Biochem Genet (2012) 50:670–672
DOI 10.1007/s10528-012-9509-1

NOTE

PICcalc: An Online Program to Calculate Polymorphic Information Content for Molecular Genetic Studies

Sándor Nagy · Péter Poczai · István Cernák · Ahmad Mousapour Gorji · Géza Hegedűs · János Taller

Received: 30 June 2011 / Accepted: 3 February 2012 / Published online: 10 May 2012

Introduction

Molecular markers have proved to be valuable tools in the characterization and evaluation of genetic diversity within and between species and populations. Marker systems differ in their information content, which depends on polymorphism. The concept of polymorphism is used to define genetic variation in a population, which has been extensively studied in recent years by several established scientific disciplines, for example, genetics, ecology, zoology, and microbiology (Mukherjee et al. 2010; Muneer et al. 2011; Rajkumar et al. 2011). Examples are numerous and obvious. For the practical design of molecular genetic studies, a few questions must be considered. How difficult will it be to find usefully polymorphic loci? How many markers are needed? How polymorphic must each marker be? These questions can be answered by measuring the information content of the markers. There are two measures of the quality or informativeness of a polymorphism as a genetic marker: heterozygosity (H) and polymorphic information content (PIC). Since its first application by Botstein et al. (1980) PIC has become the most widely applied

S. Nagy · G. Hegedűs
Department of Economic Methodology, Georgikon Faculty, University of Pannonia, Pethe Ferenc 7, Keszthely 8360, Hungary

P. Poczai (✉)
Plant Biology, University of Helsinki, P.O. Box 65, 00014 Helsinki, Finland
e-mail: peter.poczai@gmail.com

I. Cernák · A. M. Gorji
Potato Research Centre, Centre of Agricultural Sciences, University of Pannonia, Festetics 7, Keszthely 8360, Hungary

J. Taller
Department of Plant Science and Biotechnology, Georgikon Faculty, University of Pannonia, Festetics 7, Keszthely 8360, Hungary

formula for genetic studies to measure the information content of molecular markers. To illustrate its wide application, we surveyed DNA fingerprinting publications of the last 20 years. This search revealed that more than one thousand published papers utilized the PIC formula.

Materials and Methods

The heterozygosity of a locus is defined as the probability that an individual is heterozygous for the locus in the population (Liu 1998) and can be calculated as:

$$H = 1 - \sum_{i=1}^{l} P_i^2$$

where P_i is the frequency for the ith allele among a total of l alleles. PIC refers to the value of a marker for detecting polymorphism within a population, depending on the number of detectable alleles and the distribution of their frequency; thus, it provides an estimate of the discriminating power of the marker. The PIC value of an l-allele locus can be calculated as

$$\mathbf{PIC} = 1 - \sum_{i=1}^{l} P_i^2 - \sum_{i=1}^{l-1} \sum_{j=i+1}^{l} 2\, P_i^2\, P_j^2$$

where P_i and P_j are the population frequency of the ith and jth allele. According to Guo and Elston (1999), PIC is defined as the probability that the marker genotype of a given offspring will allow deduction, in the absence of crossing over, of which of the two marker alleles of the affected parents it received. In other words, PIC is a modification of the heterozygosity measure that subtracts from the H value an additional probability that an individual in a linkage analysis does not contribute information to the study (Speer 1999).

Results and Discussion

For the accurate design of genetic studies, such estimates must be calculated to describe the informativeness of the markers, but presently there are no easily accessible calculators for that purpose. To simplify the work of molecular studies, we have developed a useful online tool (http://w3.georgikon.hu/pic/english/default.aspx) to facilitate the calculation of H and PIC values. This program, PICcalc, can calculate these values from manually uploaded allelic frequencies or from a given file containing binary data. The latter option allows the user to calculate the values for a given number of loci from a simply prepared text file, ensuring the estimation of PIC and H for a primer or primer sets used in the analysis with different genetic marker systems dealing with binary data.

Dominant and codominant markers are routinely used in molecular genetic studies. For multilocus methods (e.g., AFLP, ISSR, RAPD), in theory, it is

presumed that fragments of equal length amplify from the corresponding loci and that they represent a single, dominant locus with two possible alleles (presence/absence). The maximum value of PIC and *H* for dominant markers is 0.5, since two alleles per locus are assumed and both are influenced by the number and frequency of the alleles (Henry 1997; De Riek et al. 2001; Bolaric et al. 2005). To consider this feature of dominant markers, a link for this calculation is implemented in the program especially for these kinds of markers.

The additional utilities made available on the Web site are for free usage, making data evaluation faster and easier. This simple online tool provides an easy way to compute PIC and *H* from binary data or from allelic frequencies.

Acknowledgments The present article was published in the frame of the projects TÁMOP-4.2.1/B-09/1/KONV-2010-0003 and TÁMOP-4.2.2/B-10/1-2010-0025. The project is realized with the support of the Hungarian Government and the European Union, with the cofunding of the European Social Fund. This work was financed by the Hungarian Grant OTKA K 76485.

References

Bolaric S, Barth S, Melchinger AE, Posselt UK (2005) Genetic diversity in European perennial ryegrass cultivars investigated with RAPD markers. Plant Breed 124:161–166

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

De Riek J, Calsyn E, Everaert I, Van Bockstaele E, De Loose M (2001) AFLP-based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theor Appl Genet 103: 1254–1265

Guo X, Elston RC (1999) Linkage informative content of polymorphic genetic markers. Hum Hered 49:112–118

Henry RJ (1997) Practical applications of plant molecular biology. Chapman and Hall, London

Liu BH (1998) Statistical genomics: linkage, mapping and QTL analysis. CRC Press, Boca Raton

Mukherjee AK, Ratha S, Dhar S, Debata AK, Acharya PK, Mandal S, Panda PC, Mahapatra AK (2010) Genetic relationships among 22 taxa of bamboo revealed by ISSR and EST-based random primers. Biochem Genet 48:1015–1025

Muneer PMA, Sivanandan R, Gopalakrishnan A, Basheer VS, Musammilu KK, Ponniah AG (2011) Development and characterization of RAPD and microsatellite markers for genetic variation analysis in the critically endangered yellow catfish *Horabagrus nigricollaris* (Teleostei: Horabagridae). Biochem Genet 49:83–95

Rajkumar S, Singh SK, Nag A, Ahuja PS (2011) Genetic structure of Indian valerian (*Valeriana jatamansi*) populations in Western Himalaya revealed by AFLP. Biochem Genet 49:674–681

Speer MJ (1999) Genetic linkage: concepts and methods. In: Albers MJ (ed) Genetics of cerebrovascular disease. Blackwell, Oxford, pp 25–26