THE DEMoMa: A SIMPLE TOOL FOR DETERMINATION OF EFFECTIVENESS OF MOLECULAR MARKERS

DEMoMa: PROSTE NARZĘDZIE DO OKREŚLANIA SKUTECZNOŚCI MARKERÓW MOLEKULARNYCH

Fatih HANCI

Faculty of Agriculture, Erciyes University, Türkiye

Summary: The determination of the type and number of molecular markers is significant for the results of molecular-genetic research. At this stage, the success of the marker type and number in finding useful polymorphic loci is the most important consideration. The degree of polymorphism should be evaluated by values of polymorphism information content and heterozygosity. Also, marker index, effective multiplex ratio, and resolving ratio are other measures for molecular markers. Various software is available to calculate these parameters. However, much of this software was designed in a very complex structure and required an internet connection. In addition, this software takes up a large amount of memory due to its complex processing patterns. This makes it impossible to use on some tablets or mobile phones. The other challenge is that researchers with years of training in agriculture or biotechnology find it difficult to learn to use this software. In this study, a calculation tool: the DEMoMa, based on Microsoft Excel software, which does not require an internet connection and is designed quite simply, is introduced. This tool processes numerical data obtained from molecular markers and displays the activity levels of the molecular markers used.

Keywords: DNA, heterozygosity, molecular marker; polymorphism

Streszczenie: Określenie rodzaju i liczby markerów molekularnych ma istotne znaczenie dla wyników badań molekularno-genetycznych. Na tym etapie najważniejszym czynnikiem jest powodzenie typu i liczby markerów w znalezieniu przydatnych loci polimorficznych. Stopień polimorfizmu powinien być oceniany przez wartości zawartości informacji o polimorfizmie i heterozygotyczności. Innymi miarami dla markerów molekularnych są również wskaźnik markera, efektywny stosunek multipleksów i stosunek rozdzielczy. Do obliczania tych parametrów dostępne jest różne oprogramowanie. Jednak większość tego oprogramowania została zaprojektowana w bardzo złożonej strukturze i wymagała połączenia z Internetem. Ponadto oprogramowanie to zajmuje dużą ilość pamięci ze względu na złożone wzorce przetwarzania. Uniemożliwia to korzystanie z niektórych tabletów lub

telefonów komórkowych. Innym wyzwaniem jest to, że naukowcom z wieloletnim stażem w rolnictwie lub biotechnologii trudno jest nauczyć się obsługi tego oprogramowania. W niniejszym opracowaniu wprowadzono narzędzie obliczeniowe: DEMoMa, oparte na oprogramowaniu Microsoft Excel, które nie wymaga połączenia z Internetem i zostało zaprojektowane w bardzo prosty sposób. To narzędzie przetwarza dane liczbowe uzyskane z markerów molekularnych i wyświetla poziomy aktywności użytych markerów molekularnych.

Słowa kluczowe: DNA, heterozygotyczność, marker molekularny; wielopostaciowość

INTRODUCTION

Polymorphism is defined as the differences in DNA sequence between individuals and populations. If polymorphisms are found in the regulatory regions of genes, they may cause phenotypic changes by affecting the transcriptional regulation of genes [1, 2]. It is possible to detect such polymorphisms with careful phenotypic screening. However, genetic polymorphisms that are not reflected in the phenotype can only be detected by laboratory applications such as molecular markers. Observation of the linkage between two variable loci is only possible from matches in which one parent is heterozygous (carrying two different alleles) for the marker or gene at each locus. Therefore, the potential of the marker to detect heterozygous individuals is essential in defining its efficacy. Based on this theory, markers with a large number of alleles or highly polymorphic markers tend to be quite informative [3].

Many DNA-based molecular markers have been developed for the detection of polymorphic regions on DNA such as DAF DNA (Amplification Fingerprinting), RAPD (Randomly Amplified Polymorphism), AFLP (Amplified Fragment Length Polymorphism), S-SAP (Sequence-Specific Amplification Polymorphism), SSR (Microsatellites or Simple Sequence Repeat), SCAR (Sequence Characterized Amplification Region), etc. [4]. It is possible to categorize molecular markers according to the type of information provided from a single locus such as bi-allelic dominant (RAPDs, AFLPs), the bi-allelic co-dominant (RFLPs, SSCPs), and the multi-allelic co-dominant (microsatellites) markers [3].

