# *Deniz*: The electronic database for ß–thalassemia mutations in the Arab world

*Ghazi O. Tadmouri, PhD, MSc, Resul I. Gulen, MSc, BSc.*

## ABSTRACT

**Objective:** Data on the distribution of ß-thalassemia mutations in Arab populations are usually destined to disparate locations and much of these become increasingly difficult for an average researcher to locate. That is why we aimed at establishing an electronic database network, called Deniz, for ß-thalassemia allele frequency distributions in the Arab world at http://biobase.fatih.edu.tr.

**Methods:** The scheme of the database combines the benefits of the relational and hierarchical systems. Detailed statistics of the frequencies of ß-thalassemia mutations are retrieved in tabular forms. Multiple permanent connections allow flexible movement within the database. Queries are processed by the system's language and sent to the user's browser as hypertext markup language documents.

**Results:** The database catalogues the frequencies of ß-thalassemia mutations in 14 Arab countries as pooled from the analysis of 3,138 chromosomes by 36 laboratories. Of the 57 ß-globin gene mutations reported in Arabs, IVS-I-110 (G-A), IVS-I-5 (G-C), IVS-I-6 (T-C), IVS-II-1 (G-A), and IVS-I-1 (G-A) are the most encountered and they account for approximately two thirds of the Arab chromosomes registered in Deniz.

**Conclusion:** In addition to its importance as a hub of updated information on the distribution of ß-thalassemia mutations in Arabs, information in Deniz may be used to predict diagnostic strategies that may be offered to natives of unstudied countries. Incidence data may also give important clues on the possible origins of ß-thalassemia in the Arab world. The integration of Deniz with other databases is currently in process and researchers are invited to contribute to the growth of the database.

**Saudi Med J 2003; Vol. 24 (11): 1192-1198**

The fact that biological data are being produced at a phenomenal rate made computers indispensable to biological research. This unexpected union between biomedical sciences and computational techniques lead to the birth of the science of bioinformatics. The main aims of bioinformatics are 3-fold: (a) The organization of biological data in a way that grants researchers access to the existing information and submit new entries promptly, (b) the development of resources that aid in the analysis of data, and (c) the interpretation of results in a biomedically meaningful manner.[1] To fulfill these aims, vast amounts of information are stored in information repositories called databases. Database types differ considerably depending, primarily, on the nature of the information stored. In biomedical applications, the hierarchical and relational schemes are widely used. A relational database stores data in 2-dimensional tables that embody different

From the Department of Biology, Bioinformatics Unit, Fatih University, Buyukcekmece, Istanbul, *Turkey*.

Received 18th May 2003. Accepted for publication in final form 9th August 2003.

Address correspondence and reprint request to: Prof. Ghazi O. Tadmouri, Department of Biology, Fatih University, 34500 Buyukcekmece, Istanbul, *Turkey*. Tel. +90 (535) 8232219. Fax. +90 (212) 8890832. E-mail: tadmouri@hotmail.com

aspects of the data. Different tables, located in different files, may then be combined as long as they have common entries. Unlike the flexible relational model, the hierarchical model is usually built with permanent hierarchical connections that are defined when the database is made. A hierarchical database can be simply thought of as a tree with permanent one-to-many connections between a parent file and offspring files.[2] The gene encoding the ß-globin polypeptide has long been a paradigm for many studies. Mutations involving the ß-globin gene are the most common cause of genetic disorders of humans. Many of these mutations result in a disorder known as ß-thalassemia, affecting millions of people around the globe.[3] Extensive and reliable data on the distribution and exact incidence of more than 200 ß-thalassemia mutations in various world populations were made available during the last 20 years. These mutations are population-specific, different sets of which are found in Mediterraneans, Asian Indians, Blacks, and Chinese.[4] We estimate that out of 7,000 international scientific articles discussing thalassemia (1966-2003), over 300 papers dealing with systematic ß-globin gene characterization in human populations have been published. Approximately 50 of these papers dealt with the distribution of ß-thalassemia mutations in Arab countries. Since access to all this information is beyond the capacity of many individual investigators, we aimed at establishing a database network for ß-thalassemia allele frequency distributions in the Arab world. The database was named Deniz, a Turkish name that means 'sea' in English or 'thalassa' in Greek, since thalassemia was also known as the 'Anemia of the (Mediterranean) Sea'. The database is currently available at the BioBase Server located in the Bioinformatics Unit at Fatih University (http://biobase.fatih.edu.tr) and can be accessed using a standard World Wide Web browser.

