GeneMarker®
by SoftGenetics
May 2009

Database Searching and Kinship Analysis of Monoecious Plant and Invertebrate Microsatellite Data

He Haiguo, David Hulce, Teresa Snyder-Leiby, Jonathan CS Liu

Introduction

Database searches for exact duplicate and near relative samples are extremely useful in applications such as determining plant and animal population diversity, relatedness of individuals within populations, identification of successful breeding individuals or clonal purity of asexually reproducing populations. Kinship analyses are powerful tools but have many challenges due to remote DNA sampling of animal populations, lack of information on known breeding pairs, and mobility of individuals (animal migration and seed dispersal).

Short Tandem Repeats (STR), simple sequence repeats (SSR) or microsatellite analysis has the ability to provide complete individual profiles. Microsatellites are variable regions in genomic DNA which are amplified with specific primers by Polymerase Chain Reaction (PCR). Many polymorphic plant and animal STR markers that follow Mendelian inheritance have been identified[1-3]. The likelihood that unrelated individuals will share the same STR profile can range from 1 in a million or more, depending on the number of loci compared between the two samples. Related individuals have more shared loci than those that are unrelated. The higher the number and diversity of loci included in the genotype the greater the significance of the likelihood ratio (LR) results. Kinship formulas have been established in the literature to calculate the relatedness between individuals based on shared loci[4].

GeneMarker is biologist friendly genotyping software with integrated Kinship analysis, using identity by descent (IBD) calculations to provide likelihood of relationship level between two individuals, and Database searching tools to identify exact duplicate and near relative samples.

I. Procedure: Database Search – exact matches and probable relatives

1. Import data files (*.FSA, *.ESD, *.RSD, *.SCF)
2. Data Analysis →Run Wizard → size and allele calls result (fig 1)
3. Select Applications →Relationship Testing
4. Select Tools →Allele Frequency (fig 2)
5. Select DataBase →Save to database
6. Select Tools →Genetic Analysis Parameters → set allowance for mutation/mistyping, limit number of files or minimum LR to report
7. Select Tools →Family Group Tool →Okay
8. Select individual node, right click and choose find family (fig 3)
9. Click on 'Report' to display all files with the same STR profile, the random match probability and files with the highest kinship scores to the sample

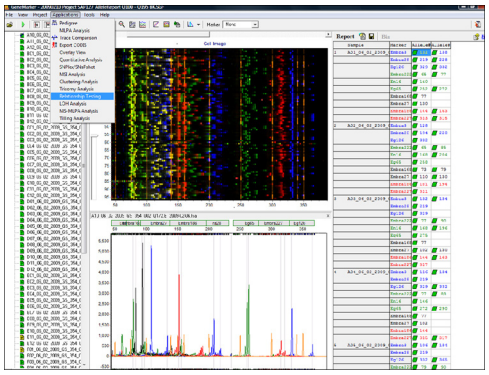

Figure 1: The main analysis screen displays sized data and allele calls. User friendly linked navigation between allele report, electropherogram and synthetic gel image aids in data review.

Results: Database Search – exact matches and probable relatives

The database search and kinship calculations for sample A01 were run under parameters of no mutation allowance, gender deselected, minimum LR=1 and maximum number of files = 96. When a file fulfills the IBD conditions for more than one relationship level with the current sample, it will be reported in the relationship level with the highest LR. Database search results in figure 3 indicate that there are no additional exact matches in the database. The random match probability for this microsatellite in the given population is 7.94×10^{18}. There are no files in the database that fulfill the IBD conditions for a parent/child relationship. Four files have potential full-sibling relationship and several have a possible half-sibling relationship.

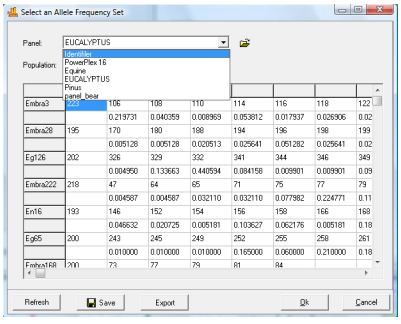

Figure 2: Species specific and population specific allele frequency tables are easily imported as .txt files and used for IBD calculations.

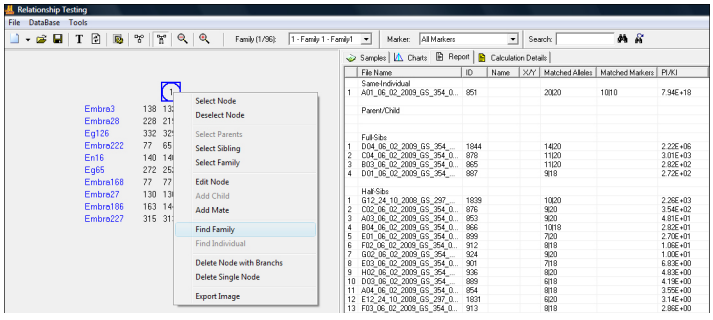

Figure 3: Results of the database search and calculation of likelihood ratios indicate that there are no files in the database with a parent child relationship to sample A01 but several potential sibling and half-siblings exist.

