J. theor. Biol. (2000) **202,** 305–314
doi:10.1006/jtbi.1999.1062, available online at http://www.idealibrary.com on IDEAL®

Is there Replication-associated Mutational Pressure in the *Saccharomyces cerevisiae* Genome?

AGNIESZKA GIERLIK*, MARIA KOWALCZUK*, PAWEŁ MACKIEWICZ*, MIROSŁAW R. DUDEK† AND STANISŁAW CEBRAT*‡

**Institute of Microbiology, Wrocław University, ul. Przybyszewskiego* 63/77, 54-148 *Wrocław, Poland and †Institute of Theoretical Physics, Wrocław University, pl. Maxa Borna* 9, 50-204 *Wrocław, Poland*

(*Received on* 8 *March* 1999, *Accepted in revised form on* 22 *November* 1999)

Compositional bias of yeast chromosomes was analysed using detrended DNA walks. Unlike eubacterial chromosomes, the yeast chromosomes did not show the specific asymmetry correlated with origin and terminus of replication. It is probably a result of a relative excess of autonomously replicating sequences (ARS) and of random choice of these sequences in each replication cycle. Nevertheless, the last ARS from both ends of chromosomes are responsible for unidirectional replication of subtelomeric sequences with pre-established leading/lagging roles of DNA strands. In these sequences a specific asymmetry is observed, resembling the asymmetry introduced by replication-associated mutational pressure into eubacterial chromosomes.

Introduction

When analysing both DNA strands, we observe equal frequencies of the bases: [A] = [T] and [G] = [C] which are called Chargaff's rules (Chargaff, 1950). It was explained by the base pairing rules of Watson & Crick (1953). But surprisingly these equalities are still observed within one strand (Lin & Chargaff, 1967). Any deviations from this intrastrand parity indicate that there is asymmetry in the substitution pattern of the two DNA strands (Sueoka, 1995; Lobry, 1995, 1996a).

There are many asymmetric processes in nucleic acids metabolism which treat the two strands of DNA unequally (for review, see Francino & Ochman, 1997; Mrazek & Karlin, 1998; Frank & Lobry, 1999). One of these processes is replication. There are many replication-related explanations of observed different mutation rates between leading and lagging strands, such as enzymological and architectural asymmetry of the replication fork, differences in processivity between lagging and leading strand complexes and different replication-associated DNA repair of these strands (i.e. Marians, 1992; Kunkel, 1992; Veaute & Fuchs, 1993; Waga & Stillman, 1994; Rosche *et al.*, 1995; Iwaki *et al.*, 1996; Fijalkowska *et al.*, 1998; Radman, 1998). The most substantial cause of spontaneous mutation could be deamination C → T (Frederico *et al.*, 1990; Echols & Goodman, 1991; Lindahl, 1993; Kreutzer & Essigmann, 1998; Frank & Lobry, 1999).

Asymmetric nucleotide substitutions could also occur during transcription of genes (Beletskii & Bhagwat, 1996; Francino *et al.*, 1996; Francino

‡Author to whom correspondence should be addressed. E-mail: cebrat@angband.microb.uni.wroc.pl

0022–5193/00/040305 + 10 $35.00/0

& Ochman, 1997; Freeman *et al*., 1998). Transcription-associated mutational pressure (i.e. substitution $C \rightarrow T$) preferentially affects the non-transcribed strand of genes which stay single-stranded during the transcription.

Moreover, biased gene distribution (Mrazek & Karlin, 1998; Freeman *et al*., 1998; McLean *et al*., 1998) may be responsible for asymmetry in total genomic DNA, because coding sequences have very strong nucleotide compositional bias (Zhang & Zhang, 1991; Dujon *et al*., 1994; Karlin & Burge, 1995; Francino *et al*., 1996; Cebrat *et al*., 1997; Mrazek & Karlin, 1998; Wang, 1998; McClean *et al*., 1998).

As a consequence of all these processes, specific compositional bias in long stretches of DNA has been observed in many eubacterial chromosomes (Lobry, 1996a, b; Blattner *et al*., 1997; Kunst *et al*., 1997; Fraser *et al*., 1997; Fraser *et al*., 1998; Anderson *et al*., 1998; Freeman *et al*., 1998; Mrazek & Karlin, 1998; Grigoriev, 1998; McLean *et al*., 1998; Rocha *et al*., 1999; Mackiewicz *et al*., 1999a, b). The asymmetry was detected also in many viruses (Daniels *et al*., 1983; Filipski, 1990; Mrazek & Karlin, 1998; Grigoriev, 1999).

In Archea generally asymmetry is not detected or it is weak (Karlin *et al*., 1998; McLean *et al*., 1998; Mrazek & Karlin, 1998; Grigoriev, 1998). These results suggest that archeal genomes, like eucaryotic ones, have many origins of replication (Olsen & Woese, 1997; Mrazek & Karlin, 1998). However, some studies provide evidence for one origin of replication in some Archea (Salzberg *et al*., 1998; Lopez *et al*., 1999).

Differences in mechanisms of replication of yeast and bacterial chromosomes can explain why the asymmetry observed in bacterial chromosomes is not observed in yeast chromosomes. Eubacterial chromosomes are usually circular and have only one origin and one terminus of replication. That is why in each replication cycle the same strand of a given DNA sequence plays always the same, lagging or leading role. In yeast chromosomes, there are many autonomously replicating sequences (ARS) which are probably in excess (Dershowitz & Newlon, 1993) and not all of them are active in initiation of replication (Dubey *et al*., 1991). Moreover, even these active ARSs are not used in each replication cycle (Deshpande & Newlon, 1992). The choice of these sequences in each replication cycle is more or less random. What is more, replication does not start simultaneously at all chosen ARSs and therefore termination points of replication are not fixed, either (Greenfeder & Newlon, 1992). All these features prevent fixation of the leading/lagging role of DNA strands in yeast chromosomes. Nevertheless, even if this explanation is true for the most of "interior" chromosomal sequences, it does not concern the sequences outside the last ARS at the end of a chromosome. In these sequences, the leading/lagging role is pre-established and cannot be changed from one replication cycle to another. Thus, if there exists any mutational pressure associated with replication of yeast chromosomes, it should be possible to show compositional bias between 3′ and 5′ ends of DNA strands at the ends of chromosomes.

