

# Comparative Analysis of Eubacterial DNA Polymerase III Alpha Subunits

Xiao-Qian Zhao<sup>1,2</sup>, Jian-Fei Hu<sup>3</sup>, and Jun Yu<sup>1,4\*</sup>

<sup>1</sup>Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 101300, China; <sup>2</sup>Graduate School of Chinese Academy of Sciences, Beijing 100049, China; <sup>3</sup>College of Life Sciences, Peking University, Beijing 100871, China; <sup>4</sup>James D. Watson Institute of Genome Sciences, Zhejiang University, Hangzhou 310008, China.

DNA polymerase III is one of the five eubacterial DNA polymerases that is responsible for the replication of DNA duplex. Among the ten subunits of the DNA polymerase III core enzyme, the alpha subunit catalyzes the reaction for polymerizing both DNA strands. In this study, we extracted genomic sequences of the alpha subunit from 159 sequenced eubacterial genomes, and carried out sequence-based phylogenetic and structural analyses. We found that all eubacterial genomes have one or more alpha subunits, which form either homodimers or heterodimers. Phylogenetic and domain structural analyses as well as copy number variations of the alpha subunit in each bacterium indicate the classification of alpha subunit into four basic groups: polC, dnaE1, dnaE2, and dnaE3. This classification is of essence in genome composition analysis. We also consolidated the naming convention to avoid further confusion in gene annotations.

**Key words:** DNA polymerase III alpha subunit, dnaE, polC

## Introduction

DNA polymerases play central roles in DNA replication and repair, whose fidelity is the primary source of DNA sequence variations. Among the five eubacterial DNA polymerases (I-V) (1-4), polymerase III is responsible for catalyzing DNA polymerization in replication, and the rest are involved in subsidiary roles in replication (I) and repair (I, II, IV, and V) (1-6). DNA polymerase III is a ten-component complex consisting of the replicase ( $\alpha$ ,  $\epsilon$ , and  $\theta$ ), the clamp loader or gamma complex ( $\gamma$ ,  $\delta$ ,  $\delta'$ ,  $\zeta$ ,  $\chi$ , and  $\psi$ ), and the sliding clamp ( $\beta_2$ ) (2, 7-17). The alpha subunit works as a dimer at the replication fork and plays a central role in the complex. However, its classification remains controversial and ambiguous due to occasional redundancy and seemingly unrelated groupings, which are mainly resulted from the horizontal gene transfer among prokaryotes. The alpha subunit has been roughly classified into polC and dnaE groups (18-22). A “confused dnaE” (23) and “a split of dnaE” (into dnaE1 and dnaE2) (24), as well as the integrated dnaE1 and dnaE2 in several cyanobacterial species were noticed previously (25).

When two intact dnaE genes were detected, they were usually named as dnaE1 and dnaE2 in an arbitrary way (25, 26). No detailed analysis has been carried out to classify these important molecules.

In this study, we collected DNA sequences from 159 representative eubacterial genomes and extracted their alpha subunit genes according to their annotations, followed by thorough searches with the BLAT tools under variable parameters. After that, we compiled amino acid sequences of these genes and studied the enzyme systematically at three different levels, including their overall phylogeny, domain structures, and identifiable motifs. As a result, we classified all the alpha subunit genes into four groups.

## Results

The distribution of the alpha subunit genes per genome among the 159 eubacterial genomes is quite variable (Table S1). According to our statistics, 77 (48.4%) genomes have one single alpha subunit gene, 75 (47.2%) genomes have two alpha subunit genes, and the remaining seven (4.4%) genomes (*Agrobacterium tumefaciens* C58 UWash, *Bacillus subtilis*, *Clostridium acetobutylicum*, *Sinorhizobium*

