Systematic Evaluation and Classification of Bacterial Genes in INSDC

Yoshio Tateno

Center for Information Biology and DNA Data Bank of Japan National Institute of Genetics Yata, Mishima 411-8540, Japan

An ever-increasing number of DNA data submissions to the International Nucleotide Sequence Database Collaboration (INSDC, http://insdc.org) have made a profound impact and contribution not only to the research community of life sciences but those of medicine, pharmacology, agriculture and others. However, the majority of the submissions contain genes that are not examined or confirmed *in vivo/vitro*, but inferred by homology search or the like *in silico*. Therefore, a great number of submitted genes have been described as being "hypothetical" or "homologous to a particular gene". Another aspect with this problem is that the *in silico* inferences were made by using various computer tools with various parameters. This aspect makes the problem even worse. To mitigate the problem we systematically evaluated all bacterial genes in INSDC *in silico* by the same tool and parameters, and classified them into the degree of reliability. Our approach and results of the evaluation and classification will be reported.