FEBS Letters 450 (1999) 72–76 FEBS 21945

A low rate of nucleotide changes in *Escherichia coli* K-12 estimated from a comparison of the genome sequences between two different substrains

Takeshi Itoh[a,b], Toshitsugu Okayama[a,c], Hiroyuki Hashimoto[a,c], Jun-ichi Takeda[b], Ronald W. Davis[d], Hirotada Mori[b], Takashi Gojobori[c,*]

[a] *Center for Information Biology, National Institute of Genetics, Yata 1111, Mishima 411-8540, Japan*
[b] *Research and Education Center for Genetic Information, Nara Institute of Science and Technology, Ikoma 630-0101, Japan*
[c] *Hitachi Software Engineering, Yokohama 231-0015, Japan*
[d] *DNA Sequencing and Technology Center, Stanford University, Stanford, CA 94305, USA*

Received 8 February 1999; received in revised form 23 March 1999

Abstract Two genome sequences of *Escherichia coli* K-12 substrains, one partial W3110 and one complete MG1655, have been determined by Japanese and American genome projects, respectively. In order to estimate the rate of nucleotide changes, we directly compared 2 Mb of the nucleotide sequences from these closely-related *E. coli* substrains. Given that the two substrains separated about 40 years ago, the rate of nucleotide changes was estimated to be less than 10^{-7} per site per year. This rate was supported by a further comparison between partial genome sequences of *E. coli* and *Shigella flexneri*.

Key words: *Escherichia coli*; Genome project; Complete genome; Rate of nucleotide substitution; Rate of nucleotide change; Insertion sequence

1. Introduction

The molecular mechanisms of mutation in bacteria have been extensively studied. However, there is a paucity of information on the evolutionary rates of bacteria when compared with that of other organisms such as mammals and viruses. Recently, two genomes of *Escherichia coli* K-12 have been independently sequenced by Japanese and American teams [1–7]. It was unfortunate that both genome project teams used the same strain K-12, however, it is fortunate that the two teams employed different substrains. In fact, the Japanese team determined two thirds of the entire genome sequence of W3110 [1–6], while the American team successfully sequenced the complete genome of MG1655 [7]. These two closely-related genomes gave us a unique opportunity to investigate the evolutionary rates of *E. coli* by directly comparing the genome sequences.

In this study, we attempted to estimate the rate of nucleotide changes by a direct comparison of the DNA sequences of two mutant derivatives of *E. coli* K-12, W3110 and MG1655. According to laboratory history records, these substrains originated from W1485 approximately 40 years ago [8]. Of course, W1485 is also the derivative of the K-12 wild-type (Fig. 1). Only a few differences were found between the two genomes in our comparative analysis. When using 40 years as the divergence time between W3110 and MG1655, we estimated the rate of nucleotide changes to be about 10^{-7} per site per year. We should note that since MG1655 has been exposed to a mutagenic agent (acridine orange) for derivation of this isolate [8], the estimated rate should be taken as the maximal rate under laboratory conditions. The rate of nucleotide changes estimated herein is relatively low compared to those of viruses [9,10]. We also compared partial genome sequences between *E. coli* and *Shigella flexneri*. The divergence time between these two enteric bacteria has been supposed to be 25 million years ago [11]. We then estimated the rate of nucleotide changes to be about 10^{-9} per site per year, which is consistent with the rate of nucleotide changes that we have estimated between the two *E. coli* substrains.

2. Materials and methods

2.1. DNA sequences of E. coli

The entire genome sequence (4.6 Mb) of MG1655 determined by Blattner et al. [7] was obtained from the ftp site at ftp://ncbi.nlm.nih.gov/genbank/genomes. The version of the sequence was M52. Another 526 kb DNA segment of MG1655 was constructed from the sequences which were determined by the group at Stanford University (DDBJ/EMBL/GenBank accession numbers U70214, U73857, U82598 and U82664). The DNA sequences of W3110, determined by the Japanese team [3–6], are available at http://www.ddbj.nig.ac.jp/e-coli/ecoli_list.html or http://www.cib.nig.ac.jp/dda/taitoh/ecomp.html. We did not use the portion spanning 0–12 min of W3110, because these regions were sequenced in the early stages of the Japanese project [1,2] and were expected to contain many errors, due to limitations of the methods used about 10 years ago. The non-redundant DNA sequences of the recent Japanese project were approximately 2090 kb in total. This is approximately 45% of the whole genome sequence. We found that the clone #320 sequence was incorrectly inverted, hence, we corrected the direction.

