NEW YORK STATE POLICE CRIME LABORATORY SYSTEM
TRUEALLELE® CASEWORK VALIDATION ADDENDUM

TABLE OF CONTENTS

VI. CONCLUSIONS

Genl. 7

NEW YORK STATE POLICE

MEMORANDUM

Station: FIC - Headquarters

Date: December 9, 2013

To: **Dr. Barry Duceman, Director of Biological Sciences**
Dr. Russell Gettig, Associate Director of Biological Sciences

From: **Jay Caponera, Forensic Scientist III**

Subject: **Proposal to incorporate four person mixtures and familial samples into the TrueAllele® Casework probabilistic genotyping software validation**

Objective:

To expand and enhance the current TrueAllele Casework interpretation guidelines by incorporating data generated from four person mixtures and familial samples.

Scope:

Current mixture interpretation guidelines for TrueAllele Casework limit comparisons to a maximum of three donors. However, an unambiguous determination of contributor number for many casework mixtures is challenging, particularly with low template and/or degraded samples. Allowing analysts the additional flexibility of solving TrueAllele cases for up to four donors will enhance our interpretational ability and increase the number of cases that may be solved with the fully continuous probabilistic genotyping approach. Interpretation of mixtures samples composed of related donors also presents a substantial challenge in forensic casework. With the goal of improving such

interpretation, this proposed work also investigates the feasibility of using TrueAllele to deconvolute familial mixture samples that are common in many sexual assault cases.

The following experiments are suggested to provide additional data on the reliability, reproducibility, and robustness of the TrueAllele Casework probabilistic genotyping software with respect to both four person mixture deconvolution and mixture deconvolution between related individuals:

1. Four Person Mixture Study (n = 32)

TrueAllele Casework performance on two separate sets of four person mixtures will be assessed using the mixture ratios listed below. Buccal swabs from eight distinct and unrelated contributors will be extracted, quantified in triplicate, and the resulting averages used to create two mixture sets with a combined target input amount of 1ng template DNA (minimum input of approx. 66.7pg and maximum input of approx. 466.7pg across all mixtures ratios). All four person mixture requests will be solved in duplicate for 50 and 100K read out cycles (burn-in cycles equal to read out for all replicates). Longer run times and/or additional replicates may be necessary after assessing convergence.

The resulting mixture weights, standard deviations, KL scores, and match statistics for all samples will be recorded for each known donor. Markov chain Monte Carlo (MCMC) convergence will be evaluated with Gelman-Rubin statistics and by assessing all mixture weight standard deviations and Markov chain histograms. Specificity of the software will be evaluated by running approximately 16 known staff profiles against all four person mixtures, with the expectation that all resulting log (LR) match statistics will be negative for non-donor staff. All weight of evidence will be expressed in log10 form. Reproducibility and precision of the software will be evaluated by quantifying the variation (measured as standard deviation) between replicated match statistics.

4 person mixtures:

SET 1	**SET 2**
1:1.5:2:2.5	1:1.5:2:2.5
3.5:3:1.5:1	3.5:3:1.5:1
5:3:2:1	5:3:2:1
7:4.5:2.5:1	7:4.5:2.5:1

Two mixture sets n = 8
Duplicate TA requests
With 2 run times n = 32

2. Familial Samples Study (n = 176)

The ability of TrueAllele Casework to deconvolute familial mixtures will be assessed by creating a suite of two person mixtures in ratios of 1:1 and 3:1 from two separate families of four non-identical full sibs (C1 v C2, C1 v C3, C1 v C4, C2 v C3, C2 v C4, C3 v C4, where C = child). The above combinations address the possible defense assertion that the true culprit was the defendant's sibling. Parent-offspring deconvolution will be assessed with an additional set of 1:1 and 3:1 two person mixtures (M v C1, M v C2, M v C3, M v C4, F v C1, F v C2, F v C3, and F v C4, where M = mother and F = father). All parent-offspring combinations will then be solved a second time using the child reference as a known donor. Buccal swabs from all related contributors will be extracted, quantified in triplicate, and the resulting averages used to create mixture sets with a combined target input amount of 1ng template DNA. All two person familial mixture requests will be solved in duplicate for 25K cycles (25K burn-in and 25K read out). Parent-offspring mixture requests using known child references (one unknown mixtures) will also be solved in duplicate for 25K cycles. Longer run times and/or additional replicates may be necessary after assessing convergence.

The resulting mixture weights, standard deviations, KL scores, and match statistics for all samples will be recorded for each known donor. MCMC convergence will be evaluated with Gelman-Rubin statistics and by assessing all mixture weight standard deviations and Markov chain histograms. Specificity of the software will be evaluated by running approximately 14 unrelated staff profiles against all mixtures in addition to all family members in the study. Reproducibility and precision of the software will be evaluated by quantifying the variation (measured as standard deviation) between replicated match statistics, with all weight of evidence expressed in log10 form.

Full sib comparisons:

C1 v C2
C1 v C3
C1 v C4
C2 v C3
C2 v C4
C3 v C4

Mixture ratios of 1:1 and 3:1	n = 12
Replicate TA requests	n = 24
Two separate families	n = 48

Parent-child comparisons (two unknown):

M v C1
M v C2
M v C3
M v C4
F v C1
F v C2
F v C3
F v C4

Mixture ratios of 1:1 and 3:1 n = 16
Replicate TA requests n = 32
Two separate families n = 64

Parent-child comparisons (one unknown):

M v C1
M v C2
M v C3
M v C4
F v C1
F v C2
F v C3
F v C4

Mixture ratios of 1:1 and 3:1 n = 16
Replicate TA requests n = 32
Two separate families n = 64

Note:

All samples will be extracted with DNA Investigator chemistry (QIAGEN) on an EZ1®Advanced XL, quantified with Quantifiler® Duo, amplified with Identifiler® Plus, and sequenced on a 3130*xl* genetic analyzer with 1.0uL DNA input in 9.0 HIDI/LIZ. An alpha level of 0.05 will be used for all statistical tests. This work is intended to serve as an addendum to both the original NYSP FIC Developmental Validation Study (TrueAllele System for Forensic Casework STR DNA Data Interpretation) and the subsequent Identifiler Plus internal validation studies completed in August 2013.

Budget: Reagents for extraction, quantification, amplification, and CE are projected to cost approximately $500.00 for the given sample size.

Timeline: The experiments outlined above combined with the final write-up and revisions to current protocols are expected to take eight to ten weeks for completion.

II. Project Summary

Experiment 1: Four Person Mixture Study

Extending the previous TrueAllele validation work on two and three person mixtures, the four person mixture data examined here exhibit similar patterns with respect to genotype specificity, reproducibility, and mixture weight inference. For the staff comparisons component of the validation, all non-donor log(LR) values for both suites of four person mixtures tested were negative, indicating that the software retains the ability to separate actual donors from non-donors in complex mixtures. The magnitude of negative non-donor log(LR) varied among the four mixture ratios tested, with a maximum of -1.91 found at the ratio with the lowest template contributor (7:4.5:2.5:1 ratio; ~66.7pg). This finding indicates that profiles with low information content may also exhibit low specificity. However, experiment-wide pairwise comparisons between donor and non-donor log(LR) were significantly different, indicating that TrueAllele Casework can effectively deconvolute true contributors from non-contributors in complex mixtures with medium to low template DNA input.

The extended specificity comparisons suggest that the findings above can be successfully extrapolated to a much larger data set with similar results. Tested against 3,000 random profiles from three separate ethnic groups, the software yielded a maximum non-donor log(LR) of 1.945 (likelihood ratio of 88) from all two, three, and four person mixtures (564,000 total pairwise comparisons). This equates to an overall false positive rate of approximately 0.00024. Additionally, the maximum non-donor log(LR) of 1.945 found above was not reproducible, and no false negatives were observed. However, a decrease in overall specificity was noted with increased donor numbers when comparing the

combined validation results from all two, three, and four person mixtures (see Results section IV.A).

With respect to reproducibility, four person mixtures were similar to the two and three person mixtures with no significant differences between replicate genotype inferences detected. Larger variations were also associated with lower template contributors. A comparison of replicate log(LR) values between 50 and 100K cycle times showed no statistical difference, suggesting that even complex mixture requests may be run at reduced cycle times with no adverse effects. With the increased complexity inherent to increased contributor number, mixture weight standard deviations were markedly higher for the four person mixtures, thus inferred mixture weights were not as sharp as those seen in the two and three person mixture sets. However, close approximations to the theoretical mixture weights were still achieved in most samples. Lastly, mean Gelman-Rubin MCMC convergence values were below 1.2 for both 50 and 100K cycle times, suggesting that acceptable mixture weight convergence can still be achieved with up to four contributors. On visual inspection, mixture weight histograms were typically multi-modal for most all contributors and Markov chain diagrams were characterized by crossing chains and non-stationary distributions; such results should be expected when assessing convergence for complex mixture profiles.

Experiment 2: Familial Mixture Study

Full-sibling pairwise comparisons demonstrate exceptional genotype specificity between actual donors and related non-donors (full-siblings and parents). Log(LR) separation between siblings and all related non-donors was statistically significant at both the 1:1 and 3:1 mixture ratios tested. Computer-inferred genotype inference for all familial samples was markedly better when full-sib donor template amounts were unequal, with mean separation improving approximately 13 orders of magnitude at the 3:1 mixture ratios when compared to 1:1 data. Cross-matching was present for some of the full-sib comparisons, and in all instances was strongly correlated with the number of shared alleles between the mixture profile and non-donor siblings. However, in all mixtures the two highest match statistics were always associated with the actual donor siblings. A clear gradient of specificity emerged from the full-sibling mixture data with the largest separation from inferred genotypes occurring in unrelated references, then related parents next, and related full-siblings showing the smallest separation (see Results Section V.A).

Parent-child two unknown donor comparisons showed an improvement of approximately three orders of magnitude in familial separation over the full-sibling comparisons. Similar to the full-sib data set, the highest specificity was observed at the 3:1 mixture ratio level, with a mean separation of over 23 log units observed between all parent-child comparisons and related non-donors (parents and children). The strongest separation between donor and non-donor related genotypes occurred in the parent-child one unknown comparisons, with an overall mean separation of over 33 log units observed.

Experiment-wide log(LR) values were highly reproducible for the familial mixture data set with no significant differences between replicates detected. The largest variations were observed in the full-sib comparisons at 3:1 mixture ratios; in some mixtures the large amount of allele sharing resulted in increased variability in log(LR) values for the minor donors. Computer-based mixture weight inference was almost identical to the theoretical mixture weights for both the 1:1 and 3:1 samples. KL values (divergence of inferred profiles from population priors) also showed no significant difference between replicated samples, indicating that the software can reproducibly infer separate genotypes from mixture samples with a high degree of allele sharing.

The findings above strongly support the application of TrueAllele Casework in DNA mixture samples where family members may be contributors. For maximal specificity and where individual case context allows, familial references should be used as known donors in mixtures involving the assumed presence of related family members.

III. Procedures and Methods

Experiment 1: Four Person Mixture Study (n =32)

TrueAllele Casework performance on four person mixtures was assessed using mixture ratios of 2.5:2:1.5:1, 3.5:3:1.5:1, 5:3:2:1, and 7:4.5:2.5:1 for two separate mixture sets. The specific ratios were chosen to provide information on medium (~500pg) to low template (<100pg) contributors. Buccal swab samples from eight unrelated staff members were extracted with DNA Investigator chemistry on an EZ1 Advanced XL, quantified in triplicate with Quantifiler Duo, and amplified with Identifiler Plus using a target DNA input of 1.0ng. Four person mixture TrueAllele requests were created and run in duplicate for both 50K cycles (50K burn-in and 50K read out) and 100K cycles. The resulting mixture weights, mixture weight standard deviations, Kullback-Leibler (KL) divergence scores, Gelman-Rubin (GR) convergence statistics, and log(LR) match statistics were recorded for each known donor. Specificity was evaluated by running 16 unrelated non-donor staff profiles against all four person mixtures. This examination was further extended to include a comparison against 3,000 individual profiles (1,000 each from CAU, BLK, and HIS populations) for all inferred profiles from the two, three, and four person mixtures generated during TrueAllele validation work. All FBI population profiles were supplied by Cybergenetics. Reproducibility and precision of the software were evaluated by quantifying the variation (measured as standard deviation) between

replicated match statistics, mixture weights, and KL scores. Markov chain Monte Carlo (MCMC) convergence was evaluated with GR statistics and by visually assessing all mixture weight histograms and Markov chain plots.

Experiment 2: Familial Mixture Study (n = 176)

DNA from two separate six person families (each containing four full-sibs and two parents) was extracted with DNA Investigator chemistry (QIAGEN) on an EZ1 Advanced XL, quantified in triplicate with Quantifiler Duo, and amplified with Identifiler Plus. All possible within-family pairwise comparisons were created (full-sibs vs. full-sibs and parents vs. full-sibs) in mixture ratios of 1:1 and 3:1 with a total DNA input of 1.0ng for all samples. Two person TrueAllele mixture requests were created and run in duplicate for all pairwise samples, and additional one unknown requests were created for all parent-child comparisons using the child as the known genotype. All requests were run for 25K cycles (25K burn-in and 25K read out). The resulting mixture weights, mixture weight standard deviations, KL divergence scores, GR convergence statistics, and log(LR) match statistics were recorded for each known donor. Specificity was evaluated by examining log(LR) values for 14 unrelated non-donor staff profiles in addition to all related non-donors (full-sibs and parents) for all familial mixtures. Reproducibility and precision of the software were evaluated by quantifying the variation (measured as standard deviation) between replicated match statistics, mixture weights, and KL scores. MCMC convergence was evaluated with GR statistics and by visually assessing all mixture weight histograms and Markov chain plots.

Note:

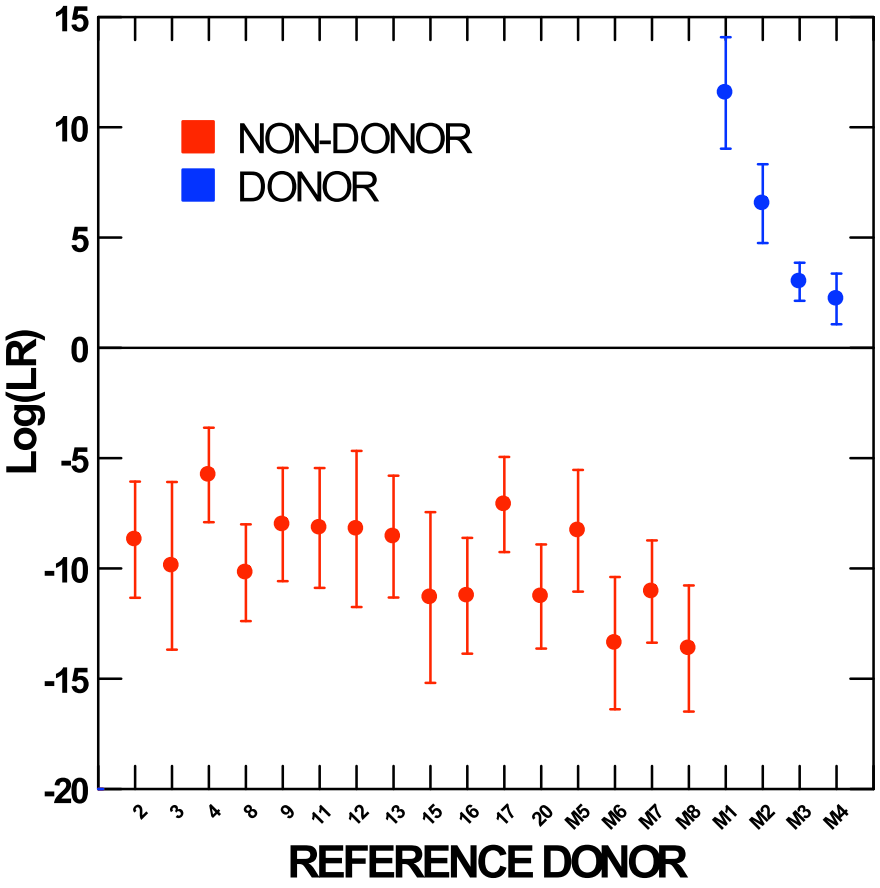
Cybergenetics TrueAllele Casework version 3.3.5148.1 (26-Nov-2013) was used for all validation experiments listed above. All TrueAllele requests used data run on a 3130*xl* using 5 second injection times for familial mixtures and 10 second injection times for the unrelated four person mixture sets. All statistical tests assume an alpha level of 0.05. ANOVA, regression, and non-parametric Kruskal Wallis tests were performed in Systat v. 13.1; Student's t tests were performed in Excel. All match statistics were calculated using a theta value of .01. All graphs were created with Systat v. 13.1, and all tables were created with Excel.

IV. Experiment 1 Results: Four Person Mixture Study

A. Specificity: Staff Comparisons

Data from mixture set 1 show a mean separation of 15.88 log units between known donor and non-donor reference profiles, and a mean separation of 16.80 log units was found for mixture set 2 (Figures IV.A.1. and IV.A.2). For all mixtures tested, non-donor log (LR) values were negative with maximum values of –2.64 for mixture set 1 and -2.78 for mixture set 2. Minimum separation between known donor and non-donor log(LR) values was 3.38 for mixture set 1 and 4.70 for mixture set 2. Further, pairwise statistical comparisons between donor and non-donor log(LR) match statistics were significantly different for both mixture set 1 (Kruskal Wallis $p < 0.001$) and mixture set 2 ($p < 0.001$), indicating strong specificity out to four person mixtures. Specificity was also similar across mixture ratios with maximal separation values obtained at the 7:4.5:2.5:1 level (Figure IV.A.3). However, specificity of genotype inference is clearly diminished with increasing contributor numbers (Figure IV.A.4).

Figure IV.A.1. Dot plot showing mean specificity of TrueAllele Casework by reference sample for mixture set 1. Error bars represent one standard deviation; dashed line is set at zero; MCMC cycle time was 50K/50K and all four mixture ratios are pooled.



Average donor log(LR) = 5.83
St. Dev. = 4.04

Average non-donor log(LR) = -10.05
St. Dev. = 4.16

Figure IV.A.2. Dot plot showing mean specificity of TrueAllele Casework by reference sample for mixture set 2. Error bars represent one standard deviation; dashed line is set at zero; MCMC cycle time was 50K/50K and all four mixture ratios are pooled.

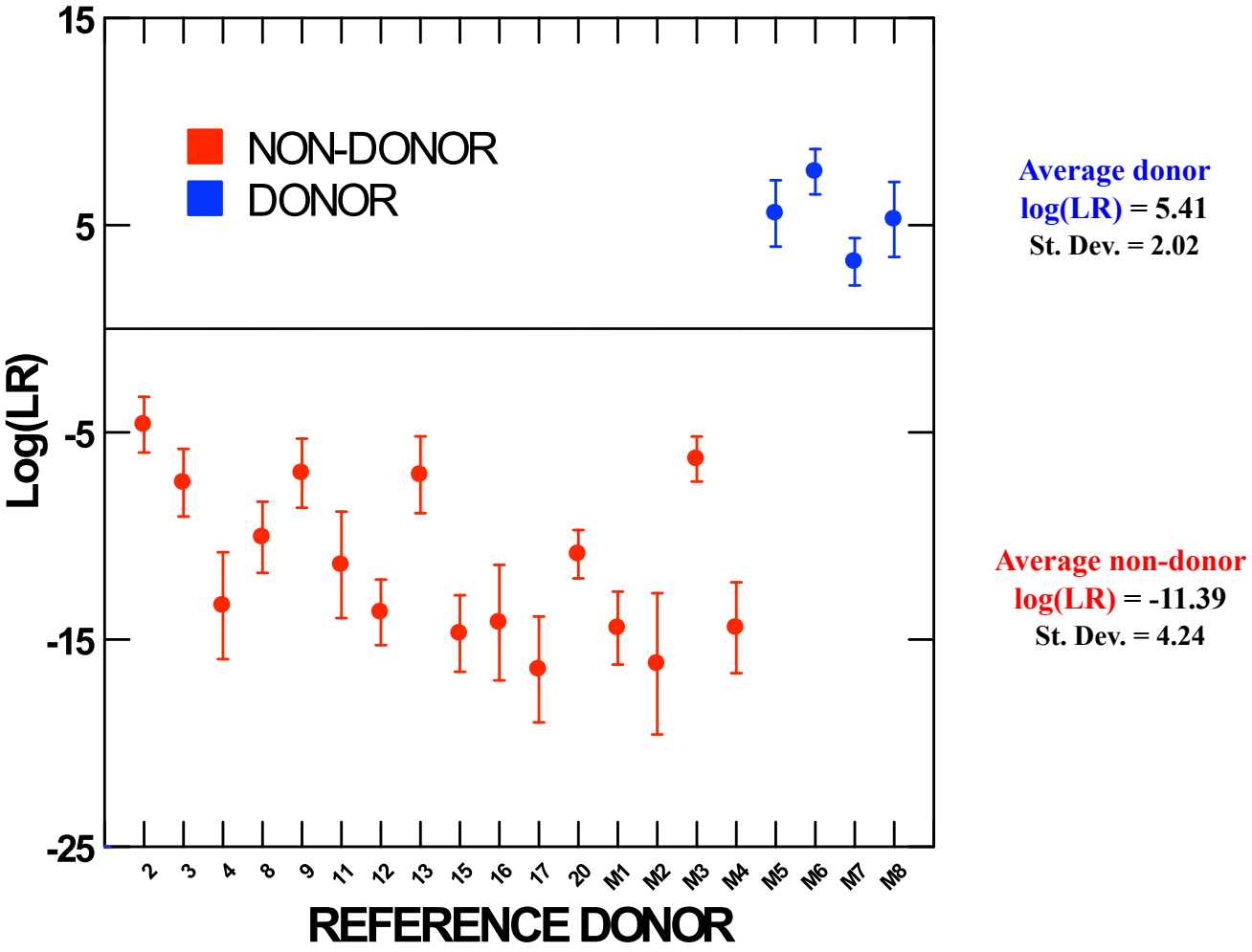


Figure IV.A.3. Boxplot of donor and non-donor log(LR) values from all four person mixtures as a function of mixture ratio. Median values are shown; asterisks indicate outliers. Mixture sets 1 and 2 are pooled (2,177 observations). Dashed line set at zero.

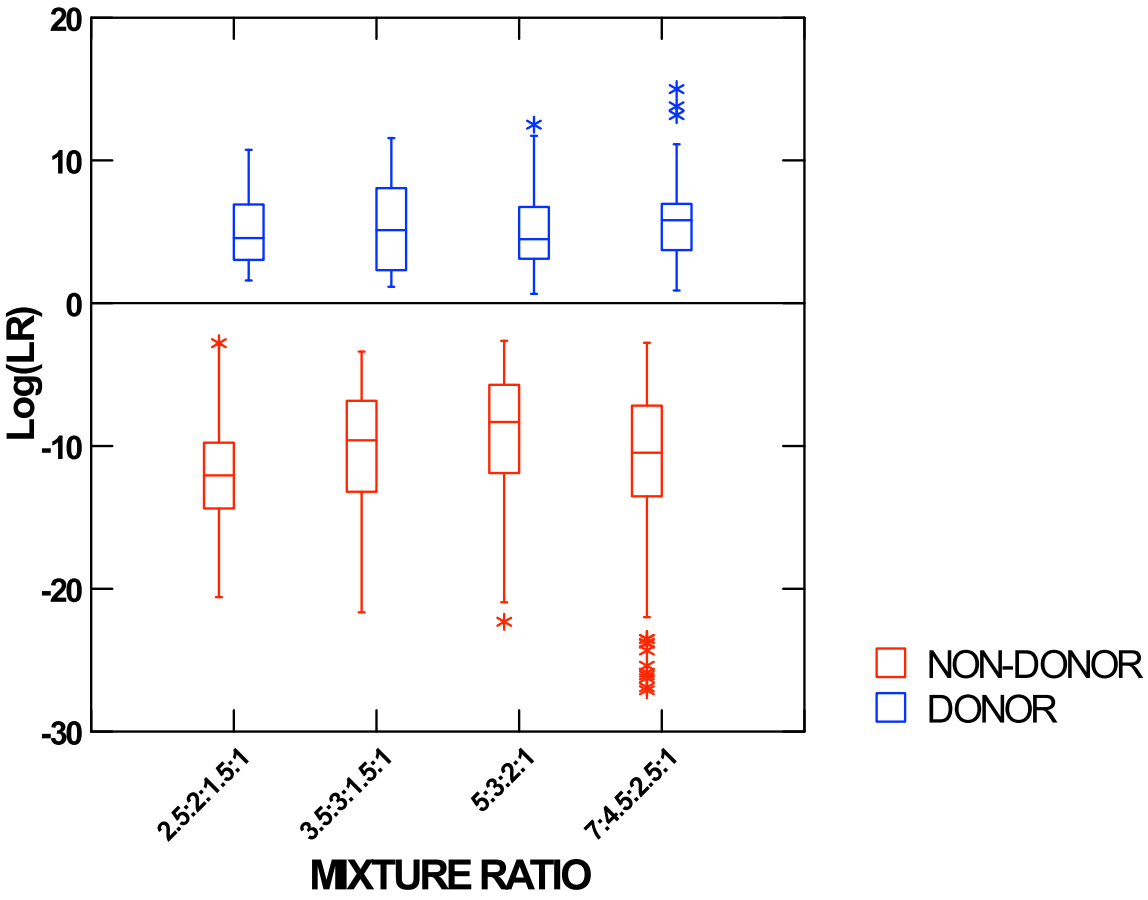
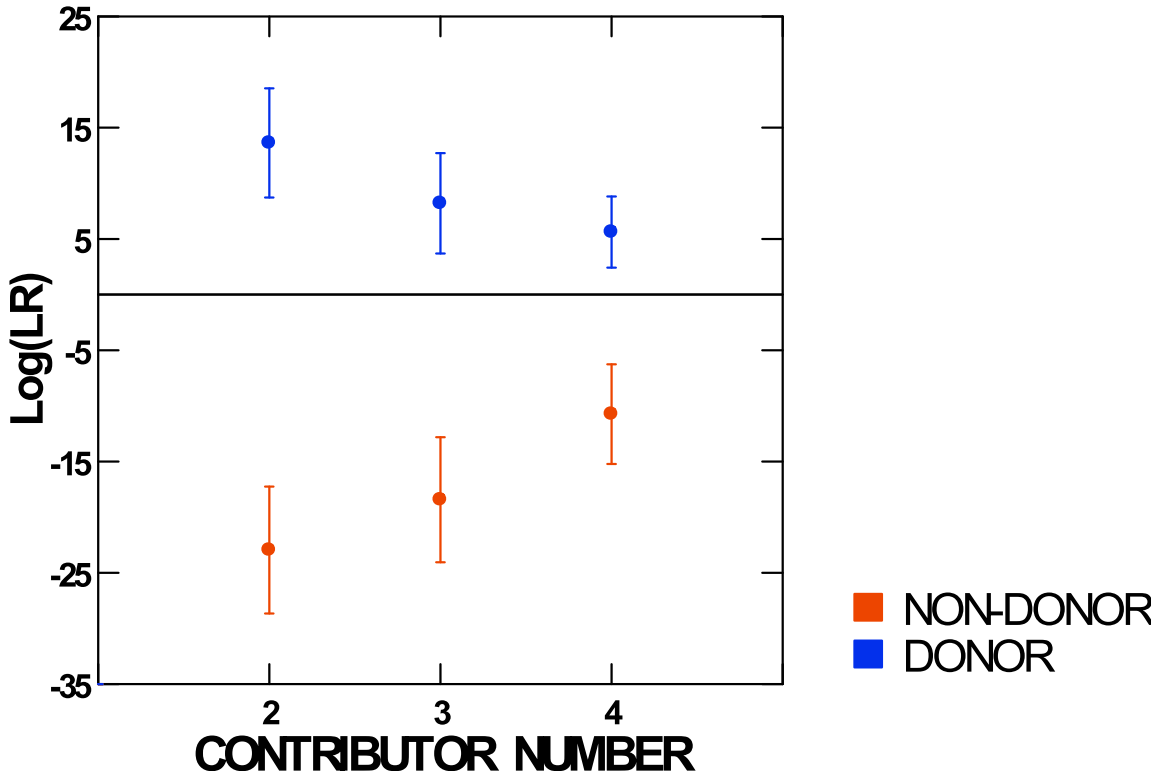


Figure IV.A.4. Dot plot showing specificity as a function of contributor number. Mean values from combined two, three, and four person mixture sets are shown from all in-house TrueAllele Casework mixture validation work. Error bars represent one standard deviation; dashed line is set at zero.



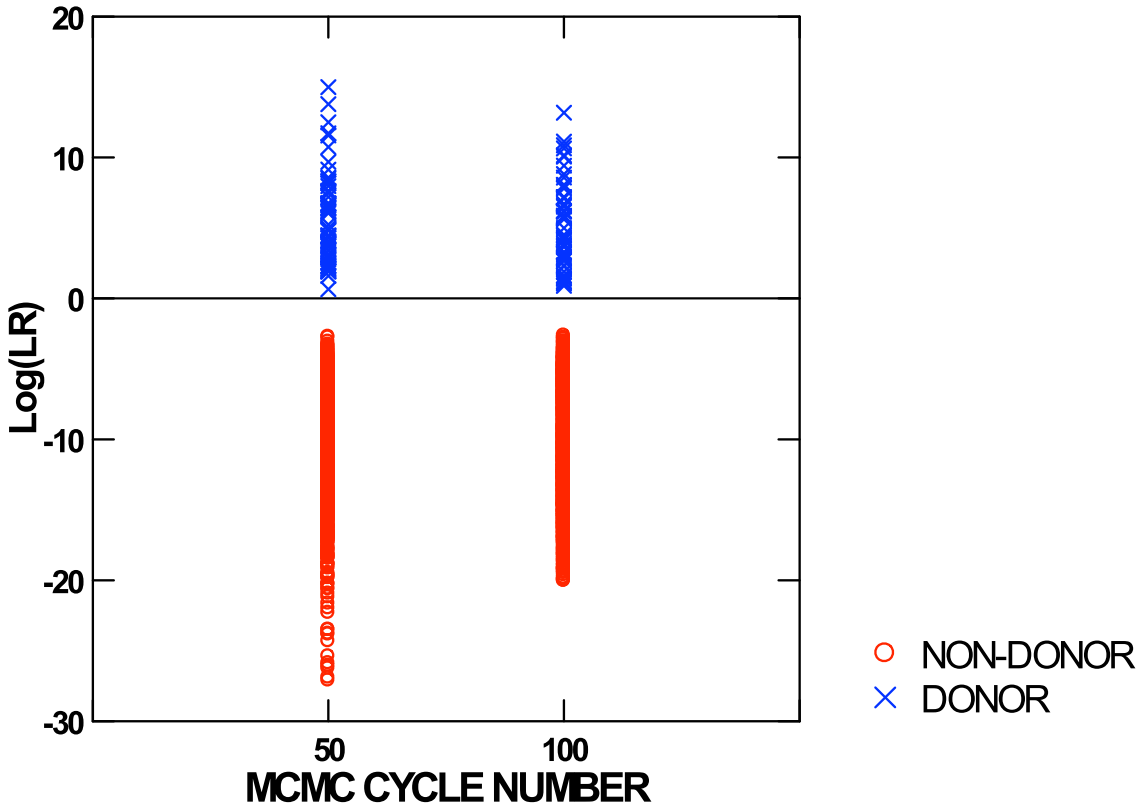
To evaluate TrueAllele Casework performance with longer MCMC cycle times, the specificity experiments above were repeated at 100K/100K burn-in/read out. Log(LR) values were statistically significant between both run times (Kruskal Wallis p = 0.024), with the shorter run times providing better mean separation (Figure IV.A.5). However, overall 100K separation was similar to the 50K results, with a mean of 14.18 log units between known donor and non-donor reference profiles for mixture set 1 and 16.73 log units for mixture set 2. Minimum separation between known donor and non-donor log(LR) values was 3.53 for mixture set 1 and 5.46 for mixture set 2 (Table IV.A.1). While standard deviations were slightly smaller with the longer run times, the data suggest that analysts may run mixtures of up to four contributors at reduced cycle times with no deleterious effect on overall specificity.

Table IV.A.1. Comparison of specificity results between 50K and 100K MCMC cycle times.

Cycle Time	Mixture Set	Mean Donor Log(LR)	Donor Std. Dev.	Mean Non-Donor Log(LR)	Non-Donor Std. Dev.	Minimum Log(LR) Separation

50K	**1**	**5.83**	**4.04**	**-10.05**	**4.16**	**3.38**
100K	**1**	**5.19**	**3.69**	**-8.99**	**3.26**	**3.53**
50K	**2**	**5.41**	**2.02**	**-11.39**	**4.24**	**4.70**
100K	**2**	**5.38**	**1.74**	**-11.35**	**4.23**	**5.46**

Figure IV.A.5. Scatterplot of log(LR) values from all four person mixtures as a function of MCMC cycle number and grouped by donor type. Mixture sets 1 and 2 are pooled (2,177 observations); dashed line set at zero.