**\*Corresponding author.**

**E-mail:** junyu@genomics.org.cn

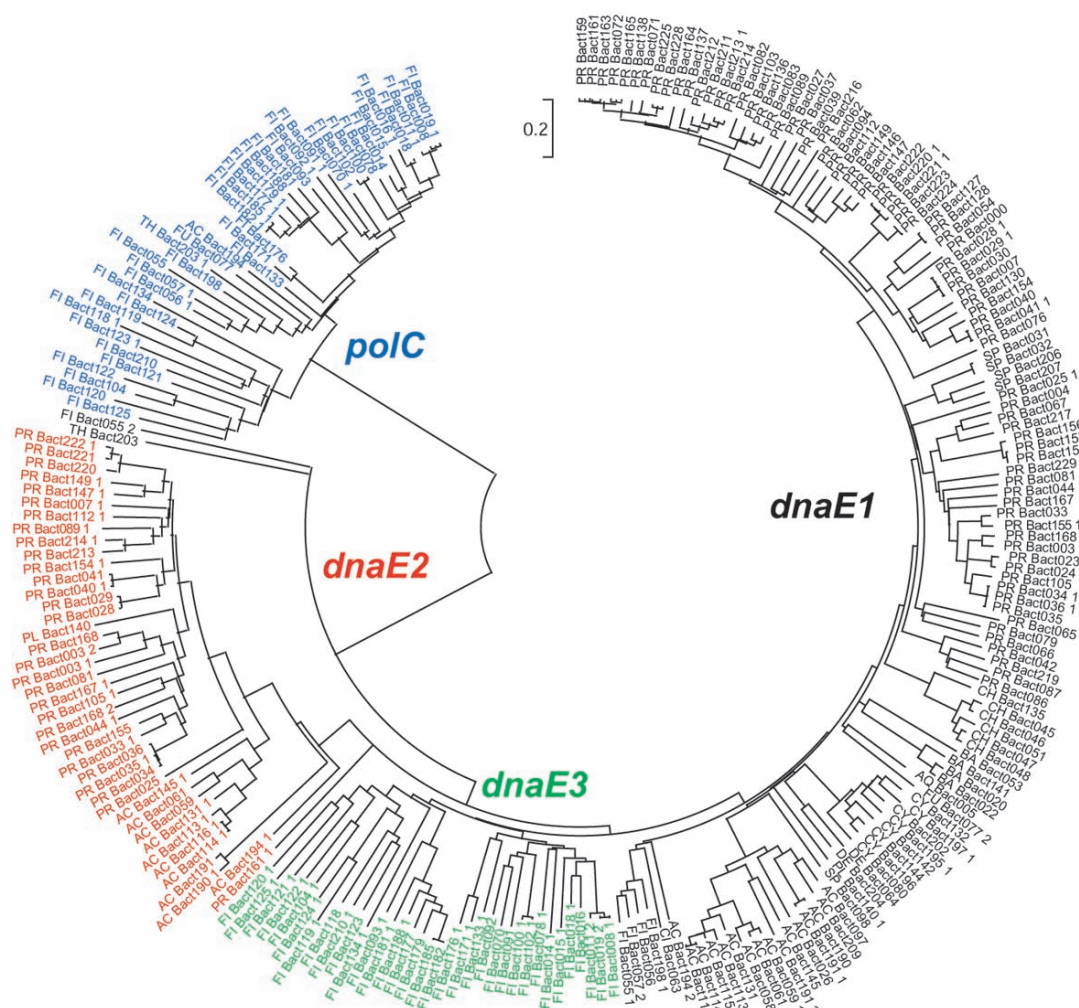
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

*meliloti*, *Staphylococcus epidermidis* RP62A, *Streptomyces coelicolor*, and *Symbiobacterium thermophilum* IAM14863) each contain three alpha subunit genes. Therefore, these genomes have 248 alpha subunit genes in total. We examined these genes carefully and did not find frame shifts, stop codons, or other anomalies. We believe that they are not pseudogenes although we lack empirical data to establish their exact functionality.

## Phylogenetic analysis

We tried several methods (neighbor-joining, minimum evolution, and maximum parsimony) to assess the

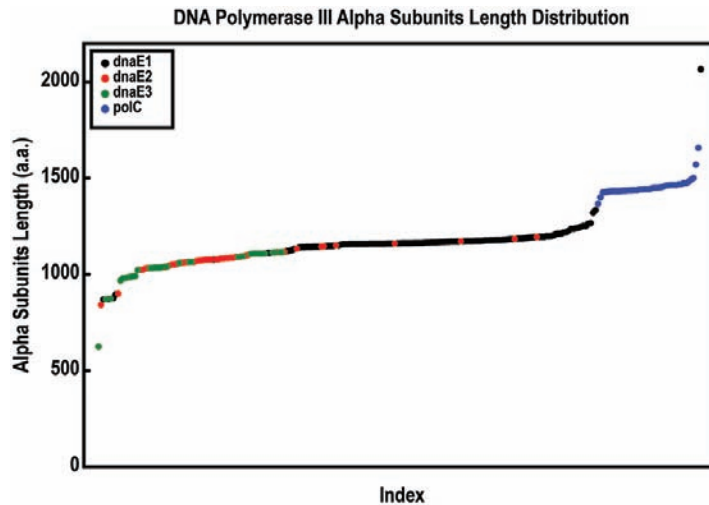
phylogeny of the alpha subunit genes and achieved similar results. Based on the phylogenetic analysis, we classified the 248 genes into four simple groups (Figure 1), namely *polC*, *dnaE1*, *dnaE2*, and *dnaE3*, regardless whether some of the bacteria have multiple alpha subunit genes or not. The *polC* group is rather distinct from the other three, in which 41 *polC* genes are clustered together, forming a single group. The lengths of these genes in this group range from 1,367 to 1,658 amino acids (a.a.), with an average of 1,455 a.a., which are longer than those of the other three groups in general. The *dnaE* genes are divided into three groups according to the cluster results. The lengths of these three groups are quite different:



