

# Real-Time DNA Investigation

Mark W. Perlin  
Cybergenetics, Pittsburgh, PA

November 4, 2005

*In the Proceedings of Promega's  
Sixteenth International Symposium on Human Identification*

Cybergenetics © 2005



Contact information:

Mark W. Perlin, PhD, MD, PhD  
Chief Executive Officer  
Cybergenetics  
160 North Craig Street  
Suite 210  
Pittsburgh, PA 15213  
USA  
(412) 683-3004  
(412) 683-3005 FAX  
[perlin@cybgen.com](mailto:perlin@cybgen.com)

## Abstract

DNA technology can connect a crime scene profile to a DNA databank reference profile, thereby identifying a previously unknown suspect to the crime. When this DNA matching process is driven by police investigations (as in the United Kingdom), criminals can be identified within days, and then rapidly apprehended. Such a "DNA-led policing" approach can reduce crime by removing active criminals from society before they can commit more crimes.

DNA laboratory automation has introduced robotic batch processing that can transform biological specimens into informative data in less than a day. However, the task of human data review, interpretation, matching and reporting often takes months. To complete DNA investigations within 48 hours, it would be useful to have a computer system that can perform these information processing tasks (and deliver investigative leads) within several hours.

Cybergenetics TrueAllele® System 3 is an automated computer-based system that interprets DNA evidence into profiles, matches these profiles against other profiles, and delivers profile and match results to end-users via Internet. The system can be distributed across multiple locations, with DNA interpretation, match and reporting occurring simultaneously for different cases on different computers that share a coordinating database. Doubling the number of interpretation computers halves the processing time. Once a case's DNA peak information has been uploaded to the TrueAllele database, all downstream processing can proceed automatically without human intervention.

We have been scientifically assessing TrueAllele System 3 in several real-time DNA investigative applications. Useful measures for comparing the computer against current manual systems include speed, accuracy, capacity, information, labor and cost. For property crimes, we have looked at the system's interpretation, match and reporting of crime scene samples against an offender database. As reported previously at the 2003 Promega Symposium, the mixture analysis functionality can rapidly infer profiles from unknown-suspect sexual assault data; these DNA profiles are highly informative (relative to human review) and can be automatically matched against offender databases. We have also applied the system to the World Trade Center (WTC) mass disaster, where the problem has analogous elements of inferring crime scene profiles from the victim remains, forming a reference database of missing person profiles (from effects and relatives), and matching the victim remain profiles against the reference profiles to make connections.

This paper introduces TrueAllele System 3 and describes its distributed Internet architecture for DNA interpretation, matching and reporting. We describe results for our real-time property crime, sexual assault and mass disaster studies. We present assessment measures that characterize how the system produces informative identifications in real time. Applying these measures to STR study data, we show that our fully automated TrueAllele computer system is *1,000 times faster, 1,000 times better and 1,000 times cheaper* than the current computer-assisted human alternative. Real-time DNA investigation can reduce crime, and provide high capacity for eliminating DNA backlogs.

## Table of Contents

Abstract .....	2
DNA Match.....	4
The TrueAllele System.....	4
Property Crime.....	5
Sexual Assault.....	6
NIJ mixture data .....	7
DNA profile uncertainty.....	7
DNA match information.....	8
Conservative human review .....	8
Aggressive human review .....	9
Mass Disaster .....	9
Conclusion .....	10
Acknowledgements.....	10
References .....	11
Figures.....	12
Table.....	18

## DNA Match

DNA identification has primarily been used in United States law enforcement for convicting the guilty and exonerating the innocent. However, large scale DNA efforts have shown that DNA match technology can also prevent crime. In the United Kingdom, which has no appreciable DNA backlog, 45% of property crime DNA profiles match an offender on their national DNA database, leading to a decrease in property crime (1). In the state of Virginia, 40% of the DNA database hits from stranger sexual assaults are to non-violent offender profiles (2). A recent cost-benefit study estimated that rapid DNA processing of all convicted offenders, property crime scenes and sexual assaults would prevent over 300 stranger rapes every day in this country (3).

There are several DNA identification scenarios that work by matching unknown profiles against reference profiles:

Property crime. DNA profiles derived from biological evidence at a property crime scene are matched against convicted offender reference profiles.

Sexual assault. Profiles inferred from biological mixtures of victim and perpetrator DNA are matched against a convicted offender reference profiles.

Mass disaster. Victim remain profiles are matched against personal effect profiles, as well as the DNA profiles of relatives.

These disparate DNA identification problems can all be addressed by a single architecture that infers DNA profiles and matches them (Figure 1). TrueAllele System 3 is an intelligent computer system that fully automates these DNA inference and match tasks.

## The TrueAllele System

The first TrueAllele automated STR genotyping paper was published in 1994, and focused on stutter deconvolution, a method for accurately calling alleles in the presence of PCR artifacts (4). That year, the first TrueAllele technology patents were filed (six have issued so far) (5, 6) and licensing discussions were initiated for genetic applications. Also in 1994, a National Institutes of Health grant was funded for applying this computer-based genotyping technology to automated STR diagnosis of neuromuscular degenerative disease (7).

In the year 2000, the UK Forensics Science Service (FSS) licensed the TrueAllele technology from Cybergenetics for forensic databank applications in England (8). The FSS TrueAllele deployment reduced their turnaround time from two weeks to eight hours, their allele calling error from about one per thousand to about one per million, and the number of people involved in databank data review from close to 100 to under 10 (9). More recently, the FSS expanded their TrueAllele technology license to include crime scene evidence (10). In the United States, TrueAllele System 2 has been described for DNA databank applications (11, 12) and validated by the New York State Police (13).

Also in the year 2000, Cybergenetics began work on TrueAllele System 3, an automated DNA data review system for casework applications. While the initial emphasis was on DNA mixture deconvolution (14), the subsequent effort focused on developing an automated DNA interpretation and matching computer system based on modern mathematics and statistical theory (15, 16) that could fully address all casework situations.

