of an intron in the V-arm has also been reported for the tDNA-Gly (CCC) of the the hyperthermophilic archaeon *Thermofilum pendens* (Kjems et al., 1989). The recent sequencing of *P. aerophilum* (A03) has revealed several introns at novel, but unspecified positions (Fitz-Gibbon et al., 2002). We found that five of these introns are located in the anticodon stem at position 29/30 and four at various positions in the T-loop (Figs. 4 and 8). Moreover, taking into account the presence of a consensus BHB motif (Lykke-Andersen et al., 1997), we also found that an intron is located in the acceptor stem at positions 3/4 of tDNA-Asp (GTC), tDNA-Glu (TTC), and tDNA-Glu(CTC) (red arrow in Fig. 4; details will be reported elsewhere).

In Eukarya, the tRNA splicing machinery recognizes the three-dimensional structure of the intron-containing pre-tRNA and cleaves exclusively at position 37/38 (reviewed in Westaway & Abelson, 1995; Belfort & Weiner, 1997; Lykke-Andersen et al., 1997). In Archaea, the splicing machinery does not require the three-dimensional structure of the pre-tRNA and cleaves introns at variable positions in pre-tRNAs within a simple motif, such as a BHB (Arn & Abelson, 1996; Trotta et al., 1997; Li et al., 1998). Recently, splicing of introns exhibiting BHB motifs in pre-mRNAs (Watanabe et al., 2002) as well as in pre-rRNA (Tang et al., 2002) has been reported in Archaea, suggesting that a unique archaeal splicing machinery handles all kinds of RNA introns.

The evolutionary origin as well as the role of the introns in eukaryotic and archaeal tDNA are still debated issues. In a few cases, the intron has been shown to play an essential role in the correct modification of the pre-tRNA by the so-called intron-dependent tRNA modification enzymes during early steps of tRNA maturation (reviewed in Grosjean et al., 1997; see also Jiang et al., 1997; Motorin & Grosjean, 1999). In the particular case of archaeal tDNA-Trp (CCA), it has been pointed out that the long intron (62 to 118 bases in length) often harbors a duplicate form of the characteristic C + D boxes that were previously found in all snoRNAs that guide the 2′-*O*-methylations in eukaryotic rRNA (Gaspin et al., 2000; also reviewed in Dennis et al., 2001). In the case of *Halobacterium volcanii* tRNA-Trp, these snoRNA-like sequences guide the methylation in *cis* of the 2′-hydroxylribose at positions 34 and 39, as demonstrated in vitro (d'Orval et al., 2001; C. Daniels, pers. comm.). However, this might not be a general rule, as two archaeons (*A. pernix* (A04) and *Methanococcus jannaschii* (A12)) do not posses such a snoRNA-like motif in the intron of their tDNA-Trp. In these organisms, it is, however, not known whether C34 and/or C39 in mature tRNA-Trp is/are methylated. Also, *P. aerophilum* (A03) is the only archaeon examined in this work that does not contain an intron in its tDNA-Trp (Fig. 8). No C + D boxes are found in other intron-containing tDNAs, not even in tDNA-Asp (GCT) of *A. pernix* (A04) harboring a 121-base-long intron. Noteworthy, not all tDNAs coding for a mature tRNA harboring 2′-*O*-methylribose at position 34, such as archaeal tRNA-iMet (CAT), contain a long intron with characteristic C + D boxes. In these cases, it may well be that the 2′-*O*-methylation reactions are guided by specific small RNAs acting in *trans* (reviewed in Dennis et al., 2001; Omer et al., 2002) or by "protein-only" enzymes acting on intron-containing tRNAs as those founded in *S. cerevisiae* and *S. pombe* (reviewed in Grosjean et al., 1997; see also Pintard et al., 2002). Alternatively, an intron may also play a role in the stepwise folding of the tRNA molecule by avoiding its misfolding during maturation (Dennis et al., 2001).

In bacterial genomes, self-splicing (autocatalytic) introns of group I were identified in the anticodon loops of a very limited number of pre-tRNAs from a few cyanobacteria such as *Synechocystis* 6803 (B04) (tDNA-iMet (CAT)) and *Anabaena* 7210 (B05) (tDNA-Leu (TAA)) (green arrows in Fig. 4) and from the purple bacteria *Agrobacterium tumefaciens* and *Azoarcu* sp. strain BH72 (Edgell et al., 2000). tRNA introns in cyanobacterial tRNA are believed to have moved through lateral transfer; indeed, the ORF inserted in the intron of tRNA-iMet (CAT) of *Synechocistis* 6803 (B04) encodes a double-strand homing endonuclease (Bonocora & Shub, 2001).

Distribution of tRNA genes in genomes and peculiarities of their transcription in Eukarya

The distribution of tDNA in the genomes and their mode of transcription are two intimately linked issues. Two types of gene distribution are encountered: closely packed arrays of codirectional genes transcribed into multiple pre-tRNA (predominant organization in Bacteria) or isolated genes transcribed separately (usual organization in Eukarya). Based on the sole criteria of intergenic distances, nearly all tDNAs are arranged in arrays in some bacteria as in *B. subtilis* (B08) or *V. cholerae* (B18), whereas in other bacteria such as *Anabaena* 7120 (B05), most tDNAs are scattered throughout the genome (data not shown). In Archaea, variable patterns were also observed, and extreme packing of genes occurs with intergenic distances even shorter than in Bacteria. For example, in *Sulfolobus tokodaii* (A08), base 73 of tDNA-Pro (CGG) and base 1 of tDNA-Ala (CGC) are separated by a single base, and so are tDNA-Asn (GTT) and tDNA-Met (CAT) of *T. acidophilum* (A09). Short intergenic distances are seldom encountered in Eukarya. As an example, in *S. pombe* (E02), four copies of tDNA-iMet (CAT) are located 5 or 7 bases 3′ to tDNA specific for Ser or Asp.

To transcribe their tRNA genes, prokaryotic cells (Bacteria and Archaea) use regular upstream promoters (Palmer & Daniels, 1995) and a single RNA polymerase that is in charge of all genes in the cell. In contrast, Eukarya have developed three specialized RNA polymerases, of which RNA polymerase III (Pol III) transcribes most of the RNA genes, including all tRNA genes (reviewed in Sprague, 1995). To recognize their tRNA genes, Eukarya use a specific transcription factor, TFIIIC, which, unlike any other transcription factor, recognizes promoter sequences internal to the RNA genes instead of upstream sequences as in prokaryotes. Once TFIIIC is bound, a second factor, TFIIIB is recruited on upstream DNA to which Pol III docks in a third step (for a review, see Paule & White, 2000, and references therein). The promoter elements that TFIIIC recognizes in the tDNA are indeed the conserved bases in and around the D- and T-loops that are common to all eukaryotic tDNA (Galli et al., 1981; Hamada et al., 2001). In the D-stem and D-loop, these bases are: T8, Y11 (or B11 in initiators), A14, G18, G19, A21 (R21 in *S. pombe* (E02)), and R24 (S24 in initiators; in lines Ee and Ei and box “A-box” of Fig. 3). Additionally, all eukaryotic tDNAs have no base at position 17a such that, whatever the base occupancy at position 17, one of the Gs located at positions 18 or 19 is always found 4 bases apart from A14, as shown in the sequences entitled “cs1,” where position 17 is occupied, and “cs2,” where position 17 is empty (box “A box” in Fig. 3). Therefore, the more conserved sequences in D-loop/A Box of Eukarya compared to Archaea and Bacteria most probably result from a functional pressure exerted by the Pol III machinery. It has been recently proposed that the difference in A box consensus sequences among Eukarya (e.g., G10 is conserved in *S. cerevisiae* (E01); Fig. 3) might, in fact, reflect the alternate use of upstream helper TATA promoter elements such as in *S. pombe* (E02) and other eukaryotes (Hamada et al., 2001). To the contrary, T-loop/B box sequences do not appear significantly more constrained in Eukarya compared to Archaea and Bacteria (Fig. 3). Taken together, these data favor the hypothesis that, during evolution, the emerging eukaryotic Pol III machinery has exploited the D-loop/A box and T-loop/B box conservations in different ways. The B box sequence conservation, because of its strong tRNA functional constraint, was used “as it was.” On the contrary, among different possible patterns in the D-loop, the A box sequences retained were those compatible with both tRNA functional constraints and sequence recognition by transcription factor TFIIIC.

Finally, in Eukarya, a higher level of organization of tDNA exists that is probably related to a novel role discovered for the Pol III transcription machinery. As a matter of fact, tDNAs are often found clustered in centromeres (yet with much larger intergenic distances than found in Archaea or Bacteria) as in *S. pombe* (E02) (Wood et al., 2002) and *C. elegans* (E03) (*C. elegans* Sequencing Consortium, 1998). Such clustering is not found in *S. cerevisiae*, in which the tDNAs are apparently scattered, although 30% are related by chromosomal duplications (Seoighe & Wolfe, 1999). Nevertheless, in *S. cerevisiae*, it has been shown that a functional tRNA gene acts as a barrier to the spreading of the heterochromatin structure (Donze & Kamakaka, 2001). As suggested by earlier results on TFIIIC-chromatin interaction (Burnol et al., 1993), this confirms that, in Eukarya, the Pol III transcription machinery also has an important function in the maintenance of chromatin.

CONCLUSION

In this work, we took advantage of the availability of all tDNAs from 50 selected genomes (out of 84 available) representative of the three domains of life in order to better highlight the relevant features of tRNA genes not only throughout different phylogenetically related genomes but also within a given genome. Using the sole criterion of the cloverleaf model, we have extracted over 4,000 nuclear, nonorganellar tRNA genes. With the exception of the yet incomplete human genome, coherent tRNA repertoires (anticodons used and number of genes for each isoacceptor) coding for the 20 common natural amino acids were identified for each genome, thus excluding tmRNAs and tRNAs specific for selenocysteine and pyrrolysine. This allowed us to propose three major anticodon-sparing strategies that are used differently in the three domains of life. Our detailed analysis not only reinforces the universality of the cloverleaf structure, but also confirms or extends several known peculiarities in the tDNAs of the three domain of life: (1) the consensus A and B box promoter sequences used by the eukaryal transcription machinery; (2) the high sequence conservation of tDNA-iMet throughout the three domains; (3) the ubiquitous existence of the special tDNA-Ile (CAT reading ATA) in Archaea and Bacteria; (4) the quasi general encoding of G(−1) in tDNA-His of Archaea and Bacteria; and (5) the universal presence of long V-arms in tDNAs-Leu, tDNAs-Ser, and bacterial tDNAs-Tyr. Special attention was given to the introns of archaeal tDNAs, for which novel locations were discovered in the acceptor stem and V-arm. All tDNA features examined are listed in Figure 9 as a function of the three domains of life. Remarkably, Archaea sequester sometimes either with Eukarya or with Bacteria depending on the tDNA feature considered (highlighted with darker gray background). No property common to Eukarya and Bacteria but not shared with Archaea was found. Thus, from an evolutionary point of view, tRNAs from Archaea appear as an intermediate domain between Eukarya and Bacteria.