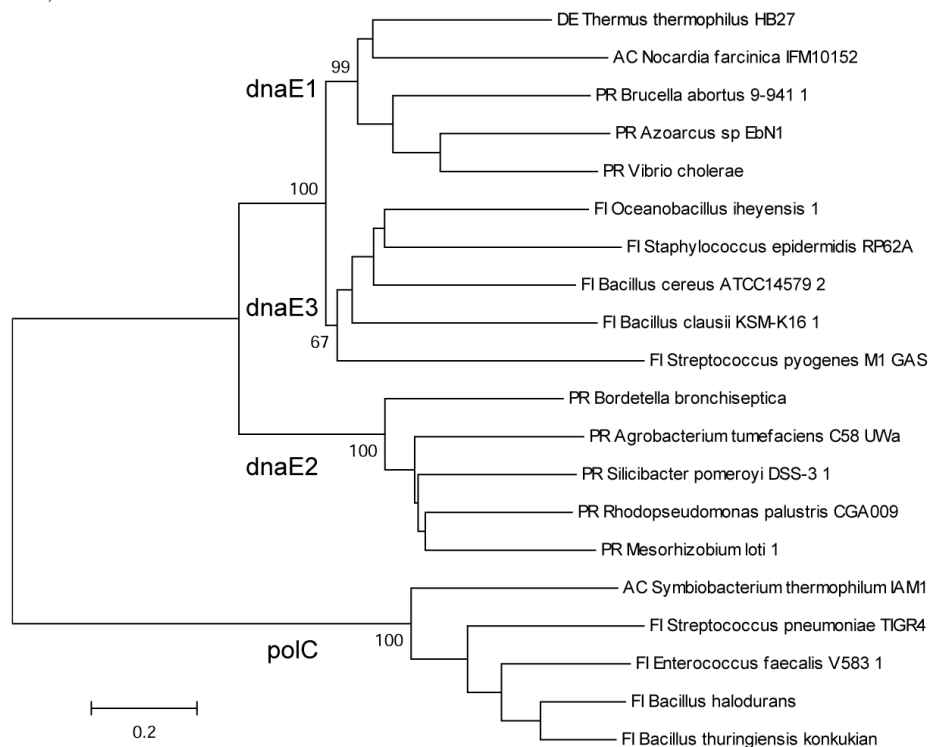
**Fig. 1** Phylogeny of the alpha subunit gene groups. Genomes are labeled in abbreviated characters for the phylum names (Table S1). The Arabian number after the bacterium number denotes whether it is classified based on the first alpha subunit gene or other co-existed isoforms. For instance, FI Bact181 1 stands for the Firmicute bacterium No. 181 in the collection, which is classified based on its first alpha subunit gene. AC, Actinobacteria; AQ, Aquificae; BA, Bacteroidetes/Chlorobi group; CH, Chlamydiae/Verrucomicrobia group; CI, Chloroflexi; CY, Cyanobacteria; DE, Deinococcus-Thermus; FI, Firmicutes; FU, Fusobacteria; PL, Planctomycetes; PR, Proteobacteria; TH, Thermotogae; SP, Spirochaetes.

dnaE1 is the longest, with an average length of 1,174 a.a., whereas dnaE2 and dnaE3 are shorter, with average lengths of 1,078 and 1,033 a.a., respectively (Figure 2). To demonstrate the unequivocal relationship,

we selected five alpha subunits randomly from each of the four groups, and grouped the subset of twenty alpha subunits into four clusters with significant bootstrap values (Figure 3).



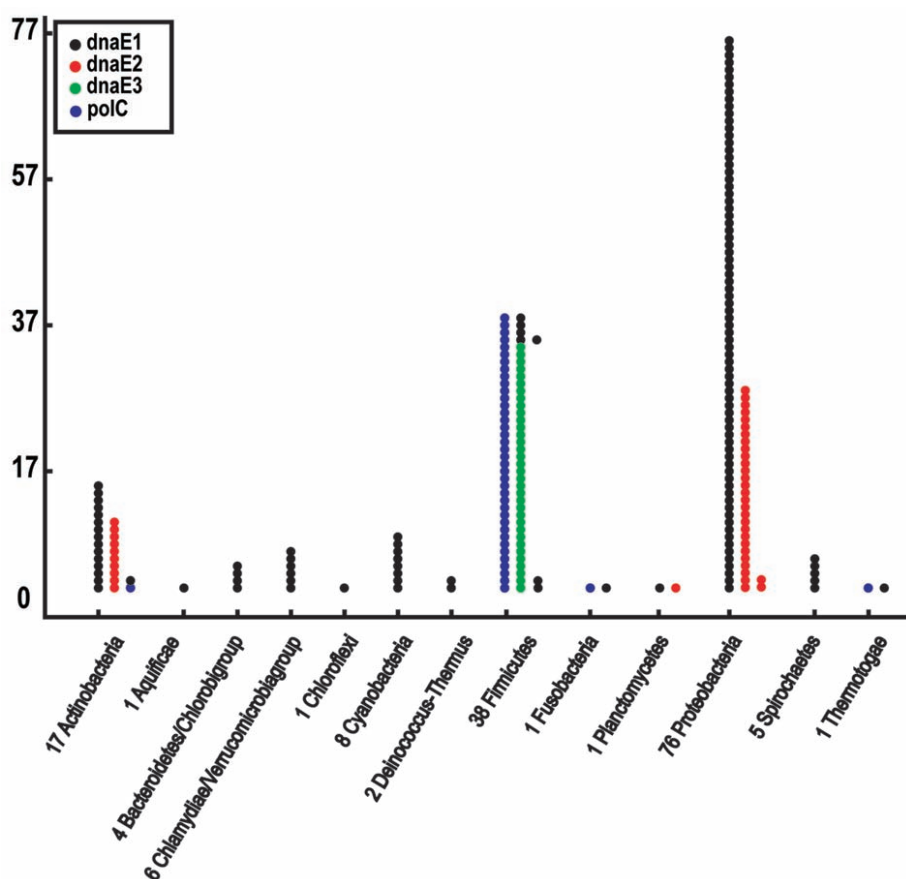
**Fig. 2** The length distribution of all alpha subunit genes in the collection. Among the 248 genes, there are 129 dnaE1 genes (average length of 1,174 a.a.), 44 dnaE2 genes (average length of 1,078 a.a.), 34 dnaE3 genes (average length of 1,033 a.a.), and 41 polC genes (average length of 1,455 a.a.). The shortest sequence is a dnaE3 gene found in *Onion yellows phytoplasma* (Bact134; 625 a.a.). The longest one is a single dnaE1 gene from *Thermus thermophilus* HB27 (Bact204; 2,067 a.a.).



