Introns

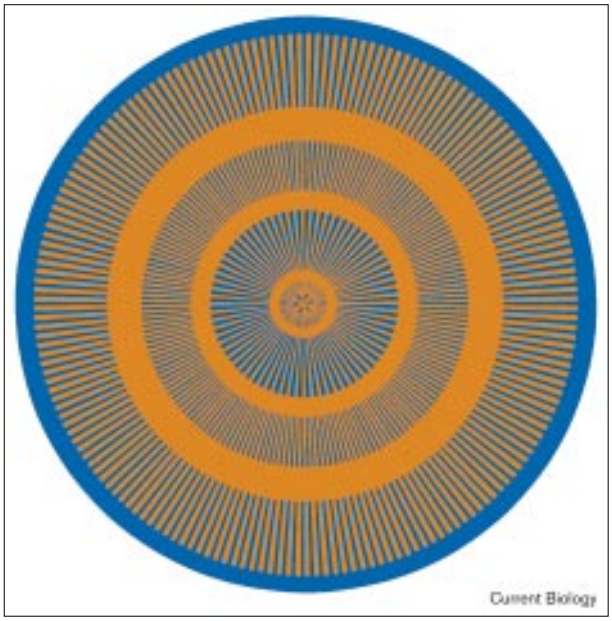
Illusions as neuro-signs

Nicholas Wade

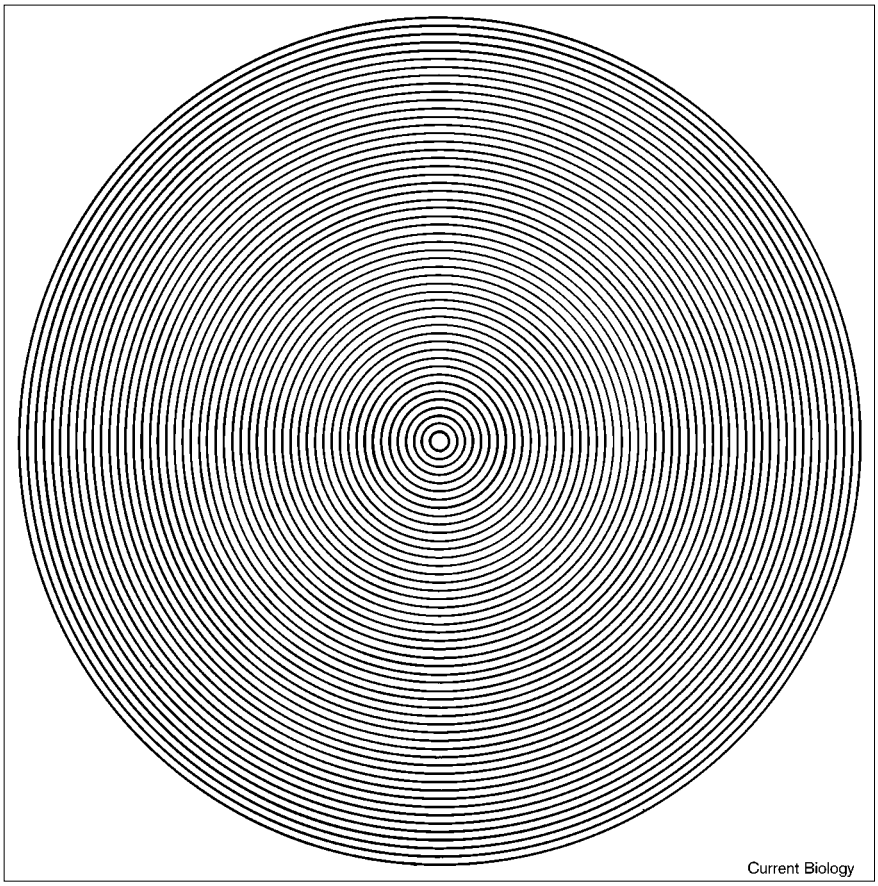
Almost the whole of visual science is concerned with illusions — if they are defined as images in which the physical description differs from the image perceived. One of the attractions of illusions, apart from their obvious appeal as geometrical abstract art, is that they direct us to an understanding of visual processing at many levels.

But before the question of why an illusion is perceived can be addressed, we must have a physical description of the image, which is not always straightforward. After all, what is the physical description of a stick that is partially immersed in water? It is straight if the stick itself is measured, but not if a photograph is taken of it.

For this reason (among others) psychologists make a distinction between the distal (or physical) stimulus and the proximal stimulus (its projection onto the retina). An illusion occurs when there is a mismatch between the proximal stimulus and perception.



The distinction between distal and proximal stimuli can be appreciated with the image of black and white concentric circles below. The distal stimulus is circular and stable, but its projection onto the retina is distorted and dynamic. Spokes appear to radiate from the centre and to rotate; they are probably a consequence of slight astigmatic changes in the curvature of the lens, because they are reduced when viewed through a pinhole, they have similar temporal dynamics to measured fluctuations in the curvature of the lens, and they do not occur in those who have no crystalline lens (aphakia). Aphakic individuals lacking a lens in one eye see such patterns as stable using that eye, but see it with radiating spokes using the eye that has a lens.



In addition, viewing the pattern from a distance will reveal a portrait of Jan Evangelista Purkinje, who first described the radiating spokes phenomenon in 1825. The portrait is carried by low spatial frequency variations in the pattern, which correspond to the coarse grain over the whole pattern; these are normally inhibited by the higher spatial frequencies (the fine grain of the thin lines) when the pattern is close to the eye. When viewed from a distance, the high spatial frequency content (thin lines) is not sharply resolved and the low spatial frequency pattern emerges.

Purkinje also demonstrated similar effects with radiating lines (see coloured image above). High-contrast coloured or black and white radiating lines are also visually dynamic — they appear to move.

The motion consists of shimmering points which are perpendicular to the lines of the distal stimulus. Thus, with a radiating pattern they are circular, and can be seen more readily within the orange annuli. If a white surface is observed following fixation on the pattern for around 30 seconds then scintillating dots will seem to be streaming in a direction orthogonal to the previously presented lines.

Regular geometrical patterns produce a range of visual effects; some are based on the optics of the eye (like the spokes) and others (such as the scintillations) could be caused by visual processing in the brain. For example, Purkinje's radiating lines illusion has been examined with the aid of neural imaging. Positron emission tomography (PET) scans indicate that a pattern of radiating lines causes the blood flow in V5 — the area of the brain that is involved in motion processing — to increase. Alternative interpretations, however, relate to eye movement instabilities during fixation. It is likely that several phenomena are induced by these patterns, with a different explanation for each one.

When we know that an image is distorted before it reaches the retina — that the proximal stimulus differs from the distal stimulus — we consider that we have some interpretation of the illusion. Thus, the bent-stick distortion is due to differential refraction through air and water. But in general, explanations for the illusion are sought in higher levels of visual processing. Such illusions might provide an index of neural function, which I have called a 'neuro-sign'.

Address: Department of Psychology, University of Dundee, Dundee DD1 4HN, UK.

This is the first in a short series of articles by Nicholas Wade, describing some common visual illusions.

Correspondence

A bacterial antibiotic resistance gene with eukaryotic origins

James R. Brown*, Jianzhi Zhang† and John E. Hodgson‡§

The rapid emergence of antibiotic-resistant pathogens is mainly due to the horizontal transfer of resistance genes among different bacterial strains or species. No transfers of antibiotic resistance genes have been found to occur between bacteria and eukaryotes, however [1]. Here, we show that a gene known to convey antibiotic resistance in bacteria was apparently transferred from an early evolved eukaryote.

