Investigation into X-STR population data for forensic purposes in South Africa

Amy-Leigh Whittaker[1] and Laura Jane Heathfield[1]

[1]Division of Forensic Medicine and Toxicology, Department of Pathology, Faculty of Health Sciences, University of Cape Town, South Africa.
whtamy003@myuct.ac.za, laura.heathfield@uct.ac.za

Introduction

Short tandem repeats (STRs) are targeted in DNA profiling to achieve forensic human identification.[1] Allele frequencies from the general population are needed to calculate the probability of a DNA profile matching a random individual by chance.[1]

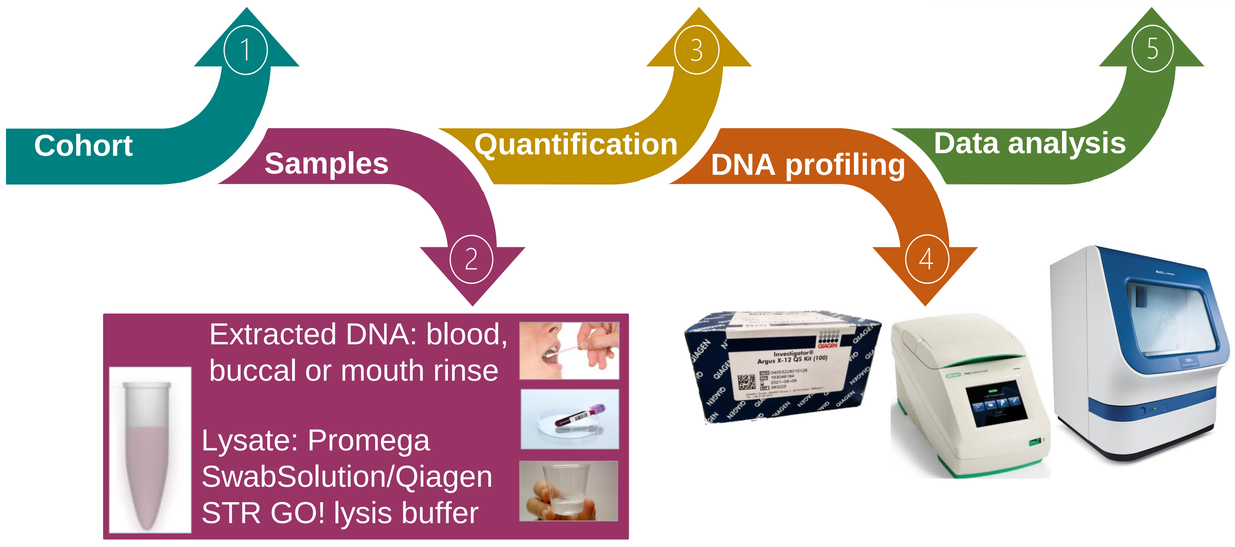
STRs located on the X-chromosome (X-STRs) are particularly useful in complex kinship and deficiency paternity testing.[2,3] However, X-STR allele frequency data is not yet available for the South African population.

AIM: Generate X-STR data for the South African population using the Investigator Argus X-12 QS kit (Qiagen, Hilden).

Materials and Methods

- A total of 781 unrelated South African individuals (collected from each of the four population groups and across all nine provinces)
- Ethics: HREC 136/2022

Data Analysis
- GeneMapper® *ID-X* Software Version 1.5
- StatsX v2.0
- Arlequin v3.5.2.2



Results and Discussion

- Samples from 264 female and 517 male South African individuals were processed using an optimised and internally validated workflow for the Investigator Argus X-12 QS kit (Qiagen, Hilden).
- Exact test performed per locus – no significant population differentiation (p > 0.05), allowing for male and female allele frequencies to be pooled for forensic parameter calculations.

Forensic parameters

Table 1: Forensic parameters for the 781 South African individuals.

Linkage group	Locus	Number of unique alleles	GD	PIC	PD_{Male}	PD_{Female}
1 (Xp22)	DXS8378	7	0.699	0.643	0.698	0.853
	DXS10135	53	0.955	0.952	0.954	0.996
	DXS10148	56	0.937	0.933	0.936	0.993
2 (Xp11)	DXS7132	9	0.751	0.709	0.750	0.896
	DXS10074	21	0.874	0.860	0.873	0.971
	DXS10079	14	0.832	0.810	0.831	0.950
3 (Xp26)	HPRTB	10	0.744	0.702	0.743	0.893
	DXS10101	27	0.917	0.910	0.916	0.987
	DXS10103	8	0.754	0.718	0.753	0.904
4 (Xp28)	DXS7423	7	0.681	0.623	0.681	0.841
	DXS10134	38	0.882	0.870	0.881	0.975
	DXS10146	41	0.917	0.910	0.916	0.987
Combined		-	-	-	0.9999	1

Abbreviations - GD: gene diversity, PIC: polymorphism information content, PD: power of discrimination

- DXS10135 was the most informative, while DXS7423 was the least informative. This is consistent with literature available for other populations. [4,5,6]

Shared haplotypes

- The haplotypes of 517 males were evaluated.
- No shared haplotypes within this South African cohort were observed for the full twelve loci haplotype, indicating diversity.
- Shared haplotypes were observed for the three-loci haplotype at each linkage group.

