

ARTICLE

The Moroccan Genetic Disease Database (MGDD): a database for DNA variations related to inherited disorders and disease susceptibility

Hicham Charoute^{1,3}, Halima Nahili¹, Omar Abidi¹, Khalid Gabi², Hassan Rouba¹, Malika Fakiri³ and Abdelhamid Barakat^{*,1}

National and ethnic mutation databases provide comprehensive information about genetic variations reported in a population or an ethnic group. In this paper, we present the Moroccan Genetic Disease Database (MGDD), a catalogue of genetic data related to diseases identified in the Moroccan population. We used the PubMed, Web of Science and Google Scholar databases to identify available articles published until April 2013. The Database is designed and implemented on a three-tier model using Mysql relational database and the PHP programming language. To date, the database contains 425 mutations and 208 polymorphisms found in 301 genes and 259 diseases. Most Mendelian diseases in the Moroccan population follow autosomal recessive mode of inheritance (74.17%) and affect endocrine, nutritional and metabolic physiology. The MGDD database provides reference information for researchers, clinicians and health professionals through a user-friendly Web interface. Its content should be useful to improve researches in human molecular genetics, disease diagnoses and design of association studies. MGDD can be publicly accessed at <http://mgdd.pasteur.ma>.

European Journal of Human Genetics (2014) 22, 322–326; doi:10.1038/ejhg.2013.151; published online 17 July 2013

Keywords: Morocco; genetic disorders; database; mutations; polymorphisms

INTRODUCTION

Modern genetic approaches have considerably evolved in the past few years. Genome-wide association studies and next-generation DNA sequencing have become the primary methods for the identification of the genetic bases of human diseases.¹ Consequently, their use generated a large number of DNA variations in the entire human genome. Therefore, there is a huge need to collect and document variations in dedicated databases to ensure their availability. Human mutation databases can be divided into three major types: general databases, locus-specific databases (LSDBs), and national and ethnic mutation databases (NEMDBs). General databases such as Online Mendelian Inheritance in Man (OMIM) and the Human Gene Mutation Database (HGMD) collect published gene mutations associated with hereditary disorders.^{2,3} LSDBs are generally developed and maintained by experts with particular interest to a single gene or chromosomal locus.⁴ NEMDBs record mutation spectrum observed in a particular population or ethnic group; they provide relevant information to study the population origin and migration flow.⁵ The Human Genome Variation Society (HGVS) maintains comprehensive and updated lists of LSDBs and NEMDBs, in a dedicated website (<http://www.hgvs.org>), which presents guidelines and recommendations to generate the content and structure of variation databases. The HGVS website displays two mutation databases dedicated to the Arab population, including the Catalogue of Transmission Genetics in Arabs (CTGA) Database, constructed by the Centre for Arab Genomic Studies, and The Lebanese National Mutation Frequency Database.^{6,7}

Morocco is located at the northwest of Africa with a land area of 710,850 km². It is bordered in the north by the Mediterranean Sea, to the south by Mauritania, to the east by Algeria and to the west by the Atlantic Ocean. According to the last Census executed in 2004, the Moroccan population number is 29 891 708 with a population growth rate of 1.4%. The main ethnic groups in the Moroccan population are Arabs, Berbers and Sahrawi. Similarly to many Arab countries, consanguineous marriage is common in Morocco, with rates ranging from 19.9 to 28%.⁸ A recent study estimates the frequency of consanguineous marriages to be 59.09% among families with autosomal recessive disorders.⁹ Despite the improvement in the health-care system in Morocco, few epidemiological data are available about the prevalence of genetic diseases. Cancer is a major health problem and a common cause of death. Lung cancer is the most common type of cancer among men, which presents 25.6% of all male cancer patients and with a mortality rate of 19.7%, followed by prostate cancer. On the other hand, breast cancer is the most common disease in women, followed by cervix uteri cancer.¹⁰ According to a National Survey during the year 2000, the prevalence of cardiovascular risk factors was high in Morocco; in particular, the prevalence of diabetes was 6.6% and similar for male and female individuals, but higher in the urban areas.¹¹

