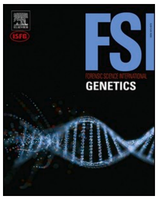
Forensic Science International: Genetics 78 (2025) 103275

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



Evaluating the effect of marker panel sizes on estimation of bio-geographical co-ancestry proportions

M. de la Puente [a,*], L. Casanova-Adán [a], J. González-Bao [a], J. Pardo-Seco [b,c], A. Mosquera-Miguel [a], A. Ambroa-Conde [a], J. Ruiz-Ramírez [a], A. Cabrejas-Olalla [a], M. Boullón-Cassau [a], A. Freire-Aradas [a], A. Rodríguez [a], C. Phillips [a,d], M.V. Lareu [a]

[a] *Forensic Genetics Unit, Institute of Forensic Sciences, Universidade de Santiago de Compostela, Spain*

[b] *Genetics, Vaccines and Pediatric Infectious Diseases Research Group (GENVIP) and Genética de Poblaciones en Biomedicina (GenPoB) Research Group, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago de Compostela, Galicia, Spain*

[c] *Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain*

[d] *King's Forensics, Faculty of Life Sciences and Medicine, King's College, London, UK*

ARTICLE INFO

Keywords:
Massively parallel sequencing
Biogeographical analysis
SNPs
Ancestry
Admixture

ABSTRACT

A large number of ancestry-informative marker panels have been developed for forensic bio-geographical ancestry (BGA) analysis during the past decade, which offer valuable investigative tools for cold cases. The developed assays for capillary electrophoresis (CE) and massively parallel sequencing (MPS) focus on the differentiation of major populations, with MPS allowing larger numbers of markers that can be multiplexed at the same time and therefore improved differentiation of more closely related Eurasian populations. One limitation of BGA inference tools is the handling of co-ancestry in individuals with admixted backgrounds, which leads to two situations being indistinguishable: (i) the individual belongs to an admixed population, or (ii) the individual has recent ancestors from different populations. Accurate and precise co-ancestry estimates can help, as first or second-degree admixture would show a ~ 50–50 % or ~ 75–25 % ratio of co-ancestry proportions. Here we compared the co-ancestry proportion estimations obtained for the set of 2504 individuals from the 1000 Genomes Project with dedicated BGA and human identification (ID) assays of different sizes compared to those obtained with the > 500,000 SNP Affymetrix Human Origins panel as the point of reference for each individual. The results of the correlation analysis performed with > 500 admixed individuals indicate that panel size plays a major role in the accuracy of the co-ancestry estimates. Therefore, the large-scale forensic MPS ID panels we evaluated constitute a valuable alternative to small- and medium-scale BGA panels, especially when admixture is expected.

1. Introduction

Forensic investigative tools, such as SNP-based phenotyping or biogeographical ancestry (BGA) informative assays, have gained increasing attention in the field due to the valuable information they can provide to police investigations when no suspect has been identified or there are no matches with databases.

A large number of ancestry-informative marker (AIM) panels have been developed for BGA analysis, focusing on the differentiation of 3–5 major continental populations (Sub-Saharan Africa, Europe, East Asia, Oceania, Native America) for capillary electrophoresis (CE) assays, comprising up to 40 markers [1,2]. Assays for massively parallel sequencing (MPS), differentiating 5–8 major populations (including genetically closer Eurasian populations such as South Asia, Middle East or North Africa) can genotype up to hundreds of markers allowing a greater degree of population resolution [3–5].

One limitation of the BGA inference tools is admixture [6]: individuals with admixed backgrounds will have differing degrees of co-ancestry contributions from two or more ancestral populations. Such patterns represent a challenge when performing BGA analysis as two scenarios cannot be easily distinguished: (i) that the individual belongs to an admixed population, or (ii) that the individual has first (parental) or second order (grandparental) antecedents from different ancestral populations. The correct estimation of the co-ancestry proportions can

---

* Corresponding author.
E-mail address: mdelcarmendela.puente@usc.es (M. de la Puente).

https://doi.org/10.1016/j.fsigen.2025.103275
Received 25 October 2024; Received in revised form 5 March 2025; Accepted 14 March 2025

help distinguish these situations by discounting the second option if estimates do not match a ~ 50–50 % or ~ 75–25 % pattern. However, several factors affect the accuracy of co-ancestry proportion estimates, including the balance of the population-specific divergence values amongst the selected SNPs and the size of the ancestry panel used.