**Fig. 3** Phylogeny of four alpha subunit groups from a partial and random selection. Five alpha subunits are randomly selected from each group. The same nomenclature is used as in Figure 1. The trees are constructed with the program MEGA (V3.1) by neighbor-joining with Kimura 2-parameter distance (scale bar). The reliability of the branching orders is estimated by bootstrapping (500 replicates), and bootstrap values (percentage after 500 iterations) for major branches are shown.

In addition to sequence-based phylogeny, there is ample taxonomic evidence that supports our four-group scheme. Firstly, *polC* and *dnaE3* frequently appear together (27) and work on the leading and the lagging strand, respectively, catalyzing the chemical reaction of DNA replication (20, 21). In contrast, *dnaE1* and *dnaE2* in general do not appear together in our collection so that *dnaE1* could work on both the leading and the lagging strand. Secondly, in the case of *dnaE2*, it pairs with *dnaE1* primarily; however, in very rare occasions it works together with *polC* (we found only one example of *S. thermophilum* IAM14863 in the 159 bacteria). It seems that *dnaE2* is not able to form a functional homodimer since it always pairs with either *dnaE1* or *polC*, even though some of the bacteria have more than one copy of the gene. Thirdly, the distribution of *dnaE1* homodimer and *polC*/*dnaE3* heterodimer is restricted to taxonomic schemes. For instance, *polC*/*dnaE3* combination is dominant in phylum Firmicutes, whereas *dnaE1* is mostly found in non-Firmicute bacteria. In

contrast, *dnaE2* is not related to bacterial taxonomy so that it has been found in different phyla, such as Actinobacteria and Proteobacteria (Figure 4). Fourthly, even for bacteria in the same genus, some have *dnaE2* while others do not. For instance, in Proteobacteria, a fraction possesses only *dnaE1* whereas others have both *dnaE1* and *dnaE2*; a similar classification result has also been observed in Actinobacteria (data not shown). The best example is the genus *Vibrio* that has four species sequenced; two of them, *Vibrio vulnificus* CMCP6 and *Vibrio parahaemolyticus*, have *dnaE1* and *dnaE2*, whereas the other two, *Vibrio fischeri* ES114 and *Vibrio cholerae*, do not have *dnaE2*. Finally, it is rather curious that how *dnaE1* functions in a genome that also has *dnaE2*, since *dnaE1* can form functional homodimers by itself. From the distribution pattern of *dnaE2*, we believe that the heterodimer of *dnaE1* and *dnaE2* overrides *dnaE1* homodimer when forming the replicase complex.



**Fig. 4** Alpha subunits in different bacterial phyla. One solid circle represents one alpha subunit. We mark the bacterial genome numbers together with phylum names; for example, “17 Actinobacteria” represents 17 bacterial genomes included in Actinobacteria.

## Domain analysis

In this part of analysis, we investigated what domain structures and their function classes separate the four alpha subunit groups. We used InterPro (28), a popular database for protein annotations, which integrates information for protein families and functional domains as well as incorporates efforts from different data-mining and data-curating projects, such as PROSITE, PRINTS, Pfam, and ProDom. We initially identified 49 domains in the alpha subunit genes among the eubacterial genomes. However, upon scrutiny on the sequences, we learnt that most of these domains are rare and only shared by less than five genomes. After removing the rare domains, we have 11 domains left. We then classified them into five basic classes based on their functions and positions in the four groups of alpha subunit genes (Table 1). The first class contains two exonuclease domains, IPR006054 (DNA polymerase III and epsilon subunit) and IPR006055 (exonuclease), which are located at the same position in all polC genes and appear to be the same sequences. In fact, the function of epsilon subunit is its exonuclease activity so that it is class-specific. The second class is the alpha subunit domains composed of IPR006308 (DNA polymerase III, alpha subunit, and Gram-positive type), IPR004805 (DNA polymerase III and alpha subunit), and IPR011708 (bacterial DNA polymerase III and alpha subunit); two of them are class-specific, IPR006308 to polC and IPR004805 to dnaE, and one, IPR011708, appears common to all alpha subunit genes. The third class is the phosphoesterase PHP domains, including IPR003141 (phosphoesterase PHP N-terminal) and IPR004013 (PHP

