MareyMap: a R-based tool with graphical interface for estimating recombination rates

Clément Rezvoy, Delphine Charif, Laurent Guéguen, Gabriel A.B. Marais
Université de Lyon; Université Lyon 1; Centre National de la Recherche Scientifique; UMR5558; Laboratoire de Biométrie et Biologie évolutive; Villeurbanne, F-69622 cedex, France

ABSTRACT
Summary: comparing genetic and physical maps (the so-called Marey map approach) is still the most widely used approach to estimate genome-wide recombination rates. Remarkably, there is no available bioinformatics tool specifically devoted to Marey map approach. Here we developed such a tool called MareyMap based on GNU R and Tk/Tk. MareyMap offers a user-friendly graphical interface and includes useful features such as data cleaning process, sophisticated interpolation methods to estimate local rates, possibility of complex queries, various range of import – export files. Moreover, MareyMap comes with ready-to-use maps for human, Drosophilia, C. elegans and Arabidopsis. MareyMap has been made so that it can be easily upgraded with new data and interpolation methods.

Availability: http://pbil.univ-lyon1.fr/software/mareymap/
Contact: rezvoy@biomserv.univ-lyon1.fr, marais@biomserv.univ-lyon1.fr

1 INTRODUCTION
Recent advances in genomics have shown that genome organization is influenced by meiotic recombination rates in Eukaryotes (see for instance Carvalho & Clark 1999, Marais et al. 2001, Bartolome et al 2002, Wright et al. 2003, Pal & Hurst 2003, Presgraves 2005). The most striking example of this is probably the non-recombining Y chromosome that is very different (i.e. chromosome size, gene content) from recombining X chromosome although these two chromosomes have the same autosomal ancestry and were identical at the early stage of their evolution (see Charlesworth et al. 2005 for review). More generally, non-recombining regions of the genome have a different structure (i.e. gene density, GC content, evolutionary rates) compared to recombining ones.
In the vast majority of the genome-analysis papers cited above, estimates of recombination rate have been obtained using Marey map approach. The Marey map approach has been proposed by A. Chakravarti (1991). Marey was a French scientist who popularized in his book “La méthode graphique” a method developed by a rail engineer called Iby. This method compared the distance in Kilometers between two stations with the time needed to travel by train from one station to another with the idea of improving global train circulation between Paris and Lyon (Marey 1885). Chakravarti used this approach to compare physical and genetic maps. Historically, genetic maps were the first maps for chromosomes. They stem on the fact that recombination frequency between two genes increases with the physical distance between these two genes. But genetic maps can be used for another purpose: if one fits a curve to the plot of genetic versus physical distances for a given chromosome then the slope of that curve will give the local recombination rates along that particular chromosome.
Although other approaches can in principles give genome-wide estimates for recombination rate (see for instance Stumpf & McVean 2003, Myers et al. 2005), Marey map is considered the more convenient approach to get such estimates (Gaut et al. 2007). Indeed, Marey map is fast, unbiased and genetic and physical maps are available for a relatively large number of organisms. Intriguingly, there are almost no tools to automatically deal with the comparison of genetic and physical maps. As far as we know, there is only one tool comparing physical and genetic maps (Cartographer, Voigt et al. 2004) and this tool has not been made to estimate recombination rates. It is actually a converter between the two types of distances (genetic and physical) used in biomedical research (Voigt et al. 2004).

2 IMPLEMENTATION
Our work aimed at developing a tool to estimate recombination rates using the Marey map approach. We wanted that tool - we called it MareyMap - to be specifically devoted to estimating recombination rates (not as Cartographer, see above), to have a user-friendly graphical interface (because Marey Map is a graphical method), to have sophisticated interpolation methods (to get slopes), to allow complex queries (to retrieve estimates at particular positions) and to have a good upgradeability (regarding data and interpolation methods). We chose GNU R to implement MareyMap because R includes many statistical functions that we needed for our interpolation methods (e.g. loess, cubic splines). We used Tk/Tk to build our graphical interface along with the Rtkplot library. However, MareyMap can also work without interface using R command lines only, which can be useful for advanced use (in a pipeline for instance). MareyMap is formed by several packages: MareyMap (work on Marey maps), MareyMapGUI (graphical interface) and one package for each interpolation method. A screen shot of MareyMap graphical interface is shown in figure 1.

*To whom correspondence should be addressed.

© The Author (2007). Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
3 DESCRIPTION

MareyMap has several advantages. It comes with physical and genetic maps from different organisms: one vertebrate (Homo sapiens), 3 sets of maps: male, female and sex average, from Kong et al. 2002, two invertebrates (Drosophila melanogaster, from Flybase at http://flybase.bio.indiana.edu/ see Marais et al. 2001, Caenorhabditis elegans, from Wormbase at http://www.wormbase.org/ see Marais et al. 2001) and one plant (Arabidopsis thaliana, from NASC at http://arabidopsis.info/ see Wright et al. 2003). Importantly, the user can include other sets of maps from the same or other species very easily. This can be done by uploading a text file with species name, chromosomes, marker names, physical position and genetic distances (see http://pbil.univ-lyon1.fr/software/mareymap/ for more details). MareyMap includes a (facultative) data cleaning process. Maps may contain errors (discrepancies between genetic and physical maps). In this case, it is possible to exclude some chosen markers temporarily or definitively. One strong advantage of MareyMap is the choice of different interpolation methods. It includes sliding window, which is the simplest and most widely used method. The idea is just sliding a window along a chromosome and getting the local estimate with the slope of the best line fit to the data in the local window. Parameters are window size and shift. We also offer more sophisticated methods: Loess, which is a sophisticated sliding window (a two-degree polynomial is fitted in each window) and cubic splines, which is probably the best method to estimate recombination rate with Marey map approach (see Yu et al. 2001, Berloff et al. 2002). More methods can be included in MareyMap by the user. Different estimators can be compared visually (in recombination rates plot) and to the original data (Marey map plot). Recombination rates for some positions in the genome can be recovered using a query window and it is also possible to upload files of preselected positions, which is very useful for large-scale genome analysis. Possible file import — export include export of pdf for plots, import of new sets of maps (see above) and export of recombination rates estimates.

REFERENCES


