Creative Education
2011. Vol. 2, No. 1, 18-21

Scientific Research
DOI:10.4236/ce.2011.21003

Large Scale Simulation for Education in Forensic DNA Science

Jason M. Kinser
Department of Bioinformatics and Computational Biology, George Mason University, Fairfax, USA
Email: jkinser@gmu.edu
Received January 12th, 2011; revised February 24th, 2011; accepted February 28th, 2011.

Forensic science education is a rapidly expanding field with several universities adding degrees in many forensic science disciplines. Concurrently, with this expansion is a new push for forensic science education in the secondary schools. This generation of students is also very adept at computer generated environments. The logical progression is therefore to provide students and instructors with a simulated environment to immerse students into forensic science investigations. The Island of Tir Ebensëa is a developing system that generates a large scale population and forensic scenarios and places the students as the investigators. Students are provided with a scenario and then generate queries to gain information about the people involved in the case. They can then draw conclusions about the scenario and compare these conclusions to the known answer. The simulation is available to educational institutions.

Keywords: Web-Based Education, Forensic Science, DNA

Introduction

Forensic science has become a very popular field in education within the last few years. One popular explanation is the advent of television shows such as CSI. Another contributing factor though is the recent development in the science itself. Modern forensic science technology has evolved rapidly in the past decade. Furthermore, the employment future in the field is bright with a predicted increase of 20% in jobs before 2018 (Bureau of Labor Statistics, 2011; ForensicScience.net, 2011; Jaroch; 2011). Colleges and universities have responded to this need as noted by a recent survey which indicates that the US now has over 70 degree programs in this field (Tebbett, Wielbo, & Khey, 2007). Likewise, opportunities for secondary school students are increasing in the terms of summer camps, specialty programs and school clubs.

This new generation of students is also very adept in using computerized social environments with the onset of social networks and simulation games. Many studies have indicated that learning experience for secondary school students is enhanced through the use of virtual environments and simulations (Akpan & Andre, 2000; Choi & Gennaro, 1987; Geban, Askar & Ozkan, 1992; Lewis, Stern & Linn) with some studies measuring improved performances (Huppert, Lomask & Lazarowitz, 2002; Mintz 1993; Willing 1988).

In the study of forensic DNA science it is necessary for students to understand the statistics of large populations as well as the inheritance properties of DNA profiles. This type of analysis is well suited for a large scale simulation which is the foundation of a project named Tir Ebensëa. The simulation provides students with forensic scenarios and several portals through which their inquiries can reveal more information about the case. Students then use spreadsheets to reach conclusions of the scenarios which can be compared to the known answers. The simulation is available to educational institutions by request to the author.

Simulation Requirements

The study of human identification through DNA profiles requires several components. The two major components are the statistical analysis of large populations and the understanding of the DNA profiles including their inheritance properties. Furthermore, realism in an investigation must include several other complicated factors. The simulation incorporates many of these properties to provide realistic scenarios for the students.

DNA Profiles

Currently, a human DNA profile used in court cases may consist of three components. The first is STR (short tandem repeats). In the nuclear DNA there are many loci in which a small DNA pattern repeats multiple times. A forensic profile identifies alleles by the number of repeats. The nuclear profile includes contributions from both biological parents and thus for each locus there are two values. As an example, a person's profile for a single locus could be the allele pair [9,11] which indicates along one strand of the DNA helix there are 9 repeats and along the other is 11 repeats. However, it is not known which parent donated which value. There is no worldwide standard yet on the set of loci used. The FBI database, named CODIS, uses 13 loci. European countries with their smaller populations often use a fewer number of loci of which some differ from the US set.

Consider a case of a single locus. The parents are [7,9] and [9,10]. A child produced from these two parents will receive one value from each. Therefore a child could be: [7,9], [7,10], [9,9] or [9,10] with equal probabilities. It is quite possible to reconstruct (at least partially) a person's STR profile from the profiles of their immediate relatives. Given a case where the mother is [7,8] and two children are [7,10] and [8,11], it is possible (excluding mutations) to reconstruct the father's profile to be [10,11]. Even though the father's DNA may not be available it is possible to reconstruct (at least in part) the fa-

ther's profile.

The second type of DNA profile use is mitochondrial DNA which is a single stranded DNA loop that exists in multiple copies in the cytoplasm of a cell. The mitochondrial DNA is inherited en masse from the mother. Statistically, it is treated as a single entity rather than a set of values such as in the STR case. The third type of DNA profile is YSTR which is a set of repeats based on the Y-chromosome. The Y chromosome exists only in males and is inherited en masse from the father to the sons. Statistically, it is treated in a manner similar to the mitochondrial DNA.

One of the requirements of the simulation is that each person has a DNA profile with STR, YSTR (males), and mitochondrial DNA. Furthermore, it necessary that people inherit their profiles from their biological parents. There is also a small possibility of some mutations that must be included.

