MolMarker 1.0 Manual

E-mail: jahnkegizella@gmail.com

by Gizella Jahnke & József Smidla

Table of Contents

1. Introduction

1.1. What is MolMarker?

MolMarker is a user-friendly software to help evaluate the research results obtained by molecular markers. The software has a graphical user interface (GUI) and is platform-independent (Java application).

1.2. System requirements

Any kind of operating system supporting JVM (Java Virtual Machine) – eg. Windows or Ubuntu (Linux)

Java SE Runtime Environment 7 or newer (download from : http://www.oracle.com/technetwork/java/javase/downloads/jre7-downloads-1880261.html)

A minimum of 64 MB RAM

At least 100 Mb free hard disk space

1.3. How to install and uninstall MolMarker?

To "install" MolMarker simly place MolMarker folder anywhere to your hard disk. Double click MolMarker.jar to run. Delete the folder MolMarker to "uninstall".

2. Project management

2.1. Creating a project

When you open molmarker, a new project (named: "New") is already exist. Rename the project as described in 2.2, or create another project by clicking the menu item: Project/New Project.

2.2. Renaming a project

To rename a project click the menu item: Project/Rename Project.

2.3. Saving a project

To save a project click Save/Project. Your project will saved to the default folder under the name: *project_name*.prj.

3. Data types and management

3.1. MolMarker data types

MolMarker can handle 2 types of molecular marker data: “molecular” and “genetic”. Molecular markers with unknown genetic backgrounds (such as biochemical merkers, eg. isozymes) belong to the molecular type. Genetic markers (with known genetic background eg. microsatellites) belong to the genetic type.

“Molecular” data are coded reflecting to the banding pattern they show. Every locus has a maximum number of bands. Every band must have a number. In a “molecular” type locus the data must reflect to the presence (1) or absence (0) of the desirable band.

3.2. Reading data from file

MolMarker employs semicolon delimited files as input. You can read both molecular and genetic type of data from sufficiently formatted input files (figure 1-2.).

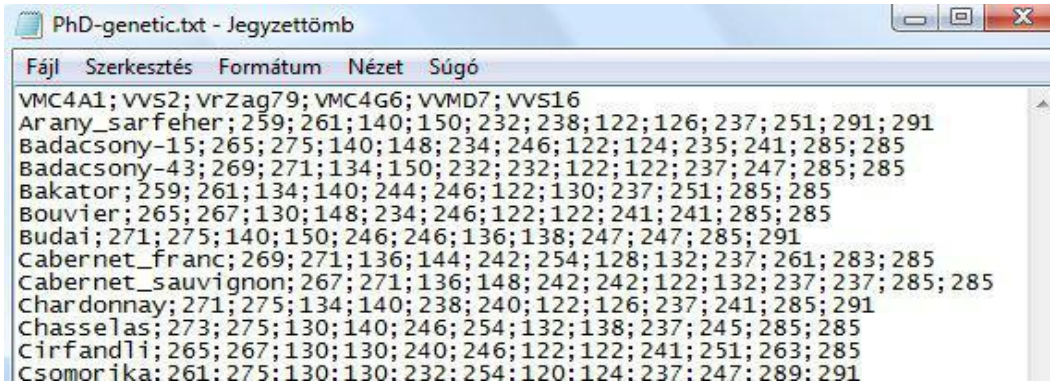


Figure S1. “Genetic” type input file format.

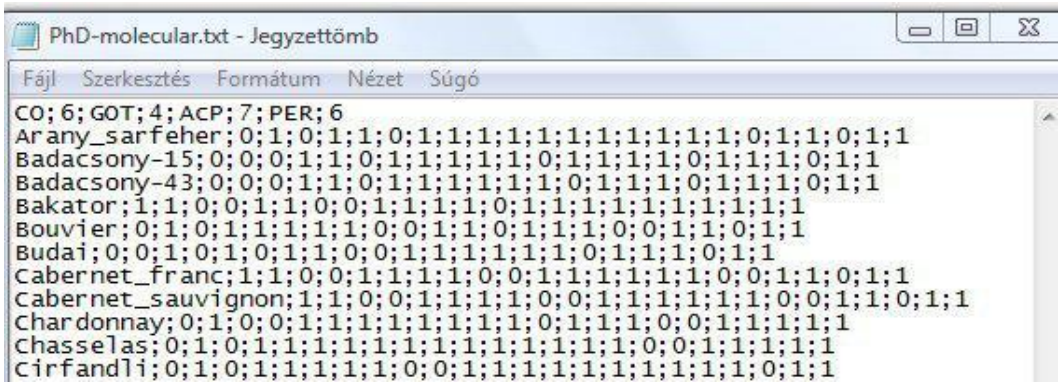


Figure S2. “Molecular” type input file format.