In this work, we investigate the effect of the number of SNPs in the panel on the accuracy of co-ancestry proportions estimations. For this purpose, we selected a range of forensic panels with different sizes and compared both BGA-dedicated and ID/kinship marker panels. We compared the co-ancestry proportion estimations obtained with these panels for the 2504 individuals of the 1000 Genomes Project, including more than 500 admixed individuals, with those obtained with the Affymetrix Human Origins panel, an array-based set of > 500,000 SNPs which have been specifically selected to reduce ascertainment bias.

2. Materials and methods

2.1. Panel selection and data acquisition

Table 1 shows the characteristics of ten forensic SNP panels [1–5, 7–11] from which co-ancestry estimates were evaluated against those obtained with the Affymetrix Human Origins panel [12]. The panels can be classified according to the number of markers included in the assay as: small-scale (< 60 SNPs), medium-scale (60–200 SNPs) and large-scale (> 1000 SNPs) and according to the criteria for SNP selection and applications as BGA or Identification (ID)/Kinship panels.

For each of the eleven SNP panels described in Table 1, genotype data for a total of 2504 individuals distributed in five major populations (as outlined in Supplementary Table S1) was gathered from publicly available high-coverage 1000 Genomes data [13]. When necessary, certain genotype data were completed with 1000 Genomes Phase III data [14].

2.2. Ancestry analysis and co-ancestry proportions comparison

Genotype data from the Human Origins panel, was transformed into corresponding bed, bim and fam input files using PLINK 1.9 [15] and analysed with the ADMIXTURE genetic cluster algorithm [16] at a preset cluster number of K= 5, under default parameters, to estimate co-ancestry proportions. The ADMIXTURE algorithm is optimized for large-scale biallelic SNP datasets, so was applied uniquely to the Human Origins genotypes, prior to analysis of forensic panel datasets with the equivalent STRUCTURE algorithm, capable of handling the multi-allelic information of SNPs included in some of the forensic panels evaluated. While the largest panels can differentiate further sub-continental components such as in Fig. 2 in [14], this is not the case with the smallest panels, which display inconsistent patterns when considering a number of components greater than the number of major populations used as the reference data. Therefore, for comparison purposes, K was kept at five for all panels.

Table 1
Characteristics of the SNP panels included in the analysis.

Panel	Refs.	CE / MPS	Number of SNPs	AIM SNPs	ID SNPs
Global AIMs Nano	[1]	CE	31	✓	
34-plex AIMs	[2]	CE	34	✓	
52-plex ID-SNPs	[7]	CE	52		✓
ForenSeq DNA Signature AIMs	[4]	MPS	56	✓	
Precision ID Identity	[7, 10]	MPS	90		✓
VISAGE Basic Tool AIMs	[5]	MPS	153	✓	
Precision ID Ancestry	[3,4]	MPS	165	✓	
MPSplex	[8]	MPS	1239		✓
FORCE	[11]	MPS	4347	✓	✓
ForenSeq Kintelligence	[9]	MPS	9933	✓	✓
Human Origins	[12]	MPS	572743	✓	✓

For the remaining panels, genotype data was transformed into STRUCTURE format using the conversion tool available at the Snipper portal (http://mathgene.usc.es/snipper/). A total of 10 iterations were performed in STRUCTURE, using parameters: 100,000 burnin steps and 100,000 MCMC steps, correlated allele frequencies under the Admixture model (no POPFLAG). The 10 iterations were merged using CLUMPAK under the StructureSelector portal [17] defining a threshold value of 0.95 for similarity scores and minimal cluster size.

Correspondence of co-ancestry proportions amongst the different panels was evaluated both visually and using descriptive statistics for all populations, by comparing the plot patterns obtained for all panels and calculating summary statistics of the co-ancestry values including: mean, standard deviation, minimum, maximum and median, with individuals grouped by superpopulation (considering the admixture) or population. Calculations and plots were made using R language v.4.4.0 [18] with packages ggplot2 [19], dplyr [20] and tidyR [21].

