

Article

Population data for 21 autosomal short tandem repeat markers in the Arabic population of the United Arab Emirates

Jones, Rebecca J, Al Tayaare, Wafa, Tay, Guan K, Alsafar, Habiba and Goodwin, William H

Available at <http://clock.uclan.ac.uk/17475/>

Jones, Rebecca J, Al Tayaare, Wafa, Tay, Guan K, Alsafar, Habiba and Goodwin, William H (2017) Population data for 21 autosomal short tandem repeat markers in the Arabic population of the United Arab Emirates. Forensic Science International: Genetics, 28 . e41-e42. ISSN 1872-4973

It is advisable to refer to the publisher's version if you intend to cite from the work.

<http://dx.doi.org/10.1016/j.fsigen.2017.02.015>

For more information about UCLan's research in this area go to <http://www.uclan.ac.uk/researchgroups/> and search for <name of research Group>.

For information about Research generally at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

Correspondence

Population data for 21 autosomal Short Tandem Repeat markers in the Arabic population of the United Arab Emirates.

Dear Editor,

The UAE lies on the east coast of the Arabian Peninsula with coastal borders of the Arabian Gulf and is neighbors with Saudi Arabia to the west and Oman to the south. The residents of the country are predominantly expatriates with only about 11.3% of approximately 9.6 million residents comprising UAE nationals (2015 census data). UAE nationals are predominantly people of Arab descent [1].

Throughout history, nomadic Arabian tribes traversed this region. The Middle East region is located at a junction where there were constant and regular human migration between the African, European and Asian continents [2,3] through the land bridge around modern Egypt. This land bridge dispersal from Africa into the Middle East was one of the two initial out of Africa migration routes. The general region around contemporary UAE, Qatar, Bahrain and Oman with borders only defined in the early 1970s, was an alternative migration route from the initial out of Africa dispersal involving migration from the Horn of Africa across the Red Sea and into Yemen [4]. Consanguineous marriages are common in societies throughout this region, potentially limiting the genetic pool. However a recent report in 2014 cited an increase in genetic diversity within the Middle East region arising from an increase in non-consanguineous marriages [5]. Studies related to genetic markers can be undertaken to appreciate the impact of cultural norms in a region that represents the hub through-which humans dispersed from Africa.

The genome era has and continues to reveal a rich vein of knowledge in disease-susceptibility and in overall genetic diversity of different populations. The focus of this study, in part, was to make observations relating to the genetic diversity within the UAE population. This study involved STR analysis of DNA samples from 477 unrelated individuals from the UAE. These individuals are Emiratis Arabs of mixed ethnic origin from the UAE. The results from these samples were then combined to a previous publication which used 519 DNA samples from unrelated Emirati individuals that are largely Emirati Bedouin [6]. It adds to the knowledge base arising from previous studies, which have also presented autosomal STR allelic data

on UAE populations [2,6,7]. With addition of new studies, such as this present study, the confidence in the information for applications in forensics improves with the iterative process and with increases in the population sample size being interrogated.

DNA samples from individuals who provided consent to have their de-identified DNA samples stored for research purposes were obtained from the Emirates Family Registry (EFR) [8]. Prior to commencement of this study, approval to undertake the work was sort from the Ethics committee of the Ministry of Health of the United Arab Emirates. Ethics Committees at each author's home institution also approved the study.

Buccal cells were collected with saliva using the Genotek's Oragene-DNA kit (Genotek, Ottawa, Canada) and DNA extracted from buccal cells using the prepIT-L2P system (Genotek) in accordance with manufacturer's instructions. The NanoDrop spectrophotometer (Thermo Scientific, Wilmington DE, USA) was used to determine the quantity and quality of the extracted DNA. The GlobalFiler® PCR Express Amplification Kit (Life Technologies, Carlsbad, CA, USA) was used to amplify 24 STR markers using half volume (total 7.6 µl). The 21 autosomal STR presented in this analysis were D3S1358, vWA, D16S539, CSF1PO, TPOX, D8S1179, D21S11, D18S51, D2S441, D19S433, TH01, FGA, D22S1045, D5S818, D13S317, D7S820, SE33, D10S1248, D1S1656, D12S291 and D2S1338. The PCR was performed using the GeneAmp® PCR System 9700 (Life Technologies). The PCR products were added to a 500 LIZ-Internal Standard (Life Technologies) and analysed using an 8 capillary ABI 3500 DNA Genetic Analyser with POP-4™ polymer (Life Technologies). GeneMapper® Software ID-X version 4.0 (Life Technologies) was used for analysis. The alleles from all 21 loci reported within this present study were labelled according to the published nomenclatures and the guidelines for performing STR analyses of the International Society for Forensic Genetics (ISFG) [9].

