Casework applications of probabilistic genotyping methods for DNA mixtures that allow relationships between contributors

Peter J. Green
UTS, Sydney, Australia
University of Bristol, UK.

Julia Mortera
Università Roma Tre, Italy.
University of Bristol, UK.

Lourdes Prieto
Forensic Sciences Institute
University of Santiago de Compostela, Spain
Comisaría General de Policía Científica, DNA Laboratory, Madrid, Spain

July 28, 2020

Abstract

In both criminal cases and civil cases there is an increasing demand for the analysis of DNA mixtures involving relationships. The goal might be, for example, to identify the contributors to a DNA mixture where the donors may be related, or to infer the relationship between individuals based on a DNA mixture. This paper applies a recent approach to modelling and computation for DNA mixtures involving contributors with arbitrarily complex relationships to two real cases from the Spanish Forensic Police.

Some key words: Coancestry, deconvolution, disputed relationship, identity by descent, kinship, DNA mixtures, likelihood ratio.

1 Introduction

In both criminal and civil cases based on relationship inference there is an increasing demand for the analysis of DNA mixtures where relatives are involved. The goal might be to identify the contributors to a mixture where the donors may or may not be related, or to determine relationships between typed individuals and one (or more) of the contributors to a mixture, also in the case that the mixture contributors themselves are related.

We analyse two real cases from the Spanish Forensic Police. In the first case we wish to identify a missing person through the analysis of DNA mixtures found on personal belongings. In many cases, the genetic profile detected on the objects is not from a single source, but might be a DNA mixture, revealing that the object was used by 2 (or more) people. In addition, very often, the contributors to these mixtures are related, mainly in cases, such as this one, where the missing person shared the dwelling with relatives.

The second case concerns a murder where a man was stabbed in his home. A DNA sample was taken from the murder weapon and appeared to be a DNA mixture from the victim and possibly a close relative of the victim.

Here we use probabilistic genotyping methods for DNA mixtures, under hypotheses about the relationships among contributors to the mixture and to other individuals whose genotype is available. Here we briefly summarise these methods and refer to Mortera (2020) which presents a review on DNA mixtures where further background can be found.

The basis for any model-based DNA mixture analysis is a joint model for the peak heights $\mathbf{z}$ in the electropherogram and genotypes represented as allele counts $\mathbf{n}$, $p(\mathbf{n}, \mathbf{z}|\psi) = p(\mathbf{n}) \times p(\mathbf{z}|\mathbf{n}, \psi)$, having parameters $\psi = (\phi, \rho, \xi, \eta)$ (Graversen 2013). Given a hypothesis on the DNA mixture contributors, the database allele frequencies, the parameters ψ, the DNA mixture model consists of two components: (a) the joint distribution $p(\mathbf{n})$ of the contributors' genotypes; (b) the conditional distribution $p(\mathbf{z}|\mathbf{n}, \psi)$ of the peak heights as observed in the electropherogram, given the genotypes. We base the analysis of the DNA mixture on the model described in Cowell *et al.* (2015). This model takes fully into account the peak heights and the possible artefacts, like stutter and dropout, that might occur in the DNA amplification process. The model can coherently analyse a combination of replicates, a combinations of different samples and a combinations of different kits. We refer to the review on DNA mixtures by Mortera (2020) for further details.

In the standard case, unknown contributors to the mixture are assumed drawn at random from the gene pool. When the contributors are related, there is positive association between their contributor genotypes. Green and Mortera (2020) present a new model aimed at making inference about complex relationships from DNA mixtures. This generalises the work in Green and Mortera (2017) which allowed inference about particular close relationships between contributors to a DNA mixture with unknown genotype and other individuals of known genotype. The new model extends the analysis to different scenarios and allows to specify arbitrary relationships between a set of actors, each of which may be mixture contributors, or have measured genotypes, or both. We can evaluate the likelihood of any such model, and compare models accordingly.

