Supplementary material

A Comparison with *DNAmixtures*

In this section we will compare the likelihood values between *EuroForMix* and *DNAmixtures* by randomly generating single source stains and two- and three-person mixtures. Note that in order to get the same results as *DNAmixtures* we let alleles with the same whole integer (e.g. allele 30, 30.2, 30.3 become allele 30) become one allele (summing their relative frequencies). Also, no degradation and sub-population structure are assumed ($\beta = 1$, $F_{st} = 0$). We used the Norwegian ESX 17 population frequencies, considering 16 markers, where new observed alleles were assigned as the smallest observed frequency. For a given number of contributors K, three random crime samples were simulated from the continuous model with parameters equal to $\sigma = 0.2$ and $\xi = 0.1$. The detection threshold used in the simulation and models was $T = 150$ RFU.

In the comparison we let *DNAmixtures* be optimized (with the true parameters as start values in the optimizer). The optimized parameters values $\theta = (\pi_1, ..., \pi_K, \mu, \sigma, \xi)$ were later inserted into the likelihood function implemented in *EuroForMix* to check that the same likelihood values were obtained. The comparison tables are shown where the values of θ are represented (5 significant digits) together with the log-likelihood values. Notice here that $\pi_K = 1 - \sum_k^{K-1} \pi_k$.

The sampled evidences were evaluated by conditioning on different types of hypotheses. We let the code xKyU denote that the model conditions on the hypothesis that the x first true references from the sampled stain are considered as known contributors, while y is number of unknown contributors in the same hypothesis. The script for carrying out the comparison can be found at `www.euroformix.com/validation`. The comparison table in Table S1 shows the likelihood values for each of the sampled evidence up to 12 decimals for each possible considered hypothesis. By comparison we found that the log likelihood values from *EuroForMix* and *DNAmixtures* are always identical up to 11 decimals.

Single-source stain (K=1) with $\mu = 200$

	DNAmixtures	*EuroForMix*	μ	σ	ξ
E1(1K0U)	-130.49052633191	-130.49052633191	200.0	0.2	0.1
E1(0K1U)	-171.261610009232	-171.261610009232	200.0	0.2	0.1
E2(1K0U)	-130.452181039142	-130.452181039142	200.0	0.2	0.1
E2(0K1U)	-175.250747463902	-175.250747463902	200.0	0.2	0.1
E3(1K0U)	-130.700911474164	-130.700911474164	200.0	0.2	0.1
E3(0K1U)	-179.675139974559	-179.675139974559	200.0	0.2	0.1

Two-person mixtures (K=2) with $\mu = 2500$ and $m_1 = 0.9$

	DNAmixtures	*EuroForMix*	m1	μ	σ	ξ
E1(2K0U)	-415.511575927926	-415.511575927926	0.90277	2477.35889	0.12543	0.10333
E1(1K1U)	-440.012776443395	-440.012776443395	0.91883	2486.27513	0.13554	0.10988
E1(0K2U)	-485.116377602029	-485.116377602029	0.91883	2486.27230	0.13554	0.10988
E2(2K0U)	-437.996105475833	-437.996105475833	0.87656	2394.87353	0.18024	0.10382
E2(1K1U)	-465.780902545946	-465.780902545946	0.87951	2400.37008	0.17808	0.11074
E2(0K2U)	-518.586686971514	-518.586686971514	0.87951	2400.36966	0.17808	0.11075
E3(2K0U)	-433.643227926185	-433.643227926185	0.91109	2384.77826	0.12932	0.10748
E3(1K1U)	-458.705779978583	-458.705779978582	0.90930	2388.84276	0.13214	0.11029
E3(0K2U)	-511.604121759721	-511.604121759722	0.90931	2388.84072	0.13215	0.11029