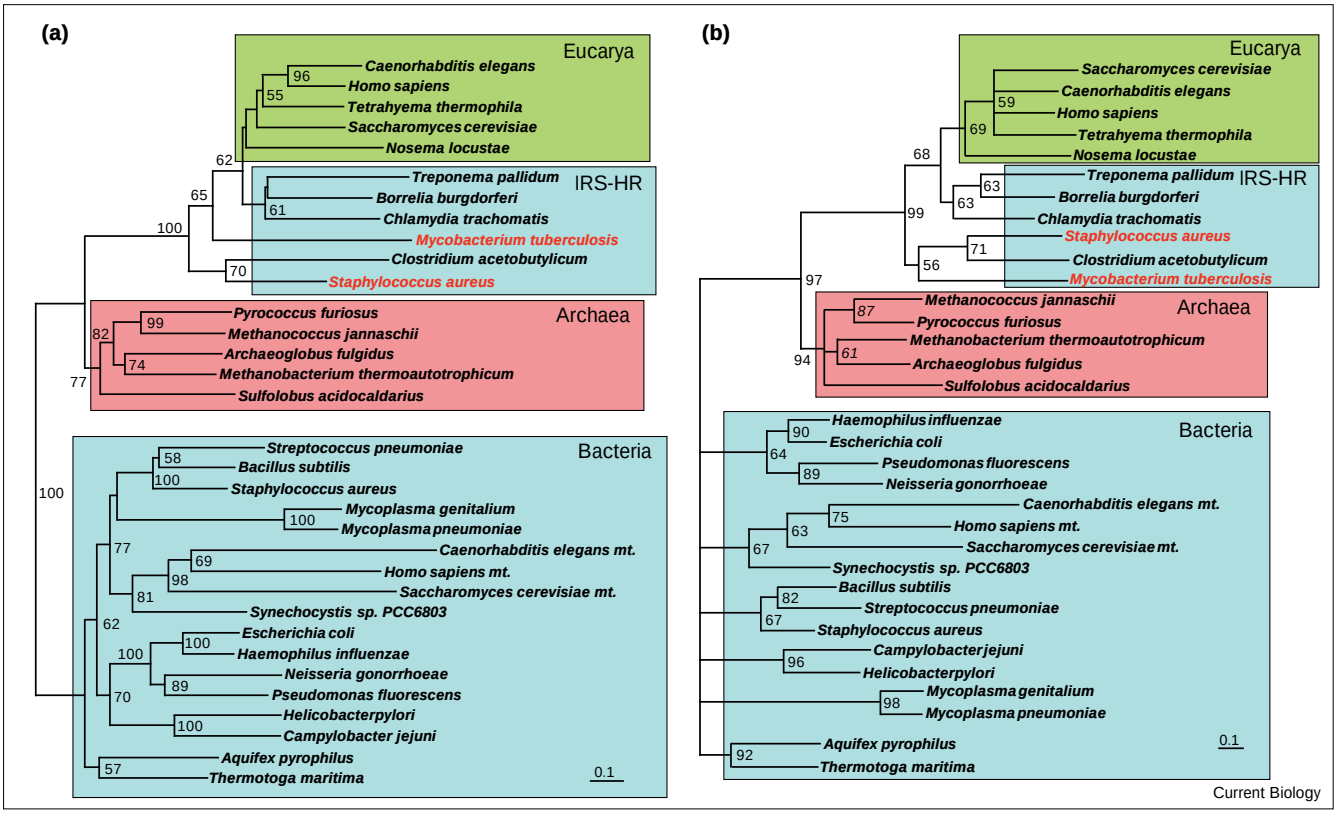
Mupirocin, a topical agent widely used since 1985 against skin surface and nasal colonization by the Gram-positive bacterium *Staphylococcus aureus*, is the natural compound pseudomonic acid produced by the bacterium *Pseudomonas fluorescens* NCIB 10586 [2]. Pseudomonic acid inhibits protein synthesis by reversibly binding to the active site of isoleucyl-tRNA synthetase (IRS) [3,4]. Sensitivity of *S. aureus* to mupirocin use *in vitro* occurs at a minimal inhibitory concentration (MIC) range of 0.12 to 1.0 mg/L [5,6]. Moderate mupirocin resistance in *S. aureus*, broadly defined as an MIC range of 8 to 256 mg/L, appears to be mediated by different single amino-acid substitutions in IRS, either within or near an ATP-binding motif with the consensus amino-acid sequence KMSKS (in the single amino-acid code) [7]. First observed clinically in the late 1980s, this level of resistance has little impact on the topical use of mupirocin.

In the early 1990s, much higher levels of mupirocin resistance in *S. aureus* appeared in some clinical settings, where estimated MICs exceed 512 mg/L [5]. Highly resistant strains have a second copy of the IRS gene (IRS-HR), located on an extrachromosomal plasmid [8], the product of which is highly divergent (34% amino-acid identity) from those of sensitive or moderately resistant IRS genes (IRS-S/MR) [9,10]. More recently, another Gram-positive bacterium, *Mycobacterium tuberculosis*, was shown to be highly resistant to mupirocin, and the product of its chromosome-located IRS gene had higher sequence similarity to *S. aureus* IRS-HR proteins than IRS-S/MR proteins [11].

Using the *S. aureus* IRS-HR sequence as a probe against public databases [12], we discovered four more bacterial IRS-HR-type genes. Phylogenetic analyses (Figure 1) show that all six bacterial IRS-HR proteins cluster significantly with eukaryotic IRS proteins that are also highly resistant to mupirocin [3,4]. The present IRS tree contradicts an earlier phylogeny, which showed the Bacteria, Archaea and Eucarya as separate monophyletic groups, with the latter two being sister groups [13]. We suggest that bacterial IRS-HR genes originated from eukaryotes through a horizontal gene transfer that occurred after the Eucarya and Archaea separated. A eukaryotic origin of bacterial IRS-HR genes is also supported by an insertion with the consensus sequence HYPFE, which is shared with eukaryotic IRS proteins but not bacterial IRS-S/MR proteins (Figure 2). Amino-acid compositions are significantly different between the IRS-HR and IRS-S/MR proteins of *S. aureus* (data not shown), further suggesting that the two genes originated from different organisms. A similar gene transfer has been suggested for glutaminyl-tRNA synthetases, for which certain proteobacteria have a eukaryote-like gene [14,15].

The deep phylogenetic position of bacterial IRS-HR proteins relative to eukaryotic IRS proteins suggests that a recent transfer of the

Figure 1



Phylogenies of isoleucyl-tRNA synthetases (IRSs) using **(a)** neighbor-joining (NJ) [17] and **(b)** maximum likelihood (ML) [18] methods. Those bacterial IRSs suggested to be highly mupirocin resistant are indicated (HR), and those species in which it has been confirmed by experiments are in red. Trees are based on the non-gapped multiple sequence alignment of amino acids flanked by class 1 aminoacyl-tRNA synthetase signature motifs HIGH and KMSKS (438 amino acids). The NJ tree was generated with the programs PROTDIST (using the Dayhoff PAM120 option) and NEIGHBOR of the PHYLIP package [19]. Maximum log likelihood of the ML tree (using the program PUZZLE 3.0.1 [18] with the JTT option) was –23,862.83 (± 514.99). Maximum parsimony (MP) analysis (using the program PAUP 3.1.1 [20]) also clustered IRS-HR sequences with eukaryotes with one minimal length tree (4295 steps) found after 100 random heuristic searches. Numbers represent the percent occurrence of branching points in 1000 bootstrap replications (NJ) or 1000 puzzling quartets (ML). Values less than 50% are not shown, and in the ML tree those nodes are collapsed. The scale bar represents an estimated 0.1 amino-acid substitutions per site. Mitochondrial targeted isoforms of eukaryotes are indicated by mt.

resistance gene from humans or animals as a consequence of intensive clinical use of mupirocin is an unlikely scenario. Plasmids mediating high-level resistance were detected in staphylococci isolated nearly twenty years before the clinical use of mupirocin [16]. The existence of mupirocin-sensitive *S. aureus* and plasmid-borne IRS-HR genes, however, indicates that IRS-HR is not fixed in the population. To reconcile these observations, we propose that a eukaryotic IRS gene was first transferred to an unknown bacterium shortly after the divergence of the Eucarya and Archaea. The gene might have been transferred through several bacteria species and only recently was it passed to *S. aureus*.

Since mupirocin (as pseudomonic acid) is a natural product of some widely found strains of *P. fluorescens*, resistance to this compound might have conferred some competitive advantages to conspecific organisms, thus driving the fixation of IRS-HR genes in different bacterial species, though the possibility of random fixation cannot be excluded. To our knowledge, *Treponema pallidum*, *Borrelia burgdorferi*, *Clostridium acetobutylicum* and *Chlamydia trachomatis* have not been tested for susceptibility to mupirocin, probably because these organisms are not involved in epidermal infections and several are intracellular, obligate human parasites. The present analysis, however, strongly suggests these organisms to be resistant to mupirocin, as their IRS proteins clearly cluster with IRS-HR proteins from species with proven resistance. Our study shows the possibility of horizontal transfers of antibiotic resistance genes from eukaryotes to prokaryotes and highlights the potential of phylogenetic analyses in medical studies.

Figure 2

```
Bacteria
            T. maritima  EKMLDTLDVWIDSGASFEYIT-------TKREDHPFPLDMYLEGSDQHRG
         A. pyrophlius  RKEEDILDVWFDSGCSHASV---------IRPLGFEKADLYLEGSDQHRG
             C. jejuni  EKVYDILDVWFDSGSTFNAVL-----NSGLYDAGEKRASMYLEGSDQHRG
              H. pylori  EKIMHILDVWFDSGSTFKAVL-----EDYHGEKGQSPSDVILEGSDQHRG
         M. genitalium  HKEIDTLDVWFDSGSSYNVL---------EINKYGSIADLYIEGSDQYRG
         M. pneumoniae  KKETDTLEVWFDSGSTYNVL---------ISNKLNFPADLYLEGSDQYRG
         Synechocystis  RKGEDTMDVWFDSGSSWAAVA------NAKNRPLKYPVDMYLEGSDQHRG
        P. fluorescens  DKISDTLDVWFDSGTTHWHVL---RGSHPMGHETGPRADLYLEGSDQHRG
     S. cerevisiae mt.  CRSQDTMDVWFDSGSSWSVIKDFYEKSLKLSKLPSPLYQVCLEGSDQHRG
        C. elegans mt.  EKNTDIMDVWLDSGLAWHAAR-------DNDTEREHVADVVLEGVDQFRG
        H. sapiens mt.  VPGQDILDIWFDSGTSWSYV----------LPGPDQRADLYLEGKDQLGG
         H. influenzae  RKVPDTLDVWFDSGSTYSSVV------ANRLEFNGQDIDMYLEGSDQHRG
                E. coli  VKVPDTLDVWFDSGSTHSSVV------DVRPEFAGHAADMYLEGSDQHRG
        N. gonorrhoeae  DKLPDTMDVWFDSGSTHYSVV-------KQREELEWPADLYLEGSDQHRG
           B. subtilis  TKEQDIMDVWFDSGSSHQAVL-------EERDDLVRPADLYLEGSDQYRG
         S. pneumoniae  KKETDIMDVWFDSGSSWNGVV-------VNRPELTYPADLYLEGSDQYRG
             S. aureus  TKETDIMDVWFDSGSSHRGVL-------ETRPELSFPADMYLEGSDQYRG
IRS-HR
             S. aureus  SRVEEVIDVWFDSGSMPFAQHHYPFD-NQKIFNQHFPADFIAEGVDQTRG
       C. acetobutylicum  TRTEEVIDCWFDSGSMPFAQLHYPFE-NKEVFENTFPAQFISEAVDQTRG
          M. tuberculosis  RRIPDVLDVWFDSGSMPYAQVHYPFE-NLDWFQGHYPGDFIVEYIGQTRG
             T. pallidum  RRVPEVLDCWFESGAMPYAQQHYPFE-HATDFERYFPAHFISEGLDQTRG
           B. burgdoferi  IRTSEVLDCWFESGAMPYASNHYPFT-NEINFKNIFPADFIAEGLDQTRG
          C. trachomatis  RRIPYVFDCWFDSGAMPYAQNHYPFE-RAEETEACFPADFIAEGLDQTRG
Eucarya
            S. cerevisiae  KRIEEVFDCWFESGSMPYASQHYPFE-NTEKFDERVPANFISEGLDQTRG
           T. thermophila  RRIDEVFDCWFESGSMPYGQQHYPFSMNEEEFSKRFPADFIGEGIDQTRG
               H. sapiens  HRISEVFDCWFESGSMPYAQVHYPFE-NKREFEDAFPADFIAEGIDQTRG
               C. elegans  KRVSEVFDCWFESGSMPYAQNHYPFE-NRKIFEDNFPADFIAEGIDQTRG
              N. locustae  RRIEEVFDCWFESGCMPYAQRHWPFE----CDNLCLPADFVAEGVDQTRG
Archaea
     M. thermoautotrophicum  KRTPDVLDVWIDSGVAGWAALHYPRE--KELFSEWFPYDFITEGHDQTRG
             M. jannaschii  KRVPDVLDVWFDSGLAPYASIGV---------KELKKADFITEGHDQVTK
               P. furiosus  RRVKDVVDVWFDSGIASWASLGYPR--NKELFEKLWPADFIVEGEDQVTK
               A. fulgidus  RRVPDVFDVWFDSGVASWGSIAYPL--RKDKFEELWPADFITEGHDQTRG
           S. acidocaldarius  RRISDVADVWFDSGVAFFASLGQDW--RKR-WSELGPVDLVLEGHDQLRG
                              518                                            566
```