In the last decade, genetic data have been gathered in Morocco in order to correlate the susceptibility of these diseases to mutations in specific, and in parallel, the analysis of Mendelian diseases led to the identifications of causative mutations. Here, we describe the Moroccan Genetic Disorders Database (MGDD), freely accessible

¹Laboratoire de Génétique Moléculaire Humaine, Département de Recherche Scientifique, Institut Pasteur du Maroc, Casablanca, Morocco; ²Département Informatique, Institut Pasteur du Maroc, Casablanca, Morocco and ³Laboratoire d'Agroalimentaire et Santé, Faculté des Sciences et Techniques, Université Hassan I, Settat, Morocco

*Correspondence: Dr A Barakat, Laboratoire de Génétique Moléculaire Humaine, Département de Recherche Scientifique, Institut Pasteur du Maroc, 1, Place Louis Pasteur, C.P., Casablanca 20360, Morocco. Tel: +212 522 434 471; Fax: +212 522 260 957; E-mail: hamid.barakat@pasteur.ma

Received 15 February 2013; revised 28 May 2013; accepted 11 June 2013; published online 17 July 2013

through internet at <http://mgdd.pasteur.ma>. The database was established for collecting and cataloguing disease-causing mutation and polymorphism frequencies. The MGDD content gives an overview about the spectrum of studied genetic diseases in the Moroccan population.

MATERIALS AND METHODS

The PubMed, Web of Science and Google Scholar databases were searched to identify available articles published until April 2013, using the following keywords: gene, allele, mutation, polymorphism, genetic disease and genetic disorder in combination with Morocco or Moroccan. The search was performed without any restriction on language. In addition, all references mentioned in the identified articles were reviewed in order to investigate additional literature that was not indexed.

The mutation nomenclature for gene variants follows the HGVS recommendations; this format was checked using the Mutalyzer software.¹² Genetic diseases registered in the database have been classified using the World Health Organization (WHO) International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) Version 2010 (<http://apps.who.int/classifications/icd10/browse/2010/en>).

The MGDD is designed and implemented on a three-tier model, in which the data presentation, the application processing and the data management are separate processes. The data storage is managed with Mysql relational database running on a Linux server located at the Pasteur Institute of Morocco. For all data retrieval and presentation, we use the PHP programming language.

RESULTS

The database web interface

The web interface provides publicly available web pages for data search using different query options and also for new data submission. Administrators can update the database content after password logging via a restricted access page. The database users can submit data via dedicated web form. The submitted data will be published online after validation by the database administrators.

The content of the MGDD can be interrogated by disease name, gene name and OMIM ID using keyword queries; a drop-down menu enables the user to select the desired information. The web interface displays suggestions when the user starts typing keywords in the search box (Figure 1). In addition, the database can be explored using alphabetical search; genes and diseases are arranged according to the first letter of their names. User can search variation data related to a given disease or gene of interest using the provided query options. The output result will display variation details that include DNA change, amino-acid change, location and mutation type. Moreover, association study data such as sample size, allelic/genotypic frequencies, *P*-value and odds ratios are also provided when available.

The MGDD provides the same essential information about each disease and gene. Disease information include OMIM number, disease name, disease category and the mode of inheritance. For each gene, details such as gene symbol, gene definition and chromosome location are provided. The reference article is displayed with a link

