Consanguinity studies and genome research in Mediterranean developing countries

Giovanni Romeo, Alessandro Gialluisi and Tommaso Pippucci

University of Bologna Medical School, Bologna, Italy

Correspondence to Giovanni Romeo, Unitá Operativa di Genetica Medica, Dipartimento di Scienze Ginecologiche, Ostetriche e Pediatriche, Policlinico Sant'Orsola Malpighi, Universitá di Bologna, Via Massarenti 9, 40136, Bologna, Italy
Tel: + 0039 051 2088420; fax: + 0039 051 5870611; e-mail: romeo@eurogene.org

Received 2 January 2011
Accepted 30 March 2011

Middle East Journal of Medical Genetics
2012, 1:1–4

Purpose
Classical studies of consanguinity have taken advantage of the relationship between the gene frequency for a rare autosomal recessive disorder (q) and the proportion of offspring of consanguineous couples who are affected with the same disorder. The Swedish geneticist Gunnar Dahlberg provided the first theoretical formulation of the inverse correlation between q and the increase in frequency of consanguineous marriages among parents of affected children with respect to marriages of the same degree in the general population. Today it is possible to develop a new approach for estimating q using mutation analysis of affected offspring of consanguineous couples. The rationale of this new approach is based on the possibility that the child born of consanguineous parents carries the same mutation in double copy (true homozygosity) or alternatively carries two different mutations in the same gene (compound heterozygosity). In the latter case the two mutations must have been inherited through two different ancestors of the consanguineous parents (in this case the two mutated alleles are not 'identical by descent').

Patients and methods
Data from the offspring of consanguineous marriages affected with different autosomal recessive disorders were collected by different molecular diagnostic laboratories in Mediterranean countries and in particular in Arab countries, where the frequencies of consanguineous marriages is high, show the validity of this approach.

Results
The proportion of compound heterozygotes among children affected with a given autosomal recessive disorder, born of consanguineous parents, can be taken as an indirect indicator of the frequency of the same disorder in the general population. Identification of the responsible gene (and mutations) is the necessary condition to apply this method.

Conclusion
The following paper from our group relevant for the present review is being published: Alessandro Gialluisi, Tommaso Pippucci, Yair Anikster, Ugur Ozbek, Myrna Medlej-Hashim, Andre Megarbane and Giovanni Romeo: Estimating the allele frequency of autosomal recessive disorders through mutational records and consanguinity: the homozygosity index (HI) annals of human genetics (in press; acceptance date 1 November 2011) In addition, our experimental data show that the causative mutation for a rare autosomal recessive disorder can be identified by whole exome sequencing of only two affected children of first cousins parents, as described in the following recent paper: Pippucci T, Benelli M, Magi A, Martelli PL, Magini P, Torricelli F, Casadio R, Seri M, Romeo G EX-HOM (EXome HOMozygosity): A Proof of Principle. Hum Hered 2011; 72:45-53.

Keywords:
autosomal recessive disorders, exome homozygosity, genomics of consanguinity, homozygosity index

Middle East J Med Genet 1:1–4
© 2012 Middle East Journal of Medical Genetics
2090-8571

The past approaches based on genetic epidemiology

Classical studies of consanguinity have taken advantage of the relationship between the frequency of a given autosomal recessive disorder and the proportion of offspring of consanguineous couples who are affected with the same disorder. The Swedish geneticist Gunnar Dahlberg provided the first theoretical formulation of the inverse correlation between the gene frequency for rare autosomal recessive disorders and the increase in frequency of consanguineous marriages among parents of affected children with respect to marriages of the same degree in the general population. This approach allows calculating the frequency of autosomal recessive disorders in an unbiased way that is free of the problems arising

Copyright © Middle East Journal of Medical Genetics. Unauthorized reproduction of this article is prohibited.

from incomplete ascertainment. Complete ascertainment means a complete survey of all individuals affected with a given disorder, which is practically impossible in most populations, especially those of developing countries.

Therefore, the three parameters that have been taken into account in classical consanguinity studies are: (a) the frequency of consanguineous parents of a given degree (e.g. first cousins) among children with a specific autosomal recessive disorder (C'); (b) the frequency of consanguineous couples of the same degree in the general population (C); and (c) the gene frequency (q) and consequently the frequency in the population of that specific autosomal recessive disorder ($q2$). If one knows two of these parameters it is possible to calculate the third one using mathematical formulas like the one used by Dahlberg and by other population geneticists. In particular, if one knows the first two parameters for any specific population (C' and C), it is possible to calculate the frequency in the population of any given autosomal recessive disorder, as done in the past in Italy for phenylketonuria, Friedreich ataxia, and cystic fibrosis (CF) (Romeo *et al.*, 1983a, 1983b, 1985; Romeo, 1984).

