Ancient duplication of DNA polymerase inferred from analysis of complete bacterial genomes

The complete genome sequences of two bacteria – *Haemophilus influenzae* with 1703 genes and *Mycoplasma genitalium* with only 468 genes – have been recently reported[1,2]. Comparative analysis of these sequences is expected to provide numerous insights into bacterial physiology and evolution. This is illustrated by the surprising observation that *M. genitalium* encodes two proteins homologous to the DNA polymerase III α subunit (PolC), the principal DNA polymerase involved in bacterial chromosome replication. Under normal conditions, key enzymes of bacterial genome replication and expression are not usually duplicated[3] and such a duplication is particularly unexpected in the 'minimal' *M. genitalium*[2,4]. One of these proteins, MG031, consisting of 1451 amino acid residues, shows the closest similarity to PolC of Gram-positive bacteria and has the same organization of the polymerase and editing-exonuclease domains[5] (Fig. 1). This is not unexpected as mycoplasmas are thought to have evolved from Gram-positive bacteria[6]. By contrast, MG261 (874 amino acids) is most closely related to PolC (DnaE) of Gram-negative bacteria and similarly, lacks the exonuclease domain (Fig. 1).

The grouping of the two predicted DNA polymerases of *M. genitalium* with the polymerases of Gram-positive and Gram-negative bacteria, respectively, is fully supported by the phylogenetic tree analysis (Fig. 2). Screening of the collection of sequenced *Mycoplasma capricolum* DNA clones[7] shows that this mycoplasma, which is only distantly related to *M. genitalium*, also encodes a Gram-positive-type and a Gram-negative-type PolC α subunit (clones MC444_1 and MC129_1, respectively). Furthermore, the biochemical analysis of DNA polymerases from *Mycoplasma pulmonis* has revealed, in addition to the cloned Gram-positive-type PolC, another activity that resembles the polymerase from Gram-negative bacteria by its sensitivity to nucleotide analogs[8].

Analysis of the proteins encoded in the complete *H. influenzae* genome ascertained that this bacterium encodes only one PolC (Refs 1, 9); this is likely to be the case also in *Escherichia coli*, for which over 75% of the genome sequence is available[3]. The most plausible explanation of these findings is an ancient duplication of the polC gene predating the radiation of Gram-positive and Gram-negative bacteria, with the subsequent elimination of one of the copies in the Gram-negative lineage (or at least in the lineage leading to the gamma subdivision of Proteobacteria, to which *H. influenzae* and *E. coli* belong).

A duplication occurring later in only the Gram-positive lineage is unlikely, as this would imply that the Gram-negative-type enzyme has been evolving much more slowly than the Gram-positive enzyme; given the conservation of the polymerase function, this is a remote possibility.

The only viable alternative to the ancient duplication hypothesis is the horizontal transfer of a *polC* gene from a Gram-negative bacterium before the radiation of the mycoplasmas.

Mycoplasma genitalium lacks the DNA polymerase I gene (*polA*) that is present in *H. influenzae*, *E. coli* and *Bacillus subtilis* and is primarily involved in DNA repair[10]. Thus, MG031 and MG261 are the only two DNA polymerases in *M. genitalium*, and one of them is likely to be responsible for the repair DNA synthesis. The best candidate is MG261, the Gram-negative-type enzyme, which belongs to a putative operon with two genes coding for repair enzymes, namely a putative exonuclease homologous to the 5′–3′ exonuclease domain of PolA (MG262) and formamidopyrimidine DNA glycosylase [a homolog of MutM from other bacteria; we have designated it MG262a (Ref. 4)].

Further sequencing of bacterial genomes will show how widespread is the DNA polymerase duplication. In addition to their significance for our understanding of bacterial genome evolution, the findings presented here immediately result in experimentally testable predictions on the role of specific *Mycoplasma* proteins in such central aspects of bacterial cell physiology as DNA replication and repair.