The case work examples in § 2 illustrate some simple scenarios, where we make inference about two-way relationships between two mixture contributors with and without information about their or their relatives' genotypes.

The software used to analyse the case work examples is the new `KinMix` `R` package (Green 2020) that extends the `DNAmixtures` `R` package (Graversen 2013) to allow for modelling DNA mixtures with related contributors. This software is general and can handle complex relationships with and between mixture contributors. Inference is not limited to two-way relationships but can be extended to relationships among 3 (or possibly more) contributors to a mixture.

2 Results of the analysis of complex DNA mixtures involving relationship testing

In this section we demonstrate the results and performance of our methods on the two case studies. For the first example we used the data gathered on 21 markers included in GlobalFiler$^{\text{TM}}$ Amplification kit (ThermoFisher) and in the second example we also used data on 16 markers in the PowerPlex®16 kit. In all examples we assume known allele frequencies taken from the Spanish allele frequency database collected on $n = 284$ individuals (García *et al.* 2012). In all the analyses presented we adopt a threshold of 50 rfus.

2.1 Example 1: Identification using personal belongings of a missing person

Background on the case Personal belongings such as toothbrushes or razor blades can be used as a source of DNA in missing person cases. In these objects, DNA from the missing person can be found since they may have been frequently used before his/her disappearance. Nevertheless, there is uncertainty about the actual donor of the DNA isolated from these objects, reason why

it is recommended to "validate" the detected profile by using a reference (known) sample from a relative of the missing person. Usually, these profiles (from objects and/or relatives) are then compared with DNA profiles of unidentified bodies that are stored in national databases (massive comparison). This is useful to know if the missing person has passed away but his body was not identified. Unfortunately, in some cases, the genetic profile detected on objects is not a single source profile but a DNA mixture, revealing that the object was used by 2 (or more) people. In addition, very often, the contributors to these mixtures are related (mainly in cases where the missing person shared the dwelling with relatives).

In this example, we present a real case related to a missing male. In this specific case, only a daughter of the missing male was available to donate a DNA sample. This is not the ideal situation since false DNA matches can be found after a massive comparison of profiles in a database when only one relative is available as a reference sample. In order to improve the reference genetic data, a toothbrush and a razor-blade, presumably used by the missing person, were also collected. DNA from both objects was recovered and analysed by using GlobalFiler kit (Thermo Fisher). The reference sample from the daughter of the missing male was also genotyped with GlobalFiler kit. Two different DNA mixtures were detected in the two objects. An excerpt of the data is shown in Table 1, showing the alleles and peak heights in the two DNA mixtures found on the toothbrush T and the razor-blade RB. The DNA profile of the daughter, denoted by D, is also shown.

Results Here we analyse the two DNA mixtures found on the toothbrush T, and a razor-blade RB, presumably used by the missing person (ante-mortem data).

Table 1: Example 1: An excerpt of the data from the toothbrush T and the razorblade RB, showing the markers, alleles and relative peak heights. The DNA profile of the daughter D of the missing person is also shown.

markers	alleles in mixture	toothbrush peak height	razorblade peak height	D
CSF1PO	10	1152	245	
	11	126	796	
	12	941	830	12
D22S1045	11	3218	334	
	15	3550	1795	15
	16		1274	
D5S818	11	5158	2141	11
	13	304	1512	13
vWA	14		945	
	16	264	853	16
	18	3664	612	18

Table 2: Example 1: Estimated parameters based on an analysis of the two mixture samples assuming that the toothbrush T contains DNA from two unknown contributors and the razor-blade RB contains DNA from three unknown contributors.

	μ	σ	ξ	ϕ_{U_1}	ϕ_{U_2}
toothbrush	2381	0.0614	0	0.926	0.074
razor-blade	1602	0.4955	0.0118	0.5002	0.4998

Table 2 shows the estimated parameters $\psi = (\mu, \sigma, \xi, \phi)$ for the analysis of the DNA mixtures found on T and RB. We assume there are 2 unknown contributors U_1 and U_2 to both T and RB.