B. Specificity: Extended Comparisons

The specificity of all computer-inferred profiles from the four person mixture sets was tested against 3,000 individual profiles for a total of 192,000 pairwise comparisons. From those, 132 were positive (false error rate of 0.00069), with a maximum non-donor

log(LR) of 1.945 (Figure IV.B.1). Extending this comparison to all two, three, and four person mixture data (564,000 pairwise comparisons) found 137 positive values with the same maximum non-donor log(LR) of 1.945 (overall combined error rate of 0.00024). No significant difference in the number of false positive results was found between two and three person mixtures (Fisher's Exact $p = 0.34$), but differences were highly significant between both two and four ($p < 0.001$) and three and four person mixtures ($p < 0.001$). The data suggest a strong inverse relationship between specificity and mixture complexity as is evident in Figure IV.B.2.

Figure IV.B.1. Histogram showing specificity of all four person mixture data run against 3,000 random profiles (1,000 from each of the three FBI population groups).

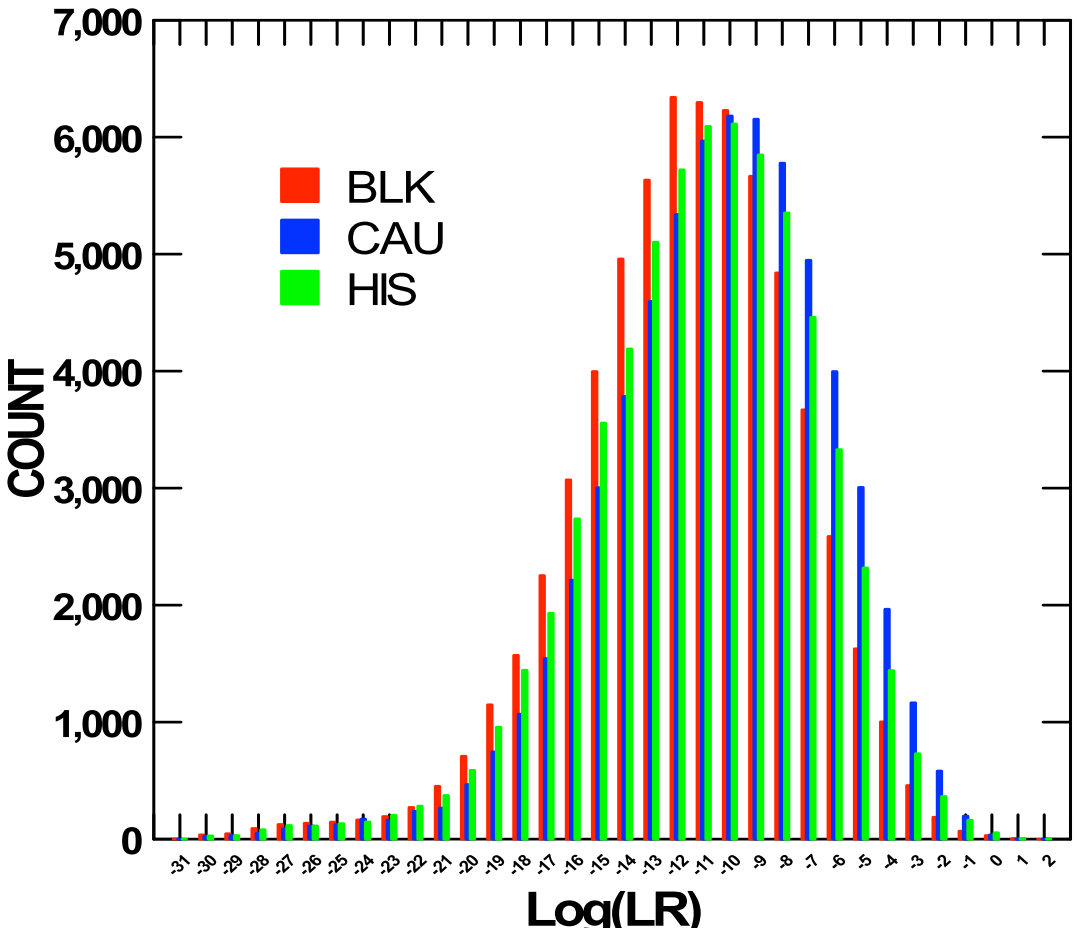
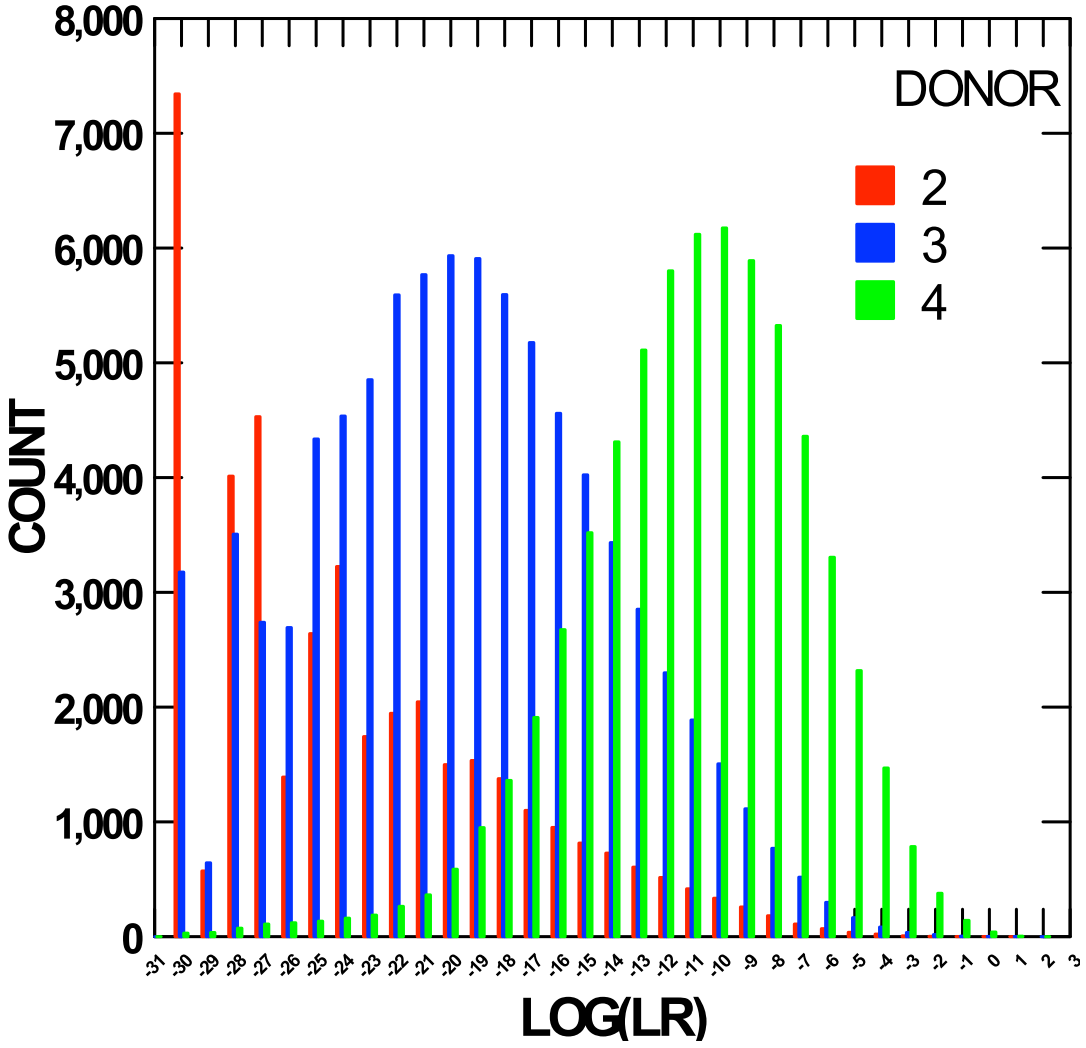


Figure IV.B.2. Histogram showing specificity of all TrueAllele Casework validation mixture data run against 3,000 random profiles as a function of contributor number. All three FBI populations are pooled.



C. Reproducibility: Log(LR), Mixture Weights, and KL

The reproducibility of TrueAllele Casework was assessed by running duplicate identification requests for all four person mixtures. Log(LR) match statistics for known donors were generally reproducible with no significant differences between replicate genotype inferences detected (ANOVA p = 0.391) (Figure IV.C.1). Similar to the two and three person mixture validation data, the largest variations were associated with the lowest template donors and with mixture profiles exhibiting minimal separation (Table IV.C.1). Reproducibility between replicate genotype inferences was also found to decrease with increasing contributor number, as is expected with increasing data complexity, but was typically within 2 orders of magnitude (Figure IV.C.2). A similar pattern is observed with mixture weight inference, with higher standard deviations associated with increased contributor number (Figure IV.C.3). No significant differences between 50 and 100K cycle times were seen across mixture weight standard deviations (Kruskal Wallis p = 0.466). However, mixture weight standard deviations were significantly different across the four ratios tested ($p < 0.001$), with the highest values observed in mixtures with the lowest template contributors (Figure IV.C.4). Lastly, KL values (divergence of inferred profiles from population priors) were not significantly different between reference donor replicates (Kruskal Wallis p = 0.812) (Figure IV.C.5).

Figure IV.C.1. Bar graph of reference donor log(LR) values as a function of replicate (A and B) from all four person mixtures. All contributors (M1-M8) are shown; 50 and 100K cycle times are pooled and error bars represent one standard deviation.

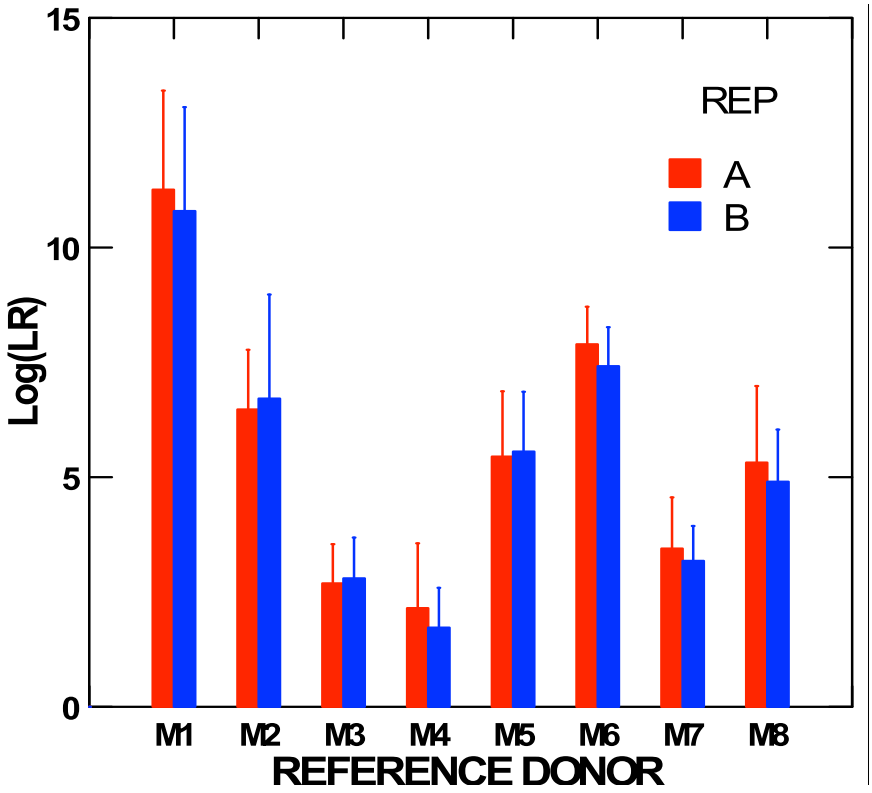


Figure IV.C.2. Comparison of mean absolute difference between duplicated known donor log(LR) values as function of contributor number. Error bars represent one standard deviation; dashed line set at a log(LR) of 2.

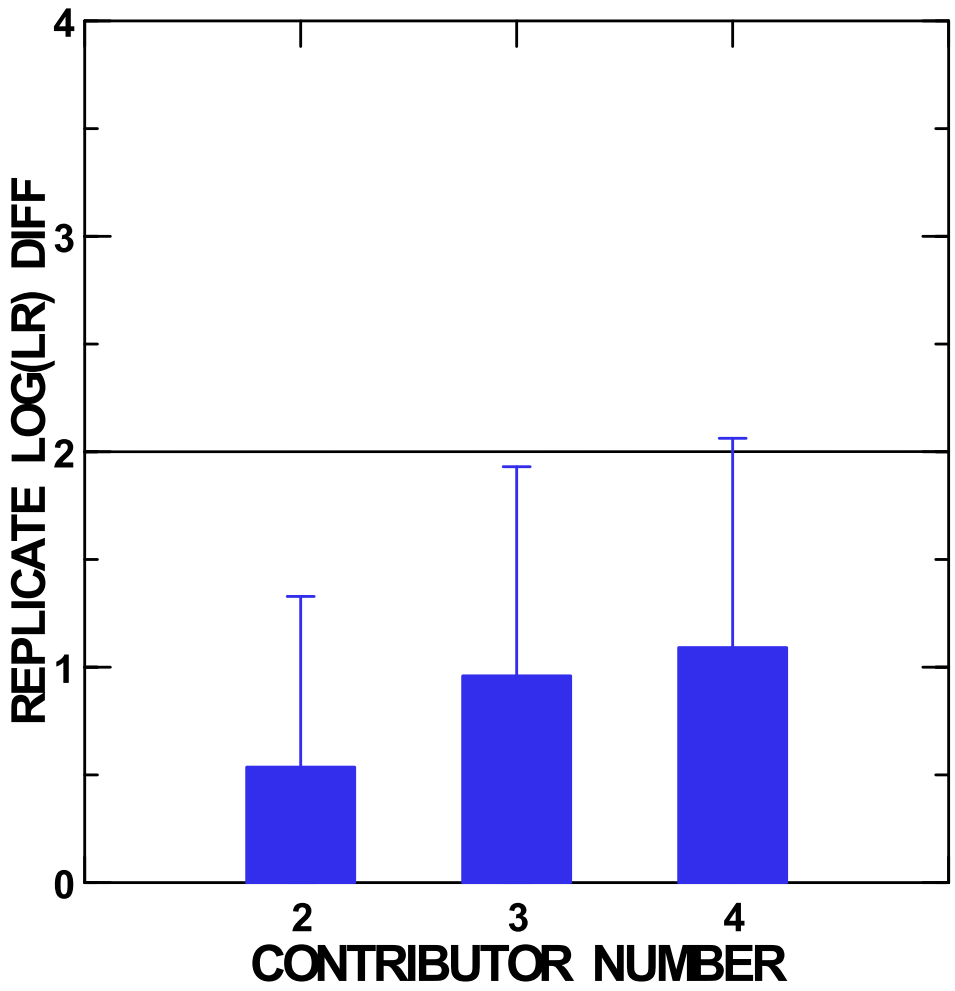


Figure IV.C.3. Comparison of mean mixture weight standard deviations as a function of contributor number. Error bars represent one standard deviation.

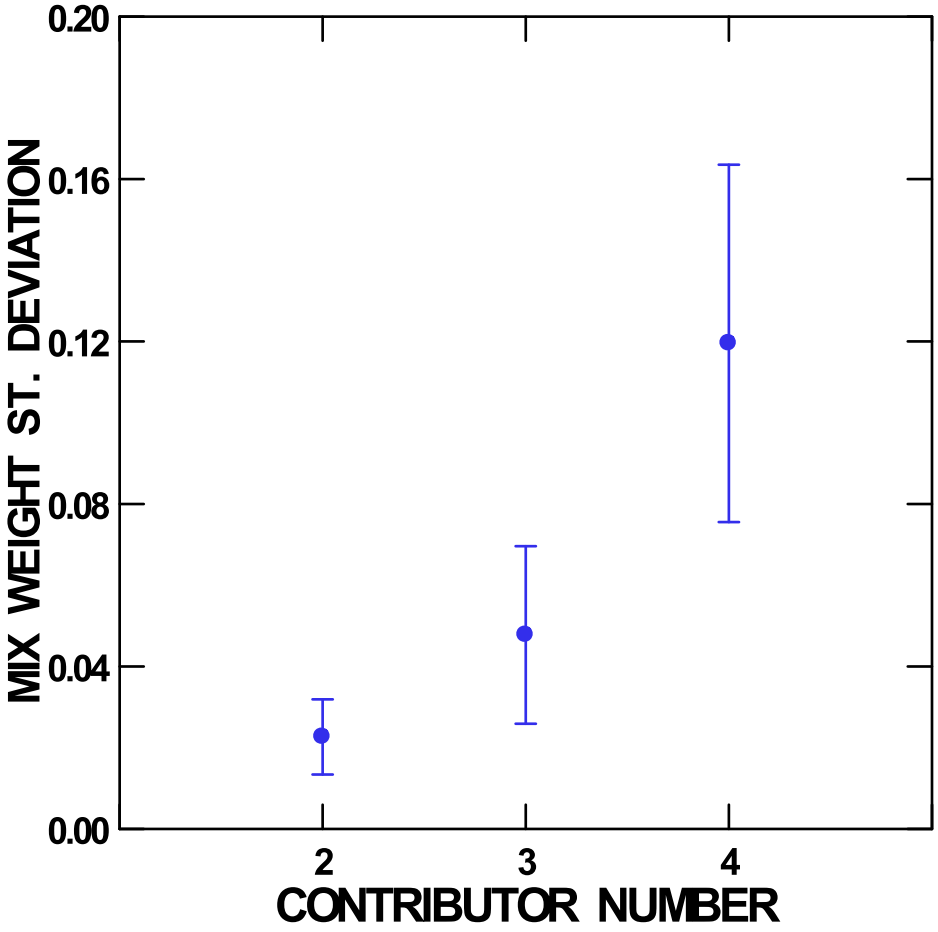


Table IV.C.1. Comparison of theoretical four person mixture weights to those inferred by TrueAllele Casework; mixture weights shown are averaged within respective mixture categories for mixture sets 1 and 2 and include data from both 50 and 100K cycle times.

Theoretical Mixture Weight Ratios	TrueAllele Inferred Mixture Weight Ratios	Mean Standard Deviation
2.5:2:1.5:1	2.7:2.3:2.4:2.2	0.088
3.5:3:1.5:1	3.0:2.8:2.2:2.0	0.133
5:3:2:1	3.2:2.7:2.1:1:8	0.154
7:4.5:2.5:1	3.4:2.6:2.3:1.8	0.142

Figure IV.C.4. Boxplot of median four person mixture weight standard deviations as a function of mixture ratio and grouped by cycle time. Asterisks and circles indicate outliers.

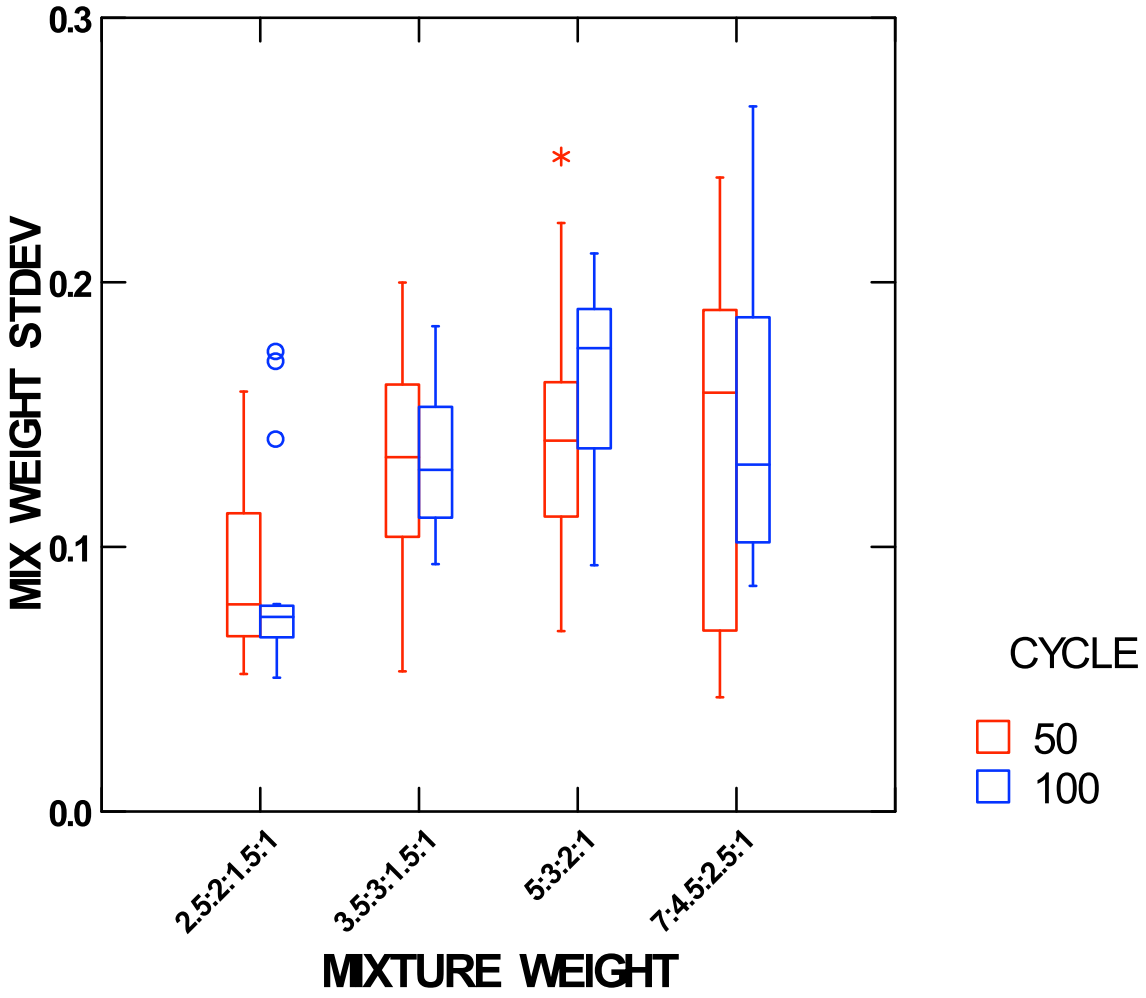
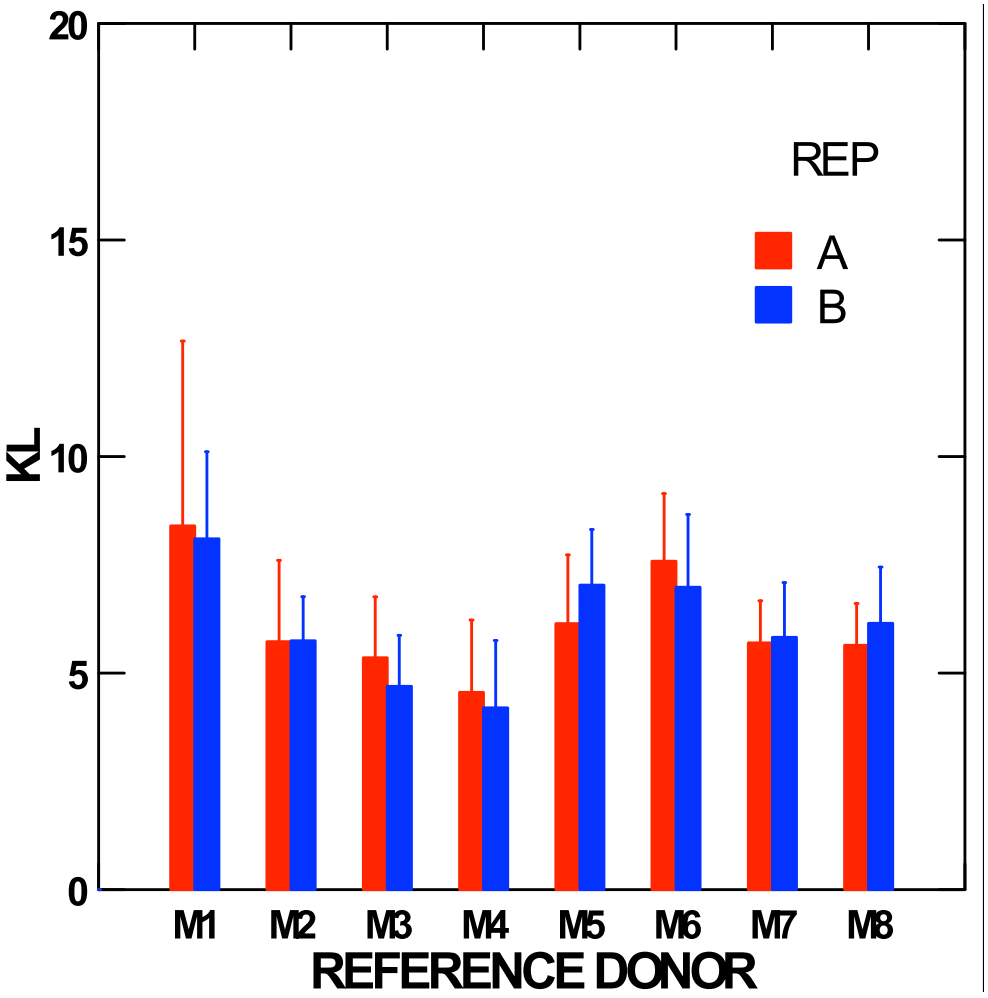


Figure IV.C.5. Bar graph of reference donor KL values as a function of replicate (A and B) from all four person mixtures. All contributors (M1-M8) are shown; 50 and 100K cycle times are pooled and error bars represent one standard deviation.



D. MCMC Convergence

Convergence was assessed with Gelman Rubin statistics (GR hereafter) and by visual inspection of the mixture weight Markov chains and histograms. For all four person mixtures tested, the mean GR was 1.14 with a standard deviation of 0.172. Examined separately, GR statistics for 50K cycles were significantly higher than 100K cycles (Kruskal Wallis p = 0.023: 50K mean = 1.18; 100K mean = 1.10) (Figure IV.D.1). However, mean values for both cycle parameters were still below 1.2, suggesting acceptable convergence (based on GR alone) was achieved for up to four contributor mixtures. Additionally, no significant difference in GR statistics was detected between mixture ratios (p = 0.174). On visual inspection, mixture weight histograms were multi-modal for almost all contributors and Markov chain diagrams were characterized by crossing chains and non-stationary distributions as may be expected for complex mixture profiles (Figure IV.D.2).

Figure IV.D.1. Bargraph of mean GR convergence scores as a function of mixture ratio and grouped by MCMC cycle time. Error bars represent one standard deviation.

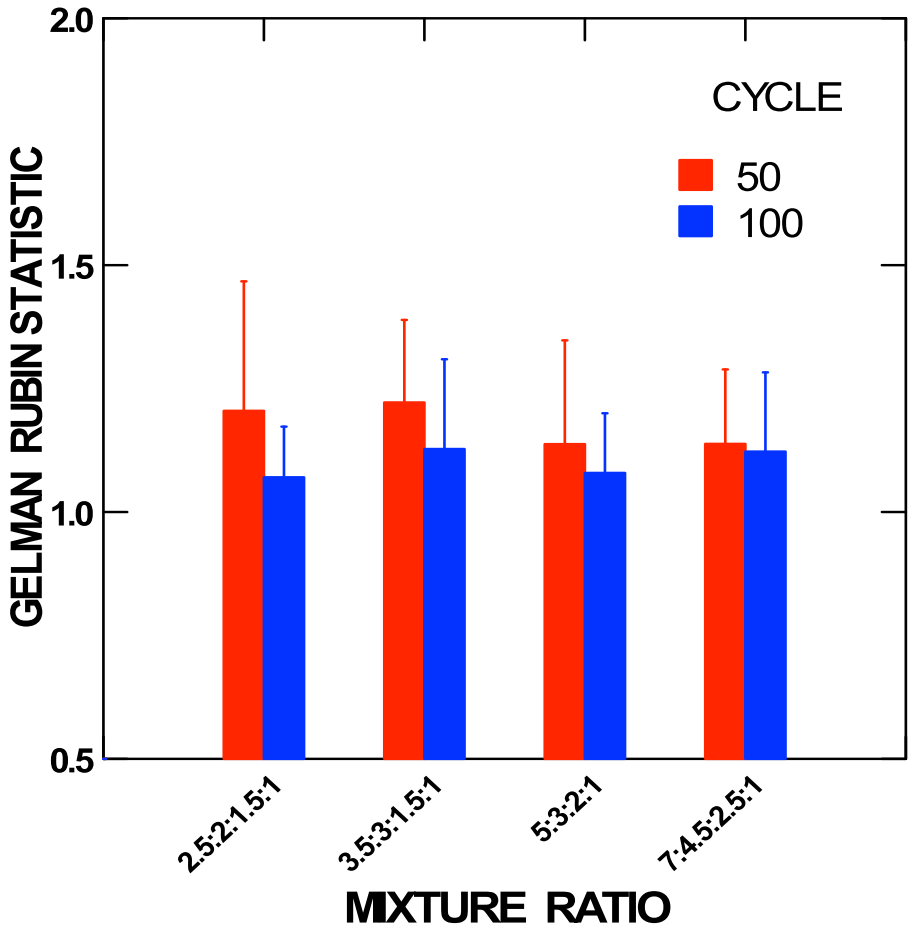
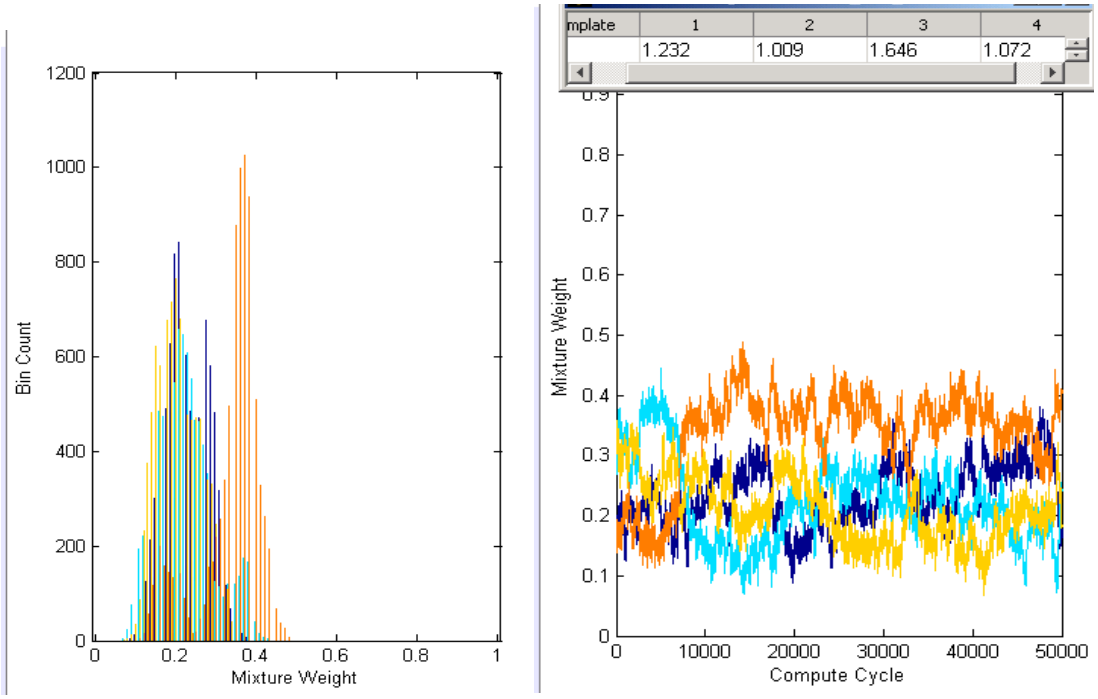
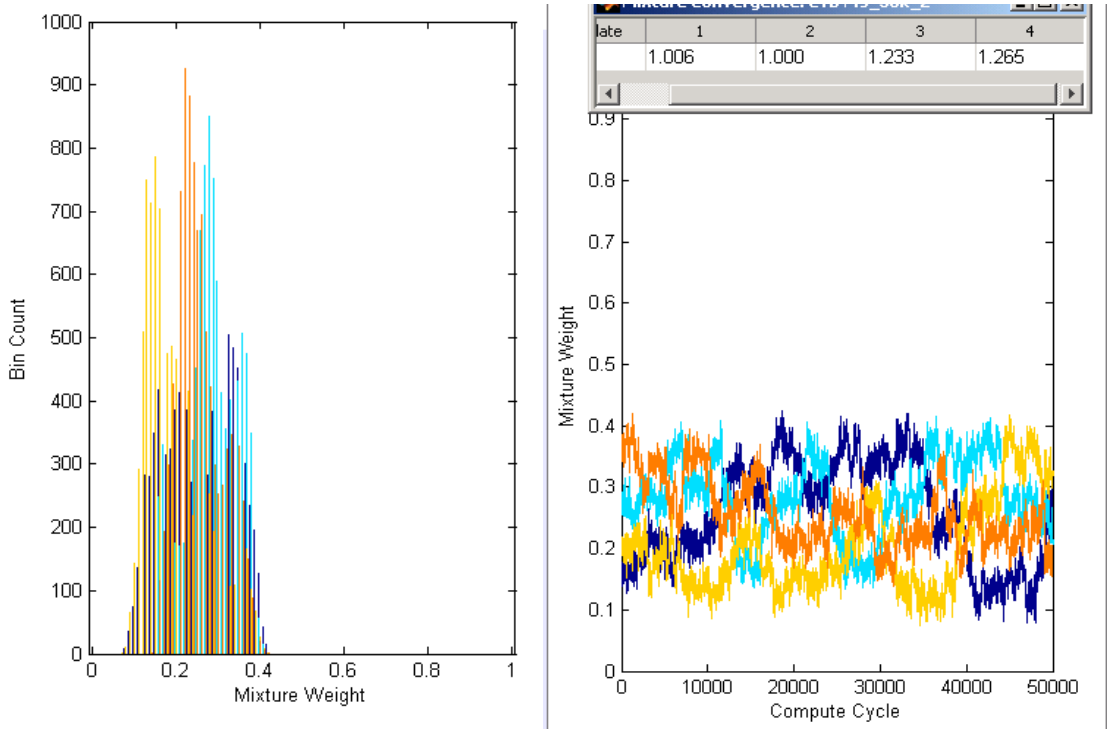


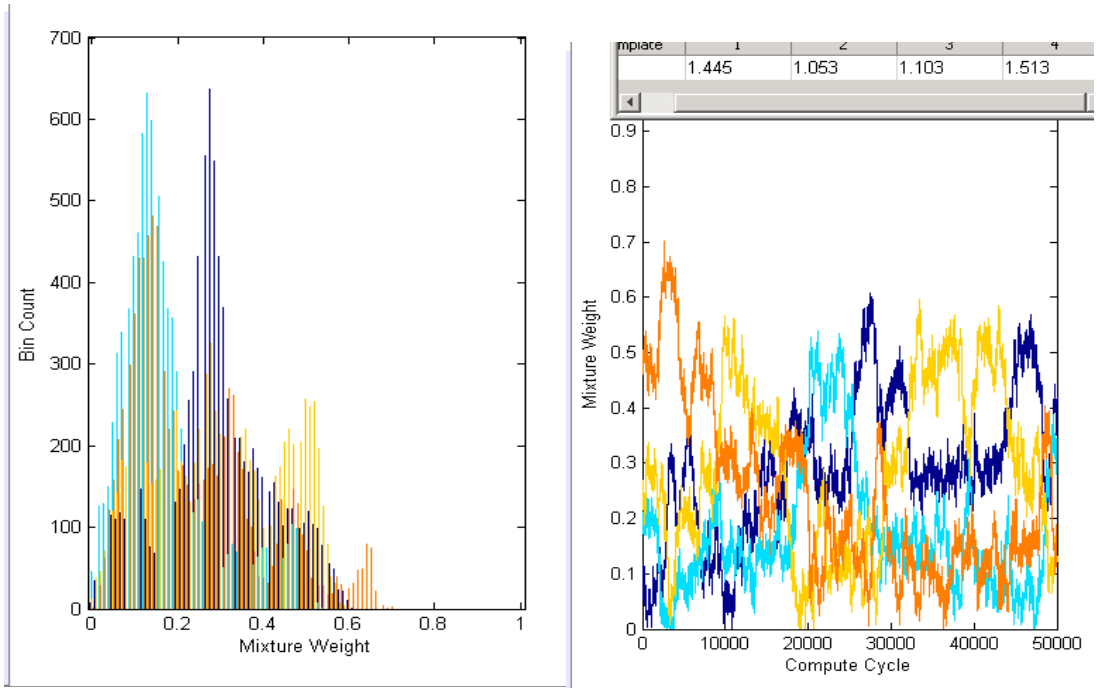
Figure IV.D.2. Mixture weight histograms (left) and associated Markov chains (right) from all four person mixture samples using 50K/50K cycle times; GR convergence scores are indicated in the Markov chains at right. 100K/100K output had similar appearances (data not shown).