2.2. DNA sequences of S. flexneri

We used the 1470 bp portion of the 3′ end of U00119 and the 1600 bp portion of the 3′ end of Z11766, which have orthologous regions in the *E. coli* genome. These two DNA segments contained the *rpoS* and *glpX* genes, respectively. In the total of 3070 bp, 18.3% were non-encoding regions.

2.3. DNA alignments

All of the DNA sequences were aligned by using the dynamic programming method [12,13]. These alignments are available at our WWW site (http://www.cib.nig.ac.jp/dda/taitoh/ecomp.html).

2.4. Terminology of evolutionary rates

The term 'mutation' is restricted to changes before natural selection operates. Thus, 'mutation rate' means the rate of all types of changes on a given nucleotide sequence before natural selection operates. On the other hand, the term 'substitution' is for changes after natural selection and its rate is 'the rate of nucleotide substitutions'. The term 'nucleotide substitution' does not include insertions or deletions (indels). When we consider both substitutions and indels after natural

*Corresponding author. Fax: (81) (559) 81 6848.
E-mail: tgojobor@genes.nig.ac.jp

Abbreviations: IS, insertion sequence; indel, insertion or deletion

PII: S0014-5793(99)00481-0

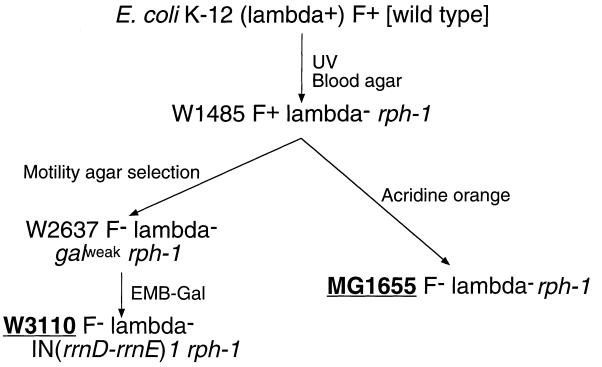

Fig. 1. Pedigree chart of substrains of *E. coli* K-12 [8].

selection operates, we call them 'nucleotide change' and its rate 'the rate of nucleotide changes.'

In this study, the number of nucleotide substitutions per nucleotide site is termed K_1 and the number of nucleotide changes per site is termed K_2. Similarly, the rate of nucleotide substitutions per site per year is k_1 and the rate of nucleotide changes per site per year is k_2. We obtain k_1 and k_2 by dividing K_1 and K_2 by a divergence time, respectively.

3. Results

3.1. Estimation of k_1 and k_2

Since a few differences were expected between the closely-related genomes, we first had to estimate the error frequency of genome-sequencing. Before Blattner et al. completed the genome-sequencing of MG1655, other DNA segments of MG1655, 526 kb in total, had been deposited in the International Databases DDBJ/EMBL/GenBank by R.W. Davis et al. at Stanford University. Therefore, we estimated the error frequency by comparing the two sequences from the identical substrain MG1655, which had been independently determined. As a result, 14 substitutions and 13 indels were found between the two sequences of Blattner et al. and Davis et al. In this study, we regarded indels of two or more consecutive nucleotides as one indel. Let N_e be the number of substitution errors and let N_t be the total number of nucleotide sites to be compared. The error frequency per nucleotide site (e) in one genome can be estimated by $e = N_e/2N_t$. The error frequency for K_1 is then estimated to be 1.33×10^{-5} per site. Likewise, when indels are taken into account, the error frequency for K_2 becomes 2.57×10^{-5} per site. Two large deletions of 102 nucleotides in the sequence from Stanford University lay in a repetitive sequence region known as BIME. These deletions could be caused by assembly errors. In the following analysis, we assume that the error frequencies have been equal among the three genome-sequencing projects.

