

STRAND SYMMETRY AND NONRANDOM DNA TRIMER ARRANGEMENT IN YEAST *SACCHAROMYCES CEREVISIAE* CHROMOSOMES

Phan Thi Huyen*, Nguyen Duc Luong

Ho Chi Minh city University of Technology, VNU-HCM, *huyencnshbk@hcmut.edu.vn

ABSTRACT: According to the Chargaff's second parity rule, the eukaryotic chromosomes have been reported to be strand symmetric. However, how the nucleotides are arranged in the chromosome and why the chromosomes possess this strand symmetric property have not been known. We found in *Saccharomyces cerevisiae* that although the frequencies of trimers and those of their respective reverse complements in the whole single stranded sequence of chromosome were almost equal, they were remarkably different in the local regions of chromosome. Results of investigation on trimer arrangement in the intergenic sequences as well as in the sense and antisense sequences of chromosome showed that the frequencies of trimers and those of their respective reverse complements in the intergenic sequences of chromosome were approximately the same, and that the trimer arrangement in the sense and antisense sequences of chromosome shaped the codon usage. The chromosomal strand symmetry is the consequence of the equal distribution of trimers and their respective reverse complements.

Keywords: *Saccharomyces cerevisiae*, intergenic sequences, sense and antisense sequences, trimer, trimer reverse complement, codon usage.

INTRODUCTION

Among the eukaryotic organisms, *Saccharomyces cerevisiae* was chosen for its genome to be first sequenced completely in 1996 [7]. Since then, its chromosomal sequences have become a valuable resource for studying the DNA structure and DNA related biological processes in the eukaryotic cell.

Structurally, the double-stranded DNA chromosomal sequence primarily interacts with histone proteins to form nucleosomes, creating the chromatin fiber. The nucleosome positioning along the yeast chromosomal sequences has been stated to regulate the expression of genes [31]. The formation of chromatin loops which facilitates the interaction of distant regions in the chromosome has also been recognized to orchestrate and regulate the transcription and the recombination [3]. Furthermore, the chromosomal DNA has been found to distribute nonrandomly into territories thereby different DNA regions in the chromosome interact with each other rather than with those in the other chromosomes. These higher-order organizations of chromosomal DNA have been found to be important consequences on the gene activity and function [10, 11, 26, 32]. Nevertheless, the principles

underlying these nonrandom organizations are not clear.

At the nucleotide level, DNA arrangement in the eukaryotic chromosome and consequent effects are not much known. Nucleotide skews, which reflect the difference in distribution of adenine (A) and thymine (T) or guanine (G) and cytosine (C) in the single stranded DNA sequences, have been found to be present in the local regions of chloroplast [14], mitochondrial [27], viral genomes [15], prokaryotic chromosomes [5, 12, 15, 17] and eukaryotic genomes [18]. These skews have also been suggested to be associated with replication and transcription [8, 16, 21, 28-30]. In contrast, nucleotide skews were found not to be present in the yeast chromosomes except the chromosome ends [6]. It was reported that the location of genes on the complementary strands of chromosome explained the presence or absence of nucleotide skews in the chromosome [22], while distribution of genes in chromosome was found to be nonrandom for reducing the gene expression noise [2]. Also, in comparison with frequency of G, McLean and Tirosh recently found that the frequency of C was much higher at the 5' end, decreased at the middle and was much lower at the 3' end of

genes. McLean and Tirosch thus stated that different biases in nucleotide distribution were associated with the initiation, elongation, and termination phases of transcription [13]. Although the nucleotide skews existed differently in the local genomic sequences of both bacterial and eukaryotic chromosomes, these chromosomes were reported to be strand symmetric, i.e., in the whole single stranded sequence of chromosome, numbers of A and T were similar as were those of G and C, showing that the Chargaff's second parity rule held true for the whole single stranded chromosomal sequence [1, 4, 25]. This rule was also found true when it was applied with dimers, trimers and tetramers for the prokaryotic chromosomes, and with oligomers of different sizes for the eukaryotic chromosomes [1, 23]. However, the reason why the chromosomes are strand symmetric remains unknown.

