Chapter 2

CTGA: The Database for Genetic Disorders in Arabs
Background

Since the 1950s, Arab countries have made considerable progress in medical services leading to better life expectancies and access to health care. Similarly, Arab scholars working in the field of biomedical sciences are giving more attention to publish their results at national or international levels. This lead to the description and extensive documentation of many genetic disorders not previously recognized.

Several publications reviewing different aspects of genetic diseases in Arab populations are available. Yet, these publications were rapidly outdated as new disorders are continuously described in Arabs. However, two significant attempts to overcome this problem are worth mentioning:

- The effort of Prof. Ahmad Teebi and Mr. Saeed Teebi to establish the Arab Genetic Disease Database (AGDDB), a curated catalog of genetic disorders found in Arab populations. The first online release of the database was populated primarily with information from the textbook ‘Genetic Disorders Among Arab Populations’, co-authored by Prof. Teebi and Dr. Talaat Farag in 1997. AGDDB was composed of data elements including clinical, genomic, reference, and population frequencies of genetic disorders. After initial indexing in 2002, AGDDB contained over 1000 unique disorder entries. The database was freely accessible at www.agddb.org until in 2004 it went offline because of technical problems.

- The attempt of Dr. Ghazi Tadmouri and Dr. Nisrine Bissar-Tadmouri to maintain offline tabular lists of genetic disorders described in Arab individuals with corresponding references by monitoring international disease databases and scanning bibliographic indices. Using this strategy, 374 entries for genetic disorders in Arabs were recorded in 1999. This number increased to 752 entries early in 2004. In March 2004, the name ‘Catalogue of Transmission Genetics in Arabs’ (CTGA) Database was coined for any future online database that may materialize out of this survey.

To this end, the Centre for Arab Genomic Studies (CAGS) adopted the proposal to launch a pilot project to construct the CTGA database with the aim to educate the medical community and raise public awareness in at-risk populations. Initially, major components of the CTGA database were constructed and tested separately using FileMaker software, which is a cross-platform, desktop, relational database application. The database was then assembled and tested with limited amount of data to decide on proper designs for offline and online interfaces used by curators (editors) and guest users, respectively (Figures 3 and 4).
The current version of CTGA is a textual database composed of sets of "flat" tables linked together through file system specifications. The structure of CTGA depends on a web-based search that uses an indexing system for rapid mining of information (See chapter 3). A detailed record in CTGA includes text-, URL- and graphic-based fields (Figure 4). The Title and Alternative Names indicate the primary title and alternative titles and symbols of the disorder or gene. A graphical map demonstrates the geographical origin of the individuals described in the entry. A disorder is categorized according to the World Health Organization International Classification of Disease (WHO-ICD) 10th revision. OMIM number is a URL-based field that takes the user to the corresponding file of the gene or disorder at the OMIM database. Information regarding Gene Map Locus is drawn primarily from OMIM. Mode of Inheritance, Description and Molecular Genetics are textual fields that contain summaries on the clinical features and genetic pathology for the corresponding entry. Epidemiology in the Arab World is the major part of an entry since it includes a detailed review of research analyses regarding the gene loci or clinical phenotypes in Arab individuals. References within an entry are linked to their corresponding PubMed abstracts except for articles from national peer-reviewed medical journals not indexed in PubMed. Following the references are two URL-based fields. Related CTGA Records takes the user to any intra-CTGA entry(ies) with a shared relationship(s) while Links anchors at external resources with additional information. Authors who contribute with additions or changes to the entry are given credit in the Contributors field along with the date when the contribution was submitted. Changes made by the editorial staff are documented in the Edit History field (Figure 4).
Hemoglobin - Alpha Locus 1

Alternative Names
HbA1
HbA
HbA1C
HbA2
HbF
HbG
HbG-2
HbG-2
do
HbG-2
G

WHO International Classification of Diseases
Disease of the blood and blood-forming organs and certain disorders involving the immune mechanism - Hematologic anemias

OMIM Number
114110

Gene Map Location
Hsnp-5 13 3

Mode of Inheritance
Autosomal dominant (Hb1 13 35 to Hb1 13 11)

Description
Thalassemia is an inherited disease of faulty production of hemoglobin. The name is derived from the Greek word "thalassa" meaning "the sea" because the condition was first described in populations living near the Mediterranean Sea.

Alpha-thalassemia are characterized by decreased hemoglobin alpha chain synthesis, alpha-zero-thalassemia being the condition where no normal alpha globin is produced, and alpha-plus-thalassemia being the condition where there is a reduced globin production. There are two alpha globin genes per haploid genome, and alpha thalassemia mutations can result from one to four gene deletions. A single alpha gene mutation leads to the most severe state (alpha-zero). The two gene mutation is a milder clinical condition, with mild hypochromic microcytic anemia.