For the six admixed populations highlighted in Supplementary Table S1, correlation analysis and the adjustment of a linear regression model between the co-ancestry proportions of forensic panels vs. Human Origins for Sub-Saharan African (AFR), European (EUR) and Native American (AMR) components was performed at three levels: (i) population level; (ii) per ancestry component; and (iii) global level. Both graphical visualization and data processing were performed in R language v.4.4.0 [18] using the packages ggplot2 [19], dplyr [20] and tidyR [21]. Pearson correlation coefficients (r) were calculated using the ‘cor’ function; the coefficient of determination (R^2) values were calculated by squaring the Pearson correlation coefficients and the linear regression coefficients and their standard error were calculated using the ‘lm’ function. Concentrating on the AFR, EUR, AMR admixture components was informed by the detailed ADMIXTURE-based population analysis of 1000 Genomes populations performed by the Project, which identified these three components as comprising > 0.95 total co-ancestry proportions in the six admixed populations (Fig. 2 in [14]). A 2-way admixture of AFR and EUR components is expected for populations ACB and ASW, while a 3-way admixture of AMR, AFR and EUR components is expected for PEL, MXL, CLM and PUR (population codes are listed in Supplementary Table S1). Note that two admixture outliers in the Peruvian in Lima, Peru (PEL) samples exist for East Asian co-ancestry, i.e. their total co-ancestry proportions for AFR, EUR and AMR are less than 0.95. These are HG02345 with ~0.1 East Asian co-ancestry proportion, and HG01944 with ~ 0.4 East Asian co-ancestry proportion.

3. Results and discussion

3.1. Visual and descriptive comparison of co-ancestry proportions

Fig. 1 shows a graphical representation of the Admixture (Human Origins) and STRUCTURE (forensic panels) co-ancestry proportion results per sample for each panel, ordered by the number of SNPs in each assay. The co-ancestry proportion values obtained are listed in Supplementary Table S2. Overall, all the panels show some degree of differentiation of the major populations, with major ancestry components corresponding to the five groups, as expected: Sub-Saharan African (AFR), European (EUR), South Asian (SAS), East Asian (EAS) and Native American (AMR), with the AMR component showing variable proportions in the four Admixed American populations (in decreasing order of variability: PEL, MXL, CLM and PUR).

The largest forensic panels of MPSplex, FORCE, and Forenseq Kintelligence) present closer patterns to the Human Origins panel, indicating an additive effect of panel size on the accuracy of co-ancestry proportion estimates. This is the case even when a panel contains ID-SNPs, such as MPSplex, and the similarities are more pronounced in admixed populations ACB, ASW, PEL, MXL, CLM and PUR (see

