Experiments in an algorithmic approach to comparative radiation hybrid mapping

Simon de Givry*, Thomas Faraut and Thomas Schiex***

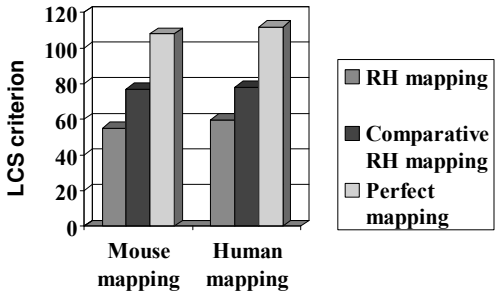
*INRA Unité de Biométrie et Intelligence Artificielle, **INRA Laboratoire de Génétique Cellulaire, Toulouse, France.

Genome mapping, either by genetic or physical mapping, including radiation hybrid (RH) mapping is a fundamental step in the study of an organism and often a prerequisite for genome sequencing. The aim of this work was to improve the RH mapping process by integrating the knowledge of closely related organisms already sequenced. The rationale for this is that closely related organisms differ by a small number of chromosomal rearrangements in comparison to the number of genes they have in common.

Materials and Methods

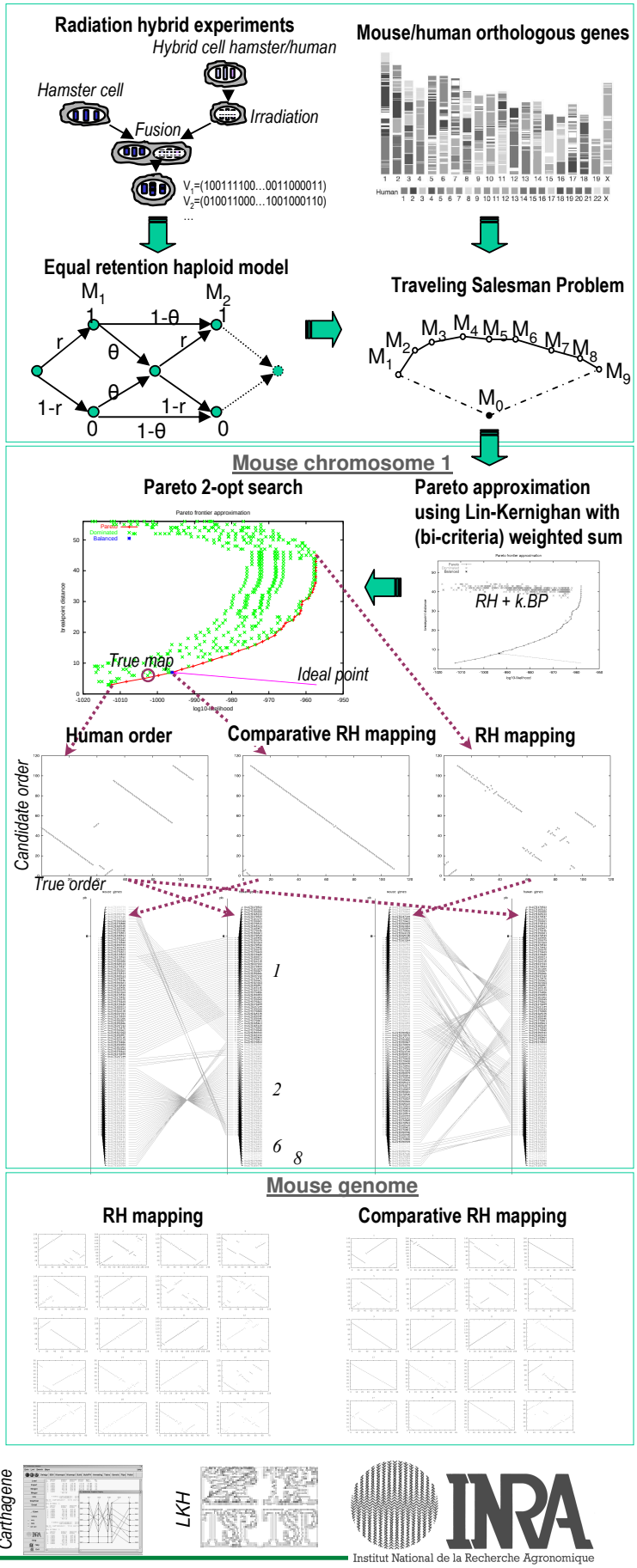
Our approach consists in identifying the orthologous genes/markers between two species (see *e.g.* NCBI Homologene) and then, searching for a map which maximizes the likelihood of the RH data while minimizing the breakpoint distance between the map order and the order of a sequenced genome. This allows to reduce the problem to a *bi-objective traveling salesman problem*. Using a weighted sum to aggregate the two objectives results in a mono-objective problem efficiently solved by Concorde* or LKH*. By varying the weight, the best maps w.r.t. the likelihood of the RH data are collected for all the possible values of the breakpoint criterion. This collection of maps represent an approximation of the so-called *Pareto frontier* (a set of non dominated solutions). It is further improved by exploring the 2-*Opt* neighborhood of every map. We offer ad-hoc and generic decision support tools in order to select the final map, including graphical representations and automatic selection by minimization of a predefined distance function.

Results



Experimental results involving mouse and human genomes showed a 40% improvement in map quality in terms of the longest common subsequence (LCS) criterion for the comparison of candidate maps with the known "correct" map. Our comparative mapping approach is currently integrated into CarthaGene, a multi-population integrated genetics and RH mapping software (http://www.inra.fr/bia/T/CarthaGene/).

We have proposed an improved RH mapping strategy that takes advantage of closely related completely sequenced genomes. Future work will take into account not only the order between markers but also their relative distance in base pairs.



Carthagene

LKH

INRA
Institut National de la Recherche Agronomique

*LKH is copyright Keld Helsgaun, *Concorde is copyright David Applegate *et al.*