68–72 Nucleic Acids Research, 1996, Vol. 24, No. 1

Compilation of tRNA sequences and sequences of tRNA genes

Mathias Sprinzl*, Clemens Steegborn, Frank Hübel[1] and Sergey Steinberg[+]

Laboratorium für Biochemie, Universität Bayreuth, 95440 Bayreuth, Germany and [1]Institut für Landwirtschaftliche Botanik, Universität Bonn, Meckenheimer Allee 176, 53115 Bonn, Germany

Received September 5, 1995; Accepted September 19, 1995

INTRODUCTION

The 1995 compilation contains 2700 sequences of tRNAs and tRNA genes. The last edition which appeared two years ago (1) was supplemented by 694 new sequences. The sequences of tRNA mutants and of tRNAs originating from transformed or differentiated cells were not considered.

The tRNAs included in the compilation are listed in Table 1. Each tRNA or tRNA gene is specified by the (abbreviated) name of the organism from which it was isolated and a four digit code: the first three digits identify the organism, the last digit specifies the particular isoacceptor. The amino acid specificity of the tRNA is indicated by a one-letter amino acid code. The tRNAs coding for selenocysteine were annotated with the letter Z. Initiator tRNAs are annotated with the letter X.

The sequences, references and footnotes of tRNAs and tRNA genes listed in Table 1 are deposited in the European Bioinformatics Institute (EBI) Data Library. The references are restricted to the first complete publication of the sequence unless additional information (e.g. base modification, corrections, etc.) was later obtained. In such cases additional references were added.

In order to facilitate a computer analysis an alignment is used which is most compatible with the tRNA phylogeny and known three-dimensional structures of tRNA. The corresponding numbering system is shown in Figure 1.

As was the case in the previous edition (1), this publication does not contain a sequence printout. Instead, the sequences have been deposited in the EBI Data Library. This publication should be therefore quoted as a reference for data obtained from the electronically accessible database. Information on how to access the sequence files can be obtained by electronic mail: send email to Netserv@ebi.ac.uk containing the command 'HELP TRNA.' The help file will contain all the information needed to obtain the requested sequence. The tRNA database is also available via anonymous FTP from ftp.ebi.ac.uk in the directory pub/databases/tRNA. It is also distributed on the EBI CD-ROM. Contact the EBI Data Library, Hinxton Hall, Hinxton, Cambridge CB10 1RQ, UK (Fax: +44 1223 494468, email: DataLib@EBI.AC.UK).

Researchers who wish to obtain the sequence information on a floppy disk, or as a hardcopy should contact M. Sprinzl, Laboratorium für Biochemie, Universität Bayreuth, D-95440 Bayreuth, Germany, Fax: +49 921 552432, email: Mathias.Sprinzl@uni-bayreuth.de. Software allowing search for tRNA sequences according to several criteria, e.g. source, partial sequence, modified nucleoside, anticodon, amino acid specificity and printout of sequences in a form of tables or cloverleaves, is also available.

Presentation of sequences

The sequences in the database are divided into three parts. The first two parts contain the sequences of the tRNA genes and tRNAs, respectively, which can be fitted into the canonical tRNA alignment. The third part contains tRNA and tRNA gene sequences, mainly of animal mitochondria, whose secondary structures differ from most tRNAs and could not be aligned according to Figure 1.

An example for sequence presentation in the database is given in Table 2. Each sequence in the compilation occupies two consecutive lines. The first line begins with the letter 'D' or 'R' and contains the six-position identification code of the sequence ('D' or 'R' for DNA or RNA, respectively; a one-letter code for the amino acid, X for methionine-initiator, Z for selenocysteine; and the four-digit code specifying the organism and isoacceptor. After this, the sequence of the anticodon (in the case of tRNA sequences in its modified form) is given, followed by the name and the kingdom of organism (Table 1), and the sequence (99 standard positions). The second line begins with the sign '+' and contains the information about base-pairing (double helical regions only, tertiary interactions are not annotated). All other lines in the compilation begin with signs other than 'D,' 'R' or '+' (usually '*') and contain comments.

Nucleotides involved in Watson–Crick pairs are marked with '=', the G·U pairs are indicated with the sign '*'. Nucleotides 26 and 44 are considered to form a base pair included in the anticodon stem (Fig. 1).

The sequences in original publications denoted as 'yeast' are assigned to *Saccharomyces cerevisiae*. The user should be aware, however, that some of these organisms have possibly been misclassified and that the original literature should be consulted.

This compilation uses a one-letter code for all nucleotides including modified ones. For standard nucleotides, adenosine, cytidine, guanosine, thymidine and uridine the usual abbreviations, A, C, G, T and U, respectively, are used. To designate

* To whom correspondence should be addressed

[+]Present address: Université de Montréal, Faculté de Médicine, Départment de Biochimie, C.P. 6128, Succursale Centre-Ville, Montréal, Québec H3C 3J7, Canada (on leave from Engelhardt Institute of Molecular Biology, Vavilova 32, Moscow 117984, Russia)

Table 1. List of tRNA sequences and sequences of tRNA genes included in the compilation