In this study, we present the biased distribution of trimers and their respective reverse complements in the yeast local chromosomal sequences. We show that the chromosomal strand symmetry is the consequence of the equal distribution of trimers and their respective reverse complements in all intergenic sequences as well as in all sense and antisense sequences of chromosome, and that the arrangement of trimers in the sense and antisense sequences of chromosome shapes the codon usage in yeast.

MATERIALS AND METHODS

DNA sequences

Chromosomal sequences of *S. cerevisiae* and data used to extract the sense and antisense sequences as well as the intergenic sequences were taken from NCBI, as files with the .fna and .ptt extensions, respectively, via ftp://ftp.ncbi.nlm.nih.gov/genomes/Fungi/Saccharomyces_cerevisiae_uid128, as of July 2012.

Trimer distribution in the chromosomes

The trimers were counted over the length of single stranded DNA sequence, in the 5' to 3' orientation and in the single-base or three-base shifting manner. The distribution of a trimer

along the chromosomal sequence was viewed by recording the physical positions of that trimer along the sequence as in the previous study [19].

The density of an individual trimer, i.e., the frequency of each trimer in the whole single stranded sequence of chromosome normalized to the frequency in 1 kb, was calculated as in the previous study [20].

Trimer distribution at three trimer positions

Three trimer positions in the sense and antisense sequences were obtained as in the previous study [19]. The frequency values were then plotted for each pair of trimer/trimer's reverse complement.

RESULTS AND DISCUSSION

Distribution of trimers along the yeast *S. cerevisiae* chromosome

We observed that the relative differences between trimer counts obtained by shifting every single base and by shifting 3 bases along the chromosomal sequences were similar (figure 1). This revealed that the trimers were not distributed randomly, but in a yet unidentified order in the chromosome. The strand symmetry in the yeast chromosomes is illustrated by the approximate equivalence between the frequencies of trimers and those of their respective reverse complements in the whole single stranded sequence of chromosome, which were obtained in the single base shifting manner and were then normalized to the frequencies in 1 kb (figure 2). We thus noted the distribution of each trimer and its reverse complement in the local regions of chromosome. As shown in figure 3A, although the frequencies of each trimer in the sufficiently long regions, i.e., regions containing 1000 individual trimers, were quite similar, the frequency of each trimer and that of its reverse complement were remarkably different in the local region of certain length. The bias in distribution of a trimer and its reverse complement was more clearly seen in the narrower local regions (figures 3B and C).

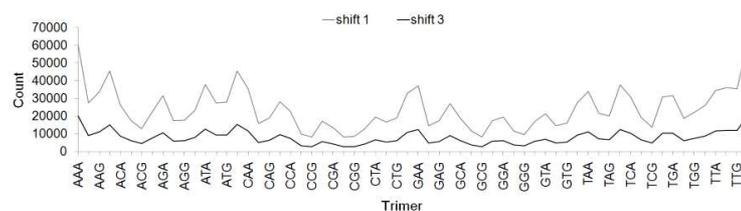


Figure 1. Counts of trimers in the *S. cerevisiae* chromosome 4. Data were plotted for all 64 trimers, however, only 32 of the 64 alphabetically listed trimers are shown due to lack of space. shift 1 and shift 3 denote the one base-pair and three base-pair shifting manners, respectively.

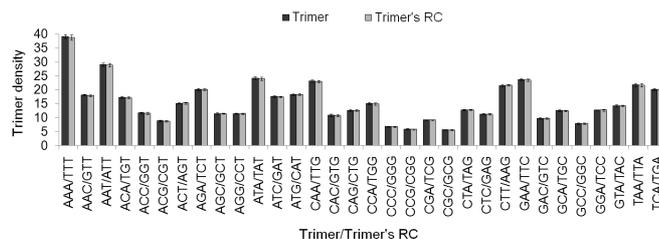


Figure 2. Mean densities of trimer/trimer's reverse complement in yeast chromosomes. Mean density values calculated from the trimer densities in each of the 16 chromosomes. RC stands for reverse complement. Data were plotted for each of 32 pairs of trimer/trimer's RC. Error bars indicate standard deviations.