Current Biology

Partial amino-acid sequence alignment showing the consensus amino-acid motif HYPFE (in red) uniquely shared among bacterial IRS-HR and eukaryotic IRS. Numbering corresponds to the first amino acid in human IRS. The di-amino-acid motif DQ (blue) potentially functions in binding the activated amino acid isoleucine-AMP [21].

Supplementary material
An alignment of multiple IRS amino-acid sequences is published with this article on the internet.

Acknowledgements
We thank K. Coleman, R. Jarvest, E. Lawlor, A. Lupas, C. Perry and J. Poupard for their comments. J.Z. was supported by NIH and NSF research grants to M. Nei.

References

1. Courvalin P: The Garrod Lecture. Evasion of antibiotic action by bacteria. *J Antimicrob Chemother* 1996, 37:855-869.
2. Fuller AT, Mellows G, Woolford M, Banks GT, Barrow KD, Chain EB: Pseudomonic acid: an antibiotic produced by ***Pseudomonas fluorescens***. *Nature* 1971, 234:416-417.
3. Hughes J, Mellows G: Interaction of pseudomonic acid A with ***Escherichia coli*** B isoleucyl-tRNA synthetase. *Biochem J* 1980, 191:209-219.
4. Farmer TH, Glibart J, Elson SW: Biochemical basis of mupirocin resistance in strains of ***Staphylococcus aureus***. *J Antimicrob Chemother* 1992, 30:587-596.
5. Poupard JA: Update on mupirocin resistance. *J Chemother* 1995, 7:71-74.
6. Sutherland R, Boon RJ, Griffin KE, Masters PJ, Slocombe B, White AR: Antibacterial activity of mupirocin (pseudomonic acid), a new antibiotic for topical use. *Antimicrob Agents Chemother* 1985, 27:495-498.
7. Yanagisawa T, Lee JT, Wu HC, Kawakami M: Relationship of protein structure of isoleucyl-tRNA synthetase with pseudomonic acid resistance of ***Escherichia coli***. *J Biol Chem* 1994, 269:24304-24309.
8. Gilbart J, Perry CR, Slocombe B: High-level mupirocin resistance in ***Staphylococcus aureus***: evidence for two distinct isoleucyl-tRNA synthetases. *Antimicrob Agents Chemother* 1993, 37:32-38.
9. Dyke KGH, Curnock SP, Golding M, Noble WC: Cloning of the gene conferring resistance to mupirocin in ***Staphylococcus aureus***. *FEMS Microbiol Lett* 1991, 77:195-198.
10. Hodgson JE, Curnock SP, Dyke KG, Morris R, Sylvester DR, Gross MS: Molecular characterization of the gene encoding high-level mupirocin resistance in ***Staphylococcus aureus*** J2870. *Antimicrob Agents Chemother* 1994, 38:1205-1208.
11. Sassanfar M, Kranz JE, Gallant P, Schimmel P, Shiba K: A eubacterial ***Mycobacterium tuberculosis*** tRNA synthetase is eukaryote-like and resistant to a eubacterial-specific antisynthetase drug. *Biochemistry* 1996, 35:9995-10003.
12. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: Basic local alignment search tool. *J Mol Biol* 1990, 215:403-410.
13. Brown JR, Doolittle WF: Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc Natl Acad Sci USA* 1995, 92:2441-2445.
14. Lamour V, Quevillon S, Diriong S, N'Guyen VC, Lipinski M, Mirande M: Evolution of the Glx-tRNA synthetase family: the glutaminyl enzyme as a case of horizontal gene transfer. *Proc Natl Acad Sci USA* 1994, 91:8670-8674.
15. Brown JR, Doolittle WF: Gene descent, duplication and horizontal transfer in the evolution of glutamyl- and glutaminyl-tRNA synthetases. *J Mol Evol* 1998, in press.
16. Rahman M, Connolly, S, Noble WC, Cookson B, Phillips I: Diversity of staphylococci exhibiting high-level resistance to mupirocin. *J Med Microbiol* 1990, 33:97-100.
17. Saitou N, Nei M: The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 1987, 4:406-425.
18. Strimmer K, von Hasseler A: Quartet puzzling: a quartet maxiumum likelihood method for reconstructing tree topologies. *Mol Biol Evol* 1996, 13:964-969.
19. Felsenstein J: PHYLIP: Phylogeny Inference Package Version 3.57c. Department of Genetics, University of Washington, Seattle 1993; http://evolution.genetics.washington.edu/phylip.html.
20. Swofford DL: PAUP: Phylogenetic analysis using parsimony Version 3.1.1. Illinois Natural History Survey, Champaign 1993.
21. Doublié S, Bricogne G, Gilmore C, Carter CW Jr: Tryptophanyl-tRNA synthetase crystal structure reveals an unexpected homology to tyrosyl-tRNA synthetase. *Structure* 1995, 3:17-31.

Addresses: *Department of Bioinformatics, SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, UP1345, PO Box 5089, Collegeville, Pennsylvania 19426-0989, USA. †Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA. ‡Department of Microbiology, Anti-Infectives Research Division, SmithKline Beecham Pharmaceuticals, 1250 South Collegeville Road, UP1345, PO Box 5089, Collegeville, Pennsylvania 19426-0989, USA. §Present address: Hoechst Marion Roussel, 102 route de Noisy, 93235 Romainville Cedex, France.

Correspondence: James R. Brown
E-mail: James_R_Brown@sbphrd.com

Supplementary material

A bacterial antibiotic resistance gene with eukaryotic origins [Correspondence]

