Assessing the Performance of the SoftGenetics® MaSTR™ Probabilistic Genotyping Software

Michael Adamowicz, Ph.D., Director, Forensic Science Program, Assoc. Professor, University of Nebraska–Lincoln

An initial assessment of the performance of the SoftGenetics® MaSTR™ probabilistic genotyping software was performed by using two previously generated synthetic mixed DNA samples. Both mixtures were created using two contributors. One mixture sample had a contributor ratio of approximately 3:1, demonstrating a fairly clear distinction in the relative peak heights of the two profiles. The other was significantly more challenging, with a contributor ratio of approximately 1:1, and showing little obvious difference in the peak heights of the two profiles. Contributor ratios were determined using qPCR quantification values and then checked and adjusted using electropherogram peak heights across multiple runs before the final mixed amplifications were performed. The MaSTR software results were scored on the accuracy of the number of deconvolved loci and the percent values associated with the deconvolved genotypes using both 5000 and 50,000 MCMC iterations.

The results of the analysis on the mixture with a contributor ratio of approximately 3:1 showed that the MaSTR software correctly deconvoluted genotypes at 13/15 STR loci with both 5000 and 50,000 iterations. The percent values associated with correctly called genotypes were all above 99%, with those performed using 50,000 iterations (e.g. 99.97% for D3S1358) being slightly higher than those performed using 5000 (e.g. 99.80% for D3S1358). The two loci that were not correctly deconvoluted (D7S820 and Penta D) both had peak heights that did not conform to those expected for the mixture of contributors used. The vagaries of amplified peak

heights are recognized and these results demonstrate the need for caution when interpreting any mixture sample. Also, specific stutter data was not used in these analyses, and its addition could impact the results.

The results of analysis on the mixture with a contributor ratio of approximately 1:1 showed that the MaSTR software correctly deconvoluted genotypes at 9/15 loci using both 5000 iterations and 50,000 iterations. The percent values associated with correctly called genotypes were much more variable than those seen in the 3:1 mixture. Some were very high, such as 100% for D7S820 (5000 iterations) to 51.08% for vWA (5000 iterations). As expected, the percent values also changed slightly when using different numbers of MCMC iterations, however there was no clear trend as there was with the 3:1 mixture sample. Sometimes the higher percent was observed with the 5000 iteration run, however as with the 3:1 sample analysis results, these differences were small. This indicates that for challenging mixtures, such as those with contributor ratios close to equal, additional MCMC iterations can be beneficial, however the percentage values should be expected to change somewhat from analysis run to analysis run.

To test the software's ability to compare known genotypes to the mixture samples, each was compared to at least one known genotype. The sample with a contributor ratio of 3:1 was compared to a known profile that was not used in its creation. In an analysis using 5000 MCMC iterations, an LR of 1.7550e-8 was calculated using allele frequency data from all groups. This value changed to 2.5117e-11 when using 50,000 iterations, demonstrating that the known was clearly not supported as a contributor to the mixture. The sample with a contributor ratio of 1:1 was compared to a known profile that was used in its creation, as this would be a

potentially more challenging comparison. In an analysis using 5000 MCMC iterations, an LR of 7.2399e+10 was calculated using allele frequency data from all groups. This value changed to 3.5792e+7 when using 50,000 iterations (see figures 1 and 2 on page 4). Similar LR values were calculated when the 1:1 mixture profile was compared to that of the second known contributor used to make the sample. Both LRs strongly supported the second genotype as a contributor to the mixture, which was the correct answer.

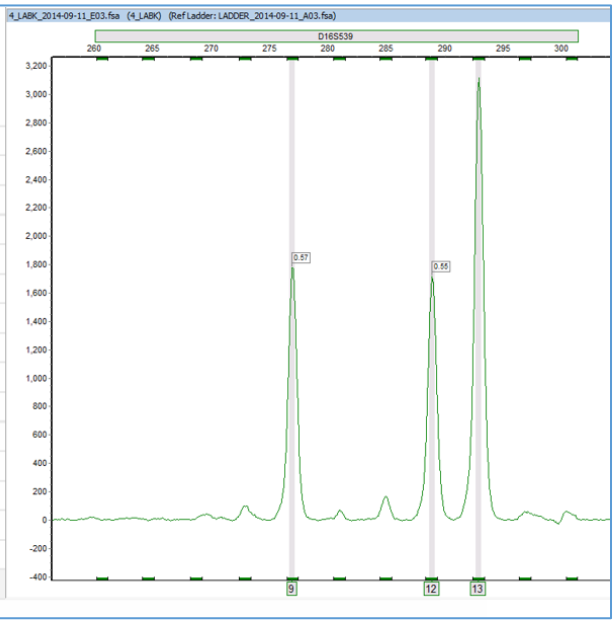
All of the analysis runs conducted took very little time. The analyses using 5000 MCMC iterations were completed in less than four minutes, those with 50,000 iterations in less than 11 minutes. The undergraduate student performing the analyses learned to operate the software over the course of a few hours and encountered no problems once familiarized with the graphical user interface. The menus were easy to identify and the Protocol Set and Model choices, once populated, were simple to use. The data was displayed in logical and concise manners, making it easy to access, use, and tailor to analyst's needs.

A full validation study with synthetic mixtures of varying concentrations and percentages of allele sharing is in progress. This validation study will incorporate more complex models, with consideration of parameters for instrument/chemistry specific stutter, heterozygous peak height variance and ancestry coefficients from NRC II recommendations (Second National Research Council Report).

With 5,000 iterations the MCMC has not reached stability at locus D16S539

Mixture_454_known_2_BK_1NG_2014-09-05_C01.txt

Marker	Genotype	LR (Unknown 1)
D3S1358	(14, 16)	0.2356
TH01	(8, 9)	20.6835
D21S11	(29, 30.2)	57.2452
D18S51	(16, 17)	0.0379
Penta_E	(7, 17)	2.6791
D5S818	(10, 12)	8.8933
D13S317	(8, 14)	94.1942
D7S820	(9, 13)	106.0843
D16S539	(9, 12)	0.0000
CSF1PO	(11, 12)	5.1229
Penta_D	(10, 13)	1.3151
vWA	(16, 18)	7.3086
D8S1179	(14, 15)	8.7829
TPOX	(8, 11)	4.1514
FGA	(19, 23)	16.0309
Overall		**7.2399e+10**



Using 50,000 iterations the MCMC is stablized at each locus, resulting in strong support of inclusion of this sample as contributor 1 and elimination as contributor 2

Mixture_454_known_2_BK_1NG_2014-09-05_C01.txt **Using 'All data' frequencies**

Marker	Genotype	LR (Unknown 1)	LR (Unknown 2)
D3S1358	(14, 16)	0.1568	5.2007
TH01	(8, 9)	20.4129	0.7333
D21S11	(29, 30.2)	55.7991	0.0077
D18S51	(16, 17)	0.0121	0.0046
Penta_E	(7, 17)	2.3537	0.0013
D5S818	(10, 12)	8.8411	0.0001
D13S317	(8, 14)	94.1509	0.0113
D7S820	(9, 13)	106.0313	0.0042
D16S539	(9, 12)	0.0022	8.3976
CSF1PO	(11, 12)	5.0685	0.0005
Penta_D	(10, 13)	1.7283	0.0016
vWA	(16, 18)	7.3685	0.4869
D8S1179	(14, 15)	8.3658	0.0360
TPOX	(8, 11)	4.1761	0.0000
FGA	(19, 23)	15.6641	0.9458
Overall		**3.5792e+7**	**1.2490e-22**