Population Genetics of Arabs, a Primer

Abdul Rezzak Hamzeh
Centre for Arab Genomic Studies, Dubai, UAE

نبذة مختصرة

يمثل العرب مجموعة من الشعوب ذات أصول عرقية متعددة، ينتشر سكانها على أراض واسعة الامتداد، تغطي ثلاثاً وعشرين دولة. يجمع ما بين هذه الشعوب لغة وتاريخ مشتركين. وقد كانت المنطقة التي يغطيها العالم العربي بشكله الحالي محط لأحداث تاريخية مهمة ومعبر لمختلف المجتمعات البشرية عبر الزمن القديم والجديد. وقد أدت هذه الأحداث إلى وجود تنوع عرقي بين السكان العرب، كما ساهمت في اكتساب العرب خليطاً متنوعاً من الجينات البشرية عبر الزمن. في هذا الفصل، نلخص بعض المفاهيم الهامة لعلم الوراثة السكانية، مثل الأنواع المختلفة من الاختلافات الجينية والعمليات الجزيئية التي تقودها، علماً أن هذه المفاهيم تنطبق على العالم العربي والسكان في جميع أنحاء العالم.

Abstract

The Arab population is a panethnic group, residing over a massive stretch of land spanning 23 countries, and whose identity is mainly defined by a common language and history. The region that corresponds to what is now known as the Arab world covers a set of highly interesting sites for old and new migrational events in the history of humanity. These events have led to the presence of the current high level of ethnic and religious diversity in many parts of the Arab world and have shaped the Arab population and contributed enormously to its genetic makeup. In this chapter, we recapitulate a few important concepts of population genetics, such as the various types of genetic variations and the molecular processes driving them, which apply to the Arab world as well as worldwide populations.

Historical Perspective

Arab identity is mainly defined by a common language and history. The region that corresponds to what is now known as the Arab world covers a set of highly interesting sites for old and new migrational events in the history of humanity. It certainly accommodates the two probable routes for humans who moved out of Africa; the Levant and Bab el-Mandab Strait. During the following centuries, the region seemed to undergo more than its fair share of major turmoils and pandemonia. These conditions accompanied the often violent introduction of various outsider population groups to the region. The latter dynamics probably underlie the high genetic diversity in Arab populations, even in areas which are regarded as isolated and homogenous (1).

In the relatively recent stages of history, the region witnessed many major human movements, including an outward migration of Arabian people towards Mesopotamia around 3500 BC. The latter movement is thought to be the start of the Assyrian/Babylonian civilization. Later migrations into the Levant produced the civilization of Canaan. Another major wave of migration took place with the spread of Islam during the 7th century, outside the confines of the Arabian Peninsula. With the firm establishment of the Islamic/Arabic civilization over vast stretches of land from western China to the Atlantic Ocean, people belonging to a myriad ethnicities mingled together, peacefully and otherwise (2). In fact, the Arab world was at the receiving end of many invasions from the Mongols, the Crusaders, and the Ottomans and each of these was accompanied with significant population admixture.

This long and complicated history explains the current high level of ethnic and religious diversity in many parts of the Arab world. There are various ethnic groups that inhabit the region, such as Kurds, Berber, Armenians and Circassians. Moreover,

many Arabic-speaking people residing in various Arab countries trace their ethnic origins to regions outside the Arab world, including modern-day Iran, Pakistan, Afghanistan, sub-Saharan Africa and the Caucasus (3,4).

As mentioned above, from day-to-day cultural and commercial exchange between the population in the Peninsula and neighboring regions, to major wars and invasions, layers upon layers of complex admixture resulted in what is now called Arab populations. One very significant unifying factor for these populations is the Arabic language; this interesting anthropological link apparently brings together a huge number of people from different ethnicities who reside over a massive stretch of land. With an area exceeding 13 million km², it ranks second after Russia in total land area. In relevant historical discourse, it is implied that Arabs originated from tribal inhabitants of the Arabian Peninsula and the Syrian Desert. It is not surprising, therefore, that throughout this vast area, indigenous Arabs had mixed with other local ethnicities to produce a multiethnic and multicultural group of people. This admixture has contributed enormously to the modern Arab population's genetic makeup and is the reason behind calling Arabs a panethnic group. It is important to remember that being Arab is not a nationality in the strict sense of the word, as an Arab person will belong to one of the following nationalities: Moroccan, Mauritanian, Eritrean, Algerian, Tunisian, Libyan, Egyptian, Sudanese, Somali, Djiboutian, Saudi, Yemeni, Omani, Emirati, Qatari, Bahraini, Kuwaiti, Palestinian, Lebanese, Syrian, Jordanian, Iraqi and Comorian.

