Computational Analysis of χ Sequences

Reina Uno [12] reina@sfc.keio.ac.jp
Yoichi Nakayama [13] ynakayam@sfc.keio.ac.jp
Kenji Yotsutani [13] t95995ky@sfc.keio.ac.jp
Takehito Mogami[13] t97991tm@sfc.keio.ac.jp
Kazuharu Arakawa [13] t98901ka@sfc.keio.ac.jp
Masaru Tomita [13] mt@sfc.keio.ac.jp

[1] Laboratory for Bioinfomatics
[2] Graduate school of Media and Governance
[3] Department of Environmental Information
Keio University, 5322 Endo, Fujisawa, Kanagawa 252–8502, Japan

1 Introduction

A χ sequence (5'-GCTGGTGG-3') is a hot spot which promotes recombination in *Escherichia coli*. RecBCD recognizes a χ sequence, and it extricates a ssDNA with its exonuclease activity. There are 1009 χ sequences whose frequency overwhelms the expected appearance. Moreover, it is also known that these χ sequences are frequently located within ORF.

2 Result of Classification

We have classified all *E. coli* genes according to their functions, and then analyzed frequencies of the χ sequence in each class of genes, normalizing by the average length of genes.

Table 1.

Clasffications	Number of Genes	Number of Chi	O/E	chi / 10000bp
Biosynthesis	257	70	1.27	2.66
Cell-Division	28	10	1.37	2.87
Cell-Killing	3	1	2.83	5.93
Chaperone	12	3	1.01	2.12
Degradation-Protein	23	14	1.87	3.92
Detoxification	11	2	0.84	1.76
Drug-resistance	33	14	1.78	3.72
IS	65	11	0.67	4.10
Metabolism	632	131	1.02	2.14
Mobility	34	17	2.67	5.99
Regulator	223	35	0.83	1.73
Replication&Repair&Recombination	103	38	1.27	2.66
Shock	14	2	0.79	1.65
Structure	112	24	1.34	2.81
Transcription	125	3	0.24	0.49
Translation	41	3	0.31	0.65
Transport	241	86	1.73	3.63
Not-Classified	1089	309	1.25	2.63
Unknown	1365	190	0.97	2.04
Total	4331	973		

The result shows that χ sequences exist preferably in Mobility and Transport genes (Table 1). The χ sequences often code for hydrophobic amino acids [1], and that is probably the reason for them to exist at high frequency in these proteins.

Interestingly, no χ sequence was found in the genes for tRNAs, rRNAs, and ribosomal subunit proteins. The reason for these genes to avoid χ sequences is not clear and needs further investigation.

3 Relationship between the χ sequences and the replication

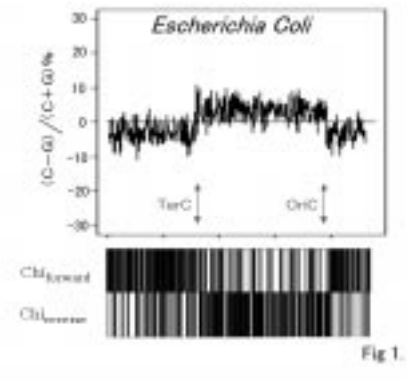

Fig 1.

Around 75% of χ sequences face the direction of replication [2]. It is widely believed that χ sequences relate to the replication in order to repair DNA damages that arrest replication folks [3]. GC skew is an inclination index of G over G+C on a DNA strand and it is known in the *E. coli* genome that the shift points of the GC skew are located in the replication origin and the termination region.

The direction of χ sequences and GC skew of the *E. coli* genome are shown in Fig. 1. The both shift points appear at the similar locations which are close to the replication origin and termination regions. Since the χ sequence contains five G's and only one C, the direction of the χ sequences depends on the inclination of G on the genome.

It has been reported that χ-like sequences in *Haemophilus influenzae* Rd have a role similar to *E. coli* χ sequence [4], whereas *H. influenzae* does not show any clear shift points of the direction of χ-like sequences or the GC skew (Fig. 2).

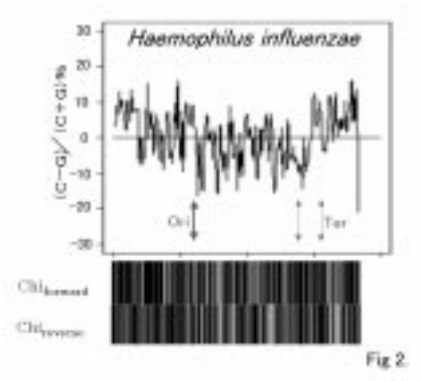

Fig 2.

It is not clearly explained why the GC skew exists in *E. coli* but not in *H. influenzae*, which is a close relative to *E. coli*. It has been believed that the directional bias of the χ sequences in *E. coli* is related to the replication mechanism. Our results, however, strongly indicate that the bias is merely due to the inclination of G, and that it is probably not due to the replication mechanism.

We are currently analyzing the directional bias of other G-rich oligonucleotide sequences as a control in order to verify the hypothesis. The result of this analysis will be presented and discussed at the meeting.

References

[1] Handa, N., Ohashi, S., and Kobayashi, I., Clustering of chi sequence in Escherichia coli genome, *Microb. Comp. Genomic*, 2(4):287–298, 1997.

[2] Blattner, F.R. et al., The complete genome sequence of *Escherichia coli* K-12, *Science*, 277:1456–1462, 1997.

[3] Colbert, T., Taylor, A.F., and Smith, G.R., Genomics, Chi sites and codons: 'islands of preferred DNA pairing' are oceans of ORFs, *Trends in Genetics*, 14(12):485-488, 1998.

[4] Sourice, S., Biaudet, V., El Karoui, M., Ehrlich, S.D., and Gruss, A., Identification of the Chi site of *Haemophilus influenzae* as several sequences related to the *Escherichia coli* Chi site, *Mol. Microbiol.*, 27:1021–1029, 1998.