J Mol Evol (2006) 63:781–787
DOI: 10.1007/s00239-006-0037-7

JOURNAL OF MOLECULAR EVOLUTION

Multiple Gene Duplication and Rapid Evolution in the *groEL* Gene: Functional Implications

Kshama Goyal,[1] Rohini Qamra,[2] Shekhar C. Mande[1]

[1] Centre for DNA Fingerprinting and Diagnostics, ECIL Road, Nacharam, Hyderabad 500 076, India
[2] Department of Biophysics, University of Delhi, South Campus, New Delhi 110021, India

Received: 15 February 2006 / Accepted: 16 August 2006 [*Reviewing Editor:* Dr. Debashish Bhattacharya]

Abstract. The chaperonins, GroEL and GroES, are present ubiquitously and provide a paradigm in the understanding of assisted protein folding. Due to its essentiality of function, GroEL exhibits high sequence conservation across species. Complete genome sequencing has shown the occurrence of duplicate or multiple copies of *groEL* genes in bacteria such as *Mycobacterium tuberculosis* and *Corynebacterium glutamicum*. Monophyly of each bacterial clade in the phylogenetic tree generated for the GroEL protein suggests a lineage-specific duplication. The duplicated *groEL* gene in Actinobacteria is not accompanied by the operonic *groES* despite the presence of upstream regulatory elements. Our analysis suggests that in these bacteria the duplicated *groEL* genes have undergone rapid evolution and divergence to function in a GroES-independent manner. Evaluation of multiple sequence alignment demonstrates that the duplicated genes have acquired mutations at functionally significant positions including those involved in substrate binding, ATP binding, and GroES binding and those involved in inter-ring and intra-ring interactions. We propose that the duplicate *groEL* genes in different bacterial clades have evolved independently to meet specific requirements of each clade. We also propose that the *groEL* gene, although essential and conserved, accumulates nonconservative substitutions to exhibit structural and functional variations.

Correspondence to: Shekhar C. Mande; *email:* shekhar@cdfd.org.in

Key words: GroEL — Paralogous genes — Functional divergence — Structure-function relationship — CIRCE — Gene duplication

Introduction

Chaperonins are large multisubunit proteins that assist proper folding of misfolded or unfolded polypeptides in an ATP-dependent manner. The 60-kDa chaperonins are classified into two groups, Group I and Group II, on the basis of amino acid sequence similarity. Group I chaperonins occur in bacteria and the endosymbiotic organelles of eukaryotes, and Group II chaperonins occur in archaea and the eukaryotic cytosols. Although chaperonins in both groups function as an oligomeric assembly, they possess distinct structural differences. Group I chaperonins form a seven-member ring, whereas Group II chaperonins assemble to form an eight- or a nine-member ring (Gutsche et al. 1999). Also, the bacterial chaperonins require a co-chaperonin for their function, whereas a similar role in group II chaperonins is played by a helical extension of the apical domain.

One of the best-studied chaperonins, GroEL, belongs to the 60-kDa group I chaperonin class. GroEL is required for the proper folding of many proteins in vitro and ~10% of newly translated polypeptides in vivo (Houry et al. 1999). GroEL is known to be overexpressed during varied stress conditions and has been shown to be indispensable during heat stress and thus named a heat shock protein. Moreover, GroEL

has been shown to be essential for growth at all temperatures (Fayet et al. 1989).

GroEL functions as a tetradecamer of two heptameric rings that stack back to back with a hydrophobic central cavity where substrate polypeptides are folded through alternate binding of the substrate and the co-chaperonin GroES (Braig et al. 1994). Each subunit of GroEL is comprised of three domains, *viz*., the equatorial, the intermediate, and the apical domains. The equatorial domain is involved in ATP binding and hydrolysis and provides most of the intra- and inter-ring contacts. The apical domain is involved in interaction with the substrate and the co-chaperonin, GroES. The intermediate domain links the apical and the equatorial domains and acts as a hinge for movement of the apical domain in response to signals provided by the equatorial domain.