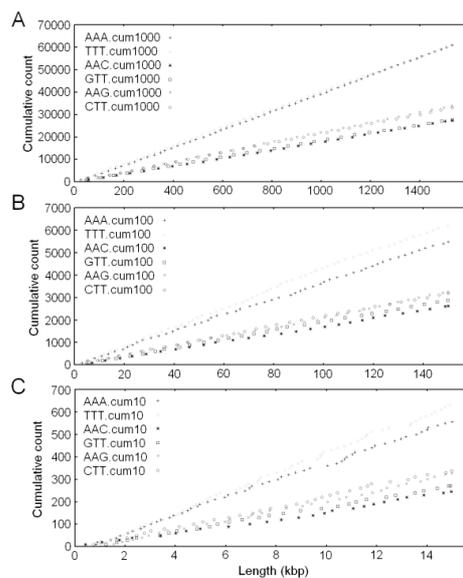


Figure 3. Distribution of trimers and their respective reverse complements along the yeast chromosome 4. The trimer distribution is shown in the whole chromosome (A), first 1/10th of chromosome (chromosomal positions from 1 to 150000) (B) and first 1/100th of chromosome (chromosomal positions from 1 to 15000) (C). For each trimer, its physical positions were recorded along the sequence. The individual trimer's cumulative count was plotted against its positions along the sequence. cum1000, cum100 and cum10 in the legend indicate that the counts were cumulated for every 1000, 100 and 10 individual trimers, respectively, standing before it. Only cumulative counts of 3 out of 32 pairs of trimer/trimer's reverse complement, i.e., AAA/TTT, AAC/GTT and AAG/CTT, are shown.

Distribution of trimers in the sense and antisense sequences and in the intergenic sequences

As we found biases in distribution of trimers and their respective reverse complements in the local regions of chromosome, we checked the distribution of these complementary trimers along the sense and antisense sequences and also along the intergenic sequences.

Along the sequence that was made by connecting all the sense and antisense sequences in the chromosome together, we

counted the occurrences of trimers at three trimer positions as in the previous study [19]. We found that in all yeast chromosomes, the frequencies of individual trimers at the first positions were almost similar to those of their respective reverse complements at the first positions. In contrast, the frequencies of trimers at the second positions were almost the same as those of their respective reverse complements at the third, while the frequencies of trimers at the third positions were almost equal to those of their respective reverse complements at the second (figure 4).

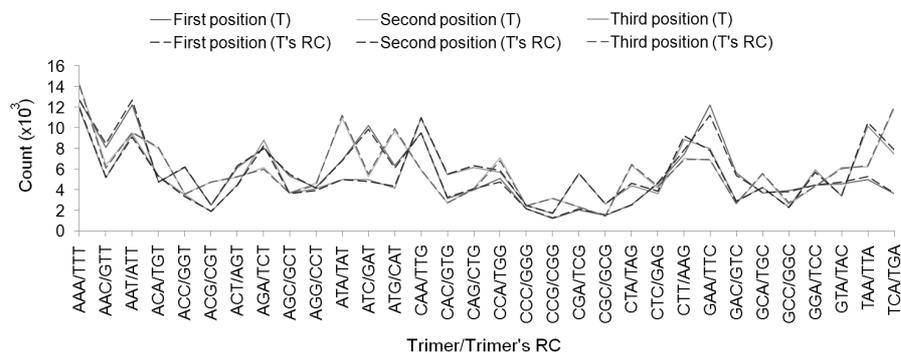


Figure 4. Counts of trimers at three trimer positions in the connected sense and antisense sequence of chromosome 4. Data were plotted for each pair of trimer (T) and its reverse complement (T's RC).