II. Procedure: Kinship Analysis

1. Follow steps 1-4 above
2. Select Tools →Kinship analysis
3. Use parameter icon to select relationship levels and report content (figure 4)
4. Use dropdown menus to select the two files for analysis from the current project, the relationship testing database or a .txt genotype file (figure 5)

Results: Kinship Analysis

Kinship analysis with no mutation or mistyping allowance (figure 6) indicates that individual B is 2.22 million times more likely to be a full sibling to individual A than a random, unrelated sample from the population. Individual B is most likely a full sibling, 790 times more likely to be full sibling than half sibling (Full Sib LR/Half Sib LR).

Kinship analysis with mutation/mistyping allowance of 1 marker (figure 7) substitutes the mutation rate for the allele frequency in the IBD calculation for marker En16. Results indicate that the probability of a parent/child relationship between the samples under the assumption of one mistyped marker is no more likely than a half-sibling relationship.

Discussion

The rigorous statistical analysis to determine levels of kinship uses identity by descent (IBD)[4,5]. GeneMarker's database search tool identifies samples with the same STR profile and calculates the random match probability (the probability that a randomly selected individual from a population will have an identical STR profile at the DNA markers tested). The same tool also searches the database and identifies files with the highest likelihood ratio for each relationship level to the experimental sample. Genetic Analysis Parameters allow setting tolerances for mistyping or mutation, limiting the number of retrieved samples by LR score or total number of samples, parent/child, sibling and half sibling search results. Allele frequencies for different species and different populations within a species can be easily uploaded and used in the relationship testing applications. The 'Save to Database' function provides easy database updates in GeneMarker; it accepts current project genotype results, previously archived genotype .txt (tab or comma delimited) or .cmf file formats. The 'Kinship Analysis' tool provides a report table with probabilities and likelihood ratios across three generations for sample pairs.

GeneMarker Relationship Testing has all of the strengths of GeneMarker including; unique pattern recognition and sizing technology providing >99% accuracy, easy linked navigation, management control and tracking, exportable LIMS reports, bulk printing capabilities, instrument compatibility with ABI (Life Technologies), MegaBACE® (GE Healthcare) and CEQ (Beckman-Coulter).

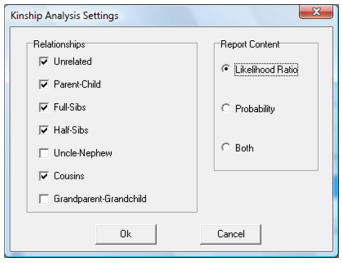

Figure 4: Select Relationship Levels and Report Content from the Kinship Analysis dialog box.

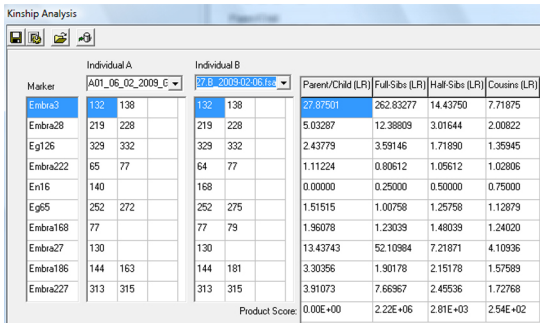

Figure 5: Results of Kinship Analysis between A01 and D04 without mutation/mistyping allowance.

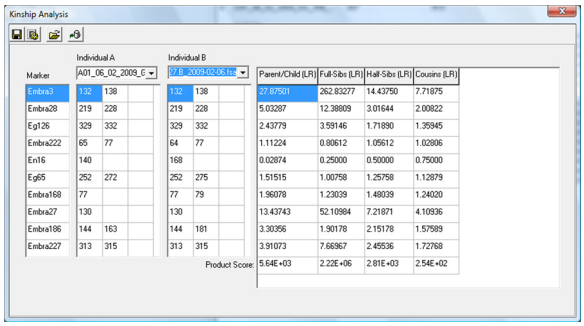

Figure 6: Results of Kinship Analysis between A01 and D04 with a mutation/mistyping allowance of 2 markers.

Acknowledgments

We would like the thank Dr. Daleen van Dyk, Forestry and Agricultural Biotechnology, University of Pretoria for data used in demonstration of the database search application for diploid, monoecious plants.

References

1. M.M. van Dyk, G. Koning, Z. Simayi, S. Booi, R. Maharaj, M.C. Selala and D.J.G. Rees. Development of microsatellite markers for marker-assisted breeding in pears (Pyrus spp.) Acta Horticulturae 2004, 671:307-313.
2. M. Morgante and A.M. Olivieri. PCR-amplified microsatellites as markers in plant genetics. The Plant Journal 1993, 3(1):175-182.
3. K. Weising, P. Winter, B. Huttel, G. Kahl. Microsatellite markers for molecular breeding. Crop Sciences: Recent Advances 1998.
4. Brenner, C. Symbolic kinship program. Genetics 1997, 145:535-542.
5. Li, C.C., and L. Sachs. The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. Biometrics 1954, 10:347-360.

SoftGenetics LLC 200 Innovation Blvd. Suite 235 State College, PA 16803 USA
Phone: 814/237/9340 Fax 814/237/9343
www.softgenetics.com email: info@softgenetics.com