Table 3: Example 1: $\log_{10} LR$ for testing whether in T and RB, $\mathcal{H}_p$ contributor (U_1 or U_2) is the father of D *vs.* $\mathcal{H}_0$ no contributor is related to D.

	$\log_{10}$ LR	
	U_1	U_2
toothbrush	10.97	4.53
razor-blade	8.442	8.444

The analysis performed for 3 unknown contributors (not shown here) yielded an almost vanishing proportion for the third contributor. These are not necessarily the same individuals contributing to T and RB. The estimated proportion of DNA for the two contributors to sample T is large for the major contributor U_1, $\phi_{U_1} = 0.93$, whereas, for item RB the estimated proportions of DNA contributed by U_1 and U_2 are roughly equal, $\phi_{U_1} \simeq \phi_{U_2} = 0.5$, implying they contributed in almost equal proportions to the mixture. As we will see in the latter case the estimation of the LR and other inference is problematic. Note that in these models, the likelihood can have a complicated shape and be difficult to safely maximise numerically. The values in Table 2 are the maximum likelihood estimates, as calculated by `DNAmixtures`.

Table 3 shows the LR and $\log_{10} LR$ for testing $\mathcal{H}_\text{p}$: D is the child of U_1 (and similarly for U_2) *vs.* $\mathcal{H}_0$: no unknown contributors are related to D. For item T, $\log_{10}$ LR $= 10.97$ is large pointing to U_1 being the father of D. It is also substantial for the hypothesis concerning U_2 being the father of D. Could this be due to the fact that the two contributors might be related? We will test this assumption later. For RB the $\log_{10}$ LR in Table 3 for the previous hypotheses is equal when testing whether D is the child of U_1 or U_2. This is probably due to the fact that the proportions are almost identical, which makes it extremely difficult to distinguish between the contributors.

Table 4: Example 1: Excerpt of marker-wise LR and overall $\log_{10}$ LR for item T, using `relMix` and `KinMix` with and without peak height information, for testing whether in T, $\mathcal{H}_p$: U_1 is the father of D *vs.* $\mathcal{H}_0$: U_1 and U_2 are random members of the population.

marker	`relMix`	`KinMix` w/o peak heights	`KinMix` with peak heights
CSF1PO	1.08	1.07	1.59
D10S1248	1.26	1.18	1.62
D5S818	2.09	2.12	1.51
vWA	2.55	2.58	3.34
partial $\log_{10}$ LR	8.35	8.42	9.94
overall $\log_{10}$ LR		9.53	10.97

Table 4 presents the marker-wise comparison between the likelihood LR and the overall $\log_{10}$ LR when using `relMix` (Hernandis *et al.* 2019) and `KinMix` with and without peak height information. `relMix` is, like `KinMix`, an `R` package that analyses DNA mixtures involving relatives, but is based only on the allele presence and does not consider the peak heights when modelling the DNA mixture. The results obtained with `relMix` and `KinMix` when not including the peak height information (columns 2 and 3) are quite similar. Small differences between `relMix` and `KinMix` when not including peak heights are to be expected since they are based on different models. For the majority of markers when including peak height information `KinMix` gave a $\log_{10}$ LR larger than when not including peak height information. When using only the markers that `relMix` is able to compute the partial $\log_{10}$ LR obtained with `KinMix` with peak heights is 9.94 and without peak heights is 8.42. The overall $\log_{10}$ LR on all the markers computed by `KinMix` with peak heights is 10.97, and without peak heights is 9.53, corresponding to a LR 27.5 times smaller.

Table 5: Example 1: For item T, $\log_{10}\text{LR}$ for $\mathcal{H}_\text{p}$: the two contributors to the mixture are related, *i.e.* U_1 has relationship R to U_2, *vs.* $\mathcal{H}_0$: the two contributors are unrelated. Several different relationships R are tested.