C-terminal). This class is universal to all alpha subunit genes but becomes ambiguous in a minor fraction of them. The fourth class is the nucleic acid-binding/OB-fold domains consisting of IPR004365 (nucleic acid binding, OB-fold, and tRNA/helicase-type) and IPR008994 (nucleic acid-binding and OB-fold). Some of these domains become more ambiguous among the genomes. The fifth class has two domains, IPR001093 (IMP dehydrogenase/GMP reductase) and IPR012337 (polynucleotidyl transferase and ribonuclease H fold); the former appears universal and the latter seems polC-specific, although the sequence-based detection of these domains is usually poor for this class of domains.

We also examined the relative positions of the identified domains in protein alignments (Figure 5; the last class was ignored in the display). The alpha subunit domains are universal but their relative positions are distinct in dnaE (in the middle) and polC (in the C-terminal portion). This most lengthy domain class, with nearly 500 a.a. in length, occupies the core of the alpha subunit gene and manifests DNA-dependent 5'-to-3' polymerase activity (29). The other two universal domain classes are the phosphoesterase PHP domains and the nucleic acid binding/OB-fold domains. They are shorter than the alpha subunit domains, having a length of nearly 100 a.a. The positions of these domains are conserved only within polC and dnaE. Comparatively, the exonuclease domain class is the most distinctive, for it is polC-specific and positioned in the middle of the gene with a length of nearly 200 a.a., which has 3'-to-5' proofreading exonuclease activity (21). Since genomes without polC always have an equivalent activity fulfilled by dnaQ, the exonuclease activity

**Table 1 Domain Analysis of the Four Groups of Alpha Subunit Genes**

Class	Domain	Domain annotation	polC (41)	dnaE1 (129)	dnaE2 (44)	dnaE3 (34)
1	IPR006054	DNA polymerase III, epsilon subunit	100%	0	0	0
	IPR006055	Exonuclease	100%	0	0	0
2	IPR006308	DNA polymerase III, alpha subunit, Gram-positive type	100%	0	0	0
	IPR004805	DNA polymerase III, alpha subunit	0	100%	100%	97%
	IPR011708	Bacterial DNA polymerase III, alpha subunit	100%	100%	100%	100%
3	IPR003141	Phosphoesterase PHP N-terminal	95%	95%	95%	85%
	IPR004013	PHP C-terminal	93%	99%	75%	76%
4	IPR004365	Nucleic acid binding, OB-fold, tRNA/helicase-type	88%	88%	95%	74%
	IPR008994	Nucleic acid-binding, OB-fold	60%	56%	55%	68%
5	IPR001093	IMP dehydrogenase/GMP reductase	33%	25%	23%	33%
	IPR012337	Polynucleotidyl transferase, ribonuclease H fold	55%	0	0	0

seems essential for the DNA polymerase III core enzyme (11, 21).

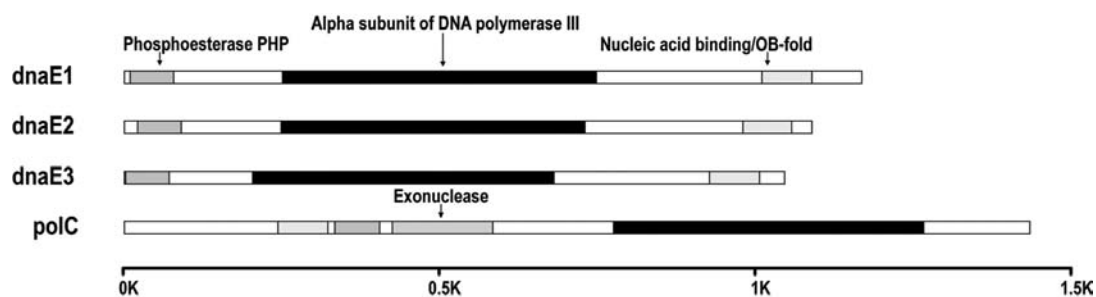
The lengths of dnaE1, dnaE2, and dnaE3 domains showed great differences in their sequences (Table 2). We analyzed the three dnaE groups to search for the common ancestor for each dnaE group, believing that dnaE1 is older than dnaE2 since it is much longer than dnaE2 and dnaE3. When deciphering three shared domains (IPR003141, IPR004365, and IPR011708) of each dnaE group, we found that their lengths are all longer in dnaE1 than in the other two dnaE groups, especially domain IPR011708, whose length is 839 a.a. in dnaE1, compared with 563 a.a. in dnaE2 and 528 a.a. in dnaE3. We suspect that the longer dnaE1 domains may form secondary structures for special functional requirement.