In the 2090 kb alignment of W3110 and MG1655, 76 substitutions and 125 indels were found. Two or more consecutive indels were regarded as one indel. Let N_o be the number of observed substitutions or changes. Given that a sequencing error and a nucleotide substitution or change could not have occurred at the same site, K_1 and K_2 can be obtained from the following formula: $K = N_o/2N_t - e$, where K is either K_1 or K_2. Thus, K_1 is 4.87×10^{-6} and K_2 is 2.24×10^{-5}. Similarly, K_2 in the encoding regions (1851 kb in total) was estimated to be 1.64×10^{-5}. On the other hand, as expected, the non-encoding regions (236 kb in total) showed a significantly higher K_2 of 6.97×10^{-5}. Since k_1 and k_2 are calculated by $k = K/T$, if the divergence time (T) is 40 years, k_1 is 1.22×10^{-7} per site per year and k_2 is 5.61×10^{-7} per site per year. However, first, it is possible that the two substrains of *E. coli* have not undergone replication events during storage, so that the exact divergence time was difficult to obtain. Therefore, we examined another possible case for a comparison (Table 1). As a result, we found that, even if their divergence time is only 10 years, k_1 is around the order of 10^{-7} and k_2 is 10^{-6}, although these rates should be taken again as maximal values because of the effects of a mutagenic agent. Second, it is also possible that the error frequency (e) estimated in this study may not be valid. Nevertheless, even if $e = 0$, k_1 and k_2 are still around the same orders as calculated above. On the contrary, if e is larger than we estimated, it dose not influence the following discussion, because we calculated k_1 and k_2 as maximal values.

3.2. Classification of indels

Classification of the indels is summarized in Table 2. Of all the indels, except insertion sequences (ISs) and one prophage, 98.3% were indels of eight or fewer nucleotides and 93.0% were indels of a single nucleotide. Only two indels were several 100 nucleotides in length. Interestingly, an indel of three nucleotides (or multiples of three nucleotides), which does not cause a frameshift error, was not observed in either genome. It appears that a large indel or an indel of three nucleotides (or multiples of) rarely occurs in the genome, while it is also possible that these indels are deleterious, so that individuals carrying such indels are excluded immediately from a population in the course of evolution. Although most indels were rather short, the number of indels was revealed to be more than that of nucleotide substitutions. This observation suggests that, even though an indel can readily disrupt a gene when it occurs in an encoding region, functional loss of such a gene is almost selectively neutral under laboratory conditions.

Moreover, we found 10 long indels of about 1000 or more nucleotides in length, which are mainly caused by ISs (Fig. 2). Eight ISs were found only in the W3110 genome, while one IS and one prophage were found only in MG1655. This prophage at 55 min appeared to be an excision of a cryptic prophage from W3110 [7]. There were alternative insertions of IS*1* and IS*5* between *flhD* and *yecG* at 43 min in MG1655 and W3110, respectively. This IS*1* was the only example of an IS inserted into MG1655 or excized from W3110.

Since the sequence examined herein is approximately 45% of the entire genome and nine large indels by ISs were found in that region, the total number of large indels by ISs between the two substrains is estimated to be 20. Given that the di-

Table 1
Rates of nucleotide substitutions (k_1) and changes (k_2) in *E. coli*

Divergence time (year)	Rate of nucleotide substitutions (k_1)	Rate of nucleotide changes (k_2)
10	4.87×10^{-7}	2.24×10^{-6}
40	1.22×10^{-7}	5.61×10^{-7}

Table 2
Classification of indels between W3110 and MG1655

Type of indels	Number of occurrences
1 Nucleotide indel	107
2 Nucleotide indel	4
5 Nucleotide indel	1
8 Nucleotide indel	1
181 Nucleotide indel	1
374 Nucleotide indel	1
Prophage	1
IS	9
Total	125

vergence time is 40 years, this indicates that an *E. coli* genome has experienced IS-derived indel events at the rate of 0.25 per year. It is of particular interest to know that *E. coli* has accepted such drastic alterations frequently, on average once in 4 years. Thus, changes of a genome structure may have depended heavily upon frequent indels by ISs, as discussed later.

4. Discussion

4.1. Comparison of the rates of nucleotide substitutions or changes

Drake [14] had experimentally estimated several spontaneous mutation rates including indels by the method of mutant accumulation. Since Drake's rates were values per nucleotide site per replication, we re-calculated the rates of spontaneous mutations per site per year in *E. coli*. Given that *E. coli* replication occurs every 20 minutes, the spontaneous mutation rates in *E. coli* estimated by Drake become $(1.08–1.81)\times10^{-5}$ per site per year (Table 3). This observation by Drake is consistent with our result that k_2 is approximately 10^{-7} or less for the following reasons. First, the spontaneous mutations dealt with by Drake must have contained deleterious mutations, which would be immediately excluded from the population by purifying selection. In particular, the natural environment should be much more severe than the laboratory conditions. Second, the replication time under the natural conditions can be much longer than 20 min, which leads to a much smaller value of k_2. These indicate that k_2 is at most 10^{-5} per site per year, that is, k_2 should be less than 10^{-5}. Hence it is quite possible that k_2 is 10^{-7} per site per year or less, as we estimated.

