RESEARCH NOTE

Quantifying genetic diversity: the starting point for population genetic studies using molecular markers

MARÍA EUGENIA BARRANDEGUY[1,2,3*] and MARÍA VICTORIA GARCÍA[1,2,3]

[1]*Departamento de Genética. Facultad de Ciencias Exactas, Químicas y Naturales. Universidad Nacional de Misiones; Posadas 3300 Misiones, Argentina*
[2]*Instituto de Biología Subtropical (UNaM – CONICET), Posadas 3300, Argentina*
[3]*Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Posadas 3300, Argentina*

Natural populations should maintain genetic diversity since it is essential for the long-term survival of species and provides the raw material for all evolutionary changes, allowing adaptation to environmental changes and thus decreasing extinction risk (Frankham *et al*. 2004). Genetic diversity represents the total genetic variation among individuals within a population. The concept of genetic markers is not a new one: in the nineteenth century, Gregor Mendel employed phenotype-based genetic markers in his experiments.

The emergence of marker systems, for the last 40 years, closely tracked developments in molecular biology. Morphological markers were largely supplanted by biochemical markers and the latest markers were supplanted by the development of markers based on DNA polymorphisms. A molecular marker is in essence a nucleotide sequence corresponding to a particular known or unknown physical location in the genome. Molecular markers and nucleotide sequence information allow uncovering of genetic diversity and also differentiation of individuals at the DNA level. A prerequisite for the beginning of population studies is to detect the genetic diversity underlying phenotypic variation. Once genetic diversity is detected its distribution within and among local populations or spatial/temporal patterns can be analysed.

Several measures of genetic diversity have been developed over the years. The seminal measure of genetic diversity from molecular data has been the number of gene alternative forms or alleles at a given locus (N_A), which is also known as gene multiplicity. As multiplicity and abundance vary independently, genetic diversity can be expressed as the effective number of alleles (N_E) (Kimura and Crow 1964). If the alleles show the same frequency, N_E will be equal to N_A. However, N_E will be a decreasing function if allelic frequency distribution is not uniform. The number of private alleles (N_P) is a related measure and represents the number of alleles which can be found in only one population (Barton and Slatkin 1986). In this way, N_P is a simple measure of genetic distinctiveness. Since the number of detected alleles in a population depends on its size, it is not advisable to compare these genetic diversity parameters among local populations with different sizes. The allelic richness estimation (R) using the rarefaction method, is useful to compare the number of alleles between samples that differ in size, because it predicts the expected number of alleles if samples have the same size (Foulley and Ollivier 2006). Forty years ago, Nei proposed the original measure of genetic diversity: Nei's gene diversity index (h) (Nei 1973). This parameter represents the probability that two alleles randomly and independently selected from a gene pool will represent different alleles. This index analyses allele frequency variation directly in terms of heterozygosity and can be applied to any population of all organisms (sexual or asexual, diploid or nondiploid) without consideration of the number of alleles at a given locus or the pattern of evolutionary forces (Nei 1973), because it has been formulated entirely in terms of alleles and genotypic frequencies in the population (Nagylaki 1998). In this way, the treatment of this index is biologically the most direct (Nagylaki 1998). Finally, the extent of DNA polymorphism for a group of nucleotide sequences sampled in a population is measured by nucleotide diversity (π), which is defined as the average number of either nucleotide differences or substitutions per site (Nei and Miller 1990).

Over the years, an evolution in the understanding and usefulness of genetic diversity quantification can be perceived. To know the evolution in these points, we surveyed the

*For correspondence. E-mail: ebarran@fceqyn.unam.edu.ar;
euge_barra@hotmail.com; vgrcia@fceqyn.unam.edu.ar.

archives of the prestigious scientific journal *Heredity* which was founded by R. A. Fisher and C. D. Darlington in 1914.

We surveyed the archives of the journal *Heredity* during 1992 to 2012 through the Scopus database using a combination of 'genetic' and 'diversity' as keywords. The search was driven through titles, abstract and keywords of research articles. In the analysed period, the search yielded 414 articles which represent around 15% of published research articles in the journal. Each article was classified according to five categories: (i) studied taxon kingdom: animal, plant or fungi; (ii) type of molecular marker: allozymes, dominant markers (RAPDs and/or AFLPs), RFLPs, microsatellites and nucleotide variation (SNPs and/or indels); (iii) title or keyword including 'genetic diversity' or 'genetic variation'; iv) scope of the study: conservation, adaptation, landscape genetics, phylogeography, classical population genetics,

breeding, evolutionary genetics and/or variability studies; and (v) estimation of Nei's gene diversity index (*h*). A binary matrix for article classification was performed considering the different analysed categories.

