



A 1000 Arab genome project to study the Emirati population

Mariam Al-Ali^{1,2} · Wael Osman¹ · Guan K. Tay^{1,3,4} · Habiba S. AlSafar^{1,2}

Received: 11 June 2017 / Revised: 21 November 2017 / Accepted: 27 November 2017 / Published online: 6 February 2018
© The Author(s) 2018. This article is published with open access

Abstract

Discoveries from the human genome, HapMap, and 1000 genome projects have collectively contributed toward the creation of a catalog of human genetic variations that has improved our understanding of human diversity. Despite the collegial nature of many of these genome study consortiums, which has led to the cataloging of genetic variations of different ethnic groups from around the world, genome data on the Arab population remains overwhelmingly underrepresented. The National Arab Genome project in the United Arab Emirates (UAE) aims to address this deficiency by using Next Generation Sequencing (NGS) technology to provide data to improve our understanding of the Arab genome and catalog variants that are unique to the Arab population of the UAE. The project was conceived to shed light on the similarities and differences between the Arab genome and those of the other ethnic groups.

Introduction

Biological evolution has been defined as “biological change through or over time” [1]. It is a continuous process, which has given rise to the contemporary populations. These have evolved from ancestral populations through adaptation and selection mechanisms [1], in which organisms with traits that confer a competitive advantage will survive and their offspring will perpetuate. Therefore, the variations observed within contemporary human populations and between different ethnic subpopulations are the consequences of this process of evolution. Each ethnic subpopulation or group can be characterized by specific morphological, anatomical, and physiological characteristics (i.e., the phenotype) that are encoded within the genome (i.e., the genotype) of the subpopulation. These genetic polymorphisms, if advantageous, are transmitted from one generation to the next [2].

As many diseases in humans involve apparently aberrant physiological processes, there is an underlying genetic basis for most diseases. Although many inherited disorders are regarded as defects in contemporary populations, the selection for these phenotypes in ancestral populations was to confer specific selective advantages. For example, the iron overload disease, hereditary hemochromatosis, is believed to have conferred a selective advantage during times of famine. As the relative content of dietary iron has increased over history, this excess now causes cirrhosis in patients who are homozygote for a series of variants in the HFE gene on chromosome 6, in particular the C282Y variant [3].

The first human genome project was completed by a collaborative international effort [4]. The subsequent 1000 genome project was an extension of the International HapMap project that aimed to sequence the whole genome of 1000 individuals from various populations [5]. The principle objective was to provide a comprehensive description of genetic variations in humans and their distribution throughout the genome and to make this data publically available. In 2016, Popejoy and Fullerton reported on the status of genome data available in the public domain, after completing an extensive audit of databases and discussed how the Arab Genome was sadly underrepresented [6]. At 0.08% of the total amount of data, it is only better than native populations [6]. This has to change in order for individuals of Arabian ancestry to benefit from the new paradigm that is precision medicine and associated improvements in standards in healthcare.

✉ Habiba S. AlSafar
habiba.alsafar@kustar.ac.ae

¹ Center of Biotechnology, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

² Department of Biomedical Engineering, Khalifa University, Abu Dhabi, United Arab Emirates

³ School of Psychiatry and Clinical Neurosciences, The University of Western Australia, Perth, Australia

⁴ School of Health and Medical Sciences, Edith Cowan University, Perth, Australia

Arab effort in sequencing the human genome

During the fossil fuel era, a number of countries in the Gulf Cooperative Council (GCC) have thrived and have built competencies and proficiencies around the financial sector, telecommunication, and aviation. These countries have in recent years looked to innovate in other technology sectors, in anticipation that oil reserves will run out and as the world energy needs shift to renewable sources. As these nations seek to diversify their economics and its base of revenues, health projects are becoming more important.

Despite the genome effort in this part of the world being largely ignored until recently, it is slowly and most certainly gaining momentum. A number of governments in the region have initiated projects. National Genome efforts have commenced in Saudi Arabia [7], Qatar, and Kuwait [8]. These projects have mainly focused on understanding the unique genetic makeup of the citizens of these respective nations. Furthermore, the studies were designed to primarily identify or verify specific genetic variants of disease that are common in their populations. For example, the study of native population of Kuwait was developed to identify deleterious polymorphisms associated with obesity and asthma [9].

Two conflicting points of view necessitate genome studies of the indigenous populations and ethnic subpopulations from Arab nations. Firstly, it has been reported that these populations are relatively homogenous groups with a highly conserved gene pool [8]. This homogeneity has presumably arisen from a number of social factors, including the high rates of consanguineous marriages (especially between first cousins) [8], endogamous unions [8], a tribal structure of the society, and the large family sizes. Contrasting this view is the suggestion that the Arab people are more varied than thought [10]. The present population residing in the Middle East region has arisen around intercontinental migration between Africa, Asia and Europe, being at the crossroads of these three continents. Migration of Semitic tribes throughout the region and interaction with traders from Far East along silk road and immigrants shunting in and out of the African continent [11] have contributed to the diversity in the region. Other factors contributing to the increase in the contemporary diversity of the Arab population include the expansion and spread of Islam between the 7th and 14th centuries, the gradual transfer of power from the Ottoman Dynasty to European rule, the Crusade wars, and more recently, migration patterns facilitated by mass air travel [12]. The history of the region suggests close social interactions between different ethnic groups.

