A mathematical duality theory and its application in forensic genetics

Hans-Jürgen Bandelt (University of Hamburg)

Isbell (1980) introduced isotropic (ternary) algebras and Mulder (1980) was the first to define quasi-median graphs (and algebras), which determine the same mathematical structure in the finite case. Ploščica (1992) then described a "natural" duality between these algebras and sets endowed with the discrete topology, with a single binary relation, a single unary relation, a constant and permutations (which together directly express gated decomposition of the algebras). There is a more parsimonious way to describe the algebra by considering some generating subset of the algebra, which dually gives rise to a data table of finite sequences over a finite alphabet. The binary relation of strong compatibility between positions of these sequences then reflects the crucial relational structure of the dual.

In this light, data tables over the DNA alphabet {A, G, C, T, —} can be viewed as the dual of a quasi-median algebra. A natural evolutionary process acting on a uniparental marker (mtDNA, or non-recombining part of the Y chromosome) produces a geneological tree connecting DNA sequences, which would be retrieved as the dual of the corresponding data table if the mutations always hit different DNA positions. Normally, however, some amount of recurrent mutations somewhat blur the genealogical signal. The strong compatibility relation is then highly sensitive to parallelisms and reversals at DNA positions. This relation helps distinguishing artificial mutational patterns caused by experimental and documentation errors from natural evolutionary patterns, when one focusses on DNA positions with known low mutation rates. Consequently, the graphical display enables curators of databases to pinpoint likely errors and thus to quickly weed out flawed datasets. Exactly this is being done routinely in the case of the worldwide forensic mtDNA database EMPOP.