With the advent of genome sequencing, more than 300 putative *groEL* homologues have been identified in several eubacteria and a few archaebacteria. Most of the bacteria possess a single copy of the *groEL* gene arranged on an operon along with the *groES* gene. However, analysis of several completely sequenced genomes that are available has shown that certain Gram-positive bacteria such as *M. tuberculosis* and *C. glutamicum* and a few Gram-negative bacteria, such as *Bradyrhizobium japonicum* and *Mesorhizobium loti* possess duplicate or multiple copies of the *groEL* genes. In Gram-positive bacteria, the second copy of the *groEL* gene mainly occurs as a stand-alone copy, while in Gram-negative bacteria almost all the multiple copies of *groEL* genes are preceded by the corresponding *groES* genes, and hence, all are bi-cistronic.

In many Gram-positive bacteria, the Controlling Inverted Repeat for Chaperonin Expression (CIRCE) element controls regulation of the *groESL* operon. These elements occur upstream of the *groESL* operon and regulate the operon through binding of the HrcA repressor (Narberhaus 1999). HrcA negatively regulates the operon, thus, at low temperature it remains bound to the CIRCE element, thereby preventing expression of the *groESL* operon. However, at high temperatures it dissociates from the repeat region and allows overexpression of the *groESL* operon.

Gene duplication is an important evolutionary force that provides an organism an opportunity to evolve new functions. Known examples of gene duplication include instances of one of the duplicated copies diverging to acquire differential regulation. Many times mutations occur in one of the duplicated genes, followed by its evolution into a protein with a new function. In the case of oligomeric proteins, the duplicate copies sometimes evolve to function as hetero-oligomers (Dickson et al. 2000). In some cases, duplication is also used as a mechanism to acquire varied substrate spectrum, as in the case of immunoglobulins, where multiple copies respond to different substrates (Eason et al. 2004). Thus, functional variations and differential regulation obtained as a result of gene duplication provide organisms an adaptive advantage and a better chance of survival.

In the present study, phylogenetic analysis of GroEL was carried out to understand the evolutionary significance of the *groEL* gene duplication in different lineages and the functional implications of the duplication event. Because GroES is essential for the proper functioning of GroEL (Fenton et al. 1994), we also analyzed the duplication of *groESL* operons in the available bacterial sequences.

Materials and Methods

BLAST (Altschul et al. 1997) search was carried out using the amino acid sequence of GroEL1 (Rv3417c) of *M. tuberculosis* with a cutoff E-value of 10^{-4} against the Swiss-Prot database (Bairoch and Apweiler 1997). Three hundred one hits, representing a wide range of the phylogenetic spectrum, were obtained in the BLAST output. Protein sequences of GroEL2 and GroEL3 of *Chlamydia trachomatis* were added manually. Sequences were retrieved and aligned using the ClustalW program (Thompson et al. 1994). The multiple sequence alignment was edited to remove redundant sequences. Sequences sharing less than 20% identity with the query were also removed from the set. As large gaps cause discrepancy in distance calculation and in inference of evolutionary relationships, sequences containing large insertions and deletions were removed from the alignment. The sequence set was realigned using ClustalW. The highly variable C-terminal region from all the sequences was not considered for further analysis.

Phylogenetic Analysis

The phylogenetic tree was generated using the distance-based and the maximum-likelihood (ML) methods of tree construction. For distance analyses, the PROTDIST program of the PHYLIP (v 3.5) package was used (Felsentein 1993), with the distances estimated from the Dayhoff PAM matrices. The tree was generated using the neighbor-joining (NJ) algorithm as implemented in the NEIGHBOR program of the PHYLIP package. One hundred bootstrap cycles were carried out to obtain statistical support for the phylogeny. The majority-rule consensus tree was determined from the bootstrapped trees using the CONSENSUS module within PHYLIP.

The PROTTEST (Abascal et al. 2005) program was used to determine the empirical model of amino acid substitutions that best fits the GroEL multiple protein sequence alignment. PROTTEST tests for 72 amino acid substitution models. The best substitution matrix suggested by PROTTEST, the WAG matrix (Whelan and Goldman 2001), was used for ML analyses, using PHYML (Guindon and Gascuel 2003). The ML analysis was used to correct the consensus tree considering a 0.016 fraction of the total sites as invariant and used a gamma distribution with $\alpha = 0.908$ as obtained from PROTTEST.

