CrossMark

GDdom: An Online Tool for Calculation of Dominant Marker Gene Diversity

Mazen Abuzayed[1] · Nourhan El-Dabba[2] ·
Anne Frary[1] · Sami Doganlar[1]

Abstract Gene diversity (GD), also called polymorphism information content, is a commonly used measure of molecular marker polymorphism. Calculation of GD for dominant markers such as AFLP, RAPD, and multilocus SSRs is valuable for researchers. To meet this need, we developed a free online computer program, GDdom, which provides easy, quick, and accurate calculation of dominant marker GD with a commonly used formula. Results are presented in tabular form for quick interpretation.

Keywords Molecular markers · PIC · Polymorphism information content

Introduction

Molecular markers are useful tools for evaluating and characterizing genetic diversity, marker-assisted selection, phylogeny studies, and genotyping of plants (Joshi et al. 1999). The usefulness of a given marker is determined by its level of polymorphism. Polymorphism for a marker locus is commonly expressed as gene diversity (GD) (Nei 1973) which is also referred to as polymorphism information content (PIC) (Shete et al. 2000). Both codominant and dominant molecular markers are used to study genetic diversity. A codominant marker identifies all the alleles at a particular locus (i.e., two alleles for a single copy marker in a diploid organism) and allows distinction between heterozygous and homozygous genotypes. In contrast, a dominant marker identifies only a single dominant allele. Despite this

✉ Sami Doganlar
samidoganlar@iyte.edu.tr

1 Department of Molecular Biology and Genetics, Izmir Institute of Technology, 35430 Urla Izmir, Turkey

2 Department of Computer Engineering, Izmir Institute of Technology, 35430 Urla Izmir, Turkey

shortcoming, dominant markers including amplified fragment length polymorphism (AFLP), random amplified polymorphic DNA (RAPD) and multilocus simple sequence repeats (SSRs) are very popular, mainly for economic reasons (Piepho and Koch 2000). Roldan-Ruiz et al. (2000) introduced a GD formula for dominant markers that has been widely used (Ling et al. 2012; Guo et al. 2014; Celik et al. 2014; Uncu et al. 2015). However, to our knowledge, no online tool for calculation of GD using Roldan-Ruiz et al. (2000) formula for dominant markers exists. Such programs are useful because manual calculation of GD may be complicated depending on the number of alleles and markers studied. Thus, the aim of this work was to construct an online program for calculation of dominant marker GD.

Materials and Methods

We used Personal Home Page (PHP) for programming and executing the calculations automatically and Hyper Text Markup Language (HTML) for designing the website form and displaying the results as tables on the website. PHP is a server-side scripting language designed for web development. It is a widely used, open source, general-purpose scripting language that is especially suited for web development and can be embedded into HTML. HTML is the standard markup language used to create web pages. We designed an algorithm that takes a CSV file as input, and its output are the tables that are displayed in the web page. Our algorithm first parses the contents of the CSV file in order to handle it as a matrix. Calculations are then done on this matrix according to the values of columns and rows. Scripts are hosted on Izmir Institute of Technology webhost under the lab URL (http://plantmolgen.iyte.edu.tr/).

Alleles for each dominant marker were scored for presence/absence and GD was calculated depending on the frequency of each allele using the formula of Roldan-Ruiz et al. (2000): $GD_i = 2f_i(1 - f_i)$, where GD_i is the gene diversity of marker 'i', f_i is the frequency of the amplified allele (band presence), and $1-f_i$ is the frequency of the null allele. According to this formula, GD for a locus can vary from 0 to 0.5. A CSV file of this binary data was then uploaded into the online GDdom program. For each marker, the numbers of missing values, total alleles, and polymorphic alleles were calculated. The average GD for each marker was also calculated with standard deviation and standard error.

Results and Discussion

GD is a measure of marker polymorphism and reveals the power of discrimination for each marker. Depending on the number of the markers and alleles under consideration, manual calculation of GD may be laborious. In this study, we have developed an online program for GD calculation (http://plantmolgen.iyte.edu.tr/GDdom/). The program summarizes all data in tabular form. Thus, the first output table describes the missing values for every sample. This allows easy determination of DNA genotypes or markers/alleles with poor amplification. The second

table shows detailed results for the number of alleles, number of polymorphic alleles, average GD value, standard deviation, and standard error. The third table gives summary statistics for the average GD value for each marker. The final table includes maximum, minimum, and average GD over all markers. The program also provides a manual with complete instructions and an example data file including present, absent and missing values.

In conclusion, we have developed this online GD calculator to help researchers analyze data for dominant markers. The online calculator saves time, avoids mistakes, and is easy to use. The program presents the results in simple tables for quick interpretation.

Compliance with ethical standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

Celik I, Gultekin V, Allmer J, Doganlar S, Frary A (2014) Development of genomic simple sequence repeat markers in opium poppy by next-generation sequencing. Mol Breed 34:323–334. doi:10.1007/s11032-014-0036-0

Guo ZH, Fu KX, Zhang XQ, Bai SQ, Fan Y, Peng Y, LK Huang, Yan YH, Liu W, Ma X (2014) Molecular insights into the genetic diversity of *Hemarthria compressa* germplasm collections native to southwest China. Molecules 19:21541–21559. doi:10.3390/molecules191221541

Joshi SP, Ranjekar PK, Gupta VS (1999) Molecular markers in plant genome analysis. Curr Sci 77:230–240

Ling Y, Zhang XQ, Ma X, Chen SY, Chen TT, Liu W (2012) Analysis of genetic diversity among wild bermudagrass germplasm from southwest China using SSR markers. Genet Mol Res 11:4598–4608. doi:10.4238/2012.October.17.5

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci 70:3321–3323. doi:10.1073/pnas.70.12.3321

Piepho HP, Koch G (2000) Codominant analysis of banding data from a dominant marker system by normal mixtures. Genetics 155:1459–1468

Roldan-Ruiz I, Dendauw J, Bockstaele EV, Depicker A, Loose MD (2000) AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.). Mol Breed 6:125–134. doi:10.1023/A:1009680614564

Shete S, Tiwari H, Elston RC (2000) On estimating the heterozygosity and polymorphism information content value. Theor Popul Biol 57:265–271. doi:10.1006/tpbi.2000.1452

Uncu AO, Gultekin V, Allmer J, Frary A, Doganlar S (2015) Genomic simple sequence repeat markers reveal patterns of genetic relatedness and diversity in Sesame. The Plant Genome 8:1–12. doi:10.3835/plantgenome2014.11.0087