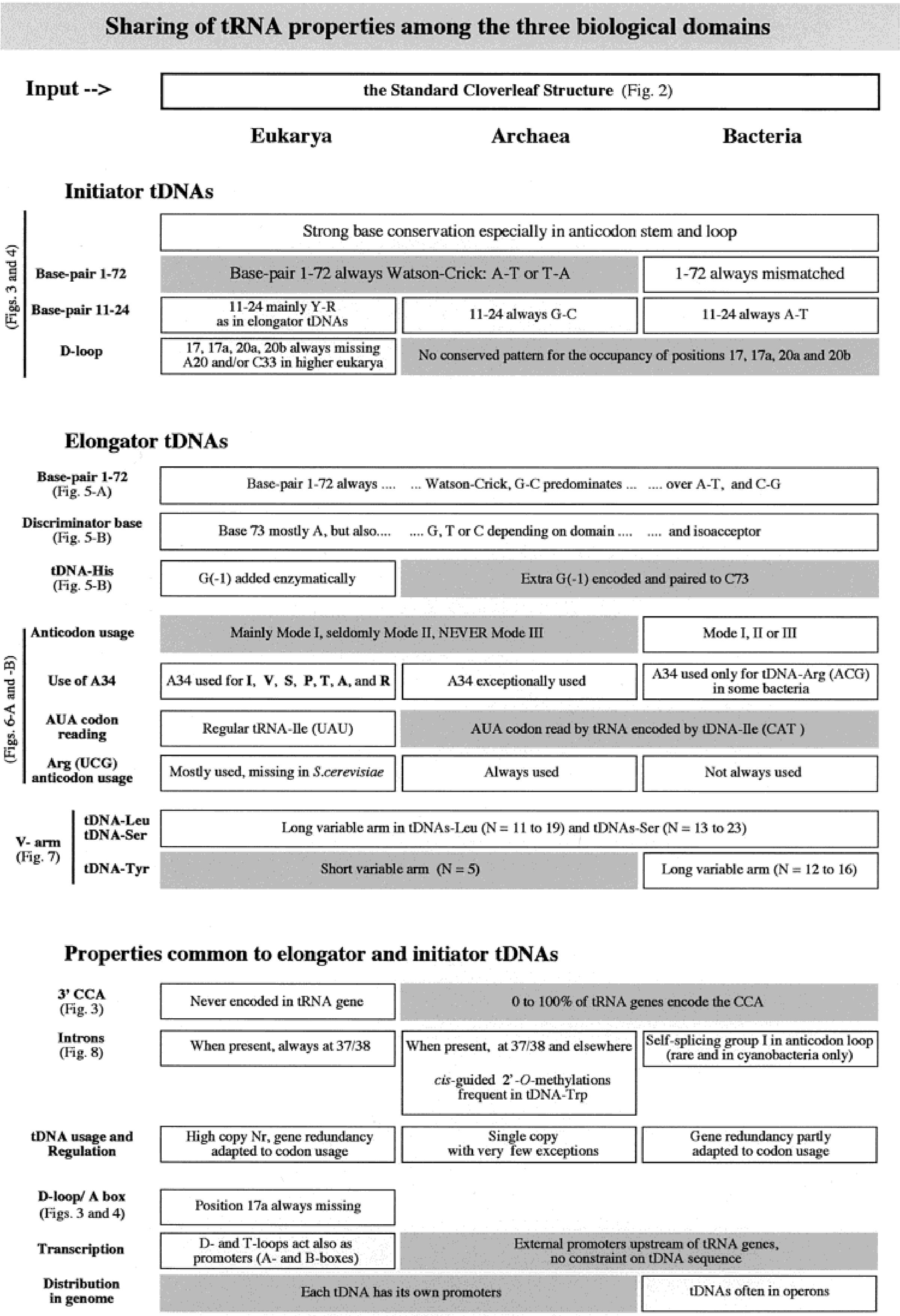
The data presented in this work also provide a benchmark for the nuclear tDNA sequences check of genomes to be sequenced. As new genomes become available, new peculiarities may be revealed. A good example is the exception recently discovered in the genome of *M. kandleri* (A01) concerning the unexpected presence of C instead of the quasi universally conserved T at position 8 in 30 tDNAs out of 34. Among future analyses, it would be interesting to compare the "tRNomics" of more numerous eukaryotic and archaeal genomes in order to find out subdomain-specific features that may exist among unicellular versus multicellular eukaryotes and among crenarchaeota versus euryarchaeota. Moreover, selected tDNA sequence specificities or structural features of the tRNA molecules could be used to reveal unexpected links between organisms, new phylogenetic characters, or more frequent than anticipated lateral genetic transfers. Finally, more systematic "two-dimensional analysis" of the tDNA sequences, as the ones presented in Figures 5A,B, 6A, and 7, should allow us to extend to other organisms the experimentally verified identity elements of a given tRNA that are important for the specific recognition by the cognate aminoacyl-tRNA synthetases, tRNA-modifying enzymes as well as any enzyme or factor interacting with it.

METHODS

Table 3 lists the 50 genomes examined in this work. A critical step, common to all search algorithms, is testing for false positives. As a first constraint, we decided to adopt the standard cloverleaf model (Holley et al., 1965; Sprinzl et al., 1998). In doing so, we accepted our losses, as a single base deletion or insertion (due, e.g., to a sequencing error) may cost the loss of one tDNA. Noncanonical sequences such as those of the selenocysteine tRNA (Hubert et al., 1998) and the tmRNA (Williams, 2002) would also be lost. The different features of the tRNA cloverleaf model were searched for by using the order used in previous software (Fichant & Burks, 1991; Pavesi et al., 1994; el-Mabrouk & Lisacek, 1996; Lowe & Eddy, 1997): first looking at the most conservative T-loop, then the T-stem, the acceptor stem, the D-stem, and the D-loop. The anticodon stem and a possible intron at position 37/38 (for eukaryotic and archaeal sequences only) as well as the variable arm were sought for as final steps. The weight matrix presented by Pavesi et al. (1994) was also used. We found that limiting the number of wobble G-T and mismatched base pairs in each of the four stems was able to efficiently eliminate false positives. The final set of all constraints used in this work are presented in Table 4.

The cloverleaf tDNA search procedure was first set up and tested using the *S. cerevisiae* (E01) genome (Blandin et al., 2000). Constraints were adjusted so that the 274 already identified tDNAs (el-Mabrouk & Lisacek, 1996; Hani & Feldmann, 1998) were correctly predicted (Fig. 2A). To easily distinguish slight se-

FIGURE 9. Sharing of tRNA properties among the three biological domains. This figure summarizes different properties (listed at left) of the cloverleaf tRNA molecule presented or discussed in this work. Properties applying to each of the three domains are listed under the headings "Eukarya", "Archaea," and "Bacteria." Horizontal boxes are used to group two or the three domains for a given property, if applicable (for clarity, some exceptions are not indicated—see text). Darker gray background is used to emphasize properties shared by Archaea with either Eukarya or Bacteria. Notice that no property common to Eukarya and Bacteria but not shared with Archaea has been found. More detailed information is presented in the Figures referenced at left.



Sharing of tRNA properties among the three biological domains

Input --> **the Standard Cloverleaf Structure** (Fig. 2)

Initiator tDNAs

<table>
<tr><th></th><th>Eukarya</th><th>Archaea</th><th>Bacteria</th></tr>
<tr><td>(Figs. 3 and 4)</td><td colspan="3">Strong base conservation especially in anticodon stem and loop</td></tr>
<tr><td><b>Base-pair 1-72</b></td><td colspan="2">Base-pair 1-72 always Watson-Crick: A-T or T-A</td><td>1-72 always mismatched</td></tr>
<tr><td><b>Base-pair 11-24</b></td><td>11-24 mainly Y-R
as in elongator tDNAs</td><td>11-24 always G-C</td><td>11-24 always A-T</td></tr>
<tr><td><b>D-loop</b></td><td>17, 17a, 20a, 20b always missing
A20 and/or C33 in higher eukarya</td><td colspan="2">No conserved pattern for the occupancy of positions 17, 17a, 20a and 20b</td></tr>
</table>

Elongator tDNAs

<table>
<tr><th></th><th>Eukarya</th><th>Archaea</th><th>Bacteria</th></tr>
<tr><td><b>Base-pair 1-72</b>
(Fig. 5-A)</td><td colspan="3">Base-pair 1-72 always .... Watson-Crick, G-C predominates ... . over A-T, and C-G</td></tr>
<tr><td><b>Discriminator base</b>
(Fig. 5-B)</td><td colspan="3">Base 73 mostly A, but also.... .... G, T or C depending on domain .... and isoacceptor</td></tr>
<tr><td><b>tDNA-His</b>
(Fig. 5-B)</td><td>G(-1) added enzymatically</td><td colspan="2">Extra G(-1) encoded and paired to C73</td></tr>
<tr><td>(Figs. 6-A and -B) <b>Anticodon usage</b></td><td colspan="2">Mainly Mode I, seldomly Mode II, NEVER Mode III</td><td>Mode I, II or III</td></tr>
<tr><td><b>Use of A34</b></td><td>A34 used for <b>I</b>, <b>V</b>, <b>S</b>, <b>P</b>, <b>T</b>, <b>A</b>, and <b>R</b></td><td>A34 exceptionally used</td><td>A34 used only for tDNA-Arg (ACG)
in some bacteria</td></tr>
<tr><td><b>AUA codon reading</b></td><td>Regular tRNA-Ile (UAU)</td><td colspan="2">AUA codon read by tRNA encoded by tDNA-Ile (CAT)</td></tr>
<tr><td><b>Arg (UCG) anticodon usage</b></td><td>Mostly used, missing in <i>S.cerevisiae</i></td><td>Always used</td><td>Not always used</td></tr>
<tr><td><b>V- arm</b> (Fig. 7) <b>tDNA-Leu</b>
<b>tDNA-Ser</b></td><td colspan="3">Long variable arm in tDNAs-Leu (N = 11 to 19) and tDNAs-Ser (N = 13 to 23)</td></tr>
<tr><td><b>tDNA-Tyr</b></td><td colspan="2">Short variable arm (N = 5)</td><td>Long variable arm (N = 12 to 16)</td></tr>
</table>

Properties common to elongator and initiator tDNAs

<table>
<tr><th></th><th>Eukarya</th><th>Archaea</th><th>Bacteria</th></tr>
<tr><td><b>3' CCA</b>
(Fig. 3)</td><td>Never encoded in tRNA gene</td><td colspan="2">0 to 100% of tRNA genes encode the CCA</td></tr>
<tr><td><b>Introns</b>
(Fig. 8)</td><td>When present, always at 37/38</td><td>When present, at 37/38 and elsewhere
<i>cis</i>-guided 2'-<i>O</i>-methylations
frequent in tDNA-Trp</td><td>Self-splicing group I in anticodon loop
(rare and in cyanobacteria only)</td></tr>
<tr><td><b>tDNA usage and Regulation</b></td><td>High copy Nr, gene redundancy
adapted to codon usage</td><td>Single copy
with very few exceptions</td><td>Gene redundancy partly
adapted to codon usage</td></tr>
<tr><td><b>D-loop/ A box</b>
(Figs. 3 and 4)</td><td>Position 17a always missing</td><td></td><td></td></tr>
<tr><td><b>Transcription</b></td><td>D- and T-loops act also as
promoters (A- and B-boxes)</td><td colspan="2">External promoters upstream of tRNA genes,
no constraint on tDNA sequence</td></tr>
<tr><td><b>Distribution in genome</b></td><td colspan="2">Each tDNA has its own promoters</td><td>tDNAs often in operons</td></tr>
</table>