In this paper, we focus on the TrueAllele casework situations of property crime, sexual assault and mass

disaster. Our goal throughout this massive research and development effort has been to create a computer system that is faster, better, and cheaper than people performing comparable data review. We present comparisons here between computers and humans for these three performance measures.

TrueAllele System 3 can be modularly configured in many ways. Since a standalone system has over 1,000 times the DNA review capacity of a typical laboratory, Cybergenetics provides System 3 in one form as a secure, centralized Internet processing resource. A schematic of the system, which includes the Internet-based architecture, is shown (Figure 2):

- The *TrueAllele Analysis* process begins the automated data review of DNA sequencer data that produces quality-checked quantitated peaks. For each sequencer run, a user spends one or two minutes checking the computer's quality and control assessment. After this quick human review, the data flow up to the TrueAllele database, and there is no further human intervention in the computer review process.
- The *TrueAllele Database* is the central repository for all DNA sequencer data and case interpretation requests. A case request (even for one contributor) can involve many STR data lanes. When an interpretation request detects that all of its specified lane data are loaded and available for processing, the request informs the database that is ready for interpretation. By coordinating the interplay between case requests and sequencer data, the TrueAllele database automates role of a human supervisor by properly allocating cases to interpretation computers.
- The *TrueAllele Interpretation* computers infer the DNA profiles for each interpretation request. For example, if the interpretation request is for the review of a databank or reference sample, the interpretation computer will infer and upload one DNA profile to the database. With a two person sexual assault mixture, the interpretation process infers and uploads two DNA profiles, one for each contributor. These interpretation computers run independently and in parallel to one another. Doubling the number of computers doubles the throughput of the system.
- The *TrueAllele Match* computer compares sets of DNA profiles and reports matches. Many match processes can be run at the same time. For example, a crime lab's burglary profiles can be compared against offender profiles, but can also be compared against other crime scene profiles. TrueAllele match processes are also useful for quality assurance and scientific validation.
- The TrueAllele profile and match results can be forwarded from the database to other computer systems, or viewed by people on a web browser via a secure connection. We will show some examples of the TrueAllele System 3 web interface later on in this paper.

Quality assurance is an integral part of TrueAllele System 3 processing. The system learns about and characterizes a laboratory's data through an initial calibration review of one or two hundred DNA samples. From this automated review, the TrueAllele system models the behavior of every marker for many important parameters, including background noise, peak variation, PCR stutter, relative amplification and sizing precision. These models capture the overall quality and reproducibility of the laboratory data, and are used by the computer to interpret data (17). This TrueAllele feature can automate much of DNA laboratory validation, as well as ongoing quality control.

## Property Crime

A DNA match process for property crime compares (a) profiles derived from a crime scene without a

known suspect against (b) known reference profiles, such as convicted offenders. A DNA profile match then associates a known individual (e.g., the convicted offender) to the unknown crime scene profile, placing a possible perpetrator at the crime scene. The unknown burglary profiles are inferred from DNA obtained from cigarette butt ends, soft drink cans, and other biological specimens left at the scene by the burglar. The collected samples may contain small quantities of (possibly degraded) DNA that produce partial profiles, and mixed DNA samples are often found. Speed can be essential in performing these analyses in order to enable the police to find a criminal, make an arrest, and prevent further serial burglaries. For example, the UK processes about 100,000 burglary samples each year, and requires DNA reporting within five days.

Cybergenetics is working with government agencies, such as the New York State Police (NYSP), to set up real time processing of property crime cases. In one representative scenario, crime scene biological specimens are processed in an automated robotic DNA laboratory, producing STR sequencer data which are uploaded to the TrueAllele System 3 database. For each crime scene, the LIMS computer uploads a case interpretation request to the database. Once a case request has all of its STR data lanes available, a TrueAllele interpretation computer infers DNA profiles from the data. These inferred profiles are uploaded to the TrueAllele database, where the TrueAllele match module automatically compares them against convicted offender profiles. The results of this DNA information process are then forwarded on to forensic DNA scientists, who can share them with the police in order to prevent further crime.

What are the time and cost of this fully automated TrueAllele DNA review process? To answer this question, we reprocessed a set of U.K. property crime cases at our [www.trueallele.net](http://www.trueallele.net) computer center. We used a test system of four interpretation computers to process 64 cases. The TrueAllele system completed its DNA profile generation and matching in 23 minutes, solving at a rate of 3 cases per minute. This annualized capacity of over 1 million cases per year exceeds the current worldwide demand for property crime DNA review.

The TrueAllele efficiency advantages are best understood by comparing them with human productivity. The human review of 1 million cases over one year, with two human reviews per case and analysts producing 200 DNA reviews a year (about one each day), would require 10,000 human DNA analysts. However, the same work can be done by 1 TrueAllele computer system. Therefore, the TrueAllele system is 10,000 times faster than people. Even if a human analyst could review 2,000 DNA cases in one year (one every hour), we can safely say that:

*TrueAllele System 3 is 1,000 times **faster** than human review.*

To assess cost, let us assume that the fully loaded cost (including salary, training, management, computer equipment and software, turnover, retirement and other benefits, and overhead costs) of a qualified North American DNA analyst is about \$100,000. The 10,000 analysts needed to review 1,000,000 property crime cases would then cost \$1,000,000,000. But (based on actual Cybergenetics contracts) a government could license the TrueAllele technology for this volume of property crimes for well under \$1,000,000. Based on this ratio of the \$1 billion human cost relative to the \$1 million TrueAllele cost, we can safely say that:

*TrueAllele System 3 is 1,000 times **cheaper** than human review.*

## **Sexual Assault**

A DNA match process for stranger rape compares (a) unknown suspect DNA profiles inferred from an

intimate swab taken from the rape victim against (b) known reference profiles, such as convicted offenders. A DNA match can then produce a suspect by associating a known criminal with the stranger rape biological specimen DNA profile. Sexual assault specimens are typically comprised of mixtures that contain both perpetrator and victim DNA. Such mixed DNA samples can pose greater challenges for the human DNA analyst because of the underlying complexity of the data. Speed should be essential for stranger rape DNA analysis, since early police intervention can effectively prevent the commission of further violent crimes against women. However, current DNA laboratory analysis and human data review processes are slow, and there remains a backlog of over 100,000 stranger rape cases (18).