Mix Set 1
2.5:2:1.5:1

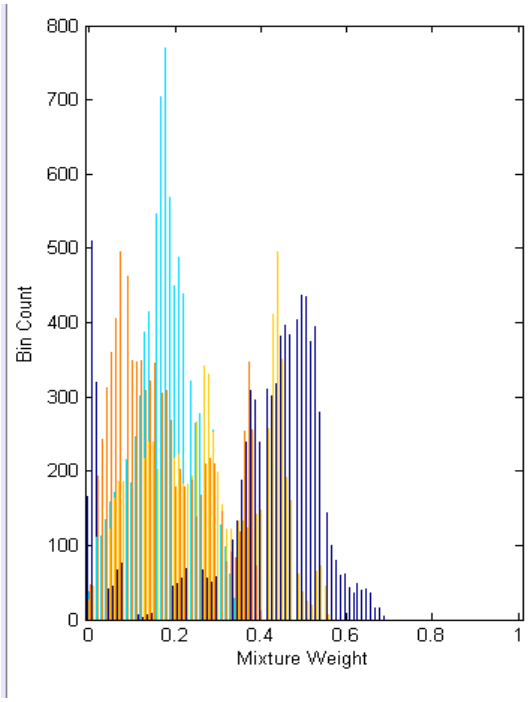


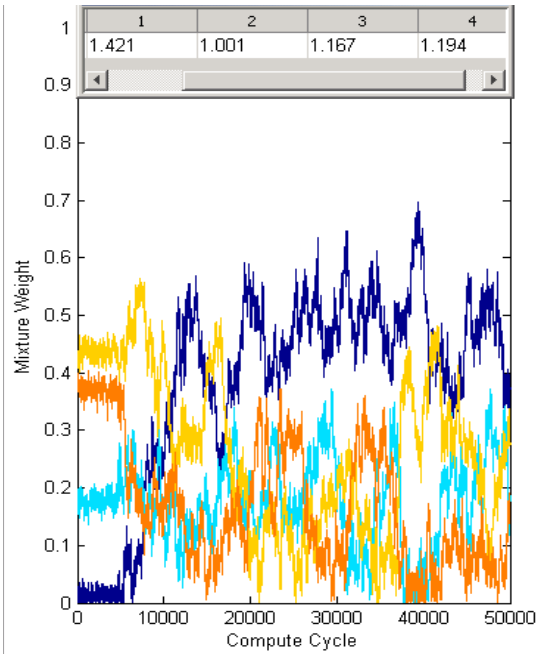
Mix Set 1 2.5:2:1.5:1 replicate

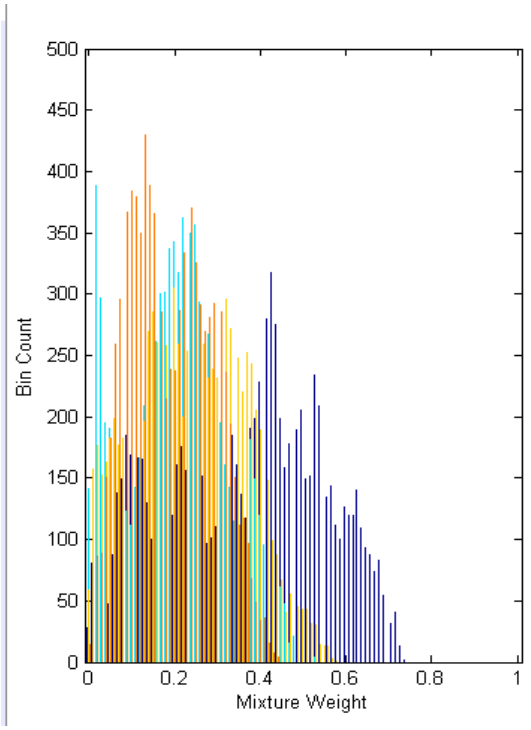


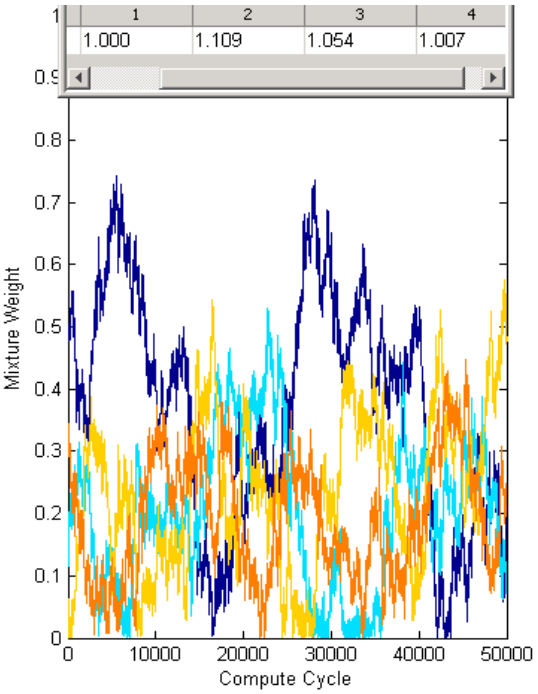
Mix Set 1 3.5:3:1.5:1

Mix Set 1 3.5:3:1.5:1

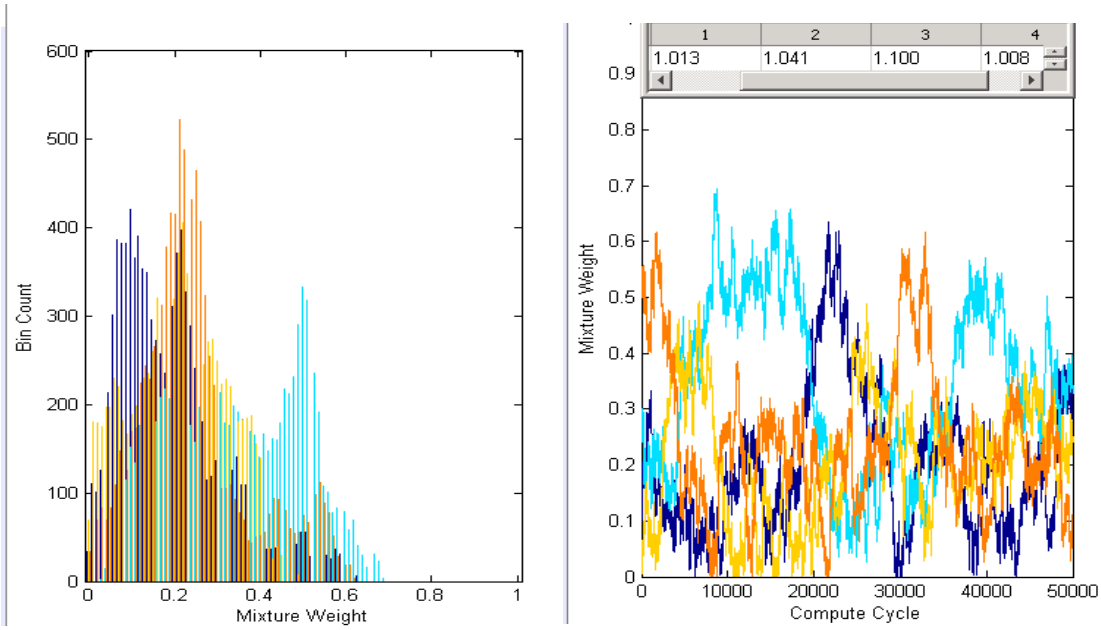




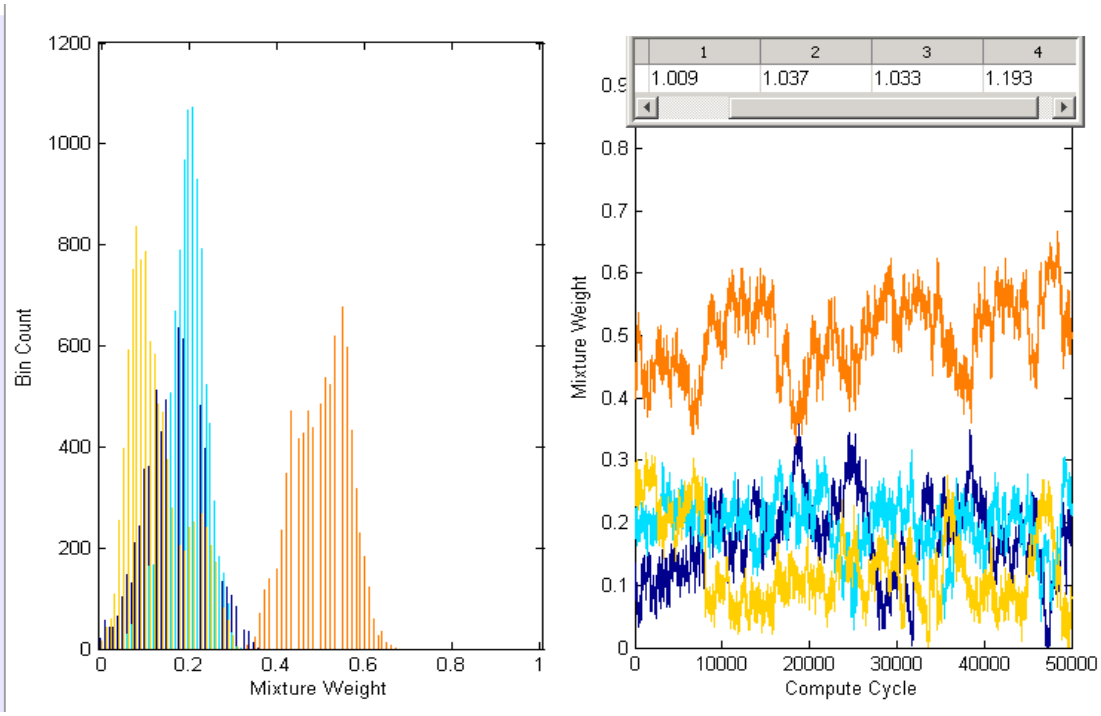




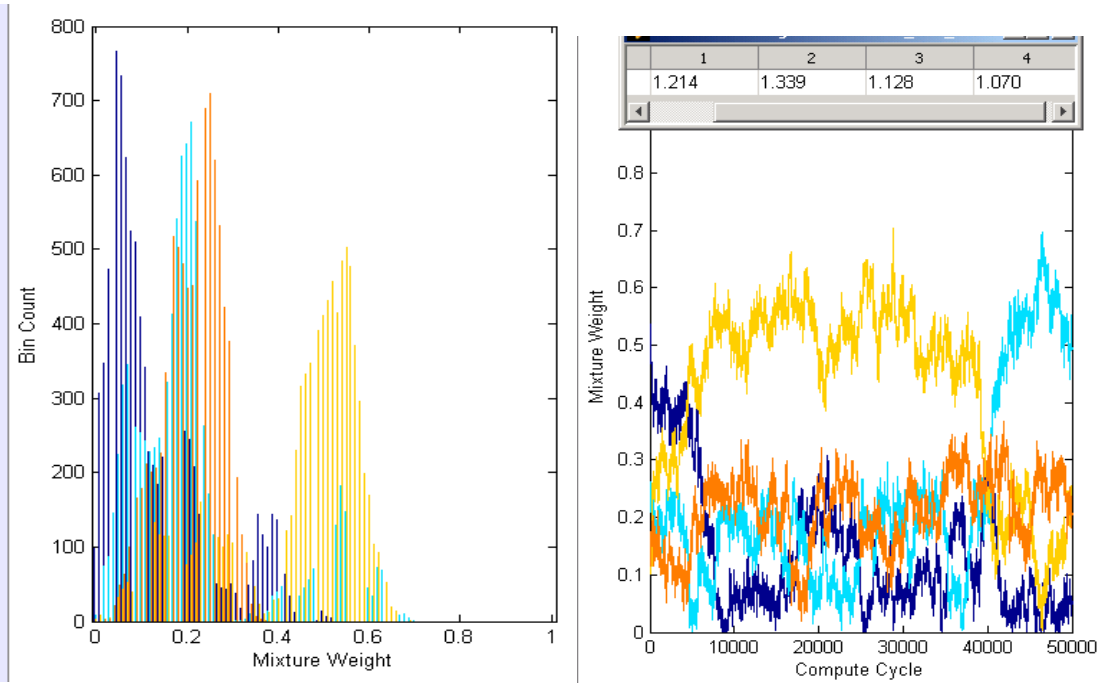
Mix Set 1
5:3:2:1



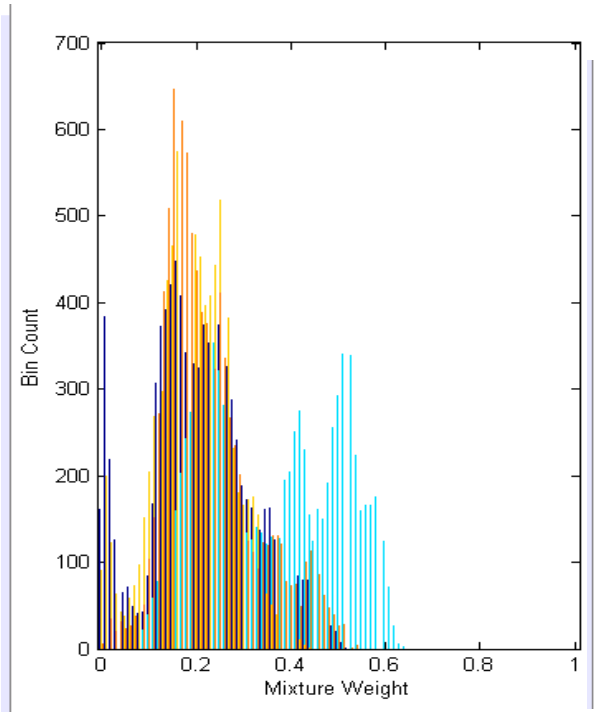
Mix Set 1
5:3:2:1
replicate

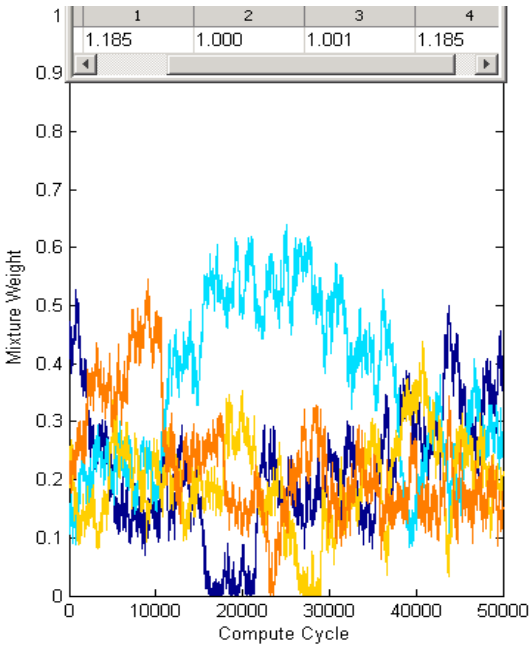


Mix Set 1
7:4.5:2.5:1

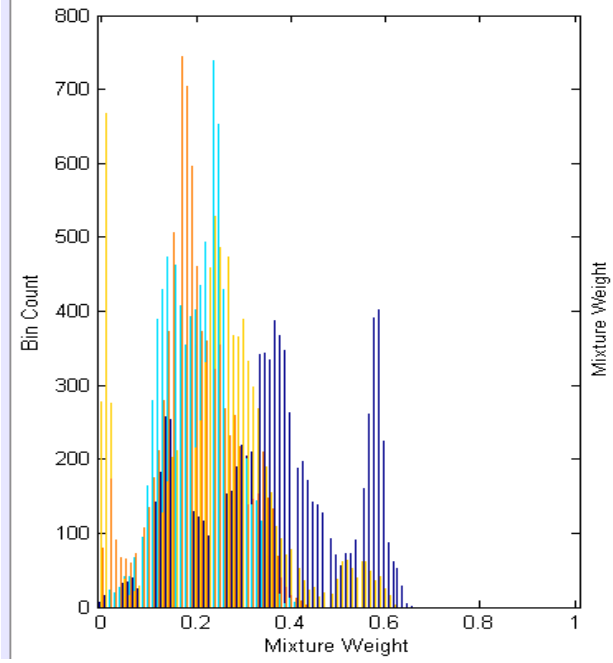


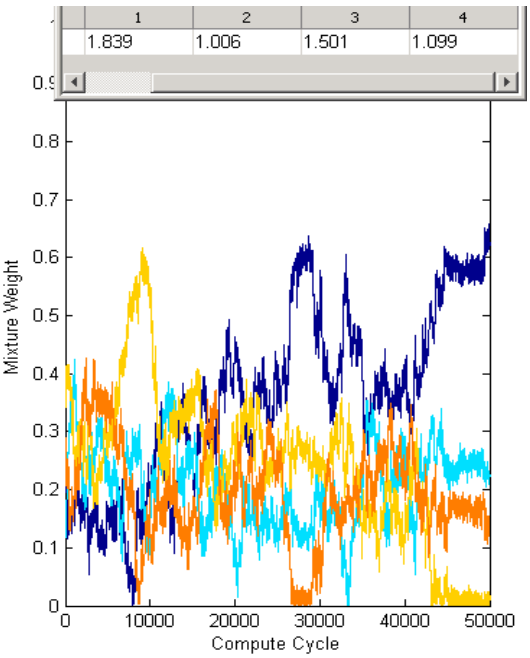
Mix Set 1
7:4.5:2.5:1
replicate



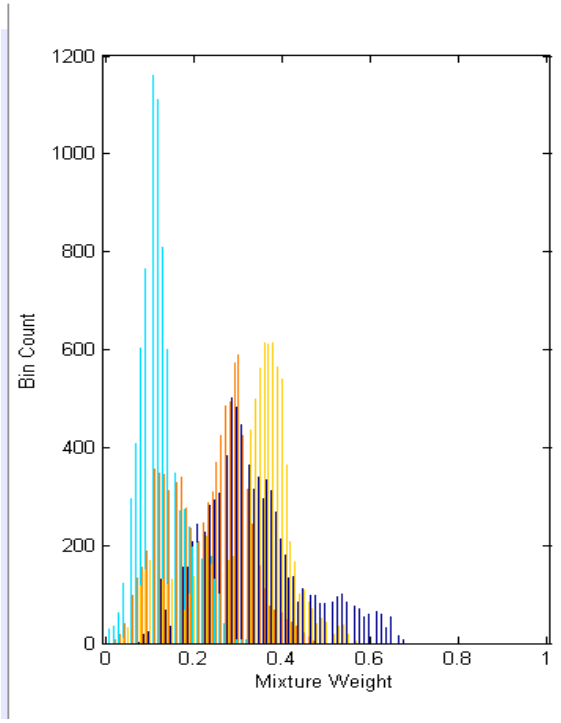


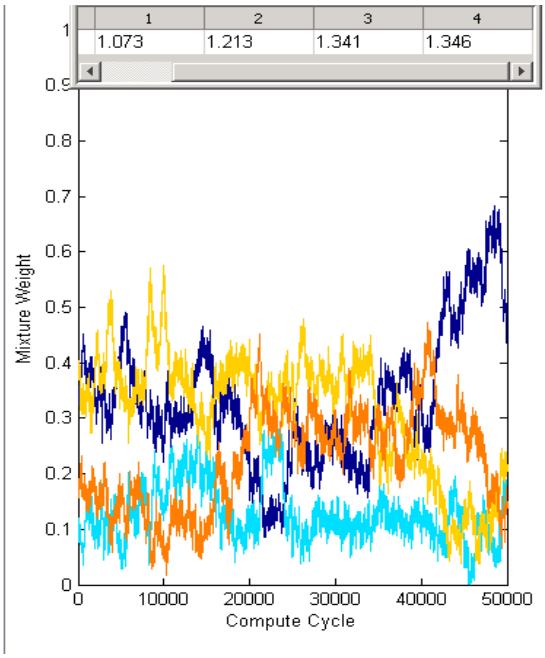
Mix Set 2
2.5:2:1.5:1



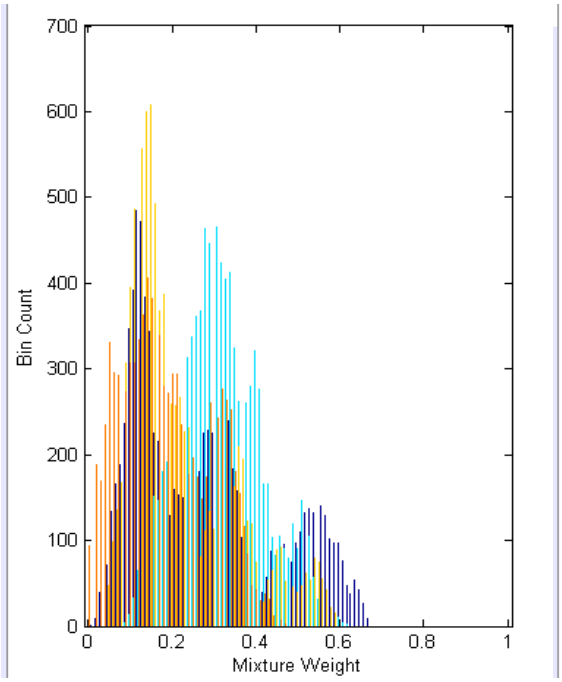


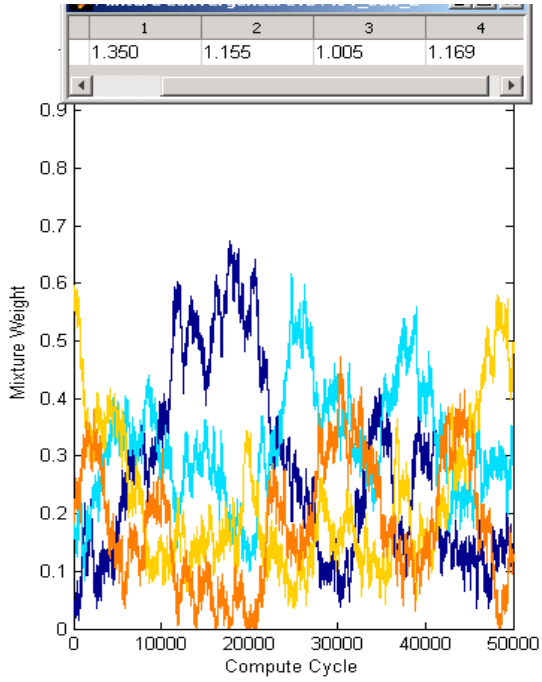
Mix Set 2
2.5:2:1.5:1
replicate



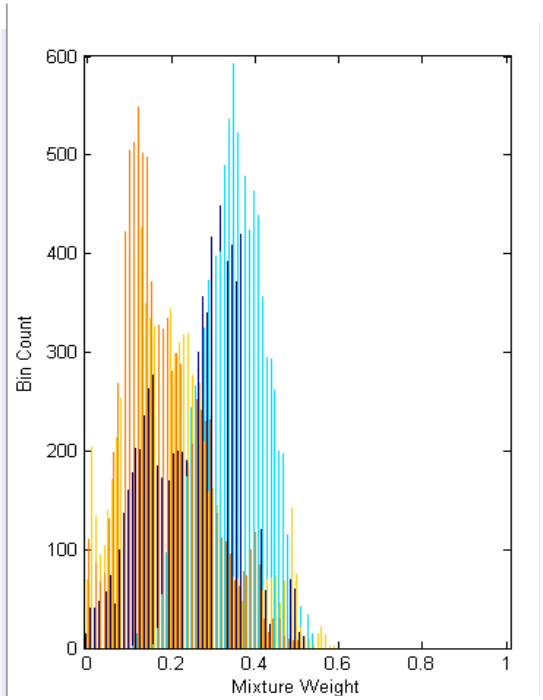


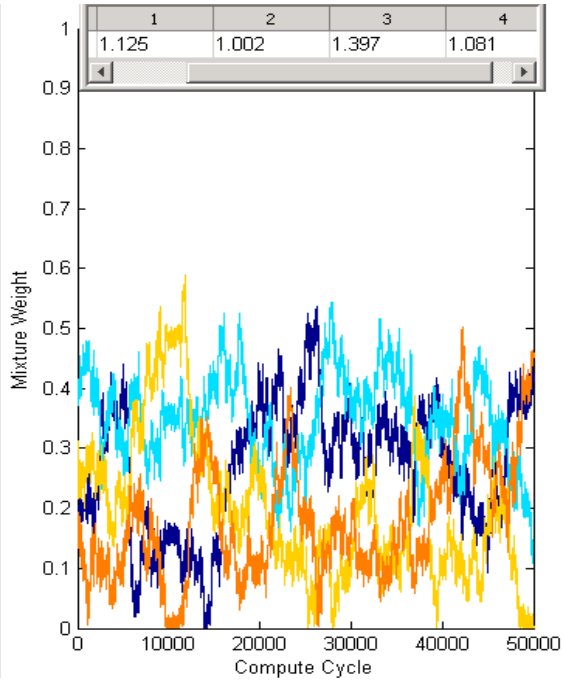
Mix Set 2
3.5:3:1.5:1



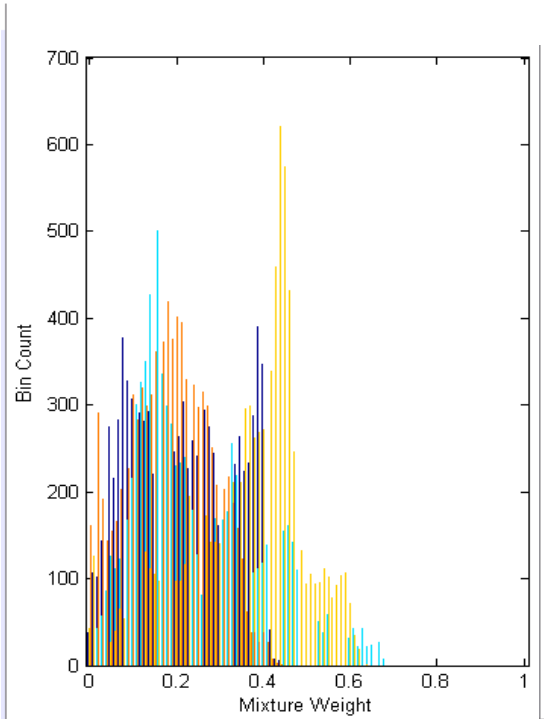


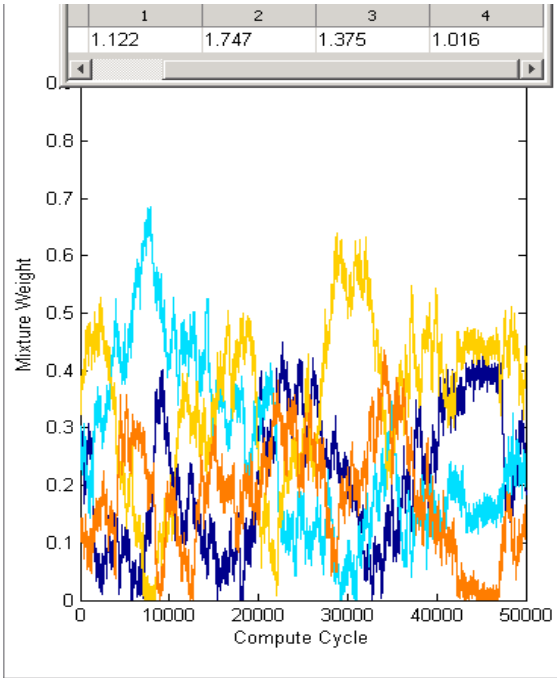
Mix Set 2
3.5:3:1.5:1
replicate



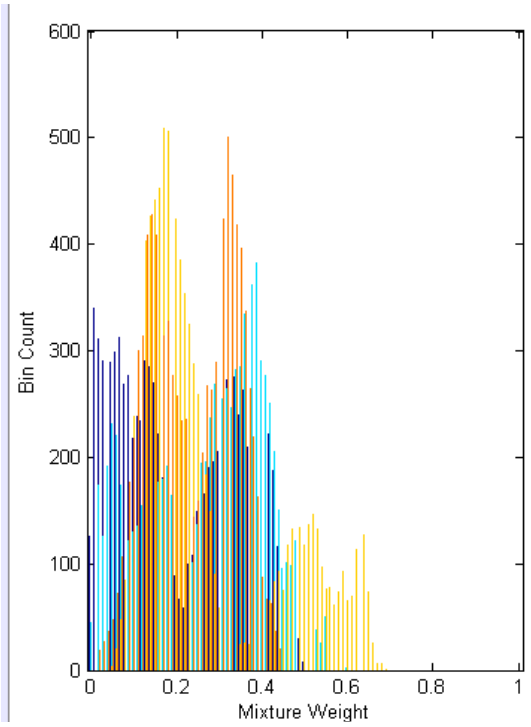


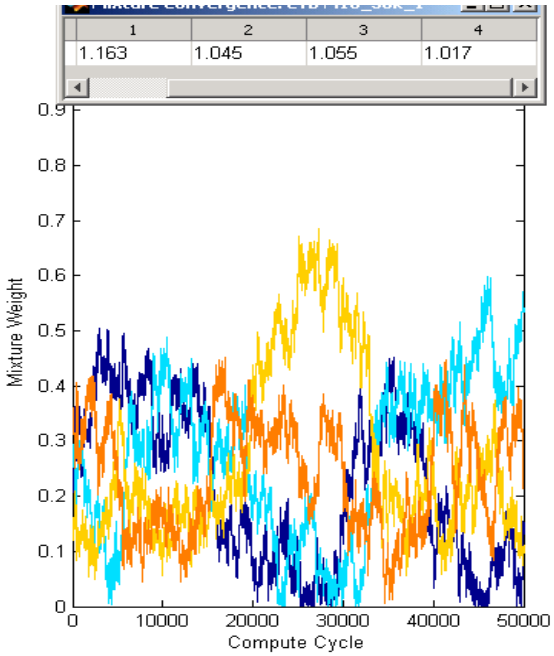
Mix Set 2 5:3:2:1



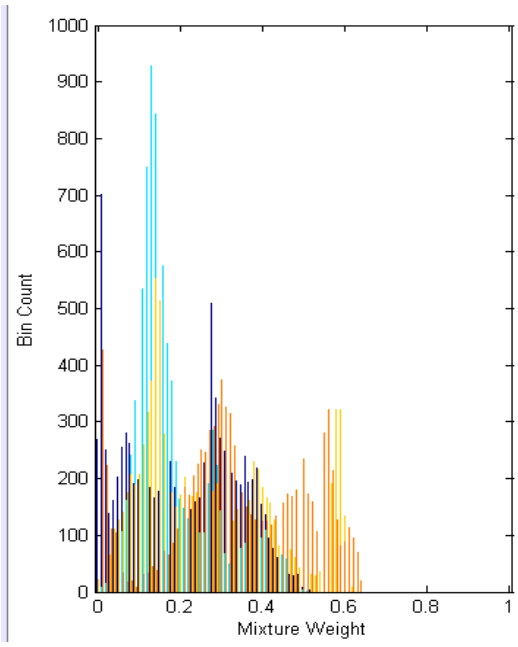


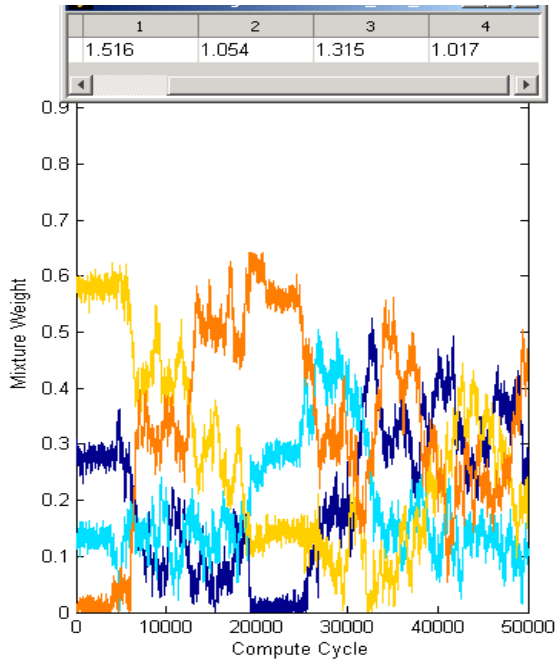
Mix Set 2 5:3:2:1 replicate





Mix Set 2
7:4.5:2.5:1





Mix Set 2
7:4.5:2.5:1
replicate

V. Experiment 2 Results: Familial Mixture Study

A. Specificity: Full-Sib Comparisons

For all pairwise full-sib comparisons at 1:1 mixture ratios, family 1 (F1 hereafter) exhibited a mean separation of 16.64 log units between full-sib donors and related non-donors (full-sibs and parents), and a mean separation of 13.18 log units was found for family 2 (F2 hereafter) (Figures V.A.1 and V.A.2). Consistent with expectations, log(LR) divergence from full-sib donors was highest for unrelated non-donors in both family datasets, with mean separations of 33.82 and 34.31 log units achieved for F1 and F2, respectively.