Three-person mixtures (C=3) with $\mu = 5000$ and $(m_1, m_2) = (0.6, 0.3)$

	DNAmixtures	*EuroForMix*	m1	m2	μ	σ	ξ
E1(3K0U)	-583.089927798733	-583.089927798732	0.58313	0.32130	5082.01259	0.11275	0.10151
E1(2K1U)	-623.069145322142	-623.069145322141	0.58567	0.32439	5080.98020	0.12075	0.10813
E1(1K2U)	-662.761440697117	-662.761440697116	0.57166	0.33105	5081.36505	0.10575	0.10480
E1(0K3U)	-692.355343509357	-692.355343509358	5.9543e-01	3.1728e-01	5.0826e+03	8.6744e-02	1.0361e-01
E2(3K0U)	-599.664027562726	-599.664027562726	0.60111	0.29599	4985.57972	0.18048	0.10699
E2(2K1U)	-617.69683952578	-617.696839525781	0.57751	0.25805	4977.95000	0.14675	0.10716
E2(1K2U)	-649.811404686069	-649.811404686068	0.57299	0.26916	4980.03152	0.13630	0.11070
E2(0K3U)	-681.724194926551	-681.72419492655	0.61305	0.23166	4981.26517	0.12217	0.10977
E3(3K0U)	-614.039509472884	-614.039509472883	5.6347e-01	3.2398e-01	4.8694e+03	1.3483e-01	9.9929e-02
E3(2K1U)	-650.84283554739	-650.84283554739	0.57955	0.31307	4867.94637	0.14258	0.11022
E3(1K2U)	-698.596535698722	-698.596535698723	0.56473	0.31811	4867.80505	0.14615	0.12143
E3(0K3U)	-734.443561553084	-734.443561553084	0.59049	0.29989	4868.43967	0.13953	0.12000

Table S1: The table compares the log-likelihood values for *EuroForMix* and *DNAmixtures* for the inserted parameter value θ based on three independent samples given as single-source stains or two- or three-person mixtures for different specified hypotheses. For K=1, *EuroForMix* obtained identical values as for *DNAmixtures*. For K=2, *EuroForMix* obtained identical values as for *DNAmixtures* except for E3(1K1U) and E3(0K2U) which were different in 12th decimal. For K=3, *EuroForMix* obtained identical value as for *DNAmixtures* except for E1(3K0U), E1(2K1U), E1(1K2U), E1(0K3U), E2(2K1U), E2(1K2U), E2(0K3U), E3(3K0U), E3(1K2U) which was different in 12th decimal.

B Technical details of model and inference

Approximation of the likelihood function using a compound allele

Some alleles who have a peak height below the detection threshold T give important information to the model parameters. Consider the situation where a peak height is presented above T for allele a, but not for $a-1$ and $a+1$. Assuming the stutter-model, a contributor with allele a will expect to add some stutter to allele $a-1$. However, since the peak height at allele $a-1$ is below the threshold, this indicates a small stutter proportion. Hence the information about the non-present peak height at allele $a-1$ is important. For a contributor at allele $a+1$ we expect some stutter proportion to allele a. However, since the peak height at allele $a+1$ is below the threshold, the contributing stutter from $a+1$ to a is expected to be less than ξT. Hence the information about potential stutters from alleles falling below the threshold is not important. Based on these arguments, we make an approximation of the likelihood function in equation (**??**) by redefining the set of alleles given in the database, $\mathbf{A}_m$, as

$$\mathbf{A}'_m = \mathbf{S}_m \cup Q_m \tag{1}$$

where $\mathbf{S}_m = \{a \in \mathbf{A}_m : Y_{m,a} \geq T\} \cup \{a-1 \in \mathbf{A}_m : Y_{m,a} \geq T\}$ is the set of alleles with peak heights above the threshold and their corresponding potential stutter alleles, and Q_m is any of the remaining alleles in the set $\mathbf{A}_m \setminus \mathbf{S}_m$, grouped together as a compound allele. As used by **?**] and **?**], we let the allele frequency of Q_m be given as $1 - \sum_{a \in \mathbf{S}_m} p_a$. For the degradation model, Q_m, is assigned to have fragment length $\max_a f_{m,a}$, the maximum fragment length at (undegraded) marker m.

The likelihood function for independent replicated samples

For this particular case we assumed that the sample has been retyped (after extraction) such that stain samples are independent replicates which are assumed to contain the same contributors and satisfy the same model assumptions (same model properties across replicates). Consider the number of replicates as R, and let the observed peak height for allele a at marker m for replicate r be given as $Y^*_{r,m,a}$. Then the probabilistic model given in equation (**??**) is extended to

$$p(E|H,\boldsymbol{\theta}) = \prod_{m=1}^{M} \sum_{\substack{g_{m,k} \in \mathbb{Q}_m \\ k=1,..,K}} \left(p(\mathbf{g}_m|H) \prod_{r=1}^{R} \prod_{a \in \mathbf{A}_m} p(Y^*_{r,m,a}|\mathbf{g}_m, \boldsymbol{\theta}) \right) \tag{2}$$

