

# Evaluation of the accuracy of the ML-relate kinship analysis program when no parents are sampled

Letícia F. Lima, Renata Schama

*Laboratório de Biologia Computacional e Sistemas, Oswaldo Cruz Institute- Fiocruz*

Knowledge of the genealogical relationships among different family groups can provide important information for the study of evolution and behavioral ecology. In parentage analysis of natural populations, numerous methods have been developed for inferring the relationship among individuals aiding conservation management in zoos and wildlife reserves. In most cases the genotypes of both offspring and potential parents are present in the sample and this data makes it much easier to assign the offspring to parents. This kind of method may also be used to better understand oviposition and migration behavior in insects. Information such as these can be of assistance for new vector control strategies. Nevertheless, in insects, most of the time the collection of samples does not involve collecting parent and offspring together; which makes the inference of kinship relationship less accurate. Also, polygamy, another characteristic of many insects, increases the difficulty of the analysis. The presence of half siblings in the sample makes it difficult to understand the family groups. In this study we use simulation data created using real data parameters (number of *loci*, alleles and genetic diversity) to test the accuracy of the program ML-relate to detect full siblings. The ML-relate program is widely used in parentage analysis and uses a maximum likelihood framework to find the most likely relationship between two individuals. To simulate the data, with the COLONY program, we used the genetic structure and diversity of the mosquito *Aedes aegypti* population from Rio de Janeiro. We simulated data containing different percentages of half-siblings (10%, 20%, 30%, 40% and 50%) to assess how the different scenarios would affect the program's classification of kinship relationships. We evaluated the error rates and accuracy of the program under the different scenarios. We found type I error to be low in all scenarios (around 0.021%). Type II errors fluctuated and were much higher (around 0.51%), indicating that for natural data a large number of full siblings would not be categorized as such. Nevertheless, accuracy was 76%, well within the range found for other programs and organisms analyzed. Since type I error was low and type II high we conclude that although the estimation of family groups will be smaller than the real ones we can be sure of the full sibling relationships estimated by the program and therefore of the inferred oviposition behavior. Nevertheless, these results will impact conclusions regarding the extent of migration by females, underestimating it.