Gene 592 (2016) 239–243

Contents lists available at ScienceDirect

Gene

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

Review

The Arab genome: Health and wealth

Hatem Zayed

College of Health and Sciences, Biomedical Sciences Department, Qatar University, PO Box 2713, Doha, Qatar

ARTICLE INFO

Article history:
Received 21 June 2016
Accepted 3 July 2016
Available online 5 July 2016

Keywords:
Arab countries
Human genome sequencing
Whole exome sequencing
Consanguinity
Endogamous marriage
Novel genes
Novel variants

ABSTRACT

The 22 Arab nations have a unique genetic structure, which reflects both conserved and diverse gene pools due to the prevalent endogamous and consanguineous marriage culture and the long history of admixture among different ethnic subcultures descended from the Asian, European, and African continents. Human genome sequencing has enabled large-scale genomic studies of different populations and has become a powerful tool for studying disease predictions and diagnosis. Despite the importance of the Arab genome for better understanding the dynamics of the human genome, discovering rare genetic variations, and studying early human migration out of Africa, it is poorly represented in human genome databases, such as HapMap and the 1000 Genomes Project. In this review, I demonstrate the significance of sequencing the Arab genome and setting an Arab genome reference(s) for better understanding the molecular pathogenesis of genetic diseases, discovering novel/rare variants, and identifying a meaningful genotype-phenotype correlation for complex diseases.

Contents

1. Introduction

The completion of the Human Genome Project (HGP) in April 2003 provided a wealth of information to scientists and clinicians. Subsequently, the world has witnessed rapid evolution in the field of human genetics and genomics (Lander et al., 2001; Venter et al., 2001). Initially, the focus of the HGP was to catalog the protein-expressing genes, which are now estimated to include approximately 20,000 to 25,000 coding genes (International Human Genome Sequencing Consortium, 2004). However, the hard work of decoding the function of many genes and their precise genotype-phenotype correlation in disease development remains.

From the publication of the first draft of the human genome, there has been fierce competition to develop sequencing technologies that are faster, more efficient and cheaper and to make the price of human genome sequencing more affordable. Thus far, whole genome/exome sequencing has provided outstanding insights into the frequency and incidence of novel variants in the human genome that are associated with disease phenotypes. This information provides opportunities to different populations in the world to be able to map the sequence variants that might be unique to their own individuals and that might be responsible for genetic disorders in their specific populations. For this purpose, the HapMap (human haplotype mapping) Project was

E-mail address: hatem.zayed@qu.edu.qa.

http://dx.doi.org/10.1016/j.gene.2016.07.007
0378-1119/

launched in 2002 (International HapMap Consortium, 2003); this project has identified a considerable number of genetic variants, providing extensive catalogs for genetic variation. The HapMap Project has also served as the basis for genome-wide association studies (GWAS). In particular, the HapMap Project has contributed to the successful mapping of more than 100 genomic regions that are associated with genetic diseases (International HapMap Consortium, 2003).

As an extension of the HapMap Project, the 1000 Genomes Project was launched in 2008 through international concerted efforts (Buchanan et al., 2012). This project aims to sequence the whole genomes of 1000 unidentified individuals from Europe, America, Africa, and Asia, and will add information to the single-nucleotide polymorphism (SNP) database already cataloged by the HapMap Project and provide a rich resource for both SNPs and structural variant haplotypes. Although this information will allow researchers to learn more about many genetic variants and genetic diseases, unfortunately, the Arab genome is greatly under-represented in the international efforts of such genomic studies; specifically, it is not included in the HGP, HapMap Project, or 1000 Genomes Project. There is no doubt that the importance of the Arab genome sequencing is significant and that this genome thus should not be omitted from the diverse collections of genomes that have already been sequenced. Therefore, I am focusing this review on elaborating upon the importance of the Arab genome and the potential contribution of the Arab genome to the genomic sciences.

2. The Arab world

The Arab world includes 22 Arabic-speaking countries (Fig. 1). According to the World Bank latest classification for 2015 (http://data.worldbank.org), the Arab countries include high-income countries (HICs) such as Bahrain, Kuwait, Oman, Saudi Arabia, Qatar, and the United Arab Emirates; middle-income countries (MICs) such as Algeria, Egypt, Iraq, Jordan, Lebanon, Libya, Morocco, Palestine, Sudan, Syria, and Tunisia; and low-income countries (LICs) such as Comoros, Djibouti, Mauritania, Somalia, and Yemen. These countries occupy a large area that extends from the Atlantic Ocean in the west to the Arabian Sea in the east, and the Arab population is approaching 0.5 billion. This region has been extensively exposed to many successive invaders from Turkey, Rome, and Europe as well as to traders and immigrants, thus contributing to mixing of the ethnic demographics of the population. However, the HICs, which include countries with the highest Gross Domestic Product (GDP) per capita worldwide (http://data.worldbank.org), spend less than 0.2% of their GDP on scientific development (Giles, 2006). This phenomenon has led to the immigration of many Arab scientists into the West to look for better opportunities. However, recently, biomedical disease-based research has received special attention from Arab governments, with the aim of improving the understanding and treatment of common diseases afflicting the Arab population. Various attempts have been made by Saudi Arabia and Qatar in particular to establish a research infrastructure, but the progress has been significantly slow relative to the amount of capital infused into such programs, and the benefits of such investments might take significant time to yield results. In this manuscript I will refer to the "Arab genome" as the genome of the 22 Arab countries.

2.1. Inbred Arab communities and rare variants discovery

There are 955 genetic diseases that have been identified in Arabs, of which 586 (60%) are reported to be recessive diseases (http://www.cags.org.ae). Arabs have one of the highest rates of consanguineous marriage worldwide, reaching up to ~70%, with an extreme prevalence of first-cousin marriage (Tadmouri et al., 2009), These factors, together with the endogamous marriage culture and large family sizes, are responsible for the spread of genetic diseases in Arab countries, with a high prevalence of rare diseases (Teebi and Teebi, 2005). Endogamous marriages approach 100% in many Arab countries, and especially the Gulf States (i.e., Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and the United Arab Emirates). For example, women in Saudi Arabia are prohibited from marrying men other than Arab men from the Gulf countries without special dispensation from the king (http://web.

(Source: http://www.arabic-keyboard.org/arabic).

Fig. 1. Arabic speaking countries according to the latest WHO classification. (Source: http://www.arabic-keyboard.org/arabic).