PART ONE: Sequences of tRNA genes

Source	Code	tRNA genes
VIRUSES	000-029	
MYCOBACTERIOPH. L5	020	NOW
PHAGE T4	022	GILRQRST
PHAGE T5	026	ACDEFGHIKLMNPQRSTVWXY
ARCHAEBACTERIA	030-109	
ARCHAEOGLOBUS FULG.	034	A
HALOBACTERIUM CUT.	038	AG
HALOBACTERIUM HAL.	042	A
HALOBACTERIUM MAR.	044	LS
HALOBACTERIUM MED.	048	W
HALOFERAX VOLCANII	050	CW
METHANOBAC.FORM.	058	A
METHANOBAC.THERM.	062	A
METHANOCOC.VANN.	066	ADEFHIKLNPQRTWY
METHANOTHRIX SOEH.	067	A
METHANOTHERM. FER.	068	ADEHIKLMNPST
RUMINOBACTER AMYLO	070	E
METHANOCOC.VOLTAE	074	DEPTY
METHANOPYRUS KAND.	076	KLQS
METHANOSPIR. HUNG.	078	A
SULFOLOBUS SOLFA.	086	FGLIYK
THERMOPLASMA ACID.	090	M
THERMOCOCCUS CELER	094	APT
THERMOFIL. PENDENS	096	GM
THERMOPROT. TENAX	098	AALIX
THERMOTOGA MARIT.	099	MMTWY
EUBACTERIA	110-239	
MYCOPLASMA CAPRIC.	114	ACDEFGHIKLLMNPQRRSTTVWWXY
MYCOPLASMA MYCOID.	118	ADEFGIMNPRRSTVX
MYCOPLASMA PNEU.	120	ACDDEGGHIKKLLMNPQRRRSSSSTTTVWWXYY
MYCOPLASMA POSS	122	KL
ACHOLEPLASMA LAID.	123	AACDEFGHIKKLLMMMQRSSTYW
SPIROPLASMA CITRI	125	DWW
SPIROPLASMA MELIF.	126	ACDHMPRSX
BORRELIA BURGDORF.	128	AI
BARTONELLA BACIL.	129	IK
STREPTOMYCES GRIS.	130	S
STREPTOMYCES COEL.	131	L
STREPTOMYCES RIM.	134	EDGEX
STREPTOMYCES LIV.	135	CDEEGGHHYGQRGWY
STREPTOMYCES AMBO.	136	P
STAPHYLOCOC. AURE.	139	DG
MYCOBACT. TUBERC.	140	P
KLEBSIELLA AEROGE.	141	N
AGROBACTER. TUME.	142	R
CLOSTRIDIUM THERM.	143	Z
DESULFOMICR. BACA.	144	Z
CLOSTRIDIUM ACETO.	145	T
PLESIOMONAS SHIGE.	146	I
ENTEROCOCCUS HIRAE	147	A
STAPHYLOCOC. AURE.	148	ACFGGGGHIKLLLMPQRSSTTVWXY
LACTOBAC. BULG.	150	DEGNPRSV
LACTOBAC.DELBRUEC.	152	S
LACTOCOCCUS LACTIS	153	AAEFGIMSX
BACILLUS SUBTILIS	154	AAAACDEFFGGGHHIIKKLLLLLLMMMNPQRSSSTTTVWXXY
BACILLUS CIRCULANS	156	P
BACILLUS SP. PS3	157	DENSV
THERMUS THERMOPHI.	158	GGTTY
RHODOTHERMUS MAR.	160	AI
THIOBACILLUS FERRO	162	AI
E.COLI	166	AACDEFGGGHIIKLLLLLMMNPPPQQRRRRRSSSSSTTTTTTVVVWXXYYZ
SALMONELLA TYPHI.	170	HLPRR
AZOSPIRILLUM LIPO.	172	KTV
TRICHODESMIUM SPEC	173	AI
PHOTOBACT. PHOSPH.	174	HP
PHOTOBAC. LEIOGNA.	175	LM
AEROMONAS HYDROPH.	178	AEHILPR
PSEUDOMONAS AER.	182	AGITTY
PSEUDOMONAS FLUOR.	184	AI
CAMPYLOBAC.JEJUNI	186	AI
CAULOBACTER CRES.	190	AI
RHIZOBIUM MELILOTI	194	L
AZOARCUS SP.BH72	195	E
BORDETELLA PERTUS.	198	L
HAEMOPHILUS INFLU.	200	GKL
ANACYSTIS NIDULANS	210	AI
SYNECHOCYSTIS SP.	214	DWX
CYANOPHORA PARAD.	218	AGGILS
PYLAIELLA LITTORA.	222	AI
STREPTOCOCCUS PN.	224	A
ORGANELLES		
CHLOROPLASTS	240-359	
CYANOPHORA PARAD.	240	AI
PYLAIELLA LITTORA.	241	AI
CHLAMYDOMONAS REIN	244	ACDEGIRRTW
CHLAMYDOMO. MOEWU.	246	T
CHLORELLA ELLIPSO.	248	AIRS
LYCOPERSICON ESCU.	249	DLY
CUCUMIS SATIVUS	250	E
ASTASIA LONGA	251	ACDGIKLMPQRSSTV
EUGLENA GRACILIS	252	AACDEFGGHIKLLMNPQRRSSTVWXY
CRYPTOMONAS SPEC.	254	AIR
SPIROGYRA MAXIMA	255	I
ANTITHAMNION SP.	257	AI
CYANIDIUM CALDAR.	258	AIK
GUSTHODISCUS LUT.	259	AI
MARCHANTIA POLYM.	260	ACDEFGGHIKLLMNPPQRRRSSSTTVVWXY
CUSCUTA REFLEXA	261	AHILMY
COLEOCHAETE ORBIC.	262	AI
HORDEUM VULGARE	264	GGMSTVX
TRITICUM AESTIVUM	268	CDEGGMPRSTWXY
ORYZA SATIVA	270	ACCDEFGGHIILLMMMNPQRRSSSTTVVWY
ZEA MAYS	272	ACFGHILLMNPRSSSTVVWX
EPIFAGUS VIRGINIA.	274	LNR
ARABIDOPSIS THAL.	276	IMP
ALLIUM PORRUM	278	R
BRASSICA OLERACEA	280	L
GLYCINE MAX	284	AIMV
MEDICAGO SATIVA	288	H
NICOTIANA TABACUM	292	ACDEFGGHIIKLLMNPQRRSSSTTVVWXY
NICOTIANA DEBNEYI	296	H
OENOTHERA SP.	300	PW
DAUCUS CAROTA	301	V
GOSSYPIUM HIRSUTUM	302	H
PELARGONIUM ZONALE	304	R
PENNISETUM AMERICA	308	I
PETUNIA HYBRIDA	312	H
PHASEOLUS VULGARIS	316	H
HELIANTHUS ANNUUS	317	HNY
PISUM SATIVUM	320	DEGHIKLNPRRSTVWXY
PINUS THUNBERGII	322	ACDEFGGHIIKLLLMMPPQRRRSSSTTVVWXY