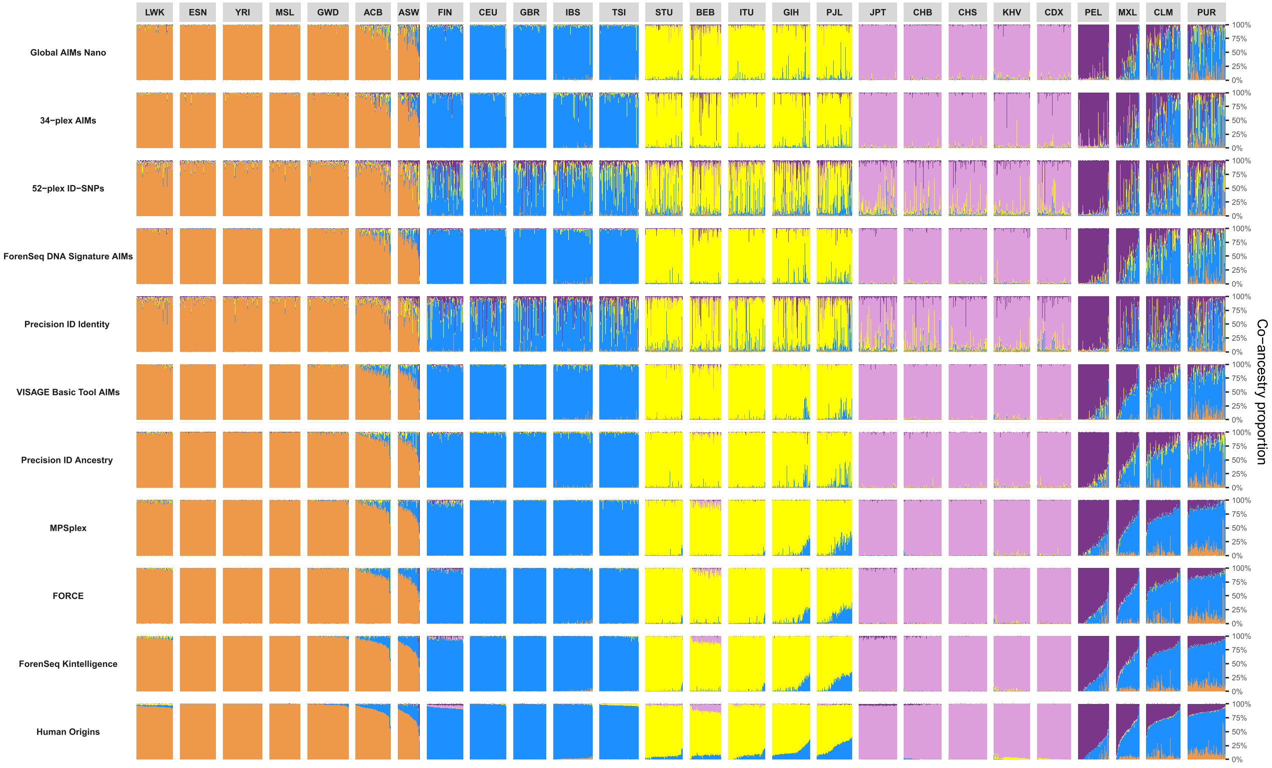


Fig. 1. Graphical representation of the ADMIXTURE (Human Origins) and STRUCTURE (other panels) co-ancestry proportion values for each panel, ordered by the number of SNPs in the assay. The different ancestry components are coloured according to Rosenberg et al. 2002 [22]: orange for Sub-Saharan African, blue for European, yellow for South Asian, pink for East Asian and purple for Native American. Population codes are listed in Supplementary Table S1.

Supplementary Table 1 for population descriptions). Moreover, the Humans Origins panel shows a small degree of EUR and EAS co-ancestry in South Asia populations, matched in the results from panels exceeding 1000 markers.

Both small-scale CE-based, and medium-scale MPS-based assays of dedicated AIM-SNPs produce less co-ancestry noise in European populations. In contrast to the MPS AIM-SNP assays, the CE assays selected for this study were not developed to differentiate South Asians, reflected in the better performance of the MPS AIM-SNPs, showing patterns that better match those of the Human Origins panel.

Small-scale ID panels, such as 52-plex or the Precision ID Identity Panel, show the highest level of 'co-ancestry noise' (or non-specific co-ancestry, which we define here as: a co-ancestry component that is not expected for the individual according to the definition of its population; not observed in, or matching the patterns in the analysis of the reference SNP panel) in comparison with dedicated forensic AIM panels of comparable size. This effect is most evident in the European and South Asian populations, where a proportion of individuals show a significant level of non-specific AMR co-ancestry.

Mean co-ancestry proportions across populations and superpopulations (considering admixed and unadmixed populations separately) are presented in Supplementary Fig. S1 and Supplementary Fig. S2, respectively. These values are listed, as well as the standard deviation, maximum, minimum and median values, in Supplementary Tables S4 and S5. Mean AMR co-ancestry proportions in European populations reached values of 0.1366 and 0.1394 for the 52-plex ID-SNPs and Precision ID Identity Panel, respectively, while maintaining levels under 0.0150 for the other panels. In the case of South Asian populations, similar AMR co-ancestry noise affected the patterns observed in the smallest BGA panels (Global AIMs Nano, 34-plex AIMs and ForenSeq DNA Signature AIMs), which can be explained by the fact that those panels were not developed to differentiate the South Asian population, resulting in a lack of informativeness. However, overall co-ancestry noise rates remain lower for BGA panels than for ID panels, with mean AMR ancestry proportions in South Asian populations reaching values of 0.0841 and 0.0592 for the 52-plex ID-SNPs and Precision ID Identity Panel, respectively; and, 0.0207, 0.0418 and 0.0191 for the Global AIMs Nano, 34-plex AIMs and ForenSeq DNA Signature AIMs, respectively; with AMR co-ancestry proportions under 0.0080 for the other panels.

Table 2 shows mean non-specific co-ancestry proportion values per superpopulation of the different panels, excluding admixed and South Asian populations, as those have previously shown specific-population patterns of co-ancestry in Fig. 2. In comparison to the other panels, showing values below the 0.050 threshold (but for Europe with the 34-plex AIMs), the two ID-SNPs panels reach values of 0.0590 and 0.0559 in Sub-Saharan Africans, 0.3106 and 0.2651 in Europe and 0.1781 and 0.1620 in East Asia, for the 52-plex ID-SNPs and Precision ID Identity Panel, respectively.

Table 2
Mean non-specific co-ancestry proportion values per superpopulation of the different panels, discounting admixed and South Asian populations.

