ActionMap: a web-based software that automates loci assignments to framework maps

G. Albini, Matthieu Falque, Johann Joets

ActionMap: a web-based software that automates loci assignments to framework maps

Guillaume Albini, Matthieu Falque and Johann Joets*

UMR de Génétique Végétale du Moulon, Ferme du Moulon, 91190 Gif-sur-Yvette, France

ABSTRACT

Genetic linkage computation may be a repetitive and time consuming task, especially when numerous loci are assigned to a framework map. We thus developed ActionMap, a web-based software that automates genetic mapping on a fixed framework map without adding the new markers to the map. Using this tool, hundreds of loci may be automatically assigned to the framework in a single process. ActionMap was initially developed to map numerous ESTs with a small plant mapping population and is limited to inbred lines and backcrosses. ActionMap is highly configurable and consists of Perl and PHP scripts that automate command steps for the MapMaker program. A set of web forms were designed for data import and mapping settings. Results of automatic mapping can be displayed as tables or drawings of maps and may be exported. The user may create personal access-restricted projects to store raw data, settings and mapping results. All data may be edited, updated or deleted. ActionMap may be used either online or downloaded for free (http://moulon.inra.fr/~bioinfo/).

INTRODUCTION

Identification of genes involved in plant traits is a first step to understanding biological processes underlying these traits. Transcriptome and proteome studies, functional annotation and ultimately co-localization of a gene with a QTL must be taken into account to track candidate genes down. As only two complete plant genome sequences are available, large efforts are still required to genetically map genes in numerous other plant species. ESTs databases are a major source for new sequences to map. The throughput of genetic mapping does not allow systematic mapping of all these ESTs, but hundreds of cDNAs can still be mapped within a year by a single laboratory. For such projects, data storage and mapping computation need to be optimized. Automation of genetic linkage map computation may increase the throughput as well as the reliability of large-scale mapping projects.

Genetic linkage map computation may be achieved by using different softwares such as MapMaker (1), JoinMap (2) or GMendel (3). MapMaker is a command line software widely-used in plant genetics. It is well suited for automation as command lines may be read directly from text files that have been written by the user or by other software. However, building a genetic map *de novo* requires too much human expertise to be easy to automate. Expert systems developed for automated construction of human genetic linkage maps remain to be improved (4), and cannot be applied to the genetic structure of usual plant mapping populations. Nevertheless, if loci are assigned to a fixed framework map, MapMaker command steps required to assign new markers to the framework are similar for each marker. We thus developed ActionMap; a software that automates the assignment of markers to a fixed framework map without adding the new markers to the map. This choice was driven by the requirements of projects involving high numbers of loci (thousands) mapped on a small number of segregating plants (less than one hundred). In such projects, three point orders between closely linked loci are not statistically significant, so it is better not to include all the markers to the map.

ActionMap performs mapping step by step by generating command line files for MapMaker. Outputs from MapMaker are captured and analyzed in order to generate the next step command line file and so on, until mapping is completed. For quality control purposes, it is critical to store all the intermediate data. Therefore, a database was integrated into our software to save the automatic mapping settings, the intermediate data and the results as well as the raw segregation data. This database can then be considered as a mapping data management system. As pointed out by Li *et al.* (5), such a database must manage the fact that segregation data may change over time (repeated experiments) and are not 100% accurate. Therefore our software provides functions for easy importation, deletion and also edition or update of the data. Reliability of the data is also an important feature to address. For instance, consistency and formats of data as well as their compatibility with MapMaker are checked when data files are transferred to ActionMap.

IMPLEMENTATION

ActionMap is packaged as a web server, which may be used online or downloaded for free (http://moulon.inra.fr/~bioinfo/).