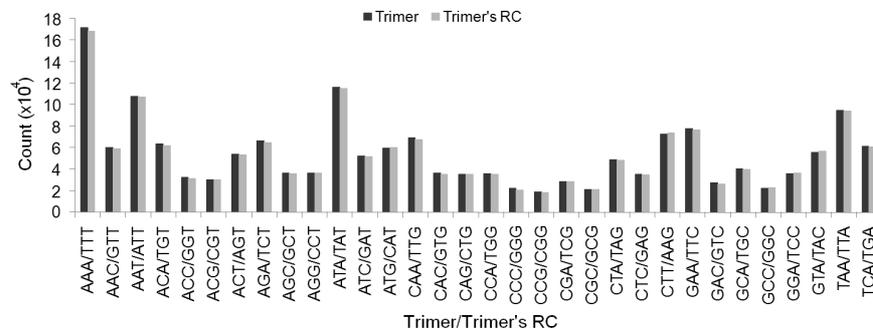


Figure 5. Trimers counts in the intergenic sequences of the yeast genome. Data were plotted for each pair of trimer (T) and its reverse complement (T's RC).

There was no such nucleotide arrangement in the intergenic sequences of chromosome. Nevertheless, the frequency of a trimer and that of its reverse complements in all intergenic sequences extracted from a chromosome were similar. Figure 5 shows the similarity in the frequencies of trimers and their respective

reverse complements in the intergenic sequences extracted from 16 yeast chromosomes.

The codon usage is conventionally defined as the proportions in frequency of individually synonymous codons that encode for an amino

acid in an organism (http://www.geneinfinity.org/sp/sp_codonusage.html). The frequencies of individual codons among every 1000 codons in the chromosome are also defined as codon usage (http://asgpb.hawaii.edu/tools/emboss/EMBOS_SGUI/cgi-bin/emboss.pl?action=manual&app=culp). Similar to the results reported in the previous study [19], the codon usage in yeast was also shaped by the trimer arrangement in the sense and antisense sequences at the chromosomal level (figure 4).

Though the nucleotide skews were not observed in the yeast local chromosomal regions except the chromosome ends [6], we found that the trimer biases, i.e, the differences between the frequencies of trimers and those of their respective reverse complements, were present in these regions. As also reported by the previous studies that the complementary strands of chromosome were symmetric on the basis of dimer and tetramer occurrences [1, 23], that the location of genes on the complementary strands of chromosome resulted in the presence or absence of nucleotide skews in the local regions of chromosome [22], and that the nucleotide skews were opposite between the beginning and the end of genes [13], we sought a rule in which the complementary trimers arranged in the chromosome. We found the similarity between the frequencies of trimers and those of their respective reverse complements not only in the whole single stranded sequence of chromosome but also in the intergenic sequences of chromosome. Consequently, the frequencies of these complementary trimers in the sense and antisense sequences of chromosome were also approximately the same. The findings that the arrangement of trimers in the sense and antisense sequences of chromosome shapes the codon usage in yeast has shed light on the relationship between the nucleotide arrangement in chromosome and the process of transcription, as biases in codon usage have been found to correlate with the level of gene expression [9], which in turn affects the translation efficiency [24]. In reality, the sense and antisense sequences accounted for about from 62 to 76 percents of the yeast

chromosome. Also, the arrangement of trimers in the sense and antisense sequences was found to shape the codon usage in the bacterium *Bacillus cereus*, where the sense and antisense sequences accounted for 85 percents of this bacterial genome [19]. The chromosomal strand symmetry is thus mainly attributed to the approximately equal distribution of codons in the coding sequences encoded by the sense sequences and in those encoded by the antisense sequences, though the arrangement of trimers in the intergenic sequences is still not clear. A relationship between the chromosomal strand symmetry and the gene expression level can now be established. However, detail about this relationship needs to be further investigated.

CONCLUSION

In the yeast *S. cerevisiae* single-stranded chromosomal sequence, the frequencies of trimers and those of their respective reverse complements were almost similar, i.e., the chromosome was strand symmetric according to the Chargaff's second parity rule extended to trimers, due to the similarity in frequencies of trimers and those of their respective reverse complements in all intergenic sequences as well as in all sense and antisense sequences of chromosome, though these trimer frequencies were biased in the local chromosomal sequences. As in the *B. cereus* bacterial chromosome, we also found that trimer arrangement in the sense and antisense sequences of chromosome shaped the codon usage in yeast.