PINUS CONTORTA	323	HK
SINAPIS ALBA	324	HKQSY
SPINACIA OLERACEA	328	ACDEHILMRSSTTVY
SPIRODELA OLIGORH.	332	NRR
VICIA FABA	336	EFHLLTY
SORGHUM BICOLOR	340	L
MITOCHONDRIA	360-599	
SINGLE CELL ORGANISMS AND FUNGI	360-419	
PROTOTHECA WICKER.	360	ACDEFGGHIIKLLMNPQRRSSTVWXY
CHLAMYDOMO. REINH.	364	MQW
TRYPANOSOMA BRUCEI	368	AA
PARAMECIUM PRIM.	372	XY
PARAMECIUM TETRA.	375	WY
PARAMECIUM AURELIA	377	PWY
TETRAHYMENA PYRIF.	380	EFHLWX
TETRAHYMENA THERM.	384	LXY
ASPERGILLUS NIDUL.	388	ACCDEFGGHIKLLMMNPQRSSTVWXY
NEUROSPORA CRASSA	392	ACMR
PODOSPORA ANSERINA	396	DMRSVW
SACCHAROMYCES CER.	400	AACDEFGHIKLMMPQRRSSTTVVWXYY
SACCHAROMYCES EXI.	401	MP
PICHIA PIJPERI	402	LMM
WILLIOPSIS MRAKII	403	KLMPQSY
SCHIZOSACCHA.POM.	404	GHLPQ
KLUYVEROMYCES LAC.	405	CKLQ
CANDIDA PARAPSILO.	406	CEFGHIKLMPRRTVWY
HANSENULA WINGEI	407	CEGHLMPQRSTVW
TORULOPSIS GLAB.	408	ACDEFGHIKLMMPQRSSTTVWXY
WILLIOPSIS SUAVE	409	M
PICHIA JADINII	410	M
TRICHOPHYTON MENT.	409	AFLMMTV
TRICHOPHYTON RUBR.	412	MQR
PENICILLIUM CHRYS.	413	MRY
PLANTS	420-459	
ARABIDOPSIS THAL.	424	EMSSY
GLYCINE MAX	428	EMX
SOLANUM LYCOPERS.	430	C
SOLANUM TUBEROSUM	431	X
LUPINUS LUTEUS	432	GIKX
BRASSICA NAPUS	434	K
OENOTHERA SP.	436	CFGHILMPSSSWXY
PHASEOLUS VULGARIS	440	MSY
HELIANTHUS ANNUUS	441	CNV
TRITICUM AESTIVUM	444	CDEFKMPQGSSSWXY
ORYZA SATIVA	446	FHNPRSW
ZEA MAYS	448	CDEHKMMPSSWXY
MARCHANTIA POLYM.	450	ACDEFGGHIKLLLMMNPQRRRSSTVWY
ANIMALS	460-599	
FASCIOLA HEPATICA	462	ADIKMPSW
ASCARIS SUUM	464	ACDEFGHIKLLNPQRSSTVWXY
CAENORHABDIT.ELEG.	466	ACDEFGHIKLLNPQRSSTVWXY
MYTILUS EDULIS	470	ACDEFGHIKLLMMNPQRSSTVWY
ARTEMIA SP.	472	EFS
LOCUSTA MIGRATORIA	476	DGKLLPST
METRIDIUM SENILE	478	X
AEDES ALBOPICTUS	480	AEFGLNRSV
APIS MELLIFERA	482	ACDDEFGHIKLLMPQRSSVWY
DAPHNIA PULEX	483	IQVWXY
DROSOPHILA MELANO.	484	CDEKLWY
DROSOPHILA YAKUBA	488	ACDEFGHIKLLNPQRSSTVWXY
DROSOPHILA VIRILIS	496	IQX
CHORISTONEURA FUM.	497	L
PISASTER OCHRACEUS	498	ACDEGLLNPQTVWXY
ASTERINA PECTINI.	500	ACDGHILLMNPQSSVWY
CERATITIS CAPITATA	501	AEFNRS
ASTERIAS FORBESII	502	ACDGLLNVWXY
CYPRINUS CARPIO	503	ACDEFGHIKKLLNPQRSSTVWXY
PARACENTROTUS LIV.	504	ACDEFFGHIKLLNPQRSSTVWXY
ANOPHELES QUADRIM.	505	ACDEFGHIKLLNPQRSSTVWXY
RAINBOW TROUT	506	FPT
ANAS PLATYRHYNCOS	507	ACDEFKLNSWY
STRONGYLOCEN.PURP.	508	ACDEFGHIKLLNPQRSSTVWXY
ACIPENSER TRANSM.	509	PT
GADUS MORHUA	510	ACFGHKLMRSWY
ACANTHAMOEBA CAST.	511	ADEIKLMPQX
XENOPUS LAEVIS	512	ACDEFGHIKLLNPQRSSTVWXY
ALLIGATOR MISSIS.	513	ACNWY
CROCODYLUS POROSUS	514	ACNWY
CARETTA CARETTA	515	ACNWY
RANA CATESBEIANA	516	ACFILNPQTWXY
MALACLEMYS TERRA.	517	ACNWY
SPHENODON PUNCTAT.	518	ACNY
EPICRATES SUBFLA.	519	ACEN
CEPHALORHYN.COM.	520	FPT
CROSSOSTOMA LACUS.	521	ACDEFGHIKLLNPQRSSTVWXY
CHICKEN	522	ACDEFGHIKLLMNPQRSSTVWY
DIDELPHIS VIRGINI.	523	DPT
ODOCOILEUS HEMIO.	524	PT
DICEROS BICORNIS	525	FP
MARMOSA SP.	526	DPT
PHILANDER SP.	527	D
RAT	528	ACDDEFGHIKKLLMNNPPQQRSSSTTVWWXXY
METACHIRUS SP.	529	D
PHALANGER SP.	530	D
CHEIROMOPHORUS UM.	531	EPT
MOUSE	532	ACDEFGHIKLLNPQRSSTVWXY
CERVUS NIPPON	533	PT
BALAENOPTERA PHYS.	534	ACDEFGHIKLLPQRTVWXY
BALAENOPTERA MUSC.	535	ACDEFGHIKLLNPQRSTVWXY
BOVINE	536	ACDEFGHIKLLNPQRSSTVWXY
HALICHOERUS GRYPUS	537	ACDEFGLNPQRRSTVWXY
PHOCA VITULINA	538	ACDEFGHIKLLNPQRSTVWXY
GADUS MORHUA	539	DEFQST
GREEN MONKEY	540	F
SIAMANG	542	ACDGKNWXY
MACACA FUSCATA	544	HLS
MACACA MULATTA	548	HLS
MACACA FASCICULA.	552	HLS
MACACA SYLVANUS	556	HLS
SAIMIRI SCIUREUS	560	HLS
TARSIUS SYRICHTA	564	HLS
LEMUR CATTA	568	HLS
CHIMPANZEE	572	ACDEHIKLNSWXY
PYGMY CHIMPANZEE	573	ACDEIKNWXY
GIBBON	576	HLS
GORILLA	580	ACDEHIKLNSWXY
ORANG UTAN	584	ACDEHIKLNSWXY