Cybergenetics is working with government laboratories to set up real time processing of sexual assault cases, and help eliminate backlogs. For example, Allegheny County in the state of Pennsylvania will use the TrueAllele technology to screen the backlog rape cases that they send out for vendor DNA analysis. Some DNA laboratories have a critical shortage of trained human DNA analysts to perform the NDIS-mandated dual review. The TrueAllele computer can preprocess the data from every case, and provide a complete solution prior to human review. By having a reliable electronic calculator initially compute genetic profiles and mixing weights, a human review that uses this pre-analyzed case information – and checks (rather than computes) the results – can reduce its review time by a factor of 10.

Which is better, conventional human data review, or automated TrueAllele computer interpretation? The answer lies in determining which system can reliably extract more match information from the same data. In this section, we describe the results of interpretation experiments which address this question.

### ***NIJ mixture data***

The mock sexual assault data were gathered in a National Institute of Justice (NIJ) funded automated mixture analysis study involving ten DNA laboratories. The underlying study design and data description were previously presented (9). In summary, the DNA of two different pairs of individuals with known profiles were examined at five different mixture ratios (1:9, 3:7, 5:5, 7:3, 9:1) and at four different dilutions (1, 1/2, 1/4, 1/8 ng), yielding two person mixture data over 40 experiments. The DNA templates were provided by the National Institute of Standards and Technology (NIST), and each participating laboratory followed Cybergenetics study protocol using their own STR chemistry and DNA sequencer.

Cybergenetics automated interpretation was conducted at its [www.trueallele.net](http://www.trueallele.net) processing center using a statistical extension of its linear mixture analysis methods (14) that infers DNA profiles and mixing weights. The DNA profile and match results for the study are publicly viewable at that website. The results for a 30% unknown contributor are shown (Figure 3).

### ***DNA profile uncertainty***

Forensic DNA scientists are used to dealing with profile uncertainty. It is important to have a proper scientific representation of this uncertainty in order to preserve DNA match information. In the field of genetics, uncertain information is typically represented using probabilities.

Consider the case of genetic inheritance (Figure 4.a). In this example, the father has the heterozygous genotype at locus of [a b], while the mother has the homozygous genotype [c c]. There are two possible genotypes (assuming strict Mendelian inheritance) for the offspring, namely [a c] and [b c]. Each of these two genotype possibilities has a probability of 1/2. A scientist would not discard this genetic information just because there is not a unique solution. Rather, the common scientific practice is to report and use these probabilities in genetic calculations (19).

Now consider the data ambiguity shown in the peak imbalance example (Figure 4.b). This data ambiguity leads to two possible genotype interpretations: [a b] and [b b]. Suppose that the laboratory modeling and quantitative peak data lead to a scientific probability determination of 2/3 for the heterozygote [a b], and of 1/3 for the homozygote [b b]. One could choose to throw out this information because the profile is ambiguous. However, best scientific practice would incorporate methods that use these probabilities to preserve the best possible match information for human identification. This profile probability approach is used in TrueAllele System 3 interpretation.

### ***DNA match information***

DNA match strength is the ratio of the probability that a specific match has occurred to the probability of a random match. For example, with a clean single source DNA profiles, the match strength is equivalent to a random person probability. When analyzing DNA mixtures, the match strength is determined for one of the contributors.

The match strength probability ratio can lead to very large numbers. With modern STR chemistry, human identification values as large as a billion billion, or  $10^{18}$  (a 1 followed by 18 0's) are often seen. It is therefore more convenient to report the logarithm of the match strength using the value in the exponent. For example a match strength of a billion billion (or  $10^{18}$ ) would be reported as 18. The match strengths of the TrueAllele interpretation of Cybergenetics one nanogram NIST mock cases are shown (Table 1).

### ***Conservative human review***

Many government laboratories follow conservative human review interpretation guidelines in order to avoid overcalling the DNA profile results. With these conservative guidelines, at each locus 0, 1, or 2 alleles are reported out. In our uncertain data example of peak imbalance (Figure 4.b), an analyst might designate one allele as having the value 'b', and let the second allele assume any value (i.e., a DNA match wildcard). Although this conservative approach loses useful DNA match information (9), it does help ensure a consistent human review.

The NYSP DNA lab used this conservative approach in their review of sexual assault cases. In our DNA interpretation comparison study, NYSP DNA analysts conducted dual human review of the ten 1 ng mock NIST sexual assault cases generated in their laboratory. Cybergenetics applied TrueAllele System 3 to the same data. The bar chart comparison (Figure 5.a) shows the dual human review for each case as the first two bars (blue and green), and the computer review as the third bar (red). The vertical axis gives the match strength against the known correct profile, expressed in logarithmic units. Note that (although not identical) the two human reviews are quite consistent with each other.

The bar chart shows that, overall, the computer is extracting more information from the same data than the human review. This improvement is made more precise by examining the ratio of the computer review match strength to the human review average match strength (Figure 5.b). In logarithmic units, the mean of this ratio is 4.29, with a 95% confidence interval of [2.34, 6.24]. Since  $10^{4.29}$  is 19,361, this comparison shows that (on average) TrueAllele System 3's scientific computer review extracts 20,000 times the information from the same data as conventional conservative human review. This result supports the hypothesis that:

*TrueAllele System 3 is 1,000 times **better** than human review.*



### ***Aggressive human review***

Some laboratories, particularly in the private sector, use a more aggressive human review of the DNA data. Their goal is to preserve DNA profile information that can be used for a more specific match. This more aggressive approach tries ruling out unlikely allele combinations, and reports out a list of feasible allele pairs at a locus. In our peak imbalance example of data uncertainty (Figure 4.b), there are two feasible allele calls: [a b] and [b b]. This result is more specific than the conservative human review above, where all combinations of allele b with any other allele were included; in particular, it excludes the unlikely combination [a a]. However, even this aggressive human review tends to lose more useful DNA match information than a scientific review that incorporates genotype probabilities (9).