Assessment of Functional Divergence

The DIVERGE program (Gu and Zhang 1997) was used to determine functionally divergent nodes (clusters) in the final ML tree. It was also used to assess the posterior probability of

divergence of each amino acid site in the alignment for the two clusters that are compared.

Results and Discussion

The final sequence alignment used for evolutionary analysis contained a total of 249 sequences. After the removal of ambiguous regions, the alignment contained 565 amino acid positions. Conservation was found to be high at 55 positions in the alignment, suggesting conserved chemical similarity of these residues across all 249 protein sequences.

Evolutionary Analysis of Multiple Gro*ELs*

In order to determine the basic phylogenetic relationship among the 249 sequences, a NJ tree was generated using the Dayhoff PAM matrix. The distances tested using PROTTEST showed that the WAG + I + G model best fits the present multiple sequence alignment. The +I feature considers a fraction of amino acid sites to be invariable due to functional and structural constraints (Reeves 1992). The +G feature infers evolutionary information by taking into account the probability of each site to belong to the given rate categories (Yang 1993). The model was selected with the minimum Akaike (1973) value of 134,688.19. It was selected with a weight of 0.94, suggesting a high probability of this model to fit the given alignment.

It was observed that the overall phylogenetic distribution was generally in agreement with the distribution in the 16S rRNA tree. Sequences belonging to the same class of bacteria were clustered together. Each bacterial clade was clearly separated in the phylogenetic tree. This therefore suggests that there has been little horizontal transfer of the *groEL* genes across different bacterial clades. However, duplication and rapid evolution are still evident from the phylogenetic tree.

Paralogy of the chaperonin gene has been observed in the archaeal and eukaryotic chaperonins (Archibald et al. 2001). Similarly, duplication of the *groEL* gene was also observed in many bacterial species. For example, Chlamydia, Cyanobacteria, Actinobacteria, and Rhizobiales each possessed two or more copies of the *groEL* gene, which were referred to as *groEL1*, *groEL2*, and *groEL3*, etc. Comparison of the GroEL protein sequences across different lineages showed that the duplicated copies of GroEL are more similar to each other within each lineage than across different lineages. In other words, the tree topology, supported by high bootstrap values, indicated that the duplication of the *groEL* gene might not have occurred in the common ancestor of Chlamydia, Cyanobacteria, and Actinobacteria. Rather, an independent lineage-specific duplication might have occurred in each of the clades. For example, the cyanobacterial node is bifurcated into GroEL1 and GroEL2 branches with a bootstrap value of 93 (Fig. 1). The GroEL1 protein sequences of all the cyanobacterial species are substantially closer to each other than to their corresponding GroEL2 sequences. The two branches, one consisting of the GroEL1 sequences and the other comprising the GroEL2 sequences, therefore form separate groups originating from a single cyanobacterial node. This suggests that the duplication of the *groEL* gene occurred in the common ancestor of all the cyanobacterial species.

