

NON-RANDOM DNA TRIMER ARRANGEMENT IN *Bacillus cereus* CHROMOSOME ATCC 10987

Phan Thi Huyen, Nguyen Duc Luong*

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

ABSTRACT: Different from the linear chromosomes of eukaryotes, most chromosomes of bacteria are circular. In this paper, we investigated the arrangement of trimers in *Bacillus cereus* chromosome by recording physical positions of individual trimers along the single-stranded sequence of this chromosome in the 5' to 3' direction. Our very first finding was that numbers of individual trimers in this chromosome were very much not the same. In an attempt to find out how the trimers were arranged along the chromosome, we found that every trimer was distributed unevenly and throughout the chromosome. Each trimer was also found to distribute differently on each side of the origin of replication. In addition, the frequency of a trimer and that of its reverse complement, such as ATG and CAT, in the local regions of chromosome were always biased. Interestingly, however, the distribution of trimers in the sense and antisense sequences of chromosome exhibited an order in their arrangement, letting us to conclude that the trimer arrangement in the sense and antisense sequences shaped the codon usage of this bacterium at the chromosomal level.

Keywords: *Bacillus cereus*, chromosome, codon usage, DNA arrangement, sense and antisense sequences, trimer, trimer reverse complement.

INTRODUCTION

Levene (1919) [8] discovered that DNA was a molecule of nucleotides, which compose of nitrogen bases, deoxyribose sugars, linked together through the phosphate groups. In 1928, DNA of the bacterium *Pneumococcus* was discovered to carry genetic information using the transformation methodology. In 1951, Ewin Chargaff found that in a double-stranded DNA segment of bacterium *Bacillus subtilis*, numbers of adenines and thymines were similar as were those of cytosines and guanines [2]. This was known as Chargaff's first parity rule and was later verified by the base-pairing DNA model developed by Watson & Crick (1953) [19]. According to Rudner et al. (1968), Forsdyke & Mortimer (2000), if the same property is found in single-stranded DNA sequences, it is referred to as Chargaff's second parity rule [4, 13]. At present, about 1783 prokaryotic genomes, of which 1664 are bacterial and 119 are archaeal, have been completely sequenced. Of the 1215 eukaryotic genomes in sequencing projects, 37 have already been sequenced completely. Freeman et al. (1998), Mradzek & Karlin (1998), Gierlik et al. (2000) found nucleotide skews, which are

deviations from the Chargaff's second parity rule, in the local regions of these genomes [5, 6, 9]. Green et al. (2003), Touchon et al. (2005), Polak & Arndt (2008) further suggested that the local DNA skews were associated with replication and transcription [7, 11, 17]. In fact, Worning et al. (2006) applied these local skews to predict the position of replication origin in the prokaryotic chromosomes [20]. At the whole organization, Niki et al. (1999) [10] reported that the *Escherichia coli* chromosome was structured into four macrodomains (MDs) and two less-structured regions. One of the MDs, the Ter MD, was constrained for reducing DNA mobility and delaying loci segregation when it was associated with the division machinery in the cell cycle, as found by Thiel et al. (2012) [16]. Niki et al. (1999) [10] and Esnault et al. (2007) [3] revealed that rearrangements of DNA within a replicore of *E. coli* perturbed the nucleoid distribution and the early steps of cytokinesis of the cell. In *Bacillus subtilis*, changing the position of origin of replication by Wang et al. (2007) [18] or making the DNA inverted within a replicore by Srivatsan et al. (2010) [15] also reduced the replication rate. The arrangement of DNA in

chromosome thus appears to play important roles in these biological processes. However, how the nucleotides are arranged in the chromosomal DNA sequence has not been known.

In this study, we examined the distribution of DNA trimers in *Bacillus cereus* genomic sequence and found that the trimers were not randomly distributed in this bacterial chromosome. In fact, the individual trimers were found to be present unevenly and throughout the chromosome. Also, trimer arrangement on one side of the replication origin or terminus was found to be different from that on the other. There were always biases in the distribution of individual trimers and that of their respective reverse complements in the local regions of chromosome. However, when considering the sense and antisense sequences, we found that the trimer

arrangement in these sequences as a whole shaped the codon usage.

MATERIALS AND METHODS

The genomic sequence of bacterium *Bacillus cereus* was taken from NCBI, via ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Bacillus_cereus_ATCC_10987_uid57673/NC_003909.fna.

