Mycoscience (2004) 45:345–350
DOI 10.1007/s10267-004-0192-x

SHORT COMMUNICATION

Kenji Sorimachi · Teiji Okayasu

An evaluation of evolutionary theories based on genomic structures in *Saccharomyces cerevisiae* and *Encephalitozoon cuniculi*

Received: March 15, 2004 / Accepted: May 27, 2004

Abstract Codon usage patterns in 16 chromosomes coincided with each other in *Saccharomyces cerevisiae*, and the same result was obtained from *Encephalitozoon cuniculi* consisting of 11 chromosomes, although each chromosome function differs. In addition, preferential codon usage in the regenerated coding systems for Leu and Lys differed between *Saccharomyces cerevisiae* and *Encephalitozoon cuniculi*. These results cannot be explained by Darwin's natural selection theory or by the neutral theory proposed against Darwin's. Furthermore, the codon usage patterns were examined in both prokaryotes and eukaryotes. The use of G or C at the third codon position was much lower than T or A in *Ureaplasma urealyticum*, whereas inversely the use of G or C at the third codon position was much higher than T or A in *Mycobacterium tuberculosis*. Additionally, *Candida albicans* and *Plasmodium falciparum* also showed a very low usage of G or C at the third codon position. It is a difficult leap to speculate that the inverse codon usage change occurred over the genome during biological evolution. Thus, the present results strongly suggest that organisms were derived from different origins, indicating that the origin of life was plural, based on genomic structures.

Key words Codon usage · *Encephalitozoon cuniculi* · Evolutionary theory · Genome · *Saccharomyces cerevisiae*

Darwin's theory of natural selection, based on his observations, is undoubtedly one of the great scientific theories (Mayr 1965, 2000); however, the opposing neutral theory based on population genetics using mathematical equations is also acknowledged as a scientific theory (Kimura 1977). Although both theories are scientifically derived, their underpinnings obviously differ, resulting in ongoing arguments between Darwinian (Dawkins 2000) and non-Darwinian (Gould 1994) evolutionary theories, indicating that perhaps neither theory completely explains biological evolution. It is noteworthy that both theories were established before complete genomic data were available. Here, we reevaluate these theories in light of the completed genome. In addition, the origin of life has been assumed to be a single event on the basis of Darwin's theory (Mayr 1965, 2000); however, an opposite theory supposing a plural origin is also acknowledged (Woese 1998; Doolittle 1999). These conclusions are obscure, because they are not based on enough scientific results that can endure scientific discussion. More definitive results are expected from the use of genomic data. In the present study, we show that a dramatic and preferential codon usage shift apparently occurred during the biological evolution of the genome, suggesting, based on genomic structures, that the origin of life was a plural form.

Codon usage patterns of the first genes in the 16 chromosomes of *Saccharomyces cervisiae* are shown in Fig. 1A. The codon usage of 64 codons differs with large variations among the 16 genes encoding 123–1860 amino acid residues. This finding indicates that it is dangerous to discuss biological evolution just with a certain protein or gene, although phylogenetic trees were obtained, for example, from cytochrome *c* (Dayhoff et al. 1972), small subunit ribosomal RNA (Sogin et al. 1986; Woese et al. 1990; Doolittle and Brown 1994), and tRNA (Maizels and Weiner 1994; DePouplana et al. 1998). We need to realize that biological evolution is carried out by living organisms that have numerous genes. However, when the largest genes in each chromosome were examined, the variations were reduced. Further, a gene assembly (putative unit) consisting of more than 10000 nucleotides showed an almost constant codon usage pattern, suggesting that the pattern of codon usage depends on the gene size (data not shown). Indeed, codon usage coincided with each other among the 16 chromosomes consisting of 105–828 genes (Fig. 1B); this is

K. Sorimachi (✉)
Department of Microbiology, Dokkyo University School of Medicine, Mibu, Tochigi 321-0293, Japan
Tel. +81-282-87-2131; Fax +81-282-86-5616
e-mail: kenjis@dokkyomed.ac.jp