The problem encountered by this type of genetic epidemiology study based on consanguinity is represented by the difficulty in having a precise and reliable estimate of the second parameter mentioned above (C), namely, the frequency in the general population of consanguineous couples for any given degree (irrespective of their being parents of an affected child). In Italy, this estimate was possible up to 1964 because of the availability in Rome of centralized Church archives that kept records of all the dispensations (or permits to marry in Church) granted by the Pope for consanguineous marriages for almost 400 years. These data were collected and organized for genetic studies in 1960 by Cavalli-Sforza and co-workers (Cavalli-Sforza *et al.*, 2004). The absence of equivalent centralized reliable data to calculate this parameter (C) has made it impossible to use this approach in any other country.

The present approaches based on mutation analysis

At present, it is possible to develop a new approach for estimating the relative frequency of autosomal recessive disorders using mutation data originated by the molecular genetic centers that have diagnosed the affected offspring of consanguineous couples. The rationale of this new approach, never experimented before, is based on the possibility that the child born of consanguineous parents carries the same mutation in double copy (true homozygosity) or alternatively carries two different mutations in the same gene (compound heterozygosity). In the latter case, the two mutations must have been inherited through two different ancestors of the consanguineous parents (in this case, the two mutated alleles are not 'identical by descent').

The allelic frequency of the pathogenic alleles in the population (q) is therefore independent of consanguinity and can be taken as an indirect indicator of the frequency of the same disorder in the general population (irrespective of consanguinity). If the disease is rare, the proportion of affected children born of consanguineous parents who are true homozygotes (C'true-hom) should be more relevant with respect to that of compound heterozygotes (C'comp-het). This relationship can be measured by the ratio (C'true-hom)/(C'true-hom) + (C'comp-het), which will vary between 0 and 1 and will take values inversely proportional to the frequency of the disorder. We call this ratio 'total homozygosity index' (THI). We confirmed this hypothesis in two different mutation data sets [for CF and familial Mediterranean fever (FMF)] collected in Lebanon by Dr André Megabarné (Saint Joseph University, Beirut), and higher THIs were found in the CF sample than in the FMF sample, according to the higher prevalence of FMF compared with CF in Lebanon.

Furthermore, in a very recent paper, Ten Kate *et al.*, (2010) showed a positive correlation between (C'comp-het) and q). On the basis of this report we developed an equation in which THI, the total allelic frequency of the pathogenic alleles for a given disease (q), and the inbreeding coefficient (F) are used to infer the population prevalence of monogenic autosomal recessive diseases (paper in preparation).

The interesting aspect of this genetic epidemiological approach is that it makes use of data already existing in all the diagnostic centers of the Mediterranean Sea basin. In other words, no investment is required to produce these laboratory data, which, following the approach just summarized, will yield a ranking order of the frequencies of the most frequent among rare monogenic disorders in each country. The final outcome of this approach will consist in measurements of the THI and mutational spectra for every autosomal recessive disorder in each Mediterranean country.

Therefore, we propose to collect mutation data from the offspring of consanguineous marriages affected with different autosomal recessive disorders from different molecular diagnostic laboratories in all the Mediterranean countries and in particular from Arab countries. We will obtain in every population a ranking order according to the different THIs, the values of which will be inversely proportional to the frequency of the disorder. This approach will have the advantage over traditional descriptive epidemiology studies to generate an unbiased estimate of the relative frequency of the different autosomal recessive disorders and will be particularly useful in the Arab world where the rates of consanguineous marriages are generally high. Moreover, this approach will not need the collection of additional mutation data from the general population because it has its own built-in population control represented by the value of C'comp-het. Finally, from a decision-making point of view, this new combined approach of molecular and genetic epidemiology based on consanguinity should become useful (not only in the Arab world) to assess the opportunity of widespread genetic screening for certain

Copyright © Middle East Journal of Medical Genetics. Unauthorized reproduction of this article is prohibited.

autosomal recessive disorders with respect to others because their relative incidences will help in establishing priorities for genetic testing at the population level.