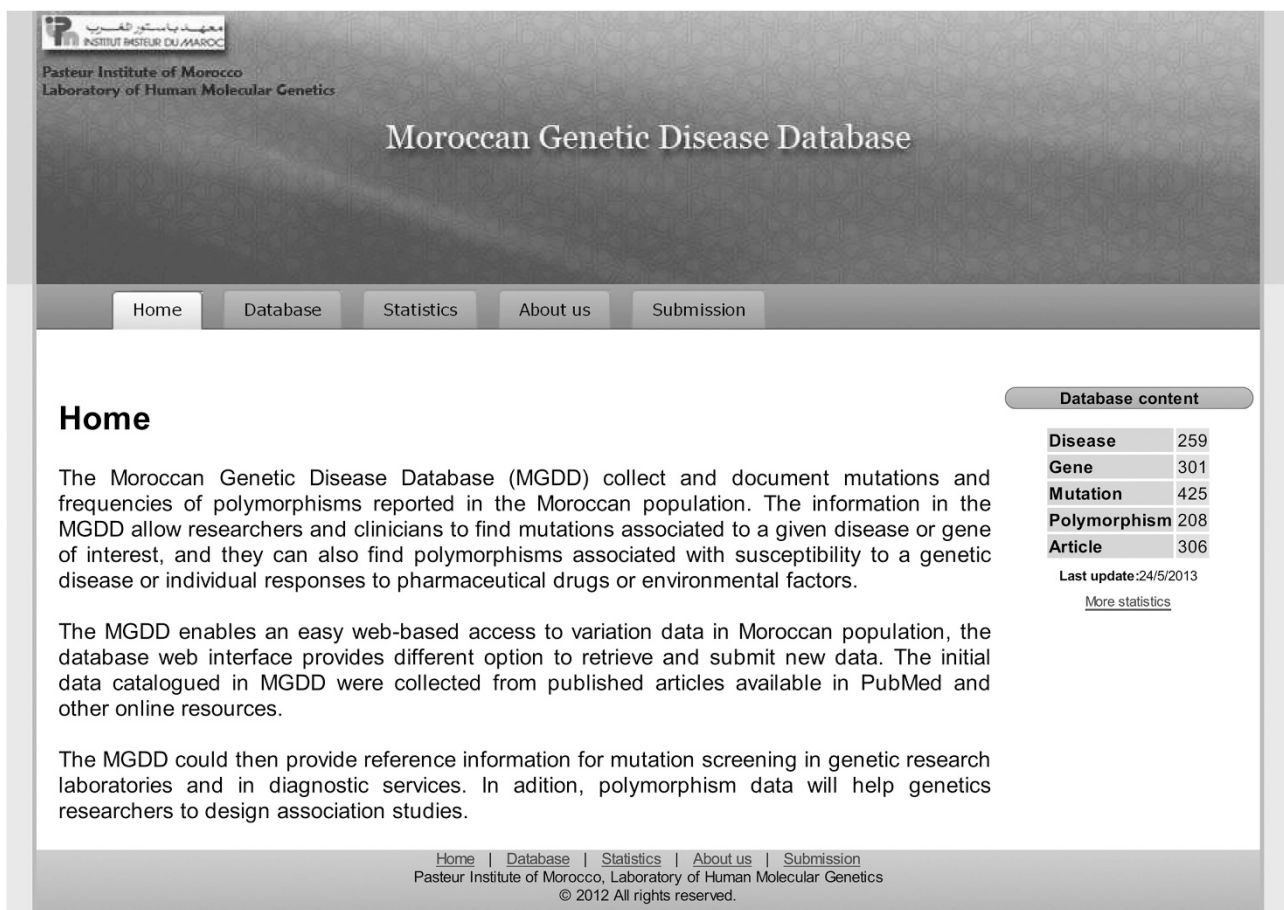


Figure 1 The web interface of the MGDD.

to their abstract, in most cases in PubMed. Furthermore, to get more information about each disease and gene, various links to other databases are available: OMIM, the portal for rare diseases and orphan drugs (Orphanet), and other online resources.

The database content

The data catalogued in the MGDD were derived from literature available in PubMed, with some exceptions of articles not indexed in PubMed. There are now 306 articles registered in the database date, ranging from 1990 to 2013, obtained from 141 different journals. To date, the MGDD contains 633 mutations/polymorphisms for 301 genes and 259 diseases. The collected diseases are divided into two major categories: Mendelian diseases and multifactorial diseases with 240 and 19 different pathologies, respectively.

Mendelian diseases

The total number of Mendelian variants included in the database is currently 425 mutations. These mutations can be categorised as 288 substitutions, 99 deletions, 14 insertions, 12 duplications, 4 repeats, 5 indels and 3 others mutations (large conversion, large lesion and balanced translocation) (Figure 2). The distribution of mutations according to DNA change location shows that 358 mutations are located in exons, 37 in introns and 8 in UTR regions (Supplementary Figure 1). The disease with the highest number of mutations is beta-thalassemia with 18 mutations, followed by Phenylketonuria and Ataxia-telangiectasia with 14 mutations (Table 1).