The Orchid Cellmark vendor laboratory used this aggressive approach in their review of sexual assault cases. Interestingly though, recent federal guidelines mandate government employee review, so such private sector reviews are no longer accepted by NDIS for direct upload into the national forensic database. In our DNA interpretation comparison study, Cellmark DNA analysts conducted dual human review of all 40 mock NIST sexual assault cases generated in their laboratory. Cybergenetics applied TrueAllele System 3 to the same data. The bar chart comparison (Figure 6.a) shows the dual human review for each case as the first two bars (blue and green), and the computer review as the third bar (red). The vertical axis gives the match strength against the known correct profile, expressed in logarithmic units.

The bar chart shows that, overall, the computer extracts more information from the same data than even the more aggressive human review. We can quantify this improvement by examining the ratio of the computer review match strength to the human review average match strength (Figure 6.b). In logarithmic units, the mean of this ratio is 2.74, with a 95% confidence interval of [1.90, 3.59]. Since  $10^{2.74}$  is 555, this comparison shows that (on average) TrueAllele System 3's scientific computer review extracts over 500 times the information from the same data as the more aggressive human review. This result supports the hypothesis that:

*TrueAllele System 3 is 1,000 times **better** than human review.*

### **Mass Disaster**

A DNA match process for mass disasters compares (a) profiles derived from unidentified victim remains against (b) known reference profiles derived from missing persons, such as personal effects or family members. A DNA profile match then associates a known individual (e.g., the missing person) to the unidentified victim remains profile, thereby identifying the victim remains. The unknown victim remains profiles are inferred from DNA obtained from a variety of biological specimens that are found at the scene of the mass disaster. The collected victim remains samples may contain small quantities of DNA or highly degraded DNA; these nonoptimal specimens often produce partial profiles. Mixed DNA samples are also found at the scene. Speed should be essential in completing these analyses in order to provide closure to be affected relatives.

Cybergenetics is working with government agencies, such as New York City's Office of the Chief Medical Examiner (OCME), to reanalyze the World Trade Center (WTC) STR data and set up real time processing of mass disaster cases. For the WTC disaster, Cybergenetics has completed the TrueAllele System 3 analysis phase of over 2,700 sequencer runs generated in over 20 different STR production processes. We are currently in the process of inferring DNA profiles from the victim remains, personal effects and family references. When that phase of the project is completed, TrueAllele System 3 will

match the victim remains profiles against the known (personal effect and kinship) reference profiles.

## **Conclusion**

TrueAllele System 3 is a dedicated computerized approach to real-time DNA investigation. The underlying computational methods are based on unique mathematics and statistics that provide a general framework for quality assurance, STR profile interpretation and DNA matching that underlie forensic DNA applications. The real-time TrueAllele architecture is organized around a relational database that permits independent computers to process many cases simultaneously.

In this paper, we focused on the real-time DNA investigation of property crime, sexual assault and mass disaster. Using data from our many collaborative studies, we demonstrated that TrueAllele System 3 is 1,000 times faster, 1,000 times better, and 1,000 times cheaper than the current human alternative.

There is increasing evidence that rapid DNA policing prevents crime. Removing the human DNA review bottleneck could prevent over 100,000 stranger rapes every year. It is now time to let dedicated computer machinery replace people in performing highly computational, time sensitive tasks. By having computers focus on DNA inference, and people focus on the uniquely human aspects of the forensic DNA process, DNA technology can help build a safer society.

## **Acknowledgements**

We would like to thank our scientific collaborators, including Jeff Ban, Robin Cotton, Cecelia Crouse, Barry Duceman, Trevor Howett, Jay Kadane, Margaret Kline and Bob Shaler. Many DNA laboratories participated in the studies, including Albany, NY, Allegheny County, PA, Orchid Cellmark, MD, The UK Forensic Science Service (FSS), Miami Valley, OH, New York, NY, the National Institute of Standards and Technology (NIST), MD, Palm Beach, FL and Richmond, VA.

The New York State Police mock case reviewers included John Brenner, Russell Gettig, Melissa Lee, Maria Mick, Shannon Morris, Urfan Mukhtar, Michael Portzer, Laura Post and Diana Seaburg. The Orchid Cellmark mock case reviewers included Christine Baer, Jason Befus, Julie Black, Paula Clifton, Kathryn Colombo, Lisa Grossweiler, Juliet Harris, Jeff Hickey, Jacki Higgins, Christopher Knickerbocker, Jason Kokoszka, Lewis Maddox, Jennifer Reynolds, Leslie Rosier, Ryan Satcher, Alissa Shofkom, Margaret Terrill, Melissa Thompson, Charlotte Word and Cindy Zimmerman.

This research was supported in part under Award number 2001-IJ-CX-K003 from the Office of Justice Programs, National Institute of Justice, Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice.