This approach based on mutation analysis in offspring of consanguineous parents can be integrated and supported by the locus specific databases (LSDBs), which have been rapidly increasing in number during the last decade (Romeo, 2010; van Baal *et al.*, 2010). Several national mutation databases for different Mediterranean populations, such as the Greek, Cypriot, Iranian, Lebanese, Israeli, Egyptian, etc. describe mutations of particular genes (for example, the globin variants and thalassemia mutations) and contain up to 50% unpublished variations with, usually, thorough phenotypic descriptions. Everyone practicing genetic counseling will agree that LSDBs are useful tools contributing toward the identification of disease-causing mutations, providing information about phenotypic patterns associated with a specific mutation and enabling researchers to define an optimal strategy for mutation detection. This explains the recent rapid growth in the number of LSDBs, now available for hundreds to thousands of human genes (sometimes with more than one database per gene).

Starting from the experience developed in the last decade, the performances of 1188 LSDBs was analyzed by Patrinos *et al.* (2010) for the presence or absence of 44 content criteria related to database features. This analysis led the authors to the conclusion that at present there is less data-content heterogeneity compared with 8 years ago when a similar analysis was performed (Claustres *et al.*, 2002).

However, the next relevant question is: how useful are the LSDBs for helping clinical genetic services in making a diagnosis? The answer from Patrinos *et al.* (2010) to this question is not equally encouraging as they state that '*current LSDBs do not support data retrieval and transmission between scientific personnel and clinicians working in diagnostic laboratories*'. Their analysis pinpoints a number of deficiencies (namely, lack of detailed disease and phenotypic descriptions for each genetic variant, etc.), which, if addressed, would allow LSDBs to better serve the clinical genetics community, patients, their families, and related associations and not just researchers. They therefore propose the concept of Clinical Genetics databases (CGDBs) and the development of a federation of network of LSDBs and CGDBs that would bridge the division between gene-centric and genome-wide approaches to databasing variation. This work should be promoted and encouraged first at the national level. In this sense, the Israeli National/Ethnic mutation database (*http://www.goldenhelix.org/server/israeli/*) is already a good model of the enhanced possibilities that other existing national LSDBs might have in providing useful information to genetic services in their own country. In contrast, the integration of tools for mutation analysis envisaged by Patrinos *et al.* (2010) may become one of the main goals of new international projects aiming at collaborative genomics for human health, as recently proposed by a group of medical geneticists from the Mediterranean region.

The future approaches based on genome sequencing of Mediterranean population samples

A recent commentary (Ozcelik *et al.*, 2010) suggests that '*sequencing a sufficient number of representative Mediterranean individuals will provide a reference and scaffold for further genetic studies in the region. Since DNA sequencing technology has advanced dramatically leading to drastic reductions in cost as well as profound improvements in efficiency over the last decade, many sequencing centers are now operational in the world including those in the Mediterranean sea basin where this capacity could be further developed and exploited to contribute to the understanding of human genetic variation in a collaborative fashion*'.

On the basis on these considerations, the authors propose to find an international collaborative Center of Excellence for Genomics Research in the Mediterranean region. They suggest that this Center of Excellence '*be decentralized and function as a network of researchers and genomics research centers whose primary remit would be to support and facilitate joint research proposals. Members of the Center would include scientists from the region and those supporting development of genomics in the region. They would engage in projects centered in Mediterranean laboratories whenever possible and involving transfer of training and technology to make the Mediterranean focus increasingly realistic with time. There is much greater strength in using resources to support science than in creating a new institution. Such a model would alleviate hurdles of bureaucracy and facilitate transnational decision-making processes. The funding programs of trans- or supra-national organizations, such as the European Commission Framework Programs, World Health Organization, International Centre for Genetic Engineering and Biotechnology and United Nations Development Fund serve as excellent models for the advancement of science and technology, and their dissemination into society. In parallel with the operational conduct of these transnational programs, national funding agencies (….) could serve as partners. These public Institutions as well as private companies will have very good scientific reasons to facilitate the generation of whole genome sequence data on representative Mediterranean populations, and gene discovery and characterization based on well-defined phenotypes in large kindreds and/or consanguineous families. A wide range of human traits, both rare and common, could be evaluated. These funding schemes and scientific objectives are excellent models and realistic goals for collaborative genomics for human health and cooperation in the Mediterranean basin, and can have major impact on public health*'.

Finally, to enable free and open access to the collected data, the authors propose that '*a virtual database named Genotheca Mediterranea could be established, with its main location within a historically important institution such as Bibliotheca Alexandrina (in Alexandria, Egypt) and with mirror sites in partner countries*'. The idea of a Genotheca Mediterranea is very much in keeping with the federation of network of LSDBs and CGDBs independently proposed by Patrinos *et al.* (2010).