T. Okayasu
Center for Medical Informatics, Dokkyo University School of Medicine, Mibu, Japan

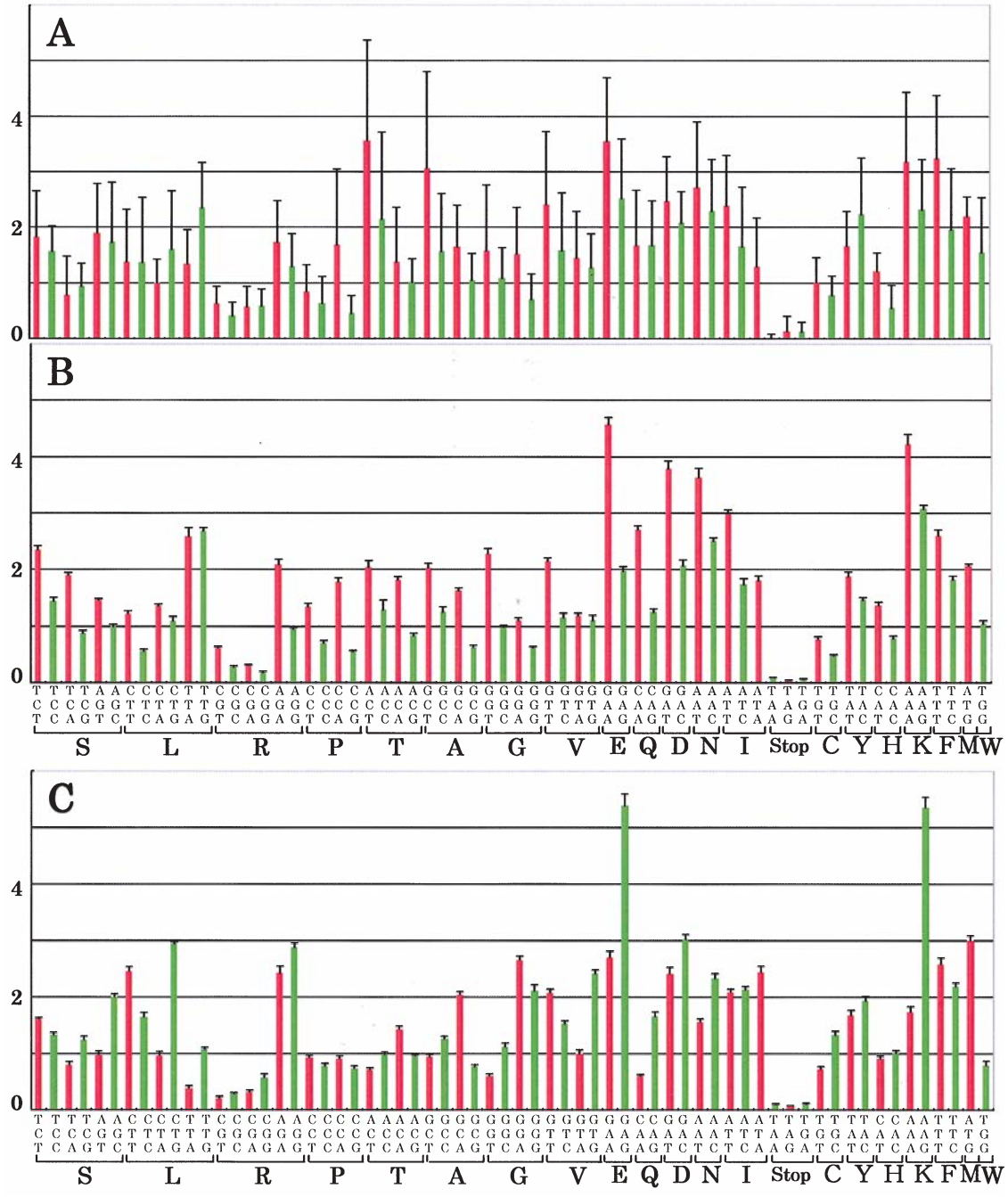
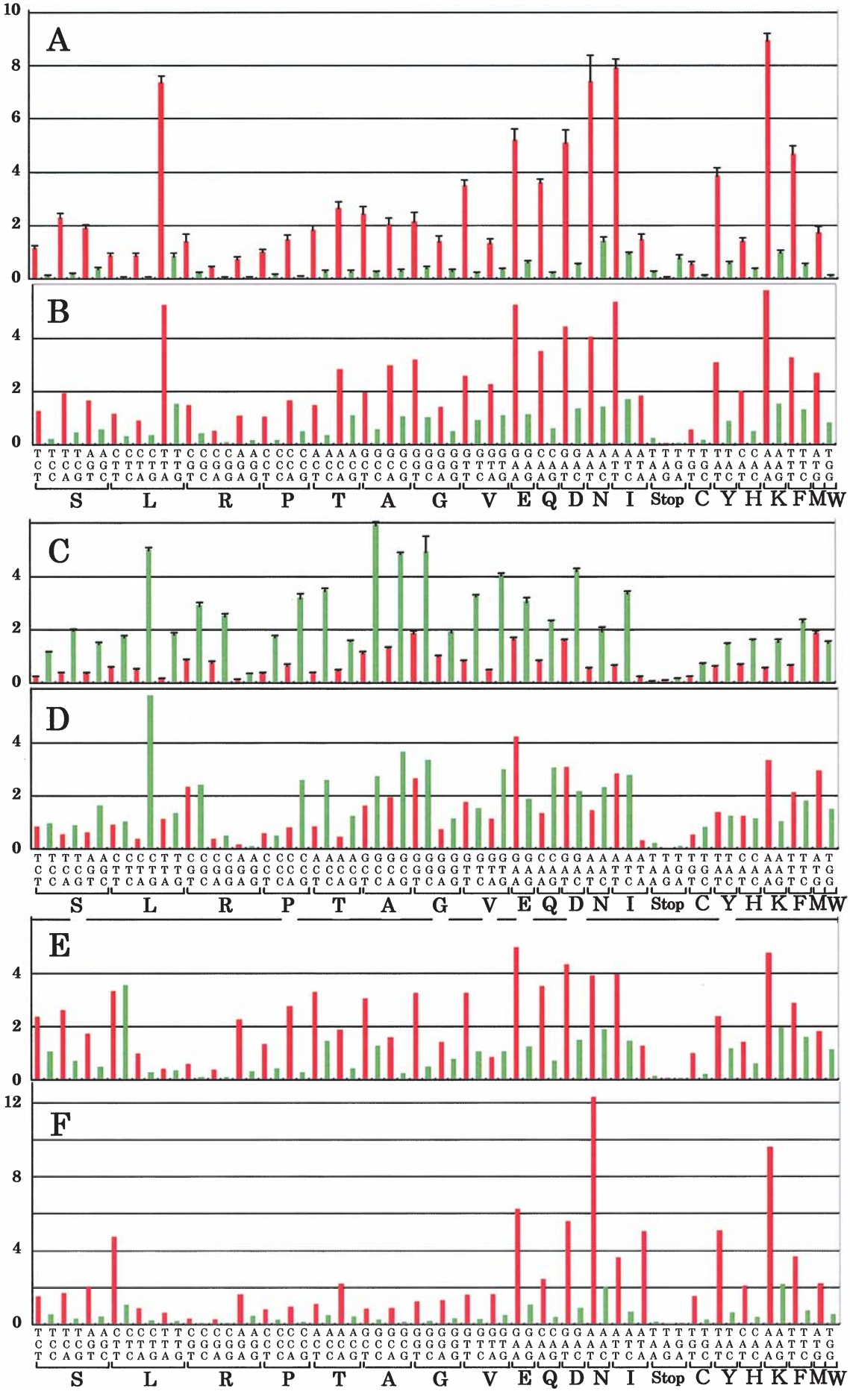