The sense and antisense sequences were extracted from this genomic sequence using data obtained from NCBI, via ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Bacillus_cereus_ATCC_10987_uid57673/NC_003909.ptt.

Trimers (table 1) were counted along the single-stranded DNA sequence of *Bacillus cereus* chromosome in 5' to 3' orientation, in the one base-pair shifting manner (figure 1a) and in the three base-pair shifting manner (figure 1b).

Table 1. List of alphabetically sorted trimers and their respective reverse complements

1	2	3	4	5	6	7	8
T	T's RC						
AAA	TTT	CAA	TTG	GAA	TTC	TAA	TTA
AAC	GTT	CAC	GTG	GAC	GTC	TAC	GTA
AAG	CTT	CAG	CTG	GAG	CTC	TAG	CTA
AAT	ATT	CAT	ATG	GAT	ATC	TAT	ATA
ACA	TGT	CCA	TGG	GCA	TGC	TCA	TGA
ACC	GGT	CCC	GGG	GCC	GGC	TCC	GGA
ACG	CGT	CCG	CGG	GCG	CGC	TCG	CGA
ACT	AGT	CCT	AGG	GCT	AGC	TCT	AGA
AGA	TCT	CGA	TCG	GGA	TCC	TGA	TCA
AGC	GCT	CGC	TCG	GGC	GCC	TGC	GCA
AGG	CCT	CGG	CCG	GGG	CCC	TGG	CCA
AGT	ACT	CGT	ACG	GGT	ACC	TGT	ACA
ATA	TAT	CTA	TAG	GTA	TAC	TTA	TAA
ATC	GAT	CTC	GAG	GTC	GAC	TTC	GAA
ATG	CAT	CTG	CAG	GTG	CAC	TTG	CAA
ATT	AAT	CTT	AAG	GTT	AAC	TTT	AAA

T and RC stand for trimer and its corresponding reverse complement, respectively. Each trimer in the column 2, 4, 6 or 8 is the reverse complement of the trimer on the same line in the column 1, 3, 5 or 7, respectively.



Figure 1. Illustration for counting trimers along the DNA sequence. Trimers such as TAA, AAT, ATG, etc. were counted along the sequence in the one base-pair shifting manner (a). Trimers, such as TAA, TGA, TCC, etc. were counted along the sequence in the three base-pair shifting manner (b). In both cases, the trimers were counted in the 5' to 3' orientation.

To view the distribution of each trimer in the chromosomal sequence, positions of the trimer along the sequence were first recorded. Over the length of chromosome, cumulative counts of every 1000 individual trimers were plotted against the chromosomal positions of those 1000th trimers. Similarly, over the length of local sequences, which were about 1/10th or 1/100th of the chromosomal sequence,

cumulative counts of every 100 or 10 individual trimers, respectively, were plotted against the chromosomal positions of those 100th or 10th trimers.

Positions of origin and terminus of replication in the chromosome were obtained from the previous study by Worning et al. (2006) [20].

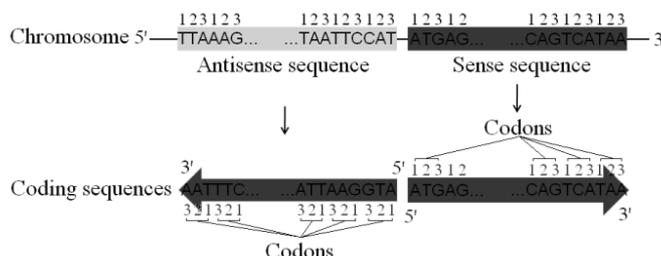


Figure 2. Illustration for counting trimers according to three trimer positions in the sense and antisense sequences. Numbers 1, 2 or 3 in red and black are first, second or third codon positions and trimer positions, respectively. In the sequence highlighted with the green color, the trimers TTA, TAA, AAA are at the first, second and third trimer positions, respectively. Counts of trimers at three trimer positions in the sense and antisense sequences of chromosome are shown in figure 7.

To count trimers at three trimer positions (figure 2), the sense and antisense sequences were first extracted from the genomic sequence and were then connected together to form a long sequence. Individual trimers that located at the first, second or third trimer positions along this long sequence were then counted. The trimers'

respective reverse complements at each of these positions were also counted.