In both of the files in the first line there are information about the loci. In the "molecular" type the first line contains of the name of the locus followed by the number of bands. From the second line the first column of the input file contain cultivar (sample) identifier, followed by the banding patterns from one locus to the another.

In the "genetic" type the first line contains only the locus names, separated by semicolons. From the second line the first column contains the identifiers as well, followed by two numbers by locus, representing the two alleles of each loci.

You can easily create input files by saving your data file in your favorite spreadsheet software (eg Microsoft Excell) to csv format.

To read data from your properly formatted data file choose Data/Read Data From File in MolMarker menu.

3.3. Input data

You can directly input data to MolMarker using the Data/Input Data menu, and following the onscreen instructions (Figure 3.).

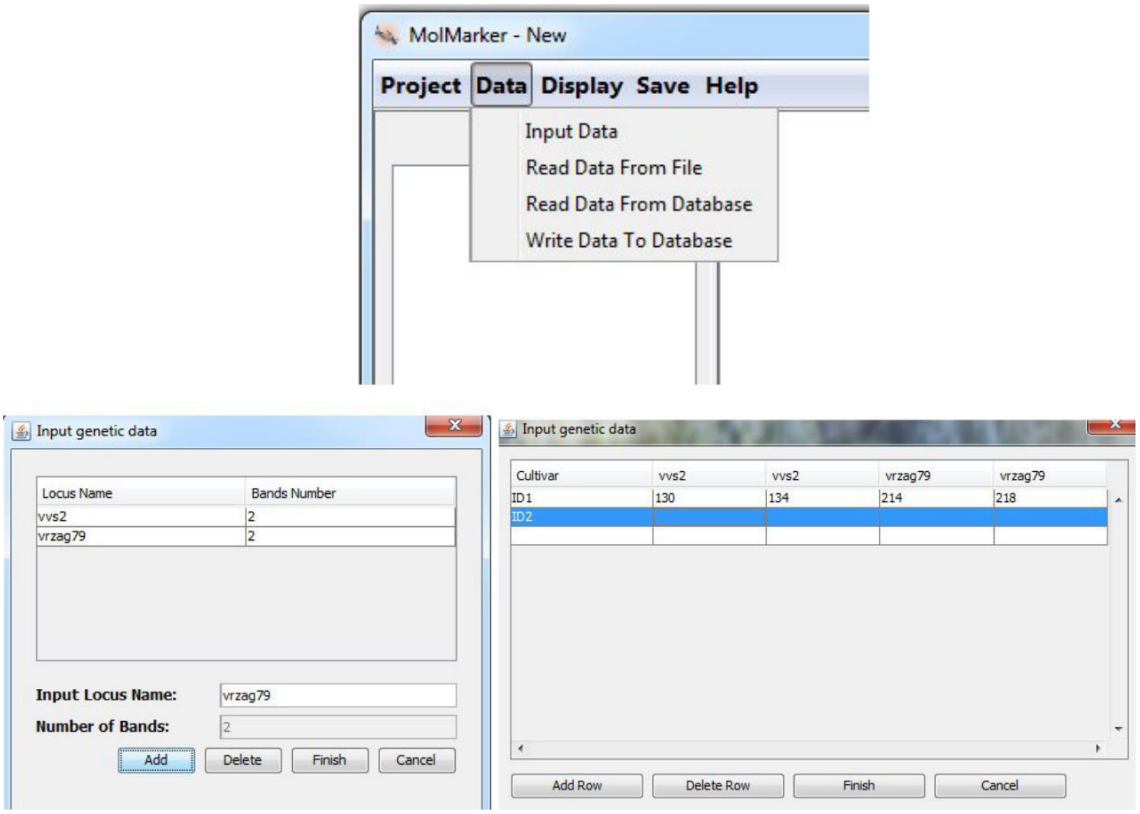


Figure S3. MolMarker Data menu and input data process

3.4. Database integration

You can save your data to database, or load your saved data from database as well. To create your database, MolMarker.sql file is provided. The ER diagram of the created database is shown in figure 4.