**Methods.** The idea to build the Deniz database went through a number of important stages. The first of these was initiated in 1994 by the tabulation of data on ß-thalassemia alleles in 5 Arab countries extracted from 8 studies.[5] A second edition published 5 years later, included data for 13 countries extracted from 19 studies.[6] Two major updates on the distribution of ß-thalassemia alleles in Arab populations were conducted in years 2001 and 2003 by means of extensive literature search. Out of several hundred papers, approximately 50 were collected, carefully reviewed, and evaluated. Single case reports were avoided and only large-scale studies were utilized. To avert the use of biased information and the inclusion of studies containing shared data in the database, 2 main precautions were taken into consideration: 1. Papers written by the same authors were reviewed and only the latest or the most extensive of these series was selected. 2. Similarly, collaborative studies sharing one or more identical author names were avoided; only one of such papers was usually selected. The subsequent publication of numerous articles on the molecular pathology underlying ß-thalassemia in Arab populations made the maintenance of the original table a difficult task and necessitated the build-up of a flexible digital copy that is user friendly and easy to update. Due to its capabilities in building and maintaining tabular sheets, Microsoft Excel was initially used to construct an original digital version of the database. Data were then exported to FileMaker Pro 6, which has important capabilities for database building, calculating statistical data, and publishing the material on the World Wide Web with minimal resources. At first, a prototype database was built using limited amount of data to decide on the best design structure. Such a strategy proved to be essential since many modifications were soon implemented after initial trials. One of the most radical adjustments applied is the conversion of the database from an extended structure to a compact one. The initial extended structure consisted of 15 minor databases, each of which is specific for a unique Arab population. Data from all these databases were retrieved by a single major relational database to calculate the overall data for Arab countries in one melting pot. Such a system suffered from a slow performance due to the considerable time spent by the major database to calculate all entries residing in several files. The system will run extremely slower if the database is to include data from world populations in the near future. Accordingly, the new design of Deniz depends on a single database file that includes several layouts, one for every Arab country (**Figure 1**). A single layout contains more than 200 records; each presenting data related to a single ß-thalassemia allele. Data within a record are organized in a tabular form with several rows each of which reveals data extracted from a single reference. Cells in each row form basic units called fields and are organized in the following order: 1. A numerical field presenting the number of chromosomes carrying a specific mutation as reported in a single study. 2. A second numerical field displaying the total number of chromosomes analyzed in that same study. 3. A textual field displaying the family name of the senior author in the study. This field can be clicked to take the visitor to a bibliographic database that includes information related to that specific reference, including an abstract and a PubMed Internet link to access the full-text of the paper, if available. 4. The fourth and last field of the row is a simple numerical field that states the year of publication. Below the table are 3 additional calculation fields. The first calculates the total number of chromosomes carrying a specific mutation for a single country as reported in all corresponding references. The second field calculates the overall total number of chromosomes analyzed in all references for that same country. The third calculation field computes a percentage based on the results of the

2 previous fields (**Figure 1**). In the same database file, a major layout retrieves data from all minor country layouts. Basically, the major layout has a similar structure to minor layouts. The main difference is that statistics in the major layout represent overall data supplied by all references used in the database. Special buttons are also available to reach detailed country data. To overcome the limitation of the hierarchical model, we allowed a flexible movement in the tree walk by multiple connections permanently available to enable the user to access a particular record, layout, or file without traversing the entire hierarchy above that file. In this case, the 'many-to-one' relationship is used. For example, while viewing data for the IVS-I-110 (G-A) mutation in a certain country, the user may chose to move 5' or 3' in the ß-globin gene to view statistics for flanking mutations or jump to another location in the ß-globin gene. Pressing a "view all" button generates a hypertext markup language (HTML) file that displays the full spectrum of ß-thalassemia mutations recorded in the corresponding country. In addition, another button takes the visitor to a "ß-Thalassemia Allele" information knowledge-based that is currently under construction.