For four RNA viruses, the rate of non-synonymous substitution per site per year was estimated to be approximately 10^{-3} [9] and for 17 RNA viruses except HTLV-1, k_1 was in a range of 10^{-1}–10^{-5} [10]. k_1 of *E. coli* seems to be low relative to those of RNA viruses (Table 4). For DNA viruses, k_1 of herpes simplex virus was reported to be 3.5×10^{-8} [18] (Table 4). The k_1 of mammals is known to be about 10^{-9} (for review, see [10]) (Table 4). The rate of spontaneous mutation in *Drosophila melanogaster* was proposed to be 2.27×10^{-8} per site per year [15,16], and its k_1 is thought to be higher than those of mammals [17]. It is possible that k_1 and k_2 of *E. coli* are as low as those in mammals and slowly-evolving DNA viruses.

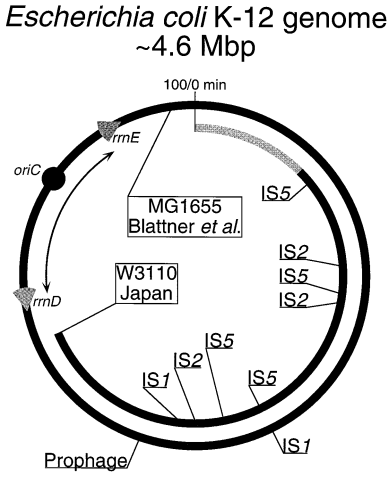

Fig. 2. Maps of unique ISs and a prophage in W3110 (inner circle) and MG1655 (outer circle), respectively. The shaded region of W3110 was not used in this study. There is an inversion between *rrnD* and *rrnE* in W3110 [26] as indicated by an arrow, but this region has not been sequenced by the Japanese team.

LeClerc et al. [19] reported that the mutation rates were high among *E. coli* and *Salmonella* pathogens. This phenomenon was explained by hitch-hiking of mutators in association with adaptive mutations [20]. Although k_2 has been suggested to be low for non-pathogenic strains under laboratory conditions, it is plausible that pathogenic strains evolve at very high mutation rates for host defense evasion and adaptive evolution. We compared the *E. coli* sequence with two large DNA fragments from a closely-related pathogenic bacteria, *S. flexneri*. Let d be the number of nucleotide differences per site, including indels, between two sequences. The percentage of differences ($d\times100$) was 3.36% between the two species. From comparisons of ecologically significant events and rRNA sequences, Ochman and Wilson [11] estimated the date of divergence between *E. coli* and *Shigella* to be 25 million years ago. k_2 is represented by the following formula: $k_2 = d/2T$. When we follow the divergence time by Ochman and Wilson, k_2 becomes 0.67×10^{-9} per site per year. Thus, it is clear that even pathogenic bacteria have not evolved at a higher rate. Therefore, this value is consistent with the rate between the two substrains of *E. coli*, less than 10^{-7} per site per year. In future studies, it will be interesting to compare the rate described here with that of a pathogen like *E. coli* O157, to elucidate the cause of the frequent emergence of pathogenic bacteria that cause food-related illnesses.

In *E. coli*, genetic heterogeneity has been discovered in various stocks of the same W3110 substrain [21]. It is worth noting that in some cases, even those small changes seem to strongly influence phenotypes, although they are indistin-

Table 3
Spontaneous mutation rates of *E. coli*

Target	Rate per site per replication (adopted from [14])	Rate per site per year (calculated in this study)
lacI[a]	4.1×10^{-10}	1.08×10^{-5}
lacI[a]	6.9×10^{-10}	1.81×10^{-5}
hisGDCBHAFE	5.1×10^{-10}	1.34×10^{-5}

[a]The mutation rate of *lacI* was measured by two different methods in reference [14]

Table 4
Rates of nucleotide substitutions per site per year (k_1)

	Rate of nucleotide substitutions per site per year
(a) *E. coli* (this study)	$< 1.22 \times 10^{-7}$
(b) RNA viruses [10]	1.0×10^{-1}–6.8×10^{-7}
(c) DNA virus (herpes simplex virus) [18]	3.5×10^{-8}
(d) Mammals [10]	$(0.5–5.0) \times 10^{-9}$

guishable under laboratory conditions. Moreover, since IS can completely disrupt a gene, it may cause a drastic change on a phenotype by only one transposition into an important gene. Accordingly, the finding of genetic heterogeneity among W3110 suggests that such a low rate of nucleotide changes, as estimated in this study, can drastically change the phenotype.