Analysed eucaryotes, including yeast, do not show strong strand bias (Karlin *et al*., 1998; Mrazek & Karlin, 1998; Grigoriev, 1998). However, Grigoriev (1998) has observed local minima of GC skew near yeast telomeres but not at other ARS sites. However, the effect of replication-associated mutational pressure near ends of yeast chromosomes was not analysed in detail.

Databases and Methods

The *Saccharomyces cerevisiae* genome sequences (Goffeau *et al*., 1997) and information on yeast telomeric and subtelomeric sequences were downloaded on June 24, 1998 from MIPS database http://speedy.mips.biochem.mpg.de. Sequence of *Borrelia burgdorferi* genome (Fraser *et al*., 1997) was downloaded on March 03, 1998 from http://www.tigr.org. After the retrieval, the data have not been updated.

Nucleotide composition of the studied genomes was analysed by DNA walks (Cebrat & Dudek, 1998; Cebrat *et al*., 1998; Mackiewicz *et al*., 1999a, b). DNA sequence was read by a virtual "walker" which moved along the analysed sequence in $5' \rightarrow 3'$ direction. Every step of the walker was associated with a unit shift in two-dimensional space. The direction of the shift depended on the type of walk and nucleotide visited. For the $G \leftrightarrow C$ walk, the shifts of the walker were (1, 1) for G, (1, −1) for C, (1, 0) for A and T. For the $A \leftrightarrow T$ walk, on the contrary,

the shifts were (1, 0) for G and C, (1, 1) for A, and (1, −1) for T. The result of the walk was a plot which visualized nucleotide composition of the analysed sequence, i.e. if a sequence was richer in C than in G, the walker in the G↔C walk went down.

Since the trends introduced by coding functions and coding ORFs into the analysed sequence are very strong and may mask a possible asymmetry of strands resulting from mutational pressure, we performed a kind of "detrended walks". In this kind of walks, each up or down movement of the walker was corrected by a factor which allowed the walker to finish the walk at value $y = 0$. In this way local trends in nucleotide composition were shown (for details see Cebrat & Dudek, 1998).

Analysing DNA walks, it is possible to distinguish between the asymmetry introduced by replication and the asymmetry introduced by transcription and coding functions (Cebrat *et al*., 1998; Mackiewicz *et al*., 1999a). DNA walks enable indication of the origin and the terminus of replication in eubacterial chromosomes. Detrended DNA walks show relative abundance of analysed parameter [A–T] and [G–C] and are different from previously, commonly used analysis of GC and AT skews in sliding windows (Lobry, 1996a). The DNA walks are less prone to fluctuation and bias in local trends and do not include strong trends introduced by coding function. The applied methods are different from cumulative skew diagrams used by Grigoriev (1998) as well. Numbers on the *Y*-axis indicate local relative cumulative abundance of A over T and G over C in the analysed sequence independently of the whole base composition of analysed sequence and its length. A more detailed description of the methods is available at our web site http://smORFland.microb.uni.wroc.pl.

Results

In all studied eubacterial chromosomes significant asymmetry in nucleotide composition has been noticed. For example in Fig. 1 a DNA walk for the Watson strand of *Borrelia burgdorferi* is shown. The *B. burgdorferi* chromosome is linear and the beginning of the plot represents one end of the chromosome. The walker starts at 5′ end of chromosome, moves along the strand in 5′ → 3′ direction and ends its analysis at the 3′ end. Numbers on the *Y*-axis represent differences from the mean value of [A–T] and [G–C] in the analysed region. This DNA walk shows specific asymmetry in G/C and A/T composition. The first part of the plot represents the lagging DNA strand which is richer in C than in G and richer in A than in T. The second part represents the leading strand which is poorer in C than in G and poorer in A than T. The extrema in the middle of the plot indicate the origin of bi-directional replication. Thus, this point represents the switch between lagging and leading fragments of the same strand.

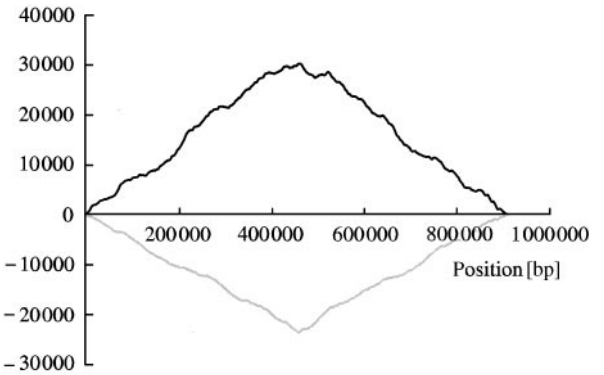

FIG. 1. DNA walks for *B. burgdorferi* genome. For G↔C walk, walkers moved one unit up when the visited nucleotide was G or one unit down when the nucleotide was C. In the A↔T DNA walks the walker moved up when the visited nucleotide was A and down when the visited nucleotide was T. In the *B. burgdorferi* the genome walker starts at the beginning of the linear genome and the extrema indicate the origin of replication. ——, (A–T); —— , (G–C).

Similar walks performed for yeast chromosomes, which have many origins of replication, did not show such strong asymmetry (Fig. 2). Changes of [A–T] and [G–C] values seem to be more random. Nevertheless, some of ARS positions correlate with extrema. It is difficult to prove that the overlapping of extrema with known ARS consensus sequences is not accidental and that the observed local asymmetries in nucleotide compositions of DNA strands are statistically significant and connected with replication-associated mutational pressure. Some authors argue that there is no asymmetry in the yeast chromosomes (Mrazek & Karlin, 1998). One, but only partial explanation could be that not all ARS are active and used as initiators of replication (Dubey *et al*., 1991).