In what follows, we recapitulate a few important concepts of population genetics, which apply to the Arab world as well as worldwide populations.

Genetic and Phenotypic Variation between Population Groups.

A considerable degree of variability exists between different population groups. Such variability affects both qualitative and quantitative traits. The latter are commonly controlled by environmental as well as genetic factors. The term heritability is used to quantify the genetic contribution to total phenotypic variability of a certain trait. In numerous qualitative traits, the genetic component is almost the sole determinant of the phenotype and variability can be attributed directly to changes at the DNA level. However, it is important to remember that the relationship between genetic changes and the corresponding phenotypes is extremely complicated. The phenomenon of incomplete penetrance sums up this complex relationship very well, and large-scale sequencing and genotyping studies of apparently healthy individuals suggest that incomplete penetrance is highly widespread (5). The first step towards addressing this issue must go through clarifying the level of genomic variation between various individuals belonging to the same population. This was originally triggered by observations made on deleterious genes (and phenotypes) (6). But with the advent of modern molecular biological techniques, the work extended to uncover other types of genetic variation.

Types of Genetic Variation

The degree of variation between parents and children can be indicative of the variation between individuals belonging to the same population or even between people from different ethnicities. Understandably, each local population harbors a somehow distinct combination of alleles. The amplitude of regions displaying genetic variation ranges widely from a single nucleotide to massive stretches of DNA.

The most common type of genetic variants are single-nucleotide polymorphisms (SNPs), where DNA sequences differ by one base occurring at a population frequency >1%. The number of SNPs identified in the latest build of the NCBI dbSNP database is more than 1.8 billion. Moving up the scale to larger genetic regions of variation, one finds submicroscopic copy number variation (CNV) of DNA segments ranging from kilobases to megabases. CNVs are stretches of DNA, at least 1 kb long, with varying copy number among individuals. CNVs comprise numerous types, including insertions, deletions, duplications and multisite variants. Towards the opposite end of this spectrum are the microscopically visible chromosome anomalies that can be identified through cytogenetic detection (7).

The clinical consequences of having a particular combination of SNPs and CNVs can be uncovered through a number of research strategies, including association studies. Importantly, our knowledge of benign and pathogenic CNVs is conspicuously minimal and more work is needed along this path (8,9). Generally, categorizing pathogenic CNVs

as such depends on whether or not they have a negative impact on the health of the individual. These CNVs can be located within genomic regions that are considered nonfunctional, and hence they are not associated with a particular phenotype. The reality, however, is far more complicated, as studies have shown that these variants can modify complex phenotypes through regulating cell growth, metabolism and cell signaling in general (10,11). Intuitively, CNVs located within coding DNA do play a role in pathological phenotypes, with established examples from a number of well-known disorders, such as susceptibility to HIV infection and progression to AIDS (12), intellectual disability (13,14), susceptibility to Crohn disease (15) and autism spectrum disorder (16).

Other structural and sequence variations in the human genome include tandem repeats and interspersed repeats. The former include minisatellites and microsatellites and represent ~10% of the genome and are usually found in heterochromatic DNA. On the other hand, interspersed repeats can be classified as long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs), and they make up a significant part of the human genome.

The Human Genome Project (HGP) and the Human Variome Project (HVP)

The road to studying variations within the human genome naturally went through establishing a "reference genome", and that mammoth task was undertaken by the Human Genome Project (HGP). HGP started in 1990, with funding from two agencies, the National Institutes of Health (NIH) and the Department of Energy (DOE) in USA. The door was opened for non-US scientists to take part in the project. Initially, it was foreseen that obtaining the complete sequence of one human genome would take 15 years at a cost of $3,000 million. At a later stage, a private US company, Celera Genomics, joined the project and a working draft of the genome was announced in 2000 (17) and a complete one (referred to as the reference genome) in 2003.