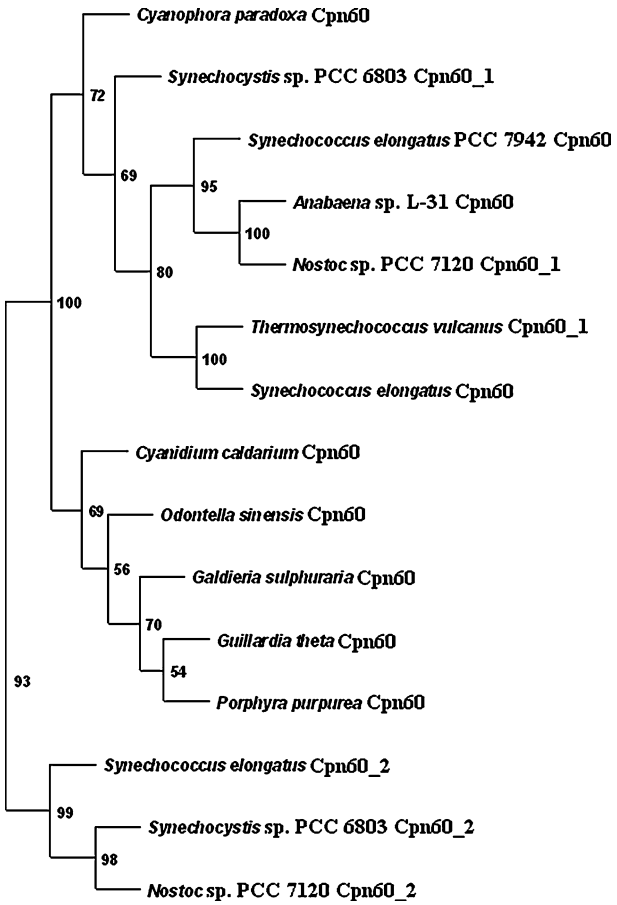

Fig. 1. The cyanobacterial clade. The tree topology strengthens the view that lineage-specific duplication has occurred in Cyanobacteria. Each branch is supported by a bootstrap value. Along with the species name, the copy number is also indicated. Cpn60_1 is GroEL1 and Cpn60_2 is GroEL2.

In the case of Actinobacteria (Fig. 2), *groEL* gene duplication was observed only in the taxa Actinomycineae and Corynebacterineae and was not observed in other taxa. The GroEL1s of Actinomycineae and Corynebacterineae form a single cluster, while the GroEL2s form a separate cluster. The clustering is supported by a bootstrap value of 94, thereby suggesting that a single duplication event occurred in the common ancestor of all Actinobacteria (Fig. 2).

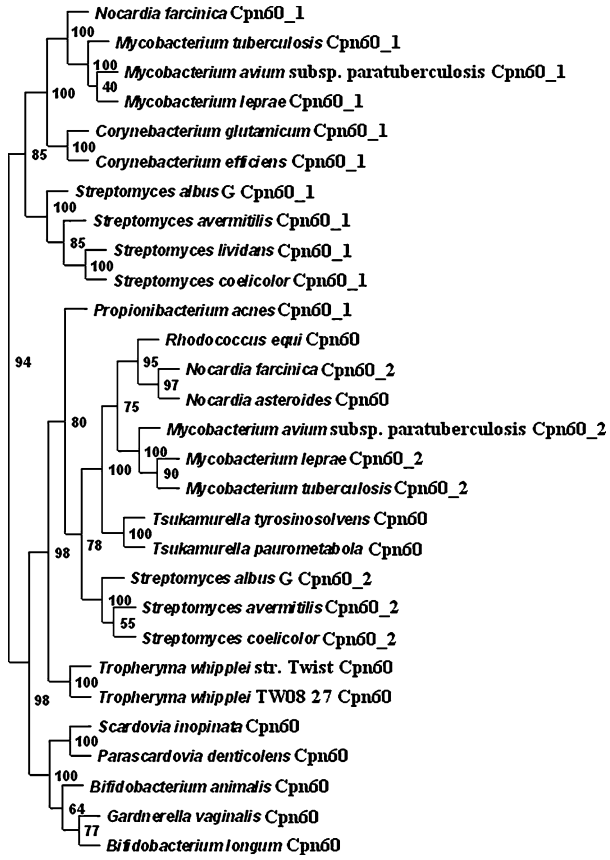

Fig. 2. The actinobacterial clade. The bootstrapped NJ tree topology clearly suggests that duplication occurred in the common ancestor of Actinobacteria. GroEL copy number is indicated as described in the legend to Fig. 1.

Interestingly, the duplicated *groEL* gene (*groEL2*) in Actinobacteria is not accompanied by the *groES* gene. The *groES* gene typically occurs between the CIRCE regulatory element and the *groEL1* gene. In the duplicated copies of the *groEL* genes in Actinobacteria, the CIRCE element is indeed present upstream of the *groEL2* gene (Hecker et al. 1996), but the *groES* gene is missing between the CIRCE element and the *groEL2* gene. Thus, the selective absence of the *groES* gene upstream of the *groEL2* genes was intriguing.