## References

1. FSS. Annual Report. Birmingham, UK: The Forensic Science Service (UK); 2005.
2. McCue C, Smith GL, Diehl RL, Dabbs DF, McDonough JJ, Ferrara PB. Why DNA databases should include all felons. *Police Chief* 2001(October):94-100.
3. Wickenheiser R. The business case for using forensic DNA technology to solve and prevent crime. *J. Biolaw & Bus.* 2004;7(3).
4. Perlin MW, Burks MB, Hoop RC, Hoffman EP. Toward fully automated genotyping: allele assignment, pedigree construction, phase determination, and recombination detection in Duchenne muscular dystrophy. *Am. J. Hum. Genet.* 1994;55(4):777-787.
5. Perlin MW. Method and system for genotyping. US Patents; 1996-2004. Report No.: #5,541,067, #5,580,728, #5,876,933, #6,054,268, and #6,750,011.
6. Perlin MW. Method for DNA mixture analysis. US Patent; 2004. Report No.: #6,807,490.
7. Perlin MW, Lancia G, Ng S-K. Toward fully automated genotyping: genotyping microsatellite markers by deconvolution. *Am. J. Hum. Genet.* 1995;57(5):1199-1210.
8. Cybergenetics. Cybergenetics awarded UK National DNA Database contract. Press release; 2000 March 24.
9. Perlin MW. Simple reporting of complex DNA evidence: automated computer interpretation. In: Promega's Fourteenth International Symposium on Human Identification; 2003; Phoenix, AZ; 2003.
10. Cybergenetics. Forensic Science Service expands license for Cybergenetics automated DNA data review technology. Press release; 2004 July 26.
11. Perlin MW. An expert system for scoring DNA database profiles. In: Promega's Eleventh International Symposium on Human Identification; 2000; Biloxi, MS; 2000.
12. Perlin MW, Coffman D, Crouse CA, Konotop F, Ban JD. Automated STR data analysis: validation studies. In: Promega's Twelfth International Symposium on Human Identification; 2001; Biloxi, MS; 2001.
13. Kadash K, Kozlowski BE, Biega LA, Duceman BW. Validation study of the TrueAllele® automated data review system. *Journal of Forensic Sciences* 2004;49(4):1-8.
14. Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. *Journal of Forensic Sciences* 2001;46(6):1372-1377.
15. Gelman A, Carlin JB, Stern HS, Rubin D. *Bayesian Data Analysis*. Boca Raton, FL: Chapman & Hall/CRC, 1995
16. Gilks, Richardson, Spiegelhalter. *Markov Chain Monte Carlo in Practice*: Chapman and Hall, 1996
17. Perlin MW. Automated quality assurance for the forensic DNA laboratory (poster). In: Promega's Sixteenth International Symposium on Human Identification; 2005; Dallas, TX; 2005.
18. Hart SV. *DNA Initiatives*. Washington, DC: National Institute of Justice; 2002 May 14.
19. Lange K, Weeks DE, Boehnke M. Programs for pedigree analysis: MENDEL, FISHER, and dGENE. *Genetic Epidemiology* 1988;5:471-472.