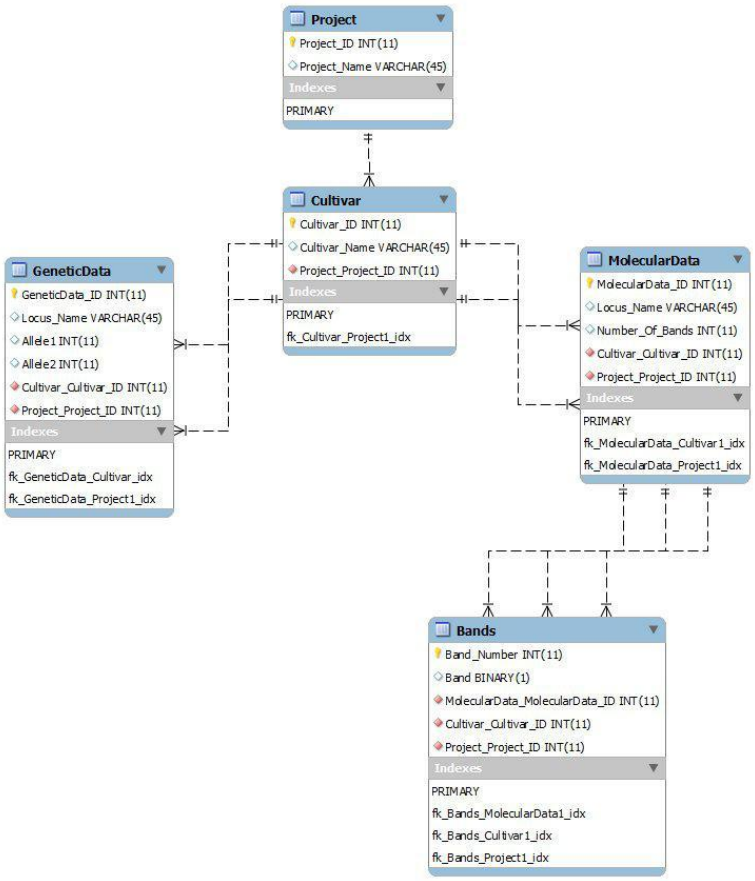


Figure S4. ER diagram of the database

Use “Read Data form Database” and “Write Data to Database” menus for database manipulation.

4. Data analyses

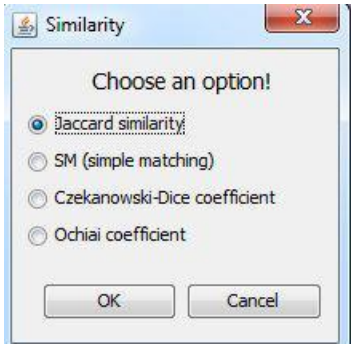
4.1. Summary statistics

Summary Statistics/Genetic menu item displays the frequency of alleles of the genetic loci. The H and PIC values for all of the loci are displayed. You can indicate, which locus has a null allele, and MolMarker calculates the frequency of null allele and updates the frequency of the visible alleles by using EM algorithm.

Summary Statistics/Molecular menu item displays the relative frequencies of bands and PIC values for molecular data.

4.2. Creating similarity matrices

MolMarker can create similarity matrices using 4 kind of index:



The displayed matrix can be saved to Excell file as well (Save/Similarity Matrix).

4.3. Creating dendograms

MolMarker can create and display UPGMA and Neighbor-joining dendograms based on any of the 4 similarity matrix. Dendograms is saved to the default folder in png, svg and MolMarker own file format (.image).