**Results.** In contrast to the relatively simple structure of biosequence data, based on a linear 4-character language (namely A, C, G, and T), the structure of knowledge-based digital libraries is very complex and necessitates the implementation of different types of information techniques.[2] The nature of results on the distribution of ß-thalassemia mutations, and the complexity of the data necessitates some unique demands on database design and development. For this reason, Deniz cannot be ascribed to a single database model but is a compound system, both relational and hierarchical: 1. Deniz is a relational database since it is based on the concept of tabulating country data in records made of multiple fields that store the various pieces of information on that country. Available tables, or records, are then combined with as many layouts as necessary. Although the database program is expected to spend a large amount of processor time to extract information from the system and to calculate statistical data, this is not observed in Deniz. This may be due to the maturity, stability, and reliability of such a model that has been adapted in a number of different applications in bioinformatics handling large amounts of data.[2] 2. Deniz is also a hierarchical database since the smallest data entities (namely fields and records) are linked to form an inverted tree with a parent layout. The relationship between parent and children is one-to-many connection, in that one parent may produce multiple children. This will be an important source of flexibility if the database is to contain future data from additional countries or other world populations. The Deniz database currently includes data on ß-thalassemia mutation frequencies in 14 Arab countries, namely, Algeria, Bahrain, Egypt, Jordan, Kuwait, Lebanon, Morocco, Oman, Palestine, Saudi Arabia, Syria, Tunisia, United Arab Emirates, and Yemen. An additional database offers statistics on ß-thalassemia alleles in Bedouins since they cannot be categorized in a specific country. So far, no studies have been reported on the characterization of ß-thalassemia alleles in Mauritania, Libya, Sudan, Somalia, Djibouti, Qatar, and Iraq. Data obtained for the Comoros were not enough to establish proper records. According to the data pooled in Deniz, approximately 36 research groups have analyzed a total of 3,138 Arab ß-thalassemia chromosomes (**Table 1**). At least 57 ß-globin gene mutations have been reported (36 point mutations, 13 small deletions, 3 insertions, and 3 large deletions of 25 bp, 290 bp, and 619 bp). Data for the mutation causing sickle cell disease were not included in Deniz due to the inconsistency in reporting this allele in the literature. The most common mutations encountered in Arabs are IVS-I-110 (G-A), IVS-I-5 (G-C), IVS-I-6 (T-C), IVS-II-1 (G-A), and IVS-I-1 (G-A). These mutations account for approximately 66.2% of all Arab ß-thalassemia chromosomes registered in Deniz (**Table 2**). The highest heterogeneity of ß-thalassemia mutations is encountered in Egypt, Jordan, United Arab Emirates, and Lebanon, where, almost 20 different mutations occur in each (**Table 1**). Of all studied ß-thalassemia chromosomes, 6% could not be linked to a specific ß-globin mutation and remain uncharacterized (**Table 2**).

**Discussion.** The recent evolution in information technology and the speed by which scientific results are communicated makes it increasingly difficult for an average researcher to locate material on the vast and unorganized web. Due to this complexity, the need to organize, store, retrieve, and analyze large amounts of information in computerized databases is becoming increasingly important. Luckily, scientists working in the fields of biomedical sciences are exceedingly well served by many high quality databases. Gaining access to these is usually a top priority for any success.[7] Research on the molecular basis of ß-thalassemia started more than 25 years ago. The limited knowledge of the ß-globin gene mutations and the 'weakness' of molecular biology techniques available at that time (namely, slow work cycle, requirement of a large workforce, and expensive expendable) limited early reports to only a handful of mutations described in small numbers of individuals. In most of the cases, importance was given to the characterization of novel mutations. Despite these factors, several groups, who reported limited spectra for ß-thalassemia mutations in several populations, recorded several significant contributions at that time.[4] The advent of the polymerase chain reaction (PCR), the subsequent development of more convenient DNA

**Table 1 -** Total numbers for ß-thalassemia mutations and chromosomes analyzed in Arab populations according to Deniz.

| Countries | Mutation | Chromosomes |
|---|---|---|
| Algeria | 17 | 301 |
| Bahrain | 12 | 66 |
| Bedouins | 3 | 12 |
| Egypt | 22 | 383 |
| Jordan | 20 | 369 |
| Kuwait | 10 | 96 |
| Lebanon | 19 | 299 |
| Morocco | 2 | 72 |
| Oman | 17 | 269 |
| Palestine | 12 | 214 |
| Saudi Arabia | 11 | 276 |
| Syria | 17 | 177 |
| Tunisia | 14 | 165 |
| United Arab Emirates | 20 | 424 |
| Yemen | 2 | 15 |