Ethnicity

In real life the distribution of allele sizes (Butler 2005; Marjanovic et al., 2005; Dutta et al., 2002; Nei 1973) varies for each ethnic group. It is possible to provide a probability of a person's ethnicity based upon their DNA profile. Therefore, the simulation must have a variety of ethnic groups each with their own distributions.

Scale

Matching DNA profiles does not prove that a DNA sample comes from a specific person. There are two conclusions that can be drawn from an analysis. The first is an exclusion where it is possible to state that a DNA sample does not come from a specific person. The second is a probability in which the researcher provides a probability of a random person having a particular profile. In some cases, this value can be so ridiculously low that it would take several times the Earth's population before there is a significant chance of a second person having the same profile. While it is not possible to conclude that a DNA sample comes from a specific person it is possible to compute that the probability of two people having the same profile is astronomical.

Before a student can make a statistical calculation from a sample population it is necessary to gather a subsample population. Even though the US population is over 300,000,000 people studies indicate that less than two hundred people are need to provide a statistically relevant sampling of the population (Chakraborty, 1992). In order to replicate this, the simulation is required to have a large population but not nearly on the same scale as the US population.

In order to replicate this type of analysis it is necessary for the simulation to have a population base with different ethnic groups with signature distributions. Furthermore, it is necessary that the size of the population be large.

Biological versus Legal Parentage

The DNA profiles are inherited from biological parents. However, many children live with adults that are legally their parents but not biologically their parents. This occurs through re-marriages, adoptions, or infidelities. In some cases, the investigator may know that the parents are not biologically related and in other cases this information is not volunteered by the family members.

Therefore, the simulation must include mechanisms by which families may be created through means other than biological evolution. It must include marriages, divorces, re-marriages, adoptions, and the occasional tryst.

Scenarios

Another major component of a simulation is the creation of scenarios that are to be solved. In this manner, the students act as the investigators. Scenarios solved by DNA analysis include missing persons, assaults, thefts, scams, law suits, and mass disasters. Each scenario presents students with a small description and then they interact with the simulation to retrieve other information that is necessary. A sample case is presented in a subsequent section.

Requirements

A simulation of forensic DNA cases must include a large population with ethnic variations. This population must have biological relationships in order to replicate inheritance, but it also must have mechanisms by which these biological relationships are disconnected. The simulation must also create scenarios suitable for students educational and maturity levels.

The Tir Ebensëa Simulation

With the requirements in hand the simulation named The Island of Tir Ebensëa has been created and is available to educational institutions. The simulation provides a large scale population, portals for inquiry, and forensic scenarios with solutions.

The Island of Tir Ebensëa

The simulation is based upon a theoretical island inhabited by four ethnic groups. The population occupies five cities as well as the country-side. Since DNA profile statistics are shown to be sensitive to regions and ethnic groups, four of the five cities contain a majority of one of the ethnic groups. This allows students to study profile distributions for global, local, ethnic and/or chronological populations.

The population is evolving at a rate of 10 years per school semester. Simulants (people in the simulation) age, die, marry, give birth, etc. during the course of a semester. A recent sample from the island from year 1627 indicates that there have been a total of 118,000 people of which 24,122 are currently living. There are four ethnic groups which do marry across racial boundaries at small rates and there are currently 1759 different surnames. Figure 1 shows the distribution of ages of the current living population. There are two spikes for the younger ages due to two recent immigration influxes.

The Simulants

Each simulant in the population contains a personality which includes the propensity to commit specific criminal acts. Figure 2 displays the distribution of one of the personality factors which controls the willingness of an individual to cooperate with the police. Two spikes at the end of the distribution indicate that there are several people that will always or never cooperate. There is a nontrivial portion of the population that may cooperate. Students requesting a DNA sample from an individ-

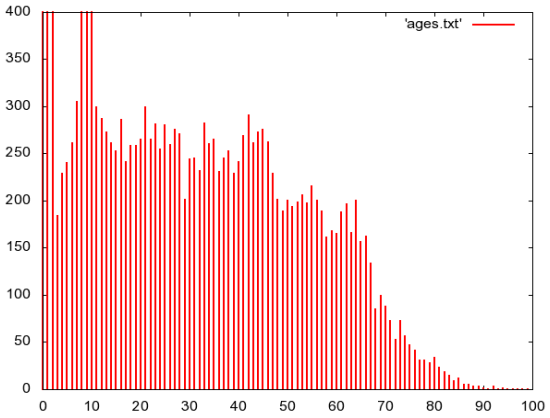

Figure 1.
Distribution of the ages of the living population.

ual may be denied this information because the simulant does not cooperate, but the students will not know if this is a temporary blockade, and requests on different days may produce a different result.

