Comparison of the effectiveness of haplotype clustering methods in intra-species variability studies

M.Attimonelli1, C.Marangi2, M.Derobertis1

1 Dipartiemento di Biochimica e Biologia Molecolare, Bari
2 Dipartimento Interateneo di Fisica, Bari

Biodiversity is one of the most popular research topics in these last decade. Indeed, the understanding of mechanisms causing diversity and the implementation of computer methods allowing the analysis of available data on bio-diversity can greatly contribute in biomedical studies as well as in anthropological studies and in agricultural applications. Biodiversity studies are all based on the observation of characters that can be classified as morphological, genetic and molecular. The observation of these data with different approaches produces discrete sets of values (quantitative and qualitative) that allow the definition of haplotypes and haplogroups. The definition of haplotypes and of haplogroups is performed through approaches based on different assumptions. The methods available for the calculation of distance are many and all based more or less on approximations. Hence, more rigorous clustering analysis methods are needed; these analyses allow to estimate distances between haplotypes in order to determine as accurately as possible the clusters, that is the haplogroups. Here we present a strategy where starting from a set of raw data it is possible to define with a high percentage of reliability the haplogroups by applying different clustering methods and comparing the results.