HUMAN	588	ACDEEFGHIKLLMNPQRSSTVVWXY
ADPYCEROS MELAMPUS	590	PV
HIPPELAPHUS TRAGOC.	591	PV
CEPHALOPHUS MAXW.	592	PV
DAMALISCUS DORCAS	593	PV
GAZELLA THOMSONI	594	PV
OBISUS ELLIPSIPRYM.	595	PV
MADOQUA KIRKI	596	PV
ORYX GAZELLA	597	PV
TRAGELAPHUS IMBER.	598	PV
EUKARYOTIC CYTOPLASM	600-699	
SINGLE CELL ORGANISMS AND FUNGI	600-629	
PLASMODIUM FALSI.	603	AILMNRRTV
TRYPANOSOMA BRUCEI	605	EKENMDDRRRTVVY
TETRAHYMENA PYRIF.	606	MQS
LEISHMANIA TARENT.	609	GIKLQRRTVW
DICTYOSTELIUM DIS.	616	AERHKILMNQRRSSSTTVVWWY
PHYSARUM POLYCEPH.	618	X
NEUROSPORA CRASSA	620	FLR
CANDIDA ALBICANS	621	LQS
PHYTOPHTHORA PAR.	623	D
PODOSPORA ANSERINA	624	SS
SACCHAROMYCES CER.	625	AACDDEEFFGGHIKKKLLLMNPQQRRRSSSSSTTTVVWWXXY
SCHIZOSACCHA.POM.	633	ADEGHIKRRSSSVXX
CANDIDA CYLINDRA.	637	S
PLANTS	670-749	
CHLAMYDIA TRACHOM.	673	TW
ARABIDOPSIS THAL.	674	AFSSSSSSVWWXYYYY
GLYCINE MAX	680	DMX
PHASEOLUS VULGARIS	688	LNP
NICOTIANA RUSTICA	706	SSSSSSSYY
PETUNIA SP.	710	N
HELIANTHUS ANNUUS	712	L
SORGHUM BICOLOR	714	G
ORYZA SATIVA	719	G
TRITICUM AESTIVUM	720	Y
TRITICUM VULGARE	724	S
SOYBEAN	730	C
ANIMALS	750-999	
CAENORHABDI. ELEG.	756	DKLLPRRWXZ
BOMBYX MORI	758	AARGE
DROSOPHILA MELANO.	774	ADEEFGGHIKKLLMNPRRSSTVVXYZ
DROSOPHILA SIMUL.	780	S
SQUID	785	K
XENOPUS LAEVIS	792	AFELMVXXYYYZ
PODOCORYNE CARNEA	793	CPGSS
CHICKEN	804	AADDFKPPWZ
MOUSE	810	ACCDDGHIKELPPXZ
RAT	816	DDEEEEEEEEEFGGKLLLFP
BOVINE	829	RZ
HUMAN	869	EEGGGIKLLNNPPQQGRSSSSSTTTVVVVVVVVXXYY