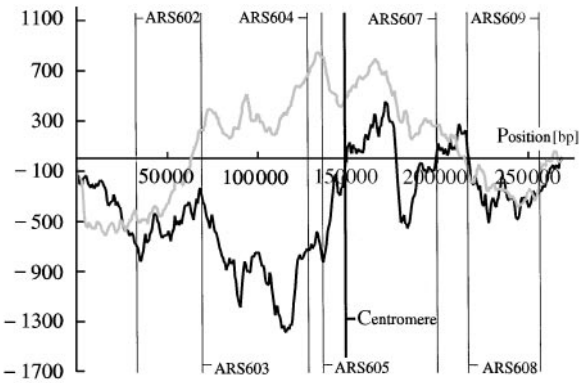

FIG. 2. DNA walks for yeast chromosome 6. Vertical lines show the positions of some ARS. Heavy vertical line indicates the position of the centromere. Numbers on X-axis indicate position in base pairs of analysed sequence. The description of DNA walks as for Fig. 1. ——, (A–T); ——, (G–C).

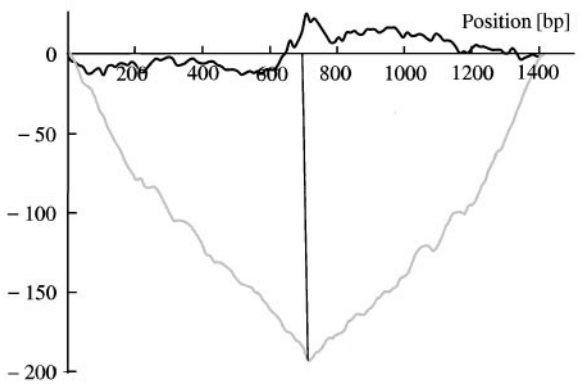

FIG. 3. DNA walks on sequence obtained by splicing two sequences from both ends of chromosome 3 located between the inner border of telomeres and the ARS closest to the telomeres. ——, (A–T); ——, (G–C).

To find out if the mutational pressure takes effect during yeast chromosome replication, we have measured compositional bias of the ends of chromosomes. The part of chromosome between the last ARS and the end has to be replicated in the same direction in each replication cycle, like a eubacterial chromosome. Thus, the roles of strands as leading or lagging are fixed. Therefore, we have performed DNA walks on two spliced sequences from the two ends of chromosome 3 located between the inner border of telomeres and the ARS closest to the telomeres (Fig. 3). On the left of the vertical line indicating the ARS closest to the telomeres is the sequence from 5′ end and on the right from 3′ end of the chromosome. The asymmetry in G/C contents between these sequences is evident. The 5′ ends correspond to the strand replicated as lagging, richer in C than G. The 3′ ends are replicated as leading and are richer in G than C.

To check how general the asymmetry at chromosomal termini is, we analysed both ends of all yeast chromosomes up to a few thousand base pairs. Analysing Watson strands, we spliced 6000 bp long sequences from the 5′ ends of all 16 chromosomes and joined them to spliced sequences from the 3′ ends of these chromosomes. Since telomeric repetitions are asymmetric and conservative in nucleotide composition: $C_{2-3}ACA_{1-6}/T_{1-6}GTG_{2-3}$ (Shampay *et al.*, 1984; Wang & Zakian, 1990) and introduced by a mechanism different to replication of the rest of chromosome sequences, we cut off telomeres from the analysed sequences. On so prepared sequence, we performed DNA walks [Fig. 4(a)] which showed asymmetry in G/C and A/T composition. We also analysed the yeast chromosomes, checking the asymmetry in regions of specific length and positions from the ends of chromosomes. We found that the asymmetry disappears at the distance of about 10 000 bp from the chromosome end. In Fig. 4(b), we have shown the DNA walks done for the sequence which was the result of splicing 6000 bp long sequences located between 12 000 and 18 000 bp from ends of chromosomes. Note that this analysis was done for the whole set of chromosomes. Individual chromosomes can differ in the length of asymmetric sequences which probably depends on location and activity of the first ARS.

However, the observed asymmetry at the ends of the chromosome shown in Fig. 4(a) may be introduced by specific subtelomeric sequences (Y′ elements, X elements and STR—A–D) which are located at the ends of chromosomes. These sequences are also asymmetric and have nonrandom directional arrangement, which means that for example Y′ element at the 5′ end of Watson strand is complementary to the Y′ element from the 3′ end of the same strand (Britten, 1998). This implicates that [G–C] values of these two elements have reciprocal signs. To show the effect of subtelomeric sequences on the analysed asymmetry, we spliced these sequences from all chromosomes, first from 5′ ends, followed by

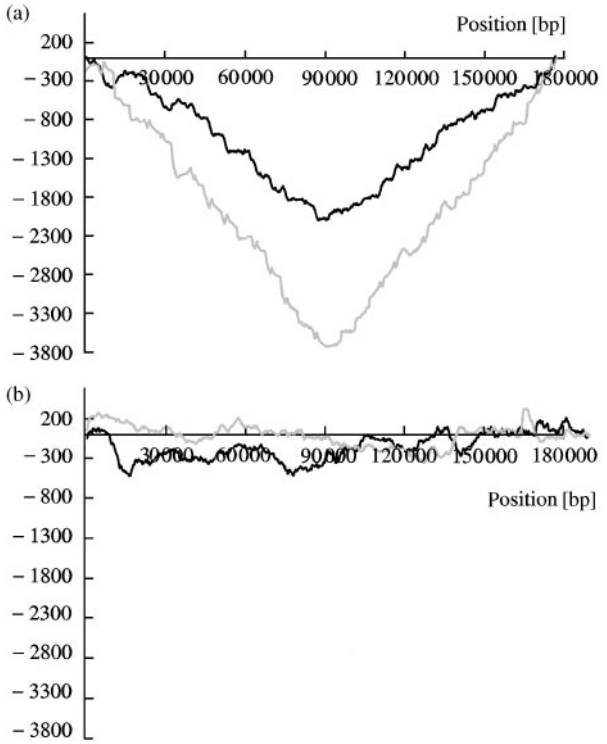

FIG. 4. DNA walks on the spliced ends of all 16 chromosomes. In the first half of the sequence the fragments from 5′ ends of Watson strand were spliced, in the second part, the fragments from 3′ ends of the same strands were spliced; *a*—sequences up to 6000 bp from the chromosome ends after cutting off telomeres; *b*—sequences from the windows between 12 000–18 000 bp from the chromosome ends. The description of DNA walks as for Fig. 1. ——, (A–T); ——, (G–C).