**Table 2 -** Frequencies of the most common ß-thalassemia mutations and uncharacterized alleles in Arab populations.

| Mutations | Type | % |
|---|---|---|
| IVS-I-110 (G-A) | ß+ | 19.6 |
| IVS-I-5 (G-C) | ß+ | 16.0 |
| Cd 39 (C-T) | ß0 | 10.2 |
| IVS-I-6 (T-C) | ß+ | 7.2 |
| IVS-II-1 (G-A) | ß0 | 6.8 |
| IVS-I-1 (G-A) | ß0 | 6.4 |
| Uncharacterized | ? | 6.4 |

A, C, G and T - the relatively simple structure of biosequence data based on a linear 4-character language

analysis methods, and the continuous accumulation of knowledge on ß-globin gene mutations, gave a great impetus to the rapid screening of large numbers of individuals. Hence, many laboratories in the world either generated or revisited their statistics with considerably larger numbers of individuals. No doubt, such studies reported results that are more accurate and paved the way to a rapidly growing knowledge of the ß-globin gene alterations in many world populations. Concurrently, many DNA analysis techniques became easily accessible leading to the establishment of advanced laboratories or collaborations in countries with limited resources.[8] This surely allowed many research groups to report considerably large amounts of data on the distribution of ß-thalassemia alleles in many world populations, including Arabs. Because of the disparate locations to which reported data are usually destined (namely, international publications, national articles, or conference abstracts), much of these become 'technically' not available to other clinicians or investigators. An important observation supporting this view is the consistent bias to refer to outdated information on the frequencies of ß-thalassemia mutations in a specific country, while updated information tend to remain unrecognized. For example, a large number of reports published after year 2000 and referring to the spectrum of Lebanese ß-thalassemia mutations depended on the original work of Chehab et al[9] published in 1987 despite the fact that a more extensive research was made available by Zahed et al.[10] The former study included 50 chromosomes that carried 8 ß-thalassemia mutations while the latter reported 17 mutations in 202 chromosomes; hence, reflecting the reality of the heterogeneity of the Lebanese population.[10] Out of this need, the idea to construct the Deniz database crystallized with the aim to condense all the information into clear trends and facts that users can readily understand. Knowing that getting data into a database is of little value unless the data can also be retrieved, we implemented data extraction methods with Web-page interfaces. In such a case, the user issues a command that is interpreted in the BioBase server. Results of the database query are then processed by the system's language and sent to the user's browser as HTML documents. Hence, access to the database is provided without any additional software and without having to "subscribe". Another application of HTML was implemented to capture complex tables with nested structures in single static diagrams (**Figure 1**). Although this could be considered as a drawback, in view of the limitations imposed by the capabilities of the www protocols, we believe that the future implementation of especially designed Java Scripts will soon allow a more flexible presentation of dynamic results without the need to the FileMaker interface. Similarly, more queries that are complex may also be formulated. Deniz is currently on the way to be integrated with 2 additional