**Table 2 Average Lengths of three Domains of the dnaE Groups**

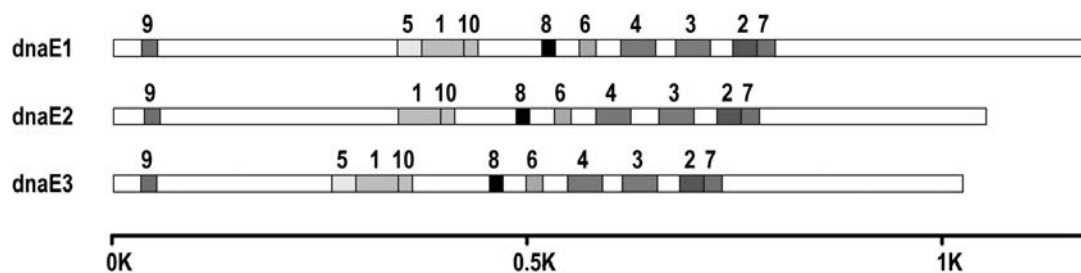
Group	Domain length (a.a.)		
	IPR003141	IPR004365	IPR011708
dnaE1	134	110	839
dnaE2	105	83	563
dnaE3	68	81	528

## Motif analysis

To find further differences among the dnaE groups, we analyzed the protein motifs, which are subsets of domain structures, by using the MEME tool for discovering motifs or highly conserved regions with its default settings (30, 31). We detected ten motifs among the dnaE genes, nine of which are shared by all members (Figure 6). Eight of the nine motifs were found in the alpha subunit domains while one (motif 9) was found in the phosphoesterase PHP domains. Motif 5 is the only one standing out that is different from the other nine motifs in its distribution among the genomes; it was identified in dnaE1 and dnaE3 genes but rarely in dnaE2 genes. Detailed analysis revealed that motif 5 is part of the alpha subunit domain and has a unique amino acid sequence, "LDVINQMGFPGYFLIVMEFIQWSKDNGIP", which is almost undetectable in dnaE2 genes. No recognizable motif was found in the nucleic acid-binding/OB-fold domains with similar parameters in the same way. In addition, we further investigated individual portions of the alpha subunit genes by varying MEME parameters but still failed in finding any meaningful motifs in the phosphoesterase PHP domains. However, this



**Fig. 5** Domain alignments of the four groups of alpha subunit genes. InterProScan has been used to identify the domains. Scale bar marks the length measured by amino acids.



**Fig. 6** Protein motifs among the three groups of dnaE genes. The MEME tool has been employed for this analysis. Each motif describes a pattern of a fixed width, and no gaps are allowed in MEME motifs. MEME numbers the motifs consecutively starting from 1 as it found them, and the most statistically significant (low E-value) motifs are usually found first. Relative to dnaE1 and dnaE3, dnaE2 appears to miss motif 5. Scale bar indicates amino acid positions.

procedure paid off in finding another motif unique to dnaE2 in the nucleic acid-binding/OB-fold domains, which is a short sequence, "VVHLVAQRLED", at the end of this domain (data not shown). Collectively, dnaE1 and dnaE3 share a common motif in the alpha subunit domain, while dnaE2 has a unique motif in its nucleic acid-binding/OB-fold domain.

## Discussion

Our results clearly pointed out two important observations about the grouping of alpha subunit genes; one is the existence of the four distinct groups in eubacterial genomes, and the other is that the grouping scheme always has exceptions albeit mostly minor. We do believe that the grouping scheme is related to the nucleotide compositional dynamics of prokaryotic genomes, which exceeds taxonomy, so that it breaks the boundaries of the bacterial taxa, even within a genus. For instance, polC is characteristic in phylum Firmicutes and co-exists with dnaE3 (89% in Firmicutes), except in four genomes in class Clostridia of Firmicutes where polC co-exists with dnaE1. Having detected these four dnaE1 genes, we found that they are clustered together in dnaE1 group and are homologous to the dnaE1 gene of *S. thermophilum* IAM14863 in phylum Actinobacteria. Meanwhile, phylogenetic analysis (Figure 1) showed that the polC gene of *S. thermophilum* IAM14863 is homologous to the polC genes of the above four genomes in Clostridia and also homologous to that of one genome in phylum Fusobacteria. The dnaE1 gene of *S. thermophilum* IAM14863 is also different from those of other bacteria in Actinobacteria; it is clustered with the dnaE1 gene of *Dehalococcoides ethenogenes* 195 (Bact063) in phylum Chloroflexi (which contains only one dnaE1) and the four dnaE1 genes of the genomes in Clostridia, whereas the other dnaE1 genes in Actinobacteria are clustered together. These results of the detection of homologous dnaE1 and polC genes in distant species suggests that horizontal gene transfer occurs frequently among prokaryotic genomes (32).