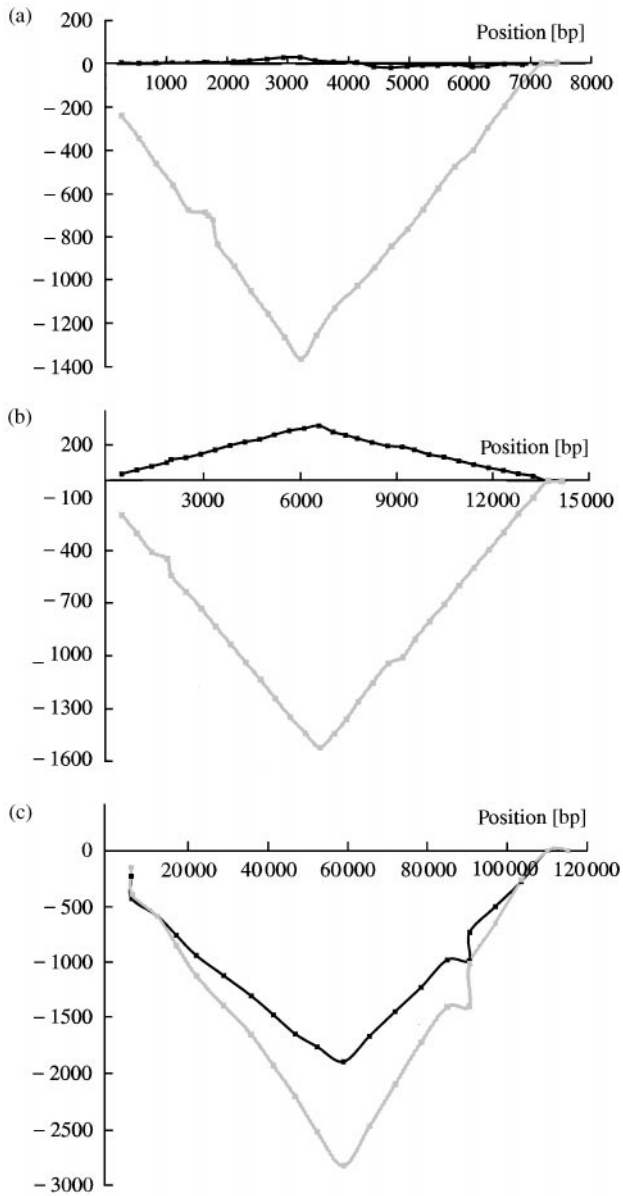

FIG. 5. DNA walks on different sub-telomeric sequences from yeast chromosomes; *a*—STR A, B, C, D elements; *b*—X elements; *c*—Y′ elements. The description of DNA walks as for Fig. 1. ——, (A–T); ——, (G–C).

these from 3′ ends. On so arranged sequences we did DNA walks. In Fig. 5(a–c), the DNA walks for STR elements, X elements, and Y′ elements, respectively, are shown. All subtelomeric sequences are strongly asymmetric in G/C composition. Y′ elements mostly contribute to the bias of chromosome ends because they are the longest. However, STR elements show the strongest asymmetry.

To show the specific asymmetry of different telomere-associated elements we have also grouped individual elements according to their types. There are two main types of chromosome ends in the yeast chromosomes. One type with Y′ element and the second one without this element. Types of elements for individual chromosomes are listed in Table 1 (terminology according to Britten, 1998). In Fig. 6, we have shown G↔C and A↔T walks for subtelomeric sequences with Y′ elements. In these walks, each subtelomeric element was read in the 5′–3′ direction starting from the end of the chromosome. Inside each subtelomeric Y′ element a local maximum in G↔C walk was found. This maximum was used to make the alignment of sequences. Using this characteristic point for alignment it is easy to visualize the homology between sequences. There are two groups of Y′ elements: the first one referred to as long Y′ and the second one as deleted or

TABLE 1
List of types of subtelomeric sequences present on termini of S. cerevisiae *chromosomes*

Type of left end	Chromosome no.	Type of right end
X2	1	X2
Y′ short	2	X2
X2	3	X2
X2	4	Y′ long
Y′ long	5	Y′ long
Y′ long	6	X2
X2	7	Y′ short
Y′ short	8	Y′ short
Y′ long	9	X2
Y′ long	10	X2
X2	11	X2
Doubled Y′ short	12	Doubled Y′ long
Y′ short	13	X2
Y′ long	14	X2
X2	15	Y′ long
Y′ long	16	Y′ short

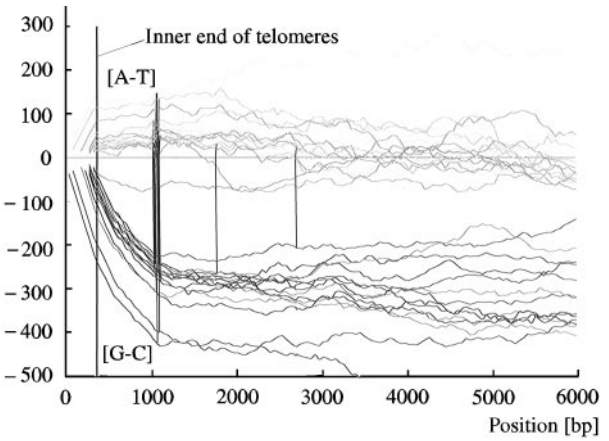

FIG. 7. DNA walks for individual chromosome termini with sub-telomeric X2 repeat elements. Each fragment was read from the chromosome end in 5′ to 3′ direction (left arm—Watson strand, right arm—Crick strand). DNA walks were aligned according to the position of inner border of telomeres (vertical bold line). Vertical lines mark positions of ARS.