Relationship R between U_1 and U_2 under $\mathcal{H}_\text{p}$	$\log_{10}\text{LR}$
parent-child	-15.54
sibs	-2.14
quadruple-half-first-cousins	-0.44
half-sibs	-0.37
first cousins	-0.10
half-cousins	-0.034

Table 6: Example 1: For item RB, $\log_{10}\text{LR}$ for $\mathcal{H}_\text{p}$: the two contributors to the mixture are related, *i.e.* U_1 has relationship R to U_2, *vs.* $\mathcal{H}_0$: the two contributors are unrelated. Several different relationships R are tested.

Relationship R between U_1 to and U_2 under $\mathcal{H}_\text{p}$	$\log_{10}\text{LR}$
parent-child	-13.65
sibs	-2.85
quadruple-half-first-cousins	-0.630
half-sibs	-0.625
first cousins	-0.15
half-cousins	-0.037

Table 5 and Table 6 show the results for testing whether the contributors U_1 or U_2 to item T and RB are related, *i.e.* $\mathcal{H}_p$*: U_2 has relationship R to U_1 versus* $\mathcal{H}_0$*: U_1 and U_2 are unrelated.* The $\log_{10}\text{LR}$s are all negative, implying that the LRs are smaller than 0.1. There is almost no evidence that the two mixture contributors have a relationship among those in $R = \{$parent-child, sibs,quadruple half-cousins, half-sib, first cousins,half-cousins$\}$.

Table 7 and Table 8 show the $\log_{10}\text{LR}$ for item T for several hypotheses $\mathcal{H}_\text{p}$ concerning different relationships R among U_1, U_2 and D, *vs.* two alternative hypotheses. The first alternative hypothesis in Table 7, $\mathcal{H}_1$: U_1 is the father of D and U_2 is unrelated to D and U_1, whereas the role of U_1 and U_2 is reversed in Table 8. The second alternative hypothesis is $\mathcal{H}_0$: U_1, U_2 and D are unrelated. The values of the $\log_{10}\text{LR}$ show that it is highly likely that the two contributors to item T are the missing father of D and D's mother. It also seems more likely that the mother is the major contributor and the father the minor contributor.

Table 7: Example 1: For item T, $\log_{10}\text{LR}$ for several hypotheses $\mathcal{H}_\text{p}$ concerning different relationships R among U_1, U_2 and D, *vs.* $\mathcal{H}_1$: U_1 is the father of D and U_2 is unrelated to D and U_1 and $\mathcal{H}_0$: U_1 and U_2 and D are unrelated.

	$\log_{10}\text{LR}$	
$\mathcal{H}_p$	$\mathcal{H}_1$	$\mathcal{H}_0$
U_1 father and U_2 mother of D	6.960	17.935
U_1 father of D and U_2 maternal aunt of D	4.605	15.579
U_1 father of D and U_2 paternal cousin of D	0.375	10.600
U_1 sib of U_2 and father of D	-2.140	8.834
U_1 father of both D and U_2	-10.990	-0.016

Table 8: Example 1: For item T, $\log_{10}$ LR for several hypotheses $\mathcal{H}_\text{p}$ concerning different relationships R among U_1, U_2 and D, *vs.* $\mathcal{H}_1$: U_2 is the father of D and U_1 is unrelated to D and U_2 and $\mathcal{H}_0$: U_1 and U_2 and D are unrelated.

	$\log_{10}$ LR	
$\mathcal{H}_p$	$\mathcal{H}_1$	$\mathcal{H}_0$
U_2 father and U_1 mother of D	13.404	17.935
U_2 father of D and U_1 maternal aunt of D	11.049	15.579
U_2 father of D and U_1 paternal cousin of D	6.069	10.600
U_2 sib of U_1 and father of D	4.303	8.834
U_2 father of D and U_1	−4.547	−0.016

Table 9: Example 1: Predicted genotypes of the major contributor U_1 with corresponding probabilities for item T and for both major and minor contributor for item RB for an excerpt of the markers. The genotype of D is also shown.