Despite the ahead-of-time completion of the project, the analysis of the uncovered DNA sequences is still ongoing. The publicity that accompanied the project expectedly raised public awareness about genetics, but it also raised the level of expectations about medical breakthroughs that were to result from our newfound knowledge about the human genome. These diagnostic and therapeutic expectations turned out to be quite unrealistic, due to the fact that health and disease occur as a consequence of a complex interaction of genetic and non-genetic factors. These interactions are very complex indeed and require a proper understanding of the functional role(s) of the genes as well as of the regulatory networks that control their expression. Conceivably, this very task has been the main focus of a huge number of researchers worldwide.

In 2006, an international initiative was launched to identify and document pathogenic and benign genomic variations worldwide; namely, the Human Variome Project (HVP). The working of the HVP involves collecting and curating genomic variation data from clinical, medical, and research laboratories. The promise of the project is to link disease prevention and diagnosis to personalized therapeutics. Through large-scale collaborative work, the HVP is expected to improve translational research strategies and clinical decision-making processes (18). HVP approached the data collection issue using two strategies; the first one was to utilize existing gene/disease specific databases, while the second strategy depended directly on country-specific nodes to collect all data within a country. As of September 2018, 26 countries had HVP nodes including Kuwait and Egypt from the Arab world (19). http://www.humanvariomeproject.org/about/international-confederation-of-countries-advisory-council.html

The Molecular Basis and Driving Force behind Variation

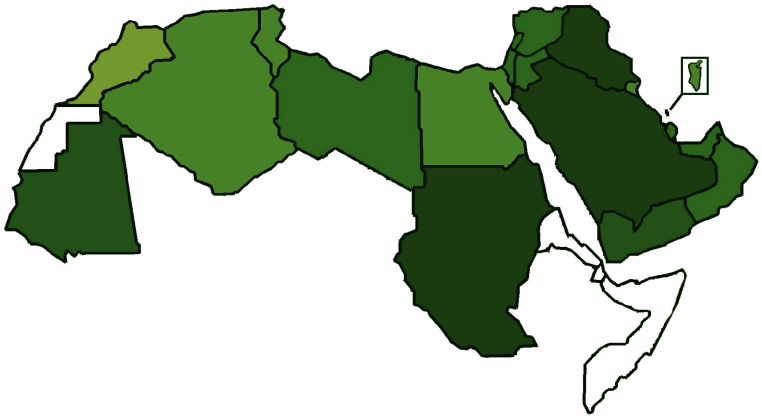
Phenotypic variation is ultimately associated with differences in the physical structure of the DNA or in the regulatory input it receives. There are a number of levels through which such variation is established. Some of these can be deduced from the central dogma in molecular biology. Changes in nucleotide sequences in coding and non-coding DNA through deletion, insertion, substitution and inversion can understandably result in phenotypic variation. This may ultimately alter the amino acid sequence of the encoded protein that may result in modifications to the structure/function of that protein. It is important to remember that changes in protein function can also result from posttranslational modifications (PTMs) such as phosphorylation, methylation and ubiquitination. Prior to the stage when proteins

may undergo PTMs, numerous regulatory steps can take place and these usually fall under the general term "regulation of gene expression". In fact, different patterns of gene expression can bring about phenotypic variation between individuals and even in different cell types within the same individual. Importantly, differential gene expression can be brought about by environmental stimuli and this provides the critical link between the organism and its surrounding. In this context, epigenetic regulation of gene expression works in a highly sophisticated manner to orchestrate varying genetic responses through processes from chromatin modifications (methylation and acetylation) to the actions of microRNAs. Some of these epigenetic modifications, such as genomic imprinting, are heritable, and a number of congenital disorders have been associated with them (20).

