

## ***Escherichia coli* molecular phylogeny using the Incongruence Length Difference test.**

LECOINTRE, G., RACHDI, L., DARLU, P. and DENAMUR, E. 1998.  
*Molecular Biology and Evolution*. 15 (12): 1685-1695.

Molecular phylogeny of the species *Escherichia coli* using the *E. coli* reference (ECOR) collection strains has been hampered by (1) the absence of rooting in the commonly used phenogram obtained from multilocus enzyme electrophoresis (MLEE) data and (2) the existence of recombination events between strains that scramble phylogenetic trees reconstructed from the nucleotide sequences of genes. We attempted to determine the phylogeny for *E. coli* based on the ECOR strain data by extracting from GenBank the nucleotide sequences of 11 chromosomal structural and 2 plasmid genes for which the *Salmonella enterica* homologous gene sequences were available. For each of the 13 DNA data sets studied, incongruence with a nonnucleotide whole-genome data set including MLEE, random amplified polymorphic DNA, and *rnn* restriction fragment length polymorphism data was measured using the incongruence length difference (ILD) test of Farris et al. As previously reported, the incongruence observed between the *gnd* and plasmid gene data and the whole-genome data was multiple, indicating numerous horizontal transfer and/or recombination events. In five cases, the incongruence detected by the ILD test was punctual, and the donor group was identified. Congruence was not rejected for the remaining data sets. The strains responsible for incongruences with the whole-genome data set were removed, leading to a "prior-agreement" approach, i.e., the determination of a phylogeny for *E. coli* based on several genes, excluding (1) the genes with multiple incongruences with the whole genome data, (2) the strains responsible for punctual incongruences, and (3) the genes incongruent with each other. The obtained phylogeny shows that the most basal group of *E. coli* strains is the B2 group rather than the A. group, as generally thought. The D group then emerges as the sister group of the rest. Finally, the A and B1 groups are sister groups. Interestingly, the most primitive taxon within *E. coli* in terms of branching pattern, i.e., the B2 group, includes highly virulent extraintestinal strains with derived characters (extraintestinal virulence determinants) occurring on its own branch.