

OneMap: software for genetic mapping in outcrossing species

G. R. A. MARGARIDO¹, A. P. SOUZA² and A. A. F. GARCIA¹

¹*Department of Genetics, Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo (USP), Piracicaba, São Paulo, Brazil*

²*Centro de Biologia Molecular e Engenharia Genética (CBMEG), Universidade Estadual de Campinas (UNICAMP), Campinas, São Paulo, Brazil*

Margarido, G. R. A., Souza, A. P. and Garcia, A. A. F. 2007. OneMap: software for genetic mapping in outcrossing species. — *Hereditas* 144: 78–79. Lund, Sweden. eISSN 1601-5223. Received March 12, 2007. Accepted April 16, 2007

OneMap is an environment for constructing linkage maps of outcrossing plant species, using full-sib families derived from two outbred parents. The analyses are performed using a novel methodology based on the maximum likelihood approach for simultaneous estimation of linkage and linkage phases (WU et al. 2002), which has been successfully applied to sugarcane (GARCIA et al. 2006). It is implemented as a set of functions for the freely distributed software R, and handles pairwise marker analysis, marker ordering and map refinement. The software is freely available at <http://www.ciagri.usp.br/~aafgarci/OneMap/>.

Antonio Augusto Franco Garcia, Escola Superior de Agricultura “Luiz de Queiroz”, Departamento de Genética, Caixa Postal 83, 13400-970, Piracicaba, SP, Brazil. E-mail: aafgarci@esalq.usp.br

Genetic maps are of great use to breeding programs, because they allow the localization of QTL's, i.e. genome regions containing loci affecting a quantitative trait (DOERGE 2002). For many species, due to their allogamous nature, it is difficult to obtain homozygous inbred lines, and therefore the traditional approaches used to construct linkage maps for F₂ and backcross populations cannot be used (GRATTAPAGLIA and SEDEROFF 1994).

Traditionally, the construction of genetic maps in F₁ populations has been achieved using the double “pseudo-testcross” strategy (GRATTAPAGLIA and SEDEROFF 1994), in which markers segregating in a 1:1 ratio (testcross configuration) are analyzed with traditional statistical approaches and softwares, such as MapMaker/EXP (LANDER et al. 1987). However, these are not suitable for markers with different segregation patterns and have the disadvantage of resulting in two separate linkage maps, one for each parent. Presently, the most widely available software for genetic mapping in F₁ populations is JoinMap (STAM 1993). It handles the analysis of markers showing different segregation patterns in both parents. However, the methods used for marker ordering in linkage groups and recombination fraction estimation are based on minimum squares approximations for multipoint estimates of distances, and not on the maximum likelihood approach, which has several advantages (WU et al. 2002). Recently, GARCIA et al. (2006) compared the results of analysis from JoinMap and using Wu's methodology for sugarcane mapping, and found better results for the latter. Actually, it was the sugarcane mapping that led us to the development of OneMap.

OneMap software was developed to facilitate the linkage analyses in outcrossing species, using the methodology proposed by WU et al. (2002). It is implemented as a set of functions for the freely distributed software R (IHAKA and GENTLEMAN 1996). Algorithms were coded in C language, while the data manipulation functions were coded directly in the R language. It is currently available as source code for Windows and Unix.

LINKAGE AND LINKAGE PHASE ESTIMATION

The core part of the software is the simultaneous maximum likelihood estimation of linkage and linkage phases between markers, according to WU et al. (2002). This methodology allows the analysis of a mixed set of different marker types containing various segregation patterns, such as 1:1:1:1, 1:2:1, 3:1 and 1:1. OneMap is capable of handling all possible cross types and, therefore, any data set can be analyzed, besides the more complex F₁ populations derived from two outbred parent plants.

To our knowledge, no other software has been implemented to handle these analyses. Thus, although being very promising, this method has currently been underused by geneticists.

CURRENT FEATURES

The first version of OneMap covers various steps from genotype data reading to genetic map refinement. There are functions to perform the basic two-point

analysis between markers via the EM algorithm (DEMPSTER et al. 1977) and, based on its results, assign markers to different linkage groups.

Next, marker ordering in linkage groups is done using the Rapid Chain Delineation algorithm (DOERGE 1996), which is not computationally intensive and only requires a matrix with recombination fractions between markers. Finally, the more informative multipoint (three-point) analysis may be used to check the local order of markers in a mapped linkage group and also to refine the map distance between adjacent markers.

There are also some data manipulation functions that make OneMap usage more flexible. The criteria used in all steps by the software are completely user-defined and it is assumed that the user has some experience with genetic mapping and the R language. Nonetheless, there is also a tutorial available, intended to get users started with OneMap. It comes with a hypothetical simulated data set, used as an example to illustrate the usage of OneMap functions.

PERSPECTIVES

OneMap is free software and can be changed and redistributed by any user under the regulations of the GNU General Public License (FREE SOFTWARE FOUNDATION 1991). Furthermore, it will be continuously updated and improved. Currently, work is being conducted in order to make available another ordering algorithm, called Seriation (BUETOW and CHAKRAVARTI 1987). Besides that, the refinement algorithm Ripple (LANDER et al. 1987) will also be implemented and released.

When the mapped markers are subjected to three-point analysis, two estimates of the recombination fraction for the same interval are obtained. These estimates can be combined using a weighted mean, with the weights being the reciprocals of the variances of the two separate estimates (RIDOUT et al. 1998, WU et al. 2002). This feature is already under development and is expected to be available soon.

In the future, we expect to release OneMap as an add-on package to R, including help files, new

functions for graphic representation of the results and more example data sets. In this sense, critics and suggestions are welcome.

REFERENCES

- Buetow, K. H. and Chakravarti, A. 1987. Multipoint gene mapping using seriation. I. General methods. – *Am. J. Human Genet.* 41: 180–188.
- Dempster, A. P., Laird, N. M. and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. – *J. R. Statist. Soc. B* 39: 1–38.
- Doerge, R. W. 1996. Constructing genetic maps by rapid chain delineation. – *J. Quantitative Trait Loci* 2, article 6.
- Doerge, R. W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. – *Nat. Rev. Genet.* 3: 43–52.
- Garcia, A. A. F., Kido, E. A., Meza, A. N. et al. 2006. Development of an integrated genetic map of a sugarcane (*Saccharum* spp.) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. – *Theor. Appl. Genet.* 112: 298–314.
- GNU General Public License. 1991. – Free Software Foundation, Inc., Boston, MA.
- Grattapaglia, D. and Sederoff, R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross mapping strategy and RAPD markers. – *Genetics* 37: 1121–1137.
- Ihaka, R. and Gentleman, R. 1996. R: a language for data analysis and graphics. – *J. Comput. Graph. Statist.* 5: 299–314.
- Lander, E. S., Green, P., Abrahamson, J. et al. 1987. MAPMAKER: an interactive computing package for constructing primary genetic linkages of experimental and natural populations. – *Genomics* 1: 174–181.
- Ridout, M. S., Tong, S., Vowden, C. J. et al. 1998. Three-point linkage analysis in crosses of allogamous plant species. – *Genet. Res.* 72: 111–121.
- Stam, P. 1993. Construction of integrated genetic linkage maps by mean of a new computer package: JoinMap. – *Plant J.* 3: 739–774.
- Wu, R., Ma, C. X., Painter, I. et al. 2002. Simultaneous maximum likelihood estimation of linkage and linkage phases in outcrossing species. – *Theor. Popul. Biol.* 61: 349–363.