There are two possible explanations for such a gene inventory: (a) the *groEL* gene and its CIRCE regulator duplicated independently, maintaining a single copy of the *groES* gene, and (b) the entire operon was duplicated, followed by selective loss of the *groES* gene due to the absence of selection pressure from the second copy of *groEL*. In order to understand the possible evolutionary event we analyzed the distance between the CIRCE element and the duplicated genes and, also, explored the functional dependence of GroEL on GroES. It was observed that in *M. tuberculosis* H37Rv the distance between the *groEL1* gene and its CIRCE element (–168 nucleotides) is approximately equal to the distance between *groEL2* and its corresponding CIRCE element (–153 nucleotides) (Stewart et al. 2002). This suggested that the entire operon might have been duplicated in the common ancestor of Actinobacteria. Furthermore, it is known that GroES is strictly required for GroEL-mediated protein folding in bacteria such as *E. coli* (Ranson et al. 1998) but might not be required in *M. tuberculosis* for the proper functioning of GroELs (Qamra et al. 2004; Qamra and Mande 2004). The lack of GroES requirement for proper functioning of GroEL2 might thus have led to the loss of selection pressure on the *groES* gene. Thus, the similar distance between the regulator CIRCE and the duplicated *groEL* genes and the ability of *M. tuberculosis* GroELs to function in a GroES-independent manner together suggested that the complete *groESL* operon might have been duplicated in the ancestor of Actinobacteria, with the subsequent loss of the second copy of the *groES* gene.

Interestingly, it was observed that not all the species belonging to Actinobacteria possessed two *groEL* copies. Species such as *Bifidobacterium longum* and *Tropheryma whipplei* possessed only one copy of the *groEL* gene. The only copy of the *groEL* gene in these organisms is significantly more similar to the *groEL2* of other Actinobacteria. For example, the similarity between *T. whipplei* GroEL and *M. tuberculosis* GroEL1 is 69%, whereas the similarity between the former and the *M. tuberculosis* GroEL2 is 79%. Moreover, the only GroELs of these species cluster with the GroEL2 sequences of other Actinobacteria with a bootstrap value of 98. Furthermore, in *B. longum* and *T. whipplei*, the *groES* gene does not accompany the only *groEL* gene on the chromosome, thus indicating that these organisms lack the canonical *groESL* operon. The *groES* gene in *B. longum* and *T. whipplei* is present elsewhere on the genome independent of the *groEL* gene, suggesting possible loss of the *groEL1* gene due to redundancy. This observation is further supported by studies in *Streptomyces albus* (Servant et al. 1993), *M. smegmatis* (Kim et al. 2003), and *C. glutamicum* (Barreiro et al. 2005), where the *groEL1* gene was found to be dispensable, while the *groEL2* gene was shown to be essential for the growth and survival. It therefore appears that after the gene duplication event in the ancestor of Actinobateria, GroEL1 underwent accelerated evolution (Hughes 1993) and was finally lost from *B. longum* and *T. whipplei* due to the redundancy of GroEL function. It also appears that the only GroELs of *T. whipplei* and *B. longum* might not require GroES for their function. Based on these observations, we propose that in Actinobacteria, the complete *groESL* operon was duplicated and the *groEL2* gene evolved to function in a GroES-independent manner. This was followed by a selective loss of the *groES* gene from the duplicated operon.

Analysis of the distribution of GroELs of organisms belonging to the Chlamydiae clade showed complex events of gene duplication. The estimated distances showed that the *groEL2* genes of *Chlamydophila caviae* and *Chlamydophila pneumoniae* have undergone accelerated evolution (Supplementary Figure). The same also holds true for the *groEL2* and *groEL3* genes of *Chlamydia trachomatis*. Thus the GroEL phylogeny, in Chlamydiae, appears to be complex, with the events of duplication followed by higher rates of substitutions in the second and the third copies of *groEL* genes. Besides, GroEL2 and GroEL3 of these organisms have many mutations even at functionally important sites, suggesting possible loss of the chaperonin function. For example, the conserved ATP-binding motif GDGTTT has acquired mutations in GroEL2 (T89A and T90K) and GroEL3 (G86A, T89V, and T90I) of *C. trachomatis*. Further, a recent report suggests that GroEL2 and GroEL3 from *C. trachomatis* are not able to complement a temperature-sensitive *E. coli groEL* mutant, although GroEL1 from *C. trachomatis* does complement the same (Karunakaran et al. 2003). Thus, GroEL2 and GroEL3 of *C. trachomatis* might not function as the canonical chaperonins or might be in the process of evolving a hitherto unknown function.