	toothbrush			razor-blade major			razor-blade minor				
markers	genotype		prob.	genotype		prob.	genotype		prob.	D	
CSF1PO	10	12	1	10	11	0.42	11	12	0.44	12	12
				11	12	0.24	10	12	0.33		
				10	12	0.22	12	12	0.23		
				11	11	0.11					
D22S1045	11	15	0.9999	11	16	0.40	15	16	0.40	15	15
				11	15	0.26	11	15	0.33		
				15	16	0.26	15	15	0.26		
				16	16	0.07					
D5S818	11	11	1	11	13	0.56	11	13	0.63	11	13
				11	11	0.35	11	11	0.22		
				13	13	0.08	13	13	0.15		
vWA	18	18	0.9999	14	16	0.42	16	18	0.43	16	18
				14	18	0.24	14	18	0.24		
				16	18	0.15	14	16	0.17		
				16	16	0.08	16	16	0.10		
				14	14	0.08	18	18	0.05		
				18	18	0.01					

Table 9 shows the deconvolution of the mixtures in items T and RB. The predicted profile of the major contributor of the T mixture has probabilities close to 1 on all markers. This profile is compatible with being the father of D as it shares at least one allele on all markers. As in the previous analyses, the deconvolution of the mixture in RB yields very uncertain predictions of the major and minor contributor's genotype as there are many candidate genotypes and all the highest ranking probabilities are smaller than 0.5.

2.2 Example 2: Analyses of a Spanish murder case

Description of the case This concerns a murder case where a man was stabbed in his home. There was a knife with blood at the crime scene. The blood was mainly on the blade, but there was also some blood on the handle. The sample from the handle turned out to be a DNA mixture, with a major profile matching the victim. We also wish to test whether the minor profile in the mixture could be a close relative of the victim (possibly a son). The DNA profile of the son was

not available. Two EPGs of the mixture were obtained by using two different kits, we denote these by EPG1 and EPG2. The kits have partially overlapping sets of markers, EPG1 was analysed on its 16 markers and EPG2 on its set of 22 markers, both include Amelogenin.

Months later, a man was arrested for a different crime (drug trafficking) and a reference DNA sample was collected. When his profile was entered in the DNA database several matches were found, among which with the DNA mixture on the handle of the knife. The matches were investigated and the identity of the person (name, date of birth, place of birth, name of the father, name of the mother) was that of the son of the victim. Table 10 gives an excerpt of the data showing the markers, alleles and relative peak heights for EPG1 and EPG2, together with the father and son's genotypes.

Table 10: Example 2: An excerpt of the data showing the markers, alleles and relative peak heights for EPG1 and EPG2, together with the father's and son's genotypes

marker	allele	EPG1 height	EPG2 height	victim	suspect
CSF1PO	10	305	625	10	10
	11	240	504	11	11
D10S1248	13		6990	13	
	14		2309		14
	16		7144	16	16
D7S820	9	606	1136	9	9
	10		686	10	
TH01	9.3	863	2654	9.3	9.3
	10	570			10

Results We analysed the data from this case to illustrate the different queries that can be analysed using the recently developed `Kinmix` code.

In particular we analyse the following different possible scenarios:

Scenario 1 Here none of the contributors are typed. The analysis of a 2-person mixture model for a prosecution hypothesis $\mathcal{H}_p$: being the two unknowns being father and son versus $\mathcal{H}_0$ the two unknown contributors are unrelated.

Scenario 2 Here only the father (the victim) is typed. Analysis of a 2-person mixture model, where father has been typed and the prosecution hypothesis is $\mathcal{H}_p$: son of father and 1 unknown are contributors versus $\mathcal{H}_0$: no contributor is related to the typed individual (the victim).

Scenario 3 Both father and son are typed. Here we analyse a 2-person mixture model where $\mathcal{H}_p$: the contributors are victim (father) and son versus $\mathcal{H}_0$: contributors to the mixture(s) are 2 unknown individuals.

