

SHORT COMMUNICATION

Kenji Sorimachi · Teiji Okayasu

Gene assembly consisting of small units with similar amino acid composition in the *Saccharomyces cerevisiae* genome

Received: June 9, 2003 / Accepted: June 17, 2003

Abstract Amino acid compositions of all genes in *Saccharomyces cerevisiae* were determined using a computer analysis of the complete genome. The amino acid composition of an assembly of several genes or a single gene consisted of 3000–7000 amino acid residues forming a certain pattern of amino acid composition. This rule was independent not only of the gene size, but also of the gene position. Thus, the small units, consisting of 3000–7000 amino acid residues, showed a similar amino acid composition, and they formed all the genes in the complete genome.

Key words Amino acid composition · Gene · Genome · *Saccharomyces cerevisiae*

The basic pattern of cellular amino acid composition is quite similar from bacteria to mammalian cells (Sorimachi 1999; Sorimachi et al. 2000, 2001). In addition, the amino acid composition determined from the complete genome of Archaea resembles that obtained from the amino acid analysis of cells, assuming that all genes are equally expressed in cells (Sorimachi et al. 2001). However, this coincidence has puzzled us for a long time, because the expression of each gene differs in a cell. Therefore, the present study was initiated to solve this puzzle.

Recently, the complete genome of *Saccharomyces cerevisiae* was analyzed by several groups (Oliver et al. 1992; Dojon et al. 1994, 1997; Feldman et al. 1994; Johnston et al. 1994, 1997; Bussey et al. 1995, 1997; Murakami et al. 1995; Galibert et al. 1996; Goffeau et al. 1996; Bowman et al. 1997; Churcher et al. 1997; Dietrich et al. 1997; Jacq et al.

1997; Philippsen et al. 1997; Tettelin et al. 1997). We used their data from GenomeNet (<http://www.genome.ad.jp>) and analyzed them with a personal computer.

The amino acid composition of a single gene consisting of 100–500 amino acid residues differs among various genes (Fig. 1A). Naturally, their amino acid sequences differ from each other. On the other hand, even a single gene consisting of more than 3000 amino acid residues showed a very simi-

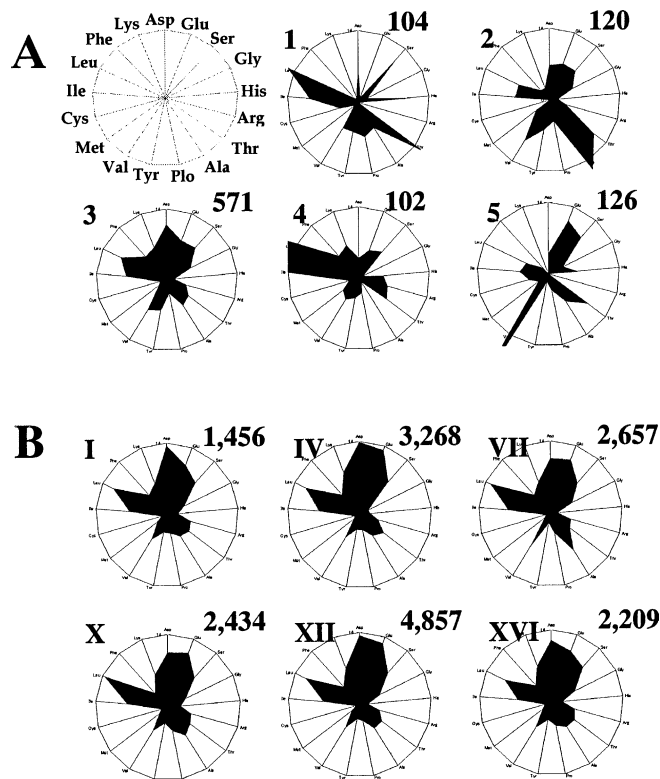


Fig. 1. Radar graphs of amino acid compositions presumed from various genes of *Saccharomyces cerevisiae*. The value is expressed as the percentage of total amino acids. **A** The first 5 genes in chromosome I; **B** the largest gene in each chromosome (I, IV, VII, X, XII, and XVI). The number on the upper right of each graph represents the total number of amino acid residues examined

K. Sorimachi (✉)
Department of Microbiology, Dokkyo University School of
Medicine, Mibu, Tochigi 321-0293, Japan
Tel. +81-28-287-2131; Fax +81-28-286-5616
e-mail: kenjis@dokkyomed.ac.jp

