GeneMarker®
by SoftGenetics

Kinship Analysis and Duplicate Sample Identification of Animal Microsatellite Markers with GeneMarker®

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

Introduction

Kinship analyses are strong tools, either separately or in complementation with observation/behavioral data, to track emigration/immigration and inbreeding in natural populations. There are many challenges to kinship analysis with natural populations due to remote DNA sampling and mobile populations. Paternity issues in animal breeding programs are common with livestock and in aquaculture. Aquaculture also requires the ability to detect escaped domestic individuals in wild populations, as is a concern with salmon.

Short Tandem Repeats (STR) analysis has the ability to provide complete individual profiles, even when the DNA samples have degraded from time or exposure to the elements. STRs are variable regions in genomic DNA which are amplified with specific primers by Polymerase Chain Reaction (PCR). Many polymorphic 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 billion or more, depending on the number of loci compared between the two samples. Related individuals have more shared loci than those that are unrelated. 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 and Database searching tools. The 'Kinship Analysis' tool provides a report table with probabilities and likelihood ratios across three generations for sample pairs. The rigorous statistical analysis to determine levels of kinship uses identity by descent (IBD), follows the methods of Brenner[4] and uses stochastic matrices of Li and Sachs.[5] GeneMarker 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 'Find Family' tool 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. The save to database function in GeneMarker can accept allele frequency tables for species specific markers and can accept previously archived genotype .cmf files, providing easy database updates.

Procedure

To Locate Duplicate STR profiles and Nearest Relatives:
1. Import data files (fsa, abi, ab1, scf)
2. Select the Run icon to launch the Run Wizard to make allele calls
3. After the data is processed, select Applications → Relationship Testing
4. Select the appropriate allele frequency
5. Select DataBase and 'Save to database'
6. Select Family Group Tool and 'Okay'
7. Select individual node, right click and choose find family
8. Click on 'Report' to display all files with the same STR profile and files with high kinship scores to the sample

To Compare Two Samples in Kinship Analysis:
1. Follow steps 1-4 above
2. Select Tools → Kinship analysis
3. Use dropdown menus to select the two files for analysis
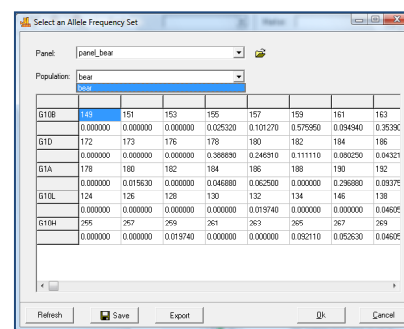4. Use parameter icon to select relationship levels and report content
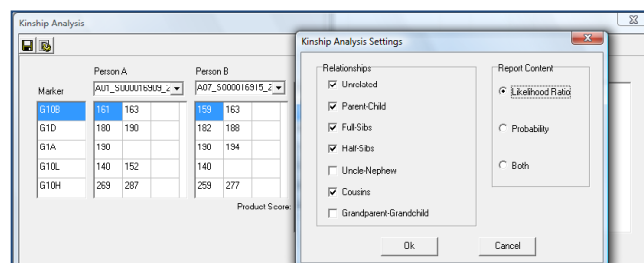

Figure 1: Allele frequency panel

Figure 2: Kinship analysis settings

Results

Each individual file is represented as a node (circle = female, square = male) with the genotype listed below the node. The find family tool activates the database search and lists all files with the same STR profile as the selected individual. Six samples from the project of 100 STR profiles are duplicates (Figure 3). The likelihood that random samples from this population will share all of these loci is 1 in 5,280,000. Kinship analysis indicates that it is 1,080 times more likely that individuals with these STR profiles are full siblings rather than unrelated individuals from the population (Figure 4).