Scenario 4 Both father and son are typed. Here we analyse a 2-person mixture model where $\mathcal{H}_p$: the contributors are victim (father) and son versus $\mathcal{H}_0$: contributors to the mixture(s) are the victim and an unknown.

In all scenarios, unless otherwise stated, when considering an unknown contributor to a mixture, he is taken to be a random member of the reference population, so unrelated to typed individuals.

For EPG1 the MLEs of the parameters under both $\mathcal{H}_p$ and $\mathcal{H}_0$ are similar and are roughly equal to $\psi = (\mu = 576, \sigma = 0.32, \xi = 0, \phi_{U_1} = 0.88, \phi_{U_2} = 0.12)$. When the victim's genotype is known the estimated proportion contributed to EPG1 is $\phi_v = 0.18$, $\phi_{U_1} = 0.82$. For EPG2 the MLEs of the parameters are roughly equal to $\psi = (\mu = 2542, \sigma = 0.97, \xi = 0, \phi_{U_1} = 0.75, \phi_{U_2} = 0.25)$. When the victim's genotype is known the estimated proportion contributed to EPG1 is $\phi_v = 0.14$, $\phi_{U_1} = 0.86$. In both EPG1 and EPG2 the victim is estimated to be the minor contributor. Note that EPG2 has a higher μ than EPG1 but this is also accompanied by a larger σ, so the coefficient of variation is similar in both EPGs. The MLEs of the mean stutter proportion ξ are zero, which indicates that the data has been preprocessed so that peaks that were classified in the laboratory as stutter have been removed. Our models do not, however, require that the data be preprocessed, thus avoiding eliminating a true peak in stutter position.

Table 11 gives the $\log_{10}$ LR for the 4 scenarios when analysing EPG1 and EPG2 separately and jointly. When combining EPGs made from the same DNA extract, as in this case, it is natural to make an assumption that contributors are the same. In Graversen *et al.* (2019) we show how results based on a combination of replicates, a combinations of different samples and a combinations of different kits improve the robustness of the analysis and help in fixing any complications relating to degradation. However, when combining profiles from different samples one needs to carefully consider whether there is perhaps only a partial overlap.

Table 11: Example 2: $\log_{10}$ LR for Scenarios 1–4 using EPG1 and EPG2 separately and in combination.

Scenario	1	2	3	4
Typed actors	none	father	father & son	
EPG1	−0.806	5.60	22.16	22.78
EPG2	−0.175	10.66	29.16	11.68
EPG1& EPG2	2.49	8.26	40.17	26.20

Table 12: Example 2: For item EPG1 and EPG2, $\log_{10}$ LR for $\mathcal{H}_\text{p}$: the two contributors to the mixture are related, *i.e.* U_1 has relationship R to U_2, *vs.* $\mathcal{H}_0$: the two contributors U_1 and U_2 are unrelated and are independent of the typed individuals. Several different relationships R are tested.

	$\log_{10}$ LR	
Relationship	EPG1	EPG2
parent-child	−0.806	−0.175
sibs	−1.270	−0.940
quadruple-half-first-cousins	−0.316	−0.022
half-sibs	−0.275	0.045
first cousins	−0.108	0.059
half-cousins	−0.046	0.040

Table 12 shows $\log_{10}$ LR for testing whether the two unknown contributors to the DNA mixture are related versus that they are unrelated. For EPG1 the LRs for testing $\mathcal{H}_\text{p}$ that the U_1 has a relationship R to U_2, *vs.* $\mathcal{H}_0$: the two contributors U_1 and U_2 are unrelated and are independent of the typed individuals, vary between $[0.16, 0.9]$ giving roughly equal weight to $\mathcal{H}_1$ versus $\mathcal{H}_0$. For EPG2 these vary between $[0.11, 0.86]$.