---

*To whom correspondence should be addressed. Tel: +33 1 69 33 23 78; Fax: +33 1 69 33 23 40; Email: joets@moulon.inra.fr

Validation	Common informations				Error informations	Position informations				Intervals informations
Check	Status	Locus Name	Linkage Group	Lod Score	Error Name	Relative Position	Proximal Bin Name	Distal Bin Name	Absolute Position	Confidence Interval
Ok : ☐	**YES**	LN1	6	7.1	Assign: Lod = 3 and Dist = 30 Unique Distorsion	84.986595174263	6.05		19.307238605898	npi280 0.00 npi561 npi561 -0.09 End
Ok : ☐	**YES**	LN3	1	9.4	Assign: Lod = 3 and Dist = 30 Unique Distorsion	7.2254335260116	1	1.01	2.5505780346821	Begin -0.02 bnl5.62a bnl5.62a 0.00 npi406
Ok : ☐	**YES**	LN4	1	7.1	Assign: Lod = 3 and Dist = 30 Unique Distorsion	Extremite		1	-4.1	Begin 0.00 bnl5.62a bnl5.62a -0.22 npi406
Ok : ☐	**YES**	LN5	1	8.2	Assign: Lod = 3 and Dist = 30 Unique Distorsion	13.310580204778	1	1.01	4.6986348122867	Begin -0.44 bnl5.62a bnl5.62a 0.00 npi406
Ok : ☐	**YES**	LN6	1	4.2	Assign: Lod = 3 and Dist = 30	43.478260869565	1	1.01	15.347826086957	Begin -2.43 bnl5.62a bnl5.62a 0.00 npi406
Ok : ☐	NO	LN7	1	2.4	Unassigned Lod Diminution Assign: Lod = 2 and Dist = 30 Unique Distorsion Try not Ok Several 0.00					Begin 0.00 bnl5.62a bnl5.62a 0.00 npi406 npi406 2.98 npi439a npi439a 1.83 dpg11 dpg11 2.82 bnl7.21a
Ok : ☐	**YES**	LN8	1	3.8	Assign: Lod = 3 and Dist = 30	53.14465408805	1	1.01	18.760062893082	Begin -2.44 bnl5.62a bnl5.62a 0.00 npi406 npi406 -2.82 npi439a

A

Parameters	Types	Descriptions
Position Information	Boolean	Visualisation of position data of mapped loci
Position within the adjacent interval	Boolean	Relative position of locus within interval
Name of the bin	Boolean	Interval name
Absolute Position in cM	Boolean	
Data Information	Boolean	Visualisation of data about mapped locides
Number of alleles of each parent	Boolean	
Segregation Distorsion Test	Boolean	
Segregation Distorsion threshold	%	
Number of Missing Data	Boolean	
Missing Data Warning	Boolean	
Threshold for Missing Data Warning	%	
Chromosome Information	Boolean	Visualisation of data about linkage groups
Chromosome length before and after mapping process	Boolean	
Warning for difference between the two lengths	Boolean	.
Threshold for Length Warning	%	
Interval Information	Boolean	Visualisation of an interval of locus position within the linkage group regarding a LOD score value.
Confidence Interval Management	Boolean	Computation of a confidence interval for locus position
Lod Threshold for confidence interval	Lod score unit	Threshold for confidence interval
Display adjacent loci	Boolean	Display adjacent loci
Display loci flanking the confidence interval	Boolean	Display interval bounding loci

B

C

Linkage Group 1
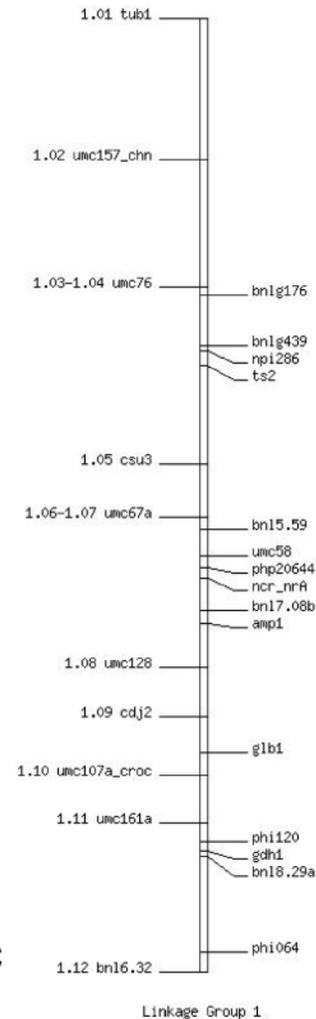
Figure 1. Mapping results display: (**A**) mapping results table; (**B**) results table optional fields; (**C**) graphical display of map. Loci of the framework are on the left side.