James R. Brown, Jianzhi Zhang and John E. Hodgson

Current Biology 21 May 1998, 8:365–367

Alignment of multiple isoleucyl-tRNA synthetase (IRS) amino-acid sequences

Columns marked with * were used in the phylogenetic analyses. Abbreviations correspond to IRS sequences from the following species with their GenBank accession numbers or source: ITma (*Thermotoga maritima*, L37104), IApy (*Aquifex pyrophilus*, L37096), ICje (*Campylobacter jejuni*, U15295), IHpy (*Helicobacter pylori*, AE000642), ISau (*Staphylococcus aureus*, X74219), ISaR (*Staphylococcus aureus Mup R* gene, X75439), ITpa (*Treponema pallidum*, The Institute for Genomic Research, personal communication), IBbu (*Borrelia burgdorferi*, AE001181), ICtr (*Chlamydia trachomatis*, Chlamydia Genome Project, personal communication), IMge (*Mycoplasma genitalium*, U39718), IMpn (*Mycoplasma pneumoniae*, 2501010), IMtu (*Mycobacterium tuberculosis*, Z74020), IBsu (*Bacillus subtilis*, Z99112), ICac (*Clostridium acetobutylicum*, Genome Therapeutics personal communication), ISpn (*Streptococcus pneumoniae*, The Institute for Genomic Research, personal communication), ISyn (*Synechocystis* sp. PCC6803, D90907), IPfl (*Pseudomonas fluorescens*, X80132), ISmt (*Saccharomyces cerevisiae* mitochondria (mt.), L38957), ICmt (*Caenorhabditis elegans* mt., Z81038), IHmt (*Homo sapiens* mt., assembled from NCBI dbEST), IHin (*Haemophilus influenzae*, U32777), IEco (*Escherichia coli*, D10483), INgo (*Neisseria gonorrhoeae*, University of Oklahoma's Advanced Center for Genome Technology, personal communication), ISce (*Saccharomyces cerevisiae*, M19992), ITth (*Tetrahymena thermophila*, M30942), IHsa (*Homo sapiens*, U04953), ICel (*Caenorhabditis elegans*, Z70310), INlo (*Nosema locustae*, L37097), IMth (*Methanobacterium thermoautotrophicum*, M59245), IMja (*Methanococcus jannaschii*, U67538), IPfu (*Pyrococcus furiosus*, L37105), IAfu (*Archaeoglobus fulgidus*, AE001061), ISac (*Sulfolobus acidocaldarius*, L37106).

```
       *****************            *******************************
#ITma DTHGLPIEHRVSQELGEKI----KEMSP-AEIRKKCEEFALRFVDIQREEFKRLGVRGDW
#IApy DTHGLPIERAVEKELSKKKIR--KESLPKTEFRKLCREYANRYVNIQKEEFIRLGVLGDW
#ICje DCHGLPIEQQVEVKLGEK-----KKSLSKKEIREFCRQHASEFVDIQREEFKNLGIIADW
#IHpy DCHGLPIEQQILERLEKEK----TSLENPTLFREKCRDHAKKFLEIQKNEFLQLGVLGDF
#ISau DTHGLPIEQALTKKGVD------RKKMSTAEFREKCKEFALEQIELQKKDFRRLGVRGDF
#ISaR DTHGLPVELEVEKKIGIKGKQ-DIEKYGIENFINECKKSVFNYEKEWRDFSKDLGYWVDM
#ITpa DCHGLPIEHLIEQELNLNSKS-DVESYGVSAFNAACRSSVLRYVKEWQRTLTRLGRWVDF
#IBbu DTHGLPVEYEVEKKLGISGKY-EIENYGIENFNKECRKIVLRYTEEWKNIILRLGRWVDF
#ICtr DCHGVPVEYEVEKSLGLTEPG-AIERFGVANFNEECRKIVFRYADEWKYFVDRIGRWVDF
#IMge DCHGLPIEHAVSKKNPSS-----YSNLSTVEKRKLCHQFALSQIAVQKEQFQRLGLLNDF
#IMpn DCHGLPIEHAVSKKDPQH-----YASLSLSEKRDLCKQFALSQIAIQKAQFQRLGLLNDF
#IMtu DTHGLPAELEVERQLGITDKSQ-IEAMGIAAFNDACRASVLRYTDEWQAYVTRQARWVDF
#IBsu DTHGLPIETALTKNKKVN-----RKEMSVAEFRKLCEEYAWKQIEGQREQFKRLGVRGDW
#ICac DTHGLPVELEIEKKLGISGKP-QIEEYGIEKFVKECKESVFSYVSLWKDMSEKLGYWVDM
#ISpn DTHGLPIEQVLSKQGVKR------KEMDLVEYLKLCREYALSQVDKQREDFKRLGVSGDW
#ISyn DCHGLPIELKVLQSLKSSER----AELTPLTLRHKARDFALKAQQEQAVGFQRYGIWGDW
#IPfl DCHGLPIEHKVEVTYGK--------NLGADKTRELCRAYATEQIEGQKSEFIRLGVLGEW
#ISmt DCHGLPIEIKALKDLSAQQ----IESISPLKIRSMALKHAQKAIKRQRETFQHFAILTDW
#ICmt DCHGLPIELKIGKQQGNAG------SRTPLETRAAARIVADEAIGKQMNAFRRWGVTADW
#IHmt DCHGLPIEIKVLSELGREA-----QNLSAMEIRKKARSFAKAAIEKQKSAFIRWGIMADW
#IHin DCHGLPIELKVEGLVGKPN-----EKISAAEFRQKCREYAAEQVEGQKKDFIRLGVLGDW
#IEco DCHGLPIELKVEQEYGKPG-----EKFTAAEFRAKCREYAATQVDGQRKDFIRLGVLGDW
#INgo DCHGLPIEVMVEKLHGKDM--------PKARFRELCREYAAEQIARQKKDFIRLGVLGDW
#ISce DTHGVPIEHIIDKKLGITGKDDV-FKYGLENYNNECRSIVMTYASDWRKTIGRLGRWIDF
#ITth DCHGLPVEYEIDKKLGITNRQEV-LKMGVDKYNAECRSIVMRYAQEWRSIVNRFGRWVDF
#IHsa DCHGLPVEYEIDKTLGIRGPEDV-AKMGITEYNNQCRAIVMRYSAEWKSTVSRLGRWIDF
#ICel DTHGLPVEYEIDKTLGISGPQDV-MKMGIANYNNECRKIVMRYSGEWEKTMGRLGRWVDF
#INlo DRHGVPVEYEIDKALGITGVKDI-HDMGIRKYNKECRKIVLRYTKEWEAVVKRMGRWIDF
#IMth DTHGLPIEHKVEGILGVRSKKDIEDKIGIEEFVRKCREFAMENKAVMTSQFQRLGVWMDW
#IMja DMHGLPIEVKVENEFGIKNKKEIETKIGVKQFIEKCKEFALKHKEIMEKQFKNLGVWLDW
#IPfu DTHGLPIEVKVEQALGLKTKKEIEEKIGVENFVQKCKEFALNNLRIMTEQFKMLGVWMDW
#IAfu DMHGLPIEVKVEQELGFRTKRDIE-SFGIDKFIERCMNYALANKDAMTEQFKSLAVWMDW
#ISac DTHGLPIEVETEKRLGIKSKAEIIEKVGVDNFISKCKEFAVNNSKSLTQNFRNLGIFMDW
```

```
       **************************************************  **** ****
#ITma ENPYITLKPDYEVKILDVFKTLVEQGNVYRSLKPIYWCPRCRTALAEAEI--EYHD-HKSP--
#IApy ENPYLTMSPEYEATEIRELGKFFEKGLAYRSKKPVYWCIYDKTAEGQAEV--EYYE-KEDP--
#ICje DKPYLTMKFEFEAAIYRTLCEIAKKGLLCERSKPVFWSWAAKSALAEAEV--EYQD-KEDY--
#IHpy EDPYKTMDFKFEASIYRALVEVAKKGLLKERHKPIYWSYACESALAEAEV--EYKM-KKSP--
#ISau NDPYITLKPEYEAAQIRIFGEMADKGLIYKGKKPVYWSPSSESSLAEAEI--EYHD-KRSA--
#ISaR DSPYITLENNYIESVWNILSTFHKKGLLYKGHKVTPYCTHDQTALSSHEVAQGYKN-VKDL--
#ITpa DNDYKTMDVCYMESVWWVVAQLWQRKLLYEGYKILPYCPRCATALSNHELNLGGYQDVSDP--
#IBbu EKGYKTMDISFMESVWWVFKNLYEKGLIYESYYVLPYSPKLATPLSNFEVNLGEYKEVNDP--
#ICtr SATWRTMDLSFMESVWWVFRSLYDQGLVYEGTKVVPFSTKLGTPLSNFEAGQNYKE-VDDP--
#IMge QNCYYTIDESFQFKELELFLQAIKKGLIFQDLKPTYWSPISRTSLAEAEI--EYKE-VNSI--
#IMpn SKYYKTIDESFQQNELDLFLQAVKKDLIFQALKPTYWSPVSRTSLAEAEI--EYKE-VKTI--
#IMtu DNDYKTLDLAYMESVIWAFKQLWDKGLAYEGYRVLPYCWRDETPLSNHEL--RMDD-DVYQ--
#IBsu ENPYVTLKPEYEAQQIRVFGEMAKRGYIYKGLKPVNWSPSSESALAEAEI--EYQDKRSAS--
#ICac ENPYVTYHNDYIESVWWALKQMWDKDLLYKGHKIVPYCPRCGTALSSHEV--AQGYKDVKE--
#ISPn ENPYVTLTPDYEAAQIRVFGEMANKGYIYRGAKPVYWSWSSESALAEAEI--EYHDLVSTS--