**Figures**

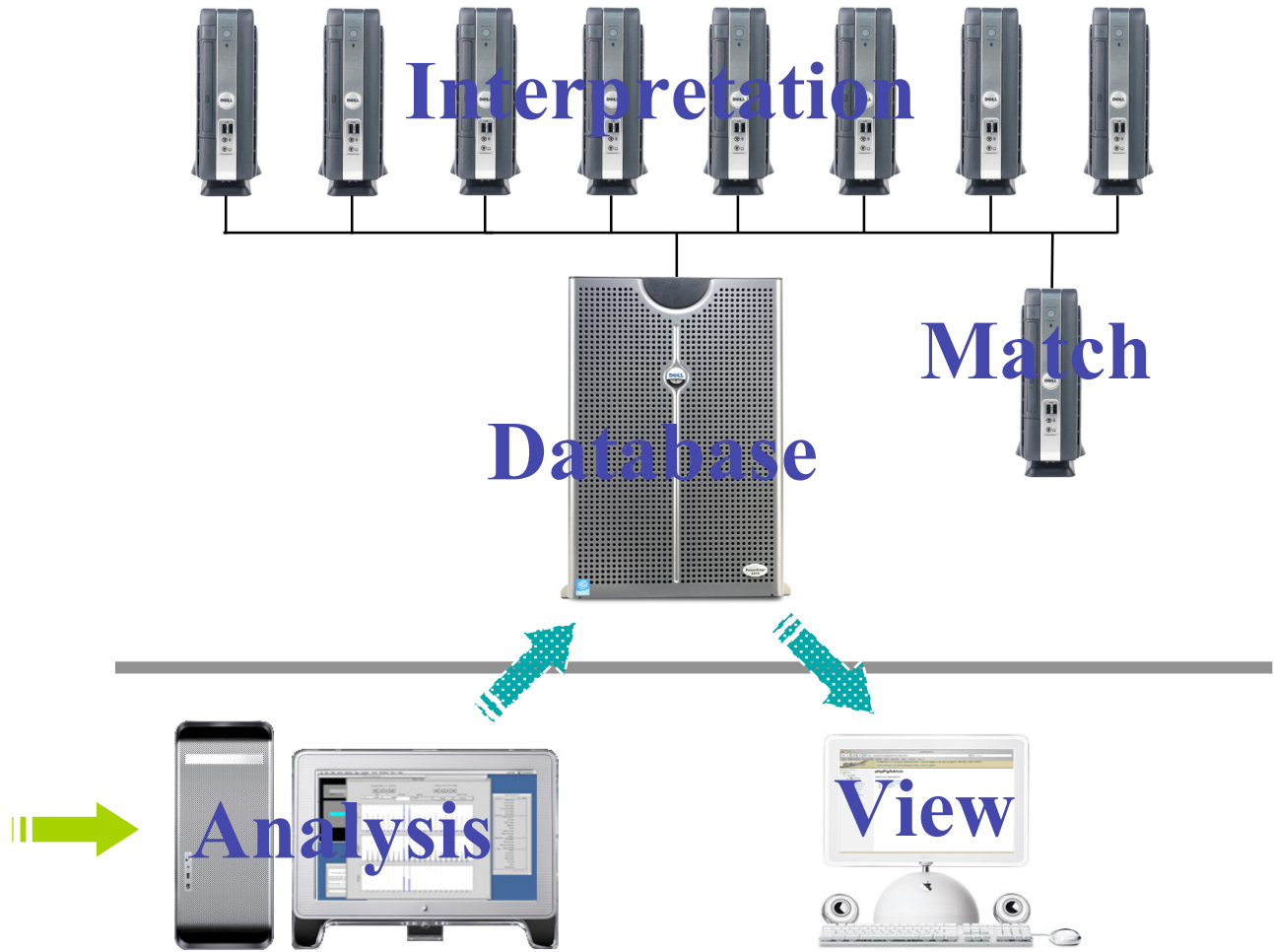
**unknown profiles**



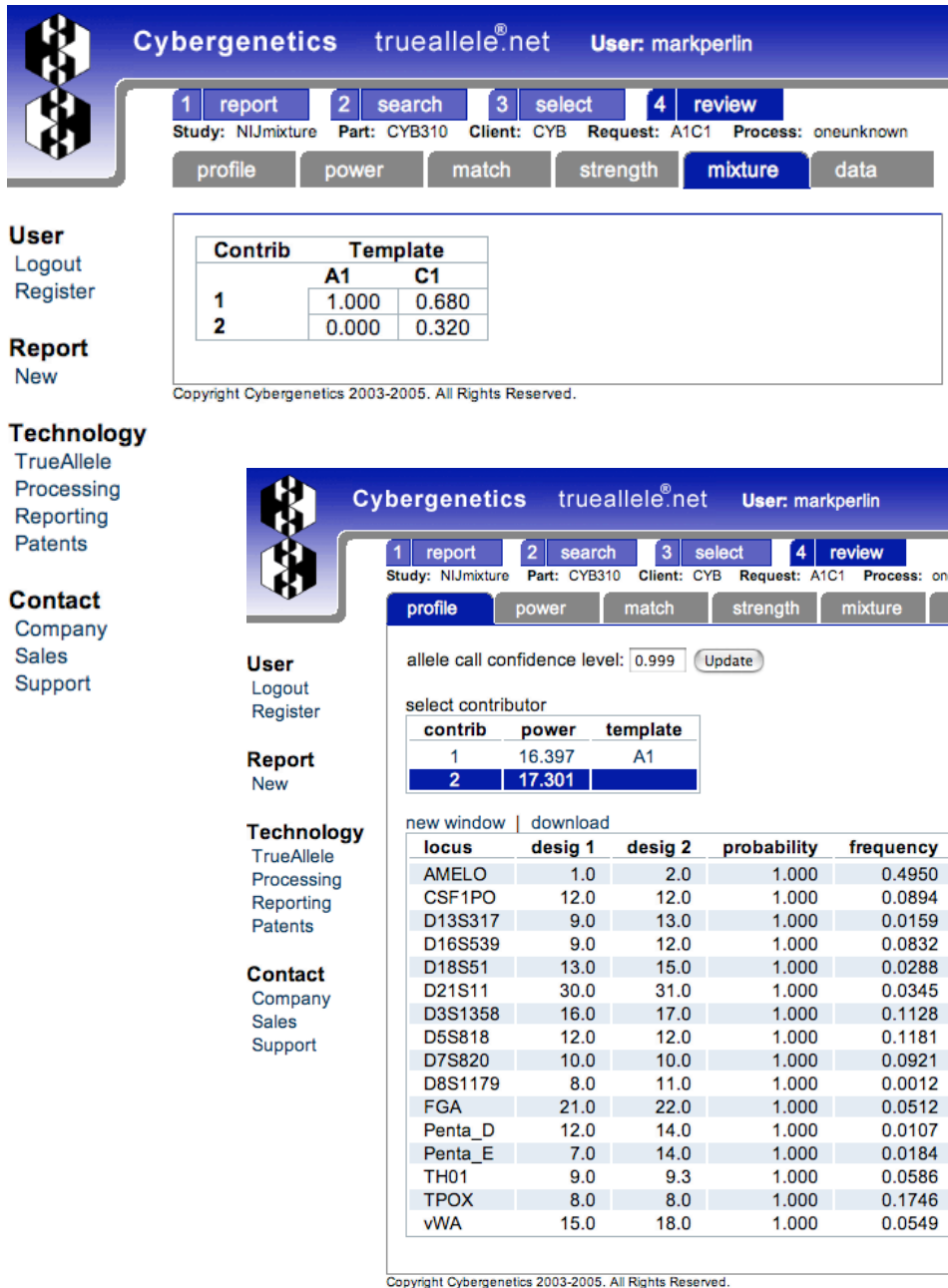
**reference profiles**



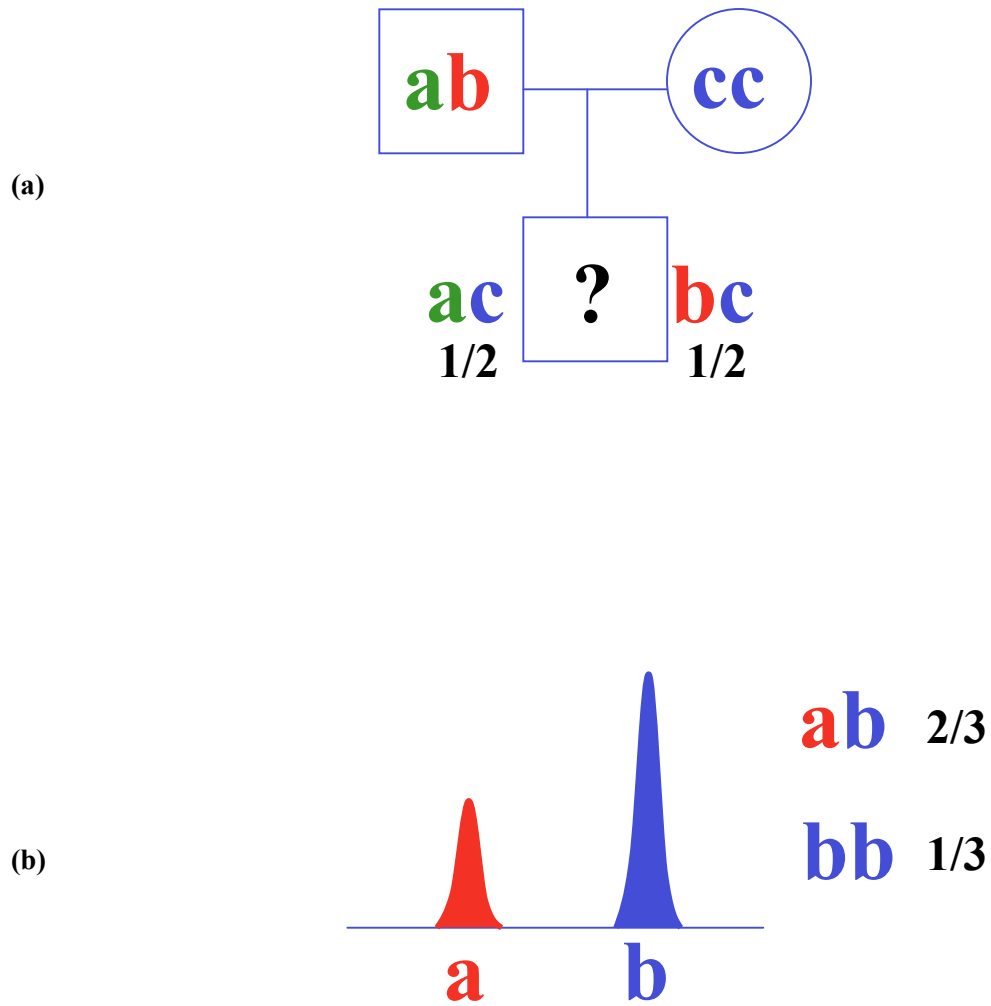
**Figure 1.** A general forensic match framework that compares one set of STR profiles with another set of STR profiles. In this framework, human identification