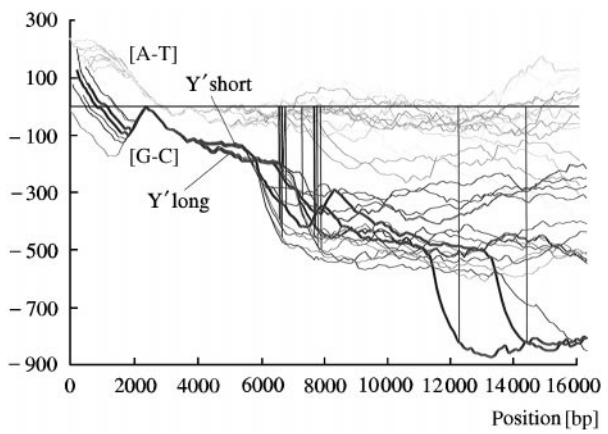

FIG. 6. DNA walks for individual chromosome termini with Y′ elements. Each fragment was read from the chromosome end in 5′ to 3′ direction (left arm—Watson strand, right arm—Crick strand). DNA walks were aligned according to the position of a local maximum in G↔C walk, i.e. around 2200 bp. The two walks marked with bold lines represent the left and the right arm of chromosome 12 with doubled Y′ short and Y′ long elements, respectively. Vertical lines mark positions of ARS. Note the regular differences in lengths of the slopes to the left of the local maximum, which are multiplicity of 72 bp (see the text).

short Y′ elements (Louis & Harber, 1990). After the alignment, all ARS consensus sequences annotated in MIPS formed two distinct groups—the first one for short Y′ and the second for long Y′. However, there are two exceptions from this localization, both for chromosome 12. This chromosome has doubled Y′ long element on its left arm and doubled Y′ short element on its right arm. The ARS are located in homologous positions in respect to the internal ends of these sequences.

In Fig. 7, DNA walks for terminal regions referred to as "just sub-telomeric sequences" and "X2" repeats (Britten, 1998) are shown. In these walks sequences were aligned according to the internal ends of telomere sequences (added by telomerase). This alignment shifts all ARS annotated in MIPS and located near this region into a very narrow range.

All ARS, associated with both X2 and Y′ elements are localized at the internal ends of sequences replicating as lagging strands and richer in C than in G. There is only one part of Y′ element (external to the local maximum seen in Fig. 6) which has reciprocal G/C skew. In this region an ORF annotated in MIPS database is localized, but in our opinion this sequence plays a structural role rather than coding. There are many premises suggesting its structural role:

- it consists of 72 bp repetitions in which 36 bp or even 12 bp repetitions are recognizable,
- inside these repetitions some positions are extremely conservative (100% consensus),

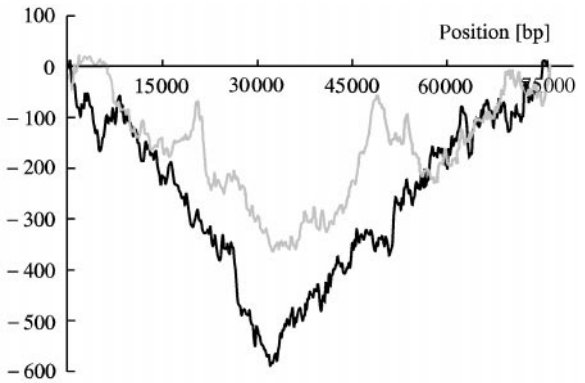

FIG. 8. DNA walks on spliced chromosome ends as for Fig. 4(a) but deprived of telomeres and any recognized subtelomeric sequences. The description of DNA walks as for Fig. 1. ▬, (A–T); ▬, (G–C).

- when these sequences are analysed in triplets as coding sequences, many conserved positions are at the third positions of codons while some, not conserved, are at the first or second codon positions,
- 72 bp repetitions are organized in higher order and, for example, the external Y′ element of left arm of chromosome 12 has two 144 bp repetitions with 135 matched positions,
- Y′ elements from different chromosomes differ in length and these differences are always the multiplicity of 72 bp (is it not one turn around the nucleosome?).

We have also repeated DNA walks on sequences shown in Fig. 4 deprived of any known (indicated in MIPS databases) subtelomeric sequences. In such sequences without telomeres and subtelomeric elements, asymmetry was still observed (Fig. 8). This asymmetry concerns G/C as well as A/T bias. However, it is possible that in-between these residue sequences some other, not recognized subtelomeric sequences selected for there function are present.

Discussion

At all types of chromosome terminals in yeast, the ARS are linked to subtelomeric sequences. Distally to these ARSes, the nucleotide asymmetry in complementary DNA strands is observed. However, the internal sequences do not show any consistent asymmetry in nucleotide composition. In all bacterial genomes, where origins and termini of replication are well defined, the asymmetry between leading and lagging strands is found. This has not been observed in the yeast genome. There are some possibilities which could explain this situation. The first one is that there is no replication-associated mutational pressure in this genome or it is very weak. It is hard to accept this hypothesis since the two strands replicate by different mechanisms and under varying living conditions the mutational pressure should be different for the two strands.

