

# GABI GAIN Data INtegration And Mining Tool

Alex Finck, Jürgen Gremmels, Robert Wagner, Birgit Kersten, Björn Usadel

www.gabiPD.org  
gabipd@mpimp-golm.mpg.de

## Introduction

Modern plant breeding research programs produce a huge amount of genomic and phenotypic data. Therefore efficient data mining becomes indispensable to exploit the generated information and to support the improvement of experimental strategies and the optimization of process design. In the scope of the GABI GAIN project, DINAMIT (Data INtegration and MIning Tool) is being developed for the integration of standardized data produced in genomics-based plant breeding. A database schema has been built into DINAMIT to store integrated breeding data of different crops in an efficient and flexible way. A corresponding exchange format, GAIN-Tab, has been defined to ease the integration of new data into DINAMIT. Based on that data, DINAMIT provides a connection with specialized software for the analysis of results of plant breeding experiments. For formats different from GAIN-Tab converters in the form of R-Scripts have been developed. Finally, the development of the graphical user interface DINAMIT provides a user with management functionality on the data, different tabular and graphical visualizations as well as the integration of various tools for planing and analysis of results of geostatistical experiments. Below we present results concerning the integration of plabqtl [1] and R Scripts for geostatistical analysis.

## GAIN-Tab

The DINAMIT graphical user interface (GUI) implements general purpose consistency checking for the GAIN-Tab format. GAIN-Tab can easily be generated using spreadsheet tools (Excel, OpenOffice), as shown on Figure 1, or using dedicated automatic converters. Datasets are subsequently validated by the DINAMIT GUI when they get imported.

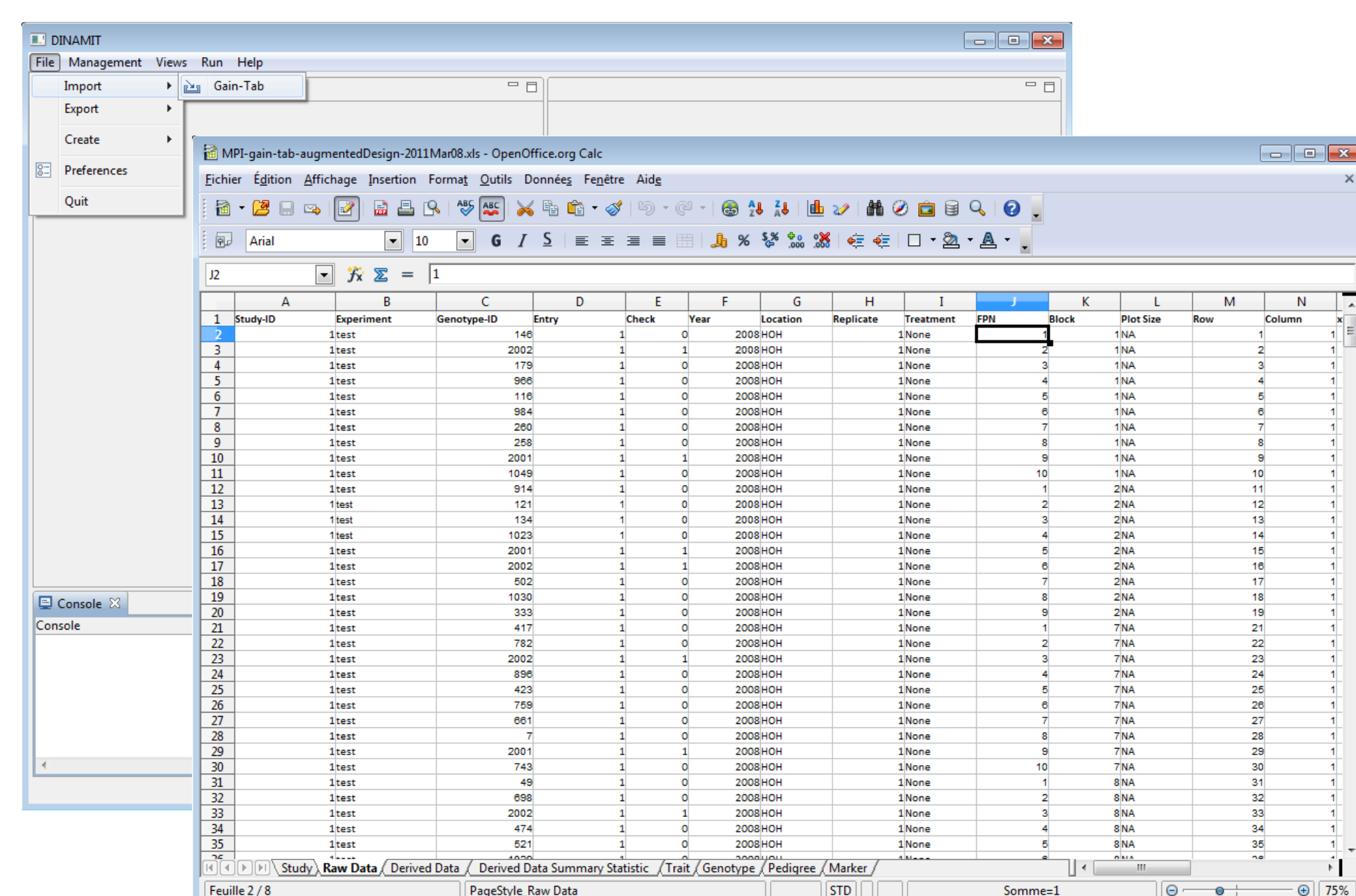


Figure 1: Definition of a GAIN-Tab dataset (foreground) and import into DINAMIT (background)