These conflicting views provide a compelling reason to study the genomic structures of Arab populations. It is

important to resolve this matter and through gathering data on the Arab genome, it will be possible to establish the relationship between Arabs and other populations.

The lack of Arab genome representation

Although the various ethnic groups of the Arab world (e.g., Arabs, Persians, Armenians, Assyrians, Bakhtiyarians, the Baluch, Beja, Berber, Copts, Gilaki, Jews, Kurds, Lurs, Mazandarani Nubians, Talyshs, and Turks) share some characteristics, they have arisen from distinctive demographical groups and geographical locations. Therefore, the genetic makeup of each population is potentially very different. An example of this difference is evident from the diversity found in conclusions drawn from a series of Genome Wide Association Studies (GWAS). For example, rs5219 SNP of the *KCNJ11* gene is known to be associated with the incidence of Type 2 Diabetes. Studies conducted on the populations from Oman [13] and Saudi Arabia [14] showed a positive association with this variant, while no or weak associations were found in the Arab populations of Tunisia [15] and Morocco [16]. To understand the significance of these differences, genome studies for each specific ethnic group in the region are needed.

Unfortunately, despite the significant efforts put into compiling genome data of ethnic groups from around the globe, information on the genome of populations from the nations of the Middle East, remains underrepresented. There is an increasing evidence suggesting that the genome structure of the individuals of Arabian descent is different to those studied to date [12]. For instance, in its December 9, 2016 edition, the Science Journal reported on the effort of Saudi Human Genome Project, which in the preceding 5 years reported nearly 200 genes that contribute to human pathologies within their population [17]. They further pointed out that the rate of inherited genetic diseases is nearly double the rate in Europe and the United States and was 10 times higher for certain disorders [17]. In addition, despite the relatively similar genetic structure between the different Arab groups, differences can be found. Another study, by Garcia-Bertrand et al. (2014), focused on admixture analysis to establish the phylogenetic relationships and ancestral populations of different Arab groups, showing the uniqueness of different Middle Eastern populations, especially those in the Gulf region [18].

The National Arab Genome project in the UAE

In spite of these genome efforts in the Arabian regions, there remains a poor level of data. As previously mentioned, only 0.08% of the information in the public domain is of Arabian origin [6]. Consequently, the National Arab genome project planned for the United Arab Emirates (UAE)

will specifically focus on the ethnic groups of the gulf region by sampling Emirati citizens. The project is intended to create reference sequences to establish databases for developing customized molecular diagnostic assays and personalized medicine strategies, for individuals of Arabian descent. From a pharmacogenomics standpoint, this should in turn maximize drug efficiency or provide an optimal dose to the patient.

The contemporary population structure of the UAE is diverse; a result of a high percentage of expatriates (80–90%, source: CIA World Fact Book) residing in the country, juxtaposed next to a relatively small but diverse indigenous population admixed with immigrants from Yemen, Oman, North Africa, Iran, Baluchistan, India, and other neighboring regions [19]. Of the approximately 11 million residents in the UAE, around 10% are Emirati citizens. The proposed study aims to sequence the genome of Emirati citizens using next-generation sequencing technology and analyze the data using advanced bioinformatics tools. The project is a collaborative effort between academia (Khalifa University of Science and Technology) and clinicians in local hospitals. The medical and health records of the participants will be stored on a secure database. The project was intended to start as a population based study, shifting to a healthcare or disease association study to identify predisposing pathogenic variants.

The expected outcomes of this study include:

1. The creation of an Emirati genome reference that identifies point and segmental (indel) polymorphisms. These polymorphisms will be examined to assess possible associations with disease that are common in Arabian populations (e.g., diabetes, cardiovascular disease, etc.). Variations in the gene copy number will also be examined.
2. The identification of novel and rare or low-frequency genetic variations (defined as those with minor allele frequencies of <1% and 1–5%, respectively) in the indigenous Emirati population. These variations will be compared with published findings. Comparative analyses may provide clues to the underlying factors behind the molecular associations and cellular mechanisms of the genetic disorders within the particular ethnic groups of Arabian descent in the UAE.
3. The identification of unique associations is expected to give rise to customized DNA-based assays to screen for the genetic polymorphisms that may assist in the early detection of disease for intervention and for improved diagnostic applications such as histocompatibility matching for transplantation.
4. The data collected will also provide opportunities to study the UAE population in detail. The data could

potentially be used as a genome panel that will serve as a reference for future imputation studies.