RESULTS AND DISCUSSION

Non-random distribution of trimers in *B. cereus* chromosome

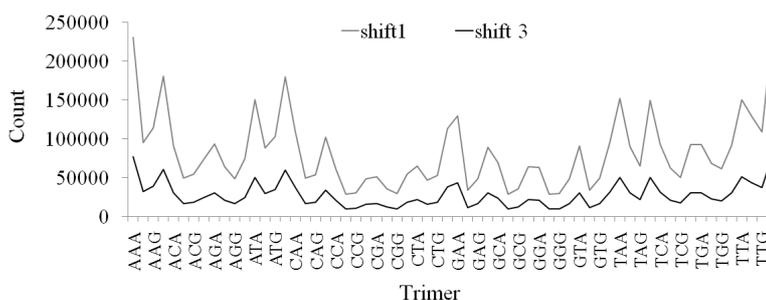


Figure 3. Counts of trimers in *B. cereus* ATCC 10987 chromosome. Data were plotted for all 64 trimers, however, only 32 of the 64 alphabetically listed trimers are shown due to lack of space. shift 1 and shift 3 denote the one base-pair and three base-pair shifting manners, respectively.

We counted the occurrences of 64 trimers over the length of single-stranded chromosomal sequence whose length was 5224283

nucleotides, in the one base-pair shifting manner. We found that the counts of different trimers were not the same. In fact, they

exhibited obvious biases. We wondered if these trimers were randomly distributed in the chromosome, we counted them in the three base-pair shifting manner. We recognized that rates of difference between the trimer counts were not changed in comparison with those obtained in the one base-pair shifting manner. The patterns of trimer counts in the two counting ways can be seen in figure 3. The rates of difference between trimer counts obtained by these two counting ways reveal that the trimers were not distributed randomly, but in a yet unidentified order in the chromosome. The trimer frequencies in chromosomes were also determined previously by Baisnée et al. (2002) [1]. However, these authors just found the approximate equal frequencies of trimers in two complementary strands of chromosome. In the followings, we present more about this nonrandom arrangement of nucleotides in *B. cereus* chromosome.

Biased and uneven distribution of trimers over the bacterial chromosome

To discover how the trimers are distributed

in *B. cereus* chromosome, we recorded the positions of individual trimers along the single-stranded chromosomal sequence. We then plotted cumulative counts of individual trimers over the sequence length. The plots in figure 4 show that each trimer was distributed throughout the chromosome. The frequency of a trimer and that of its reverse complement were observed to be biased in the local regions of chromosome. This means that the Chargaff's second parity rule applied to trimers was violated. In addition, degrees of this bias among pairs of trimers and their respective reverse complements in the local regions of chromosome were greatly different. Beside the biased distribution of each trimer compared with its reverse complement along the sequence, we observed that the trimers exhibited biases in their distribution on different sides of the replication origin or terminus. For instance, the trimer AAG was distributed more frequently from the origin to the terminus of replication and less frequently from the terminus to the origin, while its reverse complement, CTT, was oppositely distributed.

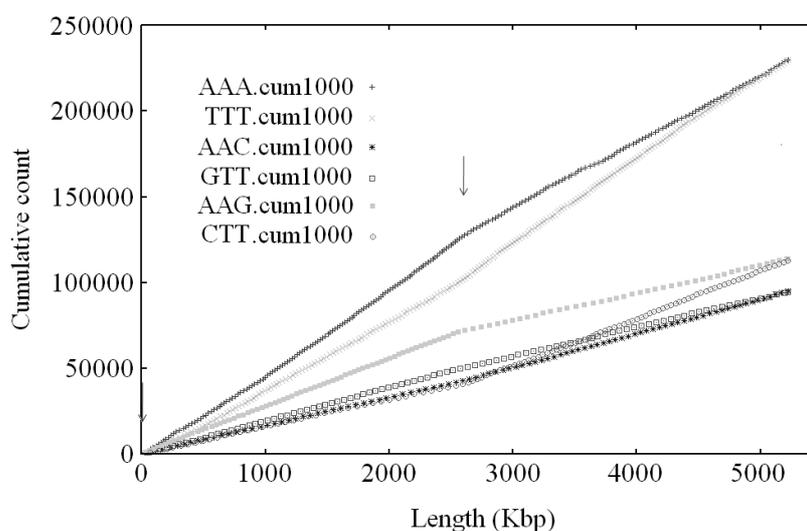


Figure 4. Distribution of trimers along the *B. cereus* ATCC 10987 chromosome. Only counts of 6 out of 64 trimers are shown. For each trimer, its positions were recorded along the chromosomal sequence. The individual trimer's cumulative count was plotted against its positions along the chromosome. Red and blue arrows indicate the positions of origin and terminus of replication, respectively; cum1000 in legend indicates that the counts were cumulated for every 1000 individual trimer standing before it.