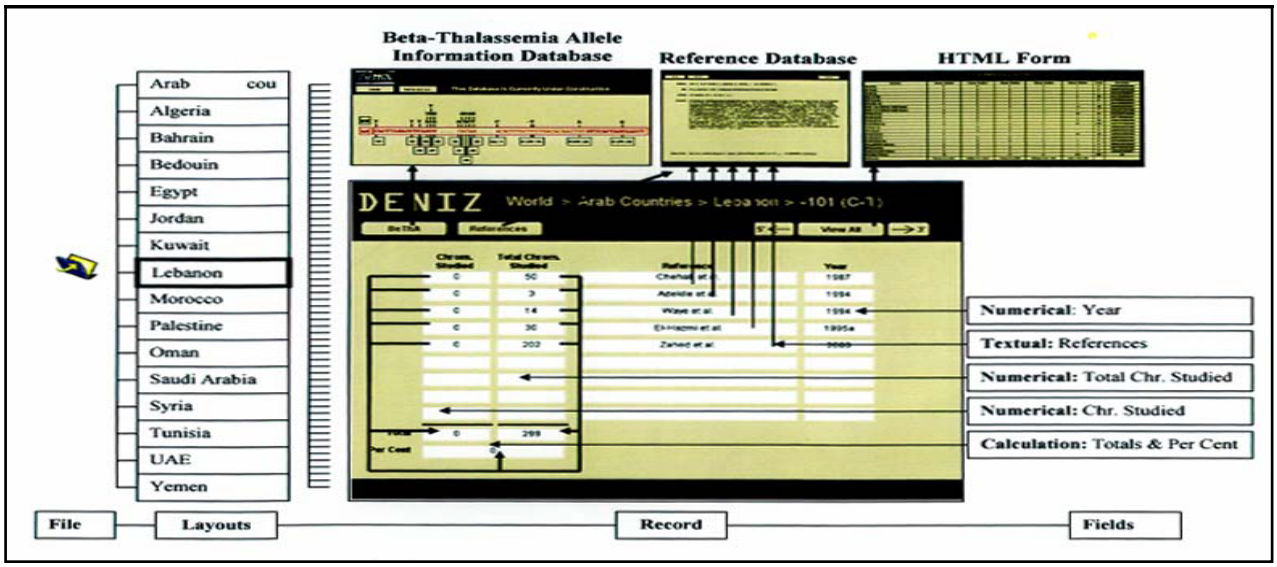


**Figure 1 -** An overview of the Deniz database structure. HTML - hypertext markup language. UAE - United Arab Emirates, Chr - chromosomes