The classification of 414 research articles showed that 203 included animal species, 203 included plant species and eight included fungi species. From 1992 to 2000, with the exception of 1994, allozyme markers were used in 40% or more of the studies. However, all types of dominant and codominant markers were used differently between 2001 and 2004, while microsatellite markers as well as the analysis of DNA sequences won the greatest popularity during 2005–2012 (figure 1a). Studies using only one molecular marker in which estimation of genetic diversity was the main aim were developed in different taxa from 1992 to 1999. The number diminished on this kind of studies between 2000 and
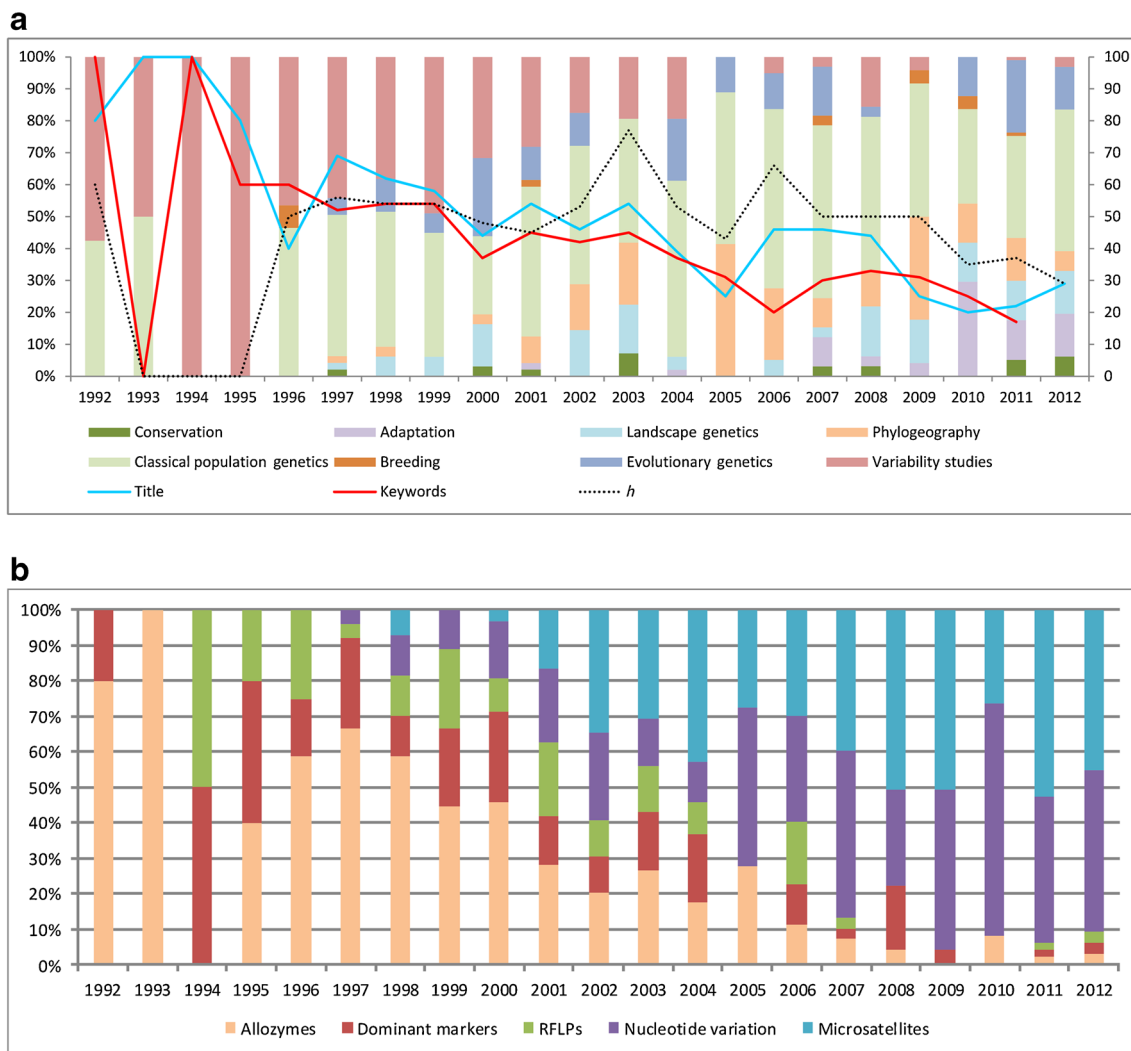

Figure 1. Classification of 414 research articles published in the journal *Heredity* between 1992 and 2012. (a) Molecular markers used in the studies expressed as percentage of total articles analysed for each year. (b) Scopes of the studies expressed as percentage of articles analysed for each year. Solid lines show percentage of articles with the words 'genetic variation' or 'genetic diversity' in either the title or keywords while the dotted line shows percentage of studies with estimation of Nei's gene diversity index.