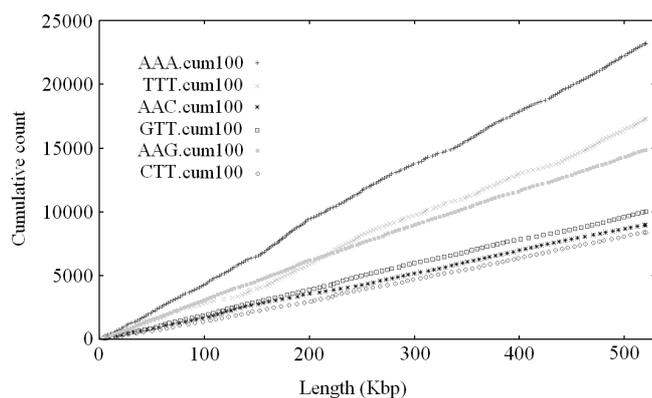


Figure 5. Distribution of trimers along 1/10th *B. cereus* ATCC 10987 chromosomal sequence. Data are shown as in figure 4, except the trimer counts were plotted after every 100 times of the trimers' appearance against their positions along the sequence.

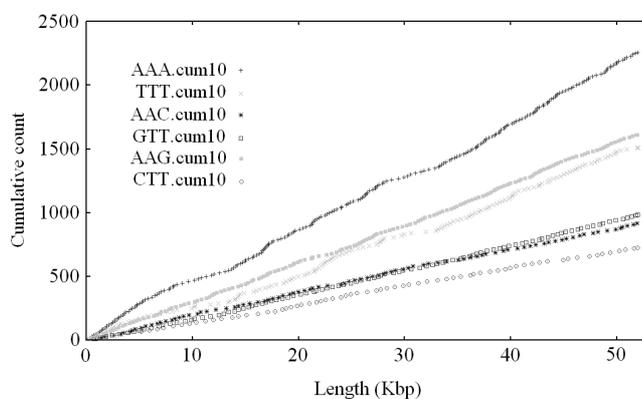


Figure 6. Distribution of trimers along 1/100th *B. cereus* ATCC 10987 chromosomal sequence. Data are shown as in figure 4, except the trimer counts were plotted after every 10 times of the trimers' appearance against their positions along the sequence.

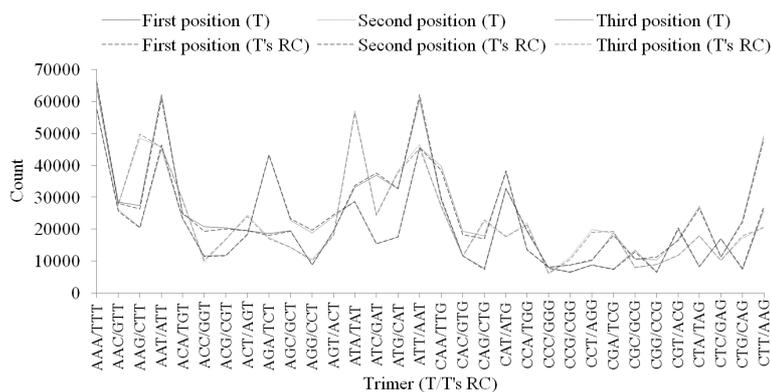


Figure 7. Counts of trimers at three trimer positions in the connected sense and antisense sequence of chromosome (see Materials and Methods). Data were plotted for each pair of trimer (T) and its reverse complements (T's RC).

On each side of the replication origin or terminus, we further clarified the distribution of each trimer and that of its reverse complement. Over the length of the 1/10th chromosomal sequence with the genomic position from 1000 to 521000, cumulative count plots of every 100 individual trimers show that along the sequence the trimers were distributed unevenly (figure 5). The unevenness in the trimer distribution can be seen more clearly when the cumulative counts of every 10 individual trimers were plotted over the length of the 1/100th chromosomal sequence, with the genomic position from 1000 to 53000 (figure 6).