FIGURE 9. *See caption on facing page.*

TABLE 3. List of genomes analyzed.[a]

Code	Short name	Name and strain	Reference
		Eukarya	
E01	*Sc*	*Saccharomyces cerevisiae*	Anonymous, 1997
E02	*Sp*	*Schizosaccharomyces pombe*	Wood et al., 2002
E03	*Ce*	*Caenorhabditis elegans*	*C. elegans* Sequencing Consortium, 1998
E04	*Dm*	*Drosophila melanogaster*	Adams et al., 2000
E05	*Hs*	*Homo sapiens*	Uncomplete[b]
E06	*Eu*	*Encephalitozoon cuniculi*	Katinka et al., 2001
E07	*At*	*Arabidopsis thaliana*	*Arabidopsis* Genome Initiative, 2000
		Archaea	
A01	*Mk*	*Methanopyrus kandleri* AV19	Slesarev et al., 2002
A02	*Pa*	*Pyrococcus abyssi* GE5	Complete[c]
A03	*Pe*	*Pyrobaculum aerophilum* IM2	Fitz-Gibbon et al., 2002
A04	*Ap*	*Aeropyrum pernix* K1	Kawarabayasi et al., 1999
A05	*Af*	*Archaeoglobus fulgidus* DSM4304	Klenk et al., 1997
A06	*Ha*	*Halobacterium* sp. NRC-1	Ng et al., 2000
A07	*Ss*	*Sulfolobus solfataricus* P2	She et al., 2001
A08	*St*	*Sulfolobus tokodaii* 7	Kawarabayasi et al., 2001
A09	*Ta*	*Thermoplasma acidophilum*	Ruepp et al., 2000
A10	*Fa*	*Ferroplasma acidarmanus*	Nearly complete[d]
A11	*Mb*	*Methanosarcina barkeri*	Nearly complete[e]
A12	*Mj*	*Methanococcus jannaschii* DSM 2661	Bult et al., 1996
A13	*Mt*	*Methanobacterium thermoautotrophicum* delta H	Smith et al., 1997
		Bacteria	
B01	*Tp*	*Treponema pallidum* subsp. *pallidum* Nichols	Fraser et al., 1998
B02	*Bb*	*Borrelia burgdorferi* B31	Fraser et al., 1997
B03	*Ct*	*Chlamydia trachomatis* D/UW-3/CX	Stephens et al., 1998
B04	*Sy*	*Synechocystis* sp. PCC 6803	Kaneko et al., 1996
B05	*An*	*Anabaena* sp. PCC 7120	Kaneko et al., 2001
B06	*Ll*	*Lactococcus lactis* ssp. *lactis* IL1403	Bolotin et al., 2001
B07	*Lm*	*Listeria monocytogenes* EGD-e	Glaser et al., 2001
B08	*Bs*	*Bacillus subtilis* 168	Kunst et al., 1997
B09	*Aa*	*Aquifex aeolicus* VF5	Deckert et al., 1998
B10	*Mt*	*Mycobacterium tuberculosis* H37Rv	Cole et al., 1998
B11	*Dr*	*Deinococcus radiodurans* R1	White et al., 1999
B12	*Nm*	*Neisseria meningitidis* MC58 (serogroup B)	Tettelin et al., 2000
B13	*Pr*	*Pseudomonas aeruginosa* PAO1	Stover et al., 2000
B14	*Bu*	*Buchnera* APS	Shigenobu et al., 2000
B15	*Bh*	*Bacillus halodurans* C-125	Takami et al., 2000
B16	*Tm*	*Thermotoga maritima* MSB8	Nelson et al., 1999
B17	*Cj*	*Campylobacter jejuni* NCTC 11168	Parkhill et al., 2000
B18	*Vc*	*Vibrio cholerae* serotype OI biotype El Tor strain N16961	Heidelberg et al., 2000
B19	*Cp*	*Clostridium perfringens* 13	Shimizu et al., 2002
B20	*Hp*	*Helicobacter pylori* 26695	Tomb et al., 1997
B21	*Rs*	*Ralstonia solanacearum* GMI1000	Salanoubat et al., 2002
B22	*Mg*	*Mycoplasma genitalium* G-37	Fraser et al., 1996
B23	*Mp*	*Mycoplasma pneumoniae* M129	Himmelreich et al., 1996
B24	*Uu*	*Ureaplasma urealyticum* (*parvum*) serovar 3	Glass et al., 2000
B25	*Xf*	*Xilella fastidiosa* CVC 8.1.b clone 9.a.5.c	Simpson et al., 2000
B26	*Hi*	*Haemophilus influenzae* KW20	Fleischmann et al., 1995
B27	*Ec*	*Escherichia coli K12* -MG1655	Blattner et al., 1997
B28	*Rp*	*Rickettsia prowazekii* Madrid E	Andersson et al., 1998
B29	*Yp*	*Yersinia pestis* CO-92 Biovar Orientalis	Parkhill et al., 2001
B30	*Sm*	*Sinorhizobium meliloti* 1021	Galibert et al., 2001

[a] Unless otherwise specified below, genomic sequences were retrieved from the GenBank ftp site at: ftp://ncbi.nlm.nih.gov/genbank/genomes/.

[b] The *Homo sapiens* (E05) sequences scanned were the ths_chrXX.fna files available at GenBank ftp site at ftp://ncbi.nlm.nih.gov/genomes/H_sapiens/CHR_XX (XX = 01 to 22, X, Y, and Un); the estimated human genome size is 3,300 Mb (Lander et al., 2001).

[c] The sequence data for *Pyrococcus abyssi* (A02) were obtained from Genethon at http://www.genoscope.cns.fr/Pab/Pabyssi_complete_genome.fasta.

[d] Preliminary sequence data for *Ferroplasma acidarmanus* (A10) were obtained from The DOE Joint Genome Institute (JGI) at http://genome.ornl.gov/microbial/faci/. Large and small contigs files were concatenated (1,932,066 bp).

[e] Preliminary sequence data for *Methanosarcina barkeri* (A11) were obtained from The DOE Joint Genome Institute at http://genome.ornl.gov/microbial/mbar/. Large and small contigs files were concatenated (5,133,054 bp).

quence variations among the members of a given isoacceptor family, we used the following criteria to define an isoacceptor subfamily: two tDNAs belong to the same isoacceptor subfamily if they are perfectly identical in the three regions from nt 1 to 37, 38 to 46, and 48 to 73. This implies that both tDNAs have the same pattern of optional bases in the D-loop at positions 17, 17a, 20a, and 20b but allows for sequence variations inside the intron (if any, always located between positions 37 and 38 in eukaryotic tDNA) and inside the variable arm (between positions 46 and 48).

In a second step, the procedure was divided into three variants specialized in each domain of life. Bacterial, archaeal, and other lower eukaryotic genomes were then added to the input set. When available, results for each genome were compared with those presented in the respective publications (listed in Table 3) and/or Web servers and also with the results produced either by running tRNAscan-SE (Lowe & Eddy, 1997) or referring to GtRDB (http://rna.wustl.edu/GtRDB/). To detect a few missing tDNAs, constraints had to be relaxed. For example, G at position 19 had to be changed into a purine (designated R19 in Fig. 1B) in order to detect 14 extra single-copy bacterial tDNAs bearing A19.

Finally, the four genomes from higher eukaryotes were added to the set (*C. elegans*, *D. melanogaster*, *A. thaliana*, and partial data from *H. sapiens*). Sorting as a function of isoacceptor subfamily (as defined above) revealed much more divergence in the sequences of the same isoacceptor tDNA in these genomes. A number of false positives had to be eliminated after visual inspection (e.g., tDNA with degenerated or undetermined bases in some genomes, tDNA present as a unique copy significantly different from a multimember family, possible sequencing errors). In practice, intermediate files were produced including all tDNAs from a given genome and these files were manually edited to

TABLE 4. Constraints and parameters used in the search for cloverleaf tDNAs.[a]

Panel A

	Intron length	V-arm length (positions 44–48)
Eukarya	1 to 60	up to 29
Archaea	6 to 121	up to 21
Bacteria	not allowed	up to 24

Panel B: Requested bases

Position	8	18	19	53	54	55	58	61
Eukarya and Bacteria	T	G	R	R	H	Y	A	Y
Archaea								
M. kandleri (A01)	Y	G	R	R	H	Y	A	Y
Other archaea	T	G	R	R	H	Y	A	Y

Panel C: Numbers of GT pairings (GT) and mismatches (mm) allowed

	Acceptor stem		D-stem		Anticodon stem		T-stem		4 stems considered together		
	GT	mm	GT	mm	GT	mm	GT	mm	GT	mm	GT + mm
Eukarya	3	1	2	2	2	1	2	1	5	2	5
Archaea	2	1	2	2	2	1	2	1	4	2	5
Bacteria	3	1	2	2	2	2	2	1	5[b]	3	5[b]

Panel D: Weight

	A box (positions 7–16)	B box (positions 52–62)
Weight	>900	>1500 (>1,475 for Archaea)

Panel E: Minimal GC percent

Eukarya	35%
Archaea	40%
Bacteria	35%

[a] A regular cloverleaf structure (see Fig. 2; Sprinzl et al., 1998) was searched for to discover tRNA genes in the genomic sequences. The acceptor stem, D-stem, anticodon stem, and T-stem are 7, 4, 5 and 5 bp long, respectively. Positions optionally occupied in the D-loop are 17, 17a, 20a, and 20b. Position 47 in the variable arm may be occupied or not, such that the variable arm is 5 or 4 bases long, respectively. The variable arm is also allowed to be longer as indicated in Panel A. In Panel C, G-T pairings searched for in the tDNA are in fact G-U pairings in the mature tRNA. Mismatches are defined as any base pair different from Watson–Crick or G-T base pairs. Weighting applied in Panel D is computed according to Pavesi et al. (1994).

[b] 6 for *C. perfringens* (B19).

discard false positives. Once again, a few selected constraints had to be relaxed: For example, C had to be allowed at position 54 instead of the quasi invariant T54 (rarely A54, designated H54 in Fig. 2B) in order to correctly detect the 10 copies of the tDNA-His of *A. thaliana* (E07). Finally, we also needed to allow for the presence of base C at position 8 (instead of the usual T) for the scanning of the recently sequenced hyperthermophilic archaeon *M. kandleri* (A01) (see Table 4).

The genome of *A. thaliana* (E07) required a specific processing because this nuclear genome is known to include insertions originating from the mitochondrial genome (*Arabidopsis* Genome Initiative, 2000). However, the tDNAs of mitochondrial origin were easily distinguished, as they formed separate singleton families. Moreover, some of these organelle-derived tDNAs used an anticodon generally not used in eukaryotes; for example, tDNA-Ser (GGA), which is never used by eukaryotes, or tDNA-Leu (GAG), which is missing in higher eukaryotes and only occasionally used in *S. cerevisiae* (E01) (see Fig. 6A). In its present state, our search procedure detected 4,204 tDNAs (4,096 elongaters and 108 initiators) corresponding to distinct tRNA genes in the 50 genomes listed in Table 3. All these tDNAs fit the cloverleaf pattern shown in Figure 2A,B and obey the constraints listed in Table 4 (several different copies of a tDNA belonging to the same family are each counted for one). “Anomalously” located archaeal introns were manually located according to literature data, when available, and their exact location was determined, and in some cases corrected, using the criteria of a correct BHB consensus (Lykke-Andersen et al., 1997). Search for selenocysteine tDNA (Hubert et al., 1998) as well as prokaryotic tmRNA (Williams, 2002) are features not yet implemented in the tDNA search algorithm.