A total of 240 Mendelian diseases were registered. The MGDD data analysis reveals that 74.17% of all the inherited diseases in the Moroccan population follow autosomal recessive mode of inheritance. Autosomal dominant are also frequent, representing 16.67% of all cases. In addition, 5% are X-linked, 0.42% Y-linked, 0.83% sporadic, 0.42% mitochondrial and 0.42% of unknown mode of inheritance. On the other hand, 2.08% of genetic diseases have multiple modes of transmission, 4 diseases are inherited as autosomal recessive and autosomal dominant, 1 disease has autosomal recessive and X-linked recessive inheritance, and 1 disease has autosomal recessive and mitochondrial mode of transmission (Figure 3).

According to the WHO ICD-10 classification of diseases, the major category of inherited diseases identified in the Moroccan population includes endocrine, nutritional and metabolic diseases (24.17%), followed by congenital malformations and chromosomal

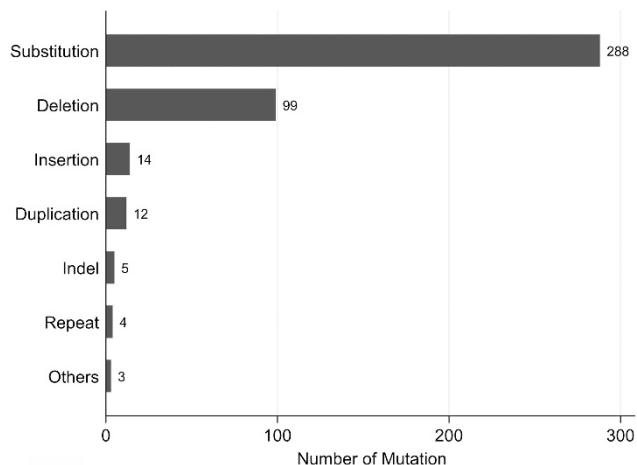


Figure 2 Distribution of mutations involved in Mendelian diseases according to DNA modification.

abnormalities (22.08%). The major affected systems are nervous (19.17%), blood and immune mechanism (10%). Diseases of the eye and adnexa (5.42%) and diseases of the ear (4.17%) are also prevalent (Figure 4).

Multifactorial diseases

The total number of polymorphisms is 208 for 74 genes. The majority of these polymorphisms were investigated in susceptibility to diabetes mellitus type 2 (T2DM), spermatogenic failure, breast-ovarian cancer and pulmonary tuberculosis (Table 1). Association study data on 19 multifactorial diseases have been deposited into the MGDD database. Neoplasms account for a major proportion with 26.32%, followed by infectious and parasitic diseases, diseases of the circulatory system and endocrine, nutritional and metabolic diseases with 15.79% (Figure 4).

DISCUSSION

We developed the MGDD for documenting DNA mutations related to inherited disorders and frequencies of polymorphisms that have been investigated in the Moroccan population. The MGDD was established with the main objective of collecting data about inherited disorders in the Moroccan population and provides them to scientists in the fields of human genetics through a user-friendly web interface.

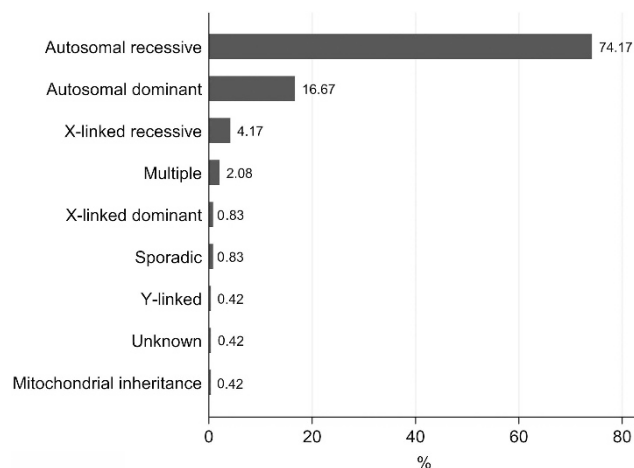


Figure 3 Distribution of Mendelian disorders according to the mode of inheritance.