4.2. Insertion sequences and genome rearrangements

Even between closely-related organisms, the genome has been shuffled frequently [22–24]. Genome rearrangements are suggested to be mediated by homologous recombination of repetitive DNA sequences [25]. In fact, it is known that W3110 has a large inversion between *rrnD* and *rrnE* (Fig. 2), probably created by recombination between highly homologous sequences [26]. In addition, as previously reported [24,27], it is likely that ISs play an important role in the genome evolution, because they provide homologous regions large enough to mediate recombination events. In fact, recent sequencing of a 93 kb plasmid pO157, which was extracted from *E. coli* O157:H7, revealed that pO157 carried 18 IS elements [28]. It also revealed that pO157 has been extensively subjected to rearrangements of the plasmid genome. Furthermore, in the *E. coli* K-12 chromosomes, we found that ISs have been actively transposed and have been interspersed in the genome during short-term evolution. Therefore, the major changes of the genome structures may be frequently caused by IS-mediated recombination events, although the rate of nucleotide changes in *E. coli* was estimated to be quite low.

4.3. Conclusion

We have compared more than 2 Mb of the genome sequence from two substrains of *E. coli* K-12. The rate of nucleotide changes was estimated to be less than 10^{-7} per site per year. Our results also suggest that there are mainly two different forces for bacterial evolution during several decades, that is, point mutations at a low rate and large indels caused by ISs. Although the rate of nucleotide changes was estimated to be very low, ISs cause drastic changes, such as disruption of an open reading frame, and they may be involved in rapid rearrangements of a genome structure.

5. Addendum

In the revising process of this paper, we noticed that R.A. Alm et al. [29] reported a whole genome comparison of two *Helicobacter pylori* strains. Most end points of rearranged genome segments in *H. pylori* were associated with repetitive sequences such as ISs. This observation supports our hypothesis that ISs play an important role in genome rearrangements. Alm's report also suggests that genome-sequencing of closely-related species is crucial for elucidating genome diversity and evolution.

Acknowledgements: We appreciate our colleagues for helpful discussion. We would like to thank S. Gaudieri and R. Chapman for proof reading of the manuscript. This work was supported in part by Grants-in-Aid from the Ministry of Education, Science, Sports and Culture of Japan.