Table 1. Summary of the languages used to develop ActionMap

ActionMap component	Language
Web interface	PHP
Functions interacting with MapMaker	Perl
Functions interacting with Oracle	PHP/Perl
Map drawing function	PHP

Users can access the database and mapping automation software via their usual web browsers (MacOS, Windows and Unix version of Explorer or Netscape). The web interface is used to input data, to set up the automatic process and to display and export the results. ActionMap was developed for Solaris platform with Perl 5.6.1, PHP 4.1.2 (Table 1) and Oracle 8i DBMS. It requires MapMaker 3 (http://www-genome.wi.mit.edu/ftp/distribution/software/).

DATA MANAGEMENT AND SETTINGS

The new user must first choose a login and a password. Each user has access to only his or her own data. Before beginning to map, the user must create a 'mapping project' including data which describe the mapping population, the framework map and settings for the automatic mapping process. These settings include the minimal LOD of assignment, the maximum genetic distance, the mapping function, as well as decision criteria to stop or carry on mapping a marker when problems are encountered. The ActionMap documentation including a guide for first use of the software and sample data files is available (see Supplementary Material). Once a project is created, the user can submit new genotypic data to ActionMap. These data can be transferred into the database from a csv (coma separated values) file which may be produced by any spreadsheet program. The user is notified by email when the mapping job is completed and the results may then be viewed via a web link pointing to a table. The user can inspect the results for each new marker and choose to validate it or not. The results are then displayed as a table or map drawing (Fig. 1). Only validated loci will appear on the map drawing. A set of web pages are dedicated to browsing, updating, editing and deleting data. The user may delete data just for one locus or for a complete 'mapping project'.

MAPPING AUTOMATION

The ActionMap automatic mapping process consists of several scripts that automate MapMaker command steps (Fig. 2). The new markers are mapped one by one and the order in which they are processed is not important because each new locus is positioned on the same fixed framework map without being added to the map. ActionMap initially builds two files. The first one contains the segregation data of the framework map markers plus the new marker to be mapped and the second one contains parameters for MapMaker like the distance function and default linkage criteria, as well as the ordered list of markers for each linkage group. Data and user-defined settings are read from the 'mapping project' that is stored in the database (see before).
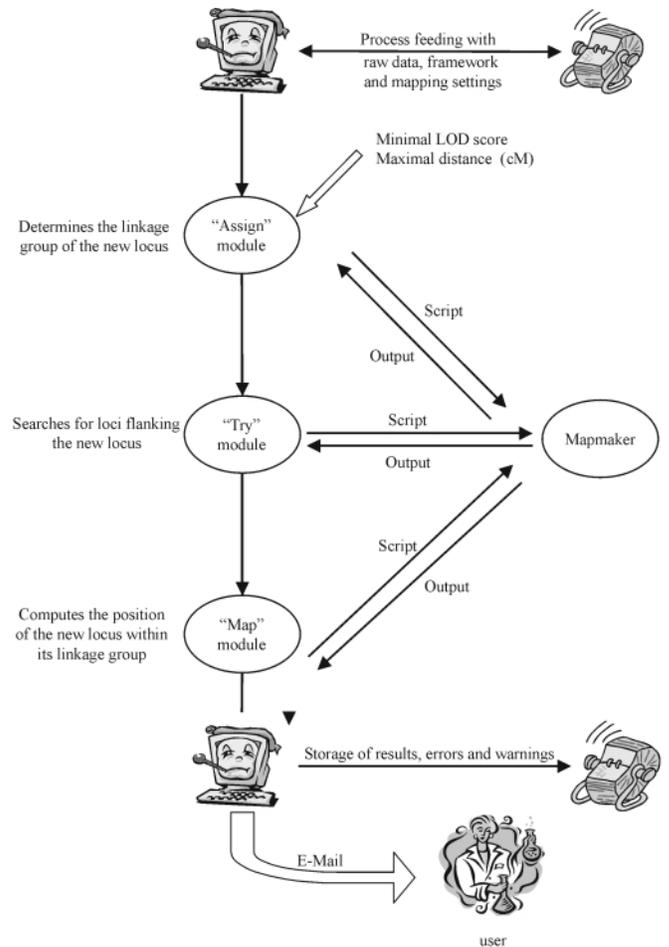
Figure 2. Flowchart of the automatic mapping process.