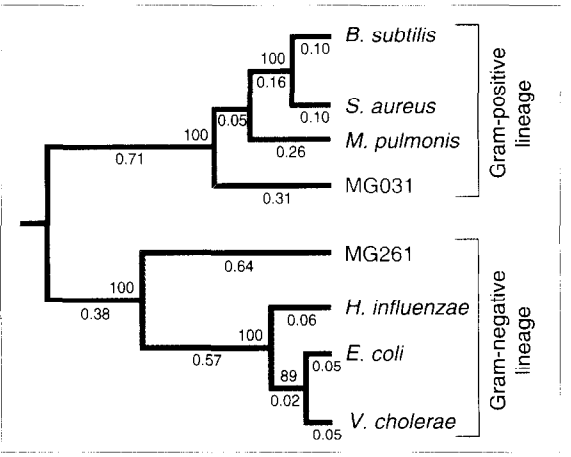

Figure 2

A tentative phylogenetic tree for the polymerase domains of DNA polymerase III (PolC). The tree was derived from the alignment of eight concatenated motifs (a total of 275 aligned amino acid residues) that are conserved in all PolC proteins; the alignment was constructed using the MACAW program[13]. A distance matrix was derived from this alignment with the PROTDIST program of the PHYLIP package[14] with the PAM120 amino acid substitution table, and the tree was constructed with the UPGMA clustering method. The percentage of bootstrap replications (out of a total of 500), in which each of the branchings was observed, is indicated next to each internal node. The branch lengths in relative units are also shown. Other methods for tree construction, namely the protein parsimony method and the Fitch–Margoliash least square method, produced the same topology supported by bootstrap analysis, assuming the root position between the Gram-positive and Gram-negative lineages. The sequences were from the GenBank database: MG031, MGU39681_1; MG261, MGU39705_1; *Bacillus subtilis* PolC, BSPOLC_1; *Staphylococcus aureus* PolC, SADNAPO3_2; *Mycoplasma pulmonis* PolC, MPU06833_1; *Haemophilus influenzae* DnaE, HIU32757_5; *Escherichia coli* DnaE, ECOLPXA_4; *Vibrio cholerae* DnaE, VCU30472_2.

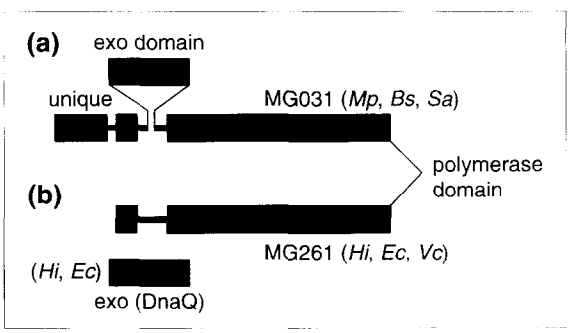

Figure 1

The structure of the two predicted DNA polymerase III enzymes in *Mycoplasma genitalium*. The polymerase domains of **(a)** the Gram-positive-type PolC (MG031) and **(b)** the Gram-negative-type PolC (MG261) are shown by differential coloring. In a database search with the BLASTP program[11], MG031 had a score of 771 (probability of matching by chance: $P = 5.9 \times 10^{-276}$) with the *Bacillus subtilis* PolC, and a score of 80 ($P = 1.6 \times 10^{-21}$) with the *Escherichia coli* PolC (DnaE); MG261 had a score of 283 ($P = 2.2 \times 10^{-80}$) with the *E. coli* PolC and a score of 62 ($P = 3.5 \times 10^{-9}$) with the *B. subtilis* PolC. The 3′–5′ exonuclease domain in the Gram-positive PolC appears to be inserted within the region that is conserved between Gram-positive and Gram-negative bacteria[12]; it is not known whether the small amino-terminal domain has a distinct function. In Gram-negative bacteria, the exonuclease domain resides in a separate protein (DnaQ). Abbreviations used: *Bs*, *Bacillus subtilis*; *Ec*, *Escherichia coli*; *Hi*, *Haemophilus influenzae*; *Mp*, *Mycoplasma pulmonis*; *Sa*, *Staphylococcus aureus*; *Vc*, *Vibrio cholerae*.

References

1 Fleischmann, R. D. *et al.* (1995) *Science* 269, 496–512

2 Fraser, C. M. *et al.* (1995) *Science* 270, 397–403

© 1996, Elsevier Science Ltd

PII: S0968-0004(96)30007-8

3 Koonin, E. V., Tatusov, R. L. and Rudd, K. E. (1995) *Proc. Natl Acad. Sci. USA* 92, 11921–11925
4 Koonin, E. V., Mushegian, A. R. and Rudd, K. E. *Curr. Biol.* (in press)
5 Barnes, M. H. *et al.* (1992) *Gene* 111, 43–49
6 Woese, C. R. (1987) *Microbiol. Rev.* 51, 221–271
7 Bork, P. *et al.* (1995) *Mol. Microbiol.* 16, 955–967
8 Barnes, M. H. *et al.* (1994) *Mol. Microbiol.* 13, 843–854
9 Tatusov, R. L. *et al. Curr. Biol.* (in press)
10 Kornberg, A. and Baker, T. (1992) *DNA Replication* (2nd edn), W. H. Freeman
11 Altschul, S. F. *et al.* (1990) *J. Mol. Biol.* 215, 403–410
12 Koonin, E. V. and Deutscher, M. P. (1993) *Nucleic Acids Res.* 21, 2521–2522
13 Schuler, G. D., Altschul, S. F. and Lipman, D. J. (1991) *Protein Struct. Funct. Genet.* 9, 180–190
14 Felsenstein, J. (1989) *Cladistics* 5, 164–166

EUGENE V. KOONIN

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA.