Fig. 1. Comparison of codon usage. The values represent the percent of total codons examined. **A** Codon usage in the first genes of the *Saccharomyces cerevisiae* chromosomes; the value is the mean ± SD of 16 chromosomes. **B** Codon usage in the *Saccharomyces cerevisiae* chromosomes; the value is the mean ± SD of 16 chromosomes. **C** Codon usage in the *Encephalitozoon cuniculi* chromosomes; the value is the mean ± SD of 11 chromosomes

Fig. 2. Comparison of codon usage. **A** Codon usage in *Ureaplasma urealyticum*. All genes (604) were grouped into 11 groups consisting of 50 genes each and 1 group consisting of 54 genes. The value is the mean ± SD of 12 groups. **B** Codon usage in the *Staphylococcus aureus* calculated from 452871 nucleotides coding 150484 amino acid residues. **C** Codon usage in *Mycobacterium tuberculosis*. The 4249 genes were put into 10 groups of 400 genes each and 1 group of 249 genes. The value is the mean ± SD of 11 groups. **D** Codon usage in *Escherichia coli* calculated from 105090 nucleotides coding 34926 amino acid residues. **E** Codon usage in *Candida albicans* calculated from 374241 nucleotides coding 373572 amino acid residues. **F** Codon usage in *Plasmodium falciparum* calculated from 2781241 nucleotides coding 2777823 amino acid residues


A
B
C
D
E
F
S L R P T A G V E Q D N I Stop C Y H K F M W

consistent with the result expressed in amino acid composition (Sorimachi and Okayasu 2003). Thus, although the chromosomes each have different functions, the synchronous codon alteration occurred independently from gene functions over the genome during biological evolution. Thus, the present result appears to fit the neutral rather than the natural selection theory; that is, data based on the genome support non-Darwinian evolution. A similar conclusion was obtained from the relationship between amino acid frequencies and codon usage, supporting random mutation (King and Jukes 1989).

Glu (E) is coded by two codons, GAA and GAG, and similarly Lys (K) is coded by AAA and AAG, while only GAA and AAA are preferentially used in every chromosome in the Glu and Lys coding systems, respectively (Fig. 1B). On the other hand, inverse codon usage, the preferential usage of GAG for Glu and AAG for Lys, was observed in *Encephalitozoon cuniculi* (Fig. 1C). Additionally, in the regenerated coding system for Leu (L), TTA and TTG were preferentially used compared with CTG in *S. cerevisiae* (Fig. 1B), whereas CTG was preferentially used compared with the other two codons in *E. cuniculi* (Fig. 1C). Therefore, the third-position nucleotide is not simply determined by the former two, although it has been reported that the former two nucleotides influence mutation of the third position (Sueoka 1988).

Similar preferential codon usage was observed in other amino acid coding systems, and similar preferential codon usage was dramatically observed in *Ureaplasma urealyticum* (Fig. 2A). The use of G or C was very much lower than that of T or A at the third codon position in any codon over the genome. It seems difficult to find functional advantages or disadvantages for organisms to preferentially use certain codons in a degenerated system among chromosomes. Therefore, preferential codon usage in the degenerated codon system cannot be explained by Darwinian evolution. Additionally, this preferential codon usage is not explained by the neutral theory.

The codon usage pattern of the gram-positive *Staphylococcus aureus* resembled that of *U. urealyticum*, although the use of G or C at the third codon position slightly increased synchronously in every codon over the genome compared with *U. urealyticum* (Fig. 2B).

The codon usage patterns of *Mycobacterium tuberculosis* were absolutely inverse to that of *U. urealyticum* (Fig. 2A,C). The usage of G or C at the third codon position was much higher than T or A in every codon, and the increase in G or C and the decrease in A or T changed synchronously compared with *U. urealyticum*. These results cannot be completely explained by either of the classical evolutional theories (Mayr 1965, 2000; Kimura 1977).