All these ambitious goals can be achieved if geneticists from the southern and northern rims of the Mediterranean sea basin decide to work together for a common endeavour, namely, that of creating the virtual international collaborative Center of Excellence for Genomic

Copyright © Middle East Journal of Medical Genetics. Unauthorized reproduction of this article is prohibited.

Research in the Mediterranean, focused on consanguinity studies. The many advantages of this idea are best understood by reading the Editorial (2010) accompanying the above Commentary, which underlines that '*the main organizational element of this Center is the pair wise collaboration between investigators within the region, a structure that sets it apart from most existing research institutions. Partners outside the region would be welcomed into these collaborations, provided the emphasis was on building local capacity. Such a bottom-up approach takes advantage of the funding available to each country, while at the same time requiring only a relatively small central council to achieve agreement on aims such as the provision of genetic tests, operating procedures and priorities for centrally used resources such as benchmark genomes and databases. With this setup, Mediterranean geneticists would be able to utilize EU, US, and other initiatives and accept new investment with minimal disruption or duplication of local effort*'.

The Nature Genetics Editorial concludes that '*this proposal is an achievable and pragmatic way to boost research productivity and improve health outcomes. Geneticists' commitment to helping families living with genetic disorders is the motive behind this initiative, and the research geneticists perform is rooted in identifying disease-causing mutations, a problem they understand how to solve. The Mediterranean region is the right frame within which similar problems can be addressed and within which results can be returned locally. The proposal is efficient in that it brings together population-rich and resource-rich countries. Beyond these aspects, there may be game-changing benefits to understanding our common genetic history and overcoming common obstacles to health and progress, but ultimately the aim here is practical rather than political*'.

Acknowledgements

Conflicts of interest

There are no conflicts of interest.

References

Cavalli-Sforza LL, Moroni A, Zei G (2004). *Consanguinity, inbreeding and genetic drift in Italy*. Princeton Un; 2004.

Claustres M, Horaitis O, Vanevski M, Cotton RG (2002). Time for a unified system of mutation description and reporting: a review of locus-specific mutation databases. Genome Res 12:680–688.

Editorial (2010). Basin of attraction. Nat Genet 42:639.

Mitropoulou C, Webb AJ, Mitropoulos K, Brookes AJ, Patrinos GP (2010). Locus-specific database domain and data content analysis: evolution and content maturation towards clinical use. Hum Mutat 10:1109–1116.

Ozcelik T, Kanaan M, Avraham KB, Yannoukakos D, Megarbane A, Tadmouri GO, *et al*. (2010). Collaborative genomics for human health and cooperation in the Mediterranean region. Nat Genet 42:641–645.

Romeo G (1984). Cystic fibrosis: a single locus disease. Invited lecture to the 9th Int. Cystic Fibrosis Congress. In: Lawson D, editor. *Cystic fibrosis horizons*.J. Wiley and Sons; 1984. pp. 155–164.

Romeo G (2010). LSDBs: promise and challanges. Hum Mutat 31:V.

Romeo G, Menozzi P, Ferlini A, Fadda S, Di Donato S, Uziel G, *et al*. (1983a). Incidence of classic PKU in Italy estimated from consanguineous marriages and from neonatal screening. Clin Genet 24:339–345.

Romeo G, Menozzi P, Ferlini A, Prosperi L, Cerone R, Scalisi S, *et al*. (1983b). Incidence of Friedreich ataxia in Italy estimated from consanguineous marriages. Am J Hum Gen 35:523–529.

Romeo G, Bianco M, Devoto M, Menozzi P, Mastella G, Giunta AM, *et al*. (1985). Incidence in Italy, genetic heterogeneity and segregation analysis of cystic fibrosis. Am J Hum Genet 37:338–349.

Ten Kate LP, Teeuw M, Henneman L, Cornel MC (2010). Autosomal recessive disease in children of consanguineous parents: inferences from the proportion of compound heterozygous. J Community Genet 1: 37–40.

Van Baal S, Zlotogora J, Lagoumintzis G, Gkantouna V, Tzimas I, Poulas K, *et al*. (2010). ETHNOS: a versatile electronic tool for the development and creation of national genetic databases. Hum Genomics 4:361–368.

Copyright © Middle East Journal of Medical Genetics. Unauthorized reproduction of this article is prohibited.