Table 13 shows the deconvolution for the major contributor to the mixture for the two EPGs. The table only indicates genotype probabilities of at least 0.001, meaning that a blank cell represents a probability of less than 0.001. We have denote by *other* the collection of alleles for which no peak

has been observed in the EPG. For EPG1 the highest ranking genotype for the major contributor U_1 on all markers has posterior probability greater than 0.99 and coincides with the genotype of the suspect (who is the son of the victim) on all markers. The deconvolution for EPG2 gives a much poorer performance. For example, on marker D7S850 the top ranking genotype for EPG2 is incorrect, the correct genotype (9,9) is ranked 3$^{\text{rd}}$ having a small probability of 0.077.

Table 13: Example 2: Predicted genotypes of U_1 with corresponding probabilities for EPG1 and EPG2 for an excerpt of the markers. An allele not observed in the EPG is denoted by *other*.

	EPG1			EPG2		
	genotype		prob.	genotype		prob.
CSF1PO	10	11	1	10	11	0.751
				10	10	0.097
				11	11	0.083
				10	*other*	0.036
				11	*other*	0.033
D13S317	12	13	0.997	12	13	0.576
	12	12	0.003	12	12	0.363
				12	*other*	0.043
				13	*other*	0.011
				13	13	0.006
D7S820	9	9	1	9	10	0.768
				10	10	0.077
				9	9	0.077
				9	*other*	0.043
				5 10	*other*	0.034
TH01	9.3	10	1	9.3	9.3	0.812
				9.3	*other*	0.185
				other	*other*	0.003

3 Conclusions

We have shown that a wide range of relationship inference problems where one or more actors appear only as contributors to a DNA mixture, can be handled coherently. We can make inference about relationships among contributors, and between contributors and typed individuals.

The new `KinMix` package (Green 2020) used in the casework examples illustrated here is a highly flexible modular software capable of solving much more complex relationships among two or more mixture contributors than those presented here. It is not limited to pairwise relationships. In Green and Mortera (2020) we show its capabilities of dealing with multi-way relationships in DNA mixtures including cases where the contributors might be inbred.

References

Cowell, R. G., Graversen, T., Lauritzen, S. L., and Mortera, J. (2015). Analysis of DNA mixtures with artefacts (with discussion). *Journal of the Royal Statistical Society: Series C*, **64**, 1–48.

García, O., Alonso, J., Cano, J., García, R., Luque, G., Martín, P., de Yuso, I. M., Maulini, S., Parra, D., and Yurrebaso, I. (2012). Population genetic data and concordance study for the kits Identifiler, NGM, PowerPlex ESX 17 System and Investigator ESSplex in Spain. *Forensic Science International: Genetics*, **6**, (2), e78–9.

Graversen, T. (2013). *DNAmixtures: Statistical Inference for Mixed Traces of DNA*. R package version 0.1-4. `http://dnamixtures.r-forge.r-project.org`.

Graversen, T., Mortera, J., and Lago, G. (2019). The Yara Gambirasio case: Combining evidence in a complex DNA mixture case. *Forensic Science International: Genetics*, **40**, 52–63.

Green, P. J. (2020). *`KinMix`: DNA mixture analysis with related contributors*. R package 2.0, `https://petergreenweb.wordpress.com/kinmix2-0`.

Green, P. J. and Mortera, J. (2017). Paternity testing and other inference about relationships from DNA mixtures. *Forensic Science International: Genetics*, **28**, 128–37. `http://dx.doi.org/10.1016/j.fsigen.2017.02.001`.

Green, P. J. and Mortera, J. (2020). Inference about complex relationships using peak height data from DNA mixtures. `https://arxiv.org/abs/2005.09365`.

Hernandis, E., Dørum, G., and Egeland, T. (2019). relMix: An open source software for DNA mixtures with related contributors. *Forensic Science International: Genetics Supplement Series*, **7**, 221–3. `https://doi.org/10.1016/j.fsigss.2019.09.085`.

Mortera, J. (2020). DNA mixtures in forensic investigations: The statistical state of the art. *Annual Review of Statistics and Its Application*, **7**, 1–34. `https://doi.org/10.1146/annurev-statistics-031219-041306`.