Table 1 Genetic diseases with the highest number of mutations and polymorphisms

Diseases	Mutation	Polymorphism	Total
Susceptibility to type 2 diabetes mellitus (T2DM)	0	43	43
Susceptibility to familial breast ovarian cancer 2	10	22	32
Susceptibility to spermatogenic failure	0	29	29
Susceptibility to familial breast ovarian cancer 1	4	21	25
Ataxia-telangiectasia	14	11	25
Beta thalassemia	18	0	18
Susceptibility to <i>Mycobacterium tuberculosis</i>	0	16	16
Phenylketonuria	14	0	14
Primary congenital A glaucoma 3	11	0	11
Retinoblastoma	11	0	11

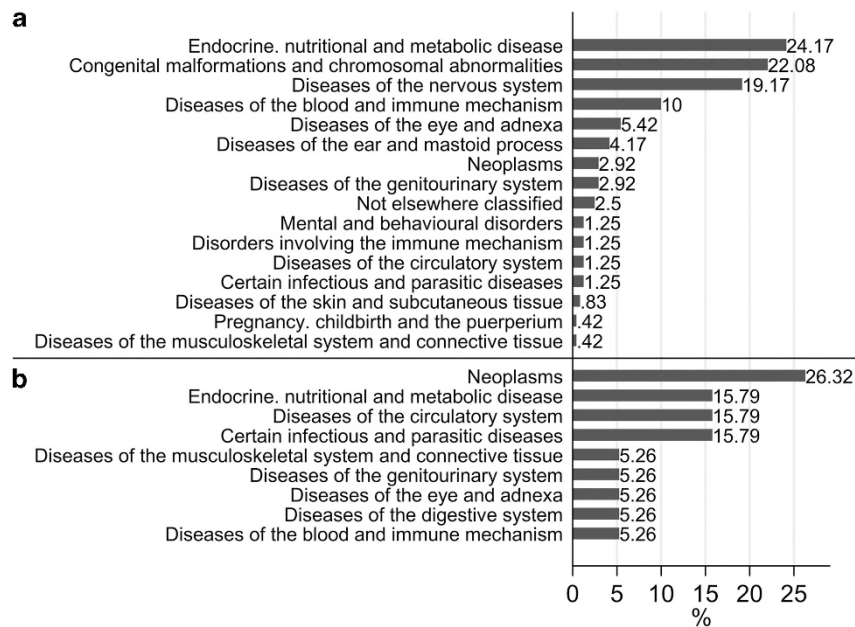


Figure 4 Classification of genetic disorders using the WHO ICD-10 system. (a) Mendelian diseases. (b) Multifactorial diseases.

There are now 425 mutations identified in 236 genes, the majority being substitutions in the coding regions. The database includes several founder mutations that have been identified in the Moroccan population. Among them, the c.35delG mutation in the GJB2 gene that causes autosomal recessive nonsyndromic hearing loss has been derived from a common founder process in the Moroccan population, its age being estimated to ~2700 years.¹³ The c.1643_1644delTG mutation in the XPC gene represents the major cause of xeroderma pigmentosum in Maghreb region (Morocco, Algeria and Tunisia); haplotype analysis revealed the presence of a common founder effect for this XPC mutation in the Mediterranean region.^{14,15} The identification of founder mutations has important implications in molecular diagnosis and genetic counselling. Furthermore, a few prevalent founder mutations enable human geneticists to make more efficient genetic screening than testing for many rare mutations.¹⁶ The database includes data on 208 polymorphisms from 74 gene loci; this includes polymorphisms associated with susceptibility to a genetic disease, or individual responses to pharmaceutical drugs or to environmental factors. In addition, polymorphisms with no significant association are also recorded. Together, these polymorphisms reflect the genetic heterogeneity of the Moroccan population and should be relevant for the optimisation and design of the association studies.