In addition to Chlamydiae, rapid evolution of the *groEL* genes was also observed in the case of rickettsials. This observation was intriguing, as duplication was not seen in this clade. When the rickettsial chaperonins were compared with the chaperonins from rhizobials (the nearest neighbors of rickettsials), the coefficient of functional divergence (θ_{ml}) between rickettsial and rhizobial clades was ~0.512 as estimated by DIVERGE. In order to identify the variant amino acid sites in the rickettsial chaperonins, the posterior probability of divergence was determined for each site. These values were mapped on the monomeric template of *E. coli* GroEL (1AON) (Xu et al. 1997) (Fig. 3). It was noticed that these mutations are mostly clustered in the apical and the equatorial domain. As the equatorial domain is involved in inter-ring interactions, variations in this domain suggest probable loss of the double-ring formation. It is therefore not surprising that the mitochondrial chaperonins that are derived from Rickettsia (Emelyanov 2001) function as single-ring structures (Nielson et al. 1999). Further, the apical domain is responsible for substrate binding and folding. Thus, digression in some of the residues from this domain suggests that the rickettsial chaperonins might not recognize identical set of substrate proteins as identified by Rhizobiales. These variations in intermediate and equatorial domains together suggest that the Rickettsiales chaperonins have undergone divergence mediating the transition from prokaryotic chaperonins to mitochondrial chaperonins.

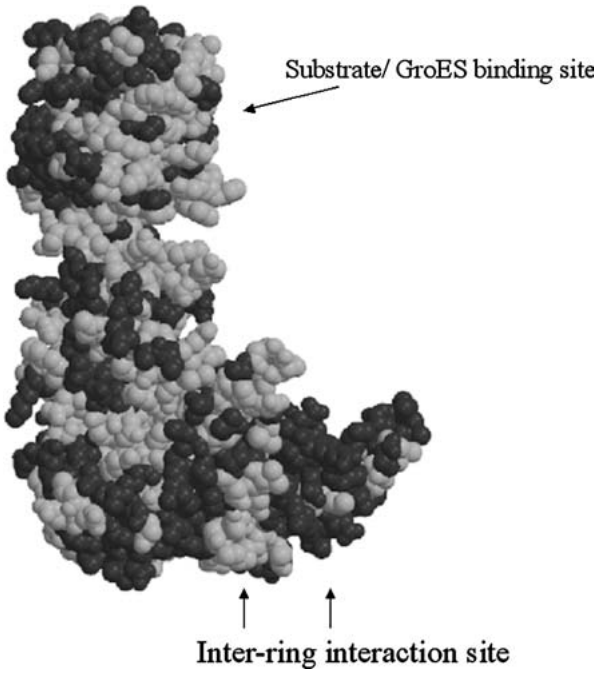

Fig. 3. Variation in the rickettsial chaperonins as mapped on the *E. coli* GroEL structure (1AON). Residues having a posterior probability of < 0.5 (i.e., less diverged) are shown in gray, and those with a high value (i.e., divergent) in black.

Finally, multiple duplication events were observed in the case of the rhizobial subclass. This led to a multifaceted clustering in this subclass, as each organism possessed five or six copies of the *groEL* genes. The tree topology (Fig. 4) clearly indicated numerous instances of gene duplication within the rhizobial subclass. However, the complex clustering suggests that, apart from gene duplications, a few horizontal transfer events also might have occurred in these organisms. Different copies of *groELs* might have diverged sequentially in a later event to attain the ability to function under different conditions of stress. Indeed it has been reported that in *B. japonicum* expression of the *groEL5* gene is regulated by cellular oxygen and that of *groEL3* is regulated by the nitrogen fixation system (Fischer et al. 1993). However, inferences regarding the chain of evolutionary events within the rhizobial clade cannot be drawn due to weak statistical support.