Minimum log(LR) separation between full-sib donors and individual related non-donors was 4.45 for F1 and 2.54 for F2 (Figures V.A.3 and V.A.4). Pairwise statistical comparisons between donor and related non-donor match statistics were significantly different for both families (Kruskal Wallis $p < 0.001$ for F1 and F2), indicating that specificity of genotype inference is sufficient to separate family members even with roughly equal mixture weights. Cross-matching to non-donor full-sibs was evident for both families, with the magnitude correlated strongly to the number of shared alleles ($R^2 = 0.839$; ANOVA $p < 0.001$) (Figure V.A.5). However, in all 1:1 full-sib mixtures tested, the two highest log(LR) values were always associated with the two siblings used to create the given mixture. Substantial differences in the degree of full-sib allele sharing were also noted between the two test families, and appear to be driven by both parent heterozygosity and random allele sharing between parents.

At 3:1 mixture ratios, inferred genotype specificity improved for all pairwise full-sib comparisons. Mean log(LR) separation between full-sib donors and related non-donors (full-sibs and parents) was 29.02 for F1 and 26.59 for F2 (Figures V.A.6 and V.A.7). Minimum log(LR) separation between full-sib donors and individual related non-donors improved to 12.22 for F1 and 8.47 for F2 (Figures V.A.8 and V.A.9). The relationship between full-sib non-donor log(LR) and number of shared alleles was also marginally significant at the 3:1 level, although the effect is diminished due to increased genotype separation ($R^2 = 0.403$; ANOVA $p < 0.043$) (Figure V.A.10). Similar to the 1:1 data, 3:1 pairwise statistical comparisons between full-sib donor and related non-donor match statistics (full-sibs and parents) were significantly different for both families (Kruskal Wallis $p < 0.001$ for F1 and F2). Overall specificity was also affected by mixture ratio, with greater divergence between full-sib donors and related non-donors evident in the 3:1 dataset (Figure V.A.11).

Combined data from all full-sib comparisons (including both mixture ratios for F1 and F2) show a mean log(LR) separation of 18.78 between non-donor and donor full sibs, and 23.68 between non-donor parents and donor full-sibs. (Figure V.A.12). Separation for unrelated non-donors was 36.43 log units. The data indicate a gradient of specificity with unrelated non-donors exhibiting the highest divergence from inferred donor profiles, related parents next highest, and full-sib non-donors showing the lowest divergence.

Figure V.A.1. Dot plot showing mean specificity for all F1 full-sib pairwise comparisons as a function of donor type at mixture ratios of 1:1. All requests were solved in duplicate (n = 12). Error bars represent one standard deviation; dashed line set at zero.

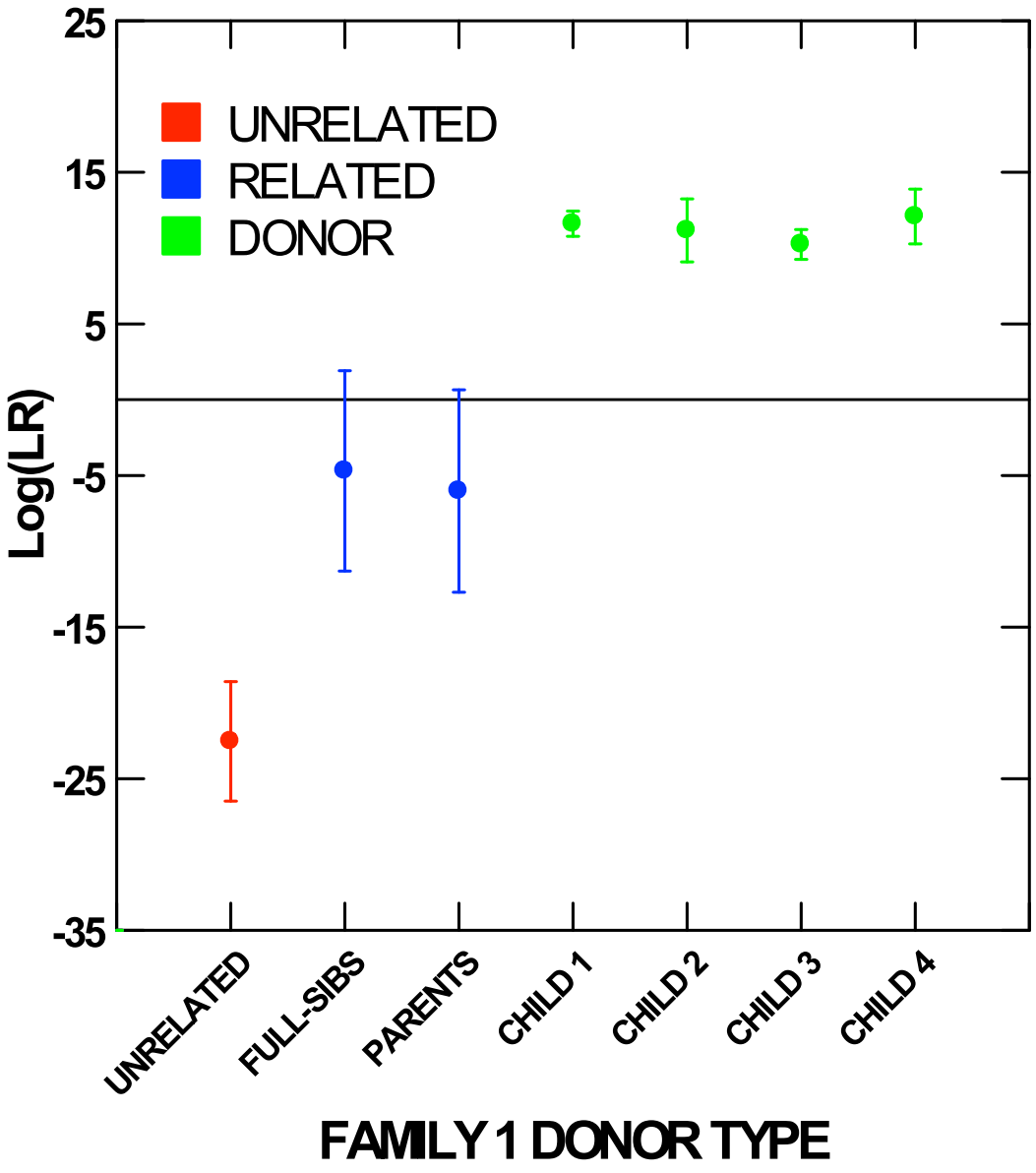


Figure V.A.2. Dot plot showing mean specificity for all F2 full-sib pairwise comparisons as a function of donor type at mixture ratios of 1:1. All requests were solved in duplicate (n = 12). Error bars represent one standard deviation; dashed line set at zero.

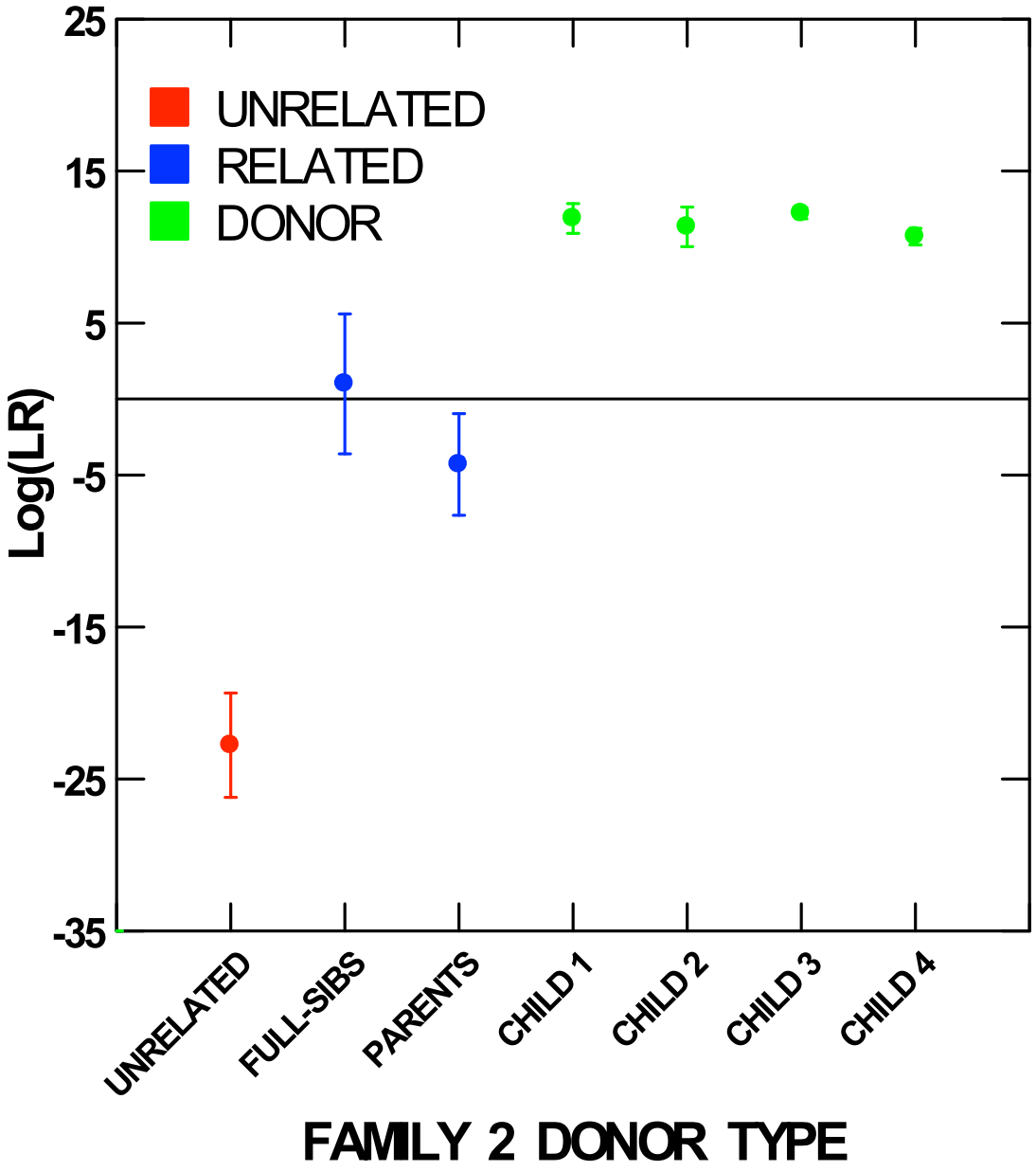
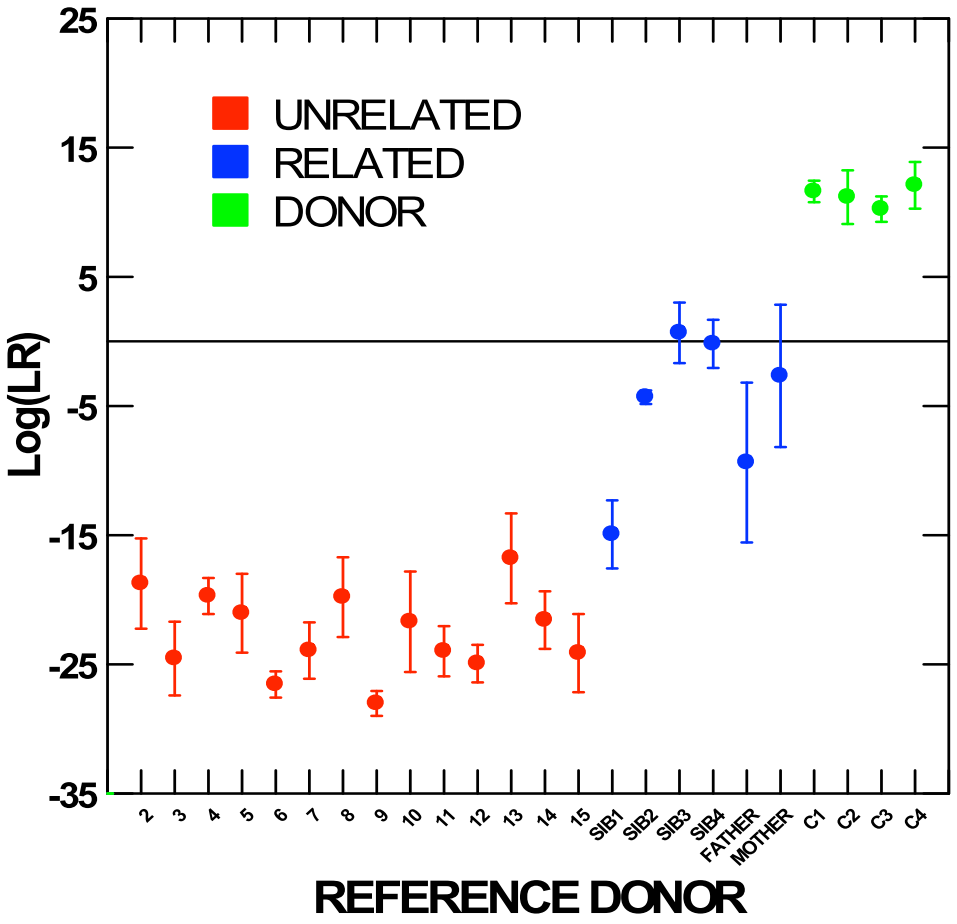


Figure V.A.3. Dot plot showing mean specificity for all F1 full-sib pairwise comparisons at mixture ratios of 1:1 by individual donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4). All requests were solved in duplicate (n = 12). Error bars represent one standard deviation; dashed line set at zero.



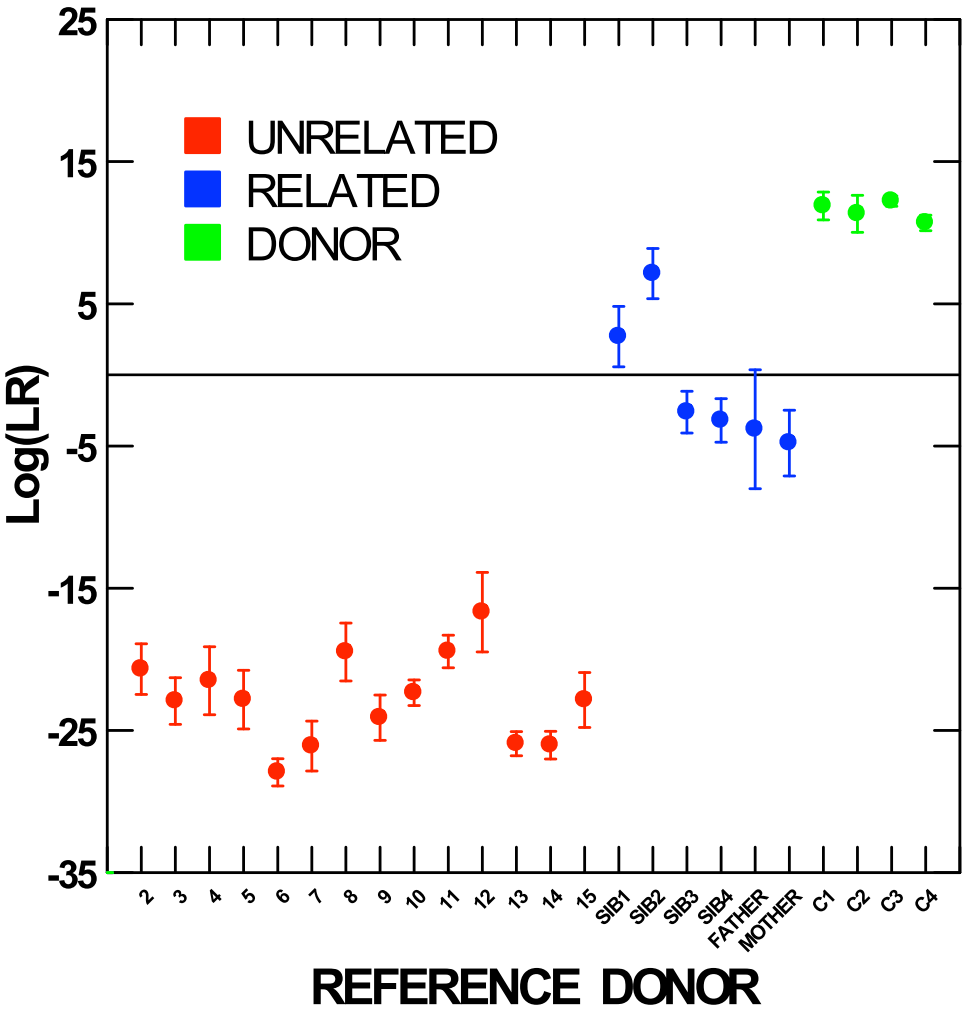
Average donor log(LR) = 11.28
St. dev. = 1.45

Average related non-donor log(LR) = -5.36
St. dev. = 6.57

Average unrelated non-donor log(LR) = -22.54
St. dev. = 3.94

Figure V.A.4. Dot plot showing mean specificity for all F2 full-sib pairwise comparisons at mixture ratios of 1:1 by individual donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 =

C4). All requests were solved in duplicate (n = 12). Error bars represent one standard deviation; dashed line set at zero.



Average donor log(LR) = 11.53
St. dev. = 0.95

Average related non-donor log(LR) = -1.66
St. dev. = 4.77

Average unrelated non-donor log(LR) = -22.78
St. dev. = 3.43

Figure V.A.5. Scatterplot showing non-donor full-sib log(LR) values as a function of the number of shared alleles for 1:1 mixture ratios. Both family datasets are combined; dashed line set at zero.

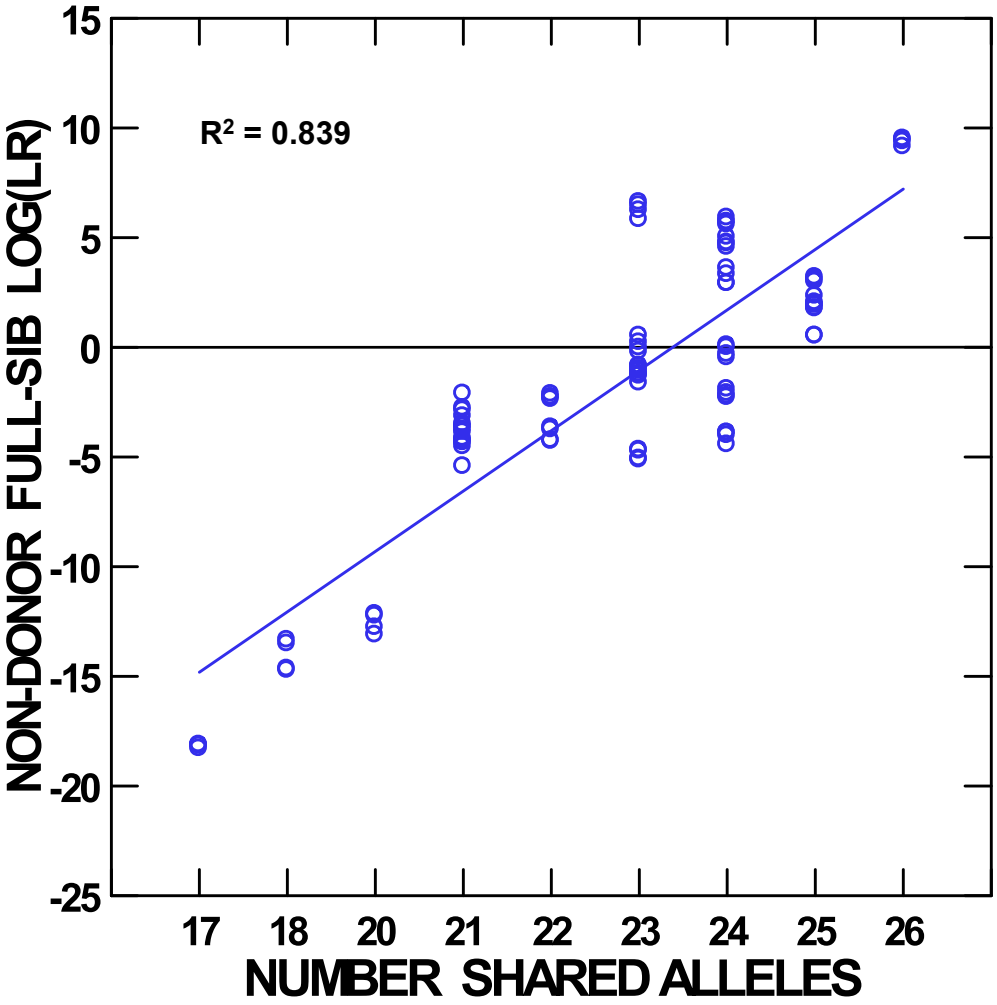


Figure V.A.6. Dot plot showing mean specificity for all F1 full-sib pairwise comparisons as a function of donor type at mixture ratios of 3:1. All requests were solved in duplicate (n = 12). Error bars represent one standard deviation; dashed line set at zero.

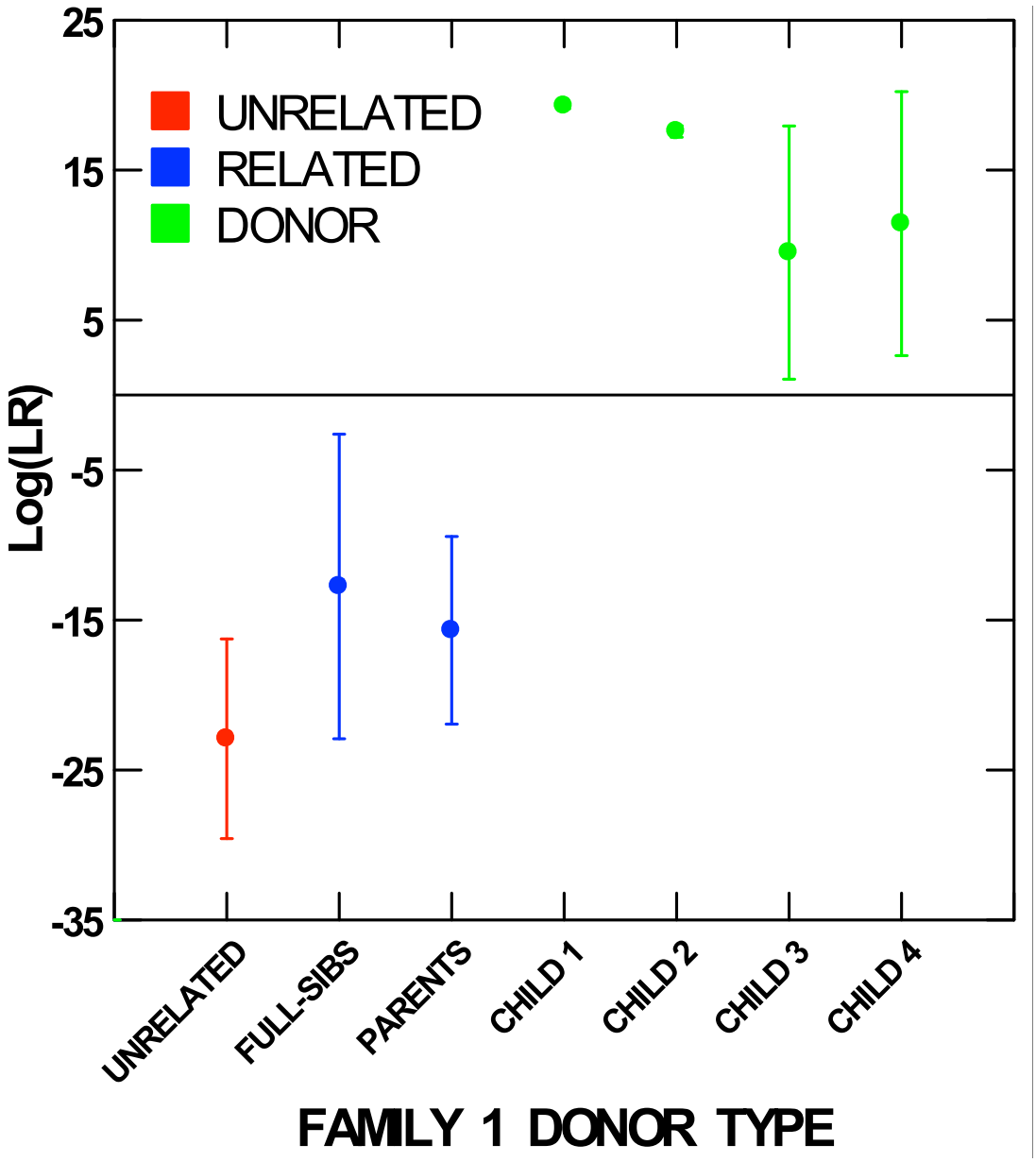


Figure V.A.7. Dot plot showing mean specificity for all F2 full-sib pairwise comparisons as a function of donor type at mixture ratios of 3:1. All requests were solved in duplicate (n = 12). Error bars represent one standard deviation; dashed line set at zero.

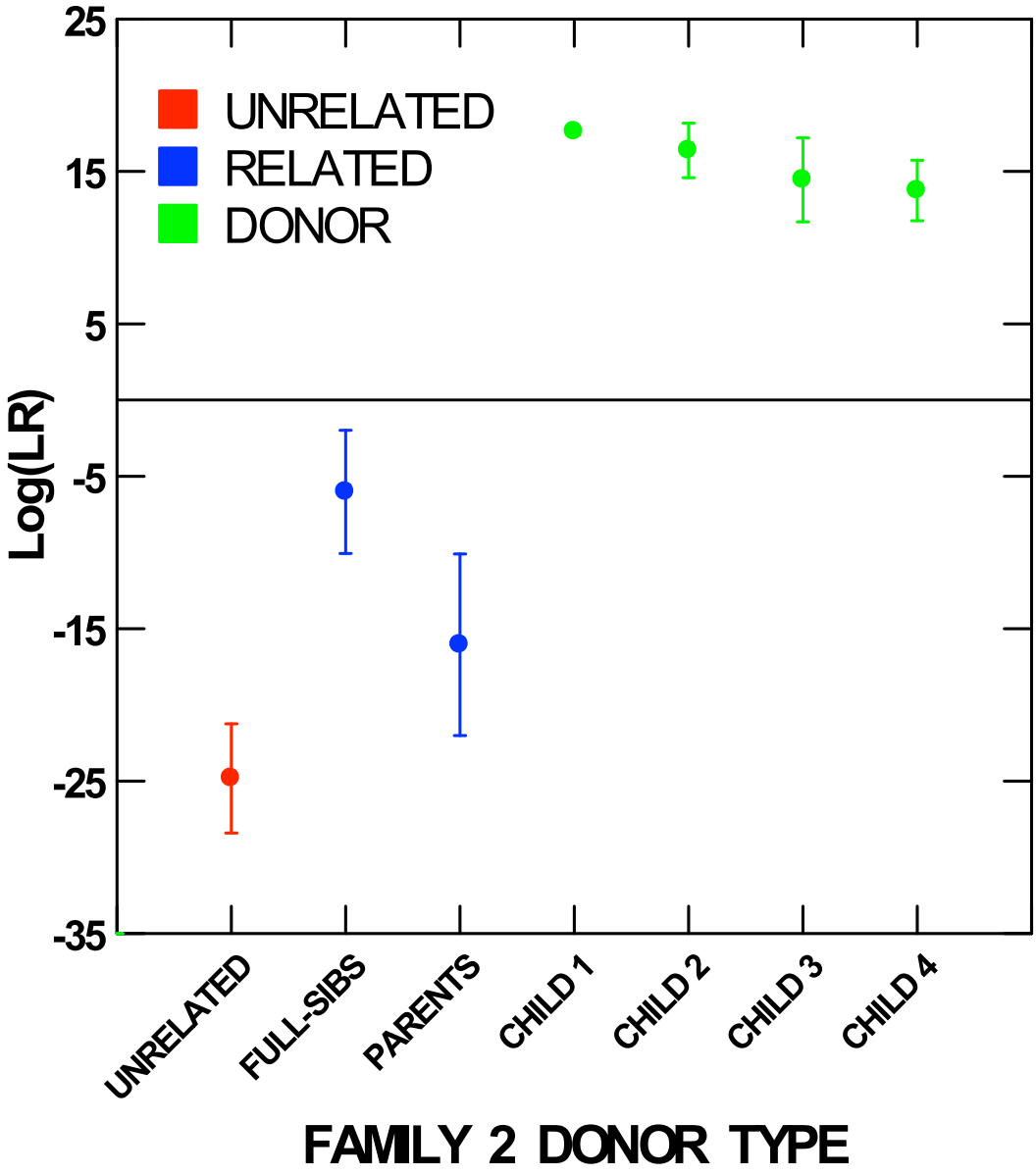
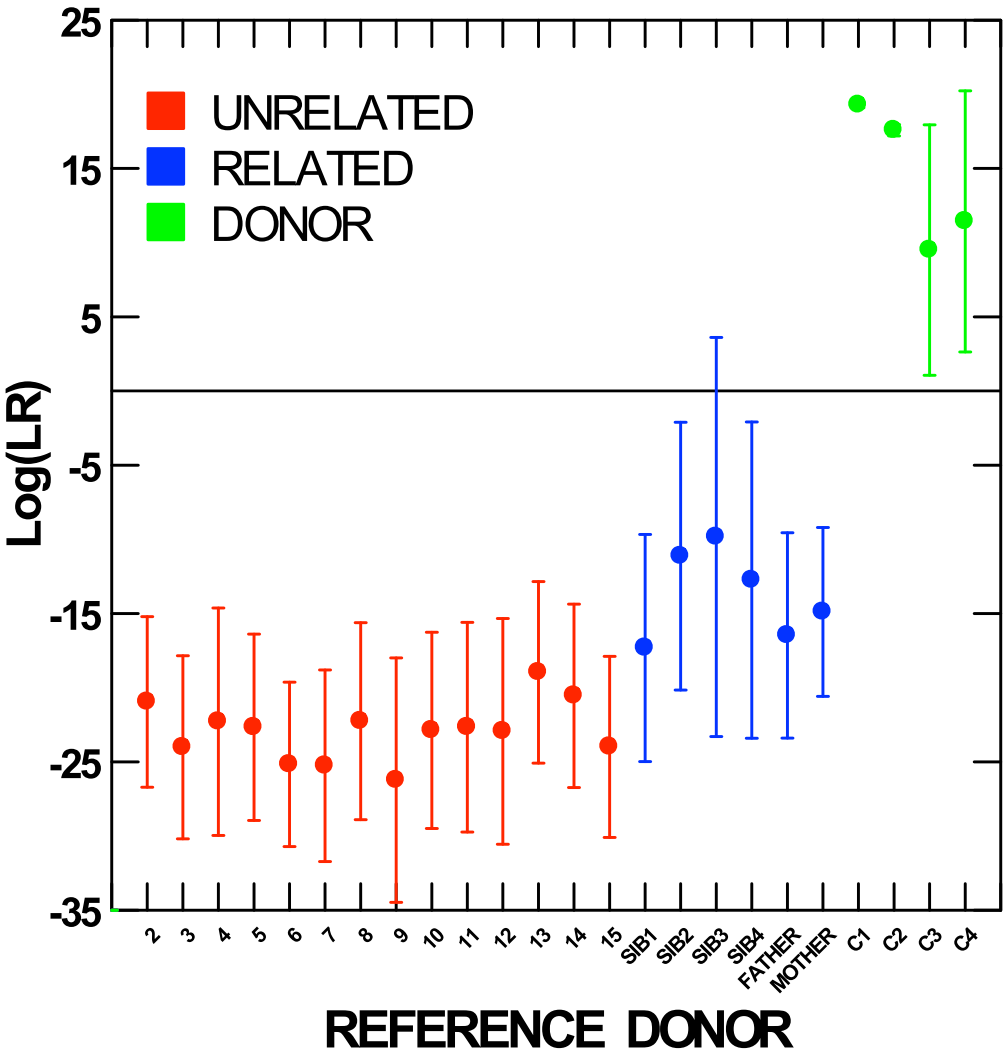


Figure V.A.8. Dot plot showing mean specificity for all F1 full-sib pairwise comparisons at mixture ratios of 3:1 by individual donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4). All requests were solved in duplicate (n = 12). Error bars represent one standard deviation; dashed line set at zero.