REFERENCES

1. Baisnée P. F., Hampson S., Baldi P., 2002. Why are complementary DNA strands symmetric?. *Bioinformatics* 18: 1021-1033.
2. Batada N. N., Hurst L. D., 2007. Evolution of chromosome organization driven by selection for reduced gene expression noise. *Nat. Genet.*, 39: 945-949.
3. Dekker J., 2008. Gene regulation in the third dimension. *Science*, 319: 1793-1794.
4. Forsdyke D. R., Mortimer J. R., 2000. Chargaff's legacy. *Gene*, 261: 127-137.

5. Freeman J. M., Plasterer T. N., Smith T. F., Mohr S. C., 1998. Patterns of genome organization in bacteria. *Science*, 279: 1827.
6. Gierlik A., Kowalczyk M., Mackiewicz P., Dudek M. R., Cebrat S., 2000. Is there replication-associated mutational pressure in the *Saccharomyces cerevisiae* genome?. *J. Theor. Biol.*, 202: 305-314.
7. Goffeau A., Barrell B. G., Bussey H., Davis R. W., Dujon B., Feldmann H., Galibert F., Hoheisel J. D., Jacq C., Johnston M., Louis E. J., Mewes H. W., Murakami Y., Philippsen P., Tettelin H., Oliver S. G., 1996. Life with 6000 genes. *Science*, 274: 546-567.
8. Green P., Ewing B., Miller W., Thomas P. J., Green E. D., 2003. Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet.*, 33: 514-517.
9. Hiraoka Y., Kawamata K., Haraguchi T., Chikashige Y., 2009. Codon usage bias is correlated with gene expression levels in the fission yeast *Schizosaccharomyces pombe*. *Genes Cells*, 14: 499-509.
10. Iyer B. V. S., Arya G., 2012. Lattice animal model of chromosome organization. *Phys. Rev. E.*, 86: 011911.
11. Iyer B. V. S., Kenward M., Arya G., 2011. Hierarchies in eukaryotic genome organization: insights from polymer theory and simulations. *BMC Biophysics*, 4: 8.
12. Lobry J. R., Sueoka N., 2002. Asymmetric directional mutation pressures in bacteria. *Genome Biol.* 3, research0058.1-0058.14.
13. McLean A. M., Tirosh I., 2011. Opposite GC skews at the 5' and 3' ends of genes in unicellular fungi. *BMC Genomics*, 12: 638.
14. Morton B. R., 1999. Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis*. *Proc. Natl. Acad. Sci. USA*, 96: 5123-5128.
15. Mradzek J., Karlin S., 1998. Strand compositional asymmetry in bacterial and large viral genomes. *Proc. Natl. Acad. Sci. USA*, 95: 3720-3725.
16. Mugal C. F., von Grünberg H. H., Peifer M., 2009. Transcription-induced mutational strand bias and its effect on substitution rates in human genes. *Mol. Biol. Evol.*, 26: 131-142.
17. Necsxulea A., Lobry J. R., 2007. A new method for assessing the effect of replication on DNA base composition asymmetry. *Mol. Biol. Evol.*, 24: 2169-2179.
18. Niu D. K., Lin K., Zhang D. Y., 2003. Strand compositional asymmetries of nuclear DNA in eukaryotes. *J. Mol. Evol.*, 57: 325-334.
19. Phan T. H., Nguyen D. L., 2012. Non-random trimer arrangement in *Bacillus cereus* ATCC 10987 chromosome. *TAP CHI SINH HOC*, 34(3): 354-361.
20. Phan T. H., Nguyen D. L., 2012. Species-specificity of DNA trimer densities in chromosomes and their use in the classification of closely related organisms. *J. Microbiol. Meth.*, 91: 30-37.
21. Polak P., Arndt P. F., 2008. Transcription induces strand-specific mutations at the 5' end of human genes. *Genome Res.*, 18: 1216-1223.
22. Poptsova M. S., Larionov S. A., Ryadchenko E. V., Rybalko S. D., Zakharov I. A., Loskutov A., 2009. Hidden chromosome symmetry: in silico transformation reveals symmetry in 2D DNA walk trajectories of 671 chromosomes. *PLoS ONE* 4: e6396.
23. Qi D., Cuticchia A. J., 2001. Compositional symmetries in complete genomes. *Bioinformatics*, 17: 557-559.
24. Robinson M., Lilley R., Little S., Emtage J.S., Yarranton G., Stephens P., Millican A., Eaton M., Humphreys G., 1984. Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nuc. Acids Res.*, 12: 6663-6671.
25. Rudner R., Karkas J.D., Chargaff E., 1968. Separation of *B. subtilis* DNA into complementary strands. III. Direct analysis.