Panel	Non-specific co-ancestry		
	Sub-Saharan Africa	Europe	East Asia
Global AIMs Nano	0.0185	0.0366	0.0330
34-plex AIMs	0.0246	0.0550	0.0459
52-plex ID-SNPs	0.0590	0.3106	0.1781
ForenSeq DNA Signature AIMs	0.0137	0.0325	0.0287
Precision ID Identity	0.0559	0.2651	0.1620
VISAGE Basic Tool AIMs	0.0117	0.0259	0.0194
Precision ID Ancestry	0.0101	0.0228	0.0227
MPSplex	0.0118	0.0206	0.0116
FORCE	0.0080	0.0176	0.0127
ForenSeq Kintelligence	0.0141	0.0214	0.0159
Human Origins	0.0149	0.0283	0.0180

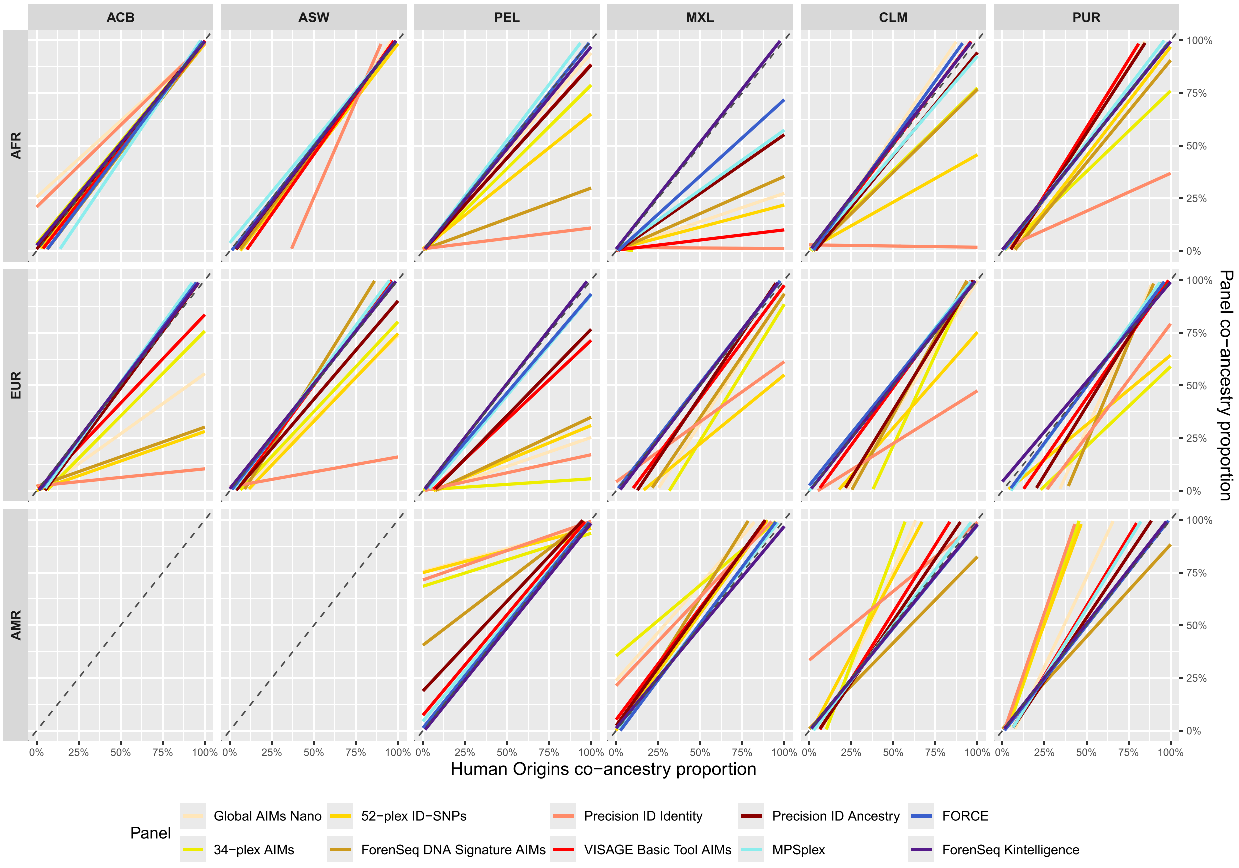


Fig. 2. Adjusted linear regression models for the different panels vs. Human Origins. The dashed line represents a perfect correlation. Population codes are listed in Supplementary Table S1. AFR: Sub-Saharan African, EUR: European, AMR: Native American. To simplify the analysis of three components of admixture in ACB and ASW admixed African populations, the AMR component was not included in the analysis. Panels are coloured according to their size: small-scale (< 60 SNPs) in yellow, medium-scale (60–200 SNPs) in red and large-scale (> 1000 SNPs) in blue.