C Peak height summary of the trace sample

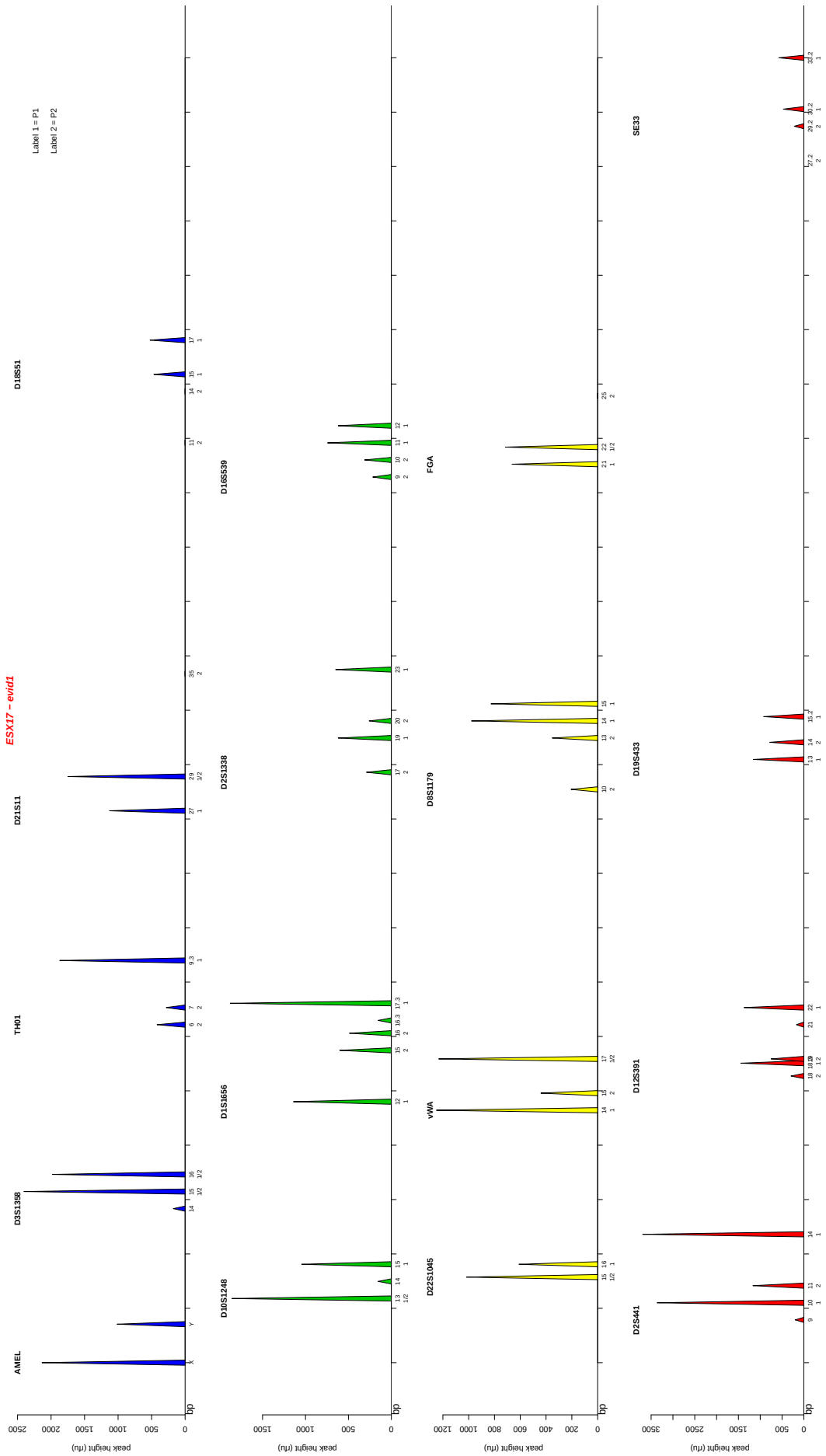


Figure S1: The plot shows the peak heights of the trace sample provided in the example, in a epg-like format.