Locus-by-locus exact tests using Arlequin v3.5.2.1 Software were carried out in this study [10]. Allele frequencies for the 21 autosomal STR markers were calculated using GeneALEx v6.5 [11,12]. Genetic parameters; such as the Power of Discrimination (PD), Power of Exclusion (PE), Match Probability (MP), typical

Paternity Index (PI) and Polymorphism Information Content (PIC); were also calculated using GeneAIEx v6.5 [11,12]. The Observed and Expected Heterozygosity values (H_o and H_e respectively) and Hardy-Weinberg Equilibrium (HWE) were calculated using an exact test through Arlequin v3.5.2.1 Software [10].

Before the two separate datasets were combined for allele frequency analysis, a locus-by-locus exact test was carried out to determine whether there were any significant variations between them (Supplementary Table 1). Three out of the 21 autosomal STRs (TPOX, FGA and D22S1045) resulted with significant differences between the two UAE population datasets ($p\text{-value} \leq 0.05$). When a hypothesis test is run and a significance threshold of 0.05 is chosen, we are essentially accepting a false positive rate of 1 in 20. In other words, for a 5% significance threshold there is a 1 in 20 chance that the result is in fact false positive. Thus, when there are multiple tests carried out in the same sample, the number of potential false positive results go up. For example, if we ran 20 tests, 1 of the results is likely to be a false positive. To correct for the impact of such 'multiple testing', we divide the initial threshold of 5% by the number of tests carried out, known as the Bonferroni correction. This study carried out 21 tests, resulting in the new significance threshold to be $0.05 \div 21$ or 0.002. After applying the Bonferroni correction, no significant deviations were observed. Furthermore, when the overall locus-by-locus AMOVA test was carried out using Arlequin, the resulting fixation index (F_{ST}) of 0.00005 showed no significant variation between the two separate datasets. Consequently, the two datasets were combined for analysis.

The calculated allele frequencies and genetic parameters for the 21 autosomal STR markers were presented (Supplementary Table 2). The STR marker with the largest allele polymorphism was SE33 (59 alleles). The least polymorphic was marker D16S539 (8 alleles). Eight of the STR markers deviated from the HWE (D21S11, D18S51, D19S433, TH01, D22S1045, SE33, D10S1248 and D2S1338). After applying the Bonferroni correction, only three markers deviated from this HWE (D19S433, SE33 and D2S1338). The H_o values ranged from 0.687 (TPOX) to 0.917 (SE33). The PD range was 0.687 (TPOX) to 0.917 (SE33) with a combined power of discrimination of 0.9999999999. The PD in correlation with MP supports the high degree of polymorphism between UAE individuals. The PE range was 0.408 (TPOX) to 0.830 (SE33) with a combined power of exclusion of 0.9999999964. The wide

range for the PE can be expected as these values do vary across individual cases [13]. The combined power of exclusion indicates that there is a higher fraction of the individuals with allele variations, highlighting genetic diversity of UAE individuals. The PIC range was 0.653 (TPOX) to 0.945 (SE33). These high informative values support the heterozygosity values indicating the high degree of genetic polymorphism. Each marker had a typical PI value larger than 1.0, which is indicative of relativeness and usefulness for paternity testing applications [13].

The data presented indicate that the 21 autosomal STR markers from the GlobalFiler® amplification kits have forensic applications for individual identification and paternity testing in the local population of the UAE. The similarity between the two datasets provides a degree of reassurance that potential errors and biases are minimal. Furthermore, consolidation of the data provides representation of the distribution of various ethnic groups that make up nationals in the UAE. As further studies of Arab populations in the region become available, it may be possible to develop a greater understanding of the relationships between the different jurisdictions on the Arabian Peninsula.