Mutation of three of the alpha genes leads to Hemoglobin H disease, characterized by macrocytic hypochromic normocytic anemia, hepatosplenomegaly, mild jaundice, and sometimes thalassemia-like bone changes. Mutation of all four alpha genes results in Hb Bart hydrops fetalis (Hb Bart) syndrome, typified by fetal onset of generalised edema, pleural and peritoneal effusions, and severe hypochromic anemia. Death usually occurs in the neonatal period. No effective treatment is available for Hb Bart syndrome. Occasional HbC transfusion may be required for patients with HBB disease.

Molecular Genetics
The alpha globin gene cluster located on chromosome 16 spans about 30 kb and includes four functional genes and three pseudogenes. HBA1 is the gene encoding alpha 1-globin and is located to the heterochromatic region of chromosome 16p. The gene is 24 Kbp in size and consists of three exons. About 90% of the mutations in HBA1 are deletions and only 10% are point mutations. Hemoglobin alpha is produced throughout fetal and adult life. Two alpha chains combine with two beta chains to constitute HBA, which in normal adult life comprises about 97% of the total hemoglobin. The expression of HBA1 is regulated by a region (RASSF) located 40 Kbp upstream of the alpha globin cluster.

Epidemiology in the Arab World
Algeria
Wajnman et al. (1972) described Hb Setif in an Algerian family.

Bakhoum et al. (1986a) described hemoglobin Lerlei [alpha 2(FP)Ala—Ser] in a 10-year-old Algerian boy born in Lerlei. The child had erythrocritosis and macrocytosis, the latter being due to iron deficiency. The oxygen loading curves at equilibrium, and the kinetic measurements demonstrated that the substitution of alpha 2(FP) Ala for a Ser results in increased oxygen affinity and decreased O2 release.

Wajnman et al. (1989) found Hb Meknès in an Algerian patient during a systematic search for hemoglobinopathy screening program in Luxembourg. Using isoelectric focusing and reverse phase-high performance liquid chromatography (RP-HPLC), Wajnman et al. (1993) determined that the molecular mutation in amino and position 141 of this HBA1 gene changed the residue from proline to serine.

Bahrain
Mohammad (1993) analysis 76 Bahraini nationals with the Hb II disease. Variability in the clinical spectrum was observed, with three different forms of clinical presentations. In a recent study, the prevalence were observed, with mean, moderate hypochromic and microcytic. Adolescents, however, were asymptomatic, behaved like heterozygous trait, and were only identified in the course of family screening. The third group was intermediate, and comprised most of the children in the age group of 1-10 years. They presented with chronic hypochromic anemia with an intermittent severity, which required regular blood transfusions. Through 40% of them had detectable epistemology, none had macrocytosis hyperchromic anemia. Since the Bahrain population is an extension of the Eastern province of Saudi Arabia, Mohammad (1993) proposed that the Hb II disease in the Bahraini population may be due to a common form of the non-deleitional alpha-globin gene defect, as reported marker for the Saudi Arabian population.

Kuwait
Adeeb et al. (1999) characterized the alpha thalassaemia determinants among Kuwaiti Arabs. PCR, hybridization and DHA sequencing techniques were used to analyze 64 alpha-thalassaemia chromosomes. Three mutations were identified in 20 chromosomes from patients with HBB disease. There were Poly A signal mutation in alpha 2-globin gene (HBA2 7%), alpha (3.7 Kb deletion, 10%) and alpha-2.1 (4.9%).

Morocco
In 1999, Wajnman et al. described Hb Nouakchott [alpha 146(M2)His—Leu] in a patient from Mauritania. The most striking fact in Hb Nouakchott was the highly increased hydrophobicity of the abnormal chain. Even though the substitution concerned a proline residue, it was without consequence on the oxygen binding and the stability of the molecule.

Morocco
Bakhoum et al. (1986b) reported the ascertainment of Hb Deun (alpha 2(Ap)−Asp—Arg) and Hb O-Arab (beta 121 [HEX]-Leu) in a healthy Moroccan man. The identification of Hb Deun was based on sequence determination of the alpha 2 chain. The percentages of the various hemoglobins showed that the doubly mutated hemoglobin Deun-O-Arab has a normal stability and suggested that the Deun mutation is carried by the alpha 1-globin. In cord blood of the premature's son, the output of the alpha Deun gene was found equivalent to that missing in the adult.

25
Source of Information

In 2004, CAGS coordinated the collection of data on genetic disorders in the Arab population of the UAE as a model system to be implemented in other Arab countries at later stages. Information in CTGA was drawn at the time from two main sources:

- **Nationally and internationally published literature**: Bibliographic databases were screened for relevant articles on genetic disorders in the UAE. Whenever possible, comprehensive manual scan of hardcopies of national peer-reviewed journals was conducted.
Laboratory records: In major hospitals of the UAE, patient records covering the last 10–15 years were studied prospectively. These hospitals included laboratories for molecular diagnostics, cytogenetics, biochemistry and others. Detailed information, including the mutation, was recorded using a standardized method. Patient records proved to be an invaluable source of information since they indicated the presence of several inherited disorders for which occurrence data had not been published before.

