On Internal Classification of the Tsezic Languages

A lexicostatistical classification is proposed for 9 languages and dialects of the Tsezic group of the North Caucasian family, based on meticulously compiled 110-item wordlists, published on-line as part of the *Global Lexicostatistical Database* project. The list of lects includes: Hunzib proper, Bezhta proper, Khoshar-Khota Bezhta, Tlyadal Bezhta, Hinukh, Kidero Dido, Sagada Dido, Khwarshi proper, Inkhokwari Khwarshi. The lexical data have been subsequently analyzed with the aid of the principal phylogenetic methods, both distance-based and character-based: Starling neighbor joining (StarlingNJ), Neighbor joining (NJ), Unweighted pair group method with arithmetic mean (UPGMA), Markov chain Monte Carlo (MCMC), Unweighted maximum parsimony (UMP). All the phylogenetic methods have yielded trees with two main branches: Eastern Tsezic (Hunzib & Bezhta) and Western Tsezic (Hinukh, Dido & Khwarshi) that agrees with the traditional classification. The methods, however, contradict each other in topology of the Hinukh-Dido-Khwarshi branch, where the first splitting off node is either Hinukh or Khwarshi. Such an incompatibility of the obtained trees is due to a number of hidden interdialectal borrowings and secondary shared innovations (contact driven homoplasy).

Phonological and morphological differences unambiguously support the group's division into Eastern and Western branches. The position of Hinukh has been debated, the researchers' proposals varying from 'a dialect of Dido' to 'a separate branch'. It will be shown that the characteristics that seem common Dido and Hinukh innovations – like the front vowels switch (Proto-Tsezic *i > e, *e > i), or Proto-Tsezic *l > r, or complete loss of nasalization, might have been independent developments and thus cannot challenge the accuracy of the lexicostatistical tree.