This manuscript of population data follows the guidelines for publication described [14,15].

Data is available upon request from contacting the authors.

Conflict of Interest

The authors have declared no conflict of interest.

Acknowledgements

We acknowledge Khalifa University, Abu Dhabi for hosting the project and the Abu Dhabi Police General Head Quarter for sponsoring the studies of Wafa Al Tayyare through a scholarship. Rebecca Jones is a Master by Research student who acknowledges the postgraduate travel award provided by University of Western Australia to work at Khalifa University. We further thank Dr Kanishka Bhattacharya for reviewing the statistical data.

References

- [1] A.E. Osman, H. Alsafar, G.K. Tay, J. Theyab, M. Mubasher, N. Sheikh, et al, Autosomal short tandem repeat (STR) variation based on 15 loci in a population from the Central Region (Riyadh Province) of Saudi Arabia, *J Forensic Res.* 6 (1) (2015) 1-5.
- [2] R. Garcia-Bertrand, T.M. Simms, A.M. Cadenas, R.J. Herrera, United Arab Emirates: Phylogenetic relationships and ancestral populations, *Gene.* 533 (1) (2014) 411-9.
- [3] S. Kundu, S.K. Ghosh, Trend of different molecular markers in the last decades for studying human migrations, *Gene.* 556 (2) (2015) 81-90.
- [4] A. Beyin, The Bab al Mandab vs the Nile-Levant: An appraisal of the two dispersal routes for early modern humans out of Africa, *Afr Archaeol Rev.* 23 (1-2) (2006) 5-30.
- [5] G.O. Tadmouri, K.S. Sastry, L. Chouchane, Arab gene geography: From population diversities to personalized medical genomics, *Glob Cardiol Sci Pract.* 2014 (4) (2014) 394-408.
- [6] O. Ali Alhmoudi, R.J. Jones, G.K. Tay, H. Alsafar, S. Hadi, Population genetics data for 21 autosomal STR loci for United Arab Emirates (UAE) population using next generation multiplex STR kit, *Forensic Sci Int.* 19 (2015) 190-1.
- [7] F. Alshamali, A.Q. Alkhayat, B. Budowle, N.D. Watson, STR population diversity in nine ethnic populations living in Dubai, *Forensic Sci Int.* 152 (2-3) (2005) 267-79.
- [8] H. Alsafar, K.A. Jama-Alol, A.A.K. Hassoun, G.K. Tay, The prevalence of Type 2 Diabetes Mellitus in the United Arab Emirates: Justification for the establishment of the Emirates Family Registry, *Int J Diabetes Developing Ctries.* 32 (1) (2012) 25-32.
- [9] P.M. Schneider, Scientific standards for studies in forensic genetics, *Forensic Sci Int.* 165 (2-3) (2007) 238-43.
- [10] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): An integrated software package for population genetics data analysis, *Evol Bioinformatics Online.* 1 (2005) 47-50.
- [11] R. Peakall, P.E. Smouse, GenAlEx 6.5: Genetic analysis in Exce: Population genetic software for teaching and research-an update, *Bioinformatics.* 28 (19) (2012) 2537-9.

- [12] R.O.D. Peakall, P.E. Smouse, GenAlEx 6: Genetic analysis in Excel: Population genetic software for teaching and research, *Mol Ecol Notes*. 6 (1) (2006) 288-95.
- [13] K. Bentayebi, F. Abada, H. Izhmad, S. Amzazi, Genetic ancestry of a Moroccan population as inferred from autosomal STRs, *Meta Gene*. 2 (2014) 427-38.
- [14] P. Lincoln, A. Carracedo, Publication of population data of human polymorphisms, *Forensic Sci Int*. 110 (2000) 3-5.
- [15] Á. Carracedo, J.M. Butler, L. Gusmão, A. Linacre, W. Parson, L. Roewer, et al, New guidelines for the publication of genetic population data, *Forensic Sci Int: Genetics*. 7 (2) (2013) 217-20.