ACKNOWLEDGMENTS

We thank Jean-Pierre Waller (Gif-sur-Yvette) for improving the manuscript and for useful critical comments. We acknowledge Hervé Philippe (Unité Mixte de Recherche (UMR) 7622 Centre National de la Recherche Scientifique (CNRS), Université Paris 6) for computing the phylogenetic data of Figure 1 and Yves Boulard (Service de Biochimie et de Génétique Moléculaire, Saclay) for help in running tRNAscan-SE. Helpful critical reading of the manuscript was done by Catherine Florentz (Institut de Biologie Moléculaire et Cellulaire, Strasbourg) and François Michel (Centre de Génétique Moléculaire, Gif-sur-Yvette). We thank Christian Vivares (UMR CNRS 6023, Université Blaise Pascal, Aubière) for communicating the *Encephalitozoon cuniculi* genome sequence before publication. We thank the U.S. Department of Energy (DOE) Joint Genome Institute for permission to use the genome data of the archaeons *Ferroplasma acidarmanus* and *Methanosarcina barkeri*. The work of the DOE Joint Genome Institute was performed under the auspices of the Office of Biological and Environmental Research, by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098, and Los Alamos National Laboratory under contract No. W-7405-ENG-36. We thank Jean Weissenbach (Généthon, Évry) and his staff for communicating the sequence of the *Pyroccocus abyssi* genome. This work was supported in part by the CNRS (Programme Interdépartemental Géomicrobiologie des Environnements Extrèmes, Géomex 2002) and in part by grants from the Ministère de l’Éducation Nationale de la Recherche et de la Technologie (Programme de Recherche Fondamentale en Microbiologie et Maladies Infectieuses, 2001–2002) to HG.

Received May 30, 2002; returned for revision June 28, 2002; revised manuscript received July 12, 2002

REFERENCES

Abelson J, Trotta CR, Li H. 1998. tRNA splicing. *J Biol Chem 273*:12685–12688.

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YH, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, et al. 2000. The genome sequence of *Drosophila melanogaster*. *Science 287*:2185–2195.

Aebi M, Kirchner G, Chen JY, Vijayraghavan U, Jacobson A, Martin NC, Abelson J. 1990. Isolation of a temperature-sensitive mutant with an altered tRNA nucleotidyltransferase and cloning of the gene encoding tRNA nucleotidyltransferase in the yeast *Saccharomyces cerevisiae*. *J Biol Chem 265*:16216–16220.

Agris PF. 1991. Wobble position modified nucleosides evolved to select transfer RNA codon recognition: A modified-wobble hypothesis. *Biochimie 73*:1345–1349.

Andachi Y, Yamao F, Iwami M, Muto A, Osawa S. 1987. Occurrence of unmodified adenine and uracil at the first position of anticodon in threonine tRNAs in *Mycoplasma capricolum*. *Proc Natl Acad Sci USA 84*:7398–7402.

Andachi Y, Yamao F, Muto A, Osawa S. 1989. Codon recognition patterns as deduced from sequences of the complete set of transfer RNA species in *Mycoplasma capricolum*. Resemblance to mitochondria. *J Mol Biol 209*:37–54.

Andersson SG, Kurland CG. 1995. Genomic evolution drives the evolution of the translation system. *Biochem Cell Biol 73*:775–787.

Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature 396*:133–140.

Anonymous. 1997. The yeast genome directory. *Nature 387*:5.

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature 408*:796–815.

Arn EA, Abelson JN. 1996. The 2′-5′ RNA ligase of *Escherichia coli*. Purification, cloning, and genomic disruption. *J Biol Chem 271*:31145–31153.

Åström SU, von Pawel-Rammingen U, Byström AS. 1993. The yeast initiator tRNA-Met can act as an elongator tRNA-Met in vivo. *J Mol Biol 233*:43–58.

Auffinger P, Westhof E. 1998. Appendix 5: Localization and distribution of modified nucleosides in tRNA. In: Grosjean H, Benne R, eds. *Modification and editing of RNA*. Washington DC: ASM Press. pp 569–576.

Auffinger P, Westhof E. 1999. Singly and bifurcated hydrogen-bonded base pairs in tRNA anticodon hairpins and ribozymes. *J Mol Biol 292*:467–483.

Auffinger P, Westhof E. 2001. An extended structural signature for the tRNA anticodon loop. *RNA 7*:334–341.

Auxilien S, Crain PF, Trewyn RW, Grosjean H. 1996. Mechanism, specificity and general properties of the yeast enzyme catalysing the formation of inosine 34 in the anticodon of transfer RNA. *J Mol Biol 262*:437–458.

Barrell BG, Anderson S, Bankier AT, de Bruijn MH, Chen E, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG. 1980. Different pattern of codon recognition by mammalian mitochondrial tRNAs. *Proc Natl Acad Sci USA 77*:3164–3166.

Becker HF, Motorin Y, Sissler M, Florentz C, Grosjean H. 1997. Major identity determinants for enzymatic formation of ribothymidine and pseudouridine in the T-psi-loop of yeast tRNAs. *J Mol Biol 274*:505–518.

Belfort M, Weiner A. 1997. Another bridge between kingdoms: tRNA splicing in archaea and eukaryotes. *Cell 89*:1003–1006.

Beuning PJ, Nagan MC, Cramer CJ, Musier-Forsyth K, Gelpi JL, Bashford D. 2002. Efficient aminoacylation of the tRNA(Ala) acceptor stem: Dependence on the 2:71 base pair. *RNA 8*:659–670.

Björk GR. 1995. Biosynthesis and function of modified nucleosides. In: Söll D, RajBhandary U, eds. *tRNA: Structure, biosynthesis, and function*. Washington DC: ASM Press. pp 165–205.

Björk GR. 1998. Appendix 6: Modified nucleosides at positions 34 and 37 of tRNAs and their predicting coding capacities. In: Grosjean H, Benne R, eds. *Modification and editing of RNA*. Washington DC: ASM Press. pp. 577–581.

Blandin G, Durrens P, Tekaia F, Aigle M, Bolotin-Fukuhara M, Bon E, Casaregola S, de Montigny J, Gaillardin C, Lepingle A, Llorente B, Malpertuy A, Neuveglise C, Ozier-Kalogeropoulos O, Perrin A, Potier S, Souciet J, Talla E, Toffano-Nioche C, Wesolowski-Louvel M, Marck C, Dujon B. 2000. Genomic exploration of the hemiascomycetous yeasts: 4. The genome of *Saccharomyces cerevisiae* revisited. *FEBS Lett 487*:31–36.

Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science 277*:1453–1474.

Bolotin A, Wincker P, Mauger S, Jaillon O, Malarme K, Weissenbach J, Ehrlich SD, Sorokin A. 2001. The complete genome sequence of the lactic acid bacterium *Lactococcus lactis* ssp. *lactis* IL1403. *Genome Res 11*:731–753.

Bonocora RP, Shub DA. 2001. A novel group I intron-encoded endonuclease specific for the anticodon region of tRNA(fMet) genes. *Mol Microbiol 39*:1299–1306.

Boren T, Elias P, Samuelsson T, Claesson C, Barciszewska M, Gehrke CW, Kuo KC, Lustig F. 1993. Undiscriminating codon reading with adenosine in the wobble position. *J Mol Biol 230*:739–749.

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, Venter JC. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science 273*:1058–1073.

Burkard U, Willis I, Söll D. 1988. Processing of histidine transfer RNA precursors. Abnormal cleavage site for RNase P. *J Biol Chem 263*:2447–2451.

Burnol A-F, Margottin F, Huet J, Almouzni G, Prioleau M-N, Mechali M, Sentenac A. 1993. TFIIIC relieves repression of U6 snRNA transcription by chromatin. *Nature 362*:475–477.

C. elegans Sequencing Consortium. 1998. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science 282*:2012–2018.

Cedergren R, Grosjean H, LaRue B. 1986. Primordial reading of genetic information. *Biosystems 19*:259–266.

Cedergren RJ, Sankoff D, LaRue B, Grosjean H. 1981. The evolving tRNA molecule. *CRC Crit Rev Biochem 11*:35–104.

Cermakian N, Cedergren R. 1995. Modified nucleosides always were: An evolutionary model. In: Söll D, RajBhandary U, eds. *tRNA: Structure, biosynthesis, and function*. Washington DC: ASM Press. pp 535–541.

Cermakian N, McClain WH, Cedergren R. 1998. tRNA nucleotide 47: An evolutionary enigma. *RNA 4*:928–936.

Chambaud I, Heilig R, Ferris S, Barbe V, Samson D, Galisson F, Moszer I, Dybvig K, Wroblewski H, Viari A, Rocha EP, Blanchard A. 2001. The complete genome sequence of the murine respiratory pathogen *Mycoplasma pulmonis*. *Nucleic Acids Res 29*:2145–2153.

Chihade JW, Hayashibara K, Shiba K, Schimmel P. 1998. Strong selective pressure to use G:U to mark an RNA acceptor stem for alanine. *Biochemistry 37*:9193–9202.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature 393*:537–544.

Colot V, Rossignol JL. 1999. Eukaryotic DNA methylation as an evolutionary device. *Bioessays 21*:402–411.

Cornish-Bowden A. 1985. Nomenclature for incompletely specified bases in nucleic acid sequences: Recommendations 1984. *Nucleic Acids Res 13*:3021–3030.

Crick FH. 1966. Codon–anticodon pairing: The wobble hypothesis. *J Mol Biol 19*:548–555.

Crothers DM, Seno T, Söll D. 1972. Is there a discriminator site in transfer RNA? *Proc Natl Acad Sci USA 69*:3063–3067.

Curran JF. 1995. Decoding with the A:I wobble pair is inefficient. *Nucleic Acids Res 23*:683–688.

Curran JF. 1998. Modified nucleosides in translation. In: Grosjean H, Benne R, eds. *Modification and editing of RNA*. Washington DC: ASM Press. pp 493–516.

Cusack S, Yaremchuk A, Tukalo M. 1996. The crystal structure of the ternary complex of *T. thermophilus* seryl-tRNA synthetase with tRNA(Ser) and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J 15*:2834–2842.

Daniels CJ, Gupta R, Doolittle WF. 1985. Transcription and excision of a large intron in the tRNATrp gene of an archaebacterium, *Halobacterium volcanii*. *J Biol Chem 260*:3132–3134.

Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olsen GJ, Swanson RV. 1998. The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature 392*:353–358.

Dennis PP, Omer A, Lowe T. 2001. A guided tour: Small RNA function in Archaea. *Mol Microbiol 40*:509–519.

Deutscher MP. 1990. Ribonucleases, tRNA nucleotidyltransferase, and the 3′ processing of tRNA. *Prog Nucleic Acid Res Mol Biol 39*:209–240.

Dieci G, Giuliodori S, Catellani M, Percudani R, Ottonello S. 2002. Intragenic promoter adaptation and facilitated RNA polymerase III recycling in the transcription of SCR1, the 7SL RNA gene of *Saccharomyces cerevisiae*. *J Biol Chem 277*:6903–6914.