PART TWO: tRNA Sequences

Source	Code	tRNA
VIRUSES	000-029	
AVIAN ONCO. VIRUS	010	M
CHICKEN ASVOAMV.RS	014	W
MOUSE M-MULV	016	RP
PHAGE T4	022	GILPQRST
PHAGE T5	026	DLNPQ
ARCHAEBACTERIA	030-109	
HALOBACTERIUM CUT.	038	AGHNDRSTVVVX
HALOPENAX VELICAMI	050	AAACDEEFGGGGHIIKKLLLLLMNPPQQRRRSSSTTVVWXY
HALOCOCCUS MORRHUA	054	X
METHANOBAC. THERM.	062	GN
SULFOLOBUS ACIDO.	082	X
THERMOPLASMA ACIDO	090	MX
EUBACTERIA	110-238	
MYCOPLASMA CAPRIC.	114	ACDERGHIKKLLLMNPQRRSSTTVWWXY
MYCOPLASMA MYCOID.	118	AGPSTYX
SPIROPLASMA CITRI	125	WW
STREPTOMYCES GRIS.	130	X
STREPTOMYCES COEL.	131	G
STAPHYLOCOC. EPID.	138	GG
MYCOBAC. SMEG.	142	X
BACILLUS STEARO.	146	RLVY
BACILLUS SUBTILIS	154	APGGGLMPRSSSTVWXYY
THERMUS THERMOPHI.	166	DFMXX
E.COLI	186	AAACDEEEPGGGHIIKKLLLMNQQPRRRRSSSSSTTVVVWXXYYZ
SALMONELLA TYPHI.	170	GGHLPP
RHODOBACIL. RUB.	202	RL
AGMENELLUM QUADR.	206	F
ANACYSTIS NIDULANS	210	LLX
SYNECHOCYSTIS SP.	214	E
ORGANELLES		
CHLOROPLASTS	240-359	
CHLAMYDOMONAS REIN	244	E
EUGLENA GRACILIS	252	F
CODIUM FRAGILE	253	GEMR
SCENEDESMUS OBLIQ.	256	MXY
LUPINUS ALBUS	263	Y
HORDEUM VULGARE	264	DDEQ
TRITICUM AESTIVUM	268	E
ZEA MAYS	272	I
GLYCINE MAX	284	LLL
NICOTIANA TABACUM	292	W
PHASEOLUS VULGARIS	316	RLLWX
SPINACIA OLERACEA	328	RLMPTVWX
MITOCHONDRIA	360-599	
SINGLE CELL ORGANISMS AND FUNGI	360-419	
TETRAHYMENA PYRIF.	380	FY
TETRAHYMENA THERM.	384	W
NEUROSPORA CRASSA	392	ALLTVWXY
SACCHAROMYCES CER.	400	PGHIKLMPRSSSTWXY
PLANTS	430-459	
SOLANUM TUBEROSUM	431	BL
OENOTHERA SP.	436	F
PHASEOLUS VULGARIS	440	RLLLMPWXY
ANIMALS	460-599	
ASCARIS SUUM	464	RMS
ARDEA ALBOPICTUS	480	DEGIKORSVX
HUMITER	534	DGRE
RAT LIVER	536	CDFKLLRVVW
BOVINE LIVER	536	DGIKLLRSSSTVWXX
HUMAN	568	S
EUKARYOTIC CYTOPLASM	600-699	
SINGLE CELL ORGANISMS AND FUNGI	600-629	
EUGLENA GRACILIS	604	DF
TETRAHYMENA THERM.	606	GGQX
SCENEDESMUS OBLIQ.	612	FXY
NEUROSPORA CRASSA	620	FX
SACCHAROMYCES CER.	628	ACDEFFGGHIKKLLLMNPPRRRSSSTTVVVWXY
SCHIZOSACCHA. POM.	632	EFXY
TORULOPSIS UTILIS	636	AILPVXY
CANDIDA CYLINDRA.	637	LLLRRRSS
PLANTS	670-749	
HORDEUM VULGARE	678	EEF
WHEAT GERM	682	FGKMRWXYY
BRASSICA NAPUS	686	F
LUPINUS LUTEUS	694	EFGHIMNP6VXY
PHASEOLUS VULGARIS	698	LLLLX
PISUM SATIVUM	702	F
SPINACIA OLERACEA	704	S
NICOTIANA RUSTICA	706	SSSSSYY
SOLANUM TUBEROSUM	707	LW
CUCUMIS SATIVUS	708	L
ANIMALS	750-999	
CAENORHABDI. ELEG.	756	L
ASTERINA AMURENSIS	762	X
BOMBYX MORI	768	AAFFGGI
DROSOPHILA MELANO.	774	RHIKSSSVVVXY
EUPHAUSIA SPIRIFA	786	X
XENOPUS LAEVIS	792	DFX
SALMON LIVER	798	X
CHICKEN	804	W
MOUSE	810	EFFRKKMQQRTVXZ
RAT	816	DDEKIKKLLNNGSSSVVX
RABBIT LIVER	822	DFEKKMY
BOVINE LIVER	828	DRLNQRRRTWYZ
CALF LIVER	834	F
COW MAMMARY GLAND	840	LL
SHEEP LIVER	846	HX
HUMAN	869	AAEFGGHLMNNQQSVXYY2

PART THREE: tRNA and rRNA gene sequences that differ from the conventional alignment

Source	Code	tRNA/rRNA gene
MITOCHONDRIA	360-599	
SINGLE CELL ORGANISMS AND FUNGI	360-419	
TRICHOPHYTON MENT.	403	E
ANIMALS	750-999	
APIS MELLIFERA	462	T
DAPHNIA PULEX	463	C
BALAENOPTERA PHYS.	534	MSS
BALAENOPTERA MUSC.	535	S
HALICHOERUS GRYPUS	537	K
PHOCA VITULINA	538	S
SIAMANG	542	S
CHIMPANZEE	572	S
PYGMY CHIMPANZEE	573	S
GORILLA	580	S
ORANG UTAN	584	S