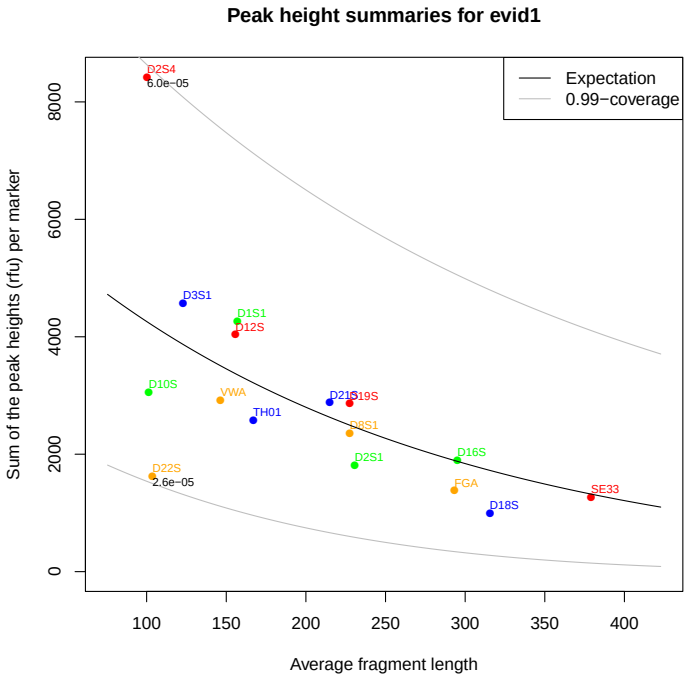
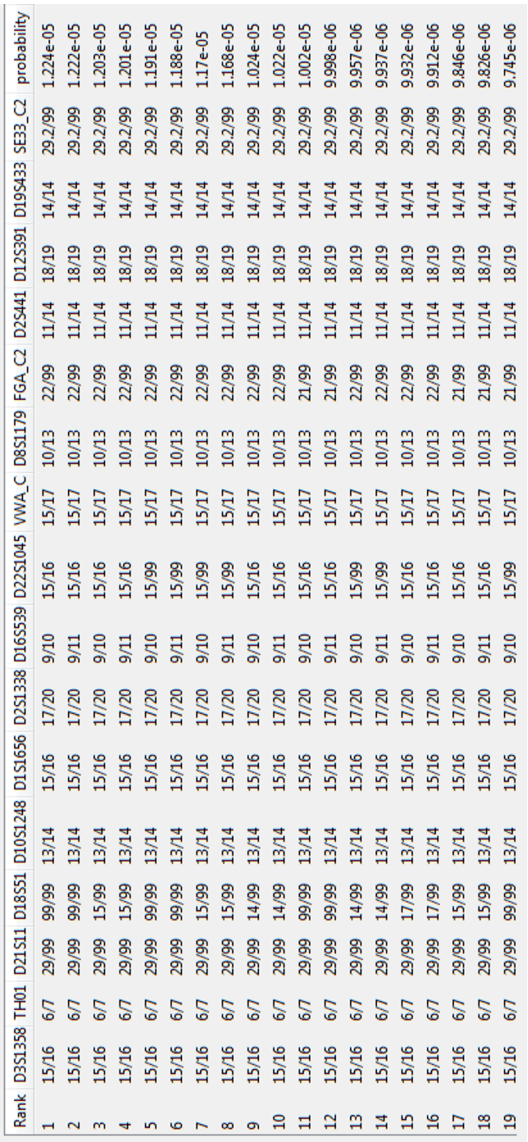


Figure S2: The plot shows the sum of the peak heights y per marker as a function of average fragment length (per marker) x for the evidence provided in the example. The lines shows the expectation, the 0.005- and 0.995-quantiles from the maximum likelihood fitted model of the underlying model $y \sim gamma(2\sigma^{-2}\beta^x, \mu\sigma^2)$ using all data. The values belonging to the points of D2S441 and D22S1045 are the probability of observing a more extreme value than observed. Each probability is based on the maximum likelihood fitted model where the corresponding marker is left out from the data.