Once these two files are created, linkage computation is performed by three Automap functions: 'Assign', 'Try' and 'Map'. The names of these functions are related to the MapMaker commands which are mainly used by each function. The 'Assign' function determines to which linkage group the new locus belongs. Then the 'Try' function determines to which interval between loci the new marker can be assigned. The last function, 'Map', computes the absolute position of the new marker in its linkage group. These functions read data from files, generate scripts with MapMaker commands and analyze MapMaker outputs. At the end of the process, the database is populated with all MapMaker outputs.

The 'Assign' function uses the MapMaker *assign* command to determine to which linkage group the locus can be the most likely assigned and indicates the corresponding LOD score. The parameters for this function are the minimal LOD score for considering assignment and the maximal genetic distance allowed. Three kinds of results are expected: 'assigned to' meaning that the locus was successfully assigned; 'unassigned' when no assignment is possible (according to LOD threshold settings); and 'conflicting data' when the probabilities of assignment of the locus to several linkage groups are not significantly different from each other. If the result is 'assigned to', the MapMaker *group* command is used to test if the new

locus is linked to only one other locus of the linkage group. If not, the assignment is considered as reliable and the automatic process can carry on, but if the locus is linked to only one other locus, a segregation distortion test is computed (chi square) and the automatic process may be stopped depending on a user-defined threshold. If the marker is 'Unassigned', the minimal assignment LOD score value may be decreased to a user-defined minimal threshold. If the result remains 'Unassigned', the automatic mapping stops. If 'Assign' produces 'conflicting data' results, the new marker is ambiguously assigned to two linkage groups. Then, depending on a user-defined parameter, the process can select the assignment having the best LOD score, but only if the difference between LOD score values exceeds a user-defined threshold. If it does not, but one of the assignments is doubtful (distortion test) then the process selects the other one. When no choice is possible, the mapping process is stopped. In all cases, a warning indicates if the assignment has a low level of reliability.

After linkage group assignment, the next step is to find the flanking loci. This is achieved by the 'Try' function which uses the MapMaker *try* command. This command indicates the LOD score values associated with the assignments of the new marker to all possible intervals between adjacent loci in the linkage group. The 'Try' function can deal with co-localization cases or ambiguous position of the new marker and indicates the closest flanking loci corresponding to a given level of confidence of assignment.

The last step is performed by the 'Map' function. This function computes the distances between all adjacent loci of the linkage group. From this result, the 'Map' function computes the absolute position of the new marker in its linkage group.

The automatic mapping is then finished and all results are stored in the database.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

REFERENCES

1. Lander,E.S., Green,P., Abrahamson,J., Barlow,A., Daly,M.J., Lincoln, S.E. and Newburg, L. (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **2**, 174–181.
2. Stam,P. (1993) Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. *Plant J.*, **3**, 739–744.
3. Holloway,J.L. and Knapp,S.J. (1994) *GMendel 3.0 Users Guide*. Department of Crop and Soil Science, Oregon State University, Corvallis, OR, USA.
4. Marinov,M., Matise,T.C., Lathrop,G.M. and Weeks,D.E. (1999) A comparison of two algorithms, MultiMap and gene mapping system, for automated construction of genetic linkage maps. *Genet. Epidemiol.*, **17** (Suppl. 1), S649–S654.
5. Li,J.L., Deng,H., Lai,D.B., Xu,F., Chen,J., Gao,G., Recker,R.R. and Deng,H.W. (2001) Toward high-throughput genotyping: dynamic and automatic software for manipulating large-scale genotype data using fluorescently labeled dinucleotide markers. *Genome Res.*, **11**, 1304–1314.