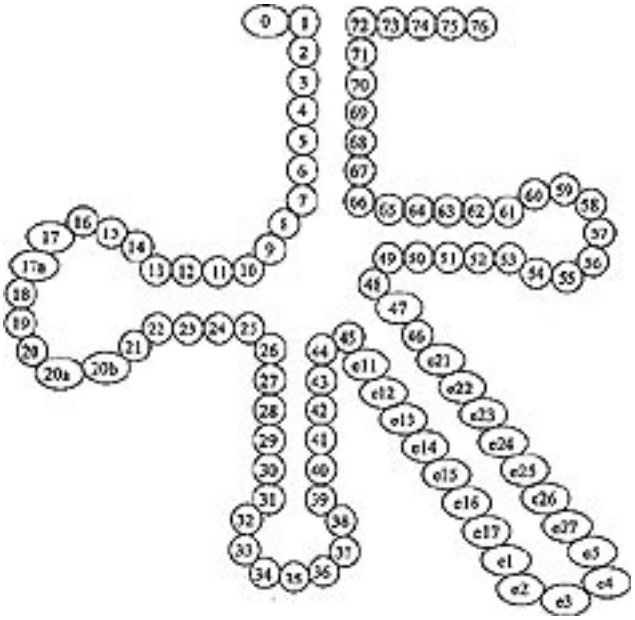

Figure 1. Numbering of nucleotides in tRNAs. Circles represent nucleotides which are always present; the ovals, nucleotides which are not present in each structure: these are nucleotides before the position 1 on the 5′-end, before and after the two invariant GMP residues 18 and 19 in the D-loop, and the nucleotides in the variable loop. The nucleotide to be added at a given site is indicated by the number of the preceding nucleotide followed by a colon and a letter in alphabetical order. The nucleotides in the variable stem have the prefix 'e' and are located between positions 45 and 46 obeying the base-pairing rules. The nucleotides in the 5′-strand and the 3′-strand are numbered by e11, e12, e13, ... and e21, e22, e23, ... respectively; and the second digit identifies the base-pair. In the case of a long variable region, the loop can be formed by up to five nucleotides: e1, e2, e3, e4 and e5.

modified nucleotides, the other ASCII signs are employed as defined in Table 3. Terminology and structure of the modified nucleosides occurring in tRNAs were used according to (2). Positions in particular sequence which are not filled (gaps in the generalised structure, Fig. 1) are indicated by a dash. All nucleotide insertions are denoted by underlining at the place of insertion.

Numbering and alignment of the variable region

The alignment of the variable region has been done in accordance with Steinberg and Kisselev (3). The extra arm is placed between nucleotides 45 and 46. The extra arm includes two double helical strands forming a stem and a loop. The annotations of the nucleotides in the extra arm positions begin with the letter 'e' (extra) followed by a one- or two-digit number. We have reserved a space for 7 base pairs in the stem and 5 nucleotides in the loop. The nucleotides in the loop are numbered from 1 to 5, whereas the nucleotides in the stem are numbered from 11 to 17 (5′-branch) and from 27 to 21, in reverse order, (3′-branch), to indicate base pair formation between nucleotides 11–21, 12–22, etc. (Fig. 1). In the tRNAs where the extra arm position 45 is empty but where the nucleotides 46–48 between the extra arm and T-domain are present, the positions will be filled in the order 48, 46, 47, i.e. tRNAs use position 48, 46 and 47 for the first, second and third nucleotide, respectively, depending on the length of the sequence in this region. A similar situation occurs in tRNAs without a long extra arm, where the most variable position 47 is deleted in many sequences.

Alignment of animal mitochondrial tRNAs

In properly aligned tRNA sequences, nucleotides occupying the same position in different tRNA sequences should play a comparable structural or functional role. Most animal mitochondrial tRNAs cannot be easily aligned with other tRNAs mainly because of the absence of information about their three-dimensional structure. Experimental data, however, point to the existence of tertiary interactions in these tRNAs. In this compilation, we use an alignment which accounts for these interactions as much as possible. Where we could do so, the animal mitochondrial tRNAs were included in Parts I and II. The alignment of animal mitochondrial tRNA is, however, not yet unambiguous.

Some animal mitochondrial tRNAs have completely unusual secondary structure and cannot be fitted in the tRNA alignment used here (Parts I and II). We treated these sequences separately including them in Part III. Here, each particular sequence has its own alignment. To this group belong the tRNAs from: (i) mitochondria of a parasitic worm lacking the T-, or D-domain; (ii) mitochondria of molluscs, insects and echinoderm, with extended anticodon and T-stems; (iii) mammalian mitochondria, lacking the D-domain.

For some tRNA genes the secondary structure pattern cannot be clearly established. We have also included these sequences in Part III. It is possible that post-transcriptional modifications of these tRNAs will result in improvement of the secondary structure.

ACKNOWLEDGEMENTS

This project was supported by Fonds der Chemischen Industrie, Deutsche Forschungsgemeinschaft, Project Sp 243/5-1, and Molekulargenetischer Arbeitskreis Rhein/Main e.V.

REFERENCES

1 Steinberg S., Misch A. and Sprinzl, M. (1993) *Nucleic Acids Res.* **21**, 3011–3015.
2 Limbach P.A., Crain P.F. and McCloskey, J.A. (1994) *Nucleic Acids Res.* **22**, 2183–2196.
3 Steinberg S.V. and Kisselev L.L. (1992) *Biochimie* **74**, 337–351.

