Market Research Matrix for STRmix

Steven P. Myers
Senior Criminalist
DOJ BFS Jan Bashinski DNA Laboratory
steven.myers@doj.ca.gov
510-620-3306

Mixture deconvolution software/systems evaluated:

1. NicheVision/ESR STRmix
2. Cybergenetics TrueAllele Casework
3. Applied Biosystems GeneMapper ID-X
4. Promega FSS-i3
5. NicheVision ArmedXpert
6. Softgenetics GeneMarker HID

Market Research Matrix

Company	Software or System	Criterion 1: Maximum # of contributors for deconvolution	Criterion 2: Automatically consider stutter as possible alleles.	Criterion 3: Employs fully continuous probabilistic assessments, not thresholds, to variable parameters.	Criterion 4: Automatically account for DNA degradation.	Criterion 5: Can be implemented on existing computers.
NicheVision and ESR	STRmix	4	Yes	Yes	Yes	Yes
Cybergenetics	TrueAllele	Unlimited (up to 6 unknown)	Yes	Yes	Yes/No	No
Applied Biosystems	GeneMapper ID-X	2	No	No	No	Yes
Promega	FSS-i3	2	No	No	No	Yes
NicheVision	ArmedXpert	3	No	No	Yes	Yes
Softgenetics	GeneMarker HID	2	No	No	No	Yes

Discussion of the criteria evaluated for market research matrix:

1. Maximum number of unknown contributors for the deconvolution of a mixture.

   In performing DNA testing on sexual assault evidence, it's common to encounter mixtures with just two or three contributors, one of whom is the victim. However, the trend in forensic science has been to expand testing to cases and samples where mixtures of more than three individuals are common (e.g., property crime, "touch" DNA samples, and clothing and gun swabs.) In some laboratories in the Bureau, these types of samples are submitted more frequently than sexual assault evidence. Unfortunately, a criminalist's ability to manually solve for mixture contributors' DNA profiles is inversely correlated to the number of contributor. Software tools are needed.

Only three of the systems examined will perform deconvolutions on mixtures of more than two people. In the other systems, mixtures with more than two contributors will be treated in as having an indeterminate number of contributors, and the population frequency estimate provided might run counter to laboratory policy.

Only STRmix and TrueAllele will deconvolute mixtures of more than three. STRmix can interpret four-person mixtures. Given the limited ability to accurately assess the number of donors when the testing results indicated the presence of more than four individuals, any benefit gained by TrueAllele's larger number of contributor options appears limited.

2. Automatically consider stutter artifacts as possible alleles when determining the DNA profile of potential contributors to a mixture.

   Stutter is the name given to the most common typing artifact encountered in the testing of Short Tandem Repeat (STR) DNA markers. Stutter is a byproduct of the DNA copying process and is a shortened form of the actual DNA (allele) being copied. The amount of stutter created by the copying process has been studied and is largely predictable. However, stutter is actual DNA and otherwise tests identically to full length STR allele results. Therefore, it is possible that a low-level contributor to a DNA mixture might have an allele that overlaps stutter from a higher-level contributor's allele. If the combined relative quantity of allele and stutter is still within the expected range for stutter alone, the presence of the low-level allele would be uncertain. In such cases, the result must be evaluated as possible stutter and as a possible allele. This concept is described and recommended in the "SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories".

   Only two of the systems examined, STRmix and TrueAllele, will perform deconvolutions that assess every peak as a possible allele, including peaks that might otherwise be ignored or edited out as stutter. The work flow of all other systems has you remove possible stutter peaks from consideration prior to performing mixture deconvolution.

3. Employs a fully continuous probabilistic system, rather than thresholds/cutoffs, to testing parameters that are known to vary (e.g., peak balance, stutter levels, mixture proportion.)

   Current testing procedures generally employ thresholds to determine whether something will be definitively identified as occurring (e.g., a particular STR allele is present in the mixture, and there must, then, be a contributor who has that allele in their DNA profile), whether it will definitely be eliminated (e.g., the majority contributor to a mixture must have one particular combination of alleles to the exclusion of all other combinations), or whether it will be treated as possible but not definitive (e.g., the result in the stutter position is below the maximum amount expected for stutter, so it might be all stutter, or it could be some combination of stutter and allele, or it could be all allele.) While this approach is valid, it removes data from the interpretation and provides a weaker answer to the scientific questions we're asking. For example, a result that is close to the maximum amount of expected stutter is more likely to at least partly be due to an allele

than it is to be just stutter. This is because we know from studies that it's rare to observe pure stutter at that maximum level. These issues also arise in the evaluation of allele balance and mixture proportion. Interpretations incorporating the relative probability of certain events are much more powerful at fully assessing the data under various hypotheses. For example, if the hypothesis is that a person was a contributor to the mixture, a threshold approach might include that person's genotype as possible, because it fits the rules of the thresholds. A probabilistic genotyping system, on the other hand, may show that it is possible but highly unlikely a person with that genotype was a contributor to the mixture. The most developed of the probabilistic systems, termed "fully continuous", use most of the data available in DNA testing results, including the relative heights of peaks and not just their present or absence.

Only two of the systems examined, STRmix and TrueAllele, will perform deconvolutions that assess the data in a fully continuous probabilistic manner.

4. Have the ability to automatically account for different levels of DNA degradation for the different contributors.

   Evidence from crime scenes regularly suffer from environmental insults such as heat, humidity, chemical exposure, and microbial action. One result of this is the random breakage of DNA molecules, termed "degradation". Our STR DNA tests require strands of evidence DNA to be of a certain length, and when the DNA is fragmented, the tests will give weaker results inversely proportional to the length of DNA they require (longer DNA requirements = weaker results.) The affects of degradation can lead to poor typing and mixture deconvolution results when using threshold-based systems, because the DNA length-based imbalances may lead to results deemed improbable or impossible. When one mixture contributor's DNA has been degraded to a different level than another's DNA, this situation is exacerbated. In the extreme, miscalls can occur where the majority contributor at one marker is actually the minority contributor at another, longer locus, but they are incorrectly identified as the majority contributor.

   Three of the systems examined, STRmix, TrueAllele, and Armed Xpert will perform deconvolutions that assess the data incorporating a model for degradation, including the allowance that the different contributors' DNA might be degraded to different extremes. STRmix always incorporates this approach. TrueAllele has that assessment available as a user option. Armed Xpert automatically assesses degradation in a module of the software.

5. Can be implemented on existing computers.

   Computer maintenance is a significant burden on any agency. When a system comes pre-installed on a software manufacturer's preferred server or workstation, it may involve hardware manufacturers and/or software platforms that are outside the experience of the laboratory. Installation into the lab's network, and routine maintenance and backup of the system, become more difficult. It is therefore a benefit to have the option of

incorporating new software into the laboratory's existing IT structure (e.g., work stations and/or network).

Except for one of the systems examined, all systems are available as software installations on workstations and/or existing networks. Only TrueAllele requires the purchase of the software and hardware as an integrated unit.