This co-existence between polC and dnaE3/dnaE1 suggests that polC can form heterodimers with both dnaE1 and dnaE3 but prefers dnaE3 in most of the cases. Besides the bacteria in Firmicutes, three bacteria in other phyla also possess polC: *Fusobacterium nucleatum* in Fusobacteria and *Thermotoga maritima* in Thermotogae both harbor polC and dnaE1. The third one, *S. thermophilum* IAM14863 in Actinobac-

teria, is even more complex, which contains one polC, one dnaE1, and one dnaE2; however, it seems that dnaE2 may pair with dnaE1 to form a heterodimer just like other bacteria in Actinobacteria. Alternatively, polC might be a latecomer that has not yet made its signature visible in this genome. It would be interesting to know what is going on in this bacterium in terms of DNA replication.

PolC and dnaE1 play a dominant role in the replication in Firmicutes and non-Firmicutes, respectively. DnaE3 is present only together with polC in Firmicutes. DnaE2 is present only in the dnaE1/dnaE2 type of bacteria (mostly Actinobacteria and Proteobacteria), and pairs with dnaE1, while very rarely with polC, and never dnaE3 (at least in our current collection), despite the fact that horizontal gene transfer is frequent. Another observable fact is the compatibility among these genes. For instance, dnaE2 and dnaE3 are never found in the same genome so that they should be incompatible.

We also reviewed the literature to see what has been noticed in alpha subunit diversity by other investigators. It is clear that the difference of dnaE genes in *E. coli* and *B. subtilis* was described previously (20). Our analysis indicated that the dnaE<sub>BS</sub> gene discovered in *B. subtilis* actually belongs to the dnaE3 group in our study. It is reported that dnaE genes contain two different isoforms; one is encoded by an intact gene, the other is encoded by a split dnaE1 and dnaE2 interrupted by an intein sequence (24). We identified these "split dnaE genes" in four Cyanobacteria genomes within our dataset: *Nostoc* sp., *Synechococcus elongatus* PCC 6301, *Synechocystis* sp. PCC6803, and *Thermosynechococcus elongates*. The four genes are all members of the dnaE1 group. The dnaE2 genes found in other studies (25, 26) are exclusive members of our dnaE2 group although they were assigned with different functions.

## Conclusion

We have classified eubacterial alpha subunit genes into four different groups: polC, dnaE1, dnaE2, and dnaE3. DnaE1 functions as homodimers and the rest form heterodimers. Phylogenetic analysis have suggested that dnaE and polC might have diverged from the same ancestor and the former further split into three subclasses. Two isoforms of dnaE1 form a homodimer that catalyzes DNA replication. DnaE2 might have subsequently lost a small but significant

motif and is flexible to dimerize with dnaE1 and polC. DnaE3 is the least flexible so that it pairs only with polC, becoming Firmicute-specific as a result. We hope that our consolidated naming convention will be used in future gene annotations for eubacterial genomes.

## Materials and Methods

We extracted alpha subunit genes from 191 bacterial sequences deposited in GenBank (May 19, 2005; ftp.ncbi.nlm.nih.gov/genomes/Bacteria/), which cover 159 species, and classified them according to the NCBI taxonomy database. For convenience, we numbered the bacteria in an alphabetic order and assigned abbreviations (Table S1). For those taxonomic groups with more than one genome sequenced, we randomly selected one to represent the species. Since incomplete annotation may lead to oversight, we also searched each genome with the extracted sequence with the BLAT tools (33). We detected the alpha subunit genes that are diverged in sequences and lengths, which are far shorter than the average sizes. As a result, we obtained 248 alpha subunit genes from the 159 genomes. For phylogenetic analysis, we used ClustalW for amino acid sequence alignments (34) and MEGA (bootstrap neighbor-joining method) (35) for phylogeny. The InterProScan (28) and MEME (30, 31) tools were used for domain and motif analyses.

## Acknowledgements

We thank Zhang Zhang and Xiang-Jun Tian for their constructive comments on this manuscript. This work was supported by a grant from the Chinese Academy of Sciences awarded to JY (KSCX2-SW-331).

## Authors' contributions

XQZ carried out sequence extraction, sequence analysis, and drafted the manuscript. JFH performed the sequence collection and biological information classification. JY designed and supervised the project as well as revised the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors have declared that no competing interests exist.