Conclusions

Our results indicate that the heat shock protein GroEL has been subjected to different selective constraints during its evolution. *groEL*, an essential gene, has undergone multiple events of gene duplication followed by sequence divergence. It is apparent from the study that the functional variations have been acquired by incorporating chemically dissimilar substitutions at functionally important residue positions.

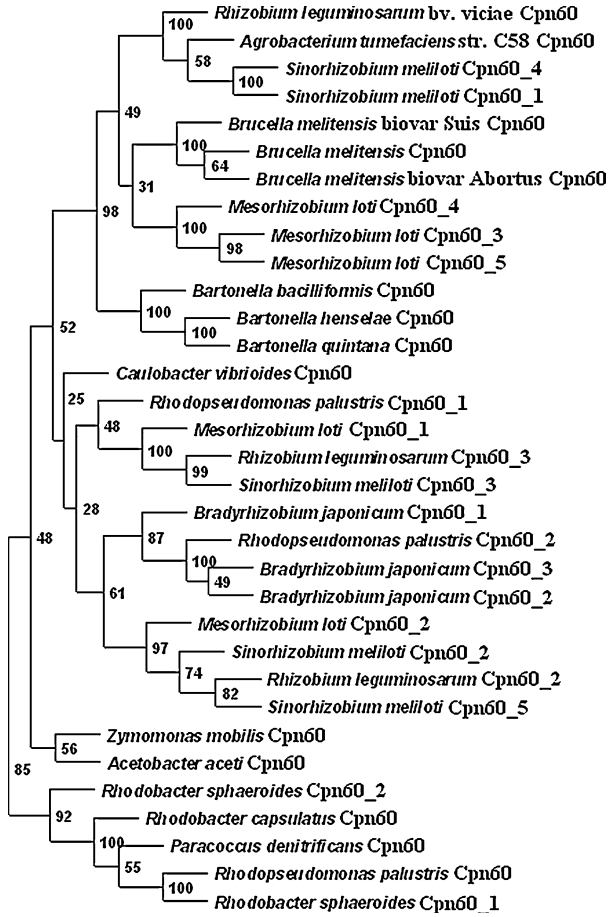

Fig. 4. The rhizobial clade. Depicting the complex evolution of multiple GroEL protein sequences supported by 100 bootstrap values in the NJ tree. Copy number is also indicated as described in the legend to Fig. 1. Cpn60_1 is GroEL1 and Cpn60_2 is GroEL2.

Actinobacteria appear to have lost the bi-cistronic *groESL*, retaining the mono-cistronic paralog that has evolved to function in the absence of the co-chaperonin GroES.

Acknowledgments. We acknowledge the overall support of S. K. Basu and S. E. Hasnain in this work. The support of the National Institute of Immunology in providing computational resources is gratefully acknowledged. We would also like to thank Debasisa Mohanty and Biju Issac for stimulating discussions. We would especially like to thank the anonymous referees for many useful comments and suggestions. K.G. is a CSIR Senior Research Fellow. S.C.M. is a Wellcome Trust International Senior Research Fellow.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21:2104–2105

Akaike H (1973) Information theory and an extension of the maximum likelihood principle. In: Proceedings of 2nd International Symposium on Information Theory. Budapest, Hungary, pp 267–281

Archibald JM, Blouin C, Doolittle WD (2001) Gene duplication and the evolution of group II chaperonins: Implication for structure and function. J Struct Biol 135:157–169

Bairoch A, Apweiler R (1997) The SWISS-PROT protein sequence database: its relevance to human molecular medical research. J Mol Med 75:312–316

Barreiro C, Gonzalez-Lavado E, Brand S, Tauch A, Martin JF (2005) Heat shock proteome analysis of wild-type *Corynebacterium glutamicum* ATCC 13032 and a spontaneous mutant lacking GroEL1, a dispensable chaperone. J Bacteriol 187:884–889

Braig K, Otwinowski Z, Hegde R, Boisvert DC, Joachimiak A, Horwich AL, Sigler PB (1994) The crystal structure of the bacterial chaperonin GroEL at 2.8 A. Nature 371:578–586