Dirheimer G, Keith G, Dumas P, Westhof E. 1995. Primary, secondary, and tertiary structures of tRNAs. In: Söll D, RajBhandary U, eds. *tRNA: Structure, biosynthesis, and function*. Washington DC: ASM Press. pp 93–126.

Dong H, Nilsson L, Kurland CG. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol 260*:649–663.

Donze D, Kamakaka RT. 2001. RNA polymerase III and RNA polymerase II promoter complexes are heterochromatin barriers in *Saccharomyces cerevisiae*. *EMBO J 20*:520–531.

Dorner M, Altmann M, Paabo S, Morl M. 2001. Evidence for import of a lysyl-tRNA into marsupial mitochondria. *Mol Biol Cell 12*:2688–2698.

d'Orval BC, Bortolin ML, Gaspin C, Bachellerie JP. 2001. Box C/D RNA guides for the ribose methylation of archaeal tRNAs. The tRNA-Trp intron guides the formation of two ribose-methylated nucleosides in the mature tRNA-Trp. *Nucleic Acids Res 29*:4518–4529.

Duret L. 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet 16*:287–289.

Edgell DR, Belfort M, Shub DA. 2000. Barriers to intron promiscuity in bacteria. *J Bacteriol 182*:5281–5289.

Edqvist J, Stråby KB, Grosjean H. 1995. Enzymatic formation of *N*2,*N*2-dimethylguanosine in eukaryotic tRNA: Importance of the tRNA architecture. *Biochimie 77*:54–61.

el-Mabrouk N, Lisacek F. 1996. Very fast identification of RNA motifs in genomic DNA. Application to tRNA search in the yeast genome. *J Mol Biol 264*:46–55.

Fey J, Weil JH, Tomita K, Cosset A, Dietrich A, Small I, Marechal-Drouard L. 2001. Editing of plant mitochondrial transfer RNAs. *Acta Biochim Pol 48*:383–389.

Fichant GA, Burks C. 1991. Identifying potential tRNA genes in genomic DNA sequences. *J Mol Biol 220*:659–671.

Filipowicz W. 2000. Imprinted expression of small nucleolar RNAs in brain: Time for RNomics. *Proc Natl Acad Sci USA 97*: 14035–14037.

Fitz-Gibbon ST, Ladner H, Kim UJ, Stetter KO, Simon MI, Miller JH. 2002. Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc Natl Acad Sci USA 99*:984–989.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science 269*:496–512.

Fraser CM, Casjens S, Huang WM, Sutton GG, Clayton R, Lathigra R, White O, Ketchum KA, Dodson R, Hickey EK, Gwinn M, Dougherty B, Tomb JF, Fleischmann RD, Richardson D, Peterson J, Kerlavage AR, Quackenbush J, Salzberg S, Hanson M, van Vugt R, Palmer N, Adams MD, Gocayne J, Venter JC, et al. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature 390*:580–586.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, et al. 1996. The minimal gene complement of *Mycoplasma genitalium*. *Science 270*:397–403.

Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R, Gwinn M, Hickey EK, Clayton R, Ketchum KA, Sodergren E, Hardham JM, McLeod MP, Salzberg S, Peterson J, Khalak H, Richardson D, Howell JK, Chidambaram M, Utterback T, McDonald L, Artiach P, Bowman C, Cotton MD, Venter JC, et al. 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science 281*:375–388.

Fromant M, Plateau P, Blanquet S. 2000. Function of the extra 5′-phosphate carried by histidine tRNA. *Biochemistry 39*:4062–4067.

Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, Bothe G, Boutry M, Bowser L, Buhrmester J, Cadieu E, Capela D, Chain P, Cowie A, Davis RW, Dreano S, Federspiel NA, Fisher RF, Gloux S, Godrie T, Goffeau A, Golding B, Gouzy J, Gurjal M, Hernandez-Lucas I, Hong A, Huizar L, Hyman RW, Jones T, Kahn D, Kahn ML, Kalman S, Keating DH, Kiss E, Komp C, Lelaure V, Masuy D, Palm C, Peck MC, Pohl TM, Portetelle D, Purnelle B, Ramsperger U, Surzycki R, Thebault P, Vandenbol M, Vorholter FJ, Weidner S, Wells DH, Wong K, Yeh KC, Batut J. 2001. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science 293*:668–672.

Galli G, Hofstetter H, Birnstiel ML. 1981. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature 294*:626–631.

Gaspin C, Cavaillé J, Erauso G, Bachellerie JP. 2000. Archaeal homologs of eukaryotic methylation guide small nucleolar RNAs: Lessons from the *Pyrococcus* genomes. *J Mol Biol 297*:895–906.

Gautheret D, Damberger SH, Gutell RR. 1995. Identification of base-triples in RNA using comparative sequence analysis. *J Mol Biol 248*:27–43.

Gautheret D, Lambert A. 2001. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *J Mol Biol 313*:1003–1011.

Gerber AP, Keller W. 2001. RNA editing by base deamination: More enzymes, more targets, new mysteries. *Trends Biochem Sci 26*:376–384.

Giegé R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res 26*:5017–5035.

Glaser P, Frangeul L, Buchrieser C, Rusniok C, Amend A, Baquero F, Berche P, Bloecker H, Brandt P, Chakraborty T, Charbit A, Chetouani F, Couve E, de Daruvar A, Dehoux P, Domann E, Dominguez-Bernal G, Duchaud E, Durant L, Dussurget O, Entian KD, Fsihi H, Portillo FG, Garrido P, Gautier L, Goebel W, Gomez-Lopez N, Hain T, Hauf J, Jackson D, Jones LM, Kaerst U, Kreft J, Kuhn M, Kunst F, Kurapkat G, Madueno E, Maitournam A, Vicente JM, Ng E, Nedjari H, Nordsiek G, Novella S, de Pablos B, Perez-Diaz JC, Purcell R, Remmel B, Rose M, Schlueter T, Simoes N, Tierrez A, Vazquez-Boland JA, Voss H, Wehland J, Cossart P. 2001. Comparative genomics of *Listeria* species. *Science 294*:849–852.

Glass JI, Lefkowitz EJ, Glass JS, Heiner CR, Chen EY, Cassell GH. 2000. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum*. *Nature 407*:757–762.

Goodman HM, Olson MV, Hall BD. 1977. Nucleotide sequence of a mutant eukaryotic gene: The yeast tyrosine-inserting ochre suppressor SUP4-o. *Proc Natl Acad Sci USA 74*:5453–5457.

Gott JM, Emeson RB. 2000. Functions and mechanisms of RNA editing. *Annu Rev Genet 34*:499–531.

Grosjean H, Auxilien S, Constantinesco F, Simon C, Corda Y, Becker HF, Foiret D, Morin A, Jin YX, Fournier M, Fourrey JL. 1996. Enzymatic conversion of adenosine to inosine and to *N*1-methylinosine in transfer RNAs: A review. *Biochimie 78*:488–501.

Grosjean H, Benne R, eds. 1998. *Modification and editing of RNA.* Washington DC: ASM Press.

Grosjean H, Cedergren RJ, McKay W. 1982. Structure in tRNA data. *Biochimie 64*:387–397.

Grosjean H, Nicoghosian K, Haumont E, Söll D, Cedergren R. 1985. Nucleotide sequences of two serine tRNAs with a GGA anticodon: The structure–function relationships in the serine family of *E. coli* tRNAs. *Nucleic Acids Res 13*:5697–5706.

Grosjean H, Szweykowska-Kulinska Z, Motorin Y, Fasiolo F, Simos G. 1997. Intron-dependent enzymatic formation of modified nucleosides in eukaryotic tRNAs: A review. *Biochimie 79*:293–302.

Gu X, Yu M, Ivanetich KM, Santi DV. 1998. Molecular recognition of tRNA by tRNA pseudouridine 55 synthase. *Biochemistry 37*:339–343.

Gu XR, Nicoghosian K, Cedergren RJ, Wong JT. 1983. Sequences of halobacterial tRNAs and the paucity of U in the first position of their anticodons. *Nucleic Acids Res 11*:5433–5442.

Guillon JM, Meinnel T, Mechulam Y, Lazennec C, Blanquet S, Fayat G. 1992. Nucleotides of tRNA governing the specificity of *Escherichia coli* methionyl-tRNA(fMet) formyltransferase. *J Mol Biol 224*:359–367.

Guindy YS, Samuelsson T, Johansen TI. 1989. Unconventional codon reading by *Mycoplasma mycoides* tRNAs as revealed by partial sequence analysis. *Biochem J 258*:869–873.

Guo Q, Gong Q, Tong Kl K, Vestergaard B, Costa A, Desgres J, Wong M, Grosjean H, Xue H, Wong JT. 2002. Recognition by tryptophanyl-tRNA synthetases of discriminator base on tRNA-Trp from three biological domains. *J Biol Chem 277*:14343–14349.

Hamada M, Huang Y, Lowe TM, Maraia RJ. 2001. Widespread use of TATA elements in the core promoters for RNA polymerases III, II, and I in fission yeast. *Mol Cell Biol 21*:6870–6881.

Hani J, Feldmann H. 1998. tRNA genes and retroelements in the yeast genome. *Nucleic Acids Res 26*:689–696.

Hao B, Gong W, Ferguson TK, James CM, Krzycki JA, Chan MK. 2002. A new UAG-encoded residue in the structure of a methanogen methyltransferase. *Science 296*:1462–1466.

Hara-Yokoyama M, Yokoyama S, Watanabe T, Watanabe K, Kitasumi M, Mitamura Y, Morii T, Takahashi S, Kushino Y, NIishimura S, Miyazawa T. 1986. Characteristic anticodon sequences of major tRNA species from an extreme thermophile: *Thermus thermophilus* HB8. *FEBS Lett 202*:149–152.

Hardt WD, Schlegl J, Erdmann VA, Hartmann RK. 1993. Role of the D arm and the anticodon arm in tRNA recognition by eubacterial and eukaryotic RNase P enzymes. *Biochemistry 32*:13046–13053.

Heidelberg JF, Eisen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Umayam L, Gill SR, Nelson KE, Read TD, Tettelin H, Richardson D, Ermolaeva MD, Vamathevan J, Bass S, Qin H, Dragoi I, Sellers P, McDonald L, Utterback T, Fleischmann RD, Nierman WC, White O. 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature 406*:477–483.

Helm M, Brulé H, Friede D, Giegé R, Pütz D, Florentz C. 2000. Search for characteristic structural features of mammalian mitochondrial tRNAs. *RNA 6*:1356–1379.

Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res 24*:4420–4449.

Hoang C, Ferré-D'Amare AR. 2001. Cocrystal structure of a tRNA Psi-55 pseudouridine synthase: Nucleotide flipping by an RNA-modifying enzyme. *Cell 107*:929–939.

Holley RW. 1965. Structure of an alanine transfer ribonucleic acid. *JAMA 194*:868–871.

Holley RW, Everett GA, Madison JT, Zamir A. 1965. Nucleotides sequences in the yeast alanine transfer ribonucleic acid. *J Biol Chem 240*:2122–2128.

Hou YM. 1994. Structural elements that contribute to an unusual tertiary interaction in a transfer RNA. *Biochemistry 33*:4677–4681.

