

The Relationship Testing Application of GeneMarker® HID: Automated Pedigree and Kinship Analysis with Database Searching Capabilities

Nicole R. Phillips, BS; Rhonda K. Roby, PhD, MPH; Jennifer L. Thomas, MS; Suzanne D. Gonzalez, PhD; John V. Planz, PhD; and Arthur J. Eisenberg, PhD
University of North Texas Center for Human Identification, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107



Introduction

DNA typing of individuals using STR analysis is a powerful tool in relationship testing. With natural and man-made mass disasters, casualties of war, and reports of missing persons, it is necessary to identify remains using reference samples from the missing individuals or their primary biological relatives (1). Additionally, verifying relationships is imperative in parentage and immigration cases. Programs which automate and expedite relationship testing are of interest to the forensic examiner; the integration of these types of programs with genotyping abilities would be of great benefit to the human identification community. GeneMarker® HID STR Human Identity Software (SoftGenetics, LLC, State College, PA) is a program designed for STR analysis (2) which meets the definition and scope of an expert system as defined by the National Institute of Justice's Expert System Tested Project (3). The STR fragment sizing and allele calling in GeneMarker® HID is directly linked to a Relationship Testing application. This module expedites kinship analyses by automating pedigree construction and calculating kinship indices for different relationship levels. Additionally, the program allows users to store profiles in a closed database, in which unknown profiles can be searched for possible biological relatives or personal effects.

Pedigree Analysis

In the Relationship Testing application of GeneMarker® HID, pedigrees can be designated manually, or the Family Group tool can automatically draw pedigree trios. Using this tool, established naming conventions are defined and used to automatically construct pedigrees for alleged paternity trios. The child's icon in the pedigree tree is linked to the electropherograms for the entire trio. Using the STR profiles for each individual, the software confirms Mendelian allele sharing (Figure 1). The software program will also automatically display loci of conflict between a parent and child or between two siblings. If a conflict in allele sharing is detected, the child's pedigree icon and conflicting loci are highlighted in red (Figure 2). If any sample contains loci with questionable data, the pedigree icon is tagged with a question mark, and the unresolved loci are listed.

In addition to paternity trios, the program can facilitate construction of large familial pedigrees. Biological parents and children can be added to any node; deceased individuals can be designated. Figure 3 depicts a family pedigree of 44 individuals, created and analyzed with the Relationship Testing application. The red icons indicate conflicts; in this pedigree, four individuals displayed mutational events.

References

1. Price, et al. 2007. DNA Commission of the ISFG. Recommendations regarding the role of forensic genetics for disaster victim identification. *Int J Leg Med* 131:22.
2. Phillips, et al. GeneMarker® HID v1.71: An expert system for high throughput analysis of single source samples. Manuscript submitted for publication.
3. Roby, et al. 2005. The National Institute of Justice's Expert System Tested Project. Proceedings of the Sixteenth Symposium on Human Identification: 2005, 2005 Nov 4. Graystone (TX), Malabar: Promega Corporation. <http://www.promega.com/resources/conferences/ehi2005/ehi2005.pdf>
4. Brenner: 1997. Synthetic kinship program. *Genetics*, 145:535-542.
5. Bodirac, et al. 2001. CODIS STR loci data from 41 sample populations. *J Forensic Sci*, 46(3): 453-489.
6. Percentage Testing Standards Program Unit. American Association of Blood Banks. Annual report summary for testing in 2003, October 2004. http://www.aabb.org/standards/standards/Paternal_Testing_Standards_ProgramSummary03.pdf

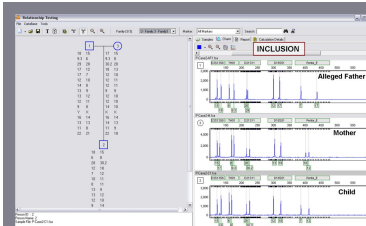


Figure 1: An automated parentage trio where the father is included. This pedigree trio was automatically created using the Family Group tool. Mendelian inheritance of alleles is confirmed automatically and is indicated by the blue pedigree icons. The child's icon is directly linked to the electropherograms of all three individuals.

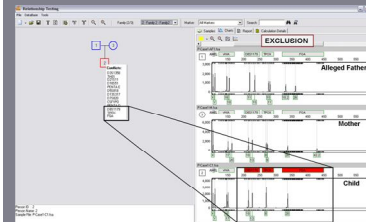


Figure 2: An automated parentage trio where the father is excluded. This figure illustrates a pedigree tree automatically created using the Family Group tool. In contrast to Figure 1, this example illustrates a paternal exclusion.