Table 2. Format of tRNA sequences in EBI databank

```
*PART ONE: SEQUENCES OF tRNA-GENES
/
*                                            accept  |D-domain           ||anticodon domain ||variable region    ||T-domain        |accep
*                                            stem                                            |extra loop      |                     stem
*Number                                      012345678911111111111222222222222233333333334444444eeeeeeeeeeeeeeeeeeeee44445555555555666666666667777777
*      |Anticodon                                     01234567789000123456789012345678901234511111111123452222222678901234567890123456789012345 6
*      |   |Organism                |Kingdom|                 a   ab                            1234567      7654321
*
DA0260 TGC PHAGE T5                 VIRUS  -GGGCGAATAGTGTCAGC-GGG--AGCACACCAGACTTGCAATCTGGTA-------------------G-GGAGGGTTCGAGTCCCTCTTTGTCCACCA
+                                            ==*=*== ===*      *=========      ======                          =====     =======*=*==
DA0340 TGC ARCHAEGLOBUS FULG.  ARCHAE -GGGCTCGTAGCTCAGC--GGG--AGAGCGCCGCCTTTGCGAGGCGGAG-------------------GCCGCGGGTTCAAATCCCGCCGAGTCCA---
+                                            ==*==== ====      ==== =====      =====                           =====     =========*==
DA0380 TGC HALOBACTERIUM CUT.  ARCHAE -GGGCCCATAGCTCAGT--GGT--AGAGTGCCTCCTTTGCAAGGAGGAT-------------------GCCCTGGGTTCGAATCCCAGTGGGTCCA---
+                                            ==*==== *===      ===* =====      =====                           =====     =========*==
DA0420 TGC HALOBACTERIUM HAL.  ARCHAE -GGGCCCATAGCTCAGT--GGT--AGAGTGCCTCCTTTGCAAGGAGGAT-------------------GCCCTGGGTTGGAATCCCAGTGGGTCCA---
+                                            ==*==== *===      ===* =====      =====                           =====     =========*==
DA0580 TGC METHANOBAC.FORMI.   ARCHAE -GGGCCCGTAGCTCAGACTGGG--AGAGCGCCGCCCTTGCAAGGCGGAG-------------------GCCCCGGGTTCAAATCCCGGTGGGTCCA---
+                                            ==*===* ====      ==== =====      =====                           =====     =====*===*==
DA0620 TGC METHANOBAC.THERM.   ARCHAE -GGGCCCGTAGCTCAGACTGGG--AGAGCGCCGCCCTTGCAAGGCGGAG-------------------GCCCCGGGTTCAAATCCCGGTGGGTCCA---
+                                            ==*===* ====      ==== =====      =====                           =====     =====*===*==
DA0660 TGC METHANOCOC.VANI.    ARCHAE -GGGCCCGTAGCTCAGTT-GGG--AGAGCGCTGCCCTTGCAAGGCAGAG-------------------GCCGTGGGTTCAAATCCCGCCGGGTCCA---
+                                            ==*==== ====      ==== =====      =====                           =*===     ===*=====*==
DA0670 TGC METHANOTHRIX SOEH.  ARCHAE -GGGCTTGTAGCTCAGCT-GGT--AGAGCGCCGCCTTTGCAAGGCGGAG-------------------GCCCTGGGTCCGAATCCCAGCAAGTCCA---
+                                            ==*==== ====      ==== =====      =====                           =====     =========*==
DA0680 TGC METHANOTHERM. FER.  ARCHAE -GGGCCCATAGCTCAGCCTGGG--AGAGCGCCGCCCTTGCAAGGCGGAG-------------------GCCCCGGGTTCAAATCCCGGTGGGTCCA---
+                                            ==*==== ====      ==== =====      =====                           =====     =========*==
DA0780 TGC METHANOSPIR. HUNG.  ARCHAE -GGGCTCGTAGCTCAGCT-GGA--AGAGCGCGGCGTTTGCAACGCCGAG-------------------GCCTGGGGTTCAAATCCCCACGGGTCCA---
+                                            ==*=*== ====      ==== =====      =====                           =====     =======*=*==
DA0940 TGC THERMOCOCCUS CELER ARCHAE -GGGCCGGTAGCTCAGCCTGGG--AGAGCGTCGGCTTTGCAAGCCGAAG-------------------GCCCCGGGTTCGAATCCCGGCCGGTCCACCA
+                                            ==*==== ====      ==== =====      =====                           =====     =========*==
DA0980 TGC THERMOPROT. TENAX   ARCHAE -GGGCCGGTAGTCTAGC--GGA--AGGACGCCCGCCTTGCGCGCGGGAG-------------------ATCCCGGGTTCGAATCCCGGCCGGTCCA---
+                                            ==*==== ===*      *=== =====      =====                           =====     =========*==
DA0981 CGC THERMOPROT. TENAX   ARCHAE -GGGCCGGTAGTCTAGC--GGA--AGGACGCCCGCCTCGCGCGCGGGAG-------------------ATCCCGGGTTCGAATCCCGGCCGGTCCA---
+                                            ==*==== ===*      *=== =====      =====                           =====     =========*==
DA1140 TGC MYCOPLASMA CAPRIC.  EUBACT -GGGCCCTTAGCTCAGCT-GGG--AGAGCACCTGCCTTGCACGCAGGGG-------------------GTCGACGGTTCGATCCCGTTAGGGTCCACCA
+                                            ==*==== ====      ==== =====      =====                           *====     ====*====*==
DA1180 TGC MYCOPLASMA MYCOID.  EUBACT -GGGCCCTTAGCTCAGCT-GGG--AGAGCACCTGCCTTGCACGCAGGGG-------------------GTCGACGGTTCGATCCCGTTAGGGTCCACCA
+                                            ==*==== ====      ==== =====      =====                           *====     ====*====*==
DA1200 TGC MYCOPLASMA PNEU.     EUBACT -GGGGATGTAGCTCAACT-GAT--AGAGCACCTGATTTGCACTCAGGAG-------------------GTTGAGGGTTTGAGACCCTTCATCTCCACCA
+                                            ==*==== ====      ==== =====      =====                           *====     ====*====*==
DA1230 TGC ACHOLEPLASMA LAID.  EUBACT -GGGGCTTTAGCTCAGCT-GGG--AGAGCGCCTGCCTTGCACGCAGGAG-------------------GTCAGCGGTTCGAT-CCGCTAAGCTCCACCA
+                                            ==*==== ====      ==== =====      =====                           =====     =========*==
DA1231 TGC ACHOLEPLASMA LAID.  EUBACT -GGGGCCTTAGCTCAGCT-GGG--AGAGCGCCTGCCTTGCACGCAGGAG-------------------GTCAGCGGTTCGAT-CCGCTAGGCTCCACCA
+                                            ==*==== ====      ==== =====      =====                           =====     =========*==
DA1260 TGC SPIROPLASMA MELIF.  EUBACT -GGGCCCGTAGCTCAGCT-GGG--AGAGCACCTGCCTTGCACGCAGGGG-------------------GTCGACGGTTCGATCCCGTTCGGGTCCACCA
+                                            ==*==== ====      ==== =====      =====                           *====     ====*====*==
DA1280 TGC BORRELIA BURGDORF.  EUBACT -GGGGGTTTAGCTCAGTT-GGCT-AGAGCATCGGCTTTGCAAGCCGAGG-------------------GTCAAGGGTTCGAGTCCCTTAACCTCCA---
+                                            ==*==== ====      ==== =====      =====                           =====     =========*==
DA1470 TGC ENTEROCOCCUS HIRAE EUBACT -GGGGCCTTAGCTCAGCT-GGG--AGAGCGCCTGCTTTGCACGCAGGAG-------------------GTCAGCGGTTCGATCCCGCTAGGCTCCA---
+                                            ==*==== ====      ==== =====      =====                           =====     =========*==
DA1480 TGC STAPHYLOCOC. AURE.  EUBACT -GGGGGCTTAGCTCAGCT-GGG--AGAGCGCCTGCTTTGCACGCAGGAG-------------------GTCAGCGGTTCGATCCCGCTAGTCTCCACCA
+                                            ==*=*== ====      ==== =====      =====                           =====     =======*=*==
```