Hou YM, Sterner T, Jansen M. 1995. Permutation of a pair of tertiary nucleotides in a transfer RNA. *Biochemistry 34*:2978–2984.

Houssier C, Grosjean H. 1985. Temperature jump relaxation studies on the interactions between transfer RNAs with complementary anticodons. The effect of modified bases adjacent to the anticodon triplet. *J Biomol Struct Dyn 3*:387–408.

Hubert N, Sturchler C, Westhof E, Carbon P, Krol A. 1998. The 9/4 secondary structure of eukaryotic selenocysteine tRNA: More pieces of evidence. *RNA 4*:1029–1033.

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol 2*:13–34.

Inagaki Y, Kojima A, Bessho Y, Hori H, Ohama T, Osawa S. 1995. Translation of synonymous codons in family boxes by *Mycoplasma capricolum* tRNAs with unmodified uridine or adenosine at the first anticodon position. *J Mol Biol 251*:486–492.

Inokuchi H, Yamao F. 1995. Structure and expression of prokaryotic tRNA Genes. In: Söll D, RajBhandary U, eds. *tRNA: Structure, biosynthesis, and function*. Washington DC: ASM Press. pp 17–30.

Jahn D, Pande S. 1991. Histidine tRNA guanylyltransferase from *Saccharomyces cerevisiae*. II. Catalytic mechanism. *J Biol Chem 266*:22832–22836.

Jiang HQ, Motorin Y, Jin YX, Grosjean H. 1997. Pleiotropic effects of intron removal on base modification pattern of yeast tRNA-Phe: An in vitro study. *Nucleic Acids Res 25*:2694–2701.

Jukes TH, Osawa S. 1990. The genetic code in mitochondria and chloroplasts. *Experientia 46*:1117–1126.

Kagawa Y, Nojima H, Nukiwa N, Ishizuka M, Nakajima T, Yasuhara T, Tanaka T, Oshima T. 1984. High guanine plus cytosine content in the third letter of codons of an extreme thermophile. DNA sequence of the isopropylmalate dehydrogenase of *Thermus thermophilus*. *J Biol Chem 259*:2956–2960.

Kanaya S, Yamada Y, Kinouchi M, Kudo Y, Ikemura T. 2001. Codon usage and tRNA genes in eukaryotes: Correlation of codon usage diversity with translation efficiency and with CG-dinucleotide usage as assessed by multivariate analysis. *J Mol Evol 53*:290–298.

Kanaya S, Yamada Y, Kudo Y, Ikemura T. 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: Gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene 238*:143–155.

Kaneko T, Nakamura Y, Wolk CP, Kuritz T, Sasamoto S, Watanabe A, Iriguchi M, Ishikawa A, Kawashima K, Kimura T, Kishida Y, Kohara M, Matsumoto M, Matsuno A, Muraki A, Nakazaki N, Shimpo S, Sugimoto M, Takazawa M, Yamada M, Yasuda M, Tabata S. 2001. Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120. *DNA Res 8*:205–213.

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, Kimura T, Hosouchi T, Matsuno A, Muraki A, Nakazaki N, Naruo K, Okumura S, Shimpo S, Takeuchi C, Wada T, Watanabe A, Yamada M, Yasuda M, Tabata S. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res 3*:109–136.

Kano A, Andachi Y, Ohama T, Osawa S. 1991. Novel anticodon composition of transfer RNAs in *Micrococcus luteus*, a bacterium with a high genomic G + C content. Correlation with codon usage. *J Mol Biol 221*:387–401.

Kano A, Ohama T, Abe R, Osawa S. 1993. Unassigned or nonsense codons in *Micrococcus luteus*. *J Mol Biol 230*:51–56.

Katinka MD, Duprat S, Cornillot E, Metenier G, Thomarat F, Prensier G, Barbe V, Peyretaillade E, Brottier P, Wincker P, Delbac F, El Alaoui H, Peyret P, Saurin W, Gouy M, Weissenbach J, Vivares CP. 2001. Genome sequence and gene compaction of the eukaryote parasite *Encephalitozoon cuniculi*. *Nature 414*:450–453.

Kawarabayasi Y, Hino Y, Horikawa H, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Otsuka R, Nakazawa H, Takamiya M, Kato Y, Yoshizawa T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Masuda S, Yanagii M, Nishimura M, Yamagishi A, Oshima T, Kikuchi H. 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain7. *DNA Res 8*:123–140.

Kawarabayasi Y, Hino Y, Horikawa H, Yamazaki S, Haikawa Y, Jin-no K, Takahashi M, Sekine M, Baba S, Ankai A, Kosugi H, Hosoyama A, Fukui S, Nagai Y, Nishijima K, Nakazawa H, Takamiya M, Masuda S, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Kikuchi H, et al. 1999. Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res 6*:83–101, 145–152.

Kawarabayasi Y, Sawada M, Horikawa H, Haikawa Y, Hino Y, Yamamoto S, Sekine M, Baba S, Kosugi H, Hosoyama A, Nagai Y, Sakai M, Ogura K, Otsuka R, Nakazawa H, Takamiya M, Ohfuku Y, Funahashi T, Tanaka T, Kudoh Y, Yamazaki J, Kushida N, Oguchi A, Aoki K, Kikuchi H. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res 5*:55–76.

Kenri T, Imamoto F, Kano Y. 1992. Construction and characterization of an *Escherichia coli* mutant deficient in the *metY* gene encoding tRNA(f2Met): Either tRNA(f1Met) or tRNA(f2Met) is required for cell growth. *Gene 114*:109–114.

Kjems J, Leffers H, Olesen T, Garrett RA. 1989. A unique tRNA intron in the variable loop of the extreme thermophile *Thermofilum pendens* and its possible evolutionary implications. *J Biol Chem 264*:17834–17837.

Klenk HP, Clayton RA, Tomb JF, White O, Nelson KE, Ketchum KA, Dodson RJ, Gwinn M, Hickey EK, Peterson JD, Richardson DL, Kerlavage AR, Graham DE, Kyrpides NC, Fleischmann RD, Quackenbush J, Lee NH, Sutton GG, Gill S, Kirkness EF, Dougherty BA, McKenney K, Adams MD, Loftus B, Venter JC, et al. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeglobus fulgidus*. *Nature 390*:364–370.

Komine Y, Adachi T, Inokuchi H, Ozeki H. 1990. Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. *J Mol Biol 212*:579–598.

Koshlap KM, Guenther R, Sochacka E, Malkiewicz A, Agris PF. 1999. A distinctive RNA fold: The solution structure of an analogue of the yeast tRNA-Phe T-Psi-C domain. *Biochemistry 38*:8647–8656.

Krupp G, Kahle D, Vögt T, Char S. 1991. Sequence changes in both flanking sequences of a pre-tRNA influence the cleavage specificity of RNase P. *J Mol Biol 217*:637–648.

Kunst F, Ogasawara N, Moszer I, Albertini AM, Alloni G, Azevedo V, Bertero MG, Bessieres P, Bolotin A, Borchert S, Borriss R, Boursier L, Brans A, Braun M, Brignell SC, Bron S, Brouillet S, Bruschi CV, Caldwell B, Capuano V, Carter NM, Choi SK, Codani JJ, Connerton IF, Danchin A, et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature 390*:249–256.

Lagerkvist U. 1986. Unconventional methods in codon reading. *Bioessays 4*:223–226.

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng

JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, et al. 2001. Initial sequencing and analysis of the human genome. *Nature 409*:860–921.

LaRue B, Cedergren RJ, Sankoff D, Grosjean H. 1979. Evolution of methionine initiator and phenylalanine transfer RNAs. *J Mol Evol 14*:287–300.

Lee CP, Seong BL, RajBhandary UL. 1991. Structural and sequence elements important for recognition of *Escherichia coli* formylmethionine tRNA by methionyl-tRNA transformylase are clustered in the acceptor stem. *J Biol Chem 266*:18012–18017.

Leontis NB, Westhof E. 2001. Geometric nomenclature and classification of RNA base pairs. *RNA 7*:499–512.

Levinger L, Bourne R, Kolla S, Cylin E, Russell K, Wang X, Mohan A. 1998. Matrices of paired substitutions show the effects of tRNA D/T loop sequence on *Drosophila* RNase P and 3′-tRNase processing. *J Biol Chem 273*:1015–1025.

Levitt M. 1969. Detailed molecular model for transfer ribonucleic acid. *Nature 224*:759–763.

Li H, Trotta CR, Abelson J. 1998. Crystal structure and evolution of a transfer RNA splicing enzyme. *Science 280*:279–284.

Lim VI, Curran JF. 2001. Analysis of codon:anticodon interactions within the ribosome provides new insights into codon reading and the genetic code structure. *RNA 7*:942–957.

Limbach PA, Crain PF, McCloskey JA. 1994. Summary: The modified nucleosides of RNA. *Nucleic Acids Res 22*:2183–2196.

Loria A, Pan T. 1997. Recognition of the T stem-loop of a pre-tRNA substrate by the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry 36*:6317–6325.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res 25*:955–964.

Lykke-Andersen J, Aagaard C, Semionenkov M, Garrett RA. 1997. Archaeal introns: Splicing, intercellular mobility and evolution. *Trends Biochem Sci 22*:326–331.

Maizels N, Weiner HM. 1999. The genomic tag hypothesis: What molecular fossils tell us about the evolution of tRNA. In: Gesteland RF, Cech TR, Atkins JF, eds. *The RNA world* (2nd ed.). Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 79–111.

Major F, Gautheret D, Cedergren R. 1993. Reproducing the three-dimensional structure of a tRNA molecule from structural constraints. *Proc Natl Acad Sci USA 90*:9408–9412.

Mandal N, Mangroo D, Dalluge JJ, McCloskey JA, RajBhandary UL. 1996. Role of the three consecutive G:C base pairs conserved in the anticodon stem of initiator tRNAs in initiation of protein synthesis in *Escherichia coli*. *RNA 2*:473–482.

Matsugi J, Murao K, Ishikura H. 1996. Characterization of a *B. subtilis* minor isoleucine tRNA deduced from tDNA having a methionine anticodon CAT. *J Biochem (Tokyo) 119*:811–816.

Mohan A, Whyte S, Wang X, Nashimoto M, Levinger L. 1999. The 3′ end CCA of mature tRNA is an antideterminant for eukaryotic 3′-tRNase. *RNA 5*:245–256.

Motorin Y, Grosjean H. 1999. Multisite-specific tRNA:m5C-methyltransferase (Trm4) in yeast *Saccharomyces cerevisiae*: Identification of the gene and substrate specificity of the enzyme. *RNA 5*:1105–1118.

Motorin Y, Keith G, Simon C, Foiret D, Simos G, Hurt E, Grosjean H. 1998. The yeast tRNA:pseudouridine synthase Pus1p displays a multisite substrate specificity. *RNA 4*:856–869.

Mueller U, Schubel H, Sprinzl M, Heinemann U. 1999. Crystal structure of acceptor stem of tRNA(Ala) from *Escherichia coli* shows unique G.U wobble base pair at 1.16 Å resolution. *RNA 5*:670–677.

Munz P, Leupold U, Agris P, Kohli J. 1981. In vivo decoding rules in *Schizosaccharomyces pombe* are at variance with in vitro data. *Nature 294*:187–188.

Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T, Yokoyama S. 1988a. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. *Nature 336*:179–181.