References

[1] Yura, T., Mori, H., Nagai, H., Nagata, T., Ishihama, A., Fujita, N., Isono, K., Mizobuchi, K. and Nakata, A. (1992) Nucleic Acids Res. 20, 3305–3308.
[2] Fujita, N., Mori, H., Yura, T. and Ishihama, A. (1994) Nucleic Acids Res. 22, 1637–1639.
[3] Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K., Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., Yano, M. and Horiuchi, T. (1996) DNA Res. 3, 137–155.
[4] Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kasai, H., Kashimoto, K., Kimura, S., Kitakawa, M., Kitagawa, M., Makino, M., Miki, T., Mizobuchi, K., Mori, H., Mori, T., Motomura, K., Nakade, S., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Sivasundaram, S., Tagami, H., Takeda, J., Takemoto, K., Takeuchi, Y., Wada, C., Yamamoto, Y. and Horiuchi, T. (1996) DNA Res. 3, 363–377.
[5] Itoh, T., Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Kasai, H., Kimura, S., Kitakawa, M., Kitagawa, M., Makino, K., Miki, T., Mizobuchi, K., Mori, H., Mori, T., Motomura, K., Nakade, S., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampei, G., Seki, Y., Sivasundaram, S., Tagami, H., Takeda, J., Takemoto, K., Wada, C., Yamamoto, Y. and Horiuchi, T. (1996) DNA Res. 3, 379–392.
[6] Yamamoto, Y., Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kimura, S., Kitagawa, M., Makino, K., Miki, T., Mitsuhashi, N., Mizobuchi, K., Mori, H., Nakade, S., Nakamura, Y., Nashimoto, H., Oshima, T., Oyama, S., Saito, N., Sampei, G., Satoh, Y., Sivasundaram, S., Tagami, H., Takahashi, H., Takeda, J., Takemoto, K., Uehara, K., Wada, C., Yamagata, S. and Horiuchi, T. (1997) DNA Res. 4, 91–113.
[7] Blattner, F.R., Plunkett, G.III, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J., Mau, B. and Shao, Y. (1997) Science 277, 1453–1474.
[8] Bachmann, B.J. (1996) in: *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology (Neidhardt, F.C., Curtiss, R.III, Ingraham, J.L., Lin, E.C.C., Low, K.B., Magasanik, B., Reznikoff, W.S., Riley, M., Schaechter, M. and Umbarger, H.E., Eds.), pp. 2460–2488, ASM Press, Washington, DC, USA.
[9] Gojobori, T., Yamaguchi, Y., Ikeo, K. and Mizokami, M. (1994) Jpn. J. Genet. 69, 481–488.
[10] Suzuki, Y. and Gojobori, T. (1998) Virus Genes 16, 69–84.
[11] Ochman, H. and Wilson, A.C. (1987) J. Mol. Evol. 26, 74–86.
[12] Needleman, S.B. and Wunsch, C.D. (1970) J. Mol. Biol. 48, 443–453.
[13] Smith, T.F., Waterman, M.S. and Fitch, W.M. (1981) J. Mol. Evol. 18, 38–46.
[14] Drake, J.W. (1991) Proc. Natl. Acad. Sci. USA 88, 7160–7164.
[15] Mukai, T. and Cockerham, C.C. (1977) Proc. Natl. Acad. Sci. USA 74, 2514–2517.
[16] Mukai, T., Harasa, K., Kusakabe, S. and Yamazaki, T. (1990) Proc. Jpn. Acad. 66, 29–32.
[17] Moriyama, E.N. (1987) Jpn. J. Genet. 62, 139–147.
[18] Sakaoka, H., Kurita, K., Iida, Y., Takada, S., Umene, K., Kim, Y.T., Ren, C.S. and Nahmias, A.J. (1994) J. Gen. Virol. 75, 513–527.
[19] LeClerc, J.E., Li, B., Payne, W.L. and Cebula, T.A. (1996) Science 274, 1208–1211.
[20] Sniegowski, P.D., Gerrish, P.J. and Lenski, R.E. (1997) Nature 387, 703–705.
[21] Jishage, M. and Ishihama, A. (1997) J. Bacteriol. 179, 959–963.
[22] Tatusov, R.L., Mushegian, A.R., Bork, P., Brown, N.P., Hayse,

W.S., Borodovsky, M., Rudd, K.E. and Koonin, E.V. (1996) Curr. Biol. 6, 279–291.
[23] Watanabe, H., Mori, H., Itoh, T. and Gojobori, T. (1997) J. Mol. Evol. 44, S57–S64.
[24] Itoh, T., Takemoto, K., Mori, H. and Gojobori, T. (1999) Mol. Biol. Evol. 16, 332–346.
[25] Himmelreich, R., Plagens, H., Hilbert, H., Reiner, B. and Herrmann, R. (1997) Nucleic Acids Res. 25, 701–712.
[26] Hill, C.W. and Harnish, B.W. (1981) Proc. Natl. Acad. Sci. USA 78, 7069–7072.
[27] Naas, T., Blot, M., Fitch, W.M. and Arber, W. (1995) Mol. Biol. Evol. 12, 198–207.
[28] Makino, K., Ishii, K., Yasunaga, T., Hattori, M., Yokoyama, K., Yutsudo, C.H., Kubota, Y., Yamaichi, Y., Iida, T., Yamamoto, K., Honda, T., Han, C.-G., Ohtsubo, E., Kasamatsu, M., Hayashi, T., Kuhara, S. and Shinagawa, H. (1998) DNA Res. 5, 1–9.
[29] Alm, R.A., Ling, L.-S.L., Moir, D.T., King, B.L., Brown, E.D., Doig, P.C., Smith, D.R., Noonan, B., Guild, B.C., deJonge, B.L., Carmel, G., Tummino, P.J., Caruso, A., Uria-Nickelson, M., Mills, D.M., Ives, C., Gibson, R., Merberg, D., Mills, S.D., Jiang, Q., Taylor, D.E., Vovis, G.F. and Trust, T.J. (1999) Nature 397, 176–180.