- Proc. Natl. Acad. Sci. USA, 60: 921-922.
26. Sa'ez-Va'squeza J., Gadal O., 2010. Genome organization and function: a view from yeast and *Arabidopsis*. Mol. Plant, 3: 678-690.
27. Shioiri C., Takahata N., 2001. Skew of mononucleotide frequencies, relative abundance of dinucleotides, and DNA strand asymmetry. J. Mol. Evol., 53: 364-376.
28. Touchon M., Arneodo A., d'Aubenton-Carafa Y., Thermes C., 2004. Transcription-coupled and splicing-coupled strand asymmetries in eukaryotic genomes. Nucleic Acids Res., 32: 4969-4978.
29. Touchon M., Nicolay S., Arneodo A., d'Aubenton-Carafa Y., Thermes C., 2003. Transcription-coupled TA and GC strand asymmetries in the human genome. FEBS Lett. 555: 579-582.
30. Touchon M., Nicolay S., Audit B., Brodie E.B., d'Aubenton-Carafa Y., Arneodo A., Thermes C., 2005. Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. Proc. Natl. Acad. Sci. USA, 102: 9836-9841.
31. Yuan G. C., Liu Y. J., Dion M. F., Slack M. D., Wu L. F., Altschuler S. J., Rando O. J., 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science, 309: 626-630.
32. Zimmer C., Fabre E., 2011. Principles of chromosomal organization: lessons from yeast. J. Cell Biol., 192: 723-733.

**ĐỐI XỨNG NHIỆM SẮC THỂ VÀ SỰ SẮP XẾP KHÔNG NGẪU NHIÊN
CỦA TRIMER TRONG CÁC NHIỆM SẮC THỂ NẤM MEN
*SACCHAROMYCES CEREVISIAE***

Phan Thị Huyền, Nguyễn Đức Lượng

Trường Đại học Bách khoa, ĐHQG tp Hồ Chí Minh

TÓM TẮT

Theo định luật thứ hai của Chargaff, nhiệm sắc thể của các sinh vật có nhân mang tính đối xứng. Tuy nhiên, các nucleotide được sắp xếp như thế nào trong nhiệm sắc thể và vì sao các nhiệm sắc thể mang tính đối xứng vẫn chưa được biết rõ. Ở nấm men *Saccharomyces cerevisiae*, chúng tôi tìm thấy mặc dù tần suất xuất hiện của các trimer và tần suất xuất hiện của các trimer bổ sung đảo ngược tương ứng gần như nhau trong mỗi sợi đơn nhiệm sắc thể nhưng chúng khác nhau rất nhiều trong các đoạn ngắn của sợi đơn này. Kết quả về sự sắp xếp của các trimer này trong các trình tự intergenic cũng như trong các trình tự sense và antisense cho thấy, tần suất xuất hiện của các trimer và tần suất xuất hiện của các trimer bổ sung đảo ngược tương ứng gần như nhau trong toàn bộ các trình tự intergenic của nhiệm sắc thể, đồng thời sự sắp xếp của các trimer trong các trình tự sense và antisense của nhiệm sắc thể định dạng sự sử dụng codon của nấm men. Sự đối xứng của nhiệm sắc thể chính là hệ quả của sự phân bố trimer và trimer bổ sung đảo ngược.

Từ khóa: *Saccharomyces cerevisiae*, trình tự intergenic, trình tự sense và antisense, trimer, trimer bổ sung đảo ngược, sự sử dụng codon.

Ngày nhận bài: 27-9-2012