#ISyn KKPYLTLTPEYEAAQIGVFGAMALKGYIYRGLKPVHWSPSSRTALAEAEL--EYPEGHTSR--
#IPfl DNPYKTMNFKNEAGEIRALAEIVKGGFVFKGLKPVNWCFDCGSALAEAEV--EYED-KKSS--
#ISmt ETPYLTMDKDYEINQLNIFKEMYERGLIKRQNKPVYWGTETRTALAEGEL--EYNE-NHKSI-
#ICmt EKPYVTKSPSYVAAQLDIFAKLVEQKLVYRSFKPVYWSPSSNTALAESEL--EYND-KHQSTS
#IHmt EIATIHLNGKYEAKQLRTFYQMYDKGLVYRSYKPVFWSPSSGLHWLEAEL--EYNP-EHVSRS
#IHin DNPYLTMNFDTEANIIRTLGKVIENGHLYKGSKPVHWCLDCGSSLAEAEV--EYED-KVSP--
#IEco SHPYLTMDFKTEANIIRALGKIIGNGHLHKGAKPVHWCVDCRSALAEAEV--EYYD-KTSP--
#INgo DNPYLTMDFKTEADTVRMLGEIYKSGYLYRGAKPVQFCLDCGSSLAEAEV--EYKD-KVSP--
#ISce DNDYKTMYPSFMESTWWAFKQLHEKGQVYRGFKVMPYSTGLTTPLSNFEAQQNYKD-VNDP--
#ITth DNDYKTLDLKFMESVWWVFKQMFDKGLVYRGCKVMPYSNGCATVLSNFETQQNYKE-VDDP--
#IHsa DNDYKTLYPQFMESVWWVFKQLYDKGLVYRGVKVMPFSTACNTPLSNFESHQNYKD-VQDP--
#ICel KHDYKTLYPWFMESVWWAFSELHKKGLVYKGVKVMPFSTACSTPLSNFEAGQNYKD-VVDP--
#INlo RNGYRTMDRSFMESIWHIFKMLYEKGSVYRGHRVMPYSTACSTPLSNFEANQNYKD-VSDP--
#IMth DDPYVTFDPAYMESCWWTLKRAHEKDLLLRDLRVITWCPRCETALALAEI--DYHE-KEDP--
#IMja ENAYMPITKEYMEIGWWTLKVAHEKGLLTRDLRVVYWCPRCETALAEHEVRGEYKE-VYDP--
#IPfu DNPYMTIKNEYIESAWFTLKWAWEKGLLEKDKRVLHWCPRCETALAEHEVRGEYKL-RKDP--
#IAfu ENPYMTIKAEYMNAAWFAIKRAHERGLLERKKMVVNWCHRCETALADAEV--EYWD-EEDP--
#ISac ENPYFTFNNDYISNSWAVIK-SIRRGLLYKGVHVLHWCSRCETTLADYEV-SEYRD-LEDP--
```

```
      **********                    ******************************
#ITma SIYVKFRSKEDPNFF--------------IVIWTTTPWTLPANVGIALHPDYEYSVVKVG-
#IApy SIYVKFPLKKEIEGKKAY-----------AVIWTTTPWTLPANLGIMVKEDADYSLVEVE-
#ICje SIFVAFDLDVKACKKLGVSKAS-------AVIWTTTPWTLVANQAIALNPNENYVITKEG-
#IHpy SIFVAFGLKKESLEKLKVKKAS-------LVIWTTTPWTLYANVAIALKKDAVYALTQKG-
#ISau SIYVAFNVKDDKGVVDADAK---------FIIWTTTPWTIPSNVAITVHPELKYGQYNVN-
#ISaR SAVVKFQLTNSKDTY--------------FLSWTTTPWTLPANVALAINKDLNYSKIRVE-
#ITpa AITVRFECTS???TC--------------FLAWTTTPWTLPCNAA?ALGPQILYVLIEAN-
#IBbu SLTIKFKIKDKNEY---------------LLVWTTTPWTLPSNLGIAVGQEIEYSKIFDK-
#ICtr SVVVKFALQDNQGF---------------LLAWTTTPWTLVSNMALAVHPELTYVRIKDK-
#IMge ALYLTFKVSKSDFLDENAN----------LLVWTTTPWTLPTNQAIAIHPDFDYLLFEYN-
#IMpn GLYLTFTVVQSAVLNSGTK----------LLVWTTTPWTLPTNQAIAVHPQFEYLLFTYN-
#IMtu SRQDPAVTVGFKVVGGQPDNGL---DGAYLLVWTTTPWTLPSNLAVAVSPDITYVQVQAG-
#IBsu IYVAFGVKDGKGVLENGE----------RIIIWTTTPWTIPANLGISVHPDLEYSVIAVG-
#ICac ATAFVKFKVKGEENK-------------YILAWTTTPWTLPSNVALAINKAYDYVEVINN-
#ISPn LYYANKVKDGKGVLDT----------DTYIVVWTTTPFTITASRGLTVGADIDYVLVQPA-
#ISyn SIYVSFPITQAGEKAAEILAPYI--DDLAVAIWTTTPWTLPGNLAVALNPELTYAVVETE-
#IPfl TIDVAFPIADDAKLAEAFGLASLA-KPAAIVIWTTTPWTIPANQALNVHPEFTYALVDVG-
#ISmt AAYVKFPLEKKSQMDLCKKLGITNNLPIYCLIWTSTPWTLLSNRAICFNQDFSYSLLRLN-
#ICmt AYFRFKLINFSSSDVIWAHSEPSKISQFFALIWTTTPWTLPLNNAISVSSAIQYSLIQFDN
#IHmt IYVKFPLLKPSQIGISYRWSPV------SILVWTTQPWTIPANEAVCYMPESKYAVVKCSK
#IHin SIYVRFPAESADEIEAKFSAQGRGQGKLSAIIWTTTPWTMPSNRAIAVNADLEYNLVQLG-
#IEco SIDVAFQAVDQDALKAKFAVSNVN-GPISLVIWTTTPWTLPANRAISIAPDFDYALVQID-
#INgo AIDVAYPFKNTVALAAAFGLAGIE-GKAFAVIWTTTPWTLPASQAVSAGADVVYQLIDTP-
#ISce AVTIGFNVIGQEKTQ--------------LVAWTTTPWTLPSNLSLCVNADFEYVKIYDE-
#ITth SLFIAFKTAEDPKTK--------------FIAWTTTPWTLPSNLALVINKDFDYVKVLDA-
#IHsa SVFVTFPLEEDETVS--------------LVAWTTTPWTLPSNLAVCVNPEMQYVKIKDV-
#ICel AVFVGFKLLDCPNRQ--------------LVAWTTTPWTLPSNLALVVHPDMLYVVTKDK-
#INlo SILVAFPLRKPFKGYSLS-----------LVAWTTTPWTLPSHMAILVNQSFIYAIFR---
#IMth SIYVKFPVSGDTY----------------ILVWTTTPWTLPANMAVAVHPDFDYAHTRLD-
#IMja SVYVKFRLANEENTY--------------IVIWTTTPWTLVANLAVTVHPDYDYAYVEVEF
#IPfu SIYVKFPVEGKENEY--------------LLIWTTTPWTLPANLAVSALPDYDYVKVGVDL
#IAfu SIYVKFPVKGEKDTY--------------IVIWTTTPWTLPANMAVAVHPSLEYAKFRAVK
#ISac SIYVKFRVKGEPNRY--------------LVIWTTTPWTLPANVFVMINKDFEYADVRVG-
```

```
            **********             **************************
#ITma ---EEKWVIATDLLDAFSK----ETGIDCSNVVEKIKGKDLEGKEFVHPIFDDRTS----------
#IApy ---GEVWIVAKELLENFFK----NIGKTYTRVLKDVKGRDLVGLEYEHPFVDRDELKGYLSE----
#ICje ------LIFASALLKSMV-----EKGLTSGEIQKELNAKEFEKLEAINPLNGRKSV----------
#IHpy ------YLVAKALHEKLA-----ALGVVDNEITHEFNSNDLEYLVATNPLNQRDSL----------
#ISau ---GEKYIIAEALSDAVAEAL--DWDKASIKLEKEYTGKELEYVVAQHPFLDRESL----------
#ISaR ---NEYYILATDLINS--------IITEKYEIIDTFSGSNLINLKYIPPFESDGLV----------
#ITpa ---DEHYILARSRLEFY------YPDSSAYRVVWEKRGEHLAGIRYRPLFSYPVFGQGPDPSVQGD
#IBbu ---TKEEILILGSKKLN----SYYDDENSYTIIEKFKGSKLEGIEYEPIFNYFLEQKDKGAF----
#ICtr -ESGDEYILGQESLPRW------FPDRESYEWIGQLSGKSLVGQSYEPLFPYFQDK----------
#IMge ----QKFVILEKLFEVFTNKL----NWTNAIKLKKFKGSNLKNSSYSHCFYNKVLP----------
#IMpn ---NEQYVVLASLFESLKTKF----GWTDAIQVQTISGSQLQNTTYKHCLYDKVNP----------
#IMtu --DRRFVLAEARLAAYAREL-GEEPVVLGTYRGAELLGTRYLPPFAYFMDWPNAF-----------
#IBsu --EDRFVVASALVENVASAC-----GFDQYEVTRTVKGKDLENIIAEHPLYGRDSL----------
#ICac --GEHLILAKALLTVLEGEY-----EVVSEFKGEKLLGMEYEQLFKFANPDKKAFY----------
#ISPn -GEARKFVVAAELTLSEKF------GWADVQVLETYRGQELNHIVTEHPWDTAVEE----------
#ISyn SHIFHRSYLIVALDLVEKLS---ETFGVKLTVKVTLQGESLENTCYQHPLFDRVSP----------
#IPfl ---DRLLVLAEEMVESCLA-----RYELQGSVIATATGSALELINFRHPFYDRLSP----------
#ISmt ----SELILVETGSIDKLGL---TTNSFETIKQFQGTHLNGLYYQNLLVDDKVGRP----------
#ICmt -EINNPTSTFYVIASKLLEEFQKSSDRKCKVVGTVKPANLIGRRYKSCWHNELGLP----------
#IHmt ---GDLYVLAADKVASVAS-----TLETTFETISTLSGVDLENGTCSHPLIPDKAS----------
#IHin ---DERVILAAELVESVAK----AVGIEHIEILGSVKGDDLELSRFHHPFYDFTVP----------
#IEco ---GQAVILAKDLVESVMQ----RIGVTDYTILGTVKGADVELLRFTHPFMGFDVP----------
#INgo ---KGKLVLAKDLAEGALK---RYGFSDGIAILAETTGDKLENLHMNHPFLERDIP----------
#ISce -TRDRYFILLESLIKTLYK----KPKNEKYKIVEKIKGSDLVGLKYEPLFPYFAEQFH--------