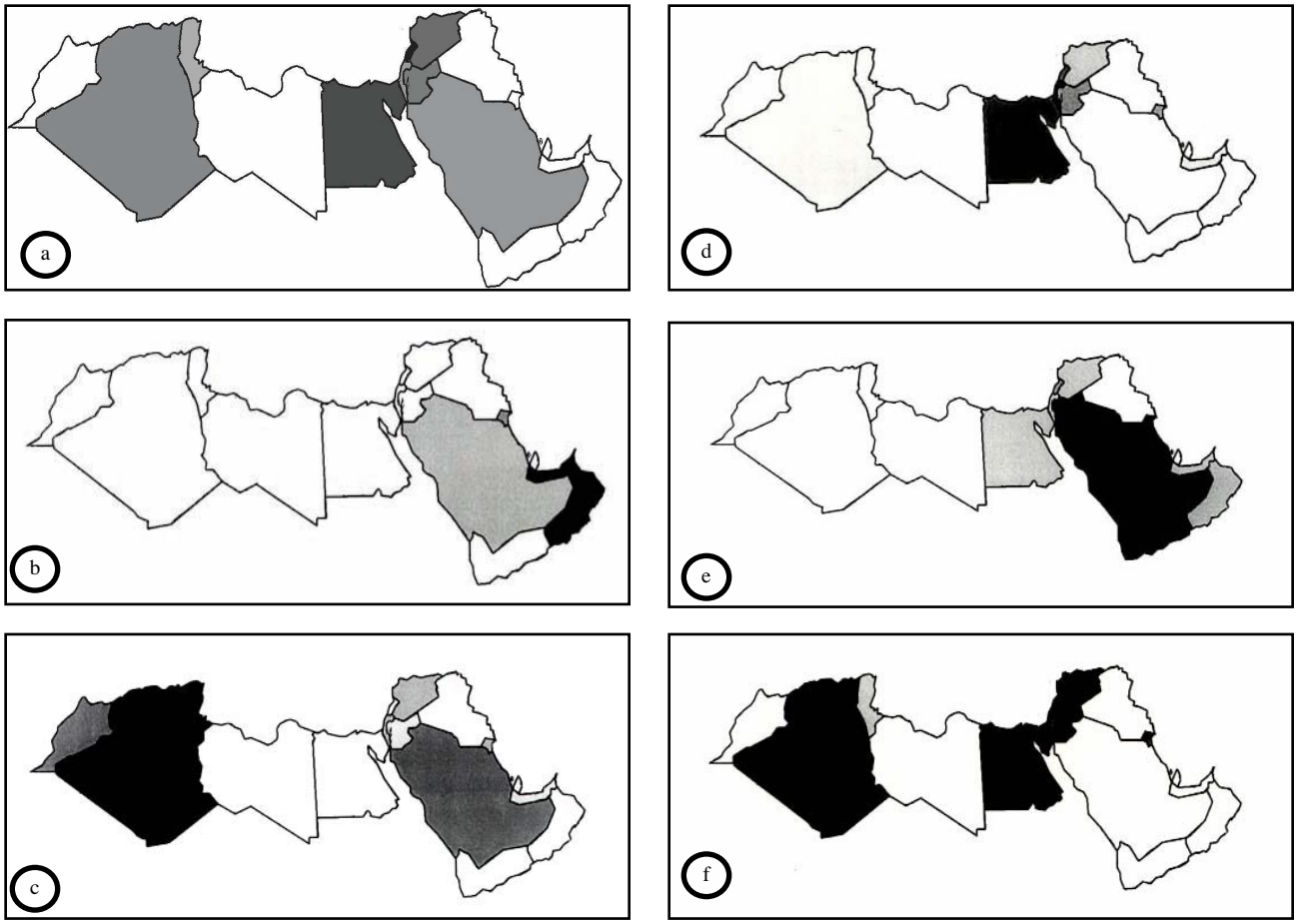


**Figure 2 -** Relative frequency distributions of the most common ß-thalassemia mutations in Arab countries: **a)** IVS-I-110 (G-A), **b)** IVS-I-5 (G-C), **c)** Cd 39 (C-T), **d)** IVS-I-6 (T-C), **e)** IVS-II-1 (G-A) and **f)** IVS-I-1 (G-A). Darker regions indicate higher frequencies for the allele in the corresponding countries. A, C, G and T - the relatively simple structure of biosequence data based on a linear 4-character language

components: 1. BeThA - The ß-thalassemia allele information database that will soon offer a multitude of information on the phenotype-genotype correlations, population genetics of ß-thalassemia alleles, graphical presentations, and direct links to corresponding information at the Globin Gene Server,[11] the Human Gene Mutation Database, and OMIM™,[7] and 2. a bibliographical database indexing references related to ß-thalassemia mutations. We believe that this version of the database contains what we feel is a very high level of experimental detail. Further experience should test whether this is excessive or appropriate and whether a less detailed scheme would be adequate. Beside its importance as a hub of updated information on the distribution of ß-thalassemia mutations in Arab people, Deniz may also have other valuable applications. A first sight on the geographical distribution of chromosomes pooled by Deniz reveals that molecular data are yet unavailable for 7 Arab countries despite indications for a high prevalence of ß-thalassemia in their populations.[12-15] Accordingly, research laboratories willing to analyze subjects originating from these countries may benefit from the information available to save a lot of time when a rapid diagnosis is needed and resources are limited. One such expectation could be to analyze Libyans for the IVS-I-110 (G-A), Cd 39 (C-T), and IVS-I-1 (G-A) mutations with a possibility of more than 50% to achieve a molecular diagnosis. The choice of the selected mutations depended on their frequent observations in neighboring Egypt, Tunisia, and Algeria. Similarly, a strategy to analyze Sudanese patients may be built dependent upon statistics from the neighboring Egypt, Saudi Arabia, and Yemen. Iraqis may then be analyzed according to the common mutations in Syria, Jordan, KSA, Kuwait, and Iran (**Figure 2**). Another possible application of Deniz falls in the field of population genetics. The considerable number of ß-thalassemia mutations recorded in Arab populations confirms the heterogeneity of Arabs because of multiple origins and history-long admixtures with other populations. With the exception of Yemenites and Moroccans, for whom limited chromosomes have been analyzed, almost all Arab countries may be considered as heterogeneous irrespective of the size of their populations. For this we see 10 ß-thalassemia mutations in Kuwait, 12 in United Arab Emirates, 14 in Tunisia, and 19 in Lebanon (**Table 1**). In each case, the heterogeneity can be easily attributed to the population admixtures in the respective countries. Similarly, a careful examination of the frequency distribution of the common ß-thalassemia mutations in Arab populations gives important clues on the possible history and origins of such mutations.

