April 19, 2013

Dear Webcast participants (or future video viewers):

We hope that you have recovered from your early morning or late night viewing (if you are not from the U.S. East Coast) of the April 12 DNA mixture interpretation. We wanted to follow-up with some additional information and clarification from last week's NIST webcast.

1) **Certificates of Participation.** Technical leaders should be receiving certificates of participation for all those that they have requested. TLs, if there is someone in your lab that watched the webcast and that you have not yet requested a certificate for, then please contact john.jones@nist.gov.

2) **Availability of EPG Maker software.** Steven Myers kindly supplied a copy of the EPG Maker program, mentioned in John Butler's first talk, in order to make it available for download. The 13 Mb Excel file can be downloaded from the STRBase mixture page: http://www.cstl.nist.gov/strbase/mixture.htm (see the software programs or information section).

3) **Availability of Bruce Heidebrecht's Excel spreadsheets for LR and mixture deconvolution**. We received requests for the Excel spreadsheets that Bruce Heidebrecht has developed for his laboratory's use in 2-person likelihood ratio calculations with the Identifiler loci using the FBI's databases for Caucasians and African Americans. Please email Bruce (Bruce.Heidebrecht@Maryland.gov) if you are interested in obtaining a copy of either program. Of course, laboratories using these programs should perform their own validation studies prior to implementing in their lab.

4) **Worked example clarification**. Bruce Heidebrecht wanted to clarify a small error in slide 38 of his Worked Examples presentation (time 49:40 to 49:52 in the video). He stated, "Even with some amount of stutter present in bin 13, this peak may contain a true sister allele to the allele 13." What should have been said was "Even with some amount of stutter present in bin 13, this peak may contain a true sister allele to the allele 10."

5) **Probabilistic genotyping clarification**. Mike Coble wanted to clarify some comments he made during his Probabilistic Genotyping presentation.

    a. On slide 28, a single replicate of the 100,000 replicates performed by TrueAllele (TA) is presented. The purpose of this slide was to show that the software goes through multiple processes of inferring potential genotypes using biological parameters such as PHR, stutter ratios, the mixture ratio, etc… This is similar to the method explained in slide 26. In slide 28, the inferred genotype of this one replicate was 20,21. As a comparison Mike listed the suspect's genotype (20,22). Mike stated that the software "got it wrong" during the presentation when in fact, the software was simply inferring a potential genotype without knowing the *a priori* genotype of the suspect. As can be seen in slide 29, the inferred genotype predicted most often during the 100,000 replicates does in fact match the suspect's genotype in this case and this information is used in the calculation of the LR.
    b. Also in the STRmix calculation for the LR using the same locus, Mike should have mentioned that the information presented in slide 35 considered different allele frequency databases used by this particular software (consisting of New Zealand populations) as opposed to US allele frequencies (used in TA). This was an unintentional oversight and not an equitable comparison of the results.
    c. Mike would like to reiterate that (in his own opinion) the purpose of **ANY** probabilistic approach should not be focused on simply "getting a bigger number" but to make better use of the data. In this particular example, exclusionary approaches (e.g. CPI) would discard potentially **all** of the data in the profile. Probabilistic methods can make better use of the data by incorporating a probability of drop-out (e.g. the semi-continuous approaches) and/or maximizing the inferred genotypes via simulations (e.g. the fully continuous approaches).
    d. If there are any additional questions or clarifications, please feel free to contact Mike at michael.coble@nist.gov.

6) **2p or not 2p (with PopStats).** Following Bruce Heidebrecht's morning presentation, a question was asked about whether or not PopStats CODIS computer software has a 2p option to make statistical adjustments for possible allele drop-out. Bruce responded that to the best of his knowledge PopStats could not make 2p calculations. Several individuals contacted us and shared that the CODIS 7 Service Pack 2 does have a 2p option.

   a. Bruce has tried the newest version of PopStats on his laboratory's CODIS computer terminal – and obtained the following:

      i. For **single-source samples**, it is an option to select the use of 2p for any locus that has a single allele entered when performing an RMP calculation. For single-source samples, the option to select the use of 2p for a locus becomes disabled once two alleles are entered. Thus, it is not possible to use a 2p calculation when you have alleles P and Q. Thus, it is possible to do a single-source RMP stat using 2pq for heterozygous loci, and either p^2 or 2p for homozygous loci on a locus-by-locus choice by using the "Rec 4.1" check mark.

      ii. For **mixture samples**, there is no option to select use of 2p for either a specific locus, or on a global scale. The "Rec 4.1" option does not exist for mixture sample stats. An example tested involved a D8S1179 mixture with alleles 10, 11, and 12. Assuming that the victim is 11,12, then the requisite allele to the unknown contributor must be allele 10. Using an unrestricted LR stat with no assumption of dropout, the unknown contributor must be 10,10 or 10,11 or 10,12. PopStats produces a LR value for this locus of 19. Using an unrestricted LR stat with the assumption of dropout, the unknown contributor must be 10,F. The LR for this locus is 4 (using a 2p calculation). PopStats will not currently perform this calculation within the mixture portion of the software.

   b. The bottom line is that what was stated in the webcast is accurate – PopStats will not perform a Likelihood Ratio stat that should include the possibility of dropout.

7) **The webcast video materials are now available for viewing.** See http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation-webcast.cfm.

8) **Polling results**. During the NIST DNA Mixture Interpretation Workshop and Webcast on April 12, 2013, a total of 20 poll questions were made available for workshop attendees and webcast viewers to answer. The responses were used as discussion points by the presenters during the event. Responses were received from individual viewers using computers or web browsers on smart phones. These anonymous responses were only presented in aggregate form as seen in this document. We are unable to determine the identity, occupation, or associated agency of the individuals that answered these poll questions. A summary of the SurveyMonkey polling results obtained during the webcast on April 12 have been compiled and are included on the conference website: http://www.nist.gov/oles/forensics/dna-analyst-training-on-mixture-interpretation.cfm. For future viewers of the webcast video, we have left open the poll questions to enable additional data to be collected.

Thank you again for your participation in last week's webcast. If you have additional feedback, we would appreciate hearing from you. Please contact John Paul Jones at john.jones@nist.gov.

Best wishes,

John Paul Jones II for all of the presenters