3.2. Correlation analysis of the co-ancestry proportions for admixed populations

A formal analysis of the correlation and the adjustment of a linear regression model between the co-ancestry proportions of the Sub-Saharan African (AFR), European (EUR) and Native American (AMR) components in admixed populations of the different panels vs. Human Origins is presented in Fig. 2 and Supplementary Table S5 (at a population level), Supplementary Table S6 (per ancestry component) and Table 3 (at a global level). The largest AIM-SNP panels show lines closer to full correlation, with global coefficient of determination (R^2) values over 0.95 for the Forenseq Kintelligence kit, and over 0.85 for MPSplex and FORCE panels (Table 3). The lowest global R^2 values were found in the small-scale ID-SNP panels, with values of 0.277 and 0.216 for 52-plex ID-SNPs and Precision ID identity panels respectively, in agreement with the analyses presented in Section 3.1. The standard error of the regression coefficient follows the same trend, with the closest values to zero in the largest panels. For each of the three ancestry components (Supplementary Table S6) the obtained values did not show any other tendencies, beyond increasing R^2 values and decreasing standard error of the regression coefficient with panel size.

Table 3
Global linear regression coefficients (β) and their standard errors (SE), Pearson correlation coefficient (r) and coefficient of determination of determination (R^2) of co-ancestry values across all panels tested (ordered by number of SNPs analysed) vs. Human Origins for the admixed population.

Panel	β	SE	r	R^2
Global AIMs Nano	1.0544	0.0258	0.8914	0.7946
34-plex AIMs	0.9275	0.0345	0.7857	0.6173
52-plex ID-SNPs	0.9190	0.0306	0.8094	0.6551
ForenSeq DNA Signature AIMs	1.0328	0.0194	0.9319	0.8684
Precision ID Identity	0.9155	0.0341	0.7872	0.6197
VISAGE Basic Tool AIMs	1.0370	0.0116	0.9729	0.9465
Precision ID Ancestry	1.0539	0.0114	0.9747	0.9500
MPSplex	1.0411	0.0050	0.9948	0.9896
FORCE	1.0249	0.0040	0.9965	0.993
ForenSeq Kintelligence	0.9976	0.0024	0.9987	0.9974

4. Concluding remarks

The results presented in this work highlight the fact that the accurate determination of co-ancestry proportions in admixed individuals remains challenging using both small- and medium-scale multiplexes, including dedicated commercial BGA MPS kits. Results from small- and medium-scale forensic ID panels showed the highest levels of co-ancestry noise in terms of the proportions of non-specific components for unadmixed population samples. Small or medium scale dedicated forensic BGA panels mainly comprise AIMs selected to show a maximum level of allele differentiation between two populations. As an alternative, focusing on SNPs that have alleles which show absolute specificity in each of the populations (with zero or near-zero allele frequencies for the population-specific allele outside of the target population) could enhance co-ancestry analysis by introducing specificity to the assignment of admixture contributors rather than just making a differential comparison of the proportion of shared alleles between the contributors.

The precision of co-ancestry estimates made in forensic BGA analyses is highly dependent on the size of the SNP panel used, with panels over > 1000 SNPs showing high correlation rates with the Human Origins panel (> 500,000 SNPs) used as the study reference. While the amount and quality of DNA necessary for array-based assays usually exceeds the requirements of forensic DNA analysis, often comprising limited and/or degraded samples, the forensic large-scale MPS panels evaluated in this study have shown much higher levels of sensitivity, which make them much more practical for forensic applications. Therefore, they constitute a valuable alternative to small- and medium-scale BGA panels, especially in the detection and analysis of individual admixture. This is particularly true in the case of BGA screening for identification of degraded remains from unknown missing persons through reference samples from surviving relatives, as the large amount of information collected by the panels with the ID SNPs can lead to accurate ancestry prediction models even when ancestry informative SNPs were not successfully genotyped or are not included. Moreover, the degree of accuracy of the co-ancestry proportion estimates offered by the largest panels is indicative that, with those panels, the analyst could discard first or second order admixture if the proportions do not match the ~ 50–50 % or ~ 75–25 % pattern.