Table 1: Haplotype forensic parameters for 517 South African males.

	LG1	LG2	LG3	LG4	Combined
	DXS8378-DXS10135-DXS10148	DXS7132-DXS10074-DXS10079	HPRTB-DXS10101-DXS10103	DXS7423-DXS10134-DXS10146	
Number of unique haplotypes	421	228	218	302	-
HD	0.999	0.994	0.994	0.996	-
PIC	0.997	0.992	0.992	0.994	-
PD_Male	0.997	0.992	0.992	0.994	0.9999
PD_Female	1.000	1.000	1.000	1.000	1
MEC_Kruger	0.984	0.981	0.981	0.983	0.9999
MEC_Kishida	0.987	0.989	0.989	0.988	0.9999
MEC_Desmarais	0.997	0.992	0.992	0.994	0.9999
MEC_Desmarais_duo	0.994	0.984	0.984	0.989	0.9999

Abbreviations – LG: linkage group, GD: gene diversity, PIC: polymorphism information content, PD: power of discrimination , MEC: mean exclusion chance (Kruger, Kishida, Desmarais and Desmarais duo – variations of the MEC calculation)..

- LG1 was the most polymorphic, while LG2 and LG3 were the least polymorphic. This was consistent with other populations. [6-9]
- The South African population also showed an increased number of unique haplotypes at LG1 compared to other populations. [5-7,9,10]

Novel alleles

31 identified

- DXS10148 had the highest number of novel alleles (n = 12), followed DXS10134 (n = 6) and DXS10135 (n = 5).
- Novel allele observations were more prevalent in individuals with African ancestry.

Conclusions

- The X-STR population data generated for this South African cohort are highly polymorphic and informative.
- This study represents the largest X-STR allele frequency dataset in Sub-Saharan Africa.
- Overall, these results will facilitate complex kinship testing in the South African population, which is one step further in addressing the burden of unidentified human remains.

References

[1] J.M. Butler, Fundamentals of Forensic DNA Typing, Academic Press, USA, 2009.
[2] J. Edelmann, R. Lessig, M. Klintschar, R. Szibor, Advantages of X-chromosomal microsatellites in deficiency paternity testing: Presentation of cases, Int. Congr. Ser. 1261 (2004) 257–259. https://doi.org/10.1016/S0531-5131(03)01722-9.
[3] C. Bobillo, M. Marino, A. Sala, L. Gusmão, D. Corach, X-STRs: Relevance in complex kinship cases, Forensic Sci. Int. Genet. Suppl. Ser. 1 (2008) 496–498. https://doi.org/10.1016/j.fsigss.2007.10.080.
[4] J.M. Salvador, D.L.T. Apaga, F.C. Delfin, G.C. Calacal, S.E. Dennis, M.C.A. De Ungria, Filipino DNA variation at 12 X-chromosome short tandem repeat markers, Forensic Sci. Int. Genet. 36 (2018) e8–e12. https://doi.org/10.1016/j.fsigen.2018.06.008.
[5] R. Curtis, D. Ward, D. Taylor, J. Henry, Investigation of X-STR haplotype diversity in the Australian Aboriginal population, Aust. J. Forensic Sci. 00 (2022) 1–13. https://doi.org/10.1080/00450618.2022.2048690.
[6] F. Guo, Population genetic data for 12 X-STR loci in the Northern Han Chinese and StatsX package as tools for population statistics on X-STR, Forensic Sci. Int. Genet. 26 (2017) e1–e8. https://doi.org/10.1016/j.fsigen.2016.10.012.
[7] C. Bini, S. Sarno, E. Tangorra, A. Iuvaro, Y.G. Tseghereda, S. Pelotti, D. Luiselli, Haplotype data and forensic evaluation of 23 Y-STR and 12 X-STR loci in eight ethnic groups from Eritrea, Int. J. Legal Med. 135 (2021) 449–453.
[8] S. Elakkary, S. Hoffmeister-Ullerich, C. Schulze, E. Seif, A. Sheta, S. Hering, J. Edelmann, C. Augustin, Genetic polymorphisms of twelve X-STRs of the investigator Argus X-12 kit and additional six X-STR centromere region loci in an Egyptian population sample, Forensic Sci. Int. Genet. 11 (2014) 26–30. https://doi.org/10.1016/j.fsigen.2014.02.007.
[9] A.O. Tillmar, Population genetic analysis of 12 X-STRs in Swedish population, Forensic Sci. Int. Genet. 6 (2012) e80–e81. https://doi.org/10.1016/j.fsigen.2011.07.008.
[10] S.A. Messaoudi, S.R. Babu, A.B. Alsaleh, M. Albujja, N.R. Al-Snan, A.R. Chaudhary, A.C. Kassab, M. Assidi, Population genetic data for 12 X-STR loci in the Central Saudi region using investigator Argus X-12 amplification kit, Ann. Hum. Biol. 48 (2021) 321–326. https://doi.org/10.1080/03014460.2021.1957147.