## References

1. Goodman, M.F. 2002. Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu. Rev. Biochem.* 71: 17-50.
2. Lewin, B. 2004. *Genes VIII*. Oxford University Press, Oxford, UK.
3. Lehman, I.R. 1989. Discovery of DNA polymerase. *J. Biol. Chem.* 278: 34733-34738.
4. Tang, M., et al. 1999. UmuD'2C is an error-prone DNA polymerase, *Escherichia coli* pol V. *Proc. Natl. Acad. Sci. USA* 96: 8919-8924.
5. Kobayashi, S., et al. 2002. Fidelity of *Escherichia coli* DNA polymerase IV. Preferential generation of small deletion mutations by dNTP-stabilized misalignment. *J. Biol. Chem.* 277: 34198-34207.
6. Patel, P.H., et al. 2001. Prokaryotic DNA polymerase I: evolution, structure, and "base flipping" mechanism for nucleotide selection. *J. Mol. Biol.* 308: 823-837.
7. McHenry, C.S. and Crow, W. 1979. DNA polymerase III of *Escherichia coli*. Purification and identification of subunits. *J. Biol. Chem.* 254: 1748-1753.
8. Kelman, Z. and O'Donnell, M. 1995. DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* 64: 171-200.
9. Herendeen, D.R. and Kelly, T.J. 1996. DNA polymerase III: running rings around the fork. *Cell* 84: 5-8.
10. McHenry, C. and Kornberg, A. 1977. DNA polymerase III holoenzyme of *Escherichia coli*. Purification and resolution into subunits. *J. Biol. Chem.* 252: 6478-6484.
11. Scheuermann, R.H. and Echols, H. 1984. A separate editing exonuclease for DNA replication: the epsilon subunit of *Escherichia coli* DNA polymerase III holoenzyme. *Proc. Natl. Acad. Sci. USA* 81: 7747-7751.
12. Maki, H., et al. 1985. The polymerase subunit of DNA polymerase III of *Escherichia coli*. I. Amplification of the dnaE gene product and polymerase activity of the alpha subunit. *J. Biol. Chem.* 260: 12982-12986.
13. Maki, H. and Kornberg, A. 1985. The polymerase subunit of DNA polymerase III of *Escherichia coli*. II. Purification of the alpha subunit, devoid of nuclease activities. *J. Biol. Chem.* 260: 12987-12992.
14. Maki, H. and Kornberg, A. 1987. Proofreading by DNA polymerase III of *Escherichia coli* depends on cooperative interaction of the polymerase and exonuclease subunits. *Proc. Natl. Acad. Sci. USA* 84: 4389-4392.
15. Maki, H., et al. 1988. DNA polymerase III holoenzyme of *Escherichia coli*. IV. The holoenzyme is an



- asymmetric dimer with twin active sites. *J. Biol. Chem.* 263: 6570-6578.
16. Stukenberg, P.T., *et al.* 1991. Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J. Biol. Chem.* 266: 11328-11334.
  17. Onrust, R., *et al.* 1995. Assembly of a chromosomal replication machine: two DNA polymerases, a clamp loader, and sliding clamps in one holoenzyme particle. *J. Biol. Chem.* 270: 13366-13377.
  18. Inoue, R., *et al.* 2001. Genetic identification of two distinct DNA polymerases, DnaE and PolC, that are essential for chromosomal DNA replication in *Staphylococcus aureus*. *Mol. Genet. Genomics.* 266: 564-571.
  19. Barnes, M.H., *et al.* 2002. DNA polymerases of low-GC gram-positive eubacteria: identification of the replication-specific enzyme encoded by dnaE. *J. Bacteriol.* 184: 3834-3838.
  20. Dervyn, E., *et al.* 2001. Two essential DNA polymerases at the bacterial replication fork. *Science* 294: 1716-1719.
  21. Bruck, I. and O'Donnell, M. 2000. The DNA replication machine of a gram-positive organism. *J. Biol. Chem.* 275: 28971-28983.
  22. Bruck, I., *et al.* 2003. The essential C family DnaE polymerase is error-prone and efficient at lesion bypass. *J. Biol. Chem.* 278: 44361-44368.
  23. Wu, H., *et al.* 1998. Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* 95: 9226-9231.
  24. Huang, Y.P. and Ito, J. 1999. DNA polymerase C of the thermophilic bacterium *Thermus aquaticus*: classification and phylogenetic analysis of the family C DNA polymerases. *J. Mol. Evol.* 48: 756-769.
  25. Galhardo, R.S., *et al.* 2005. An SOS-regulated operon involved in damage-inducible mutagenesis in *Caulobacter crescentus*. *Nucleic Acids Res.* 33: 2603-2614.
  26. Boshoff, H.I., *et al.* 2003. DnaE2 polymerase contributes to *in vivo* survival and the emergence of drug resistance in *Mycobacterium tuberculosis*. *Cell* 113: 183-193.
  27. Koonin, E.V. and Bork, P. 1996. Ancient duplication of DNA polymerase inferred from analysis of complete bacterial genomes. *Trends Biochem. Sci.* 21: 128-129.
  28. Apweiler, R., *et al.* 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29: 37-40.
  29. Lewin, B. 2000. *Genes VII*. Oxford University Press, Oxford, UK.
  30. Grundy, W.N., *et al.* 1996. ParaMEME: a parallel implementation and a web interface for a DNA and protein motif discovery tool. *Comput. Appl. Biosci.* 12: 303-310.
  31. Grundy, W.N., *et al.* 1997. Meta-MEME: motif-based hidden Markov models of protein families. *Comput. Appl. Biosci.* 13: 397-406.
  32. Lawrence, J.G. and Ochman, H. 1998. Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA* 95: 9413-9417.
  33. Kent, W.J. 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12: 656-664.
  34. Ebedes, J. and Datta, A. 2004. Multiple sequence alignment in parallel on a workstation cluster. *Bioinformatics* 20: 1193-1195.
  35. Kumar, S., *et al.* 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* 5: 150-163.
- Supporting Online Material**  
<http://evolution.genomics.org.cn/supp/supp.htm>  
Table S1