#ITth -KTQEHYILAECRLPELYK-----KDKDGYKILEKFKGSELVGREYEPLFPYFLSRKQ--------
#IHsa -ARGRLLILMEARLSALYK------LESDYEILERFPGAYLKGKKYRPLFDYFLKCKE--------
#ICel -TTGIEYVVLEERLGELKN--------DNLEVIEKLAGSQLKDLRYEPLFPYFAYMRE--------
#INlo -LKSAFFIMQRDRVNVYFK---------DAVVISEVKGHELLHLEHDQPFMYYNHYRE--------
#IMth ---GETYIMAEALVEKVLG--------EEAEIIKTVRGSELEGLTYRHPLDEEVPCHR--------
#IMja DDKKEVWIIAEKLVEEVINKAKKFHNIKNYKIIKKVKGKELEGIKYIHPLLEENERQKEFAE----
#IPfu NGREEYWILAKALVEKVLG-----DIGVKGEVVEEFKGKELEGLRYVHILMDEYPRQKEFRE----
#IAfu DGKVEYLILAKELADSVLG----KGDYDSWEVVETYLGEDLEGLEYEHPLADEVPLQK--------
#ISac -----DEILVIAKDRVKEL--MKEARIKEYKILRVYKGEELLGLEYEHPLADIVSAQS--------
```

```
             ***** ************************** ****     ************
#ITma ------RVILAD-YVSLETGTGCVHIAPGHGEEDYIYGHVQYGL-----PIVSPVDEEGRF----------
#IApy ETLKNMWRIYPSEFVSLDTGTGLVHMAPGHGQEDYTVGK-RYNL-----EPYAPLDDSGRF----------
#ICje -------LIMGE-HVLMDGGSGLVHTAPGHGEDDYYACL-KYGI-----EVLMPVDDSGCYDETLRAKRLL
#IHpy -------VALGE-HVGLEDGTGAVHTAPGHGEEDYYLGL-RYNL-----EVLMSVDEKGCYDEGIIHNQLL
#ISau -------VINGD-HVTTDAGTGCVHTAPGHGEDDYIVGQ-KYEL-----PVISPIDDKGVF----------
#ISaR ---NAYYVVDGE-FVTNSEGTGIVHIAPAHGEDDYQLVL-ERDL-----DFLNVITREGVY----------
#ITpa SEEGLFCTRVAD-FVSTEDGTGVVHVAPAFGEDDYEVFK-DAGI-----SIQCPLDAECRF----------
#IBbu ------KVHTAD-YVTTDDGTGIVHIAPFGEEDYRILKK-HTNV-----DIIDPLDAECKF----------
#ICtr KELEAFRILPAD-FIEESEGTGIVHMAPAFGEADFFACQ-EHNV-----PLVCPVDNQGCY----------
#IMge -------VLMGI-HVVDNEGTGIVHSSPAFGIDDFYLCQ-KNKI----KEVLISIDEKGVF----------
#IMpn -------VLLGN-HVLCNEGTGLVHTSPAYGLDDFYLCK-QNKL----NEALVSLDEKGVF----------
#IMtu ------QVLAGD-FVTTDDGTGIVHMAPAYGEDDMVVAE-AVGI-----APVTPVDSKGRF----------
#IBsu -------VMLGE-HVTTDAGTGCVHTAPGHGEDDFIIGQ-KYGL-----DVLCPVDEKGVM----------
#ICac -------VVHGD-FVTLSDGTGIVHIAPAYGEDDNMLGK-KYDL-----PLINLVNGEGKF----------
#ISPn ------LVILGD-HVTTDSGTGIVHTAPGFGEDDYNVGI-ANNL-----EVAVTVDERGIM----------
#ISyn ------IVIGGD-YVTTESGTGLVHTAPGHGQEDYVVGQ-RYGL-----PILSPVDAAGNL----------
#IPfl -------VYLAD-YVELGSGTGIVHCSPAYGVDDFVICK-KYGM-----VNDDIINPVQSN----------
#ISmt -------LLHGA-HVTSGTGTGLVHTAPGHGQDDYLIGI-QNGL--EIYSPVDHQGRYQLN----------
#ICmt -------IYEGP-HVTDTVGTGLVHTAFAHGFQDYDVAI-SKGD--RVESFVDSRGCYTRH----------
#IHmt -------PLLPPNHVTMAKGTGLVHTAPAHGMEDYGVAS-QHNL-----PMDCLVDEDGVF----------
#IHin -------VILGD-HVTTDGGTGLVHTAPDHGLDDFIVGK-QYDL-----PMAGLVSNDGKF----------
#IEco -------AILGD-HVTLDAGTGAVHTAPGHGPDDYVIGQ-KYGL-----ETANPVGPDGTY----------
#INgo -------MLNGE-HVTTDAGTGLVHTAPAHGLEDYAVCN-KYGI-----ELYNPVNAEGKY----------
#ISce --ETAFRVISDD-YVTSDSGTGIVHNAPAFGEEDNAACL-KNGVISEDSVLPNAIDDLGRF----------
#ITth --DGCFRILAGD-FVTADAGTGIVHCAPGFGDDDYKVSV-ANNIIKPDDPPV-PVDENGHF----------
#IHsa --NGAFTVLVDN-YVKEEEGTGVVHQAPYFGAEDYRVCM-DFNIIRKDSLPVCPVDASGCF----------
#ICel --ERNAFRVLNDTFVTSDSGTGVVHQAPYFGEIDFQVCV-ANGVIAKDQKMICPVDESGKY----------
#INlo --KGFFRVYHAD-FVSEIDGTGVVHCSPGFGEDDYKAMT-MPGLIKENELLPSPLDENGRF----------
#IMth --DMEHRVILGD-HVTLTEGTGCVHTAPGHGPEDFEIGK-EYGL-----PVFCPVDEAGVF----------
#IMja -LENAHTVILGE-HVTLEGGTGLVHTAPGHGEEDFEVGK-KYNL-----PIYSPIDDEGKY----------
#IPfu KYEWVHRVILAD-FVTLEEGTGLVHTAPGHGEEDFEVGQ-KYGL-----PVYSPVDDQGKY----------
#IAfu ---NWKHQVFFA-DFVTAENTGCVHIAPGHGVEDYELGV-EKGL-----EVFNPVDDRGVY----------
#ISac KINNHHKVLDGGEAVTLQEGTGLVHSAPGHGDVDFEIGK-KYDM-----PVVMLVNDKGEF----------
```

```
                        *************          *******************************
#ITma --TEEAGKY-KGMFIE-DANEVIIEDLKRK--------GILVHASSITHSYPHCWRCKGPVIFRATEQWFIS
#IApy --VEPAPEFIRGVRVF-DANKLIIALLKEK--------GYLVHEARIRHSYPHCWRCKNPVIFRATPQWFIG
#ICje -PSHLLEEFI-GLHIF-KANEKILELLGEK----------LLHSSKFIHSYPFCWRTHKPVIYRATKQWFIL
#IHpy -----DESYL-GEHVF-KAQKRIIEQLGDS----------LLLEQEIEHSYPHCWRTHKPVIYRATTQWFIL
#ISau --TEEGGQF-EGMFYD-KANKAVTDLLTEK--------GALLKLDFITHSYPHDWRTKKPVIFRATPQWFAS
#ISaR --NDRFPEL-VGNKAK-NSDIEIIKLLSKK--------QLLYKKQKYEHNYPHCWRCGNPLIYYAMEGWFIK
#ITpa --TAEVADY-QGLFVK-AADKAIIARVQKQ--------GALFRREQISHAYPHCWRCASPLIYRAVHSWFVA
#IBbu --TNQVKDF-KGLFVK-DADKKIIENLKLR--------NFLFKRENYLHRYPFCYRTNCPIIYRPISSWFVN
#ICtr --TAEVKDF-VGEYIK-SADKGIARRLKNE--------NKLFYQGTVRHRYPFCWRTDSPLIYKAVNSWFVA
#IMge --NNLLNDKELENCFYLKANDLIINRLKQN--------NSFIFSEVISHREPHDWRSKTPVIYRASKQLFIK
#IMpn --NDTLNDPVLTGLFYLKANDVIIERLKQH--------HNFVFSESFLHREPHDWRSKTPVIYRASKQLFIK
#IMtu --DVTVADY-QGQHVF-DANAQIVRDLKTQSGPAAVNGPVLIRHETYEHPYPHCWRCRNPLIYRSVSSWFVR
#IBsu --TSEAPGF-EGMFYD-DANKAITQQLDEK--------GALVKLEFITHSYPHDWRTKKPTIFRATAQWFAS
#ICac --VDEVEPW-KGLFVK-KADPKILEYMKEN--------GTLYKSEKFTHSYPHCWRCDTPLLYYPRDSWFVR
#ISPn --MKNAGPEFEGQFYE-KVVPTVIEKLGNP----------PSCQEEISHSYPFDWRTKKPIIWRAVPQWFAS
#ISyn --TEEAGKF-AGLNVLNDANEAIINALQDQ--------KVLLKEEAYEHKYPYDWRTKKPTIFRATEQWFAS
#IPfl -GVYVPSLEFFGGQFIFKADQPIIEKLREV--------GALMQTAAIQHSYMHCWRHKTPLIYRATAQWFIG
#ISmt ELPQSVRSIVRDEGDLTKGRQVLDAETAKIIPCKLSDLNLLYKSHEYTHSYPYDWRSKKPVIIRATPQWFAD
#ICmt -----LGHD-LDGKEV-LGEGQKIALRLLN--------HDIVHVSKHVHSYPYDWRTKKPVIIRSSEQWFID
#IHmt --TDVAGPE-LQNKAV-LEEGTDVVIKMLQ------TAKNLLKEEKLVHSYPYDWRTKKPVVIRASKQWFIN
#IHin --ISTTEFF-AGKGVF-EANPLVIEKLQEV--------GNLLKVEKIKHSYPHCWRHKTPIIFRATPQWFIG
#IEco -LPGTYPTL-DGVNVF-KANDIVVALLQEK--------GALLHVEKMQHSYPCCWRHKTPIIFRATPQWFVS
#INgo --ISETPRV-AGMSVW-EANPVILQWPEET--------GNLLASSKIEHSYAHCWRHKTPLIYRATGQWFVG
#ISce --TKDVPDF-EGVYVK-DADKLIIKYLTNT--------GNLLLASQIRHSYPFCWRSDTPLLYRSVPAWFVR
#ITth --TNVVSDF-AGVYIK-EADKLIRKNLKER--------GLLLVDSSFKHNYPFCWRSDTPLIYKAVHCWFIK