Figure 2 depicts on a log *y* scale the propensity for individuals in the population to commit a specific criminal act (crime type 1). In this sample, more than 10,000 people have absolutely no tendency to commit this act. A few hundred people (to the right of $x = 90$) are quite capable of committing this act. Several other personality and criminal propensities are used to describe the personality of each simulant and some of these qualities are partially inherited. Figure 3 displays the distribution amongst the population of a specific criminal tendency where a larger *x* value indicates a higher propensity to commit this type of crime. Cases are developed based on a person's criminal tendency profile thus creating individuals that are recidivists.

Student Interface

Students are required to have two computer tools in order to participate. The first is access to a web browser to interface with the simulation and the second is a spreadsheet. The web sites provide portals in which the students can submit queries. The results are returned as grids which they copy into their spreadsheets. Example sheets are available to demonstrate the methods in which a spreadsheet can be used to complete the computation. Currently, the material that the student turns in to the instructor is a small report and the spreadsheet.

Initially, students request a population sampling which they use to create their base profile distributions. A tool is provided to create these tables from raw data. Students store this information in a spreadsheet which will be used in almost all scenarios. This process is performed just once. When students receive a scenario they create a new spreadsheet file and store results from their queries. They also create a copy of the sheets from the population sample and add a few cells to compute the probabilities.

While these steps could be automated, the spreadsheet method provides a better teaching tool. Students can see the statistics and how they are created and combined to provide a solution. Students need to have a basic knowledge of statistics (averages and standard deviations) as well as an introduction to a few tools used in the forensic industry (Hardy-Weinberg and upper bounds). These formulae are well within the educational level of science-minded middle school students.

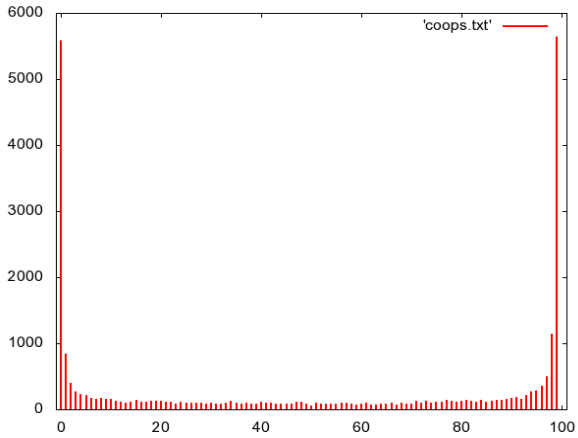

Figure 2.
Distribution of cooperation factors.

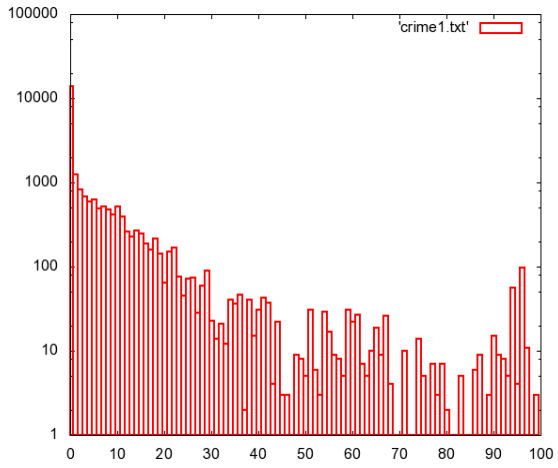

Figure 3.
Distribution or propensities to commit crime type 1.

Students formulate their conclusions which may include exclusions or probabilities. Then the actual solution to the scenario is made available through their instructor. One of the advantages of using simulation data over real world cases is that the solution is definitely known.

Finally, tracking software has been installed to follow the line of inquiry by each student. Information that is easily obtained from this includes the number of queries and the simulants that are being investigated by the students. A simple argument is that students that achieve results with a minimal number of queries have performed a better investigation than students that generate unnecessary queries. However, there are complicating factors that come with an evolving population. For example, queries performed on different days may provide different results since some of the simulants may have died or been placed in the jail. So, the number of inquiries is deemed to be important but not the only metric.

Sample Case

This section presents a case recently used that indicates the type of logic that students will need in order to solve a scenario. In this case, the victim was a discovered body in the forest. Evidence provided to the students was that the victim was an adult male, the ethnic group was identified, and the DNA profile of the victim was identified.

The following steps were the ones necessary for the proper solution.

1) Request a list of missing persons from the Missing Persons Bureau. Exclude from consideration all those that were not adult males.

2) Prioritize the remaining persons according to location and ethnicity.

3) For each person on the list, contact their immediate families and request DNA samples.

4) Exclude from the candidate list those whose DNA profiles had several mismatches with the victim.