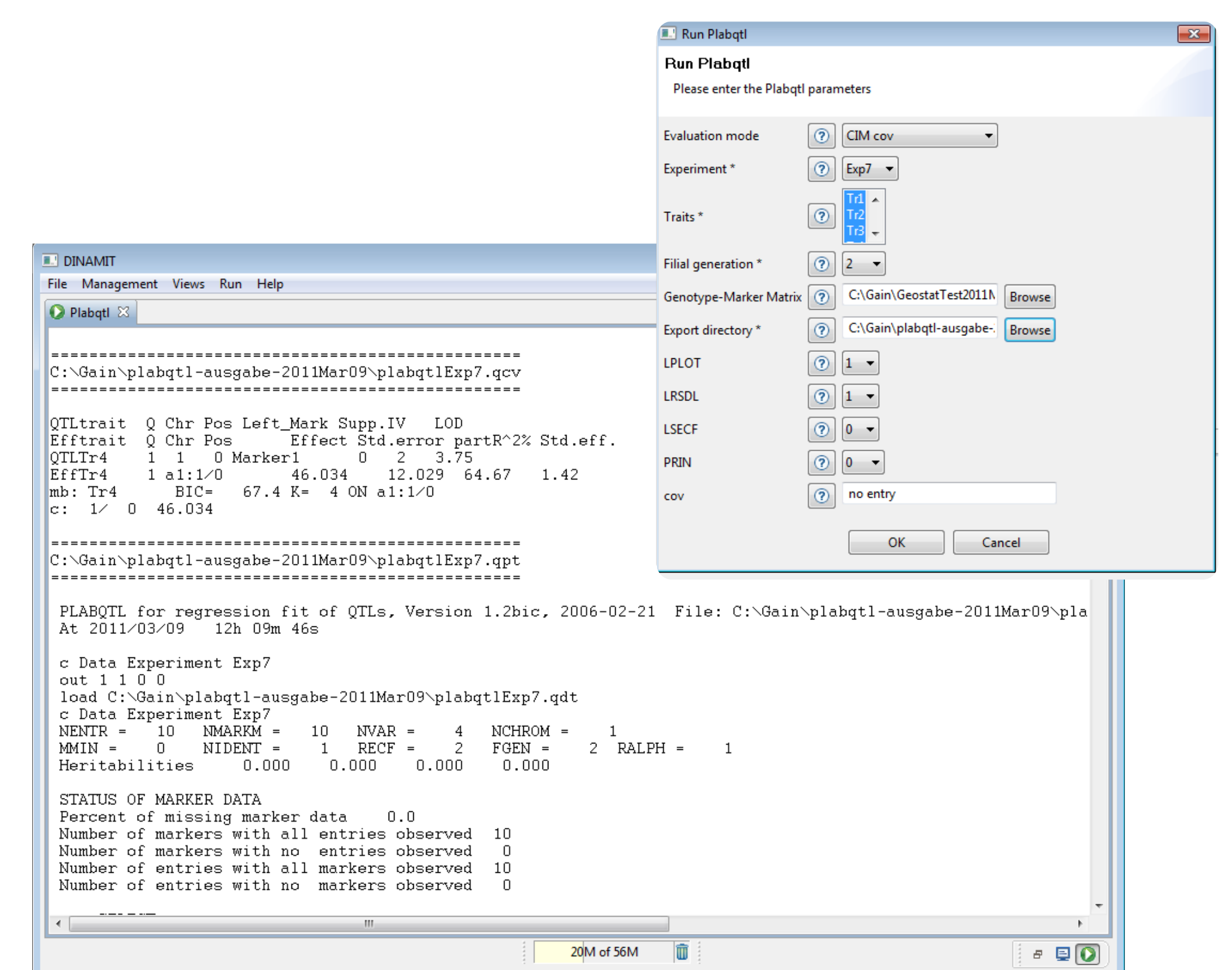


Figure 2: Configuration dialog for plabqtl (upper right) and corresponding results (lower left)

## QTL detection with plabqtl

In DINAMIT a configuration of plabqtl is specified by the user interactively using a dialog shown in Figure 2. A user can configure plabqtl to search for QTLs using Simple Interval Mapping (SIM) and Composite Interval Mapping (CIM) with or without cofactors as advised by the plabqtl documentation. Input data corresponding to a GAIN-Tab dataset originally chosen by the user is then extracted automatically from the built-in database of DINAMIT and results are presented to the user as shown on Figure 2.