Dickson R, Weiss C, Howard RJ, Alldrick SP, Ellis RJ, Lorimer G, Azem A, Viitanen PV (2000) Reconstitution of higher plant chloroplast chaperonin 60 tetradecamers active in protein folding. J Biol Chem 275:11829–11835

Eason DD, Cannon JP, Haire RN, Rast JP, Ostrov DA, Litman GW (2004) Mechanisms of antigen receptor evolution. Semin Immunol 16:215–226

Emelyanov VV (2001) Evolutionary relationship of Rickettsiae and mitochondria. FEBS Lett 501:11–18

Fayet O, Ziegelhoffer T, Georgopoulos C (1989) The *groES* and *groEL* heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. J Bacteriol 171:1379–1385

Felsentein J (1993) PHYLIP—phylogeny inference package (version 3.5c). Department of Genome Sciences, University of Washington, Seattle

Fenton WA, Kashi Y, Furtak K, Horwich AL (1994) Residues in chaperonin GroEL required for polypeptide binding and release. Nature 371:614–619

Fischer HM, Babst M, Kaspar T, Acuna G, Arigoni F, Hennecke H (1993) One member of a *groESL*-like chaperonin multigene family in *Bradyrhizobium japonicum* is co-regulated with symbiotic nitrogen fixation genes. EMBO J 12:2901–2912

Gu X, Zhang J (1997) A simple method for estimating the parameter of substitution rate variation among sites. Mol Biol Evol 14:1106–1113

Gutsche I, Essen LO, Baumeister W (1999) Group II chaperonins: new TRiC(k)s and turns of a protein folding machine. J Mol Biol 293:295–312

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Hecker M, Schumann W, Volker U (1996) Heat-shock and general stress response in Bacillus subtilis. Mol Microbiol 19:417–428

Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU (1999) Identification of *in vivo* substrates of the chaperonin GroEL. Nature 402:147–154

Hughes AL (1993) Contrasting evolutionary rates in the duplicate chaperonin genes of *Mycobacterium tuberculosis* and *M. leprae*. Mol Biol Evol 10:1343–1359

Karunakaran KP, Noguchi Y, Read TD, Cherkasov A, Kwee J, Shen C, Nelson CC, Brunham RC (2003) Molecular analysis of the multiple GroEL proteins of Chlamydiae. J Bacteriol 185:1958–1966

Kim AI, Ghosh P, Aaron MA, Bibb LA, Jain S, Hatfull GF (2003) Mycobacteriophage Bxb1 integrates into the *Mycobacterium smegmatis* groEL1 gene. Mol Microbiol 50:463–473

Narberhaus F (1999) Negative regulation of bacterial heat shock genes. Mol Micro 31:1–8

Nielsen KL, McLennan N, Masters M, Cowan NJ (1999) A single-ring mitochondrial chaperonin (Hsp60-Hsp10) can substitute for GroEL-GroES *in vivo*. J Bacteriol 181:5871–5875

Qamra R, Mande SC (2004) Crystal structure of the 65 kDa heat shock protein, chaperonin 60.2 of *Mycobacterium tuberculosis*. J Bacteriol 186:8105–8113

Qamra R, Srinivas V, Mande SC (2004) *Mycobacterium tuberculosis* GroEL homologues unusually exist as lower oligomers and retain the ability to suppress aggregation of substrate proteins. J Mol Biol 342:605–617

Ranson NA, White HE, Saibil HR (1998) Chaperonins. Biochem J 333:233–242

Reeves JH (1992) Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. J Mol Evol 35:17–31

Servant P, Thompson C, Mazodier P (1993) Use of new *Escherichia coli*/*Streptomyces* conjugative vectors to probe the functions of the two groEL-like genes of *Streptomyces albus* G by gene disruption. Gene 134:25–32

Stewart GR, Wernisch L, Stabler R, Mangan JA, Hinds J, Laing KG, Young DB, Butcher PD (2002) Dissection of the heat-shock response in *Mycobacterium tuberculosis* using mutants and microarrays. Microbiology 148:3129–3138

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18:691–699

Xu Z, Horwich AL, Sigler PB (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP)7 chaperonin complex. Nature 388:741–750

Yang Z (1993) Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol 10:1396–1401