The project was conceived to provide genome data for future healthcare applications in the UAE. The deliverables of this project will influence the future medical practice in the UAE, which is expected to include a focus on correlating phenotype and genotype for many complex diseases that afflict the local population. Optimistically, future collaborations and training opportunities will take place between the UAE and other efforts in the Arabia, as well as with the international research community. Improvement in the quality and quantity of genome data of Arab origin is also expected to improve our understanding of relationships between the genotype and phenotype in other ethnic groups around the world, since diversity will provide opportunities to contrast.

The data generated from the UAE National Genome Project is expected to present clinicians, specifically in the UAE, and broadly across the Middle East region with detailed information that is relevant to the health status of their people. Consequently, improved diagnosis can be expected, and more informed decisions could be made. This project will provide the opportunity to improve our understanding of a genetics including more in-depth appreciation of pharmacogenomic factors that impact on UAE and Arab patients. A patient's genetic makeup will be used to determine the optimal doses of treatments for use and the therapeutic response to certain drugs in the presence of the appropriate candidate loci [20]. By assuring efficient drug intake, it is possible to reduce or completely eliminate adverse side effects [20] and eliminate wastage, thus reducing the cost of healthcare in the United Arab Emirates.

Despite being largely ignored to date, the rise of Next-Generation Sequencing technologies has provided an opportunity for smaller nations to join in the genome effort. Cost and time are not prohibitive factors as they were in the past, and the National UAE Genome project is poised to deliver information not only for the local effort but to the international community.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, and provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's

Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

References

- Rice PC, Moloney N. Biological anthropology and prehistory: exploring our human ancestry, 2nd edn, Pearson/Allyn & Bacon (Boston, New York, London, 2008).
- Verhoeven KJ, Macel M, Wolfe LM, Biere A. Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proc R Soc Lond B*. 2011;278:2–8.
- Feder J, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy D, Basava A, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet*. 1996;13:399–408.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;291:1304–51.
- 1000 Genomes Project Consortium, et al. A global reference for human genetic variation. *Nature*. 2015;526:68.
- Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538:161.
- Saudi Mendeliome Group. Comprehensive gene panels provide advantages over clinical exome sequencing for Mendelian diseases. *Genome Biol*. 2015;16:1–14.
- Zayed H. The Arab genome: health and wealth. *Gene*. 2016;592:239–43.
- John SE, Thareja G, Hebbar P, Behbehani K, Thanaraj TA, Alsmadi O. Kuwaiti population subgroup of nomadic Bedouin ancestry—whole genome sequence and analysis. *Genom Data*. 2015;3:116–27.
- Teebi AS, Teebi SA. Genetic diversity among the Arabs. *Community Genet*. 2005;8:21–26.
- Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, et al. Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. *Am J Hum Genet*. (ed. Ahmad S. Teebi) 3–34 (Springer, Berlin, Heidelberg, 2010).
- Teebi AS. In: Genetic disorders among Arab populations (ed. Ahmad S. Teebi). Springer, Berlin, Heidelberg; 2010. p. 3–34.
- Al-Sinani S, Woodhouse N, Al-Mamari A, Al-Shafie O, Al-Shafae M, Al-Yahyaee S, et al. Association of gene variants with susceptibility to type 2 diabetes among Omanis. *World J Diabetes* 2015;6:358.
- Alsmadi O, Al-Rubeaan K, Wakil SM, Imtiaz F, Mohamed G, Al-Saud H, et al. Genetic study of Saudi diabetes (GSSD): significant association of the KCNJ11 E23K polymorphism with type 2 diabetes. *Diabetes Metab Res Rev*. 2008;24:137–40.
- Ezzidi I, Mtraoui N, Cauchi S, Vaillant E, Dechaume A, Chaieb M, et al. Contribution of type 2 diabetes associated loci in the Arabic population from Tunisia: a case-control study. *BMC Med Genet*. 2009;10:33.
- Benrahma H, Charoute H, Lasram K, Boulouiz R, Atig RK-B, Fakiri M, et al. Association analysis of IGF2BP2, KCNJ11, and CDKAL1 polymorphisms with type 2 diabetes mellitus in a Moroccan population: a case-control study and meta-analysis. *Biochem Genet*. 2014;52:430–42.
- Kaiser J. Saudi gene hunters comb country's DNA to prevent rare diseases. <http://www.sciencemag.org/news/2016/12/saudi-gene-hunters-comb-countrys-dna-prevent-rare-diseases>.
- Garcia-Bertrand R, Simms TM, Cadenas AM, Herrera RJ, United Arab Emirates: phylogenetic relationships and ancestral populations. *Gene*. 2014;533:411–9.
- Al-Gazali L, Ali BR. In: Genetic disorders in the United Arab Emirates (eds Al-Gazali. L & Ali. B.R.) 639–76 (Springer, Berlin, Heidelberg).
- Zhang W, Dolan ME, Impact of the 1000 genomes project on the next wave of pharmacogenomic discovery. *Pharmacogenomics*. 2010;11:249–56.