## Geostatistical analysis of plant breeding experiments

R scripts developed in the context of Gabi GAIN WP C [2] have been integrated into DINAMIT. Thus DINAMIT allows a geostatistical analysis of plant breeding experiments performed using either a replicated alpha design or an unreplicated augmented design. A unified dialog allows a user to choose between the two different designs. Here model dependent consistency checks on a GAIN-Tab dataset originally chosen by the user can also be specified. After confirmation of the dialog and consistency of the data has been asserted results are presented to the user as shown on Figure 3.

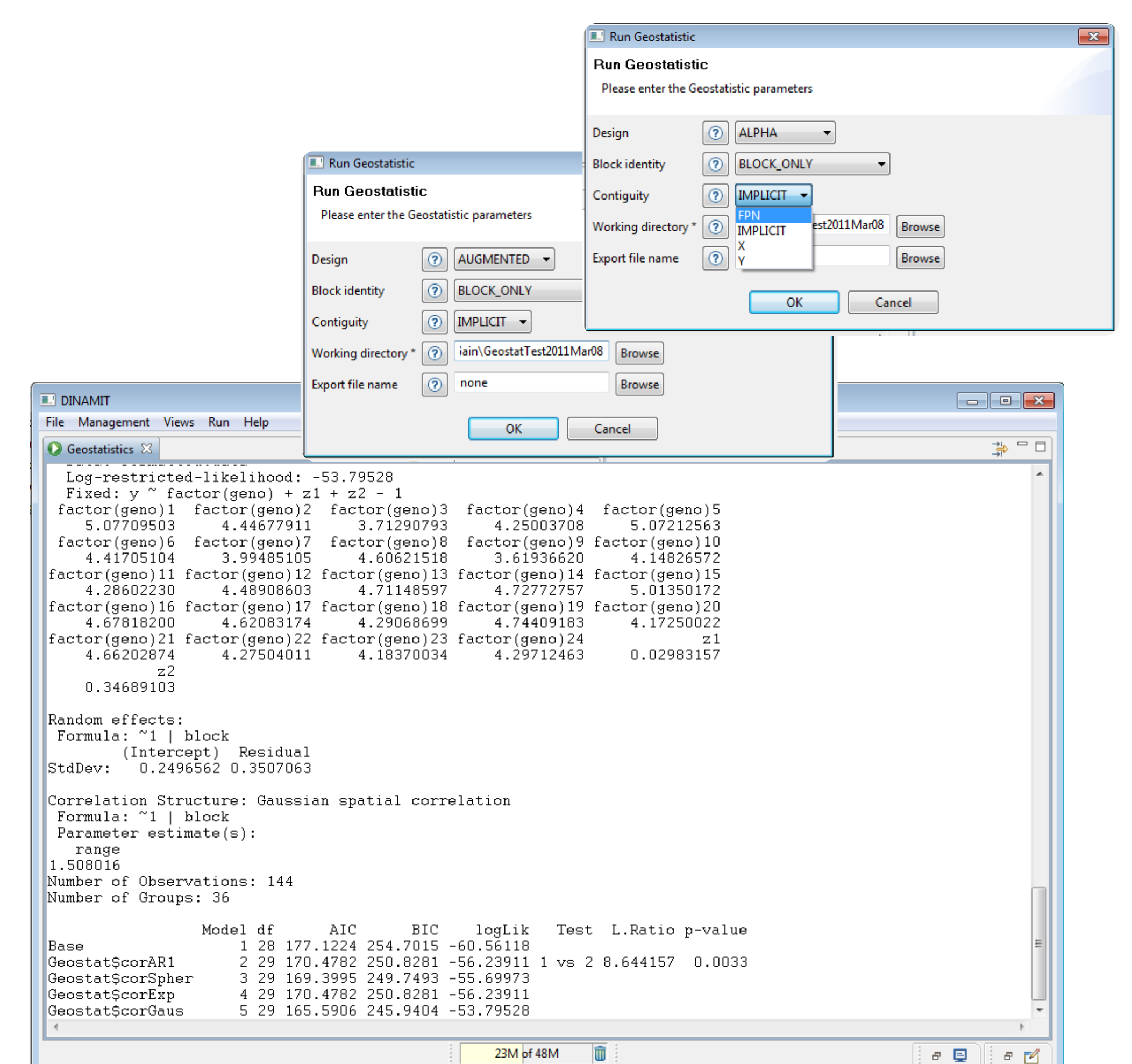


Figure 3: Configuration dialog for geostatistical analysis of datasets produced using an alpha design (upper right) or an augmented design (middle) and corresponding results (below)

## Availability

DINAMIT is provided as a locally installable application for partners within the GABI GAIN project to optimize genomics-based plant breeding

## References and Publications

- [1] Utz & Melchinger (1996), 'PLABQTL: A program for composite interval mapping of QTL', Journal of Agricultural Genomics 2, 1.
- [2] Müller, S. & Piepho (2010), 'Arrangement of check plots in augmented block designs when spatial analysis is used', Plant Breeding 129, 581-589.

## Acknowledgments

This work is supported by the German Federal Ministry of Education and Research BMBF (GABI-FUTURE grant 0315072C)