Analysis With Microsatellite And SSRs: Applications Using Relationship Testing & Database Search Tools

GeneMarker®

E. Gus Cothran[1], Eleanore Conant[1], Rytis Juras[1], David Hulce[2], Haiguo He[2], Kevin LeVan[2], Wan Ning[2], Teresa Snyder-Leiby[2], CS Jonathan Liu[2]
[1]Animal Genetics Lab, VIBS, CVM, Texas A&M University, TAMU 4458, College Station, TX 77843-4458 USA
[2]SoftGenetics, LLC 200 Innovation Blvd., Suite 235 State College, PA, 16803 USA

ABSTRACT

Kinship analyses are excellent tools to track emigration/immigration, estimate inbreeding and number of successful breeders in a population in addition to many other areas within behavior, evolution, conservation and agriculture research. There are many challenges to kinship analysis with natural animal populations due in part to remote DNA sampling and mobile populations. GeneMarker® microsatellite genotyping software is widely used in plant and animal research for reliable fragment allele calling and mutation detection via SNPlex™/SNaPshot® SNPWave™ and TILLING® analyses. Newly developed Relationship Testing and Database Search Tools in GeneMarker identify replicate samples and display samples with high likelihood ratios for parent, sibling and half-sibling relationship levels. The results enable more accurate estimations of population diversity such as: actual number of individuals in a population, number of breeders and the amount of inbreeding. The Kinship Analysis tool provides probabilities, likelihood ratios and LOD values for parent/child, sibling, half-sibling and cousin relationship levels when comparing two microsatellite profiles. Calculations are based on identity by descent (IBD), and allow for the possibility of mistyping in parents or offspring, mutation and incomplete profiles. Project appending and extended pedigree drawing features enable building pedigrees as data becomes available over time in extended population studies. Applications of these Relationship Testing Tools in GeneMarker for plant research (population heterogeneity, gene flow, assessing genetic diversity through pedigree analysis and seed stock contamination) and animal research (population density, relatedness, kinship analysis) will be presented.

INTRODUCTION

Short Tandem Repeats (STR) or Microsatellite analysis has the ability to provide complete individual genetic 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 Brenner4 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 previously archived genotype .cmf or .txt files, providing easy database updates. For this analysis, a population of 115 horses has been chosen. These horses are from the Southeastern US and are currently being organized into a breed registry. Little pedigree information was available,therefore DNA testing will provide help in setting up the initial registry stud book as well as in confirmation of presumed parentages. In addition to pedigree information, the DNA testing will be used to help develop management practices for this heritage horse population which is considered a rare breed by the American Livestock Breeds Conservancy.

METHODS

DNA EXTRACTION/AMPLIFICATION

For this study Equine DNA was extracted from hair root bulbs: five hair bulbs in 50µl total volume of solution containing 1xPCR buffer and 5µl of Proteinase K (20 mg/ml) were incubated for 45 minutes at 57ºC, followed by 15 minutes at 95ºC and finishing with cooling sample to 4ºC. The DNA typing panel consisted of 13 microsatellites: AHT4, AHT5, ASB2, HMS3, HMS6, HMS7, HTG4, HTG10, VHL20, ASB17, ASB23, LEX33 and LEX3. Amplification of microsatellites in multiple PCR reactions was performed in 25µl total volume reactions containing 30 ng of genomic DNA, 0.07 to 0.8 pmol of primers, 1xPCR buffer, 2.5mM MgCl2, 0.2 mM dNTPs and 1U AmpliTaq. For microsatellite amplification thirty cycles of 1-minute at 95ºC, 30s at 560C, 30s at 72ºC followed by final extension at 72ºC for 10 minutes were used. Reaction products were analyzed using an ABI 377 DNA sequencer. Fragment sizes were determined using the computer software STRand. Archived .txt files of genotypes were added to the database for relationship testing. Bear microsatellite profiles were provided by Dr. Kyle,Trent University. DNA typing panel consisted of six microsatellites: G10B, G1D, G1A, G10L, G10H and AMEL. PCR products were analyzed on an ABI 3730, genotyped with GeneMarker and added to the Relationship Testing database directly from genotyping.

RELATIONSHIP TESTING

Relationship Testing
Kinship Formula:

$[P_2(xy) \times \Phi_2] + [P_1(xy) \times \Phi_1] + [P_0(xy) \times \Phi_0]$

Where:

$\Phi_2\ \Phi_1\ \Phi_0$ = Identity by descent coefficients for sharing 2, 1 or 0 alleles
$P_2(xy)$ = The probability of genotype y given genotype x with 2 of their alleles IBD
$P_1(xy)$ = The probability of genotype y given genotype x with 1 of their alleles IBD
$P_0(xy)$ = The probability of genotype y given genotype x with 0 of their alleles IBD

Formulas for each possible combination of alleles, derived from stochastic matrices of Li and Sachs.[5]

$P_A\ P_B\ P_C\ P_D$ = Probability of that allele for a given population

Genotype combination	Frequency
AB AB	$\Phi_2 + 0.5\,\Phi_1(P_A + P_B) + 2\,\Phi_0 P_A P_B$
AA AA	$\Phi_2 + \Phi_1 P_A + \Phi_0 P_A^2$
AA AB	$\Phi_1 P_B + 2\Phi_0 P_A P_B$
AB AC	$0.5\,\Phi_1 P_C + 2\Phi_0 P_A P_C$
AB CD	$2\Phi_0 P_C P_D$
AA BB	$\Phi_0 P_B^2$
AA BC	$2\,\Phi_0 P_B P_C$

Relationship Testing Applications

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

Blind testing of the equine profiles was performed in GeneMarker Relationship Testing for comparison with known relationship values for this herd.