for property crime, sexual assault and mass disaster are all the same scientific problem.



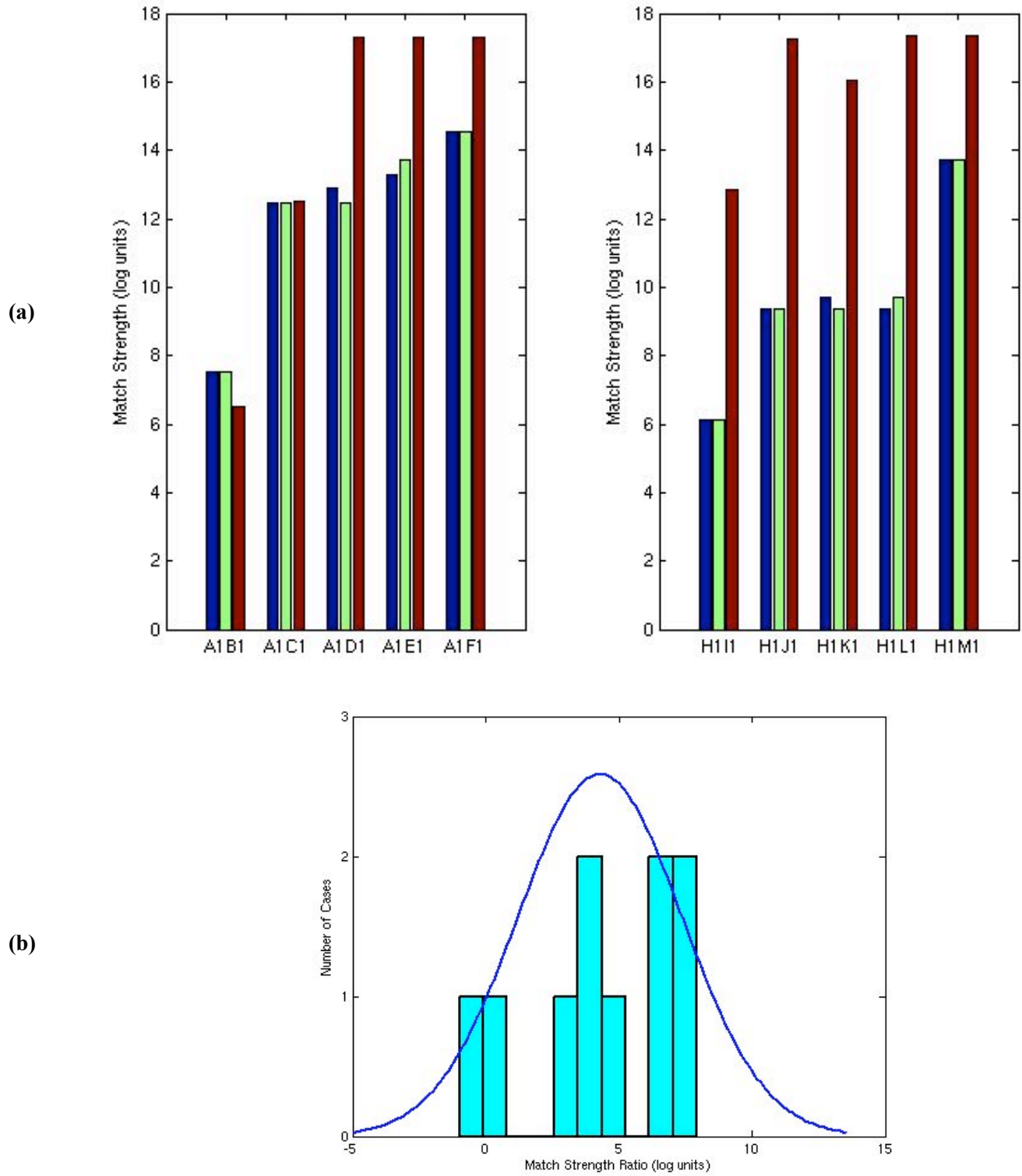
**Figure 2.** The component modules in the TrueAllele System 3 computer architecture. The components above the gray bar can all operate independently on the Internet without any human intervention.