The second hypothesis states that the replication-associated mutational pressure exists but the mode of yeast chromosome replication renders it possible to avoid the asymmetry caused by this pressure. The whole interior of chromosomes has no pre-established leading/lagging role for strands. It is due to a random choice of ARS for a given replication cycle. ARS linked to the terminal regions seem to be active at least when introduced into plasmid construct (Chan & Tye, 1980). If we assume that in some cycles they are active and in some they are "replicated through" when ARS located more inside chromosome start replication earlier, the only sequence with pre-established leading/lagging role is the sequence distal to the terminus linked ARS. Then only the short terminal sequences distal to the ARS linked to subtelomeric sequences are prone to cumulate mutations introduced by replication-associated mutational pressure. However, it is difficult to resolve if the asymmetry observed in subtelomeric sequences is simply the result of selective pressure on their function or only mutational pressure. Most probably specific base composition evolved under constant mutational pressure and composition of subtelomeric sequences has adapted to this mutational pressure.

Other source of compositional bias may be asymmetric nucleotide substitutions occurring during transcription of genes. The single-stranded non-transcribed (sense) strand is more exposed to mutational pressure (i.e. substitution C → T) and the transcribed (antisense) strand is preferentially repaired (Beletskii & Bhagwat, 1996; Francino *et al*., 1996; Francino & Ochman, 1997; Freeman *et al*., 1998; Hanawalt, 1991). This cause of compositional bias of subtelomeric

sequences may be excluded because ORF in these regions are probably non-coding (Mackiewicz *et al*., 1999c).

The asymmetry introduced by replication may decrease in inner parts of chromosomes not only by activity of many origins of replication and their random choice in replication cycle. Also high level of rearrangements or horizontal transfer of foreign DNA may diminish established compositional bias and introduce fluctuations (Grigoriev, 1998). Traces of many duplications and interchromosomal recombinations events have been found in the yeast genome (Lalo *et al*., 1993; Wolfe & Shields, 1997; Seoighe & Wolfe, 1998; Keogh *et al*., 1998; Skrabanek & Wolfe, 1998).

In the eubacterial chromosome the asymmetry connected with leading/lagging role of DNA strands is so strong that codon composition and even amino acid composition of proteins coded by different strands are significantly different (McInerney, 1998; Lafay *et al*., 1999; Rocha *et al*., 1999; Mackiewicz *et al*., 1999a). In the *B. burgdorferi* genome whose asymmetry is shown in Fig. 1, it is even relatively easy to recognize, from the nucleotide composition of a gene, whether it is located on leading or lagging DNA strand. This relation between gene composition and its location on chromosome is probably a strong constraint against translocation, illegitimate recombination and inversion and can give the effect of gene instability in a new position. That is why the general maps of eubacterial genomes are rather conservative (Wilkins, 1988). These effects should be weaker in eukaryotic genomes rendering them more flexible. The lack of asymmetry and the mode of replication with many ARS in eukaryotes may be an adaptation to replication-associated mutational pressure. Avoiding the state with long stretches of DNA with unbalanced base composition may be an advantage for maintaining structure of long chromosomes.

This work was supported by The State Committee for Scientific Research, grant no. 6 P04A 030 14.

REFERENCES

ANDERSSON, S. G., ZOMORODIPOUR, A., ANDERSSON, J. O., SICHERITZ-PONTEN, T., ALSMARK, U. C., PODOWSKI, R. M., NASLUND, A. K., ERIKSSON, A. S., WINKLER, H. H. & KURLAND, C. G. (1998). The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396,** 133–140.

BELETSKII, A. & BHAGWAT, A. S. (1996). Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in *Escherichia coli*. *Proc. Nat. Acad. Sci. U.S.A.* **93,** 13919–13924.

BLATTNER, F. R., PLUNKETT III, G., BLOCH, C. A., PERNA, N. T., BURLAND, V., RILEY, M., COLLADO-VIDES, J., GLASNER, J. D., RODE, Ch. K., MAYHEW, G. F., GREGOR, J., DAVIS, N. W., KIRKPATRICK, H. A., GOEDEN, M. A., ROSE, D. J., MAU, B. & SHAO, Y. (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1462.

BRITTEN, R. J. (1998). Precise sequence complementarity between yeast chromosome ends and two classes of just subtelomeric sequences. *Proc. Nat. Acad. Sci. U.S.A.* **95,** 5906–5912.

CEBRAT, S. & DUDEK, M. R. (1998). The effect of DNA phase structure on DNA walks. *Eur. Phys. J. B* **3,** 271–276.

CEBRAT, S., DUDEK, M. R., MACKIEWICZ, P., KOWALCZUK, M. & FITA, M. (1997). Asymmetry of coding versus non-coding strands in coding sequences of different genomes. *Microb. Comp. Genomics* **2,** 259–268.

CEBRAT, S., DUDEK, M. R., GIERLIK, A., KOWALCZUK, M. & MACKIEWICZ, P. (1998). Effect of replication on the third base of codons. *Physica A* **265,** 78–84.

CHAN, C. S. & TYE, B. K. (1980). Autonomously replicating sequences in *Saccharomyces cerevisiae*. *Proc. Nat. Acad. Sci. U.S.A.* **77,** 6329–6333.

CHARGAFF, E. (1950). Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **6,** 201–240.

DANIELS, D. L., SANGER, F. & COULSON, A. R. (1983). Features of bacteriophage lambda: analysis of the complete nucleotide sequence. *Cold Spring Harb. Symp. Quant. Biol.* **47,** 1009–1024.

DERSHOWITZ, A. & NEWLON, C. S. (1993). The effect on chromosome stability of deleting replication origins. *Mol. Cell. Biol.* **13,** 391–398.

DESHPANDE, A. M. & NEWLON, C. S. (1992). The ARS consensus sequence is required for chromosomal origin function in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12,** 4305–4313.

DUBEY, D. D., DAVIS, L. R., GREENFEDER, S. A., ONG, L. Y., ZHU, J., BROACH, J. R., NEWLON, C. S. & HUBERMAN, J. A. (1991). Evidence suggesting that the ARS elements associated with silencers of the yeast mating type locus, HML, do not function as chromosomal DNA replication origins. *Mol. Cell Biol.* **11,** 5346–5355.