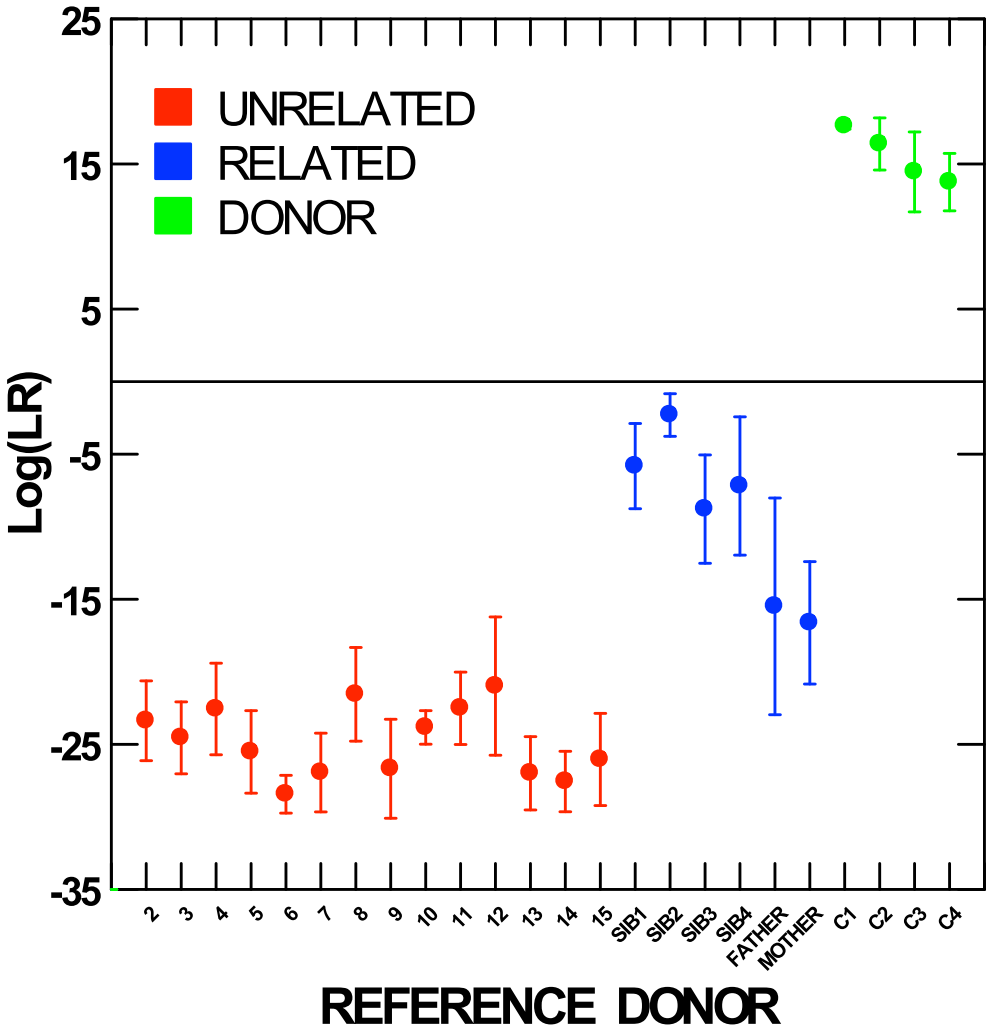


Average donor log(LR) = 14.79
St. dev. = 6.33

Average related non-donor log(LR) = -14.23
St. dev. = 8.43

Average unrelated non-donor log(LR) = -22.92
St. dev. = 6.65

Figure V.A.9. Dot plot showing mean specificity for all F2 full-sib pairwise comparisons at mixture ratios of 3:1 by individual donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4). All requests were solved in duplicate (n = 12). Error bars represent one standard deviation; dashed line set at zero.



Average donor log(LR) = 15.55
St. dev. = 2.25

Average related non-donor log(LR) = -11.04
St. dev. = 7.11

Average unrelated non-donor log(LR) = -24.82
St. dev. = 3.58

Figure V.A.10. Scatterplot showing non-donor full-sib log(LR) values as a function of the number of shared alleles for 3:1 mixture ratios. Both family datasets are combined; dashed line set at zero.

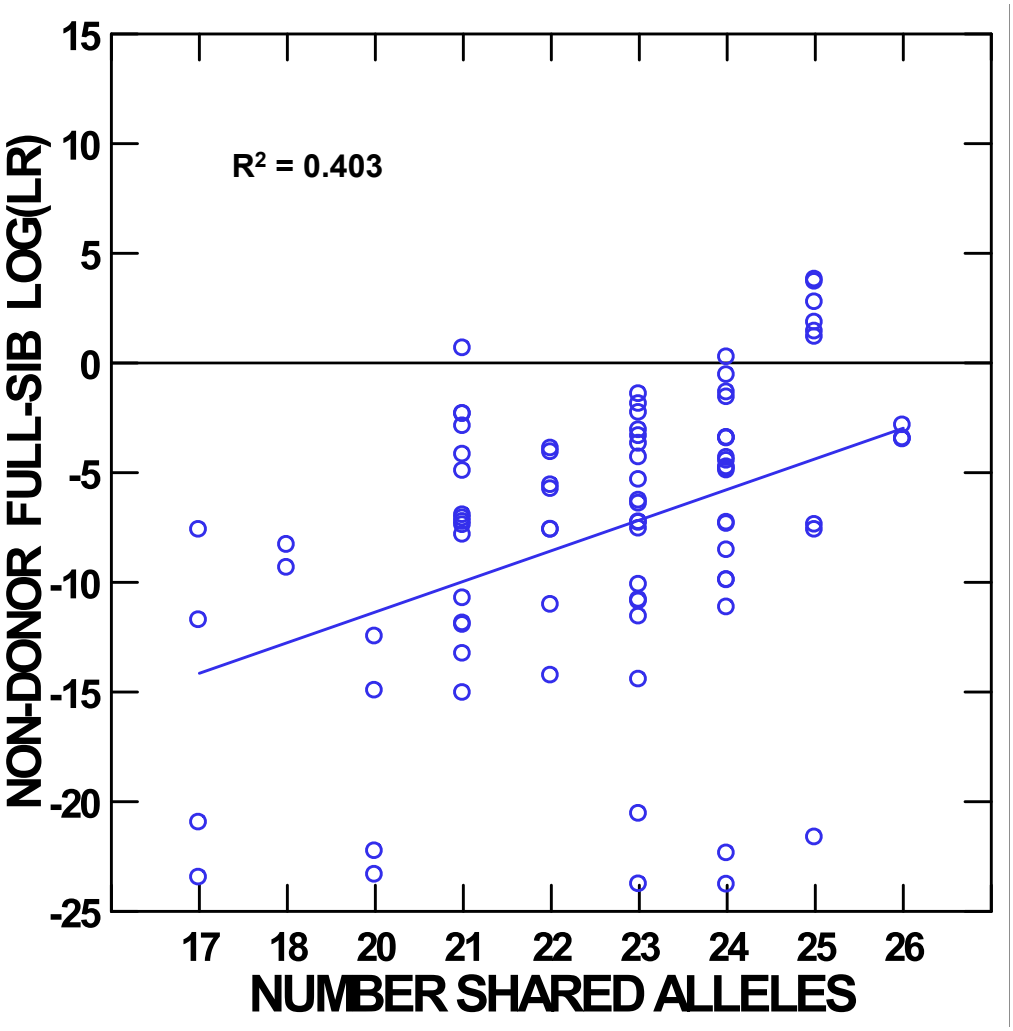


Figure V.A.11. Boxplot showing median log(LR) values for full-sib comparisons combined as a function of donor type and grouped by mixture ratios. Both family datasets are combined. Dashed line set at zero; asterisks and circles denote outliers.

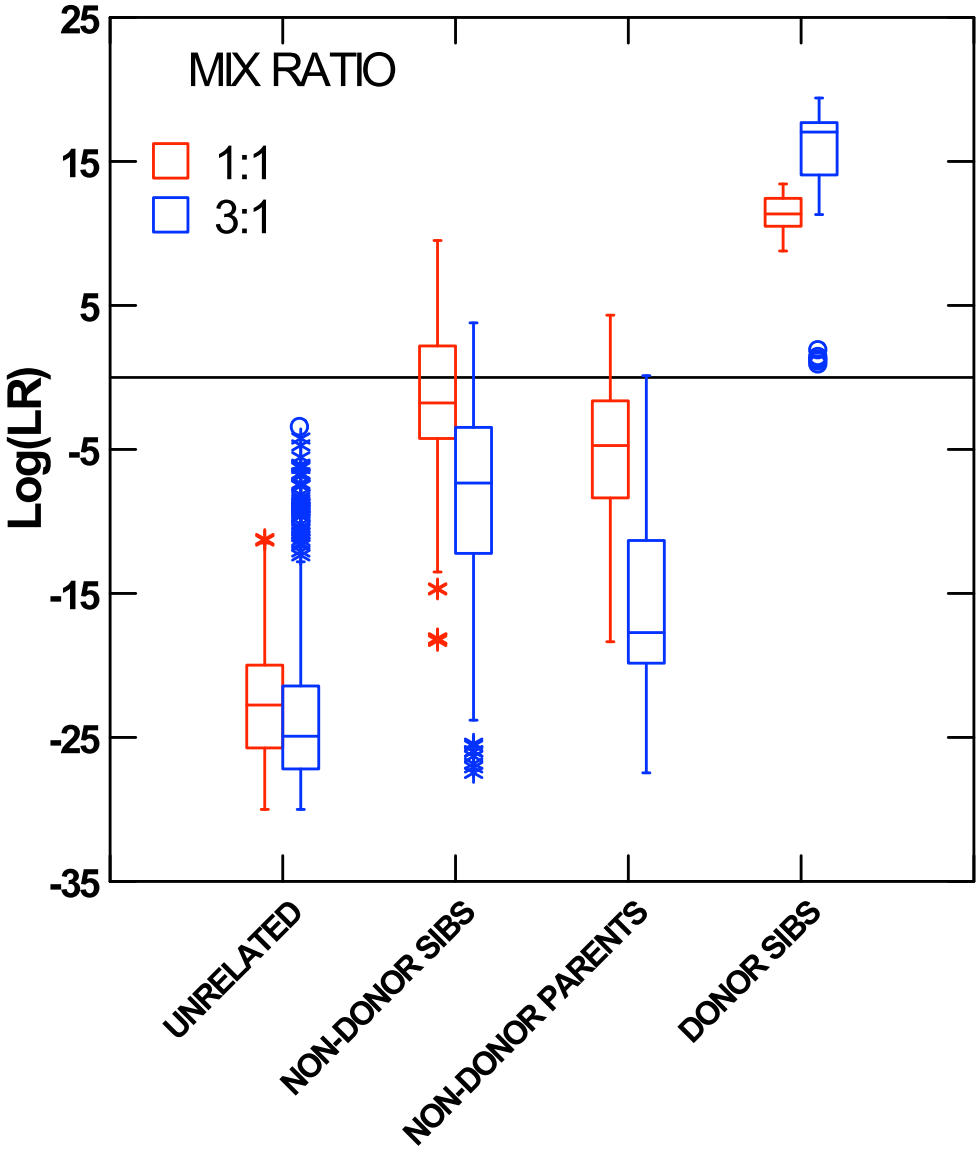
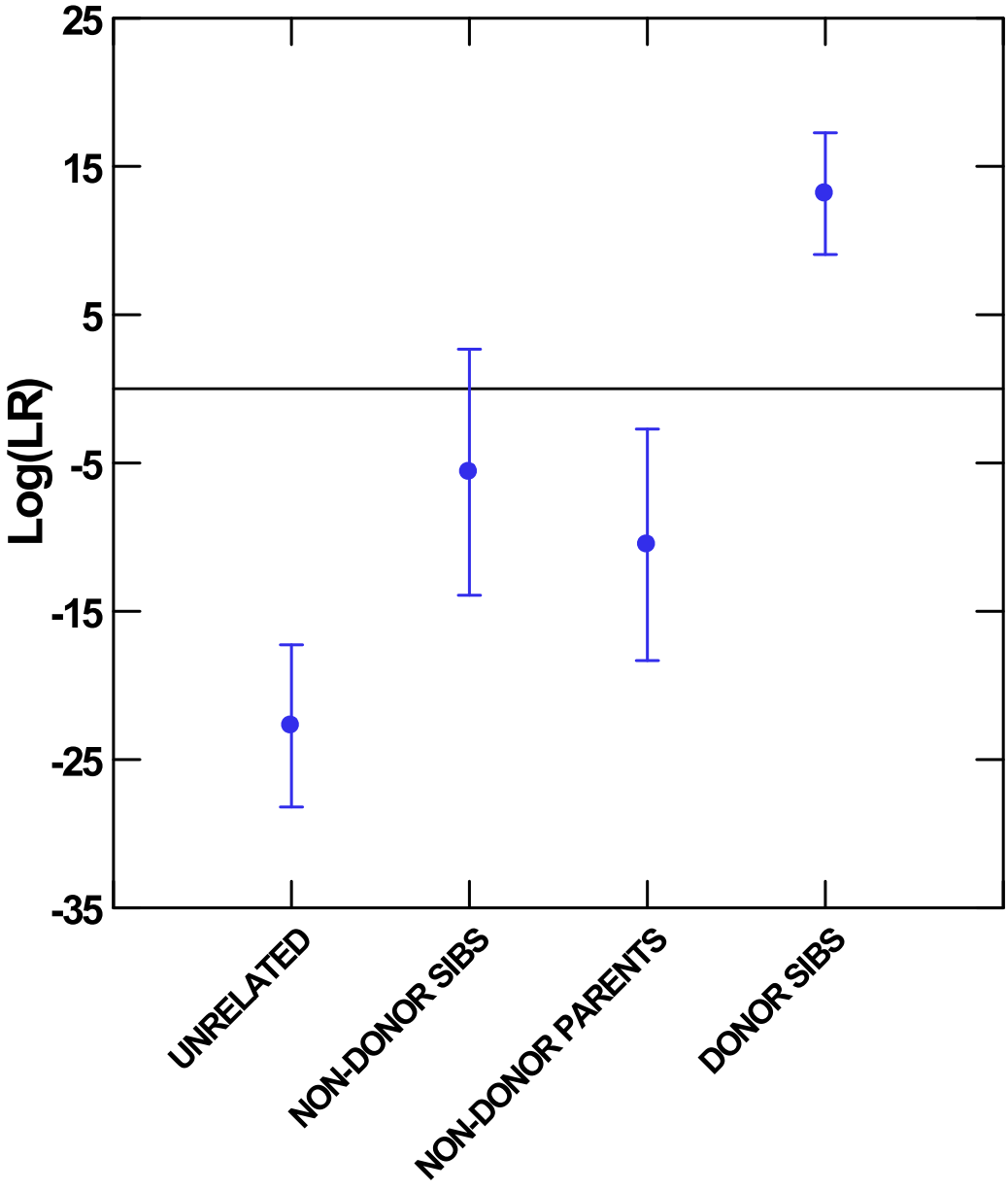


Figure V.A.12. Dot plot showing mean experiment-wide log(LR) values from all full-sib comparisons as a function of donor type. F1 and F2 datasets and both mixture ratios are combined. Error bars represent one standard deviation; dashed line set at zero.



B. Specificity: Parent-Child Comparisons (Two-Unknown Donors)

At 1:1 mixture ratios for both families, specificity of genotype inference was greater for parent-child comparisons than for full-sib comparisons. Family 1 exhibited a mean separation of 18.9 log units between full-sib donors and related non-donors (full-sibs and parents), and 13.36 log units was observed for family 2 (Figures V.B.1 and V.B.2 and see Table V.B.1). Minimum individual separation was 5.76 log units for family 1 and 2.83 for family 2
(Figures V.B.3 and V.B.4). Similar to the full-sib comparison dataset, log(LR) divergence from full-sib donors was highest for unrelated non-donors in both family datasets, with mean separations of 34.19 and 34.11 log units achieved for F1 and F2, respectively.

At 3:1 mixture ratios, inferred genotype specificity improved for all pairwise full-sib comparisons. Mean log(LR) separation between full-sib donors and related non-donors (full-sibs and parents) was 34.61 for F1 and 28.64 for F2 (Figures V.B.5 and V.B.6). Minimum individual separation approximately doubled from the 1:1 mixture ratio level, with 11.04 log units observed for family 1 and 5.50 seen for family 2 (Figures V.B.7 and V.B.8). Similar to the full-sib comparison dataset, specificity was significantly greater at the 3:1 ratios than at 1:1 ratios (Kruskal Wallis $p < 0.001$ for both families) (Figure V.B. 9).

Combined data from all parent-child comparisons (including both mixture ratios for F1 and F2) show a mean log(LR) separation of 23.88 between related non-donors (parents and children) and related donors (Figure V.B.10). Mean separation for unrelated non-donors was 37.89 log units.

Table V.B.1. Specificity statistics for all parent-child familial mixture comparisons.

Comparison Type	Mean Related Non-Donor Log(LR) Séparation	Mean Unrelated Non-Donor Log(LR) Separation	Minimum Related Non-Donor Log(LR) Separation
F1 Parent-child 1:1	**18.9**	**34.25**	**5.76**
F1 Parent-child 3:1	**34.61**	**42.75**	**11.04**

F2 Parent-child 1:1	**13.36**	**34.10**	**2.83**
F2 Parent-child 3:1	**28.64**	**40.47**	**5.50**

Figure V.B.1. Dot plot showing mean specificity for all F1 parent-child pairwise comparisons as a function of donor type at mixture ratios of 1:1. All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

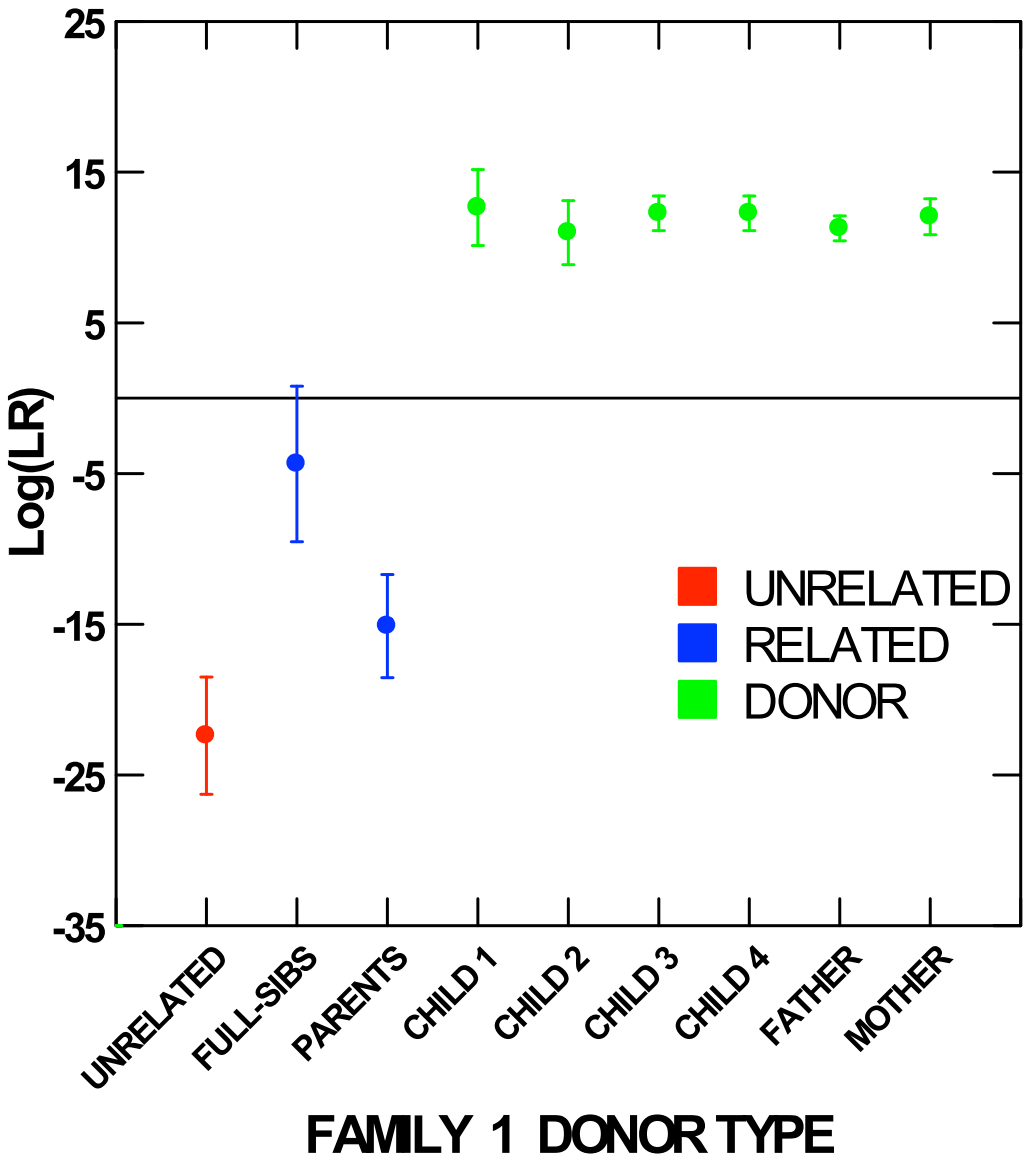


Figure V.B.2. Dot plot showing mean specificity for all F2 parent-child pairwise comparisons as a function of donor type at mixture ratios of 1:1. All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

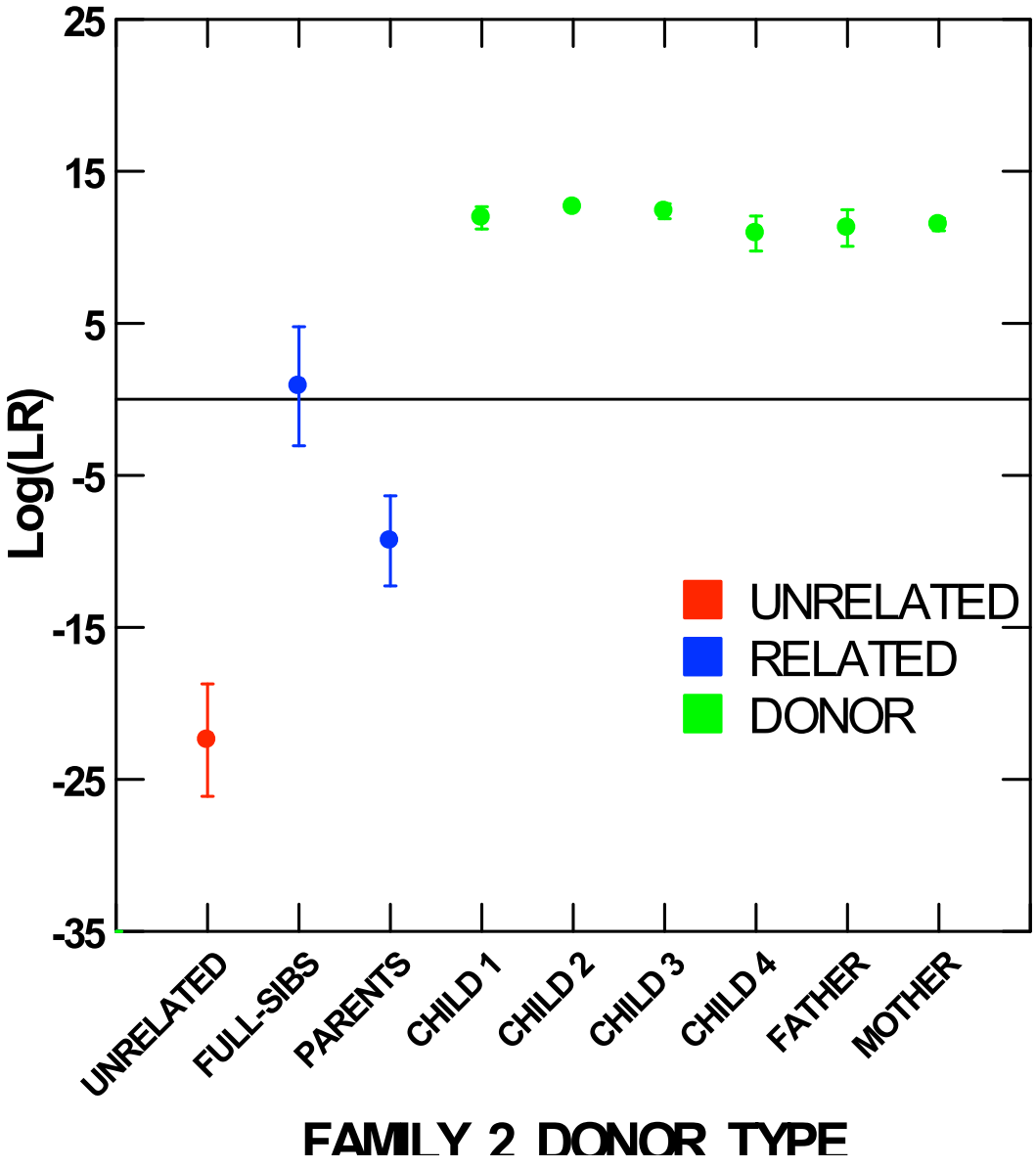


Figure V.B.3. Dot plot showing mean specificity for all F1 parent-child pairwise comparisons at mixture ratios of 1:1 by individual reference donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4, FATHER = P1, MOTHER = P2). All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

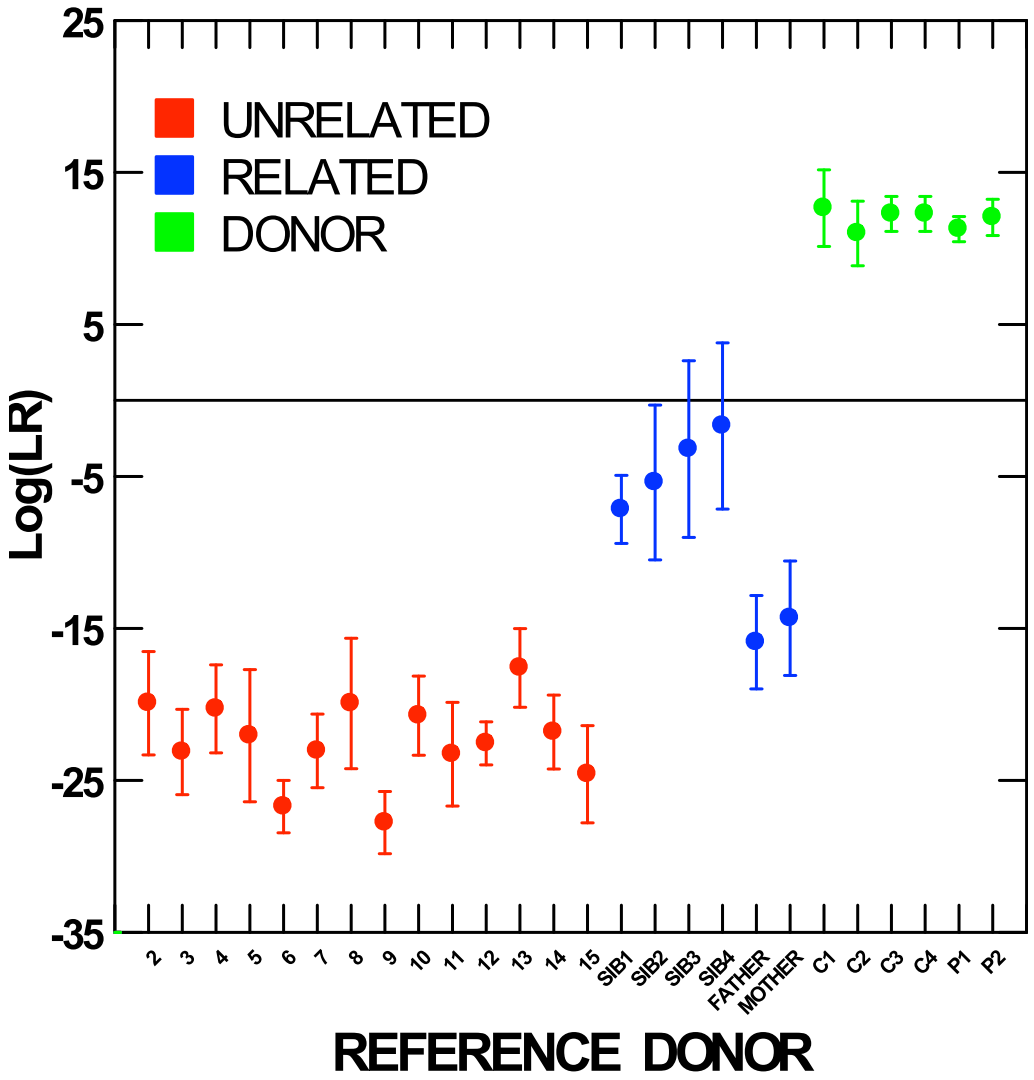


Figure V.B.4. Dot plot showing mean specificity for all F2 parent-child pairwise comparisons at mixture ratios of 1:1 by individual reference donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4, FATHER = P1, MOTHER = P2). All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

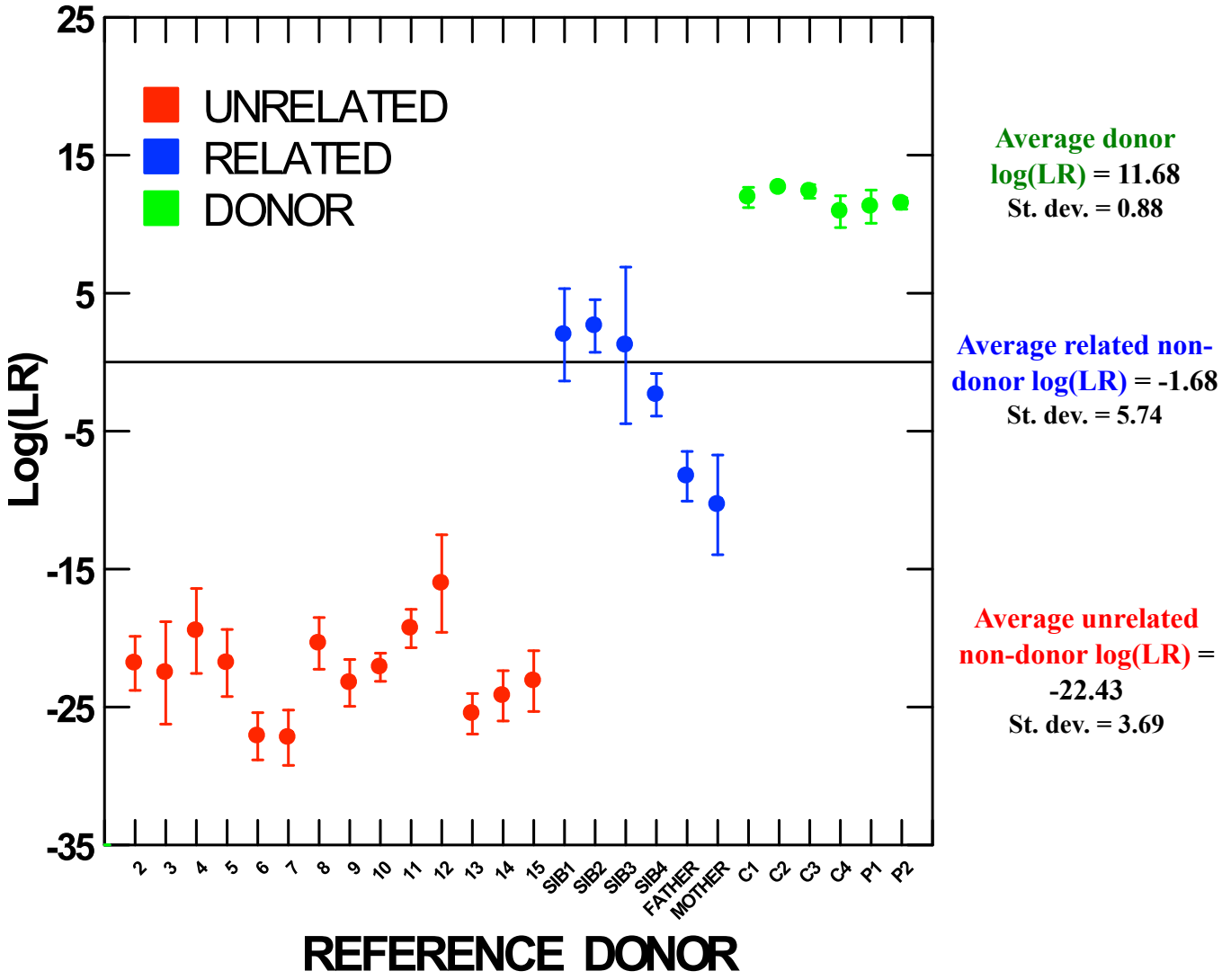


Figure V.B.5. Dot plot showing mean specificity for all F1 parent-child pairwise comparisons as a function of donor type at mixture ratios of 3:1. All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