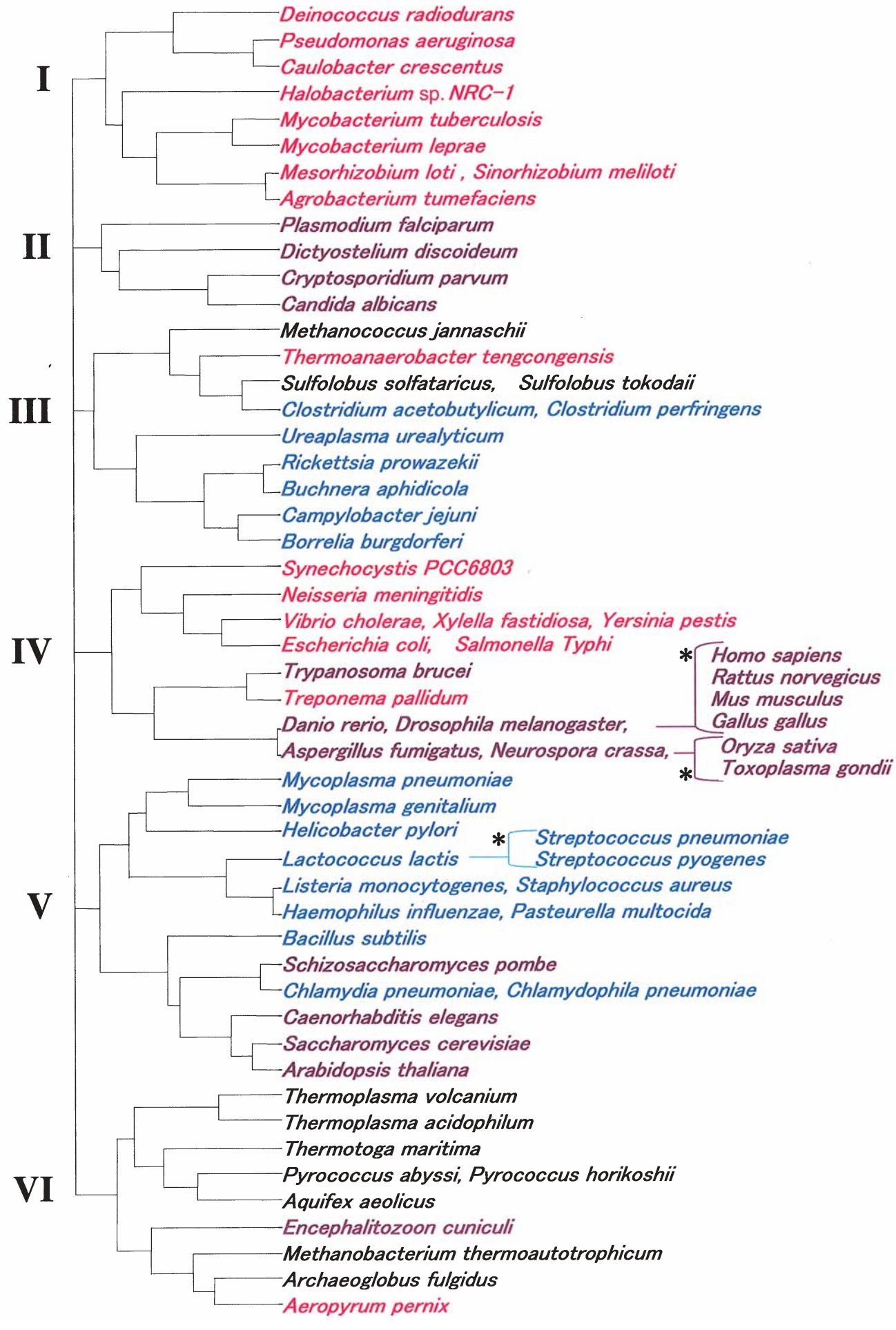
The codon usage pattern of the gram-negative *Escherichia coli* differed from those of *U. urealyticum*, *M. tuberculosis*, and *S. aureus*; the usage of G or C was almost equal to those of T or A at the third codon position, except for the preferential usage of G at the third codon position in the degenerated coding system for L (Leu) (Fig. 2D).

Eukaryotes, *Candida albicans* (fungus, budding yeast), and *Plasmodium falciparum* (protist) showed a very low usage of G or C at the third codon position in any codon (Fig. 2E,F), as observed in *U. urealyticum* (Fig. 2A). Additionally, these patterns differ from those of either *S. cerevisiae* or *E. cuniculi* (see Fig. 1).