DUJON, B., ALEXANDRAKI, D., ANDRE, B., ANSORGE, W., BALADRON, V., BALLESTA, J. P., BANREVI, A., BOLLE, P. A., BOLOTIN-FUKUHARA, M., BOSSIER, P. (1994). Complete DNA sequence of yeast chromosome XI. *Nature* **369,** 371–378.

ECHOLS, H. & GOODMAN, M. F. (1991). Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* **60,** 477–511.

FIJALKOWSKA, I. J., JONCZYK, P., MALISZEWSKA TKACZYK, M., BIALOSKORSKA, M. & SCHAAPER, R. M. (1998). Unequal fidelity of leading strand and lagging strand DNA replication on the *Escherichia coli* chromosome. *Proc. Nat. Acad. Sci. U.S.A.* **95,** 10 020–10 025.

FILIPSKI, J. (1990). Evolution of DNA sequences. Contributions of mutational bias and selection to the origin of chromosomal compartments. In: *Advances in Mutagenesis Research* 2 (Obe, G., ed.), pp. 1–54. Berlin: Springer Verlag.

FRANCINO, M. P. & OCHMAN, H. (1997). Strand asymmetries in DNA evolution. *Trends Genet*. **13,** 240–245.

FRANCINO, M. P., CHAO, L., RILEY, M. A. & OCHMAN, H. (1996). Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science* **272,** 107–109.

FRANK, A. C. & LOBRY, J. R. (1999). Asymmetric substitution patterns: a review of possible underlying mutational or selective mechanisms. *Gene* **238,** 65–77.

FRASER, C. M., CASJENS, S., HUANG, W. M., SUTTON, G. G., CLAYTON, R., LATHIGRA, R., WHITE, O., KETCHUM, K. A., DODSON, R., HICKEY, E. K., GWINN, M., DOUGHERTY, B., TOMB, J. F., FLEISCHMANN, R. D., RICHARDSON, D., PETERSON, J., KERLAVAGE, A. R., QUACKENBUSH, J., SALZBERG, S., HANSON, M., VAN VUGT, R., PALMER, N., ADAMS, M. D., GOCAYNE, J. & VENTER, J. C. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390,** 580–586.

FRASER, C. M., NORRIS, S. J., WEINSTOCK, G. M., WHITE, O., SUTTON, G. G., DODSON, R., GWINN, M., HICKEY, E. K., CLAYTON, R., KETCHUM, K. A., SODERGREN, E., HARDHAM, J. M., MCLEOD, M. P., SALZBERG, S., PETERSON, J., KHALAK, H., RICHARDSON, D., HOWELL, J. K., CHIDAMBARAM, M., UTTERBACK, T., MCDONALD, L., ARTIACH, P., BOWMAN, C., COTTON, M. D., VENTER, J. C. (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281,** 375–388.

FREDERICO, L. A., KUNKEL, T. A. & SHAW, B. R. (1990). A sensitive genetic assay for the detection of cytosine deamination: determination of rate constants and the activation energy. *Biochemistry* **29,** 2532–2537.

FREEMAN, J. M., PLASTERER, T. N., SMITH, T. F. & MOHR, S. C. (1998). Patterns of genome organization in bacteria. *Science* **279,** 1827–1830.

GOFFEAU, A., AERT, R., AGOSTINI-CARBONE, M. L., AHMED, A., AIGLE, M., ALBERGHINA, L., ALBERMANN, K., ALBERS, M., ALDEA, M., ALEXANDRAKI, D. *et al.* (1997). The yeast genome directory. *Nature* **387** (Suppl.), 5–105.

GREENFEDER, S. A. & NEWLON, C. S. (1992). A replication map of a 61 kb circular derivative of *Saccharomyces cerevisiae* chromosome III. *Mol. Biol. Cell* **3,** 999–1013.

GRIGORIEV, A. (1998). Analyzing genomes with cumulative skew diagrams. *Nucleic Acids Res.* **26,** 2286–2290.

GRIGORIEV, A. (1999). Strand-specific compositional asymmetries in double-stranded DNA viruses. *Virus Res.* **60,** 1–19.

HANAWALT, P. C. (1991). Heterogenity of DNA repair at the gene level. *Mutat. Res.* **247,** 203–211.

IWAKI, T., KAWAMURA, A., ISHINO, Y., KOHNO, K., KANO, Y., GOSHIMA, N., YARA, M., FURUSAWA, M., DOI, H. & IMAMOTO, F. (1996). Preferential replication-dependent mutagenesis in the lagging DNA strand in *Escherichia coli*. *Mol. Gen. Genet*. **251,** 657–664.

KARLIN, S. & BURGE, C. (1995). Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet*. **11,** 283–290.

KARLIN, S., CAMPBELL, A. M. & MRAZEK, J. (1998). Comparative DNA analysis across diverse genomes. *Annu. Rev. Genet*. **32,** 185–225.

KEOGH, R. S., SEOIGHE, C. & WOLFE, K. H. (1998). Evolution of gene order and chromosome number in *Saccharomyces*, *Kluyveromyces* and related fungi. *Yeast* **14,** 443–457.

KREUTZER, D. A. & ESSIGMANN, J. M. (1998). Oxidized, deminated cytosines are a source of C → T transitions *in vivo*. *Proc. Nat. Acad. Sci. U.S.A.* **95,** 3578–3582.

KUNKEL, T. A. (1992). Biological asymmetries and the fidelity of eukaryotic DNA replication. *Bioessays* **14,** 303–308.

KUNST, F., OGASAWARA, N., MOSZER, I., ALBERTINI, A. M., ALLONI, G., AZEVEDO, V., BERTERO, M. G., BESSIERES, P., BOLOTIN, A., BORCHERT, S. *et al.* (1997). The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390,** 249–256.