Muramatsu T, Yokoyama S, Horie N, Matsuda A, Ueda T, Yamaizumi Z, Kuchino Y, Nishimura S, Miyazawa T. 1988b. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. *J Biol Chem 263*:9261–9267.

Muto A, Osawa S. 1987. The guanine and cytosine content of genomic DNA and bacterial evolution. *Proc Natl Acad Sci USA 84*:166–169.

Nameki N. 1995. Identity elements of tRNA (Thr) towards *Saccharomyces cerevisiae* threonyl-tRNA synthetase. *Nucleic Acids Res 23*:2831–2836.

Nelson KE, Clayton RA, Gill SR, Gwinn ML, Dodson RJ, Haft DH, Hickey EK, Peterson JD, Nelson WC, Ketchum KA, McDonald L, Utterback TR, Malek JA, Linher KD, Garrett MM, Stewart AM, Cotton MD, Pratt MS, Phillips CA, Richardson D, Heidelberg J, Sutton GG, Fleischmann RD, Eisen JA, Fraser CM, et al. 1999. Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of *Thermotoga maritima*. *Nature 399*:323–329.

Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD, Lasky SR, Baliga NS, Thorsson V, Sbrogna J, Swartzell S, Weir D, Hall J, Dahl TA, Welti R, Goo YA, Leithauser B, Keller K, Cruz R, Danson MJ, Hough DW, Maddocks DG, Jablonski PE, Krebs MP, Angevine CM, Dale H. 2000. Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci USA 97*:12176–12181.

Oba T, Andachi Y, Muto A, Osawa S. 1991. CGG: An unassigned or nonsense codon in *Mycoplasma capricolum*. *Proc Natl Acad Sci USA 88*:921–925.

Odell L, Huang V, Jakacka M, Pan T. 1998. Interaction of structural modules in substrate binding by the ribozyme from *Bacillus subtilis* RNase P. *Nucleic Acids Res 26*:3717–3723.

Ohama T, Muto A, Osawa S. 1990a. Role of GC-biased mutation pressure on synonymous codon choice in *Micrococcus luteus*, a bacterium with a high genomic GC-content. *Nucleic Acids Res 18*:1565–1569.

Ohama T, Osawa S, Watanabe K, Jukes TH. 1990b. Evolution of the mitochondrial genetic code. IV. AAA as an asparagine codon in some animal mitochondria. *J Mol Evol 30*:329–332.

Ohama T, Suzuki T, Mori M, Osawa S, Ueda T, Watanabe K, Nakase T. 1993. Non-universal decoding of the leucine codon CUG in several *Candida* species. *Nucleic Acids Res 21*:4039–4045.

Omer AD, Ziesche S, Ebhardt H, Dennis PP. 2002. In vitro reconstitution and activity of a C/D box methylation guide ribonucleoprotein complex. *Proc Natl Acad Sci USA 99*:5289–5294.

Orellana O, Cooley L, Söll D. 1986. The additional guanylate at the 5′ terminus of *Escherichia coli* tRNA-His is the result of unusual processing by RNase P. *Mol Cell Biol 6*:525–529.

Osawa S, Collins D, Ohama T, Jukes TH, Watanabe K. 1990. Evolution of the mitochondrial genetic code. III. Reassignment of CUN codons from leucine to threonine during evolution of yeast mitochondria. *J Mol Evol 30*:322–328.

Osawa S, Jukes TH. 1988. Evolution of the genetic code as affected by anticodon content. *Trends Genet 4*:191–198.

Osawa S, Jukes TH, Muto A, Yamao F, Ohama T, Andachi Y. 1987. Role of directional mutation pressure in the evolution of the eubacterial genetic code. *Cold Spring Harb Symp Quant Biol 52*:777–789.

Osawa S, Jukes TH, Watanabe K, Muto A. 1992. Recent evidence for evolution of the genetic code. *Microbiol Rev 56*:229–264.

Osawa S, Ohama T, Jukes TH, Watanabe K. 1989. Evolution of the mitochondrial genetic code. I. Origin of AGR serine and stop codons in metazoan mitochondria. *J Mol Evol 29*:202–207.

Palmer JR, Daniels CJ. 1995. In vivo definition of an archaeal promoter. *J Bacteriol 177*:1844–1849.

Pande S, Jahn D, Söll D. 1991. Histidine tRNA guanylyltransferase from *Saccharomyces cerevisiae*. I. Purification and physical properties. *J Biol Chem 266*:22826–22831.

Pannucci JA, Haas ES, Hall TA, Harris JK, Brown JW. 1999. RNase P RNAs from some Archaea are catalytically active. *Proc Natl Acad Sci USA 96*:7803–7808.

Parkhill J, Wren BW, Mungall K, Ketley JM, Churcher C, Basham D, Chillingworth T, Davies RM, Feltwell T, Holroyd S, Jagels K, Karlyshev AV, Moule S, Pallen MJ, Penn CW, Quail MA, Rajandream MA, Rutherford KM, van Vliet AH, Whitehead S, Barrell BG. 2000. The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hypervariable sequences. *Nature 403*:665–668.

Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MT, Prentice MB, Sebaihia M, James KD, Churcher C, Mungall KL, Baker S, Basham D, Bentley SD, Brooks K, Cerdeno-Tarraga AM, Chillingworth T, Cronin A, Davies RM, Davis P, Dougan G, Feltwell T, Hamlin N, Holroyd S, Jagels K, Karlyshev AV, Leather S, Moule S, Oyston PC, Quail M, Rutherford K, Simmonds M, Skelton J, Stevens K, Whitehead S, Barrell BG. 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. *Nature 413*: 523–527.

Paule MR, White RJ. 2000. Survey and summary: Transcription by RNA polymerases I and III. *Nucleic Acids Res 28*:1283–1298.

Pavesi A, Conterio F, Bolchi A, Dieci G, Ottonello S. 1994. Identification of new eukaryotic tRNA genes in genomic DNA databases by a multistep weight matrix analysis of transcriptional control regions. *Nucleic Acids Res 22*:1247–1256.

Percudani R. 2001. Restricted wobble rules for eukaryotic genomes. *Trends Genet 17*:133–135.

Percudani R, Pavesi A, Ottonello S. 1997. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *J Mol Biol 268*:322–330.

Perreau VM, Keith G, Holmes WM, Przykorska A, Santos MA, Tuite MF. 1999. The *Candida albicans* CUG-decoding ser-tRNA has an atypical anticodon stem-loop structure. *J Mol Biol 293*:1039–1053.

Pintard L, Lecointe F, Bujnicki JM, Bonnerot C, Grosjean H, Lapeyre B. 2002. Trm7p catalyzes the formation of two 2′-O-methylriboses in yeast tRNA anticodon loop. *EMBO J 21*:1811–1820.

Price DH, Gray MW. 1999. Confirmation of predicted edits and demonstration of unpredicted edits in *Acanthamoeba castellanii* mitochondrial tRNAs. *Curr Genet 35*:23–29.

Quigley GJ, Rich A. 1976. Structural domains of transfer RNA molecules. *Science 194*:796–806.

RajBhandary U, Chow CM. 1995. Initiator tRNAs and initiation of protein synthesis. In: Söll D, RajBhandary U, eds. *tRNA: Structure, biosynthesis, and function*. Washington DC: ASM Press. pp 511–528.

Reuven NB, Zhou Z, Deutscher MP. 1997. Functional overlap of tRNA nucleotidyltransferase, poly(A) polymerase I, and polynucleotide phosphorylase. *J Biol Chem 272*:33255–33259.

Rich A, RajBhandary UL. 1976. Transfer RNA: Molecular structure, sequence, and properties. *Annu Rev Biochem 45*:805–860.

Robb FT, Maeder DL, Brown JR, DiRuggiero J, Stump MD, Yeh RK, Weiss RB, Dunn DM. 2001. Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: Implications for physiology and enzymology. *Methods Enzymol 330*:134–157.

Romby P, Carbon P, Westhof E, Ehresmann C, Ebel JP, Ehresmann B, Giegé R. 1987. Importance of conserved residues for the conformation of the T-loop in tRNAs. *J Biomol Struct Dyn 5*:669–687.

Rudinger J, Blechschmidt B, Ribeiro S, Sprinzl M. 1994. Minimalist aminoacylated RNAs as efficient substrates for elongation factor Tu. *Biochemistry 33*:5682–5688.

Rudinger J, Hillenbrandt R, Sprinzl M, Giegé R. 1996. Antideterminants present in minihelix(Sec) hinder its recognition by prokaryotic elongation factor Tu. *EMBO J 15*:650–657.

Ruepp A, Graml W, Santos-Martinez ML, Koretke KK, Volker C, Mewes HW, Frishman D, Stocker S, Lupas AN, Baumeister W. 2000. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature 407*:508–513.

Salanoubat M, Genin S, Artiguenave F, Gouzy J, Mangenot S, Arlat M, Billault A, Brottier P, Camus JC, Cattolico L, Chandler M, Choisne N, Claudel-Renard C, Cunnac S, Demange N, Gaspin C, Lavie M, Moisan A, Robert C, Saurin W, Schiex T, Siguier P, Thebault P, Whalen M, Wincker P, Levy M, Weissenbach J, Boucher CA. 2002. Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature 415*:497–502.

Santos MA, Perreau VM, Tuite MF. 1996. Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. *EMBO J 15*:5060–5068.

Santos MA, Ueda T, Watanabe K, Tuite MF. 1997. The non-standard genetic code of *Candida* spp.: An evolving genetic code or a novel mechanism for adaptation? *Mol Microbiol 26*:423–431.

Schneider A, Marechal-Drouard L. 2000. Mitochondrial tRNA import: Are there distinct mechanisms? *Trends Cell Biol 10*:509–513.

Schultz DW, Yarus M. 1994. tRNA structure and ribosomal function. II. Interaction between anticodon helix and other tRNA mutations. *J Mol Biol 235*:1395–1405.

Schurer H, Schiffer S, Marchfelder A, Mörl M. 2001. This is the end: Processing, editing and repair at the tRNA 3′-terminus. *Biol Chem 382*:1147–1156.

Seoighe C, Wolfe KH. 1999. Updated map of duplicated regions in the yeast genome. *Gene 238*:253–261.

Sharp PM, Li W-H. 1987. The Codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res 15*:1281–1295.

She Q, Singh RK, Confalonieri F, Zivanovic Y, Allard G, Awayez MJ, Chan-Weiher CC, Clausen IG, Curtis BA, De Moors A, Erauso G, Fletcher C, Gordon PM, Heikamp-de Jong I, Jeffries AC, Kozera CJ, Medina N, Peng X, Thi-Ngoc HP, Redder P, Schenk ME, Theriault C, Tolstrup N, Charlebois RL, Doolittle WF, Duguet M, Gaasterland T, Garrett RA, Ragan MA, Sensen CW, Van der Oost J. 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. *Proc Natl Acad Sci USA 98*:7835–7840.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature 407*:81–86.

Shimizu T, Ohtani K, Hirakawa H, Ohshima K, Yamashita A, Shiba T, Ogasawara N, Hattori M, Kuhara S, Hayashi H. 2002. Complete genome sequence of *Clostridium perfringens*, an anaerobic flesheater. *Proc Natl Acad Sci USA 99*:996–1001.