There are now 240 inherited diseases registered in the database; autosomal recessive disorders represent the major proportion (74.17%) followed by autosomal dominant disorders (16.67%). The high rate of consanguineous marriages in the country may contribute to these observations. In a previous study, the rate of consanguineous marriages in Morocco was shown to correlate with the high prevalence of autosomal recessive disorders.⁹ This is similar to the situation in Tunisia, where frequencies of autosomal recessive and dominant disorders were 62.9% and 22.9%, respectively.¹⁷ The disease classification reveals that endocrine, nutritional and metabolic diseases, congenital malformations and neurodegenerative diseases are the major categories of Mendelian disorders in the Moroccan population, with similar prevalence as in the Tunisian population.¹⁷

We plan to collect and add new mutation data from diagnostic services and research laboratories to our database continuously. MGDD encourages the human genetic research community and clinicians to submit their genetic data via dedicated web forms.

CONCLUSION

The central goal of the MGDD consists of cataloguing the molecular data of inherited disorders found in the Moroccan population through a user-friendly web interface. The database can benefit the human genetics researchers, clinicians and health professionals interested in a particular genetic disorder or gene to gather an overview of mutations reported in the Moroccan population. In addition, the database could help designing the diagnostic tests to detect mutations in molecular diagnostic services, as well as the implementation of epidemiological approaches for estimating the prevalence of genetic diseases in the Moroccan or more general Arab population. On the other hand, polymorphism data will be useful for the optimisation and design of genetic association studies. The MGDD provides relevant information not only to the local scientific community in Morocco but also to researchers in countries with similar ethnic backgrounds. We will keep the database updated with additional data from online submissions and literature searches.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

- Luo L, Boerwinkle E, Xiong M: Association studies for next-generation sequencing. *Genome Res* 2011; **21**: 1099–1108.
- Hamosh A, Scott AF, Amberger JS *et al*: Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 2005; **33**: D514–D517.
- Stenson PD, Ball EV, Mort M *et al*: Cooper DN Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 2003; **21**: 577–581.
- Claustres M, Horaitis O, Vanevski M, Cotton RG: Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. *Genome Res* 2002; **12**: 680–688.

- 5 Patrinos GP: National and ethnic mutation databases: recording populations' genography. *Hum Mutat* 2006; **27**: 879–887.
- 6 Tadmouri GO, Al Ali MT, Al-Haj Ali S, Al Khaja N: CTGA: the database for genetic disorders in Arab populations. *Nucleic Acids Res* 2006; **34**, Database issue D602–D606.
- 7 Megarbane A, Chouery E, van Baal S, Patrinos GP: The Lebanese National Mutation Frequency database. *Eur J Hum Genet* 2006; **14**(Suppl 1): 365.
- 8 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.
- 9 Jaouad IC, Elalaoui SC, Sbiti A, Elkerh F, Belmahi L, Sefiani A: Consanguineous marriages in Morocco and the consequence for the incidence of autosomal recessive disorders. *J Biosoc Sci* 2009; **41**: 575–581.
- 10 Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM: Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010; **127**: 2893–2917.
- 11 Tazi MA, Abir-Khalil S, Chaouki N *et al*: Prevalence of the main cardiovascular risk factors in Morocco: results of a National Survey, 2000. *J Hypertens* 2003; **21**: 897–903.
- 12 Wildeman M, van Ophuizen E, den Dunnen JT, Taschner PE: Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Hum Mutat* 2008; **29**: 6–13.
- 13 Abidi O, Boulouiz R, Nahili H *et al*: The analysis of three markers flanking GJB2 gene suggests a single origin of the most common 35delG mutation in the Moroccan population. *Biochem Biophys Res Commun* 2008; **377**: 971–974.
- 14 Soufir N, Ged C, Bourillon A *et al*: A prevalent mutation with founder effect in xeroderma pigmentosum group C from north Africa. *J Invest Dermatol* 2010; **130**: 1537–1542.
- 15 Senhaji MA, Abidi O, Nadifi S *et al*: c.1643_1644delTG XPC mutation is more frequent in Moroccan patients with xeroderma pigmentosum. *Arch Dermatol Res* 2012; **305**: 53–57.
- 16 Pastinen T, Perola M, Ignatius J *et al*: Dissecting a population genome for targeted screening of disease mutations. *Hum Mol Genet* 2001; **10**: 2961–2972.
- 17 Romdhane L, Abdelhak S: Genetic diseases in the Tunisian population. *Am J Med Genet A* 2011; **155**: 238–267.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)