The determination of the type and number of molecular markers is significant for the results of molecular-genetic research. At this stage, the success of the marker type and number in finding useful polymorphic loci is the most important consideration. The amount of marker needed for this purpose, and the degree of polymorphism for each marker, are other important considerations. Defining all these success criteria is possible by making some predefined evaluations. The heterozygosity (H) and the polymorphic information content (PIC) are the two main definitions of this evaluations [5]. In addition to these definitions, Effective

multiplex ratio (EMR), Marker index (MI), and Resolving power (RP) values can also be used to evaluate the efficiency of a molecular marker [6, 7, 8].

There are several available tools to help scientists in determining the effectiveness of molecular markers such as PICcalc, I MEC, Gene-Calc [5, 9, 10]. Each of the tools is appropriate for a particular aim and restricted in other areas. The most significant limitation is that some tools are not freely available. Besides, others have a complex design and require programming skills. These considerations restrict the tool's conformity for some users. This paper introduces DEMoMa, a very simple and offline tool developed to test the effectiveness of molecular markers (dominant or co-dominant).

MATERIALS AND METHOD

DEMoMa is a tool based on Microsoft Excel software. It does not contain any additional software nor any "macro" extensions. It is sufficient to download it to a computer once for its use. It does not require any internet connection during its use. It can also be used on smartphones or tablets if the necessary sub-application is found. The DEMoMa can be downloaded from the Erciyes University website (https://avesis.erciyes.edu.tr/fatihhanci/dokumanlar). The formulas used in DEMoMa are given below. Also, the references of the formulas for each value are shared as "unprotected" on the "References" page in the tool. In this way, users can easily copy these references and paste them under the manuscript.

PIC for the co-dominant markers was calculated according to the following formula [11]:

$$\mathrm{PIC} = 1 - \sum_{i=1}^{n} P_i^2 - 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} P_i^2 P_j^2 = 1 - \sum_{i=1}^{n} P_i^2 - \left(\sum_{i=1}^{n} P_i^2\right)\left(\sum_{i=1}^{n} P_i^2\right) + \sum_{i=1}^{n} p_i^4$$

where Pi and Pj are the population frequency of the ith and jth allele.

PIC for the dominant markers was calculated according to the following formula [12]:

$$\mathrm{PIC} = 1 - \sum P_i^2$$

where "Pi" represents the frequency of the ith accession.

To calculate EMR and MI, the following formula was used [6]:

$$\mathrm{EMR} = \mathrm{p} \times \mathrm{pf},$$

where "p" is the number of polymorphic bands of a primer; pf is the frequency of polymorphic bands.

$$\mathrm{MI} = \mathrm{EMR} \times \mathrm{PIC},$$

where EMR is the effective multiplex ratio of a marker, PIC is the polymorphism information content of a primer.

Resolving power (Rp) was assessed according to following formula [13]:

$$\mathrm{Rp} = \sum \mathrm{Ib}$$

where Ib is the band informativeness with Ib = 1 – [2 x (0.5-p)] where p is the proportion of accessions containing the polymorphic band.

The user must score the results obtained from each molecular marker as "1 or 0" in the binary system before measuring the effectiveness of the marker set. Then, the user should enter into DEMoMa the number of genotypes/individuals examined, the number of different alleles obtained for each marker, the number of polymorphic alleles, and the number of "1" scores for all the alleles as in the entire gene pool. If the co-dominant marker system is used, the number of alleles from each marker must be selected at an interface. After this selection, the user is directed to the page designed according to the relevant number of alleles. In addition to the previously requested information, the frequency values for each allele must be entered on this page. The frequency values can be easily calculated by dividing the "1" scores of each allele by the total number of genotypes. The most important thing to note is that, due to the nature of co-dominant markers, the sum of frequency values should equal "1". After the completion of all these data entries, polymorphic information content, heterozygosity, marker index, resolving power and effective multiplex ratio values, which summarize the marker efficiency quantitatively, appear on the same page as four digits after the comma.

RESULTS

The analysis of the data of two sample studies conducted with both co-dominant and dominant markers with DEMoMa is performed as follows. When the Excel file is opened, the user is asked to select the marker type in the first interface. In addition, a link showing the literature source of the formulas used is also shared on this page. After the marker type is selected (dominant marker is first example), it is requested to enter the data of the study: Such as Number of genotypes examined (number of different organisms from which DNA isolation was made. 30 tomato genotypes for example); different size allele (number of bands) from the first primer used (for example, 7 different size alleles were detected); how many of the alleles from this primer are polymorphic (5 for example); and finally how many scores (112 as "1") were scored (for a total of 30 different tomato genotypes) across the entire gene pool examined from this primer. When all these data are entered, PIC, EMR, MI, and RP values appear in the same window without any other operation.

