dnaJ is a useful phylogenetic marker for alphaproteobacteria

Ana Alexandre[1], Marta Laranjo[1], J. Peter W. Young[2], Solange Oliveira[1]*

[1]Laboratório de Microbiologia do Solo - Instituto de Ciências Agrárias Mediterrânicas (ICAM), Universidade de Évora, Portugal

[2]Department of Biology, University of York, UK

*Corresponding author. Mailing address: Universidade de Évora, Departamento de Biologia, Apartado 94, 7002-554 Évora, Portugal. Phone: +351 266 760878. Fax: +351 266 760914. E-mail: ismo@uevora.pt.

Running title: *dnaJ*, a phylogenetic marker for alphaproteobacteria

Contents Category: Evolution, Phylogeny and Biodiversity

The GenBank/EMBL/DDBJ accession numbers of new sequence data for *atpD*, *dnaJ*, *gapA, gyrB, recA, rplB* and 16S rRNA genes are given in Table 1 (in bold).

Summary

In the past, bacterial phylogeny relied almost exclusively on 16S rRNA gene sequence analysis. More recently, multilocus sequences have been used to infer organismal phylogenies. In this study, the *dnaJ* chaperone gene was investigated as a marker for phylogeny studies in alphaproteobacteria. Preliminary analysis of G+C and G+C3s contents showed no clear evidence of horizontal transfer of this gene in proteobacteria. *dnaJ*-based phylogenies were then analysed at three taxonomic levels: *Proteobacteria*, *Alphaproteobacteria* and the genus *Mesorhizobium*. Dendrograms based on DnaJ and 16S rRNA gene sequences revealed the same topology already described for *Proteobacteria*. These results indicate that the DnaJ phylogenetic signal is able to reproduce the accepted relationships among the five *Proteobacteria* classes. At a lower taxonomic level, using 20 alphaproteobacteria species, the 16S rRNA gene-based phylogeny is distinct from the one based on the DnaJ sequence analysis. Although the same clusters are generated, only the topology of the DnaJ tree is consistent with broader phylogenies from recent studies based on concatenated alignments of multiple core genes. For example, the DnaJ tree shows the two *Rhizobiales* clusters, closely related, as expected, while the 16S rRNA gene-based phylogeny shows them distantly related. In order to evaluate the phylogenetic performance of *dnaJ* at the genus level, a multilocus analysis based on five housekeeping genes (*atpD, gapA, gyrB, recA, rplB*) was performed for ten *Mesorhizobium* species. In contrast to the 16S rRNA gene, the DnaJ sequence analysis generated a tree similar to the multilocus dendrogram. For identification of chickpea mesorhizobium isolates, a *dnaJ* nucleotide-based tree was used. Despite different topologies, 16S rRNA gene- and *dnaJ*-based trees led to the same species identification. This study suggests that the *dnaJ* gene is a good phylogenetic marker, particularly for the *Alphaproteobacteria* class, since its phylogeny is consistent with phylogenies based on multilocus approaches.