STR Copyright Agreement
Agreed to this program use possibly limited by the following restrictions: 2003-2004 SoftGenetics, Inc. All rights reserved.
This software is provided under a license to use for non-commercial purposes only. The user is not permitted to copy, distribute, or otherwise disseminate this software or its contents without the express written permission of SoftGenetics, Inc.
Contact Information:
SoftGenetics, Inc., 1400 University Ave., State College, PA 16801-1000, USA
Phone: 717-339-9300
Fax: 717-339-9301
E-mail: info@softgenetics.com

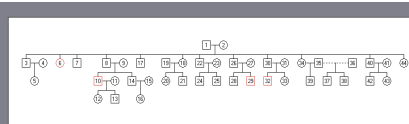


Figure 3: A large pedigree created and analyzed with the Relationship Testing application. Using this application, large pedigrees can also be constructed. Inconsistent allele sharing is highlighted with a red symbol on the pedigree. Kinship indices and probabilities for various relationship levels are automatically calculated between any two individuals using the Kinship Analysis tool (shown in Figure 4). In this example, four individuals display mutational events.

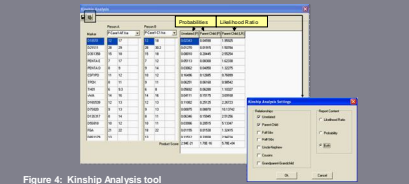


Figure 4: Kinship Analysis tool. The Kinship Analysis tool allows the user to select the sample file of any two individuals (Person A and Person B), and the software automatically reports the likelihood ratios and/or probabilities of their a possible relationship levels. The highlighted button allows the user to change the settings for analysis in the pop-up dialog box.

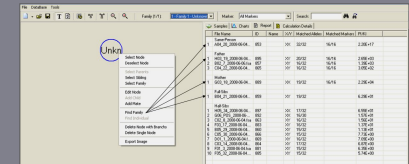


Figure 5: Database searching capabilities. The user can add reference profiles to a closed database. The Find Family feature initiates statistical analyses between the allele profile of the selected individual and the profiles in the database. In this example, the 44 family members from Figure 3 were previously entered into the database. Individual 20 from Figure 3 has been designated as Unknown (blue). The report shows above is returned, listing all matched samples first followed by possible primary relatives. The results are ranked by likelihood ratio. The report indicates a correct identification of the individual (Same-Person), father, mother, and sibling (Full-Sibs).

Kinship Analysis

The Relationship Testing application also includes a Kinship Analysis tool. This module uses STR allele frequencies to determine the probability of observing the genetic results from two individuals for a designated relationship versus the probability of observing the same genetic results under the scenario that the two individuals are unrelated. A kinship index is the likelihood ratio of these two probabilities. Statistical analyses (4) are automated with GeneMarker® HID, reporting the probabilities and kinship indices for different relationship levels. These statistics are calculated using allele frequencies from major U.S. populations (5) and mutation rates specified by the American Association of Blood Banks (6). The different population sets used for allele frequencies can be selected (e.g., Caucasian, African American). Figure 4 depicts the Kinship Analysis tool and the Kinship Analysis Settings menu. Person A is the alleged father and Person B is the child. The alleles for both individuals are listed, and the statistical analyses for the proposed parent-child relationship are shown. The display is set to report the probability of observing these genetic results under the scenario that the two are Unrelated and the probability of observing the genetic data results under the scenario that the two individuals are Parent-Child. In this example, the genetic results are approximately 57,800 times more likely under the scenario that the two individuals are parent-child as opposed to the scenario that the two individuals are unrelated.

Databasing Capabilities

An additional feature of this software program is that it enables the user to create a database of STR profiles, such as those assembled for Missing Persons Programs at the University of North Texas Center for Human Identification for the identification of missing persons and victims of mass disasters. When identifying the victims of disasters, it is recommended that a centralized database be used for data comparisons. Additionally, it is recommended that data be imported electronically to prevent human error (1). GeneMarker® HID allows for seamless integration of profiling and database searching. Once reference profiles are entered, a search can be conducted to rank likelihood ratios of duplicate and potentially related individuals (Figure 5). The program is also capable of uploading profiles to the database in the form of .cmf files.

Conclusion

The Relationship Testing application in GeneMarker® HID offers many user-friendly features which expedite the process of kinship testing. Pedigrees of any size can be easily drawn and automatically analyzed for consistent allele sharing between the defined individuals. Kinship statistics are automated using the Kinship Analysis tool. Users can create a closed database of reference profiles and search against the unknown sample profile. This database searching capability is a powerful tool for disaster victim identification and/or missing persons identification. Since the software directly links the STR analysis to the databasing feature, human error during data entry is eliminated and time is saved on the data importing and formatting process. Additionally, pedigree trees and electropherograms can be exported as a bitmap images for incorporation into other documents. All reports can be easily saved and exported as text files. Field testing of software packages such as GeneMarker® HID may help identify additional tools (e.g., Kinship Analysis tool, Family Group tool, ...) which may improve our ability to identify missing persons and human remains.