#IHsa --TTEVTDF-AGQYVK-DADKSIIRTLKEQ--------GRLLVATTFTHSYPFCWRSDTPLIYKAVPSWFVR
#ICel --TSEVPDY-QGVYVK-DADKLIIKRLKEM--------GNLVRQAEVKHSYPFCWRSDTPLLYKAVPSWFIN
#INlo --NEEVPEY-KGMHVK-DADQSIIKDLGKK----------ILYEGKVYHRYPFCWRSDTPLIYKLVPNWFVR
#IMth --TEDAGKY-RGLFVK-DADSDIIDDLRSK--------NLLLRAETISHRYGFCWRCKTPIIYLATEQWFLK
#IMja --VE--GKW-KGVFVK-DADAEIIETLKNK--------GLLVYAGKIKHSYPHCWRCKTPLLFRATEQWFLE
#IPfu --VE--GKW-KGVYVK-EADPQIIEHLKEK--------GYLVKAGEIEHKYPHCWRCKTPLIFRATDQWFLK
#IAfu --TEEAGKY-AGKHVK-EANDDIIDDLYRK--------DLLLAEERIVHRYGHCWRCKTPIIYRATEQWFIK
#ISac --TQDSGKY-AGKYVR-SASEEIISDLKQR--------SALLHASKIVHRYPVCWRCKTPLILRAIEQWFIA
```

```
                 ********************     *********************************
#ITma -----VDHNNLRQKVLEEIDKVKWIPEW-----GRNRIRSMVEERPDWCISRQRVWGTPIPAVKC
#IApy MDIEYEGKTLSGES-LEEIEKVKWIPEY-----GKNRIKSMVENRPDWCISRQRFWGVPITVFYC
#ICje MDEPKLQGKTLREC-AKEQLLKTTFYPQ----SGVKRIGSMVENRPDWCISRQRDWGTPIAFFRD
#IHpy MDEPFIQNDGSQKT-LREVALDAIEKVEFVPSSGKNRLKTMIENRPDWCLSRQRKWGVPLAFFID
#ISau I------SKVRQDI-LDAIENTNFKVNW-----GKTRIYNMVRDRGEWVISRQRVWGVPLPVFYA
#ISaR ------TTNFKNEI-INNNNNIEWFPSH----IKEGRMGNFLENMVDWNIGRNRYWGTPLNVWIC
#ITpa ------VEKIKDKM-LAANASICWQPSH----IRDGRFGKWLVCARDWAISRDRYWGNPLPIWRC
#IBbu ------VEKIKTKL-LEVNEKINWMPAH----LKKGRFGKWLENAKDWAISRNRFWGNPIPIWIC
#ICtr ------VEKVKSKM-LKANESIHWTPEH----IKQGRFGKWLEGARDWAISRNRYWGTPIPIWRS
#IMge -------TKSIKKQLKKQINQVNFLNSK-----NQLRLKEMLLQRDEWCISRQRVWGLPIPIVYA
#IMpn -------TKSIQSK-LKRQIKRVKFVNN----KNKERLQEMLLQRAEWCISRQRVWGLPIPLIYA
#IMtu V------TDFRDRM-VELNQQITWYPEH----VKDGQFGKWLQGARDWSISRNRYWGTPIPVWKS
#IBsu I------KDFRSDL-LDAIKETKWVPEW-----GEQRLHNMVRDRGDWCISRQRAWGVPIPVFYA
#ICac M------TSLRDKL-VENNNKIHWYPDN-----IRTGRFGKFVENVIDWGISRDRYWGTPLPIWQ
#ISPn V-------SKFRQEILDEIEKVKFHSEW-----GKVRLYNMIRDRGDWVISRQRAWGVPLPIFYA
#ISyn V------EGFRDQA-LKAIKEVTWIPTQ-----GENRITPMVGDRSDWCISRQRAWGVPIPVFYD
#IPfl MDKEPTSGDTLRVRSLKAIEDTKFVPSW-----GQARLHSMIANRPDWCISRQRNWGVPIPFFLN
#ISmt -------LHDVKNLALESISRVKFCPKR-----GYSRLSSFMKSRNEWCISRQRSWGIPILSFYK
#ICmt VD-----EIGKRASMMLDDISVAAGDSD-----LRASLKQLVTTRKSWCISRQRVWGTPIPALVD
#IHmt -------ITDIKTAAKELLKKVKFIPGS-----ALNGMVEMMDRRPYWCISRQRVWVVPIPVFHH
#IHin ME-----TQGLRQQALGEIKQVRWIPDW-----GQARIEKMVENRPDWCISRQRTWGVPMTLFVH
#IEco MD-----QKGLRAQSLKEIKGVQWIPDW-----GQARIESMVANRPDWCISRQRTWGVPMSLFVH
#INgo MDKAGSDGKTLRDKAIKAVDDTEFFPPW-----GRARLESMIEGRPDWVVSRQRYWGTPMTFFVH
#ISce V------KNIVPQM-LDSVMKSHWVPNT----IKEKRFANWIANARDWNVSRNRYWGTPIPLWVS
#ITth V------TALKDDL-LANNKKAYWVPKF----AQEGRFNNWLQNVSDWCFSRSRFWGNPIPIWVS
#IHsa V------ENMVDQL-LRNNDLCYWVPEL----VREKRFGNWLKDARDWTISRNRYWGTPIPLWVS
#ICel V------ETLIPRL-LANNDETYWVPAF----VKEKRFANWLRDARDWAVSRNRFWGTPINLWVS
#INlo V------KREIPRL-LESNSTINWIPES----IGEHKFKNWLSEARDWSISRNRFWGTPIPLWHC
#IMth I------TEIKDKM-LSELDRVQWIPSW----AGESRFRNWIENARDWTISRQRYWGIPIPIWVC
#IMja I------SKIKDNI-IEHAKTVQWIPHW-----VETRYINGVKFVGDWNISRQRYWGIPIPVWVC
#IPfu V------SKVKEKIIKENDEKVTWYPEW-----VKIRFDNGVRDSGDWVISRQRYWGIPLPIWQS
#IAfu I------SELKDEM-LEEIDKVMWIPEW----AGSARFKDWVSNAKDWCISRQRYWGIPIPVWIC
#ISac V------SKLKDHL-MGEIDRVRWIPDW-----GKTRIGNMVKEVRDWVISRQRFWGTPLPIWVC
```

```
             ************                              ***********
#ITma KECGEV-VLDPKVIEHFMKIVEK--------------------EGTNAWFEKEVEELIPED
#IApy ENCGEV-IKDKEVFERIASLVEK-------------------HPGGTDVWFEKSPEEILPEG
#ICje KNTKEV-IFDDELFDFVAAIFEK---------------------HGADAWWEFEIKDLIPTN
#IHpy KRTNKP-CFESEVLEHVANLFEK---------------------KGCDVWWEYSVKDLLPPS
#ISau EN-GEI-IMTKETVNHVADLFAE---------------------HGSNIWFEREAKDLLPEG
#ISaR ND-CNH-EYAPSSIKDLQNNSIN----------------------KIDEDIELHRPYVDNI-
#ITpa VH-CGA-TDCIGSRTQLYERSGM-------------------------LLEDLHKHVVDM-
#IBbu SK-TGK-KICIGSKKELENLSGQ-------------------------KIEDLHKDQIDK-
#ICtr -D-DGE-LLVIGSIQELEALSGQ-------------------------KIVDLHRHFIDE-
#IMge NN-KPL-LDFSTIQYTIKQLKKH----------------------GIDSWFEKDVTCFL---
#IMpn DN-QPL-LDVTTIKYTIQQLKKY----------------------GIDSWFEKDINFFL---
#IMtu DD-PAY-PRIDVYGSLDELERDF---------------------GVRPANLHRPYIDELTRP
#IBsu EN-GEP-VITDETIEHVSELFRQ---------------------HGSNIWFEKEAKDLLPEG
#ICac CE-CGH-RDCVGSIEELKEKGIN-----------------------VPENIELHKPYIDEV
#ISPn ED-GTA-IMVAETIEHVAQLFEE---------------------HGSSIWWERDAKDLLPEG
#ISyn EETSEP-LLTEETINHVQQIIAE---------------------KGSDAWWELTVEELLPEA
#IPfl KESGELHPRTVELMEVVAQRVEQ---------------------QGIEAWFKLDAAELL---
#ISmt KSEPDSVLMNSEILAHAIEKIKQ---------------------KGINAWFNDKDNDMKEWL
#ICmt DNGGSYTSRKLIEWVAKLTRERE----------------------NTDVWWEIDVKEILENE
#IHmt KTKDEYLINSHTTEHIVKLVEQH----------------------GSDIWWTLPPLLPKEVL
#IHin KETEELHPRTLDLLEEVAKRVER---------------------AGIQAWWDLDEKELL---
#IEco KDTEELHPRTLELMEEVAKRVEV---------------------DGIQAWWDLDAKEIL---
#INgo KETGELHPNSAELLEKVAQRIEE---------------------KGIEAWFSLDKSELL---
#ISce DDFEEV--VCVGSIKELEELT------------------------GVRNITDLHRDVIDKL-
#ITth EDFEEV--VCIGSVEELKKLT------------------------GATEITDLHKDFIDHL-
#IHsa DDFEEV--VCIGSVAELEELS------------------------GA-KISDLHRESVDHL-
#ICel EDGEEV--VCVGSIAELEELS------------------------GQ-KITDLHRESVDDV-
#INlo DG--KY--ICIGSIEELSRLS------------------------GR-KIDDIHRENVDDV-
#IMth EDCDSI--HVVGSIGELRELAV----------------------EGQLEGDFIHRPHVDRI-
#IMja EKCGKY--IVVGSVEELEEKMI--------------------NKDEVGEINDLHKPTVDKI-
#IPfu ED-GEI--YVVGSWKELVELTVAIEVNSERIDLPESYEEKLKVIEEKLGPEDLHRPYVDAF-
#IAfu EKCGEM--KVVGSINEIEWE---------------------------NDLDLHRPKIDAV-
#ISac SNCQNI--IVVGGVDELGKIPI------------------------NQVPQDLHRPWIDSV-