The GeneMarker Relationship Testing uses rigorous statistical methods to determine kinship between an unidentified sample and members within a database. The DataBase can be updated by using save to database in the Relationship Testing Application. Text files can be used to update the Allele Frequency functions in one step. The Relationship Testing automatically generates reports that include: likelihood ratios of nearest relatives in the database, identifiers, number of alleles matched and a list of exact copies located when comparing the unidentified sample to the database.

The instant listing of duplicate copies of a genotype along with the random match probability in the report provides excellent verification of the actual number of animals in a natural population. This information is also helpful in deciding which samples should be tested with additional markers for higher statistical confidence in relationship level testing. Additional applications include quantification of non-harem breeding, population diversity and inbreeding studies.

Figure 3: Search results of samples with identical STR profiles

Figure 4: Kinship Analysis between two samples

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, MegaBACE and Beckman-Coulter.

Acknowledgements

1. Banks, M.A. et al. (2000) Analysis of microsatellite DNA resolves genetic structure and diversity of Chinook salmon (Oncorhynchus tshawytscha) in California's central valley. Can. J. Fish. Aquat. Sci. 57, 915–927
2. Jones, K.L. et al. (2002) Refining the whooping crane studbook by incorporating microsatellite DNA and leg-banding analyses. Conserv. Biol. 16, 789–799
3. Kyle, C.J., T.J.Karels, B. Clark, C.Strobeck, D.Sl Hik and C.S. Davis (2004) Isolation and characterization of microsatellite markers in hoary marmots (Marmota caligata) Molec. Ecol. Notes 4, 749-751
4. Brenner, C. Symbolic kinship program, (1997) Genetics 145, 535-542
5. Li, CC and L Sachs (1954)The derivation of joint distribution and correlation between relatives by the use of stochastic 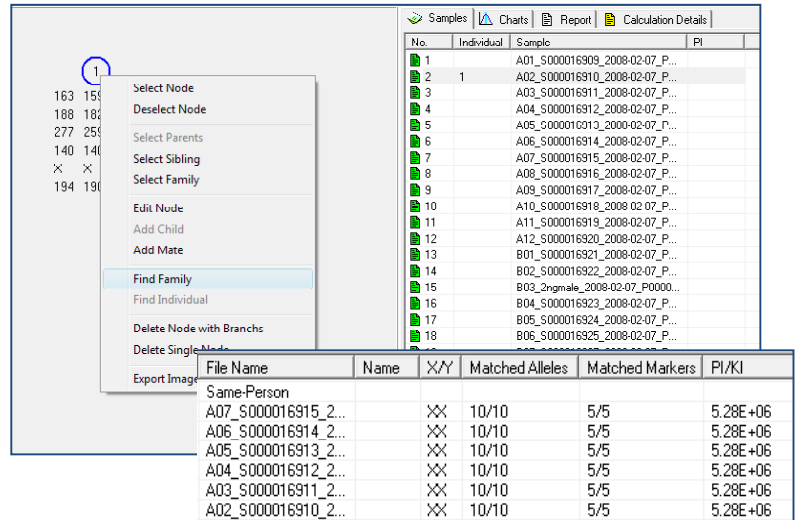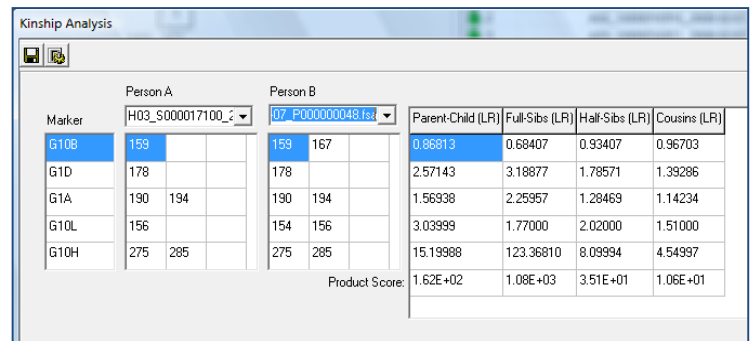matrices. Biometrics 10:347-360