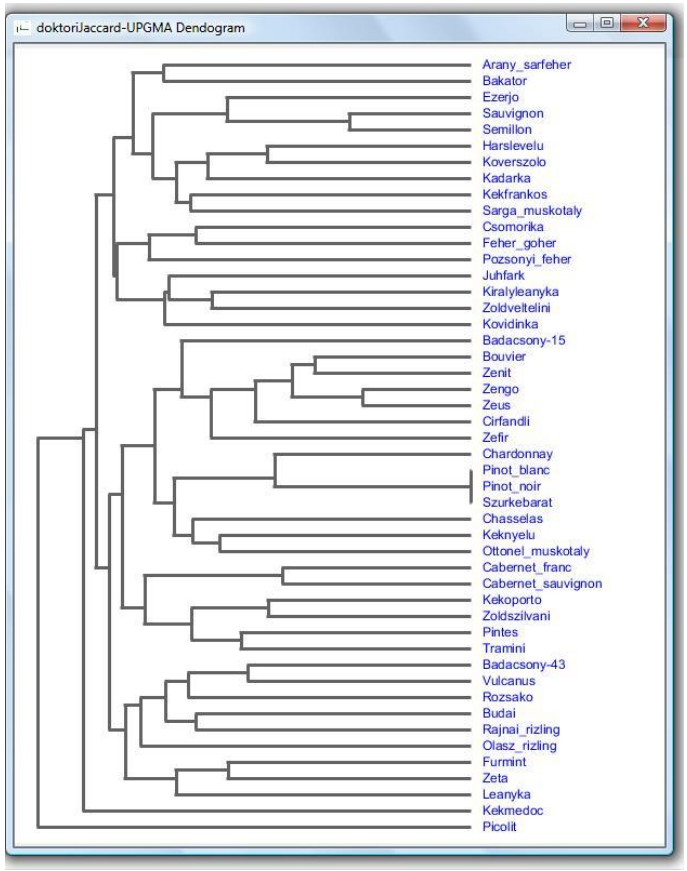


Figure S5. UPGMA dendogram displayed by MolMarker

4.4. Parentage analyses

Menu item Phylogeny/Parentage gives a list of possible parent-offsprings, and likelihood ratio statistics corresponding to the detected combinations.

The following likelihood ratios are provided (per locus and combinedover loci):

1. The ratio of the probability that th e proposed parents gave rise to the offspring's genotype versus the probability that two random individuals give rise to the offspring's genotype. (Proposed parents) versus (two random cultivars) = X x Y in output file

2. Likelihood ratio for: (Proposed parents) versus (random individual x proposed parent 1) = X x (1)

3. (Proposed parents) versus (close relative of proposed parent 2 x proposed parent 1) = rel(2) x (1)

4. (Proposed parents) versus (Proposed parent 2 x random cultivar) = (2) x X

5. (Proposed parents) versus (Proposed parent 2 x close relative of proposed parent 1) = (2) x rel(1)

5. References

1. Bowers, J.E.; Meredith, C.P. The parentage of a classic wine grape, Cabernet Sauvignon. *Nat. Genet.* **1997,** *16*, 84–87
2. Gascuel, O.; Bryant, D.; Denis, F. Strengths and Limitations of the Minimum Evolution Principle. *Syst. Biol.* **2001**, *50*, 621–627.
3. Nagy, S.; Poczai, P.; Cernák, I.; Gorji, A.M.; Hegedűs, G.; Taller, J. PICcalc: An Online Program to Calculate Polymorphic Information Content for Molecular Genetic Studies. *Biochem. Genet.* **2012**, *50*, 670–672. https://doi.org/10.1007/S10528-012-9509-1.
4. Rzhetsky, A.; Nei, M. A Simple Method for Estimating and Testing Minimum-Evolution Trees. *Mol. Biol. Evol.* **1992**, *9*, 945–945. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040771.
5. Saitou, N.; Nei, M. The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* **1987**, *4*, 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.
6. Sokal, R.; Michener, C. A Statistical Method for Evaluating Systematic Relationships. *Univ. Kansas, Sci. Bull.* **1958**, *38*, 1409–1438; ISBN 0001948000237.
7. Speed, T. Neighbour Joining Method (Saitou and Nei, 1987). 2006.