D Resulting deconvolution table



Rank	D3S1358	TH01	D21S11	D18S51	D10S1248	D1S1656	D2S1338	D16S539	D22S1045	VWA_C	D8S1179	FGA_C2	D2S441	D12S391	D19S433	SE33_C2	probability
1	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.224e-05
2	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.222e-05
3	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.203e-05
4	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.201e-05
5	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/10	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.191e-05
6	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/11	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.188e-05
7	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/10	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.17e-05
8	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/11	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.168e-05
9	15/16	6/7	29/99	14/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.024e-05
10	15/16	6/7	29/99	14/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	1.022e-05
11	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	1.002e-05
12	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	9.998e-06
13	15/16	6/7	29/99	14/99	13/14	15/16	17/20	9/10	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	9.957e-06
14	15/16	6/7	29/99	14/99	13/14	15/16	17/20	9/11	15/99	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	9.937e-06
15	15/16	6/7	29/99	17/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	9.932e-06
16	15/16	6/7	29/99	17/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	22/99	11/14	18/19	14/14	29.2/99	9.912e-06
17	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/10	15/16	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	9.846e-06
18	15/16	6/7	29/99	15/99	13/14	15/16	17/20	9/11	15/16	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	9.826e-06
19	15/16	6/7	29/99	99/99	13/14	15/16	17/20	9/10	15/99	15/17	10/13	21/99	11/14	18/19	14/14	29.2/99	9.745e-06

Figure S3: The plot shows the 19 first ranked genotypes with corresponding probabilities for the unknown profile under a model with two contributors, but where individual P1 is known to be one of the contributors. In addition the model assumes degradation and stutters, but does not assume allele drop-in or sub-population structure. Allele "99" represents any allele not presented in the sample (a compound allele).

E Experimental results

To check the practical performance of *EuroForMix*, we simulated three random DNA profiles where one, two, three and four individuals contributed. The purpose of the experiment was to discover the time taken for the methods in this article to return results for different scenarios. The experiment was carried out by first simulating alleles for K number of contributors using the population frequencies and the corresponding allelic peak heights using the model from section 2.2 with specified model parameters. We then considered the first sampled contributor as the person of interest S. To calculate the likelihood ratio (LR) we compare hypotheses H_p :"S contributed to the sample" versus H_d :"S did not contribute to the sample". We calculated both the maximum likelihood based LR, LR_F, and the Bayesian based LR, LR_B. All samples were simulated with peak height expectation $\mu = K * 1000$, peak height coefficient of variation $\sigma = 0.2$, stutter proportion $\xi = 0.07$ and degradation slope parameter $\beta = 0.7$. Each sample was generated with different specified mixture proportions $(\pi_1, ..., \pi_K)$ given in Table S2. For the maximum likelihood estimations we required 5 random start points for the optimization. For the integrals we required relative error $\delta = 0.2$. The upper boundary of the parameters in the uniform priors were 10000 for μ, 0.5 for σ and 0.2 for ξ. For the model we assumed K number of contributors, detection threshold $T = 150$ RFU, no drop-in ($C = 0$) and no sub-population structure ($F_{st} = 0$). The degradation and stutter model was considered with the corresponding parameters β and ξ treated as unknown. The total computing time to achieve a LR value was registered with an Intel Core i7-2600 3.4 GHz processor.

From Table S2 it can be observed how the calculation time grows exponentially with number of contributors, and how the Bayesian based LR tends to be more time-consuming than the maximum likelihood based LR.

K	$\pi_1/.../\pi_K$	$log_{10}LR_F$	time (min)	$log_{10}LR_B$	time (min)
1	1	22.92	0.03	22.92	0.02
1	1	22.75	0.01	22.43	0.01
1	1	22.41	0.01	22.21	0.02
2	0.5/0.5	13.42	1.2	13.02	2.7
2	0.25/0.75	13.26	0.6	13.09	2.5
2	0.1/0.9	8.24	0.6	7.61	1.8
3	0.5/0.25/0.25	14.68	113.9	13.83	1051.6
3	0.25/0.5/0.25	6.83	102.8	4.97	179.8
3	0.1/0.5/0.4	6.93	71.72	5.9	559.1
4	0.5/0.2/0.2/0.1	18.46	7395.9	15.72	11470.3
4	0.3/0.4/0.2/0.1	9.55	7089.8	8.15	11982.9
4	0.1/0.5/0.2/0.2	4.74	9134.99	0.42	12716.9

Table S2: The table shows the resulting likelihood ratio quantities for the maximum likelihood approach, LR_F, and the Bayesian approach, LR_B, for different simulated DNA samples. K is the true and assumed number of contributors to the sample, π_x is the specified mixture proportions for contributor x (here π_1 is the mixture proportion of P). The time is total number of minutes taken to calculate the LR quantity.