Estimating Estimators: An Empirical Comparison of Non-Invasive Analysis Methods

Authors : Yan Torres, Fernanda Simoes, Francisco Petrucci-Fonseca, Freddie-Jeanne Richard

Abstract : The non-invasive samples are an alternative of collecting genetic samples directly. Non-invasive samples are collected without the manipulation of the animal (e.g., scats, feathers and hairs). Nevertheless, the use of non-invasive samples has some limitations. The main issue is degraded DNA, leading to poorer extraction efficiency and genotyping. Those errors delayed for some years a widespread use of non-invasive genetic information. Possibilities to limit genotyping errors can be done using analysis methods that can assimilate the errors and singularities of non-invasive samples. Genotype matching and population estimation algorithms can be highlighted as important analysis tools that have been adapted to deal with those errors. Although, this recent development of analysis methods there is still a lack of empirical performance comparison of them. A comparison of methods with dataset different in size and structure can be useful for future studies since non-invasive samples are a powerful tool for getting information specially for endangered and rare populations. To compare the analysis methods, four different datasets used were obtained from the Dryad digital repository were used. Three different matching algorithms (Cervus, Colony and Error Tolerant Likelihood Matching - ETLM) are used for matching genotypes and two different ones for population estimation (Capwire and BayesN). The three matching algorithms showed different patterns of results. The ETLM produced less number of unique individuals and recaptures. A similarity in the matched genotypes between Colony and Cervus was observed. That is not a surprise since the similarity between those methods on the likelihood pairwise and clustering algorithms. The matching of ETLM showed almost no similarity with the genotypes that were matched with the other methods. The different cluster algorithm system and error model of ETLM seems to lead to a more criterious selection, although the processing time and interface friendly of ETLM were the worst between the compared methods. The population estimators performed differently regarding the datasets. There was a consensus between the different estimators only for the one dataset. The BayesN showed higher and lower estimations when compared with Capwire. The BayesN does not consider the total number of recaptures like Capwire only the recapture events. So, this makes the estimator sensitive to data heterogeneity. Heterogeneity in the sense means different capture rates between individuals. In those examples, the tolerance for homogeneity seems to be crucial for BayesN work properly. Both methods are user-friendly and have reasonable processing time. An amplified analysis with simulated genotype data can clarify the sensibility of the algorithms. The present comparison of the matching methods indicates that Colony seems to be more appropriated for general use considering a time/interface/robustness balance. The heterogeneity of the recaptures affected strongly the BayesN estimations, leading to over and underestimations population numbers. Capwire is then advisable to general use since it performs better in a wide range of situations.

Keywords : algorithms, genetics, matching, population