***1. IVS-I-110 (G-A).*** Is a severe ß$^+$-thalassemia mutation,[4] commonly encountered in Arabs. One of every 5 Arab ß-thalassemia chromosomes carries IVS-I-110 (**Table 2**). The highest frequency reported for this mutation is in Lebanon, where the incidence is 41.5%. This figure declines when moving South to Egypt, then West in North Africa. IVS-I-110 is seldom observed in Gulf countries (**Figure 2a**). Such data are highly indicative for an Eastern Mediterranean origin for IVS-I-110. Recent reports regarding restriction fragment length polymorphisms (RFLPs) and sequence haplotype data for IVS-I-110 in Lebanon demonstrated its association with a single RFLP and 2 sequence haplotypes.[16] In contrast, in Turkey, IVS-I-110 is associated with 6 distinct sequence haplotypes and 4 distinct RFLP haplotypes, suggesting that the mutation probably emerged there. Following its emergence in Turkey few thousand years ago, the IVS-I-110 mutation was probably introduced into the Arab world by migrations or settlements in Lebanon. Subsequently, the mutation was later transported to Egypt and North Africa possibly through Phoenician trade routes.[17]

***2. IVS-I-5 (G-C).*** In contrast to the IVS-I-110 (G-A) mutation, the geographical distribution of IVS-I-5 in the Arab world is almost totally restricted to Gulf countries (**Figure 2b**). Single cases carrying this mutation were reported in neighboring countries (Jordan, Lebanon, Egypt, and Algeria). World data clearly indicate the Southern Asian origin of the IVS-I-5. It's introduction to the Gulf countries, thus, could be easily attributed to the Spice trading routes established a few thousand years ago between the Arabs and South Asian populations. This possible maritime route is best supported by the observation that the most affected countries are Oman and United Arab Emirates, both at the southeastern side of the Gulf much close to South Asia than other countries (**Figure 2**).

***3. Cd 39 (C-T).*** Despite its lower occurrence in the Arab world, the distribution of the Cd 39 mutation seems to be the most widely spread mutation since it was observed in almost all Arab countries. Cd 39 is primarily classified as a Western Mediterranean abnormality, this explains the high incidences for this mutation in Tunisia, Algeria, and Morocco (**Figure 2c**). Haplotype data from Algeria support this proposal.[18] However, the many recently reported cases of Cd 39 occurrence in natives of Gulf countries (especially, Bahrain and KSA) raises questions on the possible origin and age for this allele (**Figure 2c**).

***4. IVS-I-6 (T-C).*** While the highest frequencies for this allele in Europe are recorded in Portugal, Italy, and Yugoslavia,[6] high occurrence is also noted in the Eastern Mediterranean Arab countries especially Palestine, Egypt, and Lebanon (**Figure 2d**). In the latter, several RFLP haplotypes were described in contrast to the relative homogeneity in Europeans;[19] hence, demonstrating an older origin for this mild mutation in the Eastern Mediterranean.

***5. IVS-II-1 (G-A).*** Similar to IVS-I-5, IVS-II-1 seems to be almost restricted to the Arabian peninsula, with less frequent occurrence in the neighboring

countries such as Lebanon, Syria, Palestine, and Egypt. This mutation is totally absent in analyzed North African populations (**Figure 2e**). Data from RFLP haplotypes indicate the association of this mutation with multiple backgrounds in world populations. This observation may suggest possible multiple origins for this mutation[20] among which one could have emerged in the southwestern part of the Arabian peninsula, and subsequently transported to neighboring areas by trading routes well established some 2000 years ago.

***5. IVS-I-1 (G-A).*** This mutation shows a restricted geographical distribution among Eastern Mediterranean Arab countries (Syria, Lebanon, Jordan, Palestine, and Egypt). Smaller numbers of chromosomes were reported elsewhere (**Figure 2f**). The haplotype homogeneity and the restricted distribution favor a recent unicentric origin for IVS-I-1 confined to the Eastern Mediterranean.[17]

The conclusion drawn from the distribution or origin correlation of the most frequent Arab ß-thalassemia mutations proves the necessity to extend Deniz to comprise a larger spectrum of world populations. That is why efforts to extend the coverage of Deniz started during the preparation of this manuscript. It is evident that a strong commitment to the current database and a considerable expansion are needed. Such expansion is not only restricted to the data offered by the database, but also to allow more tasks to be performed too (namely, statistical tools, graphical presentation of data, integrate links to other reference databases).

The task of following continuously produced world results for ß-thalassemia mutations are demanding. That is why we would like to call researchers to contribute with their data as part of the continued maintenance and growth processes of the database, thereby allowing the database to expand as a result of a community effort. Such investigator-initiated entry will surely be accomplished with strict editorial review.

**Acknowledgment.** We express our gratitude to Associate Prof. P. Perrin (University of Claude-Bernard, Lyon I, France) and Dr. F. Le Mort (Maison de l'Orient Méditerranéen - Jean Pouilloux – Lyon, France) for their generous support in the establishment of the database system. We would like also to thank Asst. Prof. N. Bissar-Tadmouri for critically reviewing this manuscript. Facilities of the Department of Biology at Fatih University, Istabul, Turkey were utilized.