RESULTS

CALCULATE LR FOR NEAREST RELATIVES

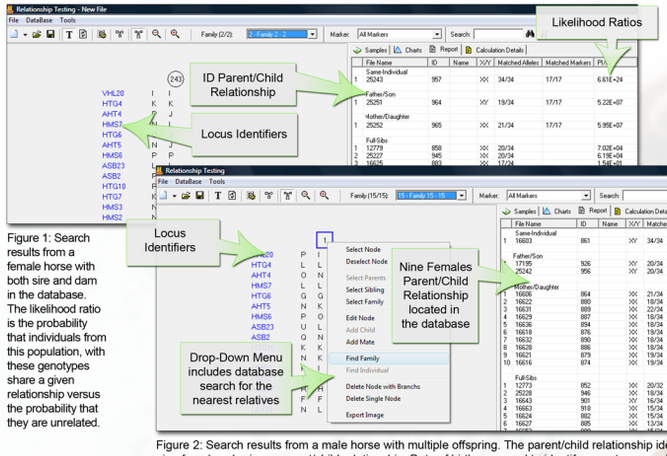

Figure 1: Search results from a female horse with both sire and dam in the database. The likelihood ratio is the probability that individuals from this population, with these genotypes share a given relationship versus the probability that they are unrelated.

Figure 2: Search results from a male horse with multiple offspring. The parent/child relationship identifies one male and nine females sharing a parent/child relationship. Date of birth was used to identify parent versus child files.

IDENTIFY DUPLICATE SAMPLES FROM REMOTE SAMPLING OF WILDLIFE POPULATIONS

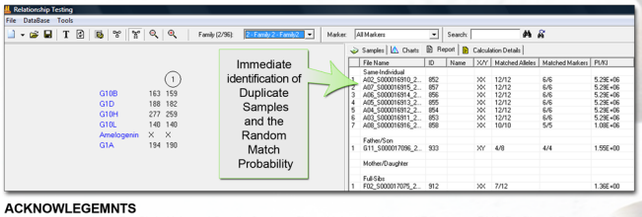

Figure 4: Search results from a female bear from a population that was remotely sampled. Results under Same Individual indicate that there is a 1 in 5,290,000 chance that another bear from this population has the same genetic profile. This information enables us to reduce the population count; providing more accurate population density estimation in wild populations.

IDENTIFY ALLELE CONFLICTS AND ALLOW FOR MUTATION IN KINSHIP CALCULATION

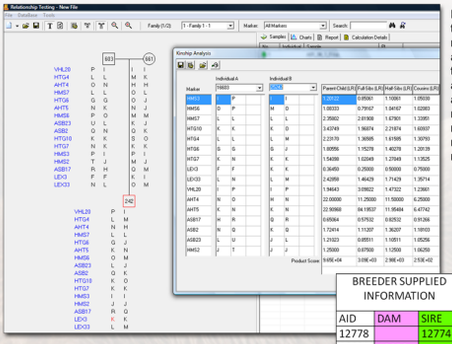

Figure 3: The LEX3 marker is X-linked, causing the appearance of allele conflict in many of the male offspring. The pedigree drawing marks any allele conflict with red. The parameter for the search was set to allow for one mutation, to avoid excluding sires due to LEX3. The genetic analysis settings allows the user to tolerate mutations. The IBD calculation substitutes the mutation rate or user provided value for mistyping, for the allele frequency when mutation tolerance is selected.

	BREEDER SUPPLIED INFORMATION			GENEMARKER RESULTS			
AID	DAM	SIRE		DAM	LR	SIRE	LR
12778		12774		25227	1.14E+06	12774	1.38E+05
16601	16656	16659		16656	1.85+08	16659	7.09e+06
24546	24547			24547	1.03E+06		
24551		24553				24553	1.16+06
24552		24553				24553	1.10E+06
24553	24550			24550	1.18E+06		
25228	12776	12774		12776	1.81E+04	12774	5.11+E04
25240		12774				12774	1.10E+06
25242	16661	16603		16661	1.34E+06	16603	9.65E+04
25243	25252	25251		25252	5.95E+07	25251	5.22E+07
25250	16614			16614	8.68E+05		

Table 1: GeneMarker Database search was in complete agreement with the parentage records for this herd. The kinship calculations used in the database search indicate high confidence in these relationship calls. Individuals 12778, 24552 and 25242 had one allele conflict at the LEX3 locus with the sire, but this is not an exclusion as the locus is X linked.

CONCLUSIONS

- 100% Concordance between GeneMarker database search parentage and known parentage
- Complete screening of the herd resulted in new parentage information for 61 additional horses that will be useful in establishing a stud book for the population
- Males that are sire to multiple offspring are readily identified (figure 1). This feature has applications in determining breeding dominance in wild populations
- Identification of duplicate samples and their random match probability provides confidence in population density estimates of wild populations (figure 4)
- The database search and kinship calculations are readily performed on data genotyped in GeneMarker and from previously archived genotype .txt files
- Pedigree and molecular kinship information have value in the genetic management of rare breeds and endangered species. GeneMarker is an accurate and intuitive program to provide these analyses.

ACKNOWLEGEMNTS

We would like to thank Dr. Chris Kyle and Smolly Coulson, Trent University, Natural Resources DNA Profiling & Forensics Centre, for use of data used in figures 3 and 4 and collaborating in the development of this application.

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.SI 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

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