(Continued in the EBI databank. See text for instructions.)

Table 3. Modified nucleosides in tRNA and their abbreviations

one-letter code of nucleotides	symbol [2]	name [2]
U	U	uridine
C	C	cytidine
A	A	adenosine
G	G	guanosine
T	T	thymine (for sequences of tRNA genes only)
-		empty position
_	(underline)	insertion (see footnote for further information)
.		unknown nucleotide
H	?A	unknown modified adenosine
"	m1A	1-methyladenosine
/	m2A	2-methyladenosine
+	i6A	N^6-isopentenyladenosine
*	ms2i6A	2-methylthio-N^6-isopentenyladenosine
=	m6A	N^6-methyladenosine
6	t6A	N^6-threonylcarbamoyladenosine
E	m6t6A	N^6-methyl-N^6-threonylcarbamoyladenosine
[	ms2t6A	2-methylthio-N^6-threonylcarbamoyladenosine
:	Am	2'-O-methyladenosine
I	I	inosine
O	m1I	1-methylinosine
^	Ar(p)	2'-O-ribosyladenosine (phosphat)
`	io6A	N^6-(cis-hydroxyisopentenyl)adenosine
<	?C	unknown modified cytidine
%	s2C	2-thiocytidine
B	Cm	2'-O-methylcytidine
M	ac4C	N^4-acetylcytidine
?	m5C	5-methylcytidine
'	m3C	3-methylcytidine
}	k2C	lysidine
>	f5C	5-formylcytidin
;	G	unknown modified guanosine
K	m1G	1-methylguanosine
L	m2G	N^2-methylguanosine
#	Gm	2'-O-methylguanosine
R	m22G	N^2,N^2-dimethylguanosine
|	m22Gm	N^2,N^2,2'-O-trimethylguanosine
7	m7G	7-methylguanosine
(	fa7d7G	archaeosine
Q	Q	queuosine
8	manQ	mannosyl-queuosine
9	galQ	galactosyl-queuosine
Y	yW	wybutosine
W	o2yW	peroxywybutosine
N	?U	unknown modified uridine
{	mnm5U	5-methylaminomethyluridine
2	s2U	2-thiouridine
J	Um	2'-O-methyluridine
4	s4U	4-thiouridine
&	ncm5U	5-carbamoylmethyluridine
1	mcm5U	5-methoxycarbonylmethyluridine
S	mnm5s2U	5-methylaminomethyl-2-thiouridine
3	mcm5s2U	5-methoxycarbonylmethyl-2-thiouridine
V	cmo5U	uridine 5-oxyacetic acid
5	mo5U	5-methoxyuridine
!	cmnm5U	5-carboxymethylaminomethyluridine
$	cmnm5s2U	5-carboxymethylaminomethyl-2-thiouridine
X	acp3U	3-(3-amino-3-carboxypropyl)uridine
,	mchm5U	5-(carboxyhydroxymethyl)uridinemethyl ester
)	cmnm5Um	5-carboxymethylaminomethyl-2'-O-methyluridine
~	ncm5Um	5-carbamoylmethyl-2'-O-methyluridine
D	D	dihydrouridine
P	Ψ	pseudouridine
]	m1Ψ	1-methylpseudouridine
Z	Ψm	2'-O-methylpseudouridine
T	m5U	ribosylthymine
F	m5s2U	5-methyl-2-thiouridine
\	m5Um	5, 2'-O-dimethyluridine