2004, but more than one type of molecular marker was used in the same study in this period and a comparison of diversity levels detected between different markers was the main aim. At the same time, new scopes started to be discussed in addition to classic population genetic studies, i.e. evolutionary genetics, landscape genetics and phylogeography. Adaptation studies were more relevant from 2007 and variability studies almost disappeared from the scene in 2010. In this way, in this short period, several scopes coexist in the research articles published in the journal *Heredity*. The presence of the words 'genetic variation' or 'genetic diversity' started to decrease in titles and keywords from 1992 to 2012 in agreement with changes in the scopes of studies. In contrast, research using Nei's gene diversity index (h) showed frequencies higher than 45% from 1996 to 2009, decreasing to 30% in 2012 (figure 1b).

In view of these data, we can conclude that estimation of genetic diversity per se was the goal of several studies 20 years ago. However, quantification of genetic diversity is nowadays an obligatory step in population genetic studies regardless of the main aim of these studies. It shows that the understanding of genetic diversity quantification has evolved in order to answer new questions. The update of these questions is consequence of the development of scientific knowledge, environmental changes, technological advances and social demand. Studies of natural populations have recently moved away from the purely descriptive analysis of the amount of variability and genetic differentiation to more sophisticated analyses that are used to attempt to infer the demographic past (Schlötterer 2004). An example of this could be seen in Hornoy *et al*. (2013) who retraced the routes of introduction and colonization in Europe of *Ulex europaeus* analysing neutral genetic diversity. This in turn enables the inference on the evolutionary potential of populations and also to find out whether nonselective processes have contributed to the evolution of the introduced populations.

Population genetics was born mainly as a theoretical science. Protein electrophoresis produced data that could be interpreted using the language and perspective of traditional population genetic theory and it led to population genetic becoming a more explicitly empirical science. The field of empirical population genetics started to be preoccupied with characterizing and quantifying genetic variation. Now, this view is evolving and the impact of population genetics has increased considerably in almost all fields of biology. Genetic variation is a huge source of information about the biology of individuals carrying a given variant. It also keeps track of the history and spatial relationships of populations composed by these individuals. This information can be extracted using appropriate genetic markers together with appropriate statistical tools. Nowadays, this last point is the real challenge for researchers in the field of population genetics.

Large amounts of molecular data are available at present as a result of next generation sequencing methods. This huge amount of genomic data could be analysed using the body of knowledge from classical population genetics strengthened with new approaches such as the model-based methods and the laws of probability together with computational development in parallel.

There is no doubt that we are at the early stages of a modern revolution in population genetics. However, this revolution is mainly based on technological advances and we still need concepts from classical theory. If we think of the scientific studies as a race track, then genetic diversity estimation would be the end of the way in the 90s, which has become the starting point for population genetics studies today.

References

Barton N. H. and Slatkin M. 1986 A quasi-equilibrium theory of the distribution of rare alleles in a subdivided population. *Heredity* **56**, 409–415.

Foulley J. L. and Ollivier L. 2006 Estimating allelic richness and its diversity. *Livest. Sci.* **101**, 150–158.

Frankham R., Ballou J. D. and Briscoe D. A. 2004 Conservations Genetics. Cambridge University Press, Cambridge, UK.

Hornoy B., Atlan A., Roussel V., Buckley Y. M. and Tarayre M. 2013 Two colonisation stages generate two different patterns of genetic diversity within native and invasive ranges of *Ulex europaeus*. *Heredity* **111**, 355–363.

Kimura M. and Crow J. F. 1964 The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

Nagylaki T. 1998 Fixation indices in subdivided populations. *Genetics* **148**, 1325–1332.

Nei M. 1973 Analysis of gene diversity in subdivided population. *Proc. Natl. Acad. Sci. USA* **70**, 3321–3323.

Nei M. and Miller J. C. 1990 A simple method for estimating average number of nucleotidesubstitutions within and between populations from restriction data. *Genetics* **125**, 873–879.

Schlötterer C. 2004 The evolution of molecular markers – just a matter of fashion? *Nat. Rev.* **5**, 63–70.