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

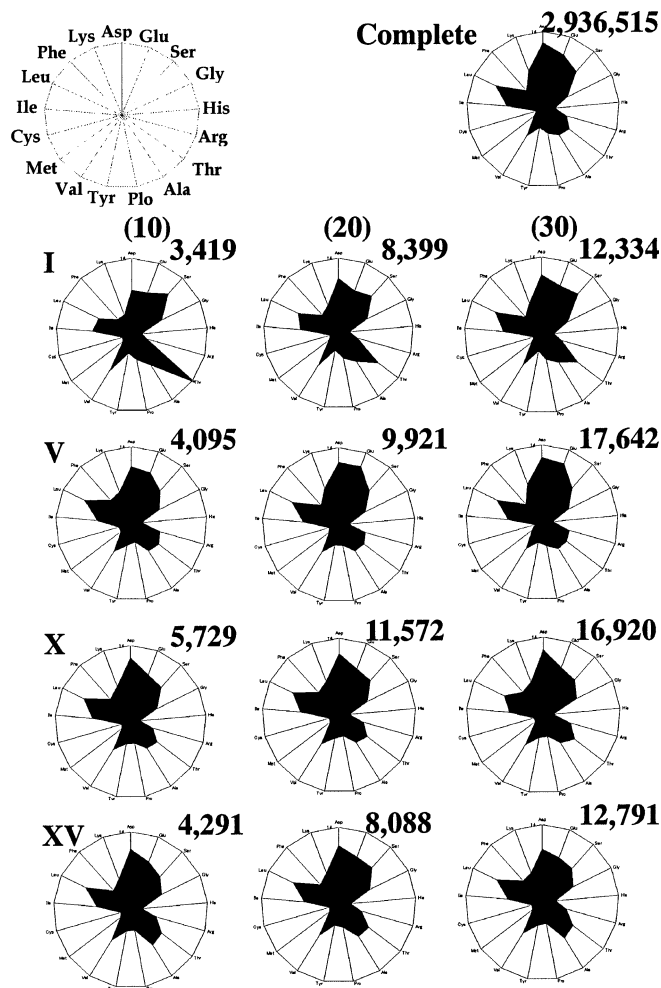


Fig. 2. Radar graphs of the amino acid compositions presumed from the assemblies of 10, 20, and 30 genes of *Saccharomyces cerevisiae*. Chromosomes I, V, X, and XV were used. The first genes were used in chromosomes I and X, and the last genes were used in chromosomes V and XV. The number on the *right upper part* of the graph represents the total amino acid residues examined.

lar amino acid composition. In a certain case, this similarity was observed in even smaller genes coding several hundred amino acid residues (data not shown).

The amino acid composition of the assembly of several proteins resembled that obtained from the amino acid analysis of cells (Sorimachi 1999), which means that the amino acid composition of an assembly of several proteins may represent the cellular amino acid composition. Thus, the amino acid composition based on 10–30 genes was examined (Fig. 2). The amino acid composition more closely resembled that based on the complete genome with the increase in the gene number or amino acid residues in each chromosome. This result suggests that genes composed of apparently small units with similar amino acid composition form the complete genome, although this unit does not have a definite restricted boundary.

The amino acid compositions of 16 chromosomes were presumed from the complete genome. As each chromosome contains many genes coding a large number of amino