LAFAY, B., LLOYD, A. T., MCLEAN, M. J., DEVINE, K. M., SHARP, P. M. & WOLFE, K. H. (1999). Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases. *Nucleic Acids Res.* **27,** 1642–1649.

LALO, D., STETTLER, S., MARIOTTE, S., SLONIMSKI, P. P. & THURIAUX, P. (1993). Two yeast chromosomes are related by a fossil duplication of their centromeric regions. *C. R. Acad. Sci. III* **316,** 367–373.

LIN, H. J. & CHARGAFF, E. (1967). On the denaturation of deoxyribonucleic acid. II. Effects of concentration. *Biochim. Biophys. Acta* **145,** 398–409.

LINDAHL, T. (1993). Instability an decay of the primary structure of DNA. *Nature* **362,** 709–715.

LOBRY, J. R. (1995). Properties of a general model of DNA evolution under no-strand-bias conditions. *J. Mol. Evol.* **40,** 326–330, **41,** 680.

LOBRY, J. R. (1996a). Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13,** 660–665.

LOBRY, J. R. (1996b). Origin of replication of *Mycoplasma genitalium*. *Science* **272,** 745–746.

LOPEZ, P., PHILIPPE, H., MYLLYKALLIO, H. & FORTERRE, P. (1999). Identification of putative chromosomal origins of replication in Archea. *Mol. Microbiol.* **32,** 883–886.

LOUIS, E. J. & HABER, J. E. (1990). The subtelomeric Y′ repeat family in *Saccharomyces cerevisiae*: an experimental system for repeated sequence evolution. *Genetics* **124,** 533–545.

MACKIEWICZ, P., GIERLIK, A., KOWALCZUK, M., DUDEK, M. R. & CEBRAT, S. (1999a). How does replication-associated mutational pressure influence amino acid composition of proteins? *Genome Res.* **9,** 409–416.

MACKIEWICZ, P., GIERLIK, A., KOWALCZUK, M., DUDEK, M. R. & CEBRAT, S. (1999b). Asymmetry of nucleotide composition of prokaryotic chromosomes. *J. Appl. Genet*. **40,** 1–14.

MACKIEWICZ, P., KOWALCZUK, M., GIERLIK, A., DUDEK, M. R. & CEBRAT, S. (1999c). Origin and properties of noncoding ORFs in the yeast genome. *Nucleic Acids Res.* **27**, 3503–3509.

MARIANS, K. J. (1992). Prokaryotic DNA replication. *Annu. Rev. Biochem.* **61,** 673–719.

MCINERNEY, J. O. (1998). Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. *Proc. Nat. Acad. Sci. U.S.A.* **95,** 10 698–10 703.

MCLEAN, M. J., WOLFE, K. H. & DEVINE, K. M. (1998). Base composition skews, replication orientation, and gene orientation in 12 prokaryote genomes. *J. Mol. Evol.* **47,** 691–696.

MRÁZEK, J. & KARLIN, S. (1998). Strand compositional asymmetry in bacterial and large viral genomes. *Proc. Nat. Acad. Sci. U.S.A.* **95,** 3720–3725.

OLSEN, G. J. & WOESE, C. R. (1997). Archeal genomics: an overview. *Cell* **89,** 991–994.

RADMAN, M. (1998). DNA replication: one strand may be more equal. *Proc. Nat. Acad. Sci. U.S.A.* **95,** 9718–9719.

ROCHA, E. P., DANCHIN, A. & VIARI, A. (1999). Universal replication biases in bacteria. *Mol. Microbiol.* **32,** 11–16.

ROSCHE, W. A., TRINH, T. Q. & SINDEN, R. R. (1995). Differential DNA secondary structure-mediated deletion mutation in the leading and lagging strands. *J. Bacteriol.* **177,** 4385–4391.

SALZBERG, S. L., SALZBERG, A. J., KERLAVAGE, A. R. & TOMB, J. F. (1998). Skewed oligomers and origins of replication. *Gene* **217,** 57–67.

SEOIGHE, C. & WOLFE, K. H. (1998). Extent of genomic rearrangement after genome duplication in yeast. *Proc. Nat. Acad. Sci. U.S.A.* **95,** 4447–4452.

SHAMPAY, J., SZOSTAK, J. W. & BLACKBURN, E. H. (1984). DNA sequences of telomeres maintained in yeast. *Nature* **310,** 154–157.

SKRABANEK, L. & WOLFE, K. H. (1998). Eukaryote genome duplication—where's the evidence? *Curr. Opin. Genet. Dev.* **8,** 694–700.

SUEOKA, N. (1995). Intrastrand parity rules of DNA bases composition and usage biases of synonymous codons. *J. Mol. Evol.* **40,** 318–325, **42,** 323.

VEAUTE, X. & FUCHS, R. P. P. (1993). Greater susceptibility to mutations in lagging strand of DNA replication in *Escherichia coli* than in leading strand. *Science* **261,** 598–600.

WAGA, S. & STILLMAN, B. (1994). Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*. *Nature* **369,** 207–212.

WANG, J. (1998). The base contents of A, C, G, or U for three codon positions and the total coding sequences show positive correlation. *J. Biomol. Struct. Dyn.* **16,** 51–57.

WANG, S. S. & ZAKIAN, V. A. (1990). Sequencing of *Saccharomyces* telomeres cloned using T4 DNA polymerase reveals two domains. *Mol. Cell. Biol.* **10,** 4415–4419.

WATSON, J. D. & CRICK, F. C. H. (1953). A structure for deoxyribose nucleic acid. *Nature* **327,** 169–170.

WILKINS, B. M. (1988). Organization and plasticity of enterobacterial genomes. *J. Appl. Bact*: *Symp. Suppl.* 51–69.

WOLFE, K. H. & SHIELDS, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387,** 708–713.

ZHANG, C. T. & ZHANG, R. (1991). Analysis of distribution of base in codon in the coding sequences by a diagrammatic technique. *Nucleic Acids Res.* **19,** 6313–6317.