The processes driving inter and intra-population variability are numerous, but the most fundamental one is mutations. The extent of physical alteration of the DNA molecule can range from those affecting single base pairs to a whole chromosome. In general, mutations introduce random changes in the genetic code; such changes can be heritable should they happen in a germ cell. It is only in the latter case that a mutation will have direct evolutionary consequences, because then it may be passed from one generation to another and by so doing it can change the frequency of the alleles in the gene pool and introduce new alleles into it. Although mutations provide the necessary starting material for genetic changes on the population level, they are not enough to cause major and sustainable changes in allele frequencies over time. Such change can result from the work of other evolutionary forces; i.e. selection, drift, gene flow which increase or decrease allele frequencies from one generation to the next (21).

Random genetic drift results from the stochastic process of "sampling" gametes that will bring together their genetic content to give rise to the next generation. This process can lead to changes in allele frequencies of some genetic variants over time. The effect of genetic drift increases upon reduction of population size; i.e. during a population bottleneck. This explains why genetic diversity in a population can be reduced substantially upon critical reduction of the size of this population. The most dramatic manifestation of this effect can be observed in the cases when a small number of individuals start a new population. The accompanying loss of diversity in the resulting population is named the founder effect.

The term selection refers to the process by which genetic variants conferring higher chances of survival for the carrier organism are perpetuated preferably over other variants. The key point in this context is not the absolute fitness of the individual; instead it is the set of characteristics that helps an individual be optimally adjusted to its environment. Therefore, the process of selection



Country	Rates for all types of consanguinity (%)
Morocco	19.9-28
Lebanon	12.8-42
Bahrain	10.9-45.5
Algeria	22.6-34
Tunisia	20.1-39.3
Syria	30.3-39.8
Palestine	17.5-66.3
Yemen	40-44.7
Kuwait	22.5-64.3
Jordan	28.5-63.7
UAE	39-54.2
Mauritania	47.2
Libya	48.4
Egypt	20.9-80.4
Iraq	47-60
Sudan	44.2-63.3
Qatar	54
Saudi Arabia	42.1-67.2
Oman	56.3-58

Figure 1: Consanguinity rates in the Arab World.

can push for changes in allele frequencies in a number of directions. New mutations introduce genetic variation into a population and natural selection determines the fate of these new mutants. This simplistic view does not take into account the interaction between different human populations, and in particular the process known as gene flow. The latter involves the movement of genetic material from one population to another through contact between various human groups. Gene flow can have a significant impact on allele frequencies (21).

Consanguinity is a very important factor in determining the level of genetic variability within a population. Unions between two individuals who are related as second cousins or closer are usually classified as consanguineous and it has been estimated that 10.4% of the current global population has come from such marriages (22). These rates are declining in many countries around the world, however current rates of consanguinity in certain countries are higher now than in earlier generations (23). This is possibly due to the fact that the progeny of consanguineous marriages do survive longer now thanks to the recent advancements in the field of medicine. Undoubtedly, this reduction in mortality is paralleled by an increase in extended morbidity if inbreeding continued on the same level from one generation to another. Studies suggest that mortality rates in the offspring of first cousin couples are 3.5% higher than in the case of unrelated couples (24). It is not surprising, therefore, to learn that inbreeding can rapidly increase the frequency of mutations in a community, regardless of whether this mutation is a recessive founder or de novo one. Close-kin unions are culturally ingrained in the Arab world and this has been attracting the attention of geneticists for some time now. Figure 1 shows the levels of consanguinity in this region.

