

## **IBS.IBD.studies: an R package for relatedness research using microsatellites and SNP data**

*Galván I.<sup>1</sup>, Graffelman J.<sup>2</sup>*

<sup>1</sup> ivan.galvan@udg.edu, Department of Computer Science, Applied Mathematics and Statistics, Universitat de Girona;

<sup>2</sup> jan.graffelman@upc.edu, Department of Statistics and Operations Research, Universitat Politècnica de Catalunya.

Relatedness is of great interest in population-based genetic association studies (Foulkes, 2009). These studies search for genetic factors related to disease. Many statistical methods used in association studies (such as standard regression models, t-tests, logistic regression) assume that the observations (individuals) are independent. These techniques can fail if independence is not satisfied. Thus, it is important to document the degree of relatedness of a pair of individuals in the database. We present **IBS.IBD.studies**, an R package for the study of relatedness based on genetic markers such as microsatellites and single nucleotide polymorphisms. This package is focused on allele sharing statistics of pairs of individuals from a given human population over a set of genetic markers. The possible number of alleles shared for each genetic marker (0, 1 or 2) can be considered identical by state (IBS) or identical by descent (IBD).

The IBS.IBD.studies package offers some graphical methods to detect family relationships by considering the proportions of sharing 0, 1 or 2 IBS alleles. On the other hand, the probabilities of sharing 0, 1 or 2 IBD alleles, which are termed Cotterman's coefficients (Cotterman, 1941), depend on the relatedness (monozygotic twins, full-siblings, parent-offspring, avuncular, first cousins, etc) and can be considered to accurately identify the type of family relationship. With the IBS.IBD.studies package, it is possible to infer relatedness by using maximum likelihood methods (Milligan, 2003; Thompson, 1991).

**Keywords:** Relatedness, allele sharing, identical by state, identical by descent, genetic marker.

### **References**

- Cotterman C.W. (1941). Relative and human genetic analysis. *Sci Monthly* 53, pp. 227-234.
- Galván, I., Graffelman J. and Barceló C. (2014). *Allele-sharing and Relationship Inference*. Master thesis, Universitat Politècnica de Catalunya (UPC).
- Foulkes, A.S. (2009). *Applied Statistical Genetics with R*. Springer.
- Milligan B.G. (2003). Maximum-likelihood estimation of relatedness. *Genetics*, 163 pp. 1153-67.
- Thompson E.A. (1991). Estimation of relationships from genetic data. In: Rao CR, Chakraborty R (Eds.) *Handbook of Statistics*. Elsevier Science: Amsterdam, Vol 8, pp. 255-269.