CRediT authorship contribution statement

María Victoria Lareu: Writing – review & editing, Supervision, Project administration, Funding acquisition. **Christopher Phillips:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Miguel Boullón-Cassau:** Writing – review & editing, Validation, Resources. **Amaia Cabrejas-Olalla:** Writing – review & editing, Validation, Resources. **Amelia Rodríguez:** Writing – review & editing, Resources. **Ana Freire-Aradas:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition. **Ana Mosquera-Miguel:** Writing – review & editing, Validation, Resources, Project administration, Funding acquisition. **Jacobo Pardo-Seco:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis. **Jorge Ruiz-Ramírez:** Writing – review & editing, Validation, Resources. **Adrián Ambroa-Conde:** Writing – review & editing, Validation, Resources. **María de la Puente Vila:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Software, Resources, Project administration, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Javier González-Bao:** Writing – review & editing, Writing – original draft, Visualization, Software, Data curation. **Lucía Casanova-Adán:** Writing – review & editing, Writing – original draft, Visualization, Software, Data curation.

Declaration of Competing Interest

The authors declare not conflicts of interest.

Acknowledgements

MdP is supported by Grant IJC2020-042638-I funded by the Gobierno de España MCIN/AEI/10.13039/501100011033 and the European Union “NextGenerationEU/PRTR”. LCA is supported by a predoctoral grant funded by “Fundación Caixa Rural Galega Tomás Notario Vacas” in collaboration agreement with the Universidade de Santiago de Compostela. JGB is supported by predoctoral grant “Ayudas para la Formación del Profesorado Universitario” (FPU22/02683) funded by Ministerio de Universidades from Gobierno de España. MBC is supported by Grant ED481A-2024-163 “Axudas de apoio á etapa predoutoral” funded by the Consellería de Cultura, Educación, Formación Profesional e Universidades from Xunta de Galicia, Spain.” MVL is supported by grant PID2019-107876RB-I00 funded by the Ministerio de Educación, Cultura y Ciencia, Spain (MCIN/AEI/ 10.13039/501100011033) and grant PID2022-141224OB-I00 funded by MCIN/AEI/10.13039/501100011033 and “ERDF A way of making Europe”. Analyses were performed by the FinisTerrae III supercomputer at the Centro de Supercomputación de Galicia, Santiago de Compostela (CESGA), Spain.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2025.103275.

References

[1] M. de la Puente, C. Santos, M. Fondevila, L. Manzo, Á. Carracedo, M.V. Lareu, C. Phillips, The global AIMs nano set: A 31-plex SNaPshot assay of ancestry-informative SNPs, Forensic Sci. Int. Genet. 22 (2016) 81–88, https://doi.org/10.1016/j.fsigen.2016.01.015.

[2] M. Fondevila, C. Phillips, C. Santos, A. Freire Aradas, P.M. Vallone, J.M. Butler, M. V. Lareu, A. Carracedo, Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies, Forensic Sci. Int. Genet. 7 (2013) 63–74, https://doi.org/10.1016/j.fsigen.2012.06.007.

[3] R. Kosoy, R. Nassir, C. Tian, P.A. White, L.M. Butler, G. Silva, R. Kittles, M. E. Alarcon-Riquelme, P.K. Gregersen, J.W. Belmont, F.M. De La Vega, M.F. Seldin, Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America, Hum. Mutat. 30 (2009) 69–78, https://doi.org/10.1002/humu.20822.

[4] K.K. Kidd, W.C. Speed, A.J. Pakstis, M.R. Furtado, R. Fang, A. Madbouly, M. Maiers, M. Middha, F.R. Friedlaender, J.R. Kidd, Progress toward an efficient panel of SNPs for ancestry inference, Forensic Sci. Int. Genet. 10 (2014) 23–32, https://doi.org/10.1016/j.fsigen.2014.01.002.

[5] M. de la Puente, J. Ruiz-Ramírez, A. Ambroa-Conde, C. Xavier, J. Pardo-Seco, J. Álvarez-Dios, A. Freire-Aradas, A. Mosquera-Miguel, T.E. Gross, E.Y.Y. Cheung, W. Branicki, M. Nothnagel, W. Parson, P.M. Schneider, M. Kayser, Á. Carracedo, M. V. Lareu, C. Phillips, null On Behalf Of The Visage Consortium, Development and evaluation of the ancestry informative marker panel of the VISAGE basic tool, Genes 12 (2021) 1284, https://doi.org/10.3390/genes12081284.

[6] E.Y.Y. Cheung, M.E. Gahan, D. McNevin, Prediction of biogeographical ancestry in admixed individuals, Forensic Sci. Int. Genet. 36 (2018) 104–111, https://doi.org/10.1016/j.fsigen.2018.06.013.

[7] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C. D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724, https://doi.org/10.1002/elps.200500671.