References

1. Garcia-Bertrand R, Simms TM, Cadenas AM, Herrera RJ. United Arab Emirates: phylogenetic relationships and ancestral populations. Gene. 2014;533(1):411-9. doi: 10.1016/j.gene.2013.09.092. PubMed PMID: 24120897.
2. Teebi AS, Teebi SA. Genetic diversity among the Arabs. Community genetics. 2005;8(1):21-6. doi: 10.1159/000083333. PubMed PMID: 15767750.
3. Omberg L, Salit J, Hackett N, Fuller J, Matthew R, Chouchane L, et al. Inferring genome-wide patterns of admixture in Qataris using fifty-five ancestral populations. BMC genetics. 2012;13:49. doi: 10.1186/1471-2156-13-49. PubMed PMID: 22734698; PubMed Central PMCID: PMC3512499.
4. Dajani R, Khader YS, Hakooz N, Fatahalla R, Quadan F. Metabolic syndrome between two ethnic minority groups (Circassians and Chechens) and the original inhabitants of Jordan. Endocrine. 2013;43(1):112-9. doi: 10.1007/s12020-012-9723-y. PubMed PMID: 22740093.
5. Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, Kehrer-Sawatzki H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. Human genetics. 2013. p. 1077-130.
6. Dobzhansky T, Wright S. Genetics of Natural Populations. V. Relations between Mutation Rate and Accumulation of Lethals in Populations of Drosophila Pseudoobscura. Genetics. 1941;26(1):23-51.
7. Feuk L, Carson AR, Scherer SW. Structural variation in the human genome. Nat Rev Genet. 2006;7(2):85-97.
8. Beckmann JS, Estivill X, Antonarakis SE. Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. Nat Rev Genet. 2007;8(8):639-46.
9. Lee C, Iafrate AJ, Brothman AR. Copy number variations and clinical cytogenetic diagnosis of constitutional disorders. Nat Genet. 2007;39(7 Suppl):S48-54.
10. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Månér S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. Science. 2004;305(5683):525-8.
11. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. Fine-scale structural variation of the human genome. Nat Genet. 2005;37(7):727-32.
12. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK. The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science. 2005;307(5714):1434-40.
13. Wagenstaller J, Spranger S, Lorenz-Depiereux B, Kazmierczak B, Nathrath M, Wahl D, Heye B, Glaser D, Liebscher V, Meitinger T, Strom TM. Copy-number variations measured by single-nucleotide-polymorphism oligonucleotide arrays in patients with mental retardation. Am J Hum Genet. 2007;81(4):768-79.
14. Potocki L, Bi W, Treadwell-Deering D, Carvalho CM, Eifert A, Friedman EM, Glaze D, Krull K, Lee JA, Lewis RA, Mendoza-Londono R, Robbins-Furman P, Shaw C, Shi X, Weissenberger G, Withers M, Yatsenko SA, Zackai EH, Stankiewicz P, Lupski JR. Characterization

of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. Am J Hum Genet. 2007;80(4):633-49.
15. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, Reinisch W, Teml A, Schwab M, Lichter P, Radlwimmer B, Stange EF. A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet. 2006;79(3):439-48.
16. Autism Genome Project Consortium Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet. 2007;39(3):319-28.
17. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann Y, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blöcker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowki J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ, Szustakowki J; International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.
18. Cotton RG, Vihinen M, den Dunnen JT; Human Variome Project Nomenclature And Standards Group. Genetic tests need the Human Variome Project. Genet Test Mol Biomarkers. 2011;15(1-2):3.
19. Patrinos GP, Smith TD, Howard H, Al-Mulla F, Chouchane L, Hadjisavvas A, Hamed SA, Li XT, Marafie M, Ramesar RS, Ramos FJ, de Ravel T, El-Ruby MO, Shrestha TR, Sobrido MJ, Tadmouri G, Witsch-Baumgartner M, Zilfalil BA, Auerbach AD, Carpenter K, Cutting GR, Dung VC, Grody W, Hasler J, Jorde L, Kaput J, Macek M, Matsubara Y, Padilla C, Robinson H, Rojas-Martinez A, Taylor GR, Vihinen M, Weber T, Burn J, Qi M, Cotton RG, Rimoin D; International Confederation of Countries Advisory Council. Human Variome Project country nodes: documenting genetic information within a country. Hum Mutat. 2012;33(11):1513-9.
20. Girardot M, Feil R, Llères D. Epigenetic deregulation of genomic imprinting in humans: causal mechanisms and clinical implications. Epigenomics. 2013;5(6):715-28.
21. Hartl DL, Clark AG. Principles of population genetics. 2007. Sinauer Associates.
22. Bittles AH. A community genetics perspective on consanguineous marriage. Community Genet. 2008;11(6):324-30.
23. Tadmouri GO, Nair P, Obeid T, Al Ali MT, Al Khaja N, Hamamy HA. Consanguinity and reproductive health among Arabs. Reprod Health. 2009;6:17.
24. Bittles AH, Black ML. The impact of consanguinity on neonatal and infant health. Early Hum Dev. 2010;86(11):737-41.