To estimate biological divergence among various organisms, we carried out a cluster analysis (Fig. 3). Utilizing Ward's method, which is a widely used practical technique for cluster analysis, each codon frequency (in total, 64 codons) was used as one of the traits to characterize each organism. All organisms examined were classified into six main clusters: I, eubacteria belonging to "E-type" (Sorimachi and Okayasu 2004); II, eukaryotes (protista and fungi); III, archaea and eubacteria belonging to "S-type" (Sorimachi and Okayasu 2004); IV, eubacteria belonging to "E-type" and eukaryotes (protist, fungi, plants, and animals); V, eubacteria belonging to "S-type" and eukaryotes (fungi, plants, and animals); and VI, archaea and eukaryotes (fungus). *S. cerevisiae* (budding yeast) was separated from *C. albicans* (budding yeast), *Aspergillus fumigatus* (filamentous fungus), and other eukaryotes including some fungi, protista, and animals. However, *S. cerevisiae* was close to *Schizosaccharomyces pombe* (fisson yeast) and to *Arabidopsis thaliana* (plant), which was separated from *Oryza sativa* (plant). On the other hand, *E. cuniculi* (microsporidia) was completely separated from other eukaryotes examined and belonged to an archaeon cluster. These results indicate that *S. cerevisiae* and *E. cuniculi* evolved in different directions. Thus, their codon usage patterns differ from each other (see Fig. 1).

In addition, according to Darwin's theory, every species was derived from a single origin. It seems difficult to speculate that *U. urealyticum* evolved to *M. tuberculosis* via certain intermediate bacteria, and that *C. albicans* or *P. falciparum* was derived from *U. urealyticum*, which has a similar codon usage pattern, or was derived from *M. tuberculosis*, which has an inverse codon usage pattern. Namely, it is very hard to speculate that the absolutely inverse codon usage change occurred over the genome during biological evolution. Thus, organisms seem to be derived from

Fig. 3. Dendrogram of various organism classifications obtained utilizing the Ward's method using 64 codons. Codon frequencies were calculated from the data obtained from GenomeNet (hhtp://www.genome.ad.jp) or the frequency data were directly used from the database (http://www.kazusa.or.jp/codon). Cluster analysis was carried out using the software Tahenryou-Kaiseki (multivariate analysis) developed by ESMI (Tokyo, Japan), as an add-in program of EXCEL. In this software, the cluster element was limited to 50, and some samples were classified into the same cluster element using more than 50 samples. *Red characters*, eubacteria belong to "E-type"; *blue characters*, eubacteria belong to "S-type"; *black characters*, archaea; *brown characters*, eukaryotes; *, included in the same cluster element as the preceding organism(s)


I
II
III
IV
V
VI
Deinococcus radiodurans
Pseudomonas aeruginosa
Caulobacter crescentus
Halobacterium sp. NRC-1
Mycobacterium tuberculosis
Mycobacterium leprae
Mesorhizobium loti , Sinorhizobium meliloti
Agrobacterium tumefaciens
Plasmodium falciparum
Dictyostelium discoideum
Cryptosporidium parvum
Candida albicans
Methanococcus jannaschii
Thermoanaerobacter tengcongensis
Sulfolobus solfataricus, Sulfolobus tokodaii
Clostridium acetobutylicum, Clostridium perfringens
Ureaplasma urealyticum
Rickettsia prowazekii
Buchnera aphidicola
Campylobacter jejuni
Borrelia burgdorferi
Synechocystis PCC6803
Neisseria meningitidis
Vibrio cholerae, Xylella fastidiosa, Yersinia pestis
Escherichia coli, Salmonella Typhi
Trypanosoma brucei
Treponema pallidum
Danio rerio, Drosophila melanogaster,
Aspergillus fumigatus, Neurospora crassa,
*
Homo sapiens
Rattus norvegicus
Mus musculus
Gallus gallus
Oryza sativa
*
Toxoplasma gondii
Mycoplasma pneumoniae
Mycoplasma genitalium
Helicobacter pylori
Lactococcus lactis
*
Streptococcus pneumoniae
Streptococcus pyogenes
Listeria monocytogenes, Staphylococcus aureus
Haemophilus influenzae, Pasteurella multocida
Bacillus subtilis
Schizosaccharomyces pombe
Chlamydia pneumoniae, Chlamydophila pneumoniae
Caenorhabditis elegans
Saccharomyces cerevisiae
Arabidopsis thaliana
Thermoplasma volcanium
Thermoplasma acidophilum
Thermotoga maritima
Pyrococcus abyssi, Pyrococcus horikoshii
Aquifex aeolicus
Encephalitozoon cuniculi
Methanobacterium thermoautotrophicum
Archaeoglobus fulgidus
Aeropyrum pernix

different origins, leading to the conclusion that this result cannot be explained by Darwin's theory that every species was derived from a single origin.