**Figure 3.** Screen snapshots from the publicly accessible www.trueallele.net web site that show the mixture weights of the contributors in each template (top) and the STR profile (bottom).

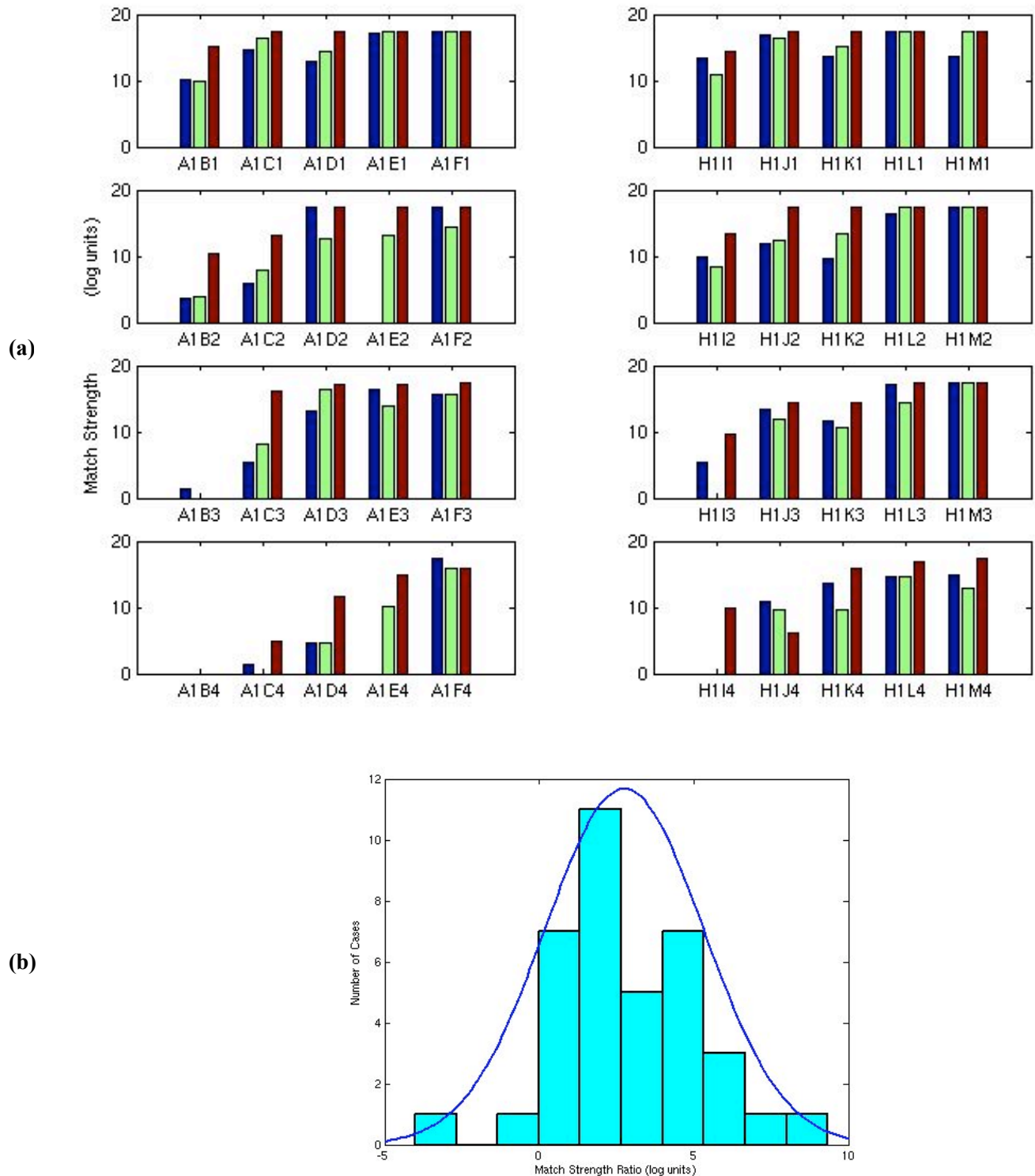


**Figure 4.** (a) Genetic uncertainty produces STR profiles that have probabilities associated with each possible genotype. (b) Data uncertainty can similarly produce STR profiles with probabilities.



**Figure 5.** Conservative human review in a government laboratory. (a) Bar charts showing the case match strength, comparing conservative human review (blue, green) with TrueAllele System 3 computer review (red). Each chart shows results for one pair of contributors, and the cases progress left to right as (10%, 30%, 50%, 70%, 90% unknown). (b) A histogram of the ratio of computer match strength to human match strength (logarithmic units).





**Figure 6.** Aggressive human review in a vendor laboratory. (a) Bar charts showing the case match strength, comparing aggressive human review (blue, green) with TrueAllele System 3 computer review (red). Each row shows progressively more dilute DNA, starting from the top as 1 ng, 1/2 ng, 1/4 ng and 1/8 ng. Each chart shows results for one pair of contributors, and the cases progress left to right as (10%, 30%, 50%, 70%, 90% unknown). (b) A histogram of the ratio of computer match strength to human match strength (logarithmic units).

**Table**

case	unknown	match strength
A1B1	10%	13.79
A1C1	30%	17.30
A1D1	50%	17.30
A1E1	70%	17.25
A1F1	90%	17.30
H1I1	10%	9.42
H1J1	30%	17.14
H1K1	50%	17.35
H1L1	70%	17.35
H1M1	90%	17.35

**Table 1.** For each of the ten 1 ng mock mixture cases, the proportion of the unknown contributor DNA is shown along with the match strength in logarithmic units.