## References

1. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? In: Haux R, Kulikowski C, editors. Yearbook of Medical Informatics 2001. Stuttgart (Germany): Schattauer Publishing Company; 2001. p. 83-99.
2. Bergeron B. Bioinformatics Computing. New Jersey (USA): Pearson Education; 2003.
3. The World Health Organization. Community control of hereditary anemias: Memorandum from a WHO meeting. ***Bull WHO*** 1983; 61: 63-80.
4. Huisman THJ, Carver MFH, Baysal E. A syllabus of thalassemia mutations (1997). Augusta (USA): The Sickle Cell Anemia Foundation; 1997.
5. Tadmouri GO. A molecular investigation of beta-thalassemia in the district of Istanbul: Application to prenatal diagnosis [dissertation]. Istanbul (Turkey): Bogazici Univ.; 1994.
6. Tadmouri GO. ß-Thalassemia in Turkey: Distribution, diversity, evolution and genotype-phenotype correlations [dissertation]. Istanbul (Turkey): Bogazici Univ.; 1999.
7. The National Center for Biotechnology Information (USA). The NCBI Handbook. Bethesda (USA): NCBI; 2003.
8. El-Kalla S, Mathews ARA. Significant ß-thalassemia heterogeneity in the United Arab Emirates. ***Hemoglobin*** 1997; 21: 237-247.
9. Chehab FF, Der Kaloustian V, Khouri FP, Deeb SS, Kan YW. The molecular basis of ß-thalassemia in Lebanon: Application to prenatal diagnosis. ***Blood*** 1987; 69: 1141-1145.
10. Zahed L, Qatanani M, Nabulsi M, Taher A. ß-Thalassemia mutations and haplotype analysis in Lebanon. ***Hemoglobin*** 2000; 24: 269-276.
11. Hardison R, Riemer C, Chui DHK, Huisman THJ, Miller W. Electronic access to sequence alignments, experimental results, and human mutations as an aid to studying globin gene regulation. ***Genomics*** 1998; 47: 429-437.
12. Deyde VM, Lo BB, Khalifa IO, Ly B, Ball A, Fattoum S. Epidemiological profile of hemoglobinopathies in the Mauritian population. ***Ann Hematol*** 2002; 81: 320-321.
13. Jain RC. Sickle cell and thalassaemic genes in Libya. ***Trans R Soc Trop Med Hyg*** 1985; 79: 132-133.
14. Bayoumi RA, Abu Zeid YA, Abdul Sadig A, Awad Elkarim O. Sickle cell disease in Sudan. ***Trans R Soc Trop Med Hyg*** 1988; 82: 164-168.
15. Al-Mondhiry H, Al-Hilali A. Severe anemia. Clinical observations in 100 patients with very low hemoglobin levels. ***Arch Intern Med*** 1979; 139: 1053-1054.
16. Zahed L, Demont J, Bouhass R, Trabuchet G, Hanni C, Zalloua P et al. Origin and history of the IVS-I-110 and codon 39 ß-thalassemia mutations in the Lebanese population. ***Hum Biol*** 2002; 74: 837-847.
17. Tadmouri GO, Garguier N, Demont J, Perrin P, Basak AN. History and origin of ß-thalassemia in Turkey: Sequence haplotype diversity of ß-globin genes. ***Hum Biol*** 2001; 73: 661-674.
18. Perrin P, Bouhassa R, Mselli L, Garguier N, Nigon VM, Bennani C et al. Diversity of sequence haplotypes associated with ß-thalassemia mutations in Algeria: Implications for their origin. ***Genet*** 1998; 213: 169-177.
19. Flint J, Harding RM, Clegg JB, Boyce AJ. Why are some genetic diseases common? Distinguishing selection from other processes by molecular analysis of globin gene variants. ***Hum Genet*** 1993; 91: 91-117.
20. Chifu Y, Nakashima H, Hara H, Yokota E, Imamura T. ß-Thalassemia major resulting from a compound heterozygosity for the ß-globin gene mutation: Further evidence for multiple origin and migration of the thalassemia gene. ***Hum Genet*** 1992; 89: 343-346.