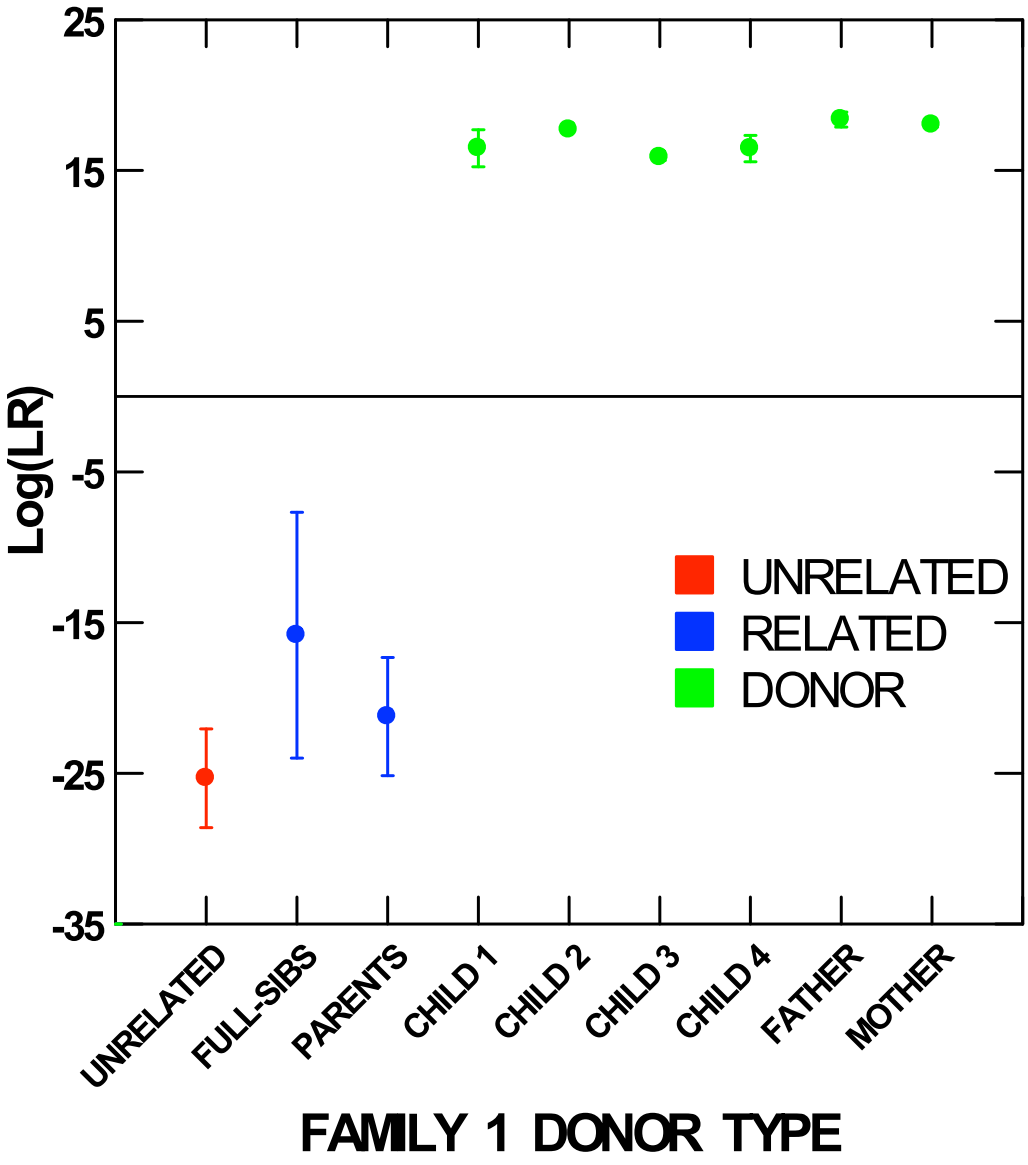


Figure V.B.6. Dot plot showing mean specificity for all F2 parent-child pairwise comparisons as a function of donor type at mixture ratios of 3:1. All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

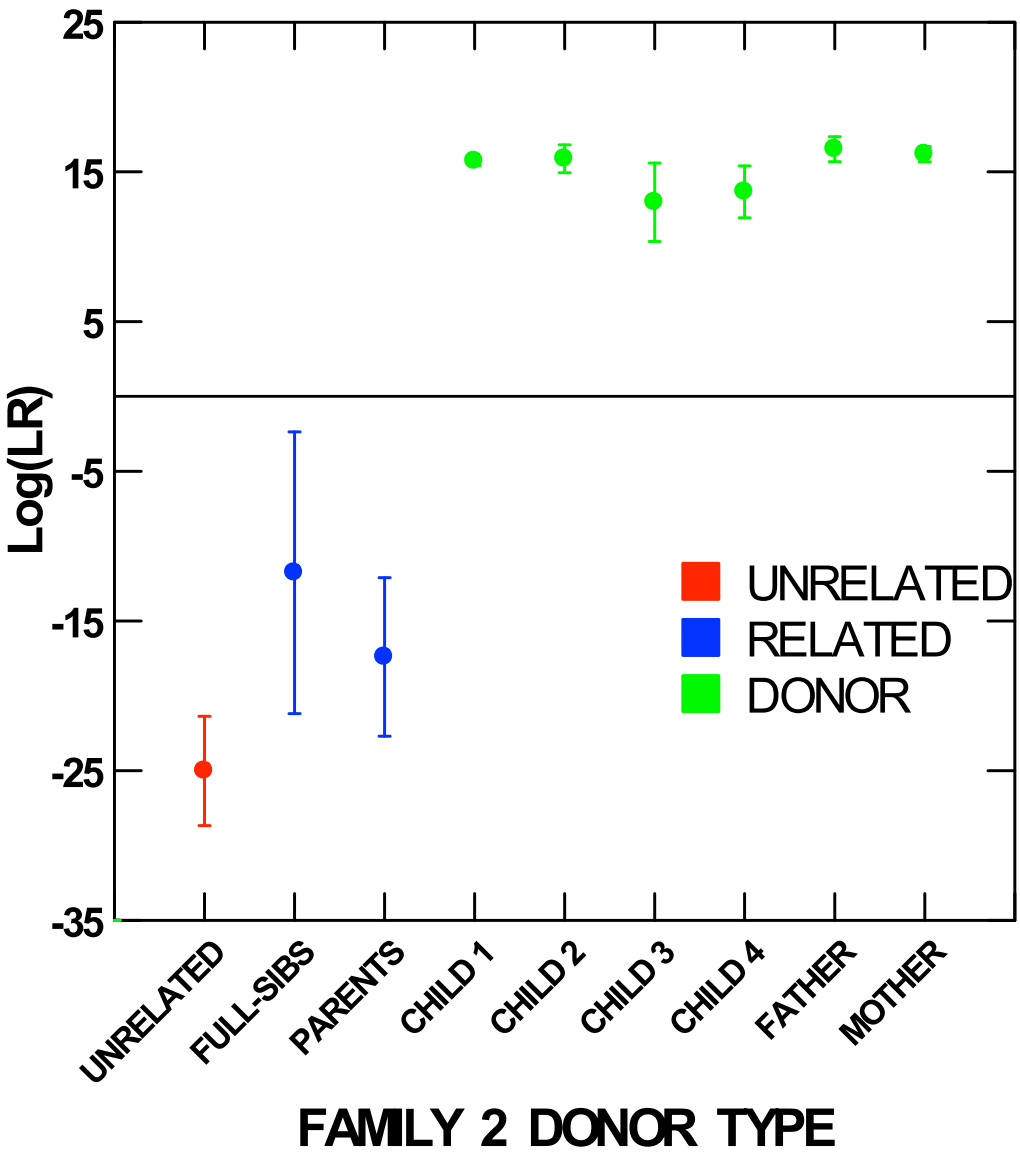


Figure V.B.7. Dot plot showing mean specificity for all F1 parent-child pairwise comparisons at mixture ratios of 3:1 by individual reference donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4, FATHER = P1, MOTHER = P2). All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

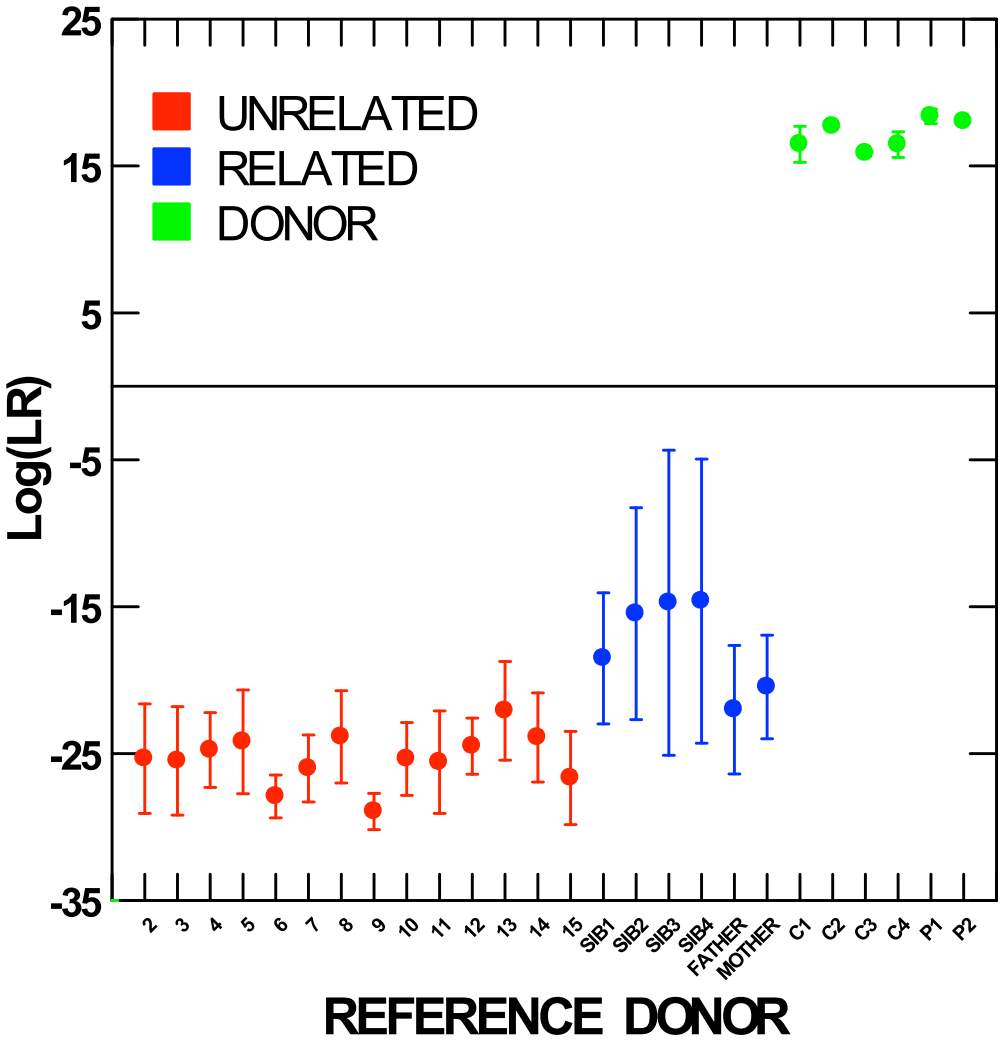


Figure V.B.8. Dot plot showing mean specificity for all F2 parent-child pairwise comparisons at mixture ratios of 3:1 by individual reference donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4, FATHER = P1, MOTHER = P2). All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

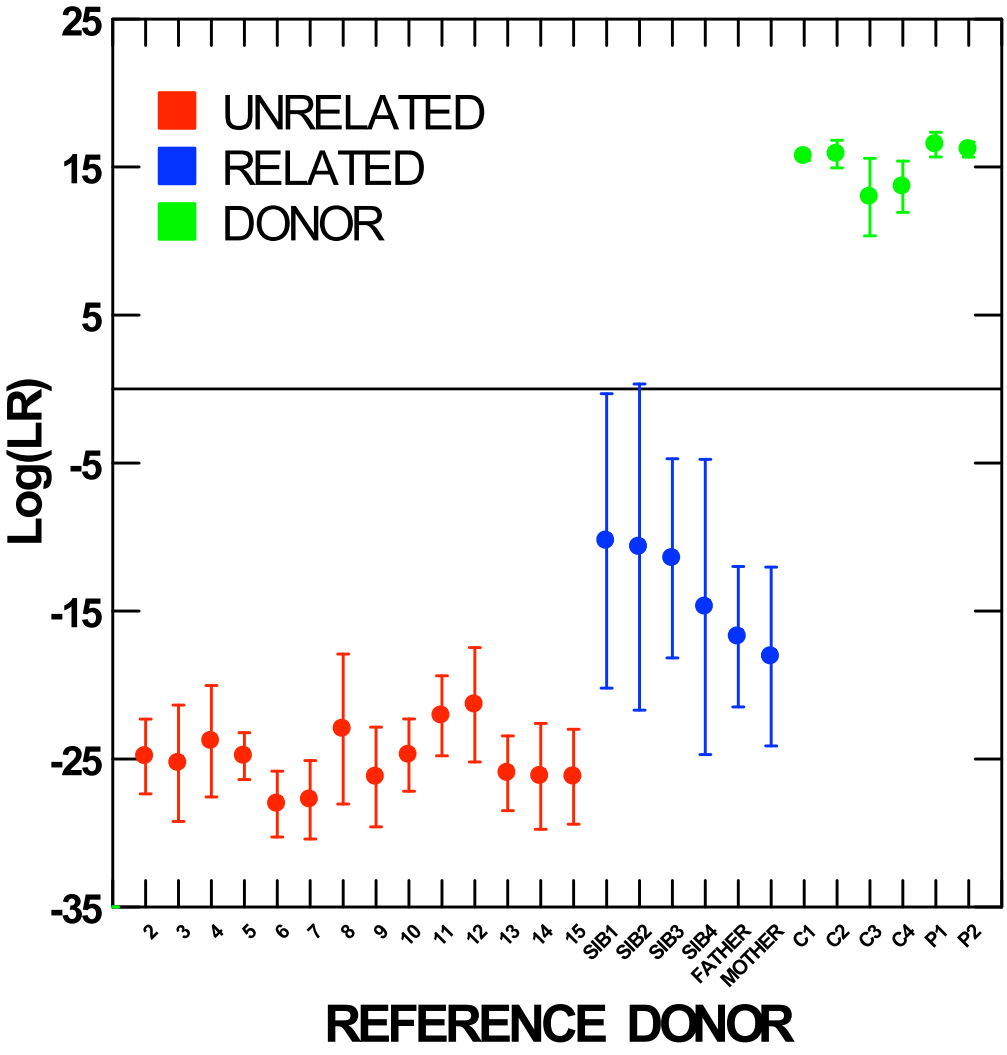


Figure V.B.9. Boxplot showing median log(LR) values for parent-child comparisons as a function of donor type and grouped by mixture ratio. Both family datasets are combined. Dashed line set at zero; asterisks and circles denote outliers.

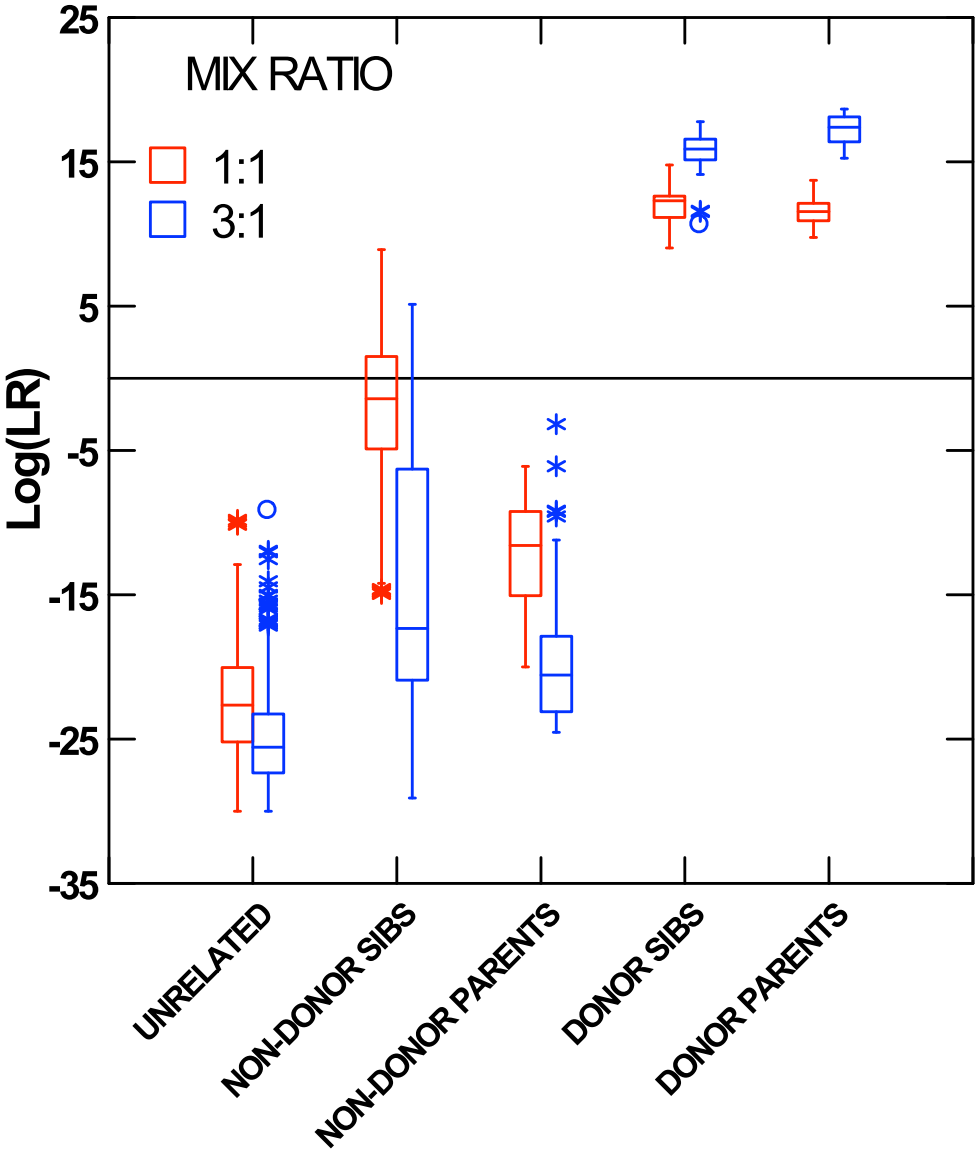
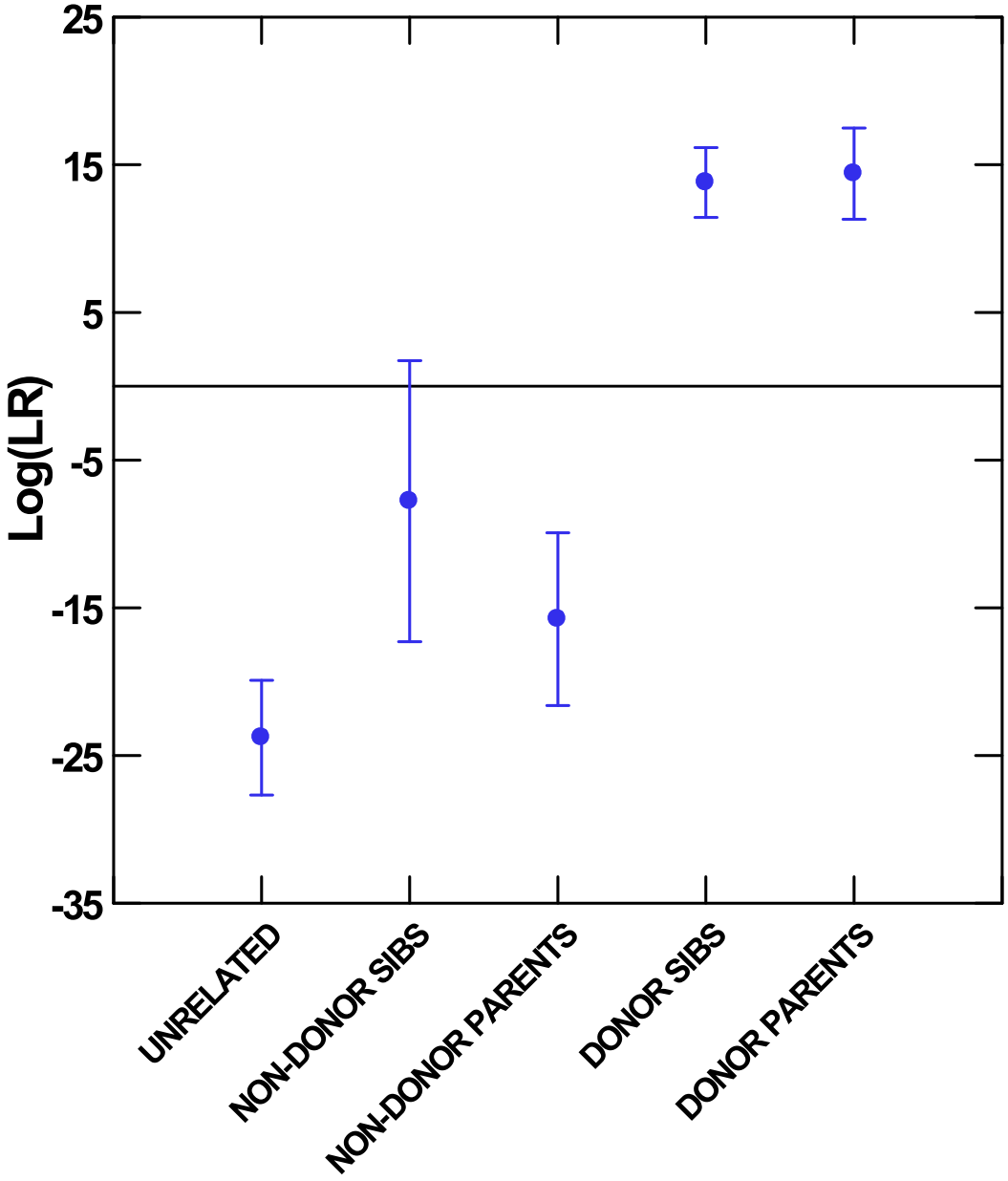


Figure V.B.10. Dot plot showing mean experiment-wide log(LR) values from all parent-child comparisons as a function of donor type. F1 and F2 datasets and both mixture ratios are combined. Error bars represent one standard deviation; dashed line set at zero.



C. Specificity: Parent-Child Comparisons (One-Unknown Donor)

Specificity for all familial mixtures was highest when solving for one unknown donor. At 1:1 mixture ratios, family 1 exhibited a mean separation of 35.23 log units between related donors and related non-donors (full-sibs and parents), and 30.38 log units was observed for family 2 (Figures V.C.1 and V.C.2 and see Table V.C.1). Minimum individual separation was 10.67 log units for family 1 and 12.24 for family 2 (Figures V.C.3 and V.C.4). Similar to the full-sib and parent-child two unknown comparisons, log(LR) divergence from all donors was highest for unrelated non-donors in both family datasets, with mean separations of 41.88 and 40.76 log units achieved for F1 and F2, respectively.

Consistent with expectations, specificity improved further still in the 3:1 mixture ratios. Mean log(LR) separation between related donors and related non-donors (full-sibs and parents) was 37.64 for F1 and 32.12 for F2 (Figures V.C.5. and V.C.6). Improvements in minimum individual separation were also seen when compared with the 1:1 data, with 16.89 log units observed for family 1 and 12.97 seen for family 2 (Figures V.C.7 and V.C. 8 and see V.C.9 for mixture ratio comparison).

Combined data from all parent-child one unknown comparisons (including both mixture ratios for F1 and F2) show a mean log(LR) separation of 33.84 between related non-donors (parents and children) and related donors (Figure V.C.10). Mean separation for unrelated non-donors was 42.26 log units. Overall, pairwise differences between related donor and related non-donor log(LR) match statistics were highly significant for all comparison types (full-sib, parent-child, and parent-child one unknown) (Kruskal Wallis $p < 0.001$) (Figure V.C.11), indicating that TrueAllele can successfully discriminate between closely related familial genotypes.

Table V.C.1. Specificity statistics for all parent-child one unknown familial mixture comparisons.

Comparison Type	Mean Related Non-Donor Log(LR) Séparation	Mean Unrelated Non-Donor Log(LR) Separation	Minimum Related Non-Donor Log(LR) Separation
F1 Parent-child 1:1 (One Unknown)	**35.23**	**41.88**	**10.67**
F1 Parent-child 3:1 (One Unknown)	**37.64**	**44.20**	**16.89**
F2 Parent-child 1:1 (One Unknown)	**30.38**	**40.76**	**12.24**
F2 Parent-child 3:1 (One Unknown)	**32.12**	**42.18**	**12.97**

Figure V.C.1. Dot plot showing mean specificity for all F1 parent-child one-unknown pairwise comparisons as a function of donor type at mixture ratios of 1:1. All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

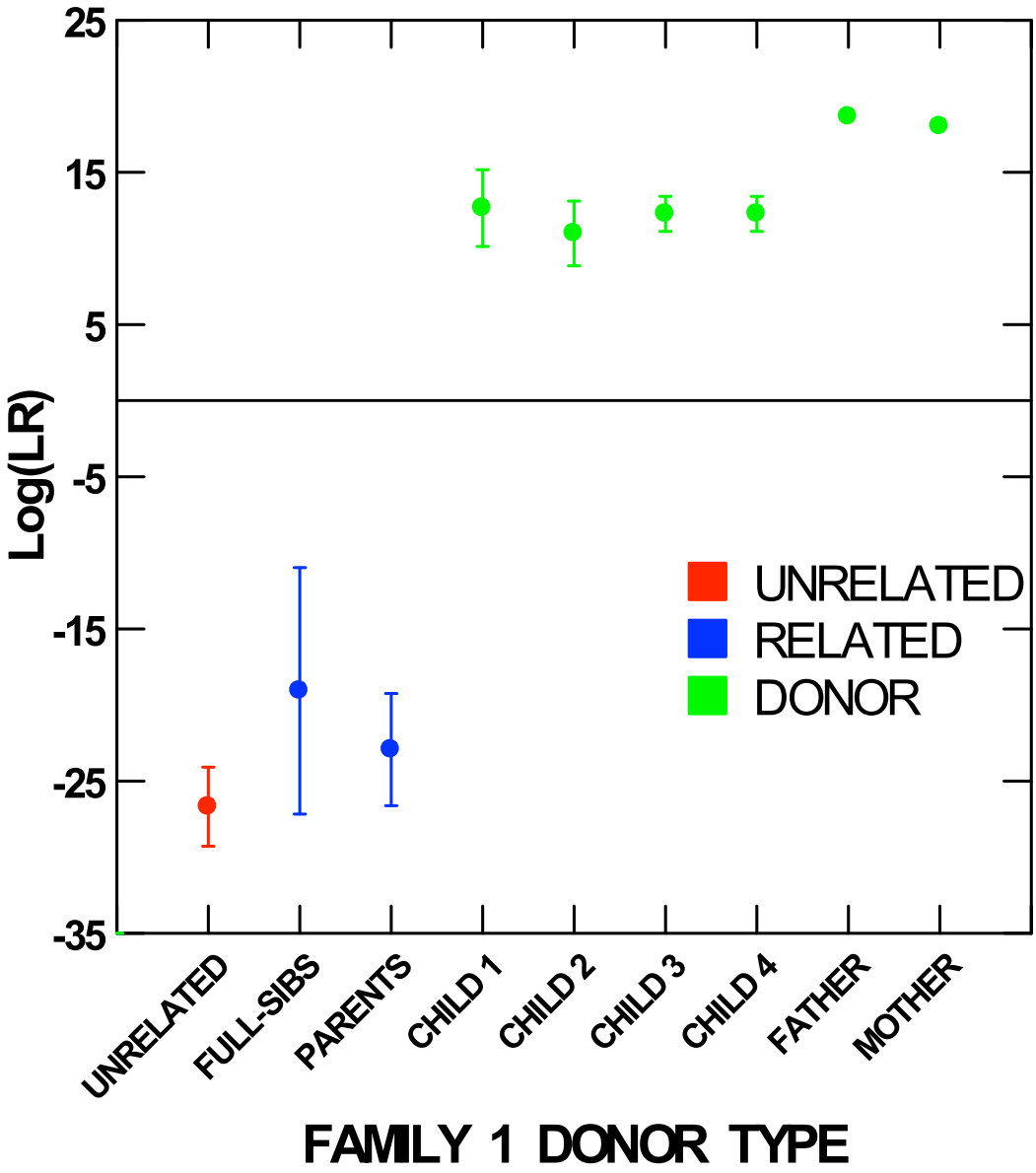


Figure V.C.2. Dot plot showing mean specificity for all F2 parent-child one-unknown pairwise comparisons as a function of donor type at mixture ratios of 1:1. All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

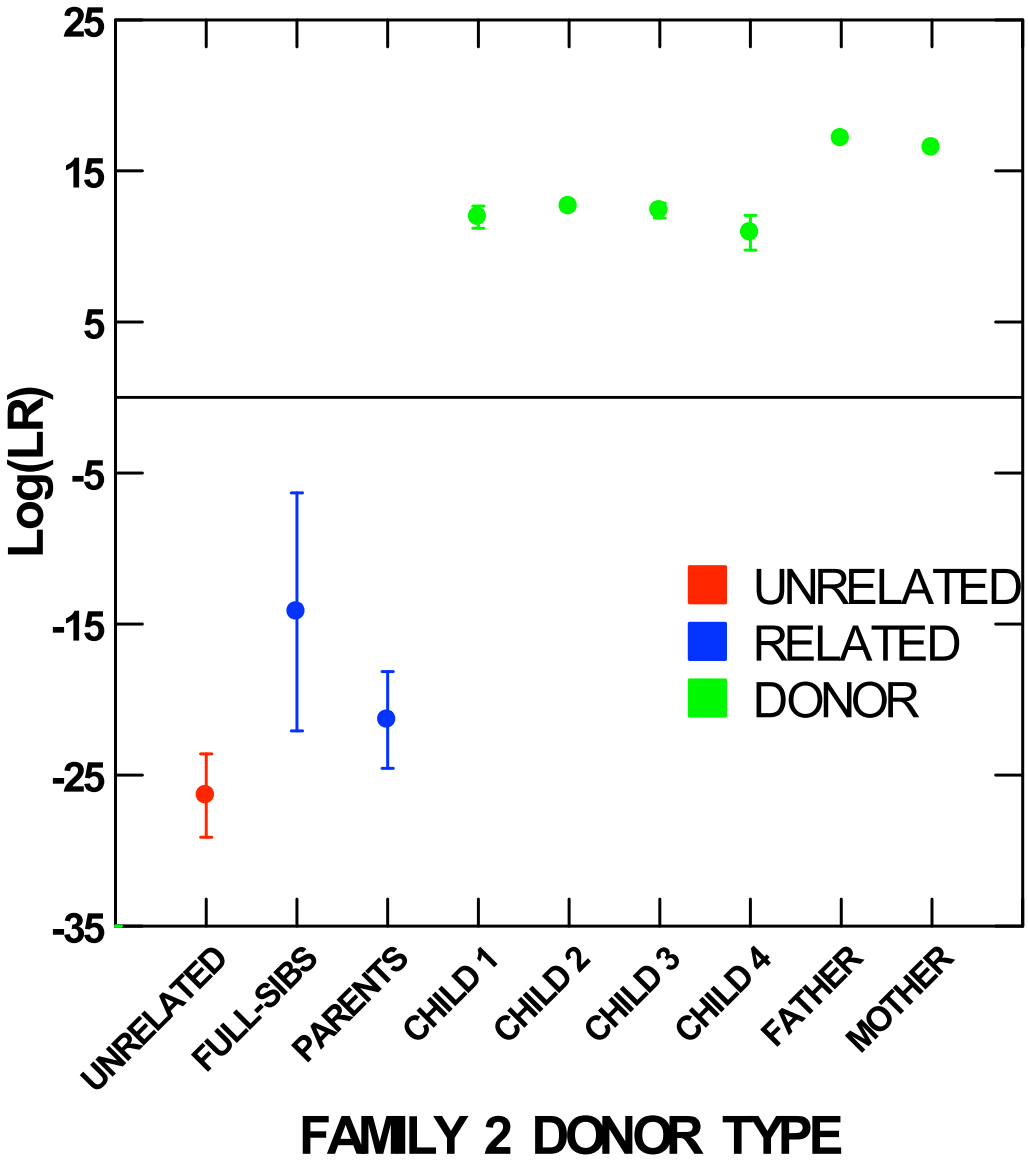


Figure V.C.3. Dot plot showing mean specificity for all F1 parent-child one unknown pairwise comparisons at mixture ratios of 1:1 by individual reference donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4, FATHER = P1, MOTHER = P2). All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