While data on genetic disorders in patients of various nationalities were collected, only those obtained from UAE nationals and other Arab patients appear in the CTGA database. Furthermore, personal communication with local geneticists provided further insight into the spectrum of inherited disorders in the UAE.

In 2006, CAGS formed its Arab Council, which includes a number of regional geneticists from Bahrain, Egypt, Jordan, Kuwait, Lebanon, Oman, Qatar, Saudi Arabia, Sudan, and Tunisia. Soon after the inception of the Arab Council of CAGS, extensive information on genetic disorders occurring in several Arab countries were submitted to CTGA. However, in late 2006 CAGS initiated a detailed data collection strategy on genetic disorders from countries neighboring the UAE (e.g., Bahrain). The strategy mainly depended on scanning nationally and internationally published literature as well as personal communications with experts in the country. Initial data indicated the presence of more than 150 genetic disorders in the relatively small Bahraini population. Succinctly, the magnitude of genetic disorders and congenital abnormalities reported from the Arab populations of the UAE and Bahrain alone (at least 250) demonstrates the efficacy of the algorithm adapted when compared to published reviews on the subject.

As of September 2006, CTGA had 719 phenotype entries and 269 related gene entries with descriptions in Arab individuals (Figure 5). In the UAE, CTGA information includes about 228 phenotypic descriptions and 28 related genes. Currently, authors at the Centre for Arab Genomic Studies create about 30 entries and update an equivalent number each month (Figure 6). Although CTGA has a short lifespan on the public domain of the Internet, it attracts 200 to 250 unique users per day. The peak of simultaneous users accessing the database usually occurs between 05:00 and 19:00 GMT and indicates a westward geographical location for most of the traffic (mainly from the Gulf region, Europe, and North America; Figure 7).

Figure 5. Total growth of CTGA Database records from December 2004 until August 2006.
Significance of CTGA

A tool for decision-making in health-related domains: The geographical distributions of genetic disorders in CTGA can either be restricted to small locales (Stuve–Wiedemann syndrome), commonly widespread (beta-thalassemia), or reflect a patchy distribution (alpha-thalassemia) although a high prevalence is expected in the region. On the other hand, the molecular/biochemical pathologies in approximately 25% of genetic disorders described in Arabs have not been determined yet. Thus, these serve as excellent candidates for linkage analyses and genotype/phenotype studies. Obviously, the interpretation of these data is an important tool for authorities to decide on future health-related strategies and to propose research directions on disorders for which information is still scant.

A hub of locally produced scientific information on genetic disorders: By publishing scientific information locally produced in peer-reviewed medical journals from the Arab World, CTGA exposes valuable local information that is not accessible to the large scientific community.

A catalyst for establishing collaborations with Arab scientific groups: The extended consanguineous family structure, commonly present in Arab societies, is an important factor leading to the propensity of severe congenital inherited diseases in most Arab populations. Incidentally, genetic disorders in Arabs tend to display peculiar distribution patterns not present in many other world populations. A major model that explains this concept is the vertical
dissemination of a genetic mutation in an Arab family, where mutation carriers mostly remain concentrated within the extended family; thus, offering great opportunities to depict the genetic nature of their disease predisposition. In view of all the above, the wealth of information that CTGA is accumulating is, in our opinion, an indispensable tool for scientists to recognize Arab colleagues working on similar domains and decide on possible collaborations or exchange of know-how.

An educational tool on genetic disorders in the region: Studies have clearly indicated that the correct dissemination of knowledge is an important step towards the eradication of genetic disorders in Arab populations. The CTGA database plays such an educational role as it addresses both the medical communities and at-risk populations.

CTGA Database Static

Recently, the Centre for Arab Genomic Studies released a new service under the name ‘CTGA Database Static’. Records on genetic disorders in Arab populations in CTGA Database Static are derived from the original CTGA Database and made available in Acrobat PDF format for easy reading, downloading, sharing, and printing (Appendix 1). Hence, CTGA Database Static represents an image copy of the original CTGA Database that will be updated once or twice every year. For this reason, users wishing to have the most updated information on genetic disorders and related genes in Arabs are advised to use the frequently updated CTGA Database. CTGA Database Static records are not searchable, but classified alphabetically according to index terms that allow easy access to information on various genetic disorders and related genes described in Arab individuals (Figure 8).

Figure 8. The Index Page of CTGA Database Static (www.cags.org.ae).