

Genotype-phenotype correlations using phylogenetic trees for large datasets

Farhat Habib, Andrew Johnson, Ralf Bundschuh, and Daniel Janies

Phylogenetic trees serve as hierarchical summaries of character change (often genomic and/or phenotypic) between ancestors and descendants. The relationship between the tree and individual character data that it is based on is commonly visualized by optimizing the character data such as sequence polymorphisms or phenotypic features onto the phylogenetic tree. While graphically interesting, the major drawback shared by visualization programs is that there are often few or no implemented means of analyzing the whole set of character change data to find emergent patterns. Here we address this gap and present a method to query whole genomes or regions within genomes to find emergent patterns of genomic changes highly correlated with a given binary phenotype using an extension of Maddison's concentrated changes test (CCT). We implement our method using the TNT scripting language and a case study is presented for the susceptibility to *Bacillus anthracis* in various strains of inbred mice. We present results that indicate the approach is useful in finding known candidate genes for this trait in mice and also suggests targets for further potential investigation.