```

```
                   ********************            ******************
#ITma FVCPKCGKRSFEKMLDTLDVWIDSGASFEYIT-------TKREDHPFPLDMYLEGSDQHRG
#IApy YKCPKCGGTSFRKEEDILDVWFDSGCSHASV---------IRPLGFEKADLYLEGSDQHRG
#ICje ---SKYNAENLEKVYDILDVWFDSGSTFNAVL-----NSGLYDAGEKRASMYLEGSDQHRG
#IHpy ---YQEDAKHYEKIMHILDVWFDSGSTFKAVL-----EDYHGEKGQSPSDVILEGSDQHRG
#ISau FTHPGSPNGTFTKETDIMDVWFDSGSSHRGVL-------ETRPELSFPADMYLEGSDQYRG
#ISaR TLSCPKCNGKMSRVEEVIDVWFDSGSMPFAQHHYPFD-NQKIFNQHFPADFIAEGVDQTRG
#ITpa VTIPCACGSVMRRVPEVLDCWFESGAMPYAQQHYPFE-HATDFERYFPAHFISEGLDQTRG
#IBbu ITWPSKDGGKFIRTSEVLDCWFESGAMPYASNHYPFT-NEINFKNIFPADFIAEGLDQTRG
#ICtr -IEINQNGKSFRRIPYVFDCWFDSGAMPYAQNHYPFE-RAEETEACFPADFIAEGLDQTRG
#IMge KPDKTKKWVKYHKEIDTLDVWFDSGSSYNVL---------EINKYGSIADLYIEGSDQYRG
#IMpn NPKKIQPGVEYKKETDTLEVWFDSGSTYNVL---------ISNKLNFPADLYLEGSDQYRG
#IMtu NPDDPTGRSTMRRIPDVLDVWFDSGSMPYAQVHYPFE-NLDWFQGHYPGDFIVEYIGQTRG
#IBsu FTHPGSPNGTFTKEQDIMDVWFDSGSSHQAVL-------EERDDLVRPADLYLEGSDQYRG
#ICac KLTCPKCGKPMTRTEEVIDCWFDSGSMPFAQLHYPFE-NKEVFENTFPAQFISEAVDQTRG
#ISpn FTHPGSPNGEFKKETDIMDVWFDSGSSWNGVV-------VNRPELTYPADLYLEGSDQYRG
#ISyn ---YRNNGRTYRKGEDTMDVWFDSGSSWAAVA------NAKNRPLKYPVDMYLEGSDQHRG
#IPfl ----GDEAPLYDKISDTLDVWFDSGTTHWHVL---RGSHPMGHETGPRADLYLEGSDQHRG
#ISmt PEKYHDVAHEYCRSQDTMDVWFDSGSSWSVIKDFYEKSLKLSKLPSPLYQVCLEGSDQHRG
#ICmt -----------EKNTDIMDVWLDSGLAWHAAR-------DNDTEREHVADVVLEGVDQFRG
#IHmt SEVGGPDALEYVPGQDILDIWFDSGTSWSYV----------LPGPDQRADLYLEGKDQLGG
#IHin ----GADAETYRKVPDTLDVWFDSGSTYSSVV------ANRLEFNGQDIDMYLEGSDQHRG
#IEco ----GDEADQYVKVPDTLDVWFDSGSTHSSVV------DVRPEFAGHAADMYLEGSDQHRG
#INgo ---SAEDCEHYDKLPDTMDVWFDSGSTHYSVV-------KQREELEWPADLYLEGSDQHRG
#ISce TIPSKQGKGDLKRIEEVFDCWFESGSMPYASQHYPFE-NTEKFDERVPANFISEGLDQTRG
#ITth TIPSQKGKGVLRRIDEVFDCWFESGSMPYGQQHYPFSMNEEEFSKRFPADFIGEGIDQTRG
#IHsa TIPSRCGKGSLHRISEVFDCWFESGSMPYAQVHYPFE-NKREFEDAFPADFIAEGIDQTRG
#ICel TIPSRSGRGVLKRVSEVFDCWFESGSMPYAQNHYPFE-NRKIFEDNFPADFIAEGIDQTRG
#INlo VICKDGEK--YRRIEEVFDCWFESGCMPYAQRHWPFE----CDNLCLPADFVAEGVDQTRG
#IMth ILECGRCGGRMKRTPDVLDVWIDSGVAGWAALHYPRE--KELFSEWFPYDFITEGHDQTRG
#IMja KLRCE-CGGEMKRVPDVLDVWFDSGLAPYASIGV---------KELKKADFITEGHDQVTK
#IPfu IIKVN--GKEMRRVKDVVDVWFDSGIASWASLGYPR--NKELFEKLWPADFIVEGEDQVTK
#IAfu TFSCQ-CGGVMRRVPDVFDVWFDSGVASWGSIAYPL--RKDKFEELWPADFITEGHDQTRG
#ISac VLDVKNVEGQARRISDVADVWFDSGVAFFASLGQDW--RKR-WSELGPVDLVLEGHDQLRG
```

```
       ************************  **********
#ITma WFHSSIFLAVAKRGSAPYKEVLT--HGFIKDEQGR----------------------------
#IApy WFQASLLESVGSYGEAPYRSVLT--HGFIVDEQGR----------------------------
#ICje WFQSSLLVGTAINESTPYESILT--HGFTTDEKGQ----------------------------
#IHpy WFQSSLLIGCVLNNQAPFKKVIT--HGFIVDEKGE----------------------------
#ISau WFNSSITTSVATRGVSPYKFLLS--HGFVMDGEGK----------------------------
#ISaR WFYSLLVISTILKGKSSYKRALS--LGHILDSNGK----------------------------
#ITpa WFYTLTILAVALFERPAFENCIV--TGLVLASDGK----------------------------
#IBbu WFYTLTILGTALFENTAFKNVIV--NGLVLSSDGR----------------------------
#ICtr WFYTLTVIAAALFDQPAFKNVIV--NGIILAEDGN----------------------------
#IMge WFNSSSNCGIIQNDLIPFKSLVS--HGFTLDENGN----------------------------
#IMpn WFNSSASCGIIQTDQLPFKSLIS--HGFTLDEHGN----------------------------
#IMtu WFYTLHVLATALFDRPAFKTCVA--HGIVLGFDGQ----------------------------
#IBsu WFNSSLSTAVAVTGKAPYKGVLS--HGFALDGEGR----------------------------
#ICac WFYTLLAISTSIFDKSSFENCIV--LGHVLDKHGL----------------------------
#ISpn WFNSSLITSVANHGVAPYKQILS--QGFALDGKGE----------------------------
#ISyn WFQSSLLTSVAVNGIAPYKTVLT--HGFVLDEKGH----------------------------
#IPfl WFHSSLLTGCAIDNHAPYRELLT--HGFTVDETGR----------------------------
#ISmt WFQSSLLTKVASSNVPVAPYEEVITHGFTLDENGL----------------------------
#ICmt WFQSLLLTSVAVQNKLPYKKIIV--HGFCIDENNN----------------------------
#IHmt WFQSSLLTSVAARKRAPYKTVIV--HGFTLGEKGE----------------------------
#IHin WFMSSLMLSTATDSKAPYKQVLT--HGFTVDGQGR----------------------------
#IEco WFMSSLMISTAMKGKAPYCQVLT--HGFTVDGQGR----------------------------
#INgo WFQSSMLTGCASSMGRAPYKQLLT-HGFVVDQNGR----------------------------
#ISce WFYTLAVLGTHLFGSVPYKNVIV--SGIVLAADGR----------------------------
#ITth WFYTLNVISTALRNSNPYKNLIV--NGIVLAADGK----------------------------
#IHsa WFYTLLVLATALFGQPPFKNVIV--NGLVLASDGQ----------------------------
#ICel WFYTLLVLSTALFNKPPFKNLIC--NGLVLASDGA----------------------------
#INlo WFYTMHVISTVLLAKPRFRNCIV--NGIVLASDKK----------------------------
#IMth WFYSQLGCGVIALDETPYRRVLM--HGFTLDEEGR----------------------------
#IMja WFYSQHALSAIVFNDIPYKKCLM--HGFTLDEHGD----------------------------
#IPfu WFYSQQAASVIAFDTVPYRAVAM--HGYVLDEKGD----------------------------
#IAfu WFYSQLGTSVVCFDKAPYKAVLM--HGFTLDEQGR----------------------------
#ISac WFFSLLRTGVILMDKAPYEAVLV--HGFMLDEQGREMHKSSGNYVEPSQVVSKYGQRYVKIMV
```

```
             ******
#ITma -------KMSKSL
#IApy -------KMSKSL
#ICje -------KMSKSK
#IHpy -------KMSKSK
#ISau -------KMSKSL
#ISaR -------KMSKSK
#ITpa -------KMSKAL
#IBbu -------KMSKSF
#ICtr -------KMSKRL
#IMge -------KMSKSL
#IMpn -------KMSKSL
#IMtu -------KMSKSL
#IBsu -------KMSKSI
#ICac -------KMSKHK
#ISpn -------KMSKSL
#ISyn -------KMSKSL
#IPfl -------KMSKSL
#ISmt -------KMSKLV
#ICmt -------KMSKSI
#IHmt -------KMSKSL
#IHin -------KMSKSI
#IEco -------KMSKSI
#INgo -------KMSKSI
#ISce -------KMSKSL
#ITth -------KMSKSK
#IHsa -------KMSKRK
#ICel -------KMSKSK
#INlo -------KMSKRL
#IMth -------KMSKSL
#IMja -------KMSKSL
#IPfu -------KMSKSL
#IAfu -------KMSKSL
#ISac TLEIPHGKMSKSH
```