In *S. cerevisiae*, both codon usage and amino acid composition patterns coincided with each other, not only in all chromosomes but also in all putative small units constructing the complete genome (Sorimachi and Okayasu 2003). Thus, genomic structure is homogeneous over the genome, not only in prokaryotes but also in eukaryotes. These results strongly suggest that the contribution of gene drifts or of lateral gene transfer to evolution might be very small. It is obvious that *U. urealyticum* and *M. tuberculosis* have completely inverse codon usage patterns (Fig. 2A,C); however, their amino acid composition patterns, represented by a "star shape," resemble each other (Sorimachi and Okayasu 2004).

We found that during biological evolution the basic pattern of cellular amino acid composition is conserved in various organisms from bacteria to mammalian cells (Sorimachi 1999). This basic pattern, which is expressed by a star shape, mostly coincided with that calculated from the complete genome (Sorimachi et al. 2001). These results indicate that codon alteration is strongly controlled internally by the amino acid composition pattern expressed by the star shape, which we hold represents all existing Earth-based organisms. As described above, evolution resulting from mutation seems to be independent of Darwin's or the neutral theory; however, inheritance after mutation is indeed controlled by Darwin's ideas of natural selection in biological evolution.

References

Dawkins R (1995) God's utility function. Sci Am 273:80–85

Dayhoff MO, Park CM, McLaughlin PJ (1972) Building a phylogenetic tree: cytochrome C. In: Atlas of protein sequence and structure, vol 5. National Biomedical Fundation, Washington, DC, pp 7–16

DePouplana LR, Turner RJ, Steer BA, Schimmel P (1998) Genetic code origins: tRNAs older than their synthetases? Proc Natl Acad Sci USA 95:11295–11300

Doolittle WF (1999) Phylogenetic classification and the universal tree. Science 284:2124–2128

Doolittle WF, Brown JR (1994) Tempo, mode, the progenote, and the universal root. Proc Natl Acad Sci USA 91:6721–6728

Gould SJ (1994) The evolution of life on the earth. Sci Am 271:85–91

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature (Lond) 267:275–276

King JL, Jukes TH (1989) Non-Darwinian evolution. Most evolutionary change in proteins may be due to neutral mutations and genetic drift. Science 164:788–798

Maizels N, Weiner AM (1994) Phylogeny from function: evidence from the molecular fossil record that tRNA originated in replication, not translation. Proc Natl Acad Sci USA 91:6729–6734

Mayr E (1965) Animal species and evolution. Harvard University Press, Cambridge

Mayr E (2000) Darwin's influence on modern thought. Sci Ame 283: 79–83

Sogin ML, Elwood HJ, Gunderson JH (1986) Evolutionary diversity of eukaryotic small subunit rRNA genes. Proc Natl Acad Sci USA 83:1383–1387

Sorimachi K (1999) Evolutionary changes reflected by the cellular amino acid composition. Amino Acids 17:207–226

Sorimachi K, Okayasu T (2003) Gene assembly consisting of small units with similar amino acid composition in the *Saccharomyces cerevisiae* genome. Mycoscience 44:415–417

Sorimachi K, Okayasu T (2004) Classification of eubacteria based on their complete genome: where does Mycoplasmataceae belong? Proc R Soc Lond B (Suppl) 271:S127–S130

Sorimachi K, Itoh T, Kawarabayasi Y, Okayasu T, Akimoto K, Niwa A (2001) Conservation of the basic pattern of cellular amino acid composition of archaeo bacteria during biological evolution and the putative amino acid composition of primitive life forms. Amino Acids 21:393–399

Sueoka N (1988) Directional mutation pressure and neutral molecular evolution. Proc Natl Acad Sci USA 85:2653–2657

Woese CR (1998) The universal ancestor. Proc Natl Acad Sci USA 95:6854–6859

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