PEER BORK

Max-Delbrück-Centre for Molecular Medicine, Berlin-Buch, Germany; and European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany.

The ethylene signal transduction pathway in *Arabidopsis*: an emerging paradigm?

Caren Chang

In the plant *Arabidopsis*, ethylene signaling involves at least two putative receptors, both of which resemble the 'two-component' regulators known almost exclusively in bacteria. Downstream in the pathway is a putative serine/threonine protein kinase related to the animal Raf protein kinases. This novel combination of signaling proteins has parallels with the postulated osmolarity-response pathway in yeast.

THE SIMPLE GAS ethylene (C_2H_4) behaves as a plant hormone with numerous effects on plant growth and development. Responses to ethylene include fruit ripening, senescence, abscission, promotion of seed germination, promotion or inhibition of cell elongation and promotion or inhibition of flowering and stress-related responses[1]. Although the steps in the ethylene biosynthetic pathway are well established[1], much less is known about the signaling pathway leading from ethylene perception to physiological responses. Various ethylene-inducible genes have been cloned and *trans*-acting factors involved in their expression are beginning to be identified[2,3].

In *Arabidopsis thaliana*, a number of ethylene-response mutants have been isolated, defining several loci in the ethylene signal transduction pathway[4]. Isolation of these mutants has been based on the conspicuous 'triple-response' phenotype, in which dark-grown seedlings treated with ethylene display inhibition of hypocotyl and root elongation, radial swelling of the hypocotyl and root, and exaggeration of the apical hook curvature[4]. Most of the mutants have been ordered into a genetic pathway as suggested by double-mutant analyses (Fig. 1; Refs 5–7; and G. Chen and A. Bleecker, pers. commun.).

The *etr* mutant (renamed *etr1-1*) was suspected to have a defect in ethylene perception owing to its reduced ethylene binding and its failure to respond to ethylene[8]. All known *etr1* mutant alleles confer ethylene insensitivity and are dominant to the wild-type allele[9]. By contrast, recessive *ctr1* mutants exhibit constitutive ethylene responses[6]. That is, *ctr1* mutants behave as though they are exposed to ethylene even in its absence, indicating that the wild-type CTR1 protein is a negative regulator of ethylene responses. From double-mutant analyses, the *CTR1* gene acts downstream of *ETR1* (Fig. 1).

Three genes in the ethylene-response pathway, *ETR1* (Ref. 9), *ERS* (an *ETR1* homolog)[5] and *CTR1* (Ref. 6) have been cloned in *Arabidopsis*. This review focuses on the emerging picture of ethylene signal transduction provided by these cloned genes.

ETR1: an ethylene receptor with similarity to bacterial 'two-component' regulators

Bacterial two-component regulators control a broad range of adaptive responses to an array of environmental stimuli such as nitrogen, carbon and osmolarity[10]. The two components of the system are known as a sensor and an associated response regulator[10] (Fig. 2a). Essentially, these components act in pairs to control responses to specific signals.

Signal perception by the amino-terminal input domain of the sensor (either by direct ligand binding or by other components) promotes autophoshorylation of a conserved histidine residue. The phosphate is then transferred from histidine to a conserved aspartate residue in the receiver domain of the cognate response regulator. The phosphorylation state of the receiver controls the attached output domain, which in turn mediates downstream steps. Many bacterial output domains are transcriptional regulators.

The *Arabidopsis ETR1* gene, isolated by map-based cloning[9], displays sequence similarity to the bacterial family of two-component regulators in the carboxy-terminal half of the deduced 738-amino acid ETR1 protein[10,11] (Fig. 2). The amino-terminal half of the deduced ETR1 protein (residues 1–325) is novel[9]. Recently, Schaller and Bleecker used a yeast expression system to demonstrate that the amino-terminal portion of ETR1 binds ethylene; the half-life for ethylene dissociation (12.5 hours) is consistent with that of ethylene dissociation rates detected in various plant extracts[12].

This result, coupled with the *etr1* mutant phenotype, provides compelling evidence that ETR1 is an ethylene receptor. Moreover, when the same experiment was performed using the mutant *etr1-1* allele, there was no detectable ethylene

C. Chang is at the Department of Plant Biology and Maryland Agricultural Experiment Station, University of Maryland, College Park, MD 20742, USA.

© 1996, Elsevier Science Ltd

PII: S0968-0004(96)10019-0