P12: A robust electronic taxonomy of prokaryotes by MLSA: Lactic acid bacteria as a test case

Sabri Naser[1], Dirk Gevers[2,3], and Jean Swings[3]

[1]Department of Biology and Biotechnology, An-Najah National University, Nablus, Palestine
[2]Laboratory of Microbiology, Ghent University, K.L. Ledeganckstraat 35, Ghent 9000, Belgium
[3] Bioinformatics and Evolutionary Genomics, Ghent University/VIB, Technologiepark 927, Ghent 9052, Belgium

Introduction

Several methods have been used for the identification of lactic acid bacteria (LAB) to the species level, e.g. SDS-PAGE of whole-cell proteins, randomly amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), rep-PCR and ribotyping. Although useful, there are some pitfalls associated with the use of these methods concerning portability, inter-laboratory reproducibility and time efficacy. Informational genes such as the 16S rRNA gene are commonly considered as reliable phylogenetic markers for assigning evolutionary relationships among species of LAB. However, 16S rRNA gene sequence data do not allow the identification of closely related species. The use of housekeeping genes is emerging as an alternative to overcome these problems. *In silico* studies based on complete genomes have provided the basis for establishing sets of housekeeping genes that can accurately predict genome relatedness and improve the accuracy of species identification. The need for alternative genomic markers that provide higher levels of discrimination than the 16S rRNA gene has led to a more systematic sequencing of housekeeping genes (Naser *et al.*, 2007).

Multilocus sequence analysis (MLSA) is a polygenic approach applied to discriminate among different species of LAB. MLSA compares the primary DNA sequences from multiple conserved protein coding loci for assessing the diversity and relation of different isolates across related taxa, thereby using an appropriate phylogenetic or cladistic approach. To be useful for species discrimination, genes must ideally be present in a single copy, evolve more rapidly than rRNA genes and be widely distributed among bacterial genomes. Those genes in which recombination might confer a selective advantage, or closely linked genes, should be avoided. Furthermore, these genes should be informative with an adequate degree of resolution and provide sufficient variability to differentiate species of a particular genus. The aim of the present study was to evaluate the use of the genes encoding RNA polymerase alpha subunit (*rpoA*), phenylalanyl-tRNA synthase (*pheS*), and the alpha subunit of ATP synthase (*atpA*) as alternative species identification tools in a multilocus sequence analysis (MLSA) scheme**.**

Materials and methods

Four hundred and seventy five well-characterized strains belonging to LAB isolated from humans, animals or food products were analysed in this study. Bacterial genomic DNA was extracted. The amplification and sequencing of *pheS* and *rpoA* genes were performed as described by Naser *et al.* (2005 a,b). Consensus sequences were determined as described by Naser *et al.* (2005 a,b). The CLUSTAL_X program was used for multiple sequence alignment. Consequently, the aligned sequences were imported into BioNumerics software version 4.5 (Applied Maths) for the calculation of similarity matrices and neighbour-joining trees. The reliability of hierarchical clustering was determined by using the bootstrapping method with 1000 resamplings.

Results

MLSA data showed that *rpoA* (553 nt), *pheS* (455 nt), and *atpA* (1102 nt) partial gene sequences are much more discriminatory than the 16S rRNA gene sequence data. At the inter-species level, all representative species were clearly differentiated on the basis of *rpoA*, *pheS*,

and *atpA* gene sequences. Evaluation of the intraspecies variation showed that *rpoA*, *pheS,* and *atpA* genes had a high degree of homogeneity among strains of the same species.

Our data convincingly prove that the simultaneous analysis of *rpoA , pheS* and atpA partial gene sequences provide an alternative tool for the rapid and reliable identification of different species. The analysis of *rpoA , pheS* and atpA gene sequences effectively allows closely related species to be differentiated at a higher discrimination level than that possible with 16S rRNA gene sequence comparisons.

The variation in the discriminatory power of the investigated genes, together with the fact that different genes might provide different closest neighbours or tree topologies, has highlighted the necessity for simultaneous analysis of several protein-coding loci for a robust identification analysis.

MLSA is attractive in that it avoids the vagaries of a single-gene approach and provides an objective method to cluster strains within a genus. Moreover, it is a high-throughput methodology that is amenable to automation and, by using digital data, allows taxonomic assignment through the internet, a solution to the burden of routine species identification. Species definition and delineation requires a taxonomic framework and suitable software tools, which can be achieved if MLSA data are augmented, in central curated databases, with additional information relating to ecological adaptation, population structure, rates of recombination, amount of horizontal gene transfer or source of isolation that would arbitrate species assignment.

In the absence of a universal species concept, MLSA presents a highly valuable 'baseline' for creating and maintaining operational protocols for species assignment.

An MLSA framework moves prokaryotic taxonomy towards a rapid, flexible and objective standard with sufficient flexibility to accommodate the vast differences in biology presented by prokaryotes. Incorporation of ecological data into the MLSA framework will allow for meaningful taxonomic assignments, where by taxa are delineated by virtue of important ecological traits as well as degrees of genetic relatedness.

References

Naser, S. M., Thompson, F. L., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Thompson, C. C., Vancanneyt, M. & Swings, J. (2005a). Phylogeny and identification of enterococci using atpA gene sequence analysis. J Clin Microbiol 43, 2224-2230.

Naser, S. M., Thompson, F. L., Hoste, B., Gevers, D., Dawyndt, P., Vancanneyt, M. & Swings, J. (2005b). Application of multilocus sequence analysis (MLSA) for rapid identification of Enterococcus species based on rpoA and pheS genes. Microbiology 151, 2141-215

Naser, S. M., Dawyndt, P., Hoste, B., Gevers, D., Vandemeulebroecke, K., Cleenwerck, I., Vancanneyt, M. & Swings, J. (2007). Identification of lactobacilli by pheS and rpoA gene sequence analyses. Int J Syst Evol Microbiol. 57, 2777–2789.