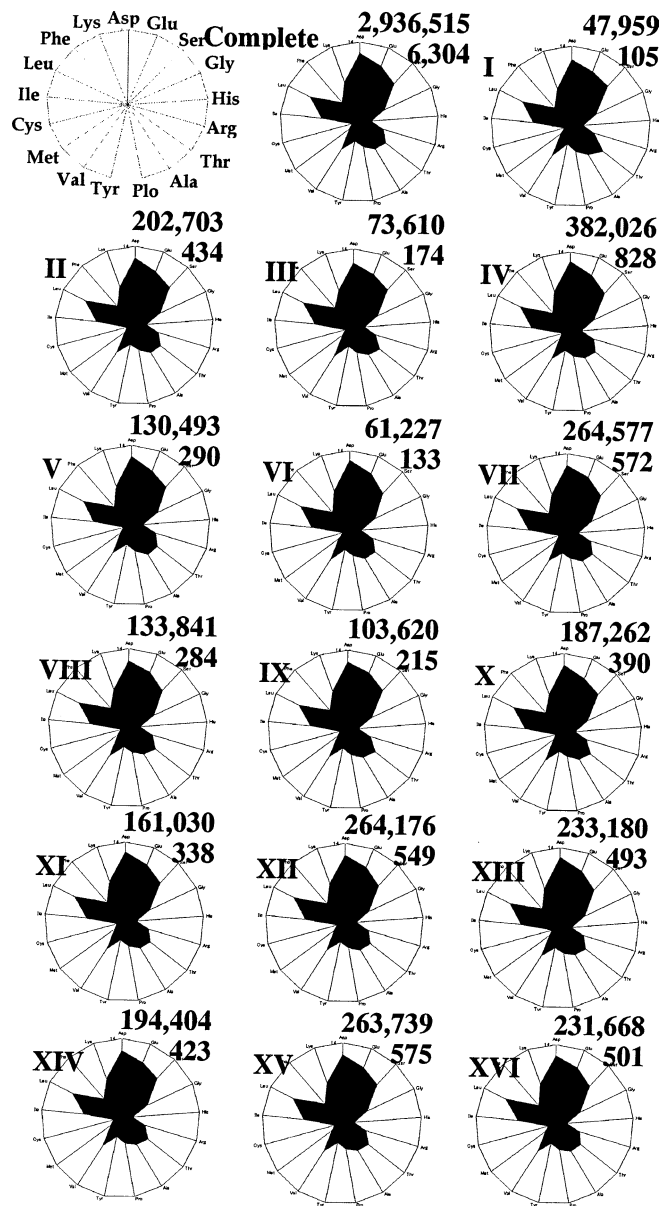


Fig. 3. Radar graphs of the amino acid compositions of *Saccharomyces cerevisiae* chromosomes. The number on the *left upper part* of the graph represents the number of chromosomes, and those on the *right upper part* of the graph represent the total amino acid residues (*upper*) and the total genes (*lower*)

acid residues (48000–265000), the amino acid composition of each chromosome evidently coincided with each other and with that based on the complete genome (Fig. 3). In addition, consistent results were obtained from the complete bacterial genome and mammalian cDNA assembly (data not shown). Thus, the complete genome consists of apparent small units without a certain restricted boundary and with similar amino acid compositions. This observation strongly suggests that synchronous changes occurred in all genes during biological evolution.

Acknowledgments The authors gratefully acknowledge Emeritus Professor Makoto Miyaji of Chiba University for his kind encouragement

and Professor J. Patrick Barron of Tokyo Medical University for reviewing the manuscript.

References

- Bowman S, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIII. *Nature (Lond)* 387:90–93
- Bussey H, et al. (1995) The nucleotide sequence of chromosome I from *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 92:3809–3813
- Bussey H, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI. *Nature (Lond)* 387:103–105
- Churcher C, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX. *Nature (Lond)* 387:84–87
- Dietrich FS, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome V. *Nature (Lond)* 387:78–81
- Dojon B, et al. (1994) Complete DNA sequence of yeast chromosome XI. *Nature (Lond)* 369:371–378
- Dojon B, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XV. *Nature (Lond)* 387:98–102
- Feldman H, et al. (1994) Complete sequence of yeast chromosome II. *EMBO J* 13:5795–5809
- Galibert F, et al. (1996) Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome X. *EMBO J* 15:2031–2049
- Goffeau A, et al. (1996) Life with 6000 genes. *Science* 274:546–567
- Jacq C, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IV. *Nature (Lond)* 387:75–78
- Johnston M, et al. (1994) Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. *Science* 265:2077–2082
- Johnston M, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XII. *Nature (Lond)* 387:87–90
- Murakami Y, et al. (1995) Analysis of the nucleotide sequence of chromosome IV from *Saccharomyces cerevisiae*. *Nat Genet* 10:261–268
- Oliver SG, et al. (1992) The complete DNA sequence of yeast chromosome III. *Nature (Lond)* 357:38–46
- Philippsen P, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XIV. *Nature (Lond)* 387:93–98
- Sorimachi K (1999) Evolutional changes reflected by the cellular amino acid composition. *Amino Acids* 17:207–226
- Sorimachi K, Okayasu T, Akimoto K, Niwa A (2000) Conservation of the basic pattern of cellular amino acid composition during biological evolution in plants. *Amino Acids* 18:193–197
- Sorimachi K, Itoh T, Kawarabayasi Y, Okayasu T, Akimoto K, Niwa A (2001) Conservation of the basic pattern of cellular amino acid composition of archaeobacteria during biological evolution and the putative amino acid composition of primitive life forms. *Amino Acids* 21:393–399
- Tettelin H, et al. (1997) The nucleotide sequence of *Saccharomyces cerevisiae* chromosome VII. *Nature (Lond)* 387:81–84