Biased distribution of trimers along the chromosome has not been reported previously, however, that of few octamers at both sides of the replication origin in the bacterial chromosomes was studied by Salzberg et al. (1998) [14]. Rocha (2004) [12] reasoned that this biased distribution was the result of asymmetric distribution of genes and corresponding regulatory signals in leading and lagging strands.

Trimer arrangement in the sense and antisense sequences shaped codon usage

So far, it has been unclear in which order the trimers are arranged along the sequence of the bacterial chromosome. We investigated the trimer frequencies in the sense and antisense sequences. Sense sequences are those that look like their messenger RNA (mRNA) sequences, whereas antisense sequences are those that serve as templates for production of mRNAs. Also, complementary to the template sequences are coding sequences. Each coding sequence contains codons, which are nucleotide triplets that constitute the genetic codes determining the insertion of specific amino acids into a polypeptide chain during protein synthesis or the signals stopping protein synthesis. For the investigation, we extracted the sense and antisense sequences from the chromosomal sequence of *B. cereus* ATCC 10987. We then determined the frequencies of trimers and those of their respective reverse complements at three trimer positions in all these sequences in one base-pair shifting manner (figures 1 and 2).

We figured out that, though biased and uneven, trimers in sense and antisense sequences of *B. cereus* ATCC 10987 chromosome were distributed in an order, in that frequencies of trimers at the first positions were almost equal to those of their respective reverse complements at the first positions. On the other hand, frequencies of trimers at the second positions were almost equal to those of their respective reverse complements at the third positions. In contrast, frequencies of trimers at the third positions were almost equal to those of their respective reverse complements at the second positions (figure 7).

By convention, codon usage of an organism is defined as frequencies of individual codons in every 1000 codons in all coding sequences in the genome. In figure 2, we see that first position count of a trimer, ATG for instance, in a sense sequence is actually frequency of codon, which is also ATG, in its coding sequence. First position count of that trimer's reverse complement, CAT, in an antisense sequence is the frequency of codon, which is the trimer's reverse complement, *i.e.* ATG in this case, in its coding sequence. This means that frequency of a codon in the coding sequences encoded by the sense sequences was almost equal to that codon's frequency in the coding sequences encoded by the antisense sequences. In other words, codon usage of this bacterium was shaped by the trimer arrangement in the sense and antisense sequences at the chromosomal level. Arrangement of DNA outside the sense and antisense sequences still remains unclear so needs to be further investigated.

Conclusions: We have investigated the arrangement of trimers in the chromosome of bacterium *B. cereus* ATCC 10987. The trimers though exhibited biases between their frequencies and those of their respective reverse complements along the chromosome, they also distributed differently and unevenly on each side of the replication origin. Regardless of these biases in their distribution, the frequencies of trimers and those of their respective reverse complements in the sense and antisense sequences were found to shape the codon usage in this bacterium. It still remains to see how this

DNA arrangement affects the essentially biological processes in the cell.