[8] C. Phillips, J. Amigo, A.O. Tillmar, M.A. Peck, M. de la Puente, J. Ruiz-Ramírez, F. Bittner, Š. Idrizbegović, Y. Wang, T.J. Parsons, M.V. Lareu, A compilation of tri-allelic SNPs from 1000 genomes and use of the most polymorphic loci for a large-scale human identification panel, Forensic Sci. Int. Genet. 46 (2020) 102232, https://doi.org/10.1016/j.fsigen.2020.102232.

[9] J. Antunes, P. Walichiewicz, E. Forouzmand, R. Barta, M. Didier, Y. Han, J. C. Perez, J. Snedecor, C. Zlatkov, G. Padmabandu, L. Devesse, S. Radecke, C. L. Holt, S.A. Kumar, B. Budowle, K.M. Stephens, Developmental validation of the ForenSeq® Kintelligence kit, MiSeq FGx® sequencing system and ForenSeq Universal Analysis Software, Forensic Sci. Int.: Genet. 71 (2024), https://doi.org/10.1016/j.fsigen.2024.103055.

[10] A.J. Pakstis, W.C. Speed, R. Fang, F.C.L. Hyland, M.R. Furtado, J.R. Kidd, K. K. Kidd, SNPs for a universal individual identification panel, Hum. Genet. 127 (2010) 315–324, https://doi.org/10.1007/s00439-009-0771-1.

[11] A. Tillmar, K. Sturk-Andreaggi, J. Daniels-Higginbotham, J.T. Thomas, C. Marshall, The FORCE panel: an all-in-one SNP marker set for confirming investigative genetic genealogy leads and for general forensic applications, Genes 12 (2021) 1968, https://doi.org/10.3390/genes12121968.

[12] N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history, Genetics 192 (2012) 1065–1093, https://doi.org/10.1534/genetics.112.145037.

[13] M. Byrska-Bishop, U.S. Evani, X. Zhao, A.O. Basile, H.J. Abel, A.A. Regier, A. Corvelo, W.E. Clarke, R. Musunuri, K. Nagulapalli, S. Fairley, A. Runnels, L. Winterkorn, E. Lowy, E.E. Eichler, J.O. Korbel, C. Lee, T. Marschall, S.E. Devine, W.T. Harvey, W. Zhou, R.E. Mills, T. Rausch, S. Kumar, C. Alkan, F. Hormozdiari, Z. Chong, Y. Chen, X. Yang, J. Lin, M.B. Gerstein, Y. Kai, Q. Zhu, F. Yilmaz, C. Xiao, Paul Flicek, S. Germer, H. Brand, I.M. Hall, M.E. Talkowski, G. Narzisi, M.C. Zody, High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios, Cell 185 (2022) 3426–3440.e19, https://doi.org/10.1016/j.cell.2022.08.004.

[14] 1000 Genomes Project Consortium, A. Auton, L.D. Brooks, R.M. Durbin, E. P. Garrison, H.M. Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, G. R. Abecasis, A global reference for human genetic variation, Nature 526 (2015) 68–74, https://doi.org/10.1038/nature15393.

[15] C.C. Chang, C.C. Chow, L.C. Tellier, S. Vattikuti, S.M. Purcell, J.J. Lee, Second-generation PLINK: rising to the challenge of larger and richer datasets, Gigascience 4 (2015) 7, https://doi.org/10.1186/s13742-015-0047-8.

[16] D.H. Alexander, J. Novembre, K. Lange, Fast model-based estimation of ancestry in unrelated individuals, Genome Res 19 (2009) 1655–1664, https://doi.org/10.1101/gr.094052.109.

[17] Li, Yulong, StructureSelector: A Web Server to Select and Visualize Genetic Clusters Based on Multiple Methods, n.d. ⟨https://lmme.qdio.ac.cn/StructureSelector/index.html⟩.

[18] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2020. ⟨https://www.R-project.org/⟩.

[19] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag New York, 2016. ⟨https://ggplot2.tidyverse.org⟩.

[20] H. Wickham, R. François, L. Henry, K. Müller, D. Vaughan, dplyr: A Grammar of Data Manipulation, 2023. ⟨https://dplyr.tidyverse.org⟩.

[21] H. Wickham, D. Vaughan, M. Girlich, tidyr: Tidy Messy Data, 2024. ⟨https://tidyr.tidyverse.org⟩.

[22] N.A. Rosenberg, J.K. Pritchard, J.L. Weber, H.M. Cann, K.K. Kidd, L. A. Zhivotovsky, M.W. Feldman, Genetic structure of human populations, Science 298 (2002) 2381–2385, https://doi.org/10.1126/science.1078311.