For a study done with a co-dominant marker, the "co-dominant" window on the first page should be clicked. Then, the number of different alleles obtained from the first primer pair should be selected (Assume 6 different alleles (bands) were detected in the example study). Although this number generally varies between 1-20, up to 30 options are provided for extreme examples. When the "6"

button is clicked, the page where data entry will be opened opens. In addition to the data entries mentioned in the dominant marker on this page, the frequencies of the alleles are also asked. How to calculate the frequencies of each allele is reminded here as additional information. For example, in this study, DNA samples were taken from 39 different peach genotypes, a total of 6 different alleles (bands) were detected from the first primer pair used, 5 of which were found to be polymorphic. This first primer pair produced 92 scores of "1" across the entire gene pool (39 peach genotypes). After entering these values, the frequencies of the 6 different alleles in the entire gene pool should be written in the right sections (for example, 0.200, 0.220, 0.160, 0.120, 0.140, 0.160). After all these data entries are completed, PIC, EMR, MI, and RP values are obtained.

DISCUSSION

DEMoMa is an offline tool that is designed to help researchers determine the effectiveness of molecular markers for genetic studies. DEMoMa has been designed to gaps in the conclusions in results of molecular-genetic studies by providing a simple design. As such, it will be of interest to both experienced researchers who rely on it to accelerate their studies and university educators who use it in their programs to demonstrate important concepts in molecular marker use and design. Some of the precious features in DEMoMa include: (1) It does not need an internet connection; (2) It does not need any software experience; (3) all data entries and results can be made on the same page; (4) All calculations are made on Microsoft Excel program. It has neither additional software nor a built-in macro application. For this reason, it does not have security risks for electronic devices. (5) It takes very little space on the hard disk (1.03 MB), so sharing is very simple and fast. It can be easily downloaded from both the Erciyes University website and the Journal website published as an "additional file". Thanks to these features, this simple and practical tool enables to calculation of PIC, H, MI, EMR, and RP from binary data or allele frequencies.

REFERENCES

[1] LUIKART G, ENGLAND PR, TALLMON D, JORDON S, TABERLET P. The power and promise of population genomics: from genotyping to genome typing, *Nat Rev Genet* 2003; **4:**981-994.

[2] SCHLÖTTERER C. Hitchhiking mapping-functional genomics from the population genetics perspective, *Trends Genet*, 2003; **19:**32-38.

[3] VIGNAL A, MILAN D, SANCRISTOBAL M, EGGEN A. A review on SNP and other types of molecular markers and their use in animal genetics, *Genet Sel* 2002; **34**:275-305.

[4] BUCKLER ES, THORNSBERRY JM. Plant molecular diversity and applications to genomics, *Curr Opin Plant Biol* 2002; **5**:107-110.
[5] NAGY S, POCZAI P, CERNÁK I, GORJ AM, HEGEDUS G, TALLER J**.** PICcalc: an online program to calculate polymorphic information content for molecular genetic studies, *Biochem Genet* 2012; **50**:670-672.
[6] POWELL W, MORGANTE M, ANDRE C, HANAFEY M, VOGEL J, TINGEY S**.** The utility of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis, *Mol Breed* 1996; **2**:225-238
[7] GILBERT J, LEWIS R, WILKINSON M, CALIGARI PDS. Developing an appropriate strategy to assess genetic variability in plant germplasm collections, *Theor Appl Genet* 1999; **98**:1125-1131.
[8] CHESNOKOV YV, ARTEMYEVA AM. Evaluation of the measure of polymorphism information of genetic diversity, *Agricultural Biology* 1999; **50**,571-578
[9] AMIRYOUSEFI A, HYVÖNEN J, POCZAI P. iMEC: Online marker efficiency calculator, *Appl Plant Sci* 2018; **6**:1159
[10] BIŃKOWSKI J, MIKS S. Gene-Calc [Computer software], URL www.gene-calc.pl 2018
[11] BOTSTEIN D, WHITE RL, SKOLNICK M, DAVIS RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *Am J Hum Genet* 1980*;* **32**:314-331
[12] NEI M. Analysis of gene diversity in subdivided populations, Proceedings of the National Academy of Sciences of the USA 1973; **70**, 3321-3323
[13] PREVOST A, WILKINSON MJ. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars, *Theor Appl Genet* 1999; **98**,107-112.

Edytor – Michał Nowicki

Received: 06.10.2022
Accepted: 03.11.2022
Faith Hanci
Faculty of Agriculture, Erciyes University
38280, Melikgazi, Kayseri, Türkiye
e-mail: fatihhanci@erciyes.edu.tr
phone: 905449161089