REFERENCES

1. Baisnée P. F., Hampson S. and Baldi P., 2002. Why are complementary DNA strands symmetric?. *Bioinformatics*, 18: 1021-1033.
2. Chargaff E., 1951. Structure and function of nucleic acids as cell constituents. *Fed. Proc.*, 10: 654-659.
3. Esnault E., Valens M., Espe'li O. and Boccard F., 2007. Chromosome structuring limits genome plasticity in *Escherichia coli*. *PLoS. Genet.*, 3: e226.
4. Forsdyke D. R. and Mortimer J. R., 2000. Chargaff's legacy. *Gene*, 261: 127-137.
5. Freeman J. M., Plasterer T. N., Smith T. F. and Mohr S. C., 1998. Patterns of genome organization in bacteria. *Science*, 279: 1827. doi: 10.1126/science.279.5358.1827a.
6. Gierlik A., Kowalczyk M., Mackiewicz P., Dudek M. R. and Cebrat S., 2000. Is there replication associated mutational pressure in the *Saccharomyces cerevisiae* genome?. *J. Theor. Biol.*, 202: 305-314.
7. Green P., Ewing B., Miller W., Thomas P. J. and Green E. D., 2003. Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet.*, 33: 514-517.
8. Levene P., 1919. The structure of yeast nucleic acid. *J. Biol. Chem.*, 40: 415-24.
9. Mradzek J. and Karlin S., 1998. Strand compositional asymmetry in bacterial and large viral genomes. *Proc. Natl. Acad. Sci. USA*, 95: 3720-3725.
10. Niki H., Yamaichi Y. and Hiraga S. 1999., Dynamic organization of chromosomal DNA in *Escherichia coli*. *Genes Develop.*, 14: 212-223.
11. Polak P. and Arndt P. F., 2008. Transcription induces strand-specific mutations at the 5' end of human genes. *Genome Res.*, 18: 1216-1223.
12. Rocha E. P. C., 2004. The replication-related organization of bacterial genomes. *Microbiol.*, 150: 1609-1627.
13. Rudner R., Karkas J. D. and Chargaff E., 1968. Separation of *B. subtilis* DNA into complementary strands. III. Direct analysis. *Proc. Natl. Acad. Sci. USA*, 60: 921-922.
14. Salzberg S. L., Salzberg A. J., Kerlavage A. R. and Tomb J. F., 1998. Skewed oligomers and origins of replication. *Gene*, 217: 57-67.
15. Srivatsan A., Tehranchi A., MacAlpine D. M. and Wang J. D., 2010. Co-orientation of replication and transcription preserves genome integrity. *PloS. Genet.*, 6: e1000810.
16. Thiel A., Valens M., Vallet-Gely I., Espéli O. and Boccard F., 2012. Long-range chromosome organization in *E. coli*: a site-specific system isolates the Ter macrodomain. *PLoS. Genet.*, 8: e1002672.
17. Touchon M., Nicolay S., Audit B., Brodie E. B., d'Aubenton-Carafa Y., Arneodo A. and Thermes C., 2005. Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. *Proc. Natl. Acad. Sci. USA*, 102: 9836-9841.
18. Wang J. D., Berkmen M. B. and Grossman A. D., 2007. Genome-wide coorientation of replication and transcription reduces adverse effects on replication in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA*, 104: 5608-5613.
19. Watson J. D. and Crick F. H. C., 1953. A structure for deoxyribose nucleic acid. *Nature*, 171: 737-738.
20. Worning P., Jensen L. J., Hallin P. F., Staerfeldt H. H. and Ussery D. W., 2006. Origin of replication in circular prokaryotic chromosomes. *Environ. Microbiol.*, 8: 353-361.

**SỰ SẮP XẾP KHÔNG NGẪU NHIÊN CỦA CÁC DNA TRIMER
TRONG NHIỄM SẮC THỂ *Bacillus cereus* ATCC 10987**

Phan Thị Huyền, Nguyễn Đức Lượng

Đại học Bách khoa, ĐHQG tp Hồ Chí Minh

TÓM TẮT

Khác với nhiễm sắc thể dạng thẳng của sinh vật bậc cao, đa số nhiễm sắc thể của vi khuẩn có dạng vòng. Chúng tôi đã tìm ra sự sắp xếp của các DNA trimer trong nhiễm sắc thể của tế bào vi khuẩn *Bacillus cereus* ATCC 10987, bằng cách xác định vị trí của mỗi trimer dọc theo chiều dài nhiễm sắc thể theo chiều từ 5' đến 3'. Ghi nhận ban đầu của chúng tôi là số lượng các trimer trong nhiễm sắc thể này khác nhau rất nhiều. Các trimer được phân bố trong toàn bộ nhiễm sắc thể và không đồng đều trong các đoạn ngắn DNA. Sự phân bố của mỗi trimer ở hai phía của điểm khởi đầu sao chép cũng khác nhau rõ rệt. Đồng thời, tần suất xuất hiện của mỗi trimer và trimer bổ sung đảo ngược của nó, chẳng hạn ATG và CAT, trong bất kỳ một đoạn DNA nào trong nhiễm sắc thể cũng khác nhau. Điều thú vị là sự sử dụng codon, tức số lượng trung bình của mỗi codon riêng biệt trên 1000 codon tổng thể của bộ gene trong tế bào, được định dạng bởi sự sắp xếp của các trimer trong tất cả các trình tự sense và antisense của nhiễm sắc thể.

Từ khóa: *Bacillus cereus*, nhiễm sắc thể, sự sử dụng codon, sự sắp xếp DNA, trình tự sense và antisense, trimer.

Ngày nhận bài: 3-1-2012