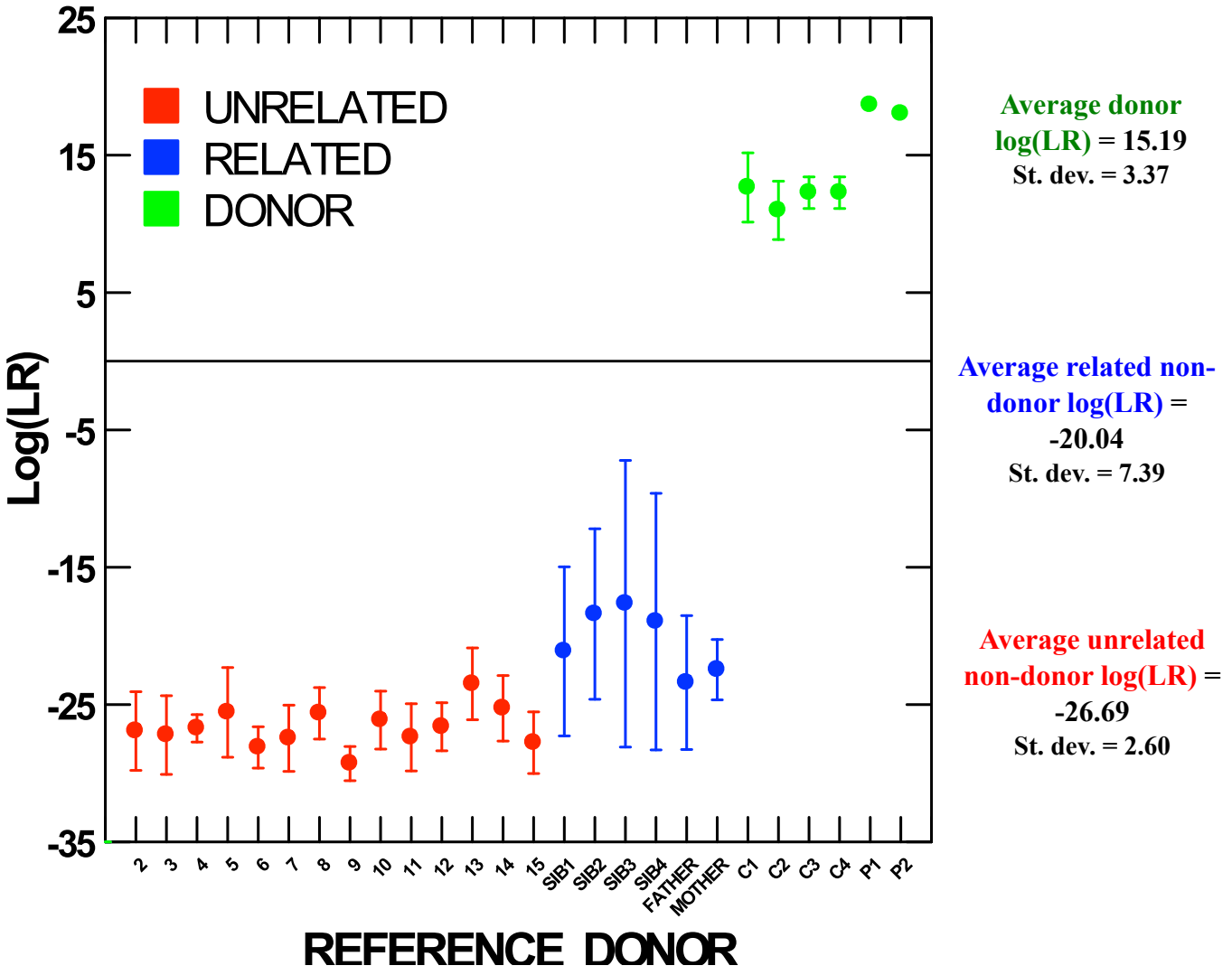
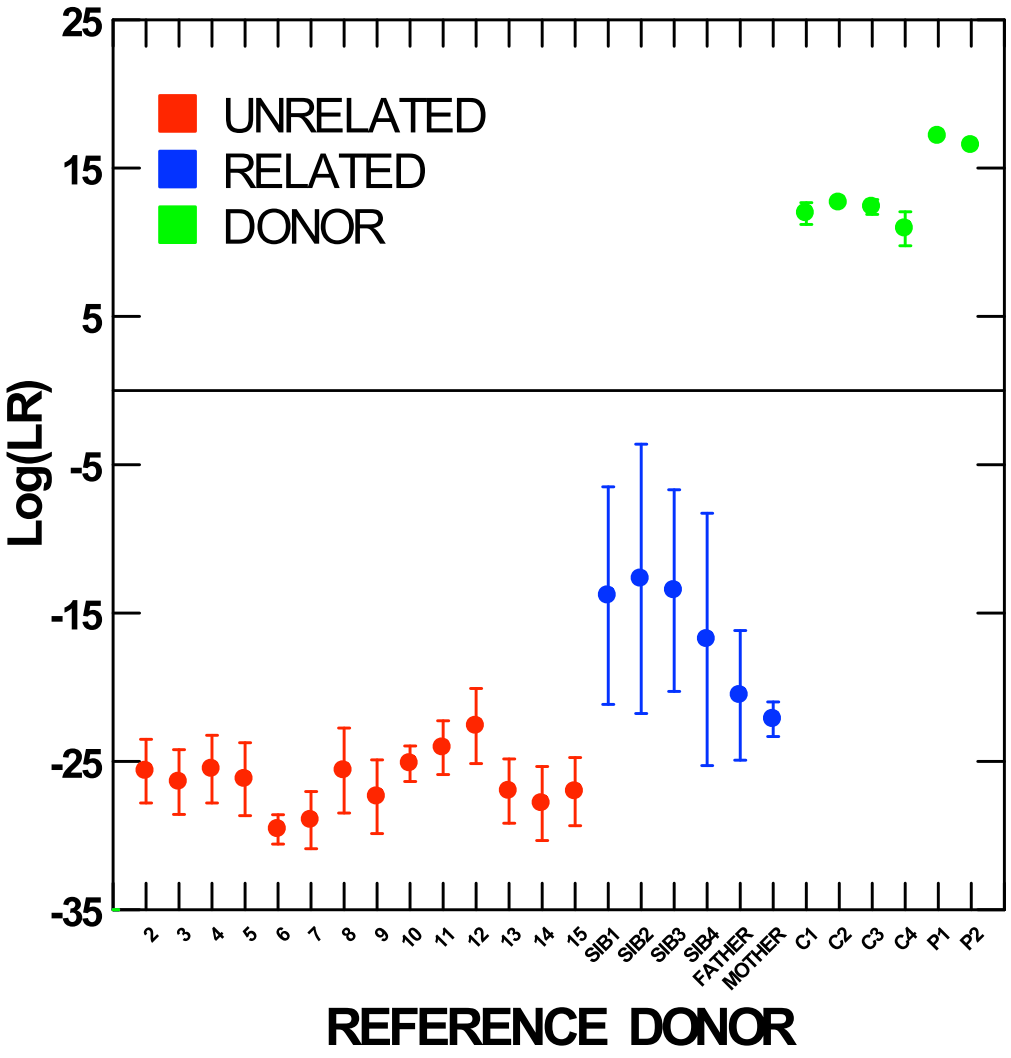


Figure V.C.4. Dot plot showing mean specificity for all F2 parent-child one unknown pairwise comparisons at mixture ratios of 1:1 by individual reference donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4, FATHER = P1, MOTHER = P2). All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.



Average donor log(LR) = 14.40
St. dev. = 2.56

Average related non-donor log(LR) = -15.98
St. dev. = 7.62

Average unrelated non-donor log(LR) = -26.36
St. dev. = 2.76

Figure V.C.5. Dot plot showing mean specificity for all F1 parent-child one-unknown pairwise comparisons as a function of donor type at mixture ratios of 3:1. All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

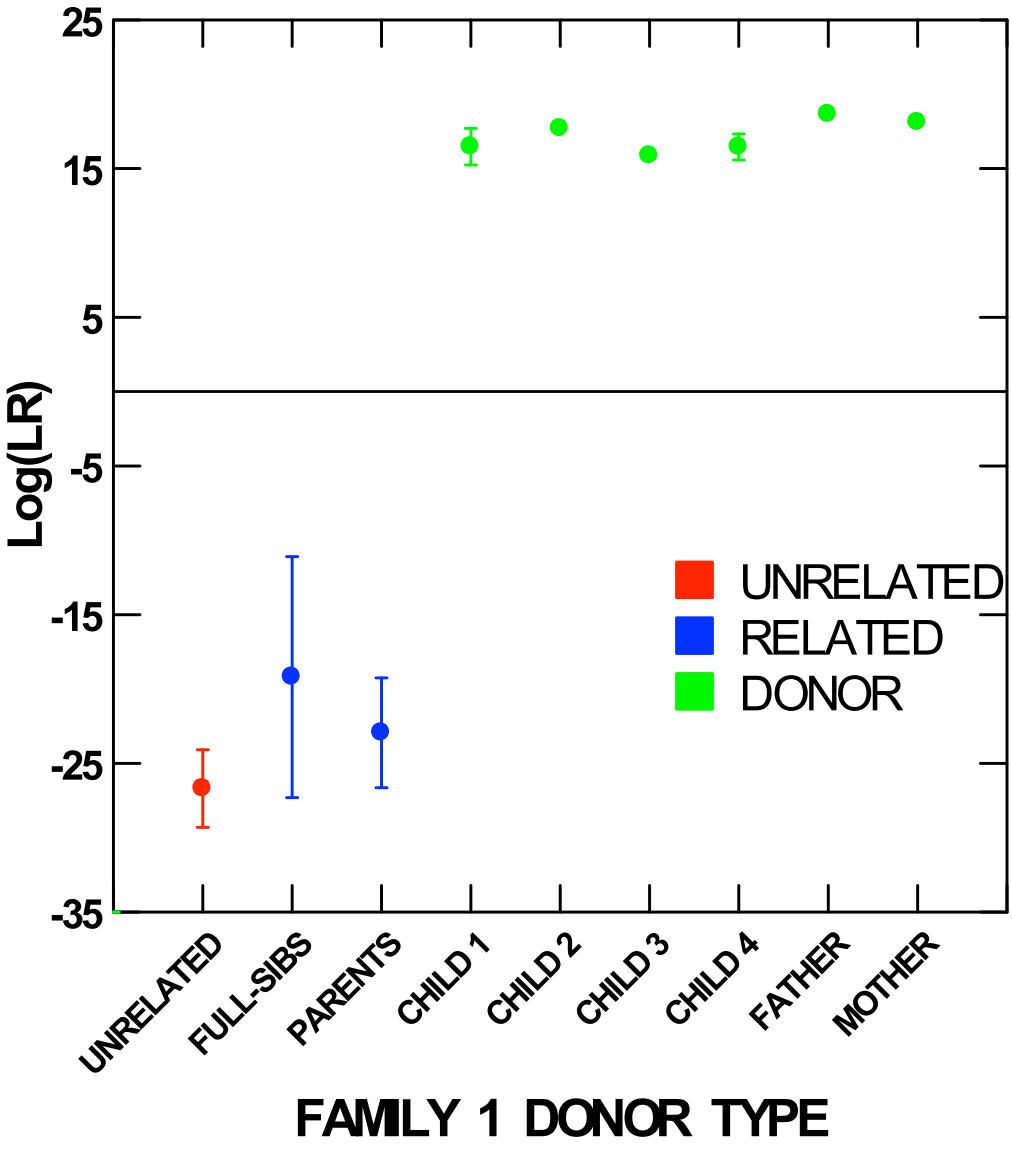


Figure V.C.6. Dot plot showing mean specificity for all F2 parent-child one-unknown pairwise comparisons as a function of donor type at mixture ratios of 3:1. All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

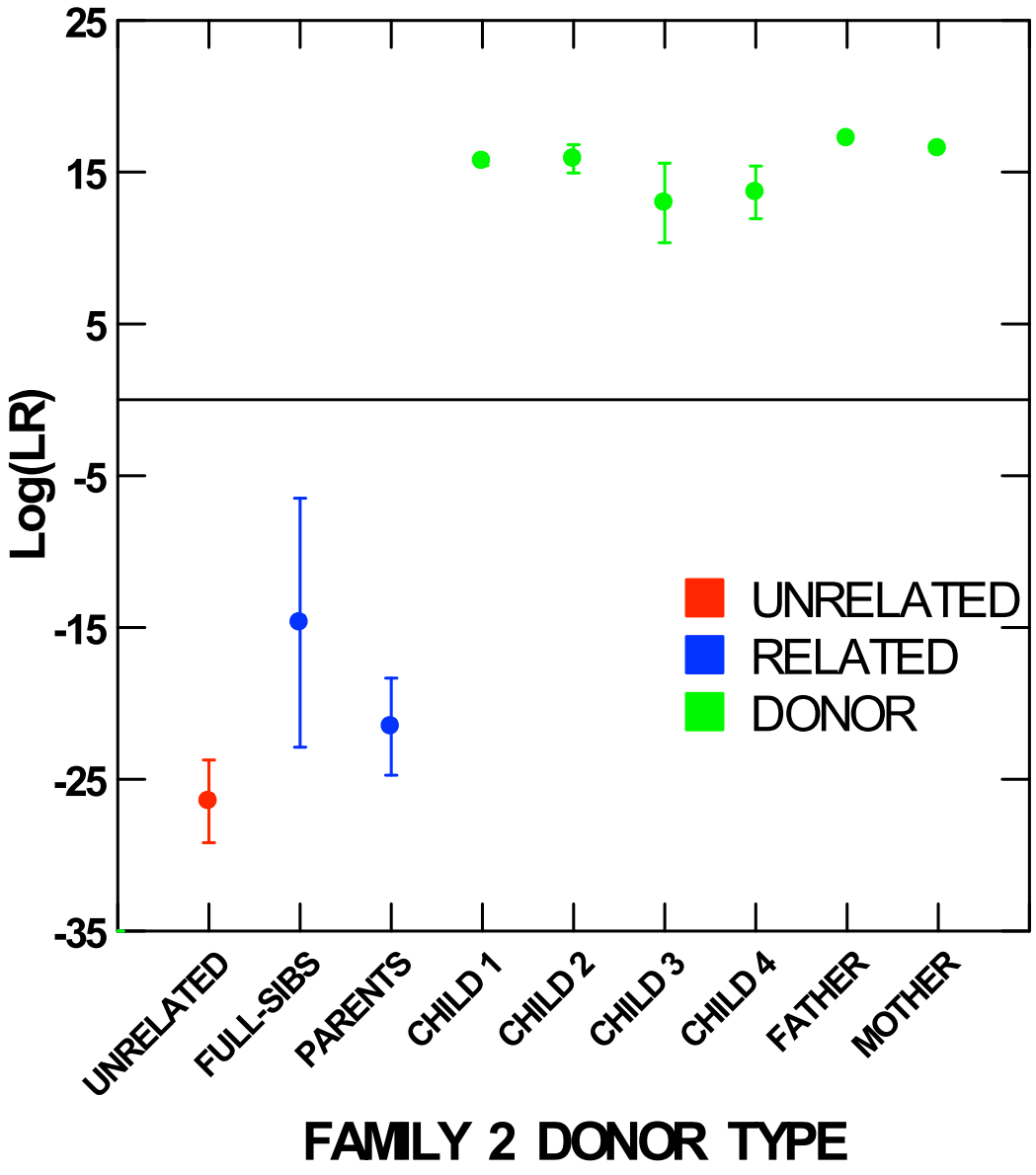
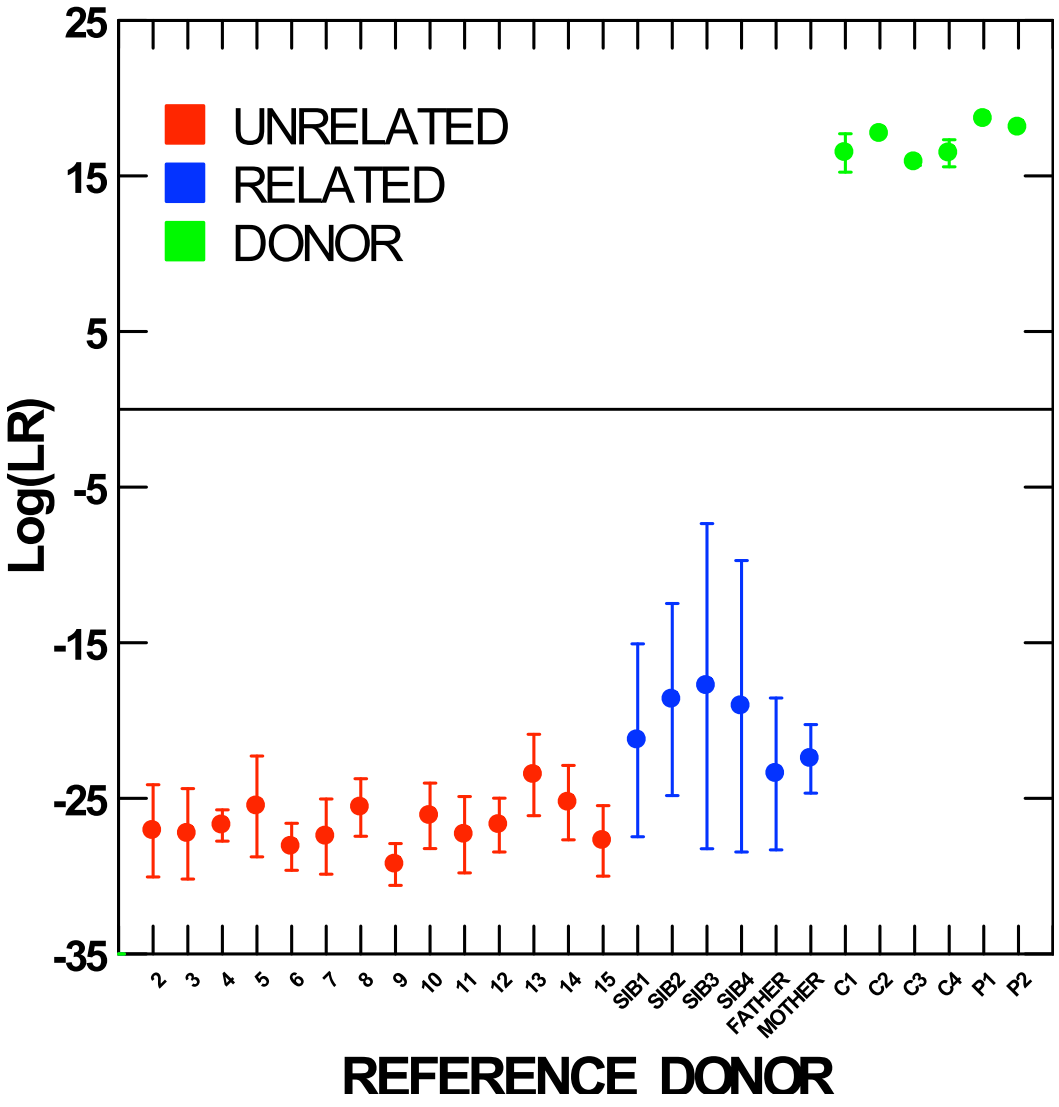


Figure V.C.7. Dot plot showing mean specificity for all F1 parent-child one unknown pairwise comparisons at mixture ratios of 3:1 by individual reference donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4, FATHER = P1, MOTHER = P2). All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.



Average donor log(LR) = 17.50
St. dev. = 1.11

Average related non-donor log(LR) = -20.14
St. dev. = 7.38

Average unrelated non-donor log(LR) = -26.70
St. dev. = 2.61

Figure V.C.8. Dot plot showing mean specificity for all F2 parent-child one unknown pairwise comparisons at mixture ratios of 3:1 by individual reference donor (SIB1 = C1, SIB2 = C2, SIB3 = C3, SIB4 = C4, FATHER = P1, MOTHER = P2). All requests were solved in duplicate (n = 16). Error bars represent one standard deviation; dashed line set at zero.

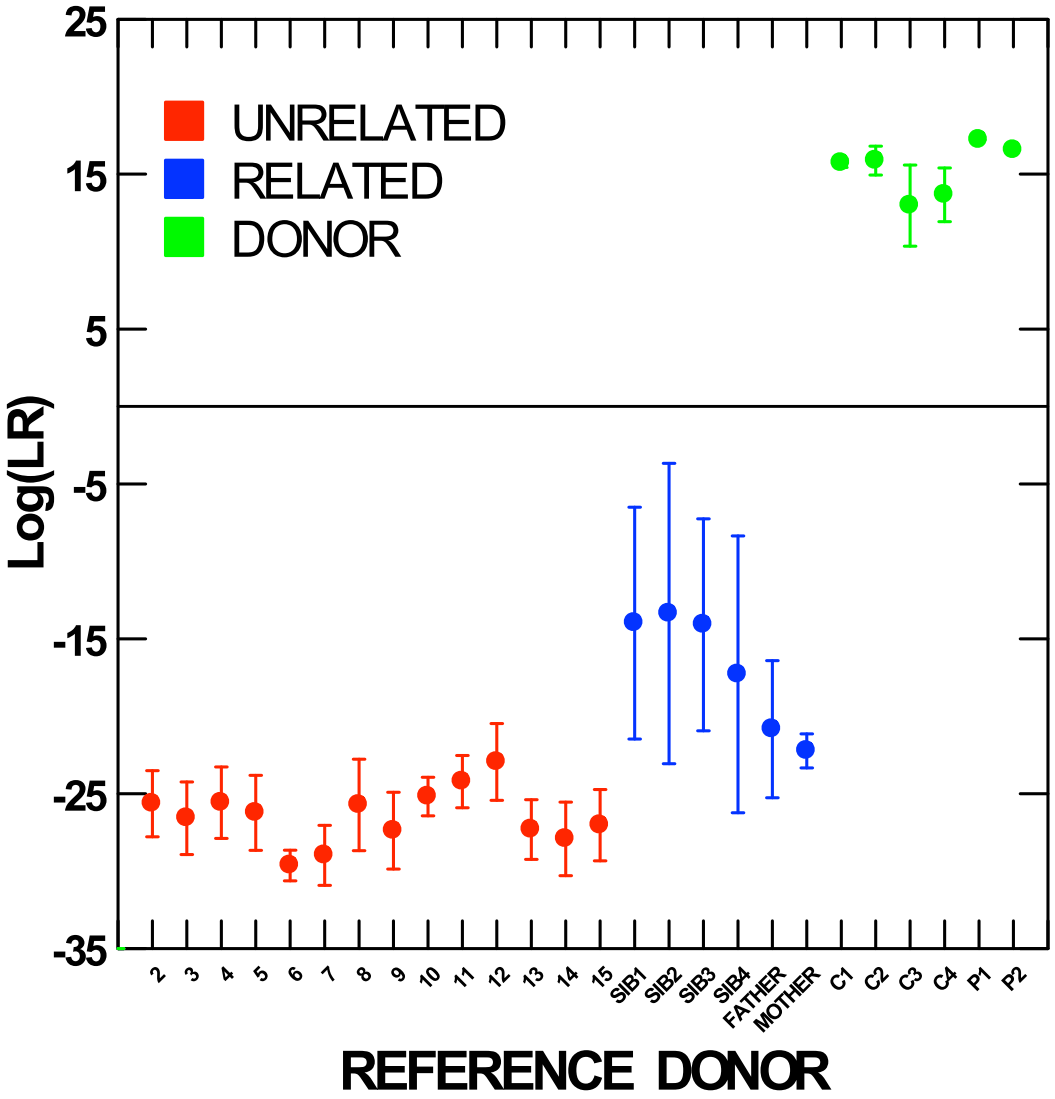


Figure V.C.9. Boxplot showing median log(LR) values for parent-child one unknown comparisons as a function of donor type and grouped by mixture ratio. Both family datasets are combined. Dashed line set at zero; asterisks and circles denote outliers.

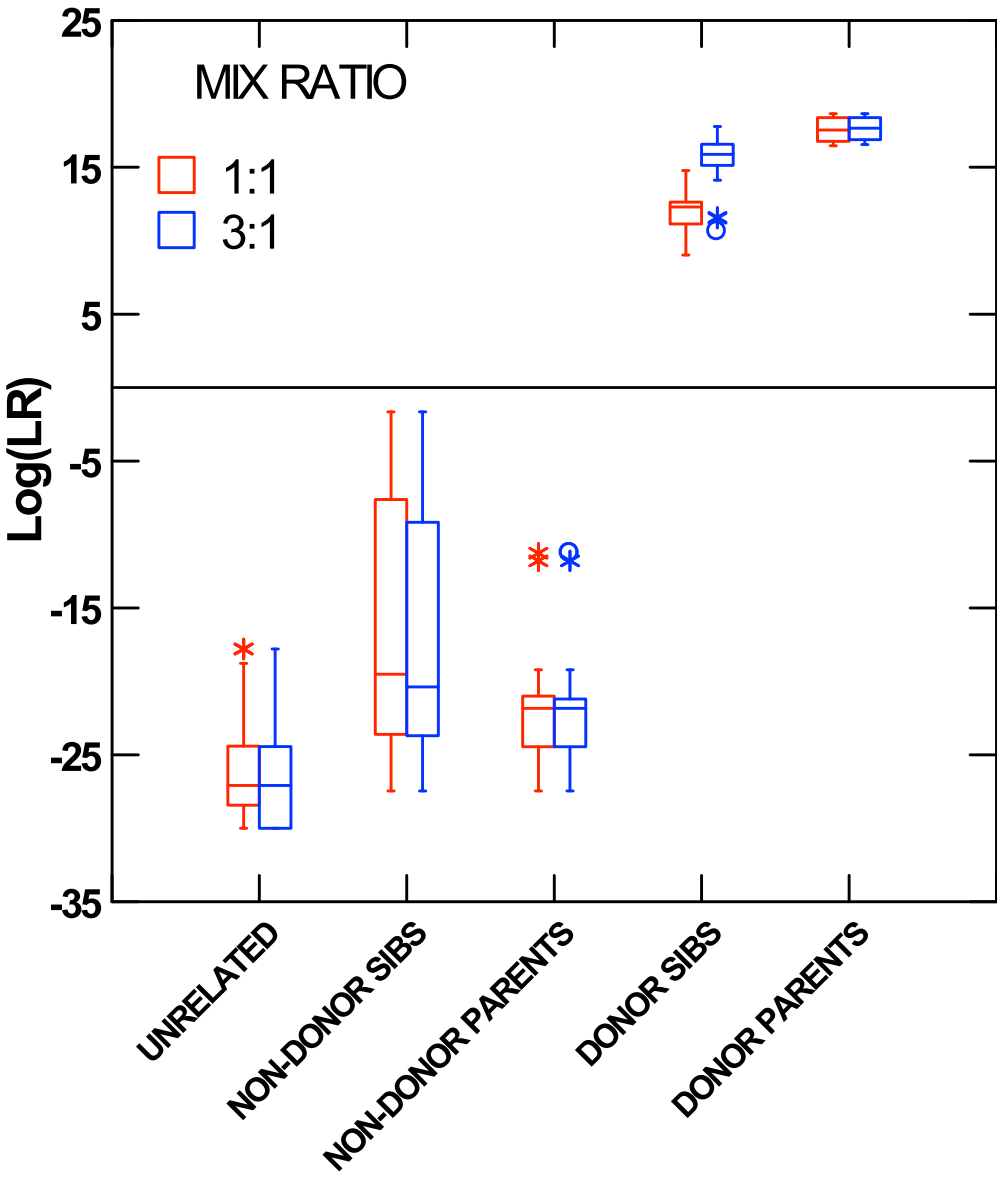


Figure V.C.10. Dot plot showing mean experiment-wide log(LR) values from all parent-child one unknown comparisons as a function of donor type. F1 and F2 datasets and both mixture ratios are combined. Error bars represent one standard deviation; dashed line set at zero.

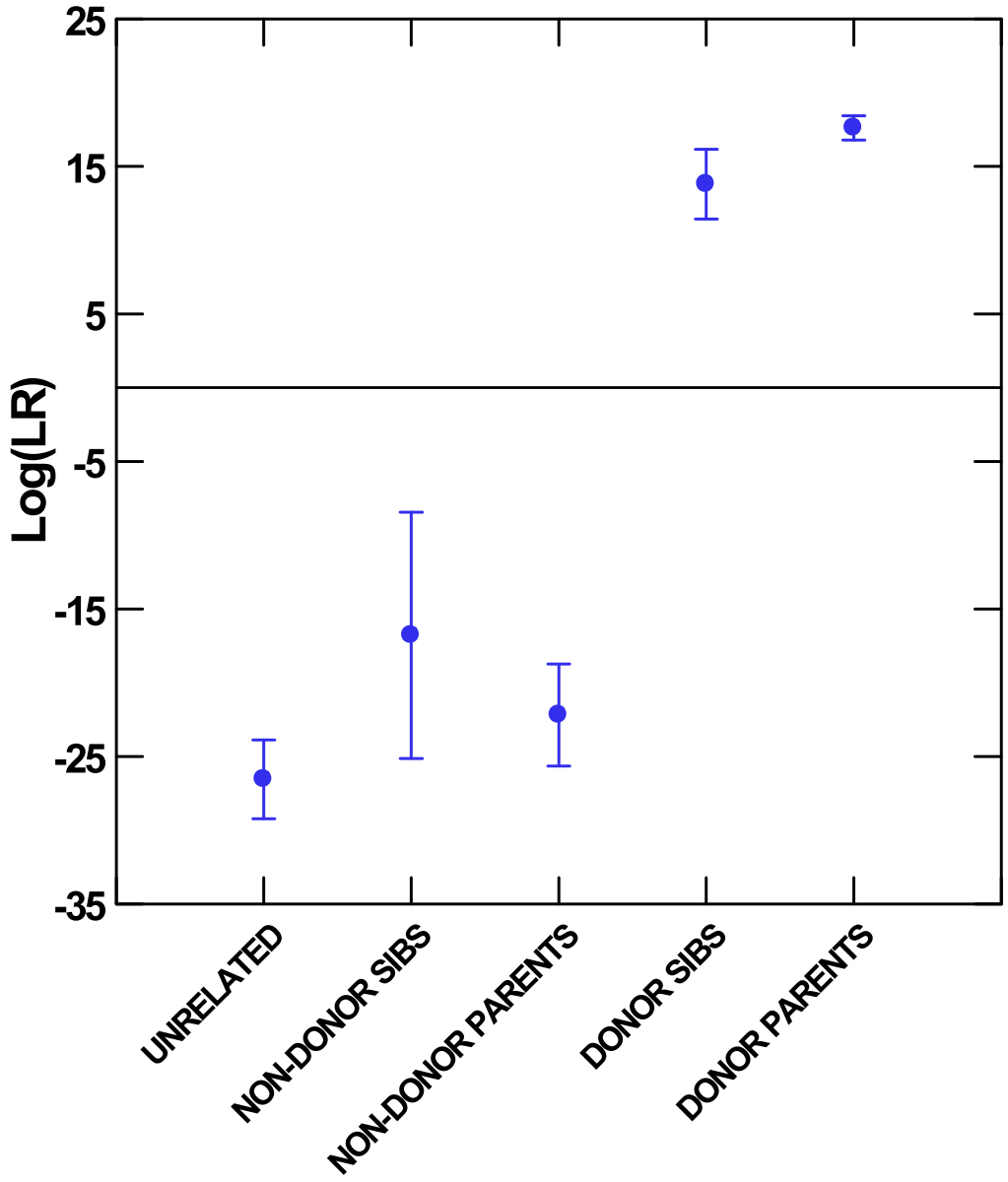
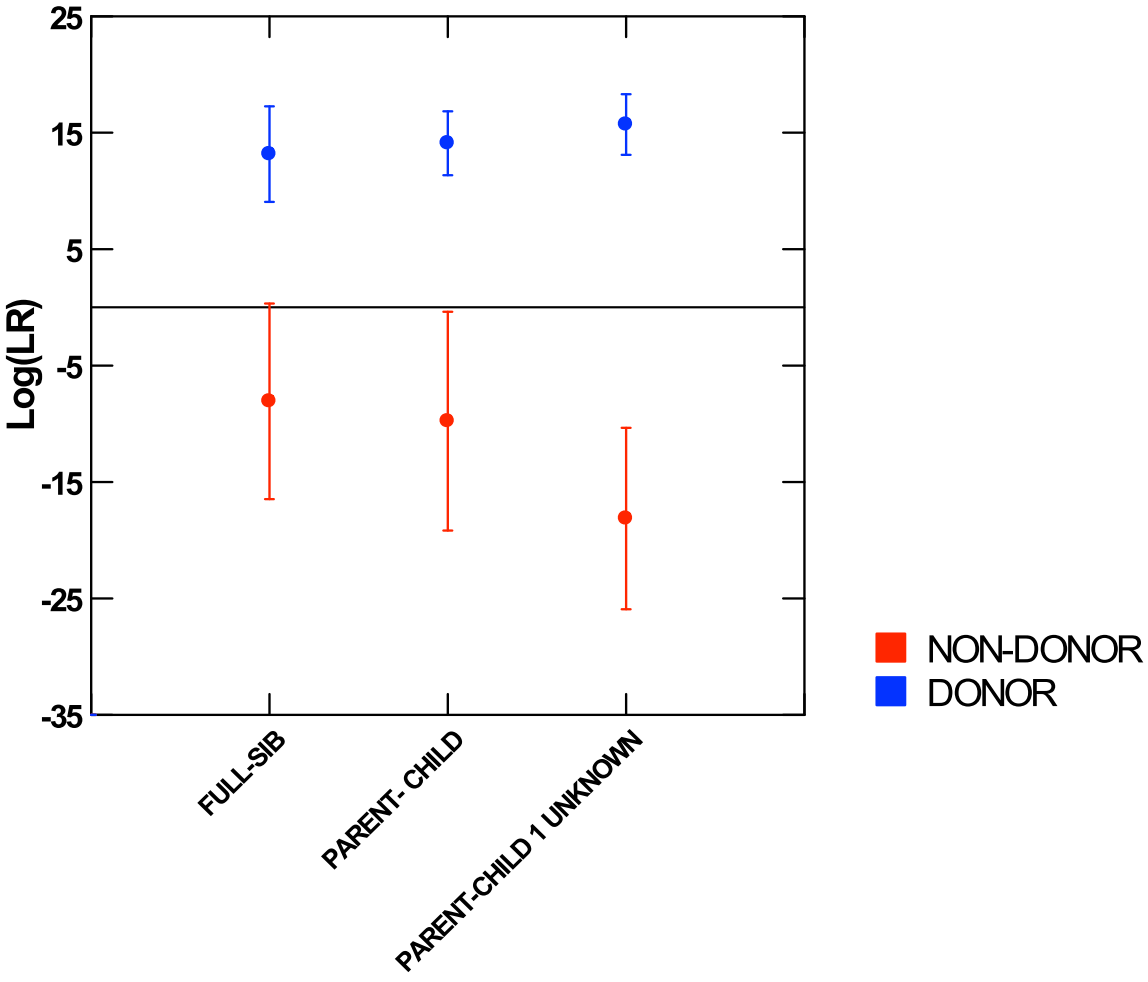


Figure V.C.11. Experiment-wide specificity between related donors and related non-donors (data from unrelated individuals not shown). Mean values from all three comparison types are displayed. Dashed line set at zero; error bars represent one standard deviation. Both family datasets and separate mixture ratios are combined within comparison types.