Simpson AJ, Reinach FC, Arruda P, Abreu FA, Acencio M, Alvarenga R, Alves LM, Araya JE, Baia GS, Baptista CS, Barros MH, Bonaccorsi ED, Bordin S, Bove JM, Briones MR, Bueno MR, Camargo AA, Camargo LE, Carraro DM, Carrer H, Colauto NB, Colombo C, Costa FF, Costa MC, Costa-Neto CM, Coutinho LL, Cristofani M, Dias-Neto E, Docena C, El-Dorry H, Facincani AP, Ferreira AJ, Ferreira VC, Ferro JA, Fraga JS, Franca SC, Franco MC, Frohme M, Furlan LR, Garnier M, Goldman GH, Goldman MH, Gomes SL, Gruber A, Ho PL, Hoheisel JD, Junqueira ML, Kemper EL, Kitajima JP, Marino CL. 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. *Nature 406*:151–157.

Slesarev AI, Mezhevaya KV, Makarova KS, Polushin NN, Shcherbinina OV, Shakhova VV, Belova GI, Aravind L, Natale DA, Rogozin IB, Tatusov RL, Wolf YI, Stetter KO, Malykh AG, Koonin EV, Kozyavkin SA. 2002. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc Natl Acad Sci USA 99*:4644–4649.

Smith D, Yarus M. 1989. Transfer RNA structure and coding specificity. II. A D-arm tertiary interaction that restricts coding range. *J Mol Biol 206*:503–511.

Smith DR, Doucette-Stamm LA, Deloughery C, Lee H, Dubois J, Aldredge T, Bashirzadeh R, Blakely D, Cook R, Gilbert K, Harrison D, Hoang L, Keagle P, Lumm W, Pothier B, Qiu D, Spadafora R, Vicaire R, Wang Y, Wierzbowski J, Gibson R, Jiwani N, Caruso A, Bush D, Reeve JN, et al. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* DH: Functional analysis and comparative genomics. *J Bacteriol 179*:7135–7155.

Soma A, Uchiyama K, Sakamoto T, Maeda M, Himeno H. 1999. Unique recognition style of tRNA(Leu) by *Haloferax volcanii* leucyl-tRNA synthetase. *J Mol Biol 293*:1029–1038.

Souciet J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, de Montigny J, Dujon B, Durrens P, Gaillardin C, Lepingle A, Llorente B, Malpertuy A, Neuveglise C, Ozier-Kalogeropoulos O, Potier S, Saurin W, Tekaia F, Toffano-Nioche C, Wesolowski-Louvel M, Wincker P, Weissenbach J. 2000. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett 487*:3–12.

Sprague KU. 1995. Transcription of eukaryotic tRNA genes. In: Söll D, RajBhandary U, eds. *tRNA: Structure, biosynthesis, and function*. Washington DC: ASM Press. pp 31–50.

Sprinzl M, Horn C, Brown M, Ioudovitch A, Steinberg S. 1998. Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res 26*:148–153.

Srinivasan G, James CM, Krzycki JA. 2002. Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA. *Science 296*:1459–1462.

Steinberg S, Ioudovitch A. 1996. A role for the bulged nucleotide 47 in the facilitation of tertiary interactions in the tRNA structure. *RNA 2*:84–87.

Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L, Mitchell W, Olinger L, Tatusov RL, Zhao Q, Koonin EV, Davis RW. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science 282*:754–759.

Sterner T, Jansen M, Hou YM. 1995. Structural and functional accommodation of nucleotide variations at a conserved tRNA tertiary base pair. *RNA 1*:841–851.

Stewart TS, Roberts RJ, Strominger JL. 1971. Novel species of tRNA. *Nature 230*:36–38.

Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, Brinkman FS, Hufnagle WO, Kowalik DJ, Lagrou M, Garber RL, Goltry L, Tolentino E, Westbrock-Wadman S, Yuan Y, Brody LL, Coulter SN, Folger KR, Kas A, Larbig K, Lim R, Smith K, Spencer D, Wong GK, Wu Z, Paulsen IT. 2000. Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen. *Nature 406*:959–964.

Sundararajan A, Michaud WA, Qian Q, Stahl G, Farabaugh PJ. 1999. Near-cognate peptidyl-tRNAs promote +1 programmed translational frameshifting in yeast. *Mol Cell 4*:1005–1015.

Suzuki T, Ueda T, Watanabe K. 1997. The "polysemous" codon—A codon with multiple amino acid assignment caused by dual specificity of tRNA identity. *EMBO J 16*:1122–1134.

Szweykowska-Kulinska Z, Senger B, Keith G, Fasiolo F, Grosjean H. 1994. Intron-dependent formation of pseudouridines in the anticodon of *Saccharomyces cerevisiae* minor tRNA(Ile). *EMBO J 13*:4636–4644.

Takami H, Nakasone K, Takaki Y, Maeno G, Sasaki R, Masui N, Fuji F, Hirama C, Nakamura Y, Ogasawara N, Kuhara S, Horikoshi K. 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res 28*:4317–4331.

Tang TH, Rozhdestvensky TS, d'Orval BC, Bortolin ML, Huber H, Charpentier B, Branlant C, Bachellerie JP, Brosius J, Huttenhofer A. 2002. RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. *Nucleic Acids Res 30*:921–930.

Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ, Nelson WC, Gwinn ML, DeBoy R, Peterson JD, Hickey EK, Haft DH, Salzberg SL, White O, Fleischmann RD, Dougherty BA, Mason T, Ciecko A, Parksey DS, Blair E, Cittone H, Clark EB, Cotton MD, Utterback TR, Khouri H, Qin H, Vamathevan J, Gill J, Scarlato V, Masignani V, Pizza M, Grandi G, Sun L, Smith HO, Fraser CM, Moxon ER, Rappuoli R, Venter JC. 2000. Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science 287*:1809–1815.

Tomb JF, White O, Kerlavage AR, Clayton RA, Sutton GG, Fleischmann RD, Ketchum KA, Klenk HP, Gill S, Dougherty BA, Nelson K, Quackenbush J, Zhou L, Kirkness EF, Peterson S, Loftus B, Richardson D, Dodson R, Khalak HG, Glodek A, McKenney K, Fitzegerald LM, Lee N, Adams MD, Venter JC, et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature 388*:539–547.

Tomita K, Weiner AM. 2001. Collaboration between CC- and A-adding enzymes to build and repair the 3′-terminal CCA of tRNA in *Aquifex aeolicus*. *Science 294*:1334–1336.

Trotta CR, Miao F, Arn EA, Stevens SW, Ho CK, Rauhut R, Abelson JN. 1997. The yeast tRNA splicing endonuclease: A tetrameric enzyme with two active site subunits homologous to the archaeal tRNA endonucleases. *Cell 89*:849–858.

Tuohy TM, Li Z, Atkins JF, Deutscher MP. 1994. A functional mutant of tRNA(2Arg) with ten extra nucleotides in its T-Psi-C arm. *J Mol Biol 235*:1369–1376.

Ushida C, Muramatsu T, Mizushima H, Ueda T, Watanabe K, Stetter KO, Crain PF, McCloskey JA, Kuchino Y. 1996. Structural feature of the initiator tRNA gene from *Pyrodictium occultum* and the thermal stability of its gene product, tRNA(imet). *Biochimie 78*:847–855.

Valenzuela P, Venegas A, Weinberg F, Bishop R, Rutter WJ. 1978. Structure of yeast phenylalanine-tRNA genes: An intervening DNA segment within the region coding for the tRNA. *Proc Natl Acad Sci USA 75*:190–194.

Watanabe Y, Yokobori S, Inaba T, Yamagishi A, Oshima T, Kawarabayasi Y, Kikuchi H, Kita K. 2002. Introns in protein-coding genes in Archaea. *FEBS Lett 510*:27–30.

Weiss GB. 1973. Translational control of protein synthesis by tRNA unrelated to changes in tRNA concentration. *J Mol Evol 2*:199–204.

Westaway SK, Abelson J. 1995. Splicing of tRNA precursors. In: Söll D, RajBhandary U, eds. *tRNA: Structure, biosynthesis, and function.* Washington DC: ASM Press. pp 79–92.

Westhof E, Auffinger P. 2001. tRNA structure. In: *Encyclopedia of Life Sciences*, New York: Nature Publishing Group. pp 1–10.

White O, Eisen JA, Heidelberg JF, Hickey EK, Peterson JD, Dodson RJ, Haft DH, Gwinn ML, Nelson WC, Richardson DL, Moffat KS, Qin H, Jiang L, Pamphile W, Crosby M, Shen M, Vamathevan JJ, Lam P, McDonald L, Utterback T, Zalewski C, Makarova KS, Aravind L, Daly MJ, Fraser CM, et al. 1999. Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science 286*:1571–1577.

Williams KP. 2002. The tmRNA Website: Invasion by an intron. *Nucleic Acids Res 30*:179–182.

Woese CR, Kandler O, Wheelis ML. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA 87*:4576–4579.

Wong JT. 1988. Evolution of the genetic code. *Microbiol Sci 5*:174–181.

Wood V, Gwilliam R, Rajandream MA, Lyne M, Lyne R, Stewart A, Sgouros J, Peat N, Hayles J, Baker S, Basham D, Bowman S, Brooks K, Brown D, Brown S, Chillingworth T, Churcher C, Collins M, Connor R, Cronin A, Davis P, Feltwell T, Fraser A, Gentles S, Goble A, Hamlin N, Harris D, Hidalgo J, Hodgson G, Holroyd S, Hornsby T, Howarth S, Huckle EJ, Hunt S, Jagels K, James K, Jones L, Jones M, Leather S, McDonald S, McLean J, Mooney P, Moule S, Mungall K, Murphy L, Niblett D, Odell C, Oliver K, O'Neil S, Pearson D, Quail MA, Rabbinowitsch E, Rutherford K, Rutter S, Saunders D, Seeger K, Sharp S, Skelton J, Simmonds M, Squares R, Squares S, Stevens K, Taylor K, Taylor RG, Tivey A, Walsh S, Warren T, Whitehead S, Woodward J, Volckaert G, Aert R, Robben J, Grymonprez B, Weltjens I, Vanstreels E, Rieger M, Schafer M, Muller-Auer S, Gabel C, Fuchs M, Fritzc C, Holzer E, Moestl D, Hilbert H, Borzym K, Langer I, Beck A, Lehrach H, Reinhardt R, Pohl TM, Eger P, Zimmermann W, Wedler H, Wambutt R, Purnelle B, Goffeau A, Cadieu E, Dreano S, Gloux S, Lelaure V, et al. 2002. The genome sequence of *Schizosaccharomyces pombe*. *Nature 415*:871–880.

Yarus M, Smith AE. 1995. tRNA on the ribosome: A waggle theory. In: Söll D, RajBhandary U, eds. *tRNA: Structure, biosynthesis, and function*. Washington DC: ASM Press. pp 443–469.

Yokobori S, Suzuki T, Watanabe K. 2001. Genetic code variations in mitochondria: tRNA as a major determinant of genetic code plasticity. *J Mol Evol 53*:314–326.

Yokogawa T, Suzuki T, Ueda T, Mori M, Ohama T, Kuchino Y, Yoshinari S, Motoki I, Nishikawa K, Osawa S, et al. 1992. Serine tRNA complementary to the nonuniversal serine codon CUG in *Candida cylindracea*: Evolutionary implications. *Proc Natl Acad Sci USA 89*:7408–7411.