

FINDbase worldwide database for clinically relevant genomic variation allele frequencies

National and Ethnic Genomic Databases comprise a well-defined niche of genomic databases, aiming to record the prevalence of alleles, mainly pathogenic and of clinical relevance, but also benign, in different populations and ethnic groups worldwide in a structured manner. These resources can be used not only in population genetic studies to study gene/mutation flow and admixture patterns, human demographic history but also in Genomic Medicine, e.g. to stratify national molecular diagnostic services or to rationalize drug use. These resources can nicely complement the data content of either central (or core) databases and/or locus-specific databases.

Fig. 1. Home page of FINDbase worldwide database for clinically relevant genomic variation allele frequencies.

In 2003, the first National and Ethnic Genomic Databases were realized in a very primitive structure, mainly to serve individual populations and ethnic groups. Couple of years later, and in an effort to maintain a homogeneous data content, a dedicated software [*ETHNOS*] was developed, allowing the development and curation of these resources. In 2006, FINDbase (Frequency of INherited Disorders database; <http://www.findbase.org>) was established as a worldwide database with the vision to comprehensively document the prevalence of clinically relevant genomic variation allele frequencies in various populations and ethnic groups worldwide. FINDbase contains data only in an aggregated manner, *i.e.*, allele frequencies stripped from any sensitive personal information of their carriers, in

order to maintain anonymity. FINDbase accommodates data pertaining to the prevalence of (a) pathogenic genomic variants, leading to monogenic disorders, (b) pharmacogenomic biomarkers for drug dose adjustment to achieve the goals of individualized treatment, and (c) the various rare genetic diseases that are documented in various populations. The above are documented in distinct data modules. Content-wise, FINDbase is the richest database among all National and Ethnic Genomic Databases currently available and, it is considered to be one of the key resources to retrieve population-specific information for clinically relevant genomic variations, as indicated by the number and origin of visitors.

FINDbase content includes several clinically relevant genomic data allele frequencies in over 90 populations and ethnic groups worldwide. In order to give a visual first impression of FINDbase data content and, at the same time, provide a quick access to the available data per population and per module, we have built a world map that is readily available at the Home page

Furthermore, the ETHNOS software has been upgraded as an “off-the-shelf” web application for National and Ethnic Genomic Database development and curation. This offers a unique opportunity to interested users, with minimal or even no (bio)informatics background, either to start a new National Genomic Database from scratch or even to “adopt” an existing one from the FINDbase collection and, together with a group of expert curators from their own country, to assume responsibility of enriching it with content, with all technical support being provided centrally by FINDbase administrators. This also contributes to harmonized National and Ethnic Genomic Database development and minimizes data content heterogeneity. The upgraded *ETHNOS* software complies with the established recommendations and guidelines to develop nation-wide projects to document the genetic heterogeneity in various countries.

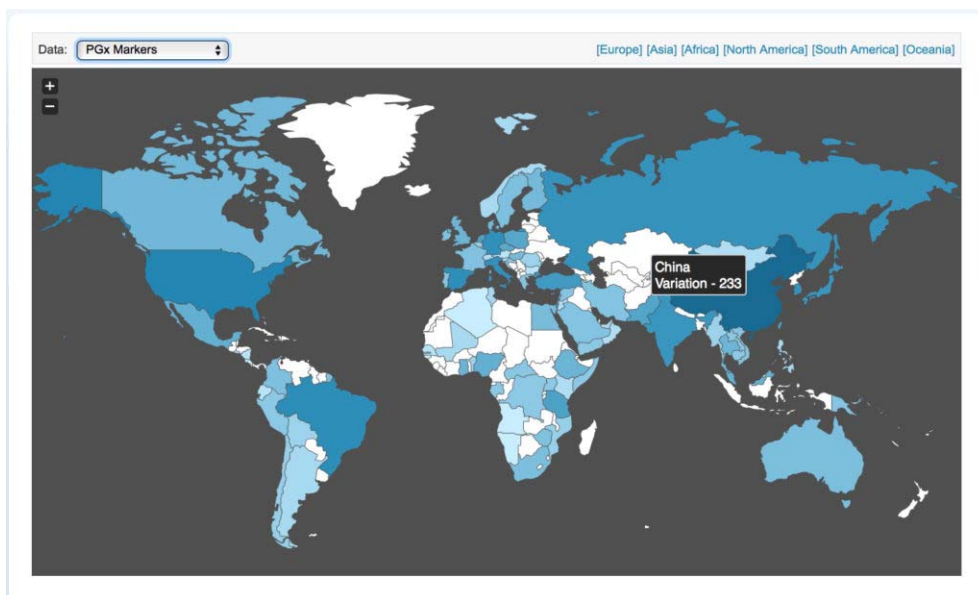


Fig. 2. Overview of the FINDbase map display visualization tool.

Ever since its establishment in 2006, FINDbase is considered a reference repository of such information worldwide. To maximize its utility, access of FINDbase content is kept free of charge, while both the data content and the battery of data visualization tools have been significantly enriched. This was made possible thanks to the uninterrupted funding since 2005, as part of major research Consortia, such as GEN2PHEN, RD-Connect and the Genomic Medicine Alliance, always interacting with the field's leading experts. FINDbase is currently being further upgraded to accommodate the anticipated increased data influx and its architecture gets further modernized, ensuring that it will always remain as up-to-date as possible.

George P. Patrinos

*University of Patras, School of Health Sciences, Department of Pharmacy,
University Campus, Rion, Patras, Greece*

Publication

[Expanded national database collection and data coverage in the FINDbase worldwide database for clinically relevant genomic variation allele frequencies.](#)

Viennas E, Komianou A, Mizzi C, Stojiljkovic M, Mitropoulou C, Muilu J, Vihinen M, Grypioti P, Papadaki S, Pavlidis C, Zukic B, Katsila T, van der Spek PJ, Pavlovic S, Tzimas G, Patrinos GP
Nucleic Acids Res. 2017 Jan 4