D. Reproducibility: Log(LR), Mixture Weights, and KL

The reproducibility of TrueAllele Casework was assessed by running duplicate identification requests for all two person familial mixtures. Log(LR) match statistics for full-sib known donors were reproducible with no significant differences between replicate genotype inferences detected (Kruskal Wallis $p = 0.866$) (Figure V.D.1). Similar results were found for both the parent-child comparisons ($p = 0.786$) and parent-child one unknown comparisons ($p = 0.877$) (Figures V.D.2 and V.D.3). The largest variations in replicate log(LR) values were observed in the full-sib comparisons at 3:1 mixture ratios (Table V.D.1). While mean 3:1 match statistics are increased compared to the 1:1 full-sib data, the large amount of allele sharing drove overestimations in mixture weight divergence with a resulting underestimation and increased variance in log(LR) values for some minor donors.

With respect to mixture weight inference, standard deviations for the 1:1 mixtures were significantly higher than for 3:1 mixtures (Kruskal Wallis $p < 0.001$) (Figure V.D.4), although in all cases values were typically in the normal range expected for two person mixtures based on previous validation work. Actual inferred mixture weights were remarkably close to the theoretical mixture weights expected for all comparison types in both mixture ratios (Table V.D.2). Lastly, KL values (divergence of inferred profiles from population priors) were not significantly different between replicates (Kruskal Wallis $p = 0.990$) (Figure V.D.5), indicating TrueAllele was able to reproducibly infer genotypes with approximately equal information content.

Table V.D.1. Mean donor log(LR) match statistics and associated within-donor standard deviations for all two person familial mixture comparisons. F1 and F2 family datasets are pooled within comparison types.

Comparison Type	Mean Donor Log(LR)	Mean Within-Donor Standard Deviations
Full-sibs 1:1	**11.40**	**1.00**
Full-sibs 3:1	**14.89**	**2.78**
Parent-child 1:1	**11.76**	**0.93**

Parent-child 3:1	**16.43**	**0.71**
Parent-child (one unknown) 1:1	**14.80**	**0.67**
Parent-child (one unknown) 3:1	**16.61**	**0.56**

Figure V.D.1. Bargraph of mean full-sub log(LR) match statistics by donor and grouped by replicate. Family F1 contains children 1-4, family F2 contains children 5-8. Both mixture ratios are combined; error bars represent one standard deviation.

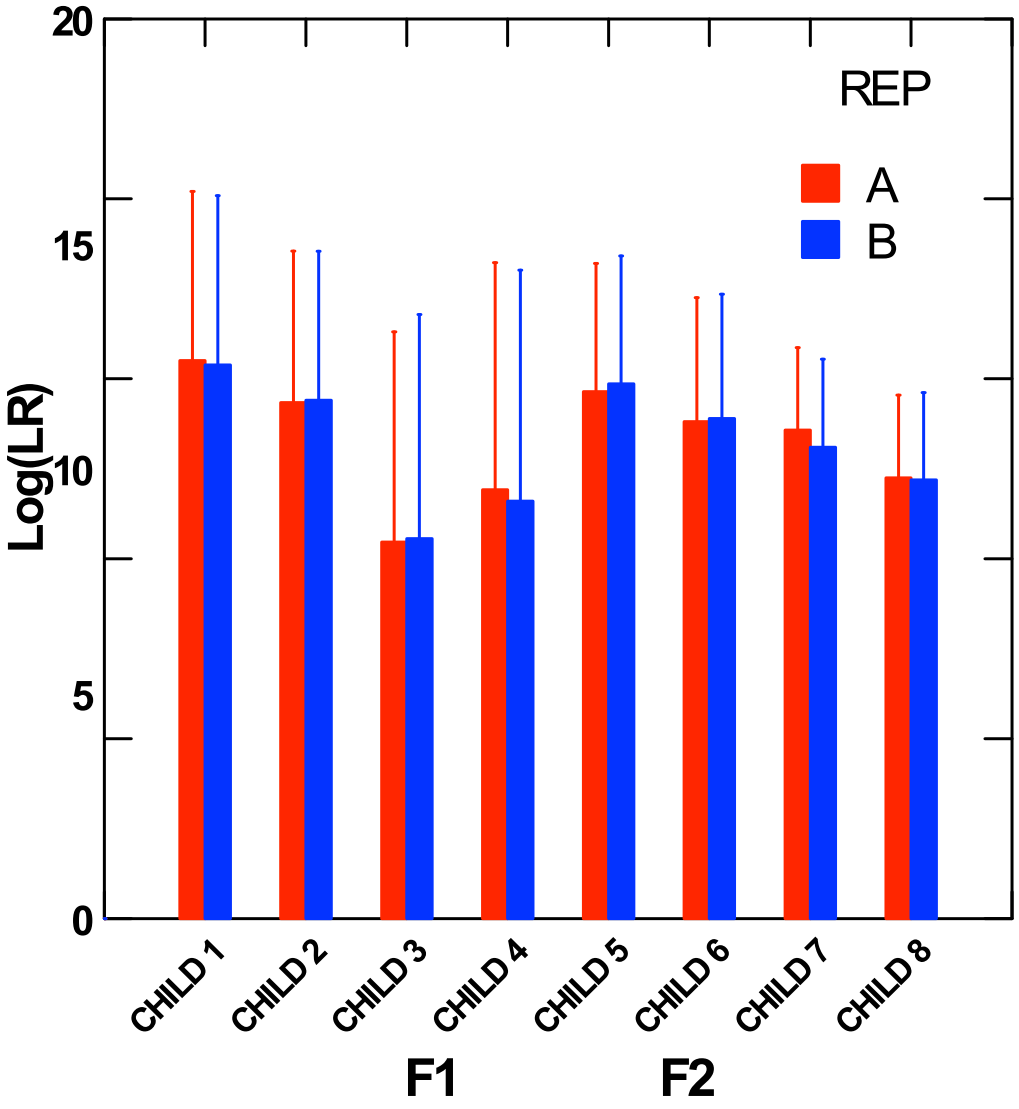


Figure V.D.2. Bargraph of mean parent-child log(LR) match statistics by donor and grouped by replicate. Family F1 contains children 1-4, family F2 contains children 5-8. Both mixture ratios are combined; error bars represent one standard deviation.

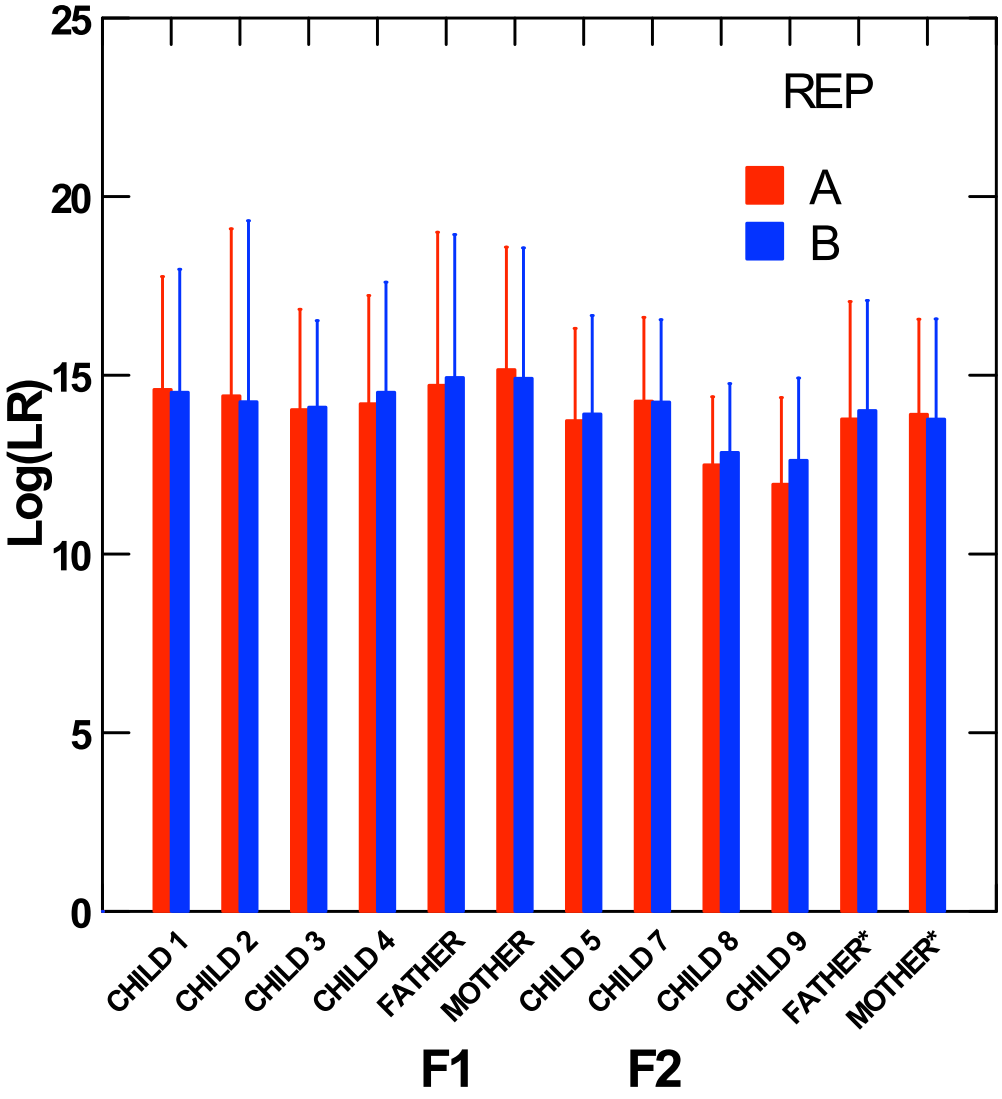

25
20
15
10
5
0
Log(LR)
REP
A
B
CHILD 1
CHILD 2
CHILD 3
CHILD 4
FATHER
MOTHER
CHILD 5
CHILD 7
CHILD 8
CHILD 9
FATHER*
MOTHER*
F1
F2

Figure V.D.3. Bargraph of mean parent-child one unknown log(LR) match statistics by donor and grouped by replicate. Family F1 contains children 1-4, family F2 contains children 5-8. Both mixture ratios are combined; error bars represent one standard deviation.

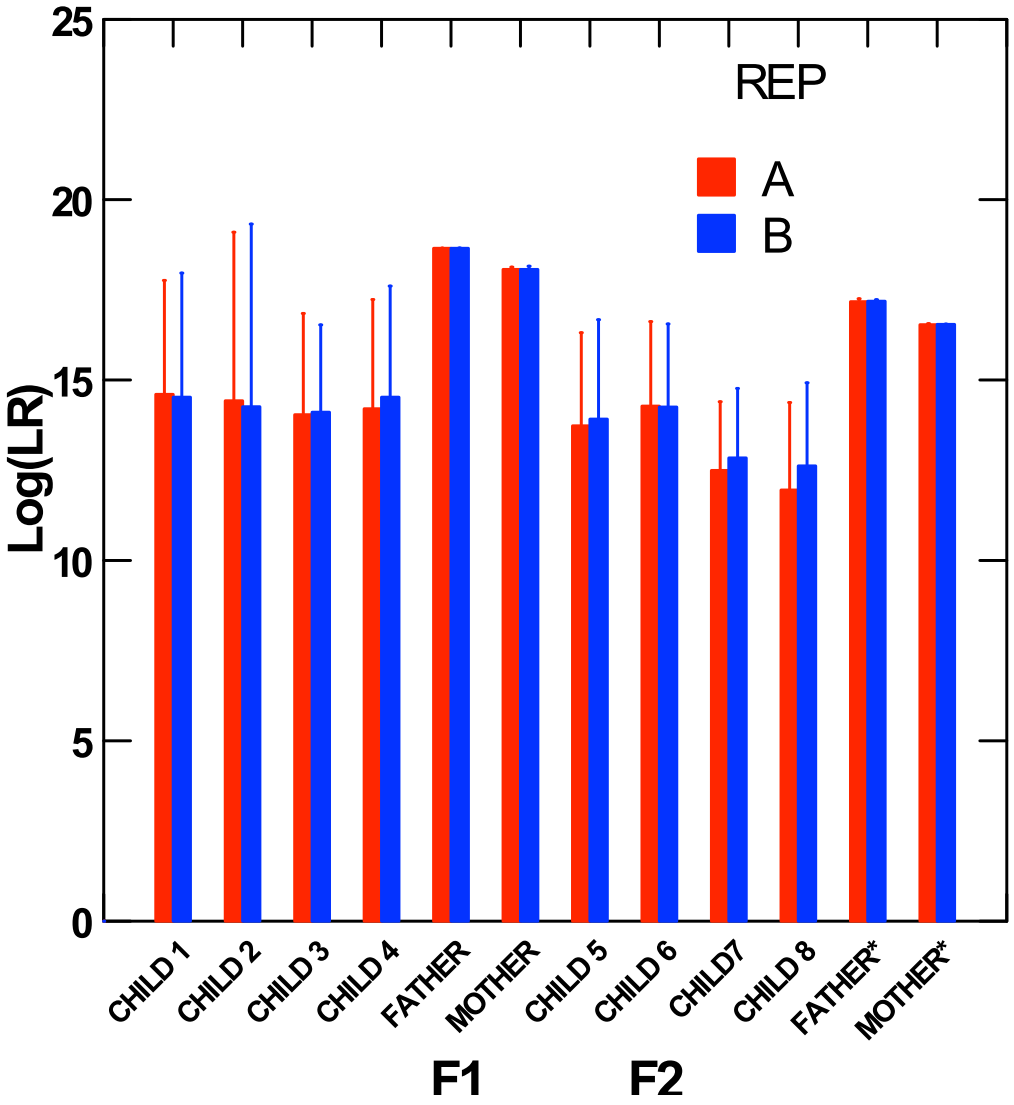


Figure V.D.4. Bargraph of two person mixture weight standard deviations as a function of familial mixture type and grouped by mixture ratio. Mean values for both families (F1 and F2) are shown; error bars represent one standard deviation.

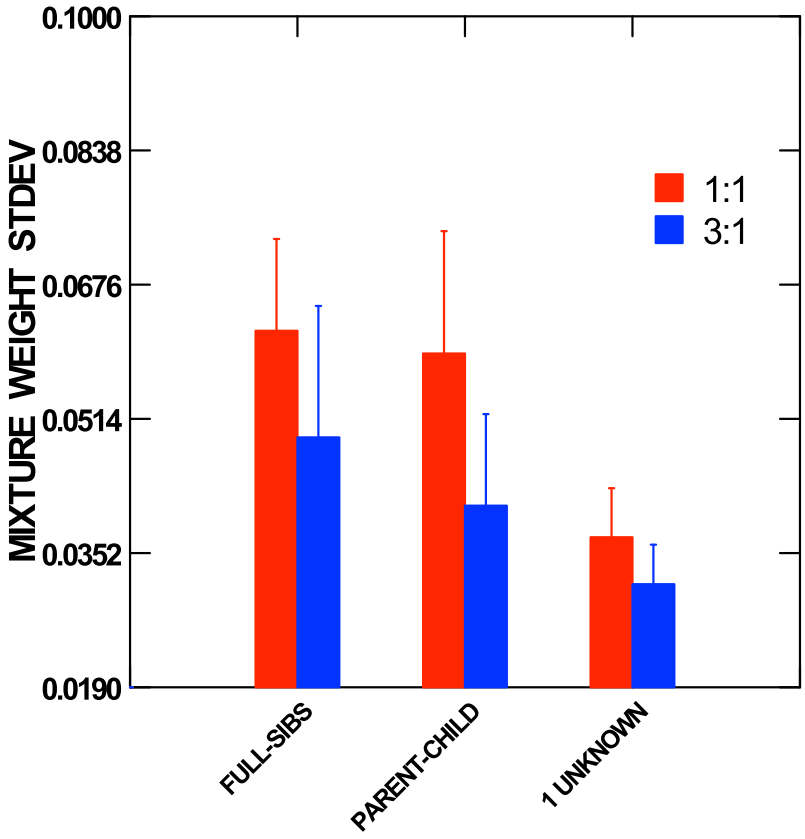
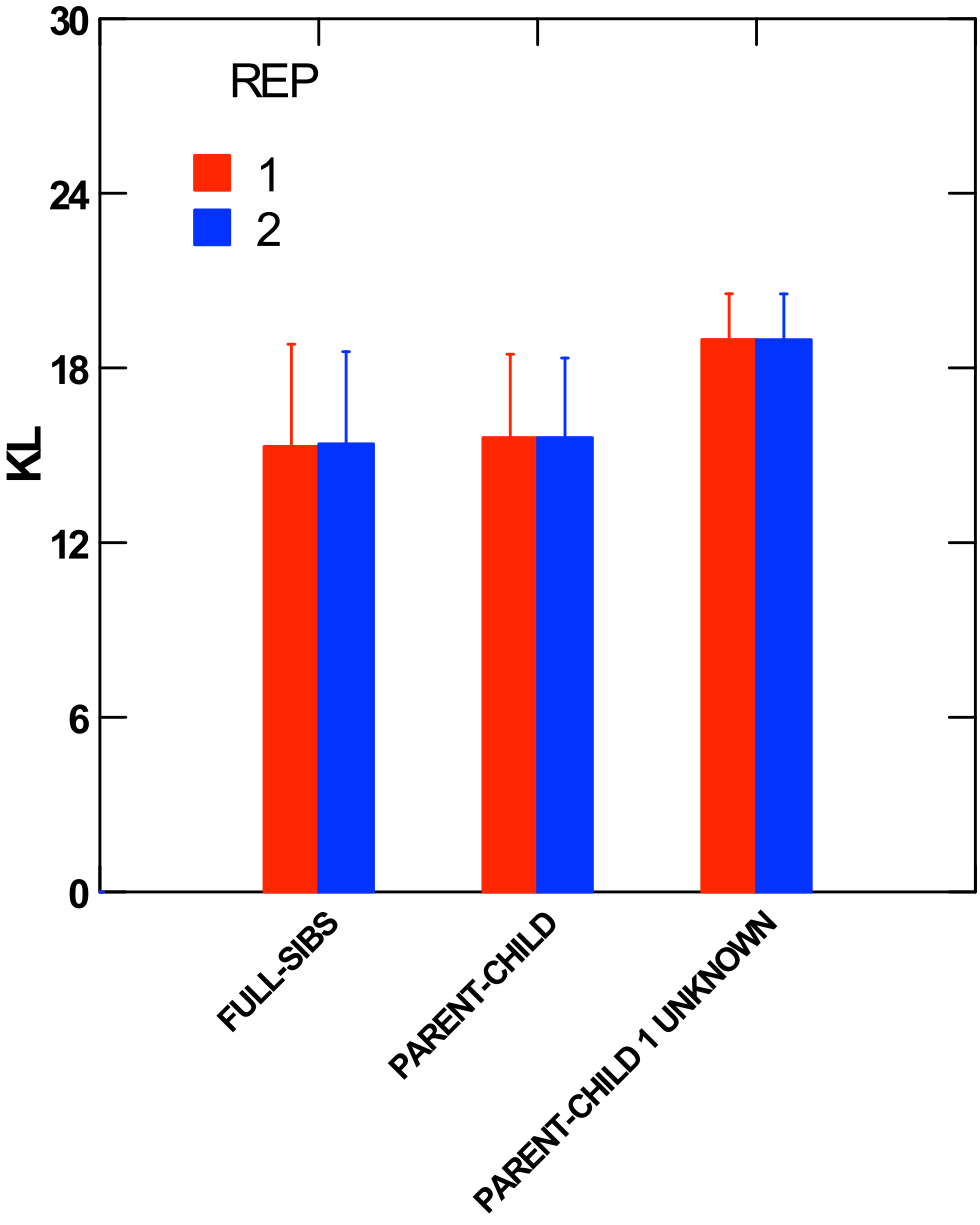


Table V.D.2. Comparison of theoretical two person familial mixture weights to those inferred by TrueAllele Casework; mixture weights shown are averaged across both families (F1 and F2).

Comparison Type	Theoretical Mixture Weights	TrueAllele Inferred Mixture Weights	Mean Standard Deviation
Full-sibs 1:1	**0.50/0.50**	**0.501/0.499**	**0.064**
Full-sibs 3:1	**0.75/0.25**	**0.797/0.203**	**0.051**
Parent-child 1:1	**0.50/0.50**	**0.503/0.497**	**0.059**
Parent-child 3:1	**0.75/0.25**	**0.757/0.243**	**0.041**
Parent-child (one unknown) 1:1	**0.50/0.50**	**0.511/0.489**	**0.037**

Parent-child (one unknown) 3:1	**0.75/0.25**	**0.754/0.246**	**0.031**

Figure V.D.5. Bargraph of mean KL values for all comparison types by replicate. Both family datasets and mixture ratios are combined; error bars represent one standard deviation.



E. MCMC Convergence

Convergence was assessed with Gelman Rubin statistics (GR hereafter) and by visual inspection of the mixture weight Markov chains and histograms. For all two person familial mixtures tested, the mean GR was 1.05 with a standard deviation of 0.096. On visual inspection, mixture weight histograms were generally in accord with the known mixture weights tested, and Markov chains typically reached stationary distributions early in the read out phase (data not shown; see previous TrueAllele validation work on two person mixtures for representative Markov chains and mixture weight histograms). While significant differences in GR statistics were detected between mixture ratios (Kruskal Wallis $p = 0.016$) and comparison types ($p < 0.001$), mean values were just slightly over 1.00, indicating acceptable convergence (Figure V.E.2). In rare instances, full-sib mixtures at the 3:1 ratio yielded mixture weights in excess of the expected 0.75/0.25 ratio (Figure V.E.1). In these cases, it appears the software over-estimated the divergence between contributors due to excessive allelic overlap($\geq$ 20 alleles). Subsequent re-runs were able to successfully infer the actual mixture weights and the phenomenon was not seen with equally weighted mixtures for any comparison type.

Figure V.E.1. Mixture weight histogram and associated Markov chain for an approximate 3:1 (0.75/0.25) mixture with 21 shared alleles between full-siblings. Inferred mixture weight was 0.938/0.062.

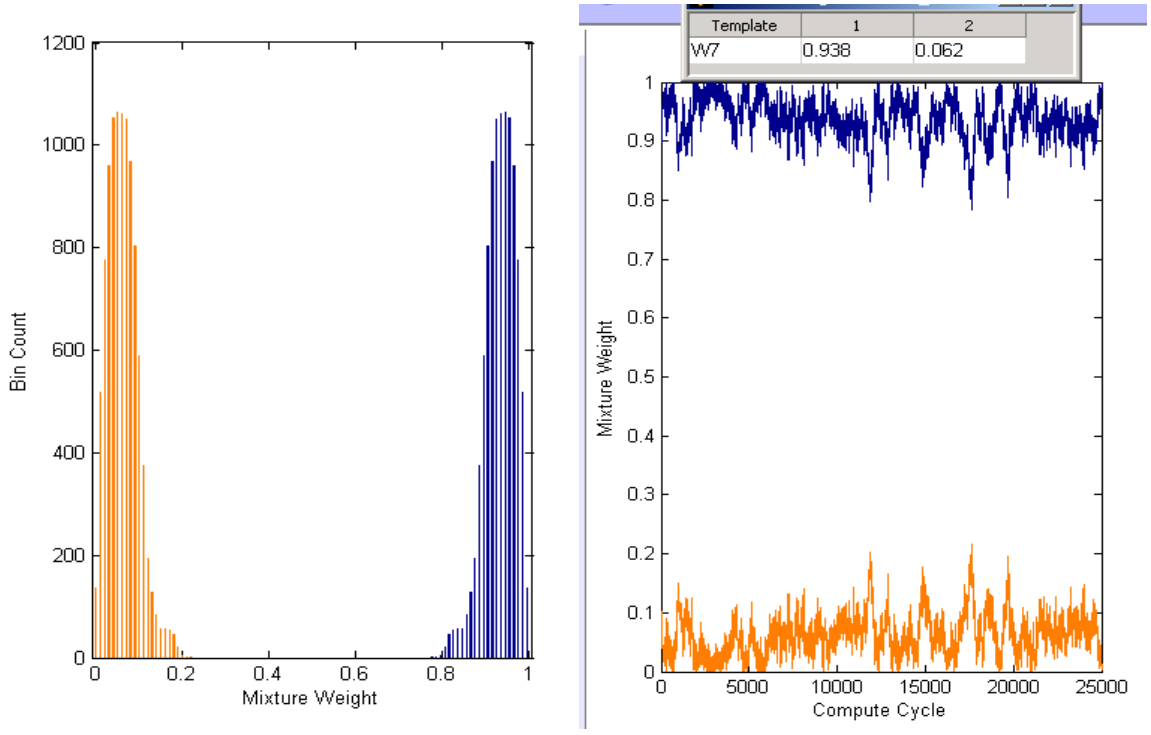
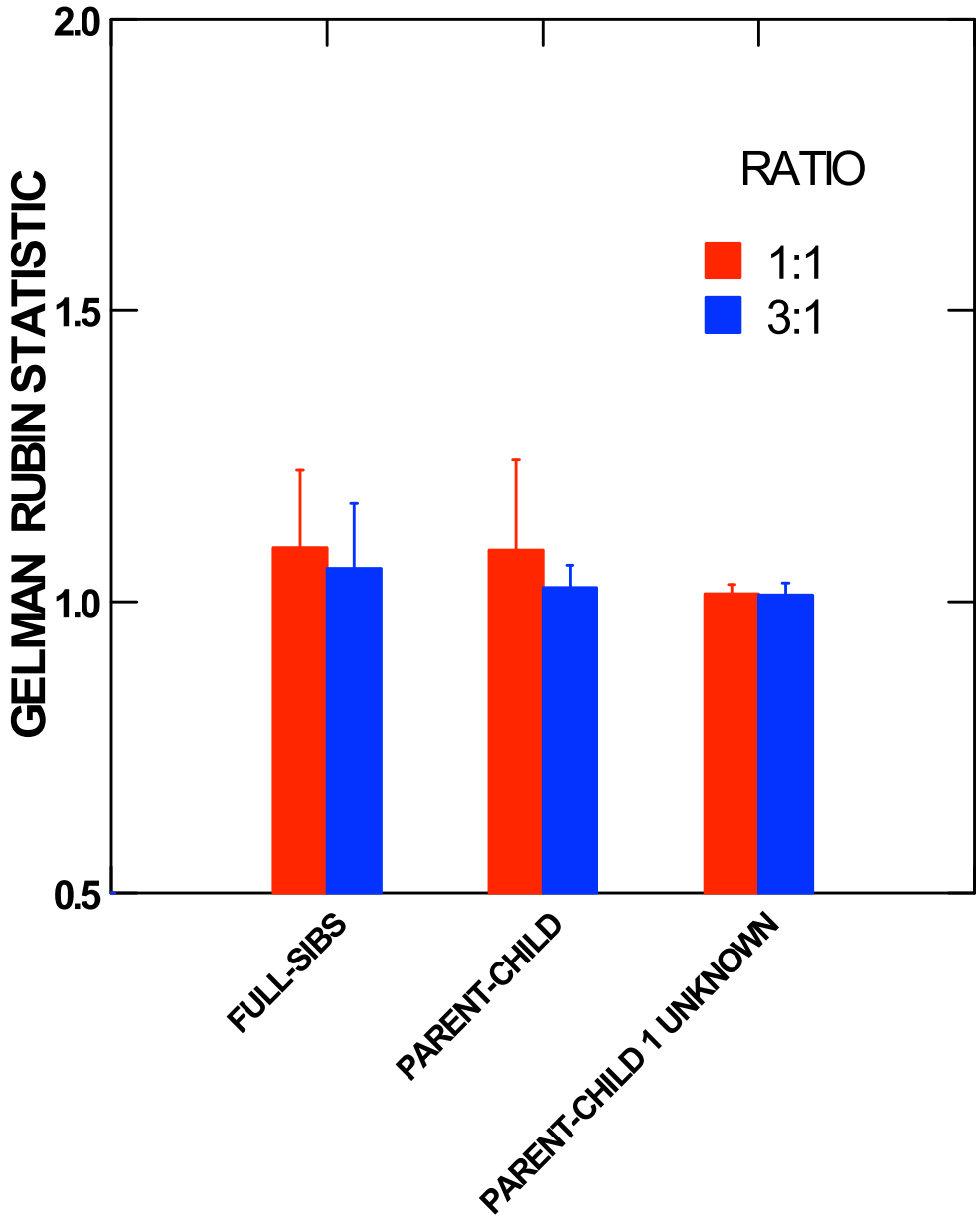


Figure V.E.2. Bargraph of mean Gelman Rubin convergence scores for each familial mixture comparison type grouped by mixture ratio. Error bars represent one standard deviation.



VI. Conclusions

Similar to previous work with up to three person mixtures, results from the four person mixture component of the validation suggest that TrueAllele Casework is still capable of providing robust, donor-specific match statistics for mixtures of up to four contributors. While such complex mixtures are currently deemed inconclusive or insufficient for interpretation, the use of probabilistic genotyping software further expands the scope of suitable evidentiary samples and enables FIC scientists to give an unbiased, full accounting of all genetic data recovered.

Results show that log(LR) match statistics for all four person mixtures tested were specific to known donors while all non-donors overwhelmingly yielded negative log(LR) values. The overall Type I error rate for all two, three, and four person mixtures was 0.00024; while several false positive matches were observed in the extended comparison data set, all were less than a log(LR) of 2. Inferred genotypes were also reproducible with respect to both information content (KL statistics) and log(LR) values. However, analysts may expect reduced match statistics, reduced specificity, and attenuation of mixture weight inference due to the increased complexity inherent to four contributor mixtures. As noted in previous TrueAllele validation work, the data support a thorough assessment of all convergence diagnostics (Gelman-Rubin statistics and Markov chain histogram and history graphs) before reporting to ensure the integrity of mixture weight inference, regardless of donor number or complexity.

The familial mixture study, the first of its kind using a fully continuous probabilistic approach, further demonstrates both the efficacy and specificity of TrueAllele Casework. While cross-matching between full-siblings and parents was observed, the highest match statistics were always associated with the true donors in all mixtures tested. Individual sibling and parent specificity was found to be a function of both the magnitude of allele sharing and mixture weight separation, with a high degree of allelic overlap and roughly equal mixture weights yielding the lowest quantitative separation. Mean log(LR) separation between donors and related non-donors was shown to improve dramatically when mixture weights were unequal and when using assumed references as known contributors. For maximal information gain and specificity, the Bayesian strategy of reducing uncertainty by using assumed known contributors should be employed where case specifics allow.

The validation findings herein support and further extend the data from multiple peer-reviewed Journal of Forensic Sciences publications from the New York State Police Forensic Investigation Center in collaboration with Cybergenetics staff, the original TrueAllele Casework validation approved May 2, 2013, and the most recent work with Identifiler Plus chemistry investigating low template samples and mixtures of up to three donors. Based on these data, it is strongly recommended that all current non-quantitative statistical methodologies for autosomal STR interpretation be replaced with the validated probabilistic genotyping approach for mixtures involving both unrelated and related individuals.