In this case, only one male (Stanton Updegraff) survived the previous pruning steps. The following steps were used to confirm the identity of the victim.

5) Determine that there were inconsistencies with the DNA of the wife (Kesha) and three children. From this analysis the students conclude that Kesha is not the biological mother of the children.

6) Through queries to other agencies students gather birth records and marriage records. From this they learn that Stanton was previously married to Serena and that the birth of the three children was during this first marriage.

7) Use Serena and the three children to reconstruct Stanton's DNA profile. In the initial analysis there are some inconsistencies in the reconstruction and the reconstructed profile does not match the victim. The early conclusion is that the victim is not Stanton. However,…

8) The inconsistencies trigger students to realize that one (or more) of the children has a different biological father. Using Y-chromosome information students conclude that the two sons have the same biological father and that the Y data matches the victim. Therefore, they consider a reconstruction without the daughter.

9) The new reconstruction shows no mismatches between Stanton and the victim.

10) Students then use statistical tools to compute the probability that a random person could have Stanton's reconstructed profile. This leads them to conclude that there is an extremely high probability that the victim is Stanton. This is the correct answer.

This case requires the students to use several tools. Students will need to be able to reconstruct DNA profiles from relatives, use Hardy-Weinberg statistics to compute probabilities, and most importantly to understand the evidence. Twice in this case students would have to understand that the evidence indicates that other people are involved in the case (first wife and another male partner). These latter two conclusions are not derived from computer tools but solely from the student's ability to understand the evidence before them.

Final Comments

The current version of the simulation is Tir3 with two new versions in the pipeline that will add other types of forensic evidence (other than DNA) and new environments. Instructors may access the simulation through a request through to the author. Access is currently controlled but not highly restrictive. Instructors wishing to participate in this project should contact the author. Sample cases of the simulation are found on the accompanying web site: http://binf.gmu.edu/kinser/fdna09/tirsimulation/.

References

Akpan, J. P., & Andre, T. (2000). Using a computer simulation before dissection to help students learn anatomy. *Journal of Computers in Mathematics and Science Teaching, 19*, 297-313.

Bureau of Labor Statistics, (2010). *Occupational outlook handbook*, (11th ed.). (accessed Jan. 11, 2011) http://www.bls.gov/oco/ocos115.htm.

Butler, J. M. (2005). *Forensic DNA typing: Biology, technology, and genetics of STR markers* (2nd ed.). London: Academic Press.

Chakraborty, R. (1992). Sample size requirements for addressing the population genetic issues of forensic Use of DNA typing. *Human Biology, 6*, 141-159.

Choi, B., & Gennaro, E. (1987). The effectiveness of using computer simulated experiments on junior high students' understanding of the volume displacement concept. *Journal of Research in Science Teaching, 24*, 539-552. doi:10.1002/tea.3660240604

Duda, R., Reddy, B. M., Chattopadhyay, P., Hasyap, V. K., & Sun, G., Deka, R. (2002). Patterns of genetic diversity at the nine forensically approved STR loci in the Indian populations. *Human Biology, 74*, 34-39.

ForensicScience.net (accessed Jan. 11, 2011). http://www. forensicscience.net/crime-scene-examiners.

Geban, O., Askar, P., & Ozkan, I. (1992). Effects of computer simulations and problem-solving approaches on high school students. *Journal of Educational Research, 86*, 5-10. doi:10.1080/00220671.1992.9941821

Huppert, J., Lomask, S. M., & Lazarowitz, R. (2002). Computer simulations in the high school: Students' cognitive stages, science process skills and academic achievement in microbiology. *International Journal of Science Education, 24*, 803-821. doi:10.1080/09500690110049150

Jaroch, L., Forensic Science Careers (accessed Jan.11, 2011). http://www.all-about-forensic-science.com/forensic-science-careers.html.

Lewis, E. L., Stern, J. L., & Linn, M. C. (1993). The effect of computer simulations on introductory thermodynamics understanding. *Educational Technology, 33*, 445-458.

Marjanovic, D., L. Kapur, N. Pojskic, & R. Hadziselimovic, (2005). DNA diversity in the studies of genetic distance among isolated populations in Bosnia. *Human Evolution, 20*, 157-166. doi:10.1007/BF02438733

Mintz, R. (1993). Computerized simulations as an inquiry tool. *School Science and Mathematics, 93*, 76-80. doi:10.1111/j.1949-8594.1993.tb12198.x

Nei, M. (1973). Analysis of gene diversity in subdivided populations, *Proceedings of the National Academy of Sciences USA, 70, Part I*, 3321-3323. doi:10.1073/pnas.70.12.3321

Tebbett, I. R., Wielbo, D. & Khey, D. (2007, summer). *The forensic examiner*.

